



**UNIL** | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

---

*Year : 2018*

## DATA-DRIVEN ANALYSIS AND MAPPING OF THE POTENTIAL DISTRIBUTION OF MOUNTAIN PERMAFROST

Deluigi Nicola

Deluigi Nicola, 2018, DATA-DRIVEN ANALYSIS AND MAPPING OF THE POTENTIAL  
DISTRIBUTION OF MOUNTAIN PERMAFROST

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB\_F417FD0D44072

### **Droits d'auteur**

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

### **Copyright**

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculty of Geosciences and Environment  
Institute of Earth Surface Dynamics

# DATA-DRIVEN ANALYSIS AND MAPPING OF THE POTENTIAL DISTRIBUTION OF MOUNTAIN PERMAFROST

Ph.D. thesis

presented at the

Faculty of the Geosciences and Environment  
of the University of Lausanne

by

NICOLA DELUIGI

B.Sc. & M.Sc. at the University of  
Lausanne (Switzerland)

To obtain the grade of

Ph.D. in Geosciences and Environment,  
subject area Geography

## Jury

Dr. Christophe Lambiel	Thesis director
Prof. Mikhail Kanevski	Thesis co-director
Prof. Alexander Brenning	Expert
Prof. Reynald Delaloye	Expert
Dr. Andreas Kellerer-Pirklbauer	Expert
Prof. Eric P. Verrecchia	President of the jury

Lausanne 2018



## IMPRIMATUR

Vu le rapport présenté par le jury d'examen, composé de

Président de la séance publique :	M. le Professeur Eric Verrecchia
Président du colloque :	M. le Professeur Eric Verrecchia
Directeur de thèse :	M. le Docteur Christophe Lambiel
Co-directeur de thèse:	M. le Professeur Mikhail Kanevski
Expert externe:	M. le Professeur Reynald Delaloye
Expert externe:	M. le Professeur Alexander Brenning
Expert externe:	M. le Docteur Andreas Kellerer-Pirklbauer

Le Doyen de la Faculté des géosciences et de l'environnement autorise l'impression de la thèse de

### **Monsieur Nicola DELUIGI**

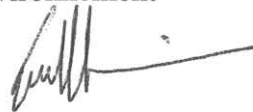
Titulaire d'une  
*Maîtrise universitaire ès Sciences en géographie mention environnements alpins  
de l'Université de Lausanne*

intitulée

### **Data-driven Analysis and Mapping of the Potential Distribution of Mountain Permafrost**

Lausanne, le 16 février 2018

Pour le Doyen de la Faculté des géosciences et de  
l'environnement



Professeur Eric Verrecchia



# Abstract

In alpine environments, mountain permafrost is defined as a thermal state of the ground and it corresponds to any lithosphere material that is at or below 0°C for at least two years. Its degradation is potentially leading to an increasing rock fall activity and sediment transfer rates. During the last 20 years, knowledge on this phenomenon has significantly improved thanks to many studies and monitoring projects, revealing an extremely discontinuous and complex spatial distribution, especially at the micro scale (scale of a specific landform; tens to several hundreds of metres).

The objective of this thesis was the systematic and detailed investigation of the potential of data-driven techniques for mountain permafrost distribution modelling. Machine learning (ML) algorithms are able to consider a greater number of parameters compared to classic approaches. Not only can permafrost distribution be modelled by using topo-climatic parameters as a proxy, but also by taking into account known field permafrost evidences. These latter were collected in a sector of the Western Swiss Alps and they were mapped from field data (thermal and geoelectrical data) and ortho-image interpretations (rock glacier inventorying). A permafrost dataset was built from these evidences and completed with environmental and morphological predictors. Data were firstly analysed with feature relevance techniques in order to identify the statistical contribution of each controlling factor and to exclude non-relevant or redundant predictors. Five classification algorithms, belonging to statistics and machine learning, were then applied to the dataset and tested: Logistic regression (LR), linear and non-linear Support Vector Machines (SVM), Multilayer perceptrons (MLP) and Random forests (RF). These techniques inferred a classification function from labelled training data (pixels of permafrost absence and presence) to predict the permafrost occurrence where this was unknown.

Classification performances, assessed with AUROC curves, ranged between 0.75 (linear SVM) and 0.88 (RF). These values are generally indicative of good model performances. Besides these statistical measures, a qualitative evaluation was performed by using field expert knowledge. Both quantitative and qualitative evaluation approaches suggested to employ the RF algorithm to obtain the best

model. As machine learning is a non-deterministic approach, an overview of the model uncertainties is also offered. It informs about the location of most uncertain sectors where further field investigations are required to be carried out to improve the reliability of permafrost maps.

RF demonstrated to be efficient for permafrost distribution modelling thanks to consistent results that are comparable to the field observations. The employment of environmental variables illustrating the micro-topography and the ground characteristics (such as curvature indices, NDVI or grain size) favoured the prediction of the permafrost distribution at the micro scale. These maps presented variations of probability of permafrost occurrence within distances of few tens of metres. In some talus slopes, for example, a lower probability of occurrence in the mid-upper part of the slope was predicted. In addition, permafrost lower limits were automatically recognized from permafrost evidences. Lastly, the high resolution of the input dataset (10 metres) allowed elaborating maps at the micro scale with a modelled permafrost spatial distribution, which was less optimistic than traditional spatial models. The permafrost prediction was indeed computed without recurring to altitude thresholds (above which permafrost may be found) and the representation of the strong discontinuity of mountain permafrost at the micro scale was better respected.

**Key words:** mountain permafrost mapping, environmental modelling, machine learning, spatial data mining.

# Résumé

Dans les environnements alpins, le pergélisol de montagne est défini comme un état thermique du sol et correspond à tout matériau de la lithosphère qui maintient une température égale ou inférieure à 0°C pendant au moins deux ans. Sa dégradation peut conduire à une activité croissante de chutes de blocs et à une augmentation des taux de transfert de sédiments. Au cours des 20 dernières années, les connaissances sur ce phénomène ont considérablement augmenté grâce à de nombreuses études et projets de suivi, qui ont révélé une distribution spatiale extrêmement discontinue et complexe du phénomène, en particulier à la micro-échelle (échelle d'une forme géomorphologique; dizaines à plusieurs centaines de mètres).

L'objectif de cette recherche était l'étude systématique et détaillée des potentialités offertes par une approche axée sur les données dans le cadre de la modélisation de la distribution du pergélisol de montagne. Les algorithmes d'apprentissage automatique (*machine learning*) sont capables de considérer un plus grand nombre de variables que les approches classiques. La distribution du pergélisol peut être modélisée non seulement en utilisant des paramètres topo-climatiques (altitude, radiation solaire, etc.), mais aussi en tenant compte de la présence et de l'absence connues du pergélisol (observations de terrain). Collectées dans un secteur des Alpes occidentales suisses, ces dernières ont été cartographiées sur la base d'investigations de terrain (données thermiques et géoélectriques), d'interprétation d'orthophotos et d'inventaires de glaciers rocheux. Un jeu de données a été construit à partir de ces évidences de terrain et complété par des prédicteurs environnementaux et morphologiques. Les données ont d'abord été analysées avec des techniques montrant la pertinence des variables permettant d'identifier la contribution statistique de chaque facteur de contrôle et d'exclure les prédicteurs non pertinents ou redondants. Cinq algorithmes de classification appartenant aux domaines des statistiques et de l'apprentissage automatique ont ensuite été appliqués et testés : Logistic regression (LR), la version linéaire et non-linéaire de Support Vector Machines (SVM), Multilayer perceptrons (MLP) et Random forests (RF). Ces techniques déduisent une fonction de classification à partir des données dites d'entraînement représentant



l'absence et la présence certaine du pergélisol. Elles permettent ensuite de prédire l'occurrence du phénomène là où elle est inconnue.

Les performances de classification, évaluées avec des courbes AUROC, variaient entre 0.75 (SVM linéaire) et 0.88 (RF). Ces valeurs sont généralement indicatives de bonnes performances. En plus de ces mesures statistiques, une évaluation qualitative a été réalisée et se base sur l'expertise géomorphologique. Les RF se sont révélées être la technique produisant le meilleur modèle. Comme l'apprentissage automatique est une approche non déterministe, il a également offert un aperçu des incertitudes de la modélisation, qui informent sur la localisation des secteurs les plus incertains dans lesquels des futures campagnes de terrain méritent d'être menées afin d'améliorer la fiabilité des cartes produites.

Finalement, RF ont démontré leur efficacité dans le cadre de la modélisation de la distribution du pergélisol grâce à des résultats comparables aux observations de terrain. L'emploi de variables environnementales illustrant la micro-topographie du relief et les caractéristiques du sol (tels que les indices de courbure, le NDVI et la granulométrie) favorise la prédiction de la distribution du pergélisol à la micro-échelle, avec des cartes présentant des variations de la probabilité d'occurrence du pergélisol sur des distances de quelques dizaines de mètres. Par exemple, dans certains éboulis, les cartes illustrent une probabilité plus faible dans la partie amont de la pente, ce qui s'avère cohérent avec les observations de terrain. La limite inférieure du pergélisol a ainsi été automatiquement reconnue à partir des évidences de terrain fournies à l'algorithme. Enfin, la haute résolution du jeu de données (10 mètres) a permis d'élaborer des cartes présentant une distribution spatiale du pergélisol moins optimiste que celle offerte par les modèles spatiaux classiques. La prédiction du pergélisol a en effet été calculée sans utiliser des seuils d'altitude (au-dessus desquels on peut trouver du pergélisol) et respecte ainsi mieux la représentation de la forte discontinuité du pergélisol de montagne à la micro-échelle.

**Mots clés:** cartographie du permafrost de montagne, modélisation environnementale, apprentissage automatique, exploration de données spatiales.

## Riassunto

Negli ambienti alpini, il permafrost di montagna è definito come uno stato termico del suolo e corrisponde a qualsiasi materiale nella litosfera che mantiene una temperatura uguale o inferiore a  $0^{\circ}\text{C}$  per almeno due anni. La sua degradazione può portare ad una crescente attività di caduta di blocchi e ad un aumento dei tassi di trasferimento dei sedimenti. Negli ultimi 20 anni, le conoscenze riguardanti il permafrost di montagna sono aumentate considerevolmente grazie ai numerosi studi e progetti di monitoraggio che hanno rivelato una distribuzione spaziale fortemente discontinua e complessa del fenomeno, in particolare alla scala della forma geomorfologica (definita come la micro scala, da decine a diverse centinaia di metri).

L'obiettivo di questa ricerca è lo studio sistematico e dettagliato delle potenzialità offerte da un approccio basato sui dati, nell'ottica di una modellizzazione della distribuzione del permafrost di montagna. Gli algoritmi di apprendimento automatico (machine learning) sono in grado di considerare più variabili rispetto agli approcci classici. La distribuzione del permafrost può essere modellizzata non solo utilizzando i parametri topo-climatici classici (altitudine, radiazione solare, ecc.), ma anche considerando esempi di presenza e assenza del permafrost (osservazioni sul campo). Raccolti in un'area delle Alpi occidentali svizzere, questi ultimi sono stati mappati sulla base di indagini di terreno (dati termici e geoelettrici), interpretazione di ortofoto e inventari di ghiacciai rocciosi. A partire dalle evidenze di terreno, è stato creato un set di dati, al quale sono stati integrati diversi predittori ambientali e morfologici. I dati sono stati dapprima analizzati con tecniche di indagine della rilevanza delle variabili; tali tecniche sono capaci di identificare il contributo statistico di ciascun fattore di controllo del permafrost e sono in grado di escludere i predittori non pertinenti o ridondanti. Sono stati, quindi, applicati e testati cinque algoritmi di classificazione appartenenti ai campi della statistica e dell'apprendimento automatico: Logistic regression (LR), la versione lineare e non lineare di Support Vector Machines (SVM), Multilayer Perceptron (MLP) e Random forest (RF). Queste tecniche deducono una funzione di classificazione dai cosiddetti dati di allenamento, che rappresentano l'assenza e la presenza certa del permafrost, e permettono in seguito di predire il fenomeno laddove è sconosciuto.

Le prestazioni di classificazione, valutate con le curve AUROC, variavano da 0.75 (SVM lineare) a 0.88 (RF). Questi valori sono generalmente indicativi di buone prestazioni. Oltre a queste misure statistiche, è stata effettuata una valutazione qualitativa. RF si è rivelata essere la tecnica che produce il modello migliore. Poiché l'apprendimento automatico è un approccio non deterministico, è stato possibile ottenere informazioni sulle incertezze della modellizzazione. Quest'ultime indicano in quali aree il modello è più incerto e, dunque, dove occorre pianificare nuove campagne di terreno per migliorare l'affidabilità delle mappe prodotte.

RF ha dimostrato la sua efficacia nella modellizzazione della distribuzione del permafrost con risultati paragonabili alle osservazioni sul campo. L'uso di variabili ambientali che illustrano la topografia e le caratteristiche del suolo (come indici di curvatura, NDVI e granulometria) aiuta a predire la distribuzione del permafrost alla micro scala, con mappe che mostrano variazioni spaziali importanti della probabilità del permafrost su distanze di poche decine di metri. In alcune falde di detrito le mappe mostrano una probabilità inferiore nella parte a monte, risultato coerente con le osservazioni sul campo. Il limite inferiore del permafrost è stato inoltre riconosciuto automaticamente dagli esempi forniti all'algoritmo. Infine, l'alta risoluzione del set di dati (10 metri) ha permesso una simulazione della distribuzione spaziale del fenomeno meno ottimistica rispetto a quella fornita dai modelli classici. La previsione del permafrost è stata, infatti, calcolata senza utilizzare delle soglie di altitudine e quindi rispetta meglio la rappresentazione dell'alta discontinuità del permafrost di montagna alla micro scala.

**Parole chiave:** cartografia del permafrost di montagna, modellizzazione ambientale, apprendimento automatico, estrazione di dati spaziali.

# Acknowledgement

Though the following manuscript is an individual work, I could never have reached the heights or explored the depths without the help, support, guidance and efforts of a lot of people. Firstly, I would like to express my sincere gratitude to my two supervisors Dr. Christophe Lambiel and Prof. Mikhail Kanevski. Christophe has believed in me since my master project, giving me the opportunity to pursue a Ph.D. on such an interesting research topic. His advices, knowledge and deep understanding of the alpine environments pushed me one step forward. He always gave me extreme freedom and confidence during the 7 years we have spent together. Prof. Mikhail Kanevski introduced me to the field of machine learning and his interest, guidance, encouragement, continuous support and constructive discussions throughout this research project were exceptional. This manuscript would not have been possible without his deep knowledge on data-driven methods and his brilliant advices.

My utmost gratitude is also extended to Prof. Reynald Delaloye for joining this project and for sharing his valuable comments during our meetings. I would also thank Reynald for having accepted to review this manuscript.

I gratefully acknowledge my two co-examiners, Prof. Alexander Brenning and Prof. Andreas Kellerer-Pirklbauer, who delightedly accepted to read and evaluate my thesis. I also thank Dr. Florence Magnin for providing the permafrost map in rock walls and for all the interesting discussions we had at ICOP and EGU meetings.

A very special thank you to Dr. Marj Tonini for her kindness, her support and for all the fruitful discussions we had in front of our numerous coffees. I would also like to acknowledge Dr. Cristian Scapozza for imparting his passion and for the excellent moments we have spent together on the field during my assistantship.

A particular thought also goes to all my closest friends and colleagues. To Jean-Baptiste Bosson and Natan Micheletti for always being two special friends before than former colleagues. Their valuable support and friendship were fundamental during the dark moments of my academic journey. I am sure that we will keep sharing a deep friendship in the years to come. To Aline Buri, Chrystelle Gabbud,

Lucien Grangier, Stephan Utz, Nico Baetz, Mario Kummert and Mariano Bonriposi for the wonderful and exciting moments I shared with them. Lastly, Elisa Giaccone and Sebastián Vivero for their friendship and the pleasant work environment that I had in our office.

I am particularly grateful for all the scientific but also the friendly support I received from my colleagues of IDYST and IGD, impossible to mention everyone, for sharing some lovely moments and coffee breaks. My gratitude also goes to Stefano Varricchio, Gabriele Leonardi and Adrien Roy for simply being great friends.

Last but not least, I would like to express my great appreciation to my family, Paola, Gabriella and Mauro, and to Simona for their continuous encouragement and love throughout!

Lausanne, February 2018

Nicola

## Financial Support

The author is grateful to the Swiss National Science Foundation for funding this project (200021-152924/2, 01/06/2014 - 31.08.2017) entitled *Analysis and modelling of the mountain ppxmafrost distribution using machine learning*.

## Credits and copyright

Deluigi, N. (2018). *Data-driven analysis and mapping of the potential distribution of mountain permafrost*. Thesis, University of Lausanne.

Posted at the University of Lausanne Open Archive (<http://serval.unil.ch>)

## Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.

## Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

## Credits

This thesis was typeset with  $\text{\LaTeX}$  2 $\epsilon$ . It uses the *Clean Thesis* style developed by Ricardo Langner. Download available at <http://cleanthesis.der-ric.de/>.



# Contents

<b>Abstract</b>	<b>x</b>
<b>Acknowledgement</b>	<b>xiv</b>
<b>List of Figures</b>	<b>xxxix</b>
<b>List of Tables</b>	<b>xxxiv</b>
<b>List of Abbreviations</b>	<b>xxxv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research context . . . . .	1
1.2 Motivation . . . . .	3
1.3 Challenges and objectives of the thesis . . . . .	5
1.4 Thesis outline . . . . .	8
<b>2 Present state of knowledge</b>	<b>9</b>
2.1 Mountain permafrost . . . . .	9
2.1.1 Permafrost controlling factors . . . . .	10
2.2 Mountain permafrost distribution . . . . .	12
2.2.1 Mountain permafrost occurrence and distribution in debris- covered surfaces . . . . .	13
2.2.2 Mountain permafrost occurrence in rock walls . . . . .	20
2.3 Permafrost distribution modelling: an overview . . . . .	21
2.3.1 First permafrost modelling attempts . . . . .	21
2.3.2 First empirical models of alpine permafrost distribution . . . . .	22
2.3.3 Empirical-statistical models of the distribution of permafrost . . . . .	23
2.3.4 Physical models of alpine permafrost distribution . . . . .	23
2.3.5 Permafrost models in rock walls . . . . .	24
2.4 Permafrost distribution maps . . . . .	25
2.4.1 Global scale . . . . .	26
2.4.2 Mountain range system scale . . . . .	27
<b>3 Materials</b>	<b>31</b>
3.1 Region of interest . . . . .	31
3.2 Mapping permafrost evidences . . . . .	34



3.3	Mountain permafrost predictors . . . . .	38
3.3.1	DEM-derived predictors . . . . .	38
3.3.2	Land-cover predictors . . . . .	41
3.3.3	Grain size . . . . .	44
3.4	Validity domain . . . . .	46
<b>4</b>	<b>Modelling approach</b>	<b>51</b>
4.1	Machine learning for environmental data analysis and modelling . .	51
4.1.1	General overview of the adopted approach . . . . .	52
4.2	Exploratory data analysis . . . . .	53
4.3	Feature relevance . . . . .	56
4.3.1	Dataset dimensionality . . . . .	56
4.3.2	Feature extraction . . . . .	57
4.3.3	Feature selection . . . . .	60
4.4	Statistical classification . . . . .	63
4.4.1	Statistical Learning Theory . . . . .	64
4.4.2	Empirical Risk Minimization . . . . .	64
4.4.3	Structural Risk Minimization . . . . .	65
4.5	Model selection and performance assessment . . . . .	66
4.6	Subset sampling strategies . . . . .	67
4.7	Selected machine learning classifiers . . . . .	72
4.7.1	Logistic regression . . . . .	72
4.7.2	Artificial neural networks . . . . .	74
4.7.3	Support Vector Machines . . . . .	76
4.7.4	Random forest . . . . .	79
4.8	Model performance assessment . . . . .	84
4.9	Used software . . . . .	86
<b>5</b>	<b>Results</b>	<b>89</b>
5.1	Visual exploratory data analysis and structure detection . . . . .	89
5.2	Statistical feature relevance . . . . .	96
5.2.1	Feature extraction . . . . .	96
5.2.2	Feature selection . . . . .	99
5.3	Subset sampling . . . . .	104
5.4	Model performances, interpretation and comparison . . . . .	106
5.4.1	Classification of permafrost data . . . . .	106
5.4.2	The effect of sampling on classification performances . . . . .	116
5.4.3	Model comparison and modelling uncertainties . . . . .	118
5.5	Permafrost mapping . . . . .	133
5.6	Synthesis . . . . .	141
<b>6</b>	<b>Discussion</b>	<b>143</b>
6.1	Permafrost data analysis and classification . . . . .	143

6.1.1	The complexity of the permafrost data . . . . .	143
6.1.2	The relevance of dimensionality reduction . . . . .	145
6.1.3	The effect of random and spatial sampling on permafrost spatial modelling . . . . .	146
6.1.4	Analysis and discussion of permafrost data classification . . .	149
6.2	Potential permafrost occurrence maps: interpretation and model comparison . . . . .	152
6.3	Advantages and limitations of the employed data-driven approach . .	176
6.4	Significance of permafrost maps and their usability . . . . .	178
6.5	Synthesis . . . . .	180
<b>7</b>	<b>Conclusion and perspectives</b>	<b>185</b>
7.1	Main achievements . . . . .	185
7.2	Research perspectives . . . . .	186
<b>A</b>	<b>Appendix</b>	<b>189</b>
<b>B</b>	<b>Published papers</b>	<b>221</b>
	<b>Bibliography</b>	<b>259</b>



# List of Figures

1.1	(A) Temperature evolution over Europe based on the CMIP3 simulations driven by the emission scenarios A2 (red), A1B (green) und B1 (blue). (B) Annual cycle of expected monthly mean change in the Alpine region of temperature (above), precipitation (below). Modified from Gobiet et al. (2014). . . . .	2
2.1	Permafrost stratigraphy and ground thermal profile (modified from Haeberli et al., 1993). . . . .	10
2.2	The Tsarmine sector (Valais Canton), a typical periglacial landscape located above the PLL. Various glacial and periglacial landforms are visible: (A) talus slope (B) rock glacier (C) rockslide (D) morainic complex (E) glacier (F) debris-covered glacier (3D map: Windows Maps ©). . . . .	13
2.3	Rock glacier controlling factors and morphological characteristics (structure) (Les Cliosses rock glacier, Hérens valley, Valais Canton; drone photo: S. Rüttimann). . . . .	15
2.4	Schematic yearly thermal regimes of a ventilated talus slope (modified from Morard et al., 2010). . . . .	17
2.5	Snow melting timelapse at the Petit Mont Rouge talus slope (Fontanesses sector, Valais Canton; UNIL webcam images). (A) October 2012: first snow (B) December 2012: thickening of the snow cover (C) February 2013: avalanches thickening the snow cover downslope (D) April 2013: appearance of melting chimneys (E) May 2013: melting windows in the upper part of the talus slope (F) June 2013: distal part starts to melt when the upper part is snow free (G) July 2013: last avalanche deposits start to melt (H) From August 2013: snow free period. . . . .	17
2.6	Glacier forefield of the Ritord and Epée glaciers (Valais Canton, Switzerland. (a & b) Marginal rock glaciers, (c) buried ice, (d, e & f) push moraines, (g) debris covered glacier and (h) talus rock glaciers (modified from Delaloye, 2004; orthophoto: swisstopo ©). . . . .	19
2.7	Modelling approach and representation of the subsurface thermal field of a mountainous massif, illustrating the differences between northern and southern exposed slopes. Example of the Matterhorn, Canton of Valais (modified from Noetzli et al., 2007). . . . .	20

2.8	MARST map and permafrost index in rock walls (Chamonix region, France; modified from Magnin et al., 2015). . . . .	25
2.9	Maps illustrating the distribution of permafrost of the North hemisphere (left), Antarctica (center) published by the CALM network and its evolution into the GPZI (right). Modified from Brown et al. (1997) and Gruber (2012). . . . .	26
2.10	Potential permafrost distribution map published by Swiss Federal Office of Environment (BAFU) and covering the Swiss Alps. Zoom illustrates the region of interest of the present thesis (modified from BAFU, 2005; hillshade: swisstopo). . . . .	28
2.11	Extract of the Alpine permafrost index map (APIM) covering the Swiss territory (modified from Boeckli et al., 2012). . . . .	28
2.12	Overview of recent maps covering the Chilean Andes (A1, A2, A3; modified from Azócar et al., 2016), the Argentinian Andes (B; modified from Angillieri, 2017) and the Maritime Southern Alps of New Zealand (C; modified from Sattler et al., 2016). . . . .	29
3.1	Situation map and extent of the region of interest (topographic map: swisstopo). . . . .	31
3.2	Extent of the region of interest and localization of (a) the Mont Fort-Mont Gelé area, (b) the Fontanesses area, (c) the Veisivi ridge and (d) the Tsaté-Moiry sector (topographic map: swisstopo). . . . .	32
3.3	The Yettes Condja active rock glaciers (Mont Gelé/Mont Fort sector; left image) and their mapping by ortho-image interpretation (right image; orthophoto: swisstopo). . . . .	35
3.4	The rock glaciers of the Tsarmine sector (left) and their mapping by ortho-image interpretation (center; orthophoto: swisstopo). Right image illustrates the InSAR signal between 23/07/2009 and 08/10/2009. . . . .	35
3.5	Survey of the Attelas talus slope (Mont Gelé-Mont Fort sector). (A) Overview of the talus slope and localization of the ERT profiles and boreholes (B) 3D view of the ERT profiles, illustrating the ground electrical resistivity (C/D) ERT cables and measuring instrument (Syscal Pro), injecting DC electrical current into the ground between one pair of electrodes and measuring the voltage between another pair (E) Interpretation and mapping of the ERT results, coupled with thermal data (boreholes and data loggers) (F) Overview of the mapped permafrost evidences. . . . .	36
3.6	Extent of the study area and localization of known (presumed) permafrost evidences (hillshaded map: swisstopo). . . . .	37
3.7	Terrain curvature types and associated landforms for Les Cliosses sector, with an active rock glacier indicating the presence of permafrost (Hérens Valley, Valais; photo: S. Rüttimann). . . . .	40

3.8	Orthophoto - primary surface map comparison for the Mont Gelé sector (orthophoto, topographic and primary surface maps: swisstopo). . . .	42
3.9	NDVI map of the Mont Gelé sector and delimitation of the vegetation extent (orthophoto, FCIR images and topographic map: swisstopo). . .	43
3.10	(A) Comparison of grain sizes for two sectors with coarse material on a rock glacier (B) and finer material at its roots (C). . . . .	44
3.11	(A) Orthophoto of the Tsarmine sector (Arolla), with various landforms and debris-covered glacier (swisstopo). (B) Gray-scaled image of the sector (25 centimetres of resolution; swisstopo). (C) Computed grain size of the sector, obtained by applying a standard deviation filter on the grey-scaled image. . . . .	46
3.12	The extent of the validity domain, delimiting the surfaces where the prediction of the permafrost occurrence was performed (hillshade: swisstopo). . . . .	48
4.1	The modelling workflow of this study, from data collection to statistical prediction. . . . .	53
4.2	Overview of the modelling steps: the raw dataset, containing permafrost evidences and environmental predictors, was analysed and resampled. Its dimensionality was then reduced before applying classification algorithms and producing predictions of the permafrost occurrence. . . .	54
4.3	Three permafrost observations in 2D can be shattered with a linear decision function. The presented combination of 4 observations cannot be shattered and it requires a classification in a different dimension (modified after Kanevski et al., 2009). . . . .	65
4.4	Variation of the bound on risk according to the empirical risk and the confidence interval for an increasing complexity of models (modified after Vapnik, 1998). . . . .	66
4.5	Underfit or overfit of a training dataset. If the model poorly fits (left image) or perfectly fits (right image) the training data, it will have poor generalization ability when applying to new data (modified after Kanevski et al., 2009). . . . .	67
4.6	Overview of the sampling strategy, coupling clustering and spatial/random cross-validation. . . . .	68
4.7	Splitting of the cleaned dataset: SOM mapping allows detecting clusters data that are then split to obtain representative sub-datasets. . . . .	69
4.8	Example of the architecture of a Self-organizing map, with the representation of the input layer connected to the competition layer by weighted vectors (modified after Carrasco Kind and Brunner, 2014). . . . .	69
4.9	SOM structure (left) and update of Best Matching Unit (BMU, right; modified after Carrasco Kind and Brunner, 2014). . . . .	70

4.10	Example of a heat map illustrating the structures in the topological 2-D representation of the data (generated with the R <i>kohonen</i> package; Wehrens, Buydens, et al., 2007). . . . .	71
4.11	Comparison between linear regression (A) and logistic regression (C) of observed probabilities $y$ on the independent variable $x$ . (D) Linear regression of the observed log odds $y$ on the independent variable $y$ . . . . .	73
4.12	An example of a feedforward neural network: multilayer perceptron with input neurons (in yellow), two hidden layers made by hidden neurons (in green) and two output neurons (in red). Connections between neurons represent the $w_{ij}$ weights (modified after Kanevski et al., 2009). . . . .	75
4.13	The largest margin hyper plane is an intuitively simple linear model with good generalization. Closest training examples to the hyperplane are called <i>training vectors</i> (modified after Kanevski et al., 2009). . . . .	77
4.14	Slack variables $\xi_i$ assigned to noisy observation lying outside their class margin (modified after Kanevski et al., 2009). . . . .	78
4.15	The application of a kernel allows the mapping of the original data into a feature space where a non-linear problem becomes linear and can be easily solved (modified after Kanevski et al., 2009). . . . .	78
4.16	A schematic tree classifiers presenting a root, two split nodes and its terminal leaf nodes. . . . .	80
4.17	Development of a decision tree providing the classification of a simple permafrost data example. . . . .	80
4.18	Development of an ID3 tree based on the highest IG splitting (entropy or gini). . . . .	82
4.19	Development of a forest of decision trees whose prediction is computed by majority vote. . . . .	83
4.20	Schematic representation of the test and the possible prediction of a binary classification. . . . .	85
4.21	The Receiver Operating Characteristics curve, as an indicator of the quality of the classification. (A) corresponds to the perfect result (no errors, AUROC = 1), (B) indicates a typical ROC curve (AUROC=0.85) and (C) refers to an inefficient classification (AUROC = 0.5). . . . .	86
5.1	Histograms of the distribution of the permafrost evidences for each of the retained environmental predictors and relative number of pixels (cell count). For a description of the variable acronyms, see table 3.1. . . . .	90
5.2	Distribution of standardized values of all the employed environmental predictors (for both positive and negative permafrost evidence). Values outside whiskers are considered as noisy data and were consequently excluded from the dataset. For a description of the variable acronyms, see table 3.1. . . . .	92

5.3	Plot of the PCA principal components illustrating the percentage of explained variance. . . . .	93
5.4	Parallel coordinate plot of the permafrost data (median is represented with a bold line, while 25 and 75 percent quartiles are shown by a dashed line). . . . .	93
5.5	2-D manifold of the permafrost data computed with t-SNE, allowing visually unravelling the presence of structures in the dataset. . . . .	94
5.6	Error curves of an AGRNN in regards to selected sigma values. . . . .	96
5.7	Parallel coordinate plots of the synthetic datasets obtained with kernel PCA, Sammon maps and LLE. . . . .	98
5.8	Information Gain variable importance, with features ranked according to their IG weights. . . . .	99
5.9	Variable importance provided by RELIEFF, with features ranked according to their weights. . . . .	100
5.10	Correlation-based feature selection for the assessment of variable importance. . . . .	100
5.11	Variable importance measures embedded in the Random forest algorithm. . . . .	101
5.12	Feature ranking resulting from a SVM recursive feature elimination procedure. . . . .	101
5.13	Correlation matrix for all environmental variables. Pie plots indicate how positively or negatively correlated each pair of prediction is. Blue pies refer to positively correlated predictors, whilst red pies indicate negatively correlated pairs. . . . .	103
5.14	SOM U-matrix and hits map of the permafrost input space. . . . .	105
5.15	The majority vote for the assessment of the optimal number of clusters (left); clustering of the SOM neurons according to the NBClust result. . . . .	105
5.16	Difference in the selection of training and test observations according to the random (A) or the spatial sampling strategy (B) for the Yettes-Condja sector (Mont Gelé area). . . . .	106
5.17	Box-plots of the odds ratios for retained features. . . . .	108
5.18	LR ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.87) and spatially sampled data (right image, AUROC: 0.84). . . . .	109
5.19	Linear SVM ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.78) and spatially sampled data (right image, AUROC: 0.75). . . . .	110
5.20	RBF-SVM ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.79) and spatially sampled data (right image, AUROC: 0.77). . . . .	112
5.21	MLP network topology with one hidden layer and 9 non-linear nodes. . . . .	113



5.22	MLP ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.91) and spatially sampled data (right image, AUROC: 0.84). . . . .	113
5.23	OOB curve used to assess the number of trees required for the dataset (applied to a randomly sampled subset). . . . .	114
5.24	RF ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.97) and spatially sampled data (right image, AUROC: 0.88). . . . .	115
5.25	Comparison of the AUROC values obtained with the employed ML algorithms for randomly and spatially sampled data sub-sets. . . . .	117
5.26	Uncertainty plots for Logistic regression classification on randomly sampled sub-sets. . . . .	120
5.27	Uncertainty plots for Logistic regression classification on spatial sampled sub-sets. . . . .	121
5.28	Uncertainty plots for linear SVM classification on randomly sampled sub-sets. . . . .	123
5.29	Uncertainty plots for linear SVM classification on spatial sampled sub-sets.	124
5.30	Uncertainty plots for Radial basis function SVM classification on randomly sampled sub-sets. . . . .	126
5.31	Uncertainty plots for Radial basis function SVM classification on spatial sampled sub-sets. . . . .	127
5.32	Uncertainty plots for Multilayer perceptron classification on randomly sampled sub-sets. . . . .	128
5.33	Uncertainty plots for Multilayer perceptron classification on spatial sampled sub-sets. . . . .	129
5.34	Uncertainty plots for Random forest classification on randomly sampled sub-sets. . . . .	131
5.35	Uncertainty plots for Random forest classification on spatial sampled sub-sets. . . . .	132
5.36	Potential permafrost distribution map obtained with Logistic regression (random sampling) and corresponding uncertainties. . . . .	134
5.37	Potential permafrost distribution map obtained with Logistic regression (spatial sampling) and corresponding uncertainties. . . . .	135
5.38	Potential permafrost distribution map obtained with Linear SVM (random sampling) and corresponding uncertainties. . . . .	136
5.39	Potential permafrost distribution map obtained with Linear SVM (spatial sampling) and corresponding uncertainties. . . . .	136
5.40	Potential permafrost distribution map obtained with RBF-SVM (random sampling) and corresponding uncertainties. . . . .	137
5.41	Potential permafrost distribution map obtained with RBF-SVM (spatial sampling) and corresponding uncertainties. . . . .	137

5.42	Potential permafrost distribution map obtained with Multilayer perceptron (random sampling) and corresponding uncertainties. . . . .	138
5.43	Potential permafrost distribution map obtained with Multilayer perceptron (spatial sampling) and corresponding uncertainties. . . . .	139
5.44	Potential permafrost distribution map obtained with Random forest (random sampling) and corresponding uncertainties. . . . .	140
5.45	Potential permafrost distribution map obtained with Random forest (spatial sampling) and corresponding uncertainties. . . . .	140
6.1	Differences of the prediction of permafrost occurrence according to a random sampling (A) and spatial sampling (B) strategy, with (a) and (b) representing two sectors with different permafrost probability of occurrence. . . . .	149
6.2	Comparison of the probability of permafrost occurrence for the talus slope of the Palanche de la Cretta. Maps carried out with LR (A), linear SVM (B), RBF-SVM (C), MLP (D) and RF (E). The swisstopo topographic map is shown in (F) for an overview of the ground truth. . . . .	151
6.3	(A) Presentation of the Mont Gelé sector: (a) <i>Lapires</i> talus slope (b) <i>Lues Rares</i> rock glacier (c) <i>Lac des Vaux</i> periglacial lobes (d) <i>Attelas</i> talus slope (e) <i>Yettes Condja</i> rock glaciers; (B) Potential permafrost occurrence map for the Mont Gelé area, based on the RF classification and localization of the the <i>Yettes Condja</i> sector (S-1), the <i>Les Attelas/Lac des Vaux</i> sectors (S-2) and the <i>Lapires/Lues Rares</i> sites (S-3). . . . .	153
6.4	(A) <i>S-1</i> sector, characterized by three active rock glaciers on the East side of the Mont Gelé peak and (B) PERMAL probability of permafrost occurrence based on RF classification. (A, B, C & F) Rock glaciers; (D & E) Coarse debris-covered surfaces. . . . .	154
6.5	(A) the Mont Gelé western slope (S-2), with the localization of the <i>Lac des Vaux</i> sector (a) and the <i>Les Attelas</i> talus slope (b). (B) PERMAL map of the permafrost extent and localization of the <i>Lac des Vaux</i> periglacial lobes A, B and C (see i.e. Scapozza, 2015a). The white arrows indicate a difference in permafrost probabilities between the mid-upper part and the mid-lower part of the <i>Attelas</i> talus slope. . . . .	155
6.6	Presentation of the S-3 sector with the localization of the (a) <i>Lapires</i> talus slope, (b) the <i>Lues Rares</i> and (c) the <i>Chassoure</i> rock glaciers (left image). Right image shows the probabilities of permafrost occurrence predicted by PERMAL and (A) the localisation of a protalus rampart on the top part of the <i>Lapires</i> . (B) Portion of the landform characterized by grater probabilities in its mid-lower part. . . . .	156

6.7	Qualitative indication of the spatial extent of permafrost occurrence according to Delaloye (2004), Lambiel et al. (2016) and Scapozza (2013), as well as the location of boreholes and GST loggers and an illustration of the coarse grain size of the Lapires talus slope. Figure reproduced from Staub et al. (2015). . . . .	157
6.8	(A) PERMAL map of the permafrost extent of the Mont Gelé sector; (B) Uncertainty map for the Mont Gelé sector, illustrating the probability variability obtained during the RF classification. (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005). (a) shows the altitude thresholds used by this empirical-statistical approach for the delimitation of the permafrost categories indicated in the BAFU legend. . . . .	158
6.9	(A) The Bec des Rosses sector: (a) La Chaux glacier forefield (b) the Gentianes moraine (c) talus slope and two inventoried rock glaciers (d) rock glacier (e) Patiéfray talus slope; (B) Potential permafrost occurrence map for the Bec des Rosses area, based on the RF classification, and delimitation of two sectors presented in figures 6.10 and 6.11. . . .	159
6.10	PERMAL map of the northern sector of the Bec des Rosses. (a) La Chaux glacier forefield (b) Gentianes moraine (c) Patiéfray talus slope. . . . .	160
6.11	PERMAL map of the southern sector of the Bec des Rosses peak; (a) talus slopes and La Perraire rock glacier 1, 2 & 3 (b) Pierrier d’Aget rock glaciers inventoried by Lambiel (1999). . . . .	160
6.12	(A) PERMAL map of the Bec des Rosses area; (B) Uncertainty map of the RF classification; (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005). . . . .	161
6.13	The Tournelon Blanc glacier sector: (a) Becca de la Lia rock walls (b) Tournelon Blanc rock walls (c & d) talus slopes (e) torrential cones; (B) Potential permafrost occurrence map for the Tournelon Blanc glacier area, based on the RF classification. . . . .	162
6.14	(A) PERMAL map of the Tournelon Blanc glacier area; (B) Uncertainty map of the RF classification; (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005). . . . .	164
6.15	(A) The Petit Mont Rouge and Tsena Réfien area: (a) the Petit Mont Rouge talus slope, with the presence of a small protalus rampart (b) relict rock glacier (c) the Tsena Réfien talus slope (d) rock walls; (B) Potential permafrost occurrence map for the Petit Mont Rouge and Tsena Réfien sector, based on the RF classification, and delimitation of the sectors presented in figures 6.16 and 6.17. . . . .	165
6.16	The Tsena Réfien talus slope with (a) a permafrost body identified by Lambiel (2006), (b) two vegetated rock glacier fronts and (c) the mid-upper part of the landform where permafrost is absent. . . . .	166

6.17	(A) Qualitative indication of the spatial extent of permafrost occurrence of the Petit Mont Rouge talus slope (reproduced from Scapozza, 2013); (B) PERMAL probabilities of permafrost occurrence and localization of (a) the protalus rampart and (b) the portion of the talus slope where permafrost absence was assessed with field data and observations (Scapozza, 2013). . . . .	167
6.18	(A) PERMAL map of the Tsena Réfien-Petit Mont Rouge area; (B) Uncertainty map of the RF classification; (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005). . . . .	168
6.19	(A) The Veisivi ridge: (a) rock glaciers (b) Tsarmine debris-covered glacier (c) Tsarmine talus slope with protalus rampart (d) Tsarmine rock glacier (e) rock glaciers (f) Perroc talus slope; (B) Potential permafrost occurrence map for the Veisivi area, based on the RF classification. . .	169
6.20	PERMAL map for the Perroc talus slope sector (F) and localization of the interpolation map of BTS measurements presented in figure 6.21 (dashed lines), of four active rock glaciers (B, C, D & E) and a relict rock glacier (A). . . . .	170
6.21	Interpolation map of BTS measurements of (A) the Perroc talus slope and (B) the Tsarmine talus slope (reproduced from Lambiel, 2006). . .	170
6.22	PERMAL probabilities of occurrence for the Tsarmine talus slope sector (a), a small protalus rampart (b), the Tsarmine rock glacier (c) and another active rock glaciers (d). Dashed lines indicates the location of the interpolation map of BTS measurements presented in figure 6.21. The limits of the protalus rampart and the two rock glaciers are reported on the orthophoto (top-left image). . . . .	171
6.23	PERMAL probabilities of occurrence for the Tsarmine debris-covered glacier, with a marginal rock glacier (a) that develops from the glacier system and an active rock glacier located on the northern slope. The limits of the rock glaciers are reported on the orthophoto (top-left image).172	172
6.24	(A) PERMAL map of the Veisivi ridge; (B) Uncertainty map of the RF classification; (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005) and localization of the Tsarmine talus slope (a) and the Tsarmine rock glacier (b). . . . .	173
6.25	(A) Presentation of the Tsaté-Moiry sector; (B) Potential permafrost occurrence map for the Tsaté-Moiry area, based on the RF classification, and delimitation of two intact rock glaciers. . . . .	174
6.26	(A) Potential permafrost occurrence map for the Tsaté-Moiry area, based on RF classification; (C) RF model uncertainty map; (D) Potential permafrost extent (BAFU, 2005). . . . .	175

A.1	Summary statistics for a the Logistic regression classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	190
A.2	Summary statistics for a the Logistic regression classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	191
A.3	Summary statistics for a the linear SVM classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	192
A.4	Summary statistics for a the linear SVM classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	193
A.5	Summary statistics for a the Radial basis function SVM classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	194
A.6	Summary statistics for a the Radial basis function SVM classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	195
A.7	Summary statistics for a the Multilayer perceptron classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	196

A.8	Summary statistics for a the Multilayer perceptron classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	197
A.9	Summary statistics for a the Random forest classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	198
A.10	Summary statistics for a the Random forest classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index). . . . .	199
A.11	Potential permafrost extent obtained with LR classification on randomly sampled data. . . . .	200
A.12	Potential permafrost extent obtained with LR classification on spatially sampled data. . . . .	201
A.13	Uncertainty map obtained with LR classification on randomly sampled data. . . . .	202
A.14	Uncertainty map extent obtained with LR classification on spatially sampled data. . . . .	203
A.15	Potential permafrost extent obtained with linear SVM classification on randomly sampled data. . . . .	204
A.16	Potential permafrost extent obtained with linear SVM classification on spatially sampled data. . . . .	205
A.17	Uncertainty map obtained with linear SVM classification on randomly sampled data. . . . .	206
A.18	Uncertainty map extent obtained with linear SVM classification on spatially sampled data. . . . .	207
A.19	Potential permafrost extent obtained with RBF-SVM classification on randomly sampled data. . . . .	208
A.20	Potential permafrost extent obtained with RBF-SVM classification on spatially sampled data. . . . .	209
A.21	Uncertainty map obtained with RBF-SVM classification on randomly sampled data. . . . .	210
A.22	Uncertainty map extent obtained RBF-SVM classification on spatially sampled data. . . . .	211

A.23	Potential permafrost extent obtained with MLP classification on randomly sampled data. . . . .	212
A.24	Potential permafrost extent obtained with MLP classification on spatially sampled data. . . . .	213
A.25	Uncertainty map obtained with MLP classification on randomly sampled data. . . . .	214
A.26	Uncertainty map extent obtained MLP classification on spatially sampled data. . . . .	215
A.27	Potential permafrost extent obtained with RF classification on randomly sampled data. . . . .	216
A.28	Potential permafrost extent obtained with RF classification on spatially sampled data. . . . .	217
A.29	Uncertainty map obtained with RF classification on randomly sampled data. . . . .	218
A.30	Uncertainty map extent obtained RF classification on spatially sampled data. . . . .	219
B.1	Permafrost lower limits in rockwalls, resulting from PERMOS (2009) field measurements. . . . .	225
B.2	Linear separable samples in 2-D (A) and 3-D (B) features space. . . . .	225
B.3	binary classification of two groups of samples by maximizing the margin $\rho$ and allowing misclassifications. . . . .	226
B.4	RBF mapping allowing non-linear classification in a linear space to be solved. . . . .	227
B.5	Labeled samples (pixels) used by Support Vector Machines (SVMs) for the simulation. . . . .	229
B.6	Support Vector Machines decision function results. Pixel values require a probabilistic transformation to be interpreted. . . . .	231
B.7	Distribution of permafrost occurrence's probabilities resulting from the Platt sigmoid function. . . . .	232
B.8	The final potential alpine permafrost distribution map for the entire Rosablanche topographic map area. . . . .	232
B.9	PERMAL simulation (A) versus the BAFU one (B) for the Mont Gelé sector. Refer to text for the meaning of the small letters. . . . .	233
B.10	PERMAL simulation (A) versus the BAFU one (B) for the Darbonneire and Tsauderys glacier circuses sector. Refer to text for the meaning of the small letters. . . . .	234
B.11	PERMAL simulation (A) versus the BAFU one (B) for the Pic d'Artsinol sector. For the significance of the small letters see the text. The highlighted zone corresponds to the zoom in Fig. B.12. . . . .	235
B.12	Rock glacier limits (A) and PERMAL simulation (B) in the northern part of the Pic d'Artsinol. . . . .	235

B.13	(A) Extent of the study area (relief map) and localization of known permafrost evidences. (B) Typical Alpine periglacial landscape characterized by active rock glaciers, debris-covered glacier, talus slopes, moraine deposits and rock walls (Arolla valley, Valais; photo: R. Delaloye)	244
B.14	Correlation plot of selected environmental variables within the validity domain. . . . .	245
B.15	Terrain curvature types and associated landforms for Les Cliosses sector, with an active rock glacier indicating the presence of permafrost (Hérens Valley, Valais; photo: S. Rüttimann). . . . .	246
B.16	Classification confusion matrix (left) and evaluation rates (bottom) used to assess the model quality. The Receiver Operating Characteristics curve also indicates the quality of the classification: (A) corresponds to the perfect result (no errors, AUROC = 1), (B) indicates a typical ROC curve (AUROC=0.85) and (C) refers to an inefficient classification (AUROC = 0.5). . . . .	252
B.17	OBB error as a function of the number of trees. . . . .	254
B.18	Permafrost distribution maps obtained with (A) Logistic regression, (B) Random Forests and (C) Support Vector Machines. (D) Ground-truth of the study area (aerial images from swisstopo) and LIA glacier forefield extent for ease of interpretation. Red dashed square indicates the location the overviews of figure B.20. . . . .	255
B.19	ROC curves for classification by Logistic regression (0.807, left), Support Vector Machines (0.848, center) and Random forests (0.884, right). . .	256
B.20	Potential permafrost distribution in the Fontanesses sector (Arolla valley) obtained with (A) Logistic regression, (B) Random Forests and (C) Support Vector Machines. . . . .	257





# List of Tables

2.1	Intact and relict rock glacier characteristics permit to interpret the potential permafrost occurrence. . . . .	14
2.2	Topo-climatic model of the PLL (possible or probable) for different terrain orientations and locations (modified from Haeberli, 1996). . . . .	22
3.1	Features contained in the raw dataset and their abbreviation, type and use. Categorical variables served at constructing the validity domain (VD). Continuous variables were used for classifying the permafrost occurrence. . . . .	47
4.1	Classification confusion matrix for binary predictions of the permafrost occurrence. . . . .	84
5.1	Portion of territory covered by training data, validity domain and region of interest in pixels and km <sup>2</sup> . . . . .	89
5.2	Summary statistics for a linear SVM and a RBF-SVM, classifying the raw permafrost dataset: precision (positive predictive value), recall (sensitivity), mean squared error (MSE) and area under the ROC curve (AUROC). . . . .	95
5.3	Linear and RBF-SVM AUCs for raw and synthetic data. . . . .	97
5.4	Feature merit. . . . .	102
5.5	List of the 15 environmental predictors retained in the clean dataset, according to their relevance computed with feature relevance techniques. The table also shows the type, as well as the used abbreviations of each variable. . . . .	103
5.6	Odds ratios in the respect of the <i>absence</i> class label. . . . .	107
5.7	Quality measures of the LR classification on randomly and spatially sampled data. . . . .	108
5.8	Linear SVM feature weights and their position with respect to the margin (positive or negative class). . . . .	109
5.9	Quality measures of the linear SVM classification on randomly and spatially sampled data. . . . .	110
5.10	Quality measures of the RBF-SVM classification on randomly and spatially sampled data. . . . .	111

5.11	Quality measures of the MLP classification on randomly and spatially sampled data. . . . .	113
5.12	Quality measures of the RF classification on randomly and spatially sampled data. . . . .	115
5.13	Comparison of the AUROC values obtained with the employed ML algorithms for randomly and spatially sampled data. . . . .	116
5.14	Performances and paired t-test outcomes (v: significantly better, *: significantly worse) for models obtained with random sampling. . . . .	118
5.15	Performances and paired t-test outcomes (v: significantly better, *: significantly worse) for models obtained with spatial sampling. . . . .	118
5.16	Parts of the validity domain predicted as <i>permafrost very likely absent</i> (0-0.25), <i>permafrost likely absent</i> (0.25-0.5), <i>permafrost likely present</i> (0.5-0.75) and <i>permafrost very likely present</i> (0.75-1) by Logistic regression (% of total grid cells). . . . .	135
5.17	Parts of the validity domain predicted as <i>permafrost very likely absent</i> (0-0.25), <i>permafrost likely absent</i> (0.25-0.5), <i>permafrost likely present</i> (0.5-0.75) and <i>permafrost very likely present</i> (0.75-1) by linear SVM (% of total grid cells). . . . .	137
5.18	Parts of the validity domain predicted as <i>permafrost very likely absent</i> (0-0.25), <i>permafrost likely absent</i> (0.25-0.5), <i>permafrost likely present</i> (0.5-0.75) and <i>permafrost very likely present</i> (0.75-1) by Radial basis function SVM (% of total grid cells). . . . .	138
5.19	Parts of the validity domain predicted as <i>permafrost very likely absent</i> (0-0.25), <i>permafrost likely absent</i> (0.25-0.5), <i>permafrost likely present</i> (0.5-0.75) and <i>permafrost very likely present</i> (0.75-1) by Multilayer perceptron (% of total grid cells). . . . .	139
5.20	Parts of the validity domain predicted as <i>permafrost very likely absent</i> (0-0.25), <i>permafrost likely absent</i> (0.25-0.5), <i>permafrost likely present</i> (0.5-0.75) and <i>permafrost very likely present</i> (0.75-1) by Random forest (% of total grid cells). . . . .	141
B.1	The list of features used for the permafrost distribution modelling in sedimentary landforms. . . . .	228
B.2	Features contained in the raw dataset. Categorical variables serve constructing the validity domain (VD), continuous variables are used for classifying the permafrost occurrence. . . . .	247
B.3	Summary statistics of the Logistic regression, Support Vector Machines and Random forests models. . . . .	253
B.4	Model performances and paired t-test outcomes (v: significantly better, *: significantly worse). . . . .	256

# List of Abbreviations

<b>ADS</b>	.....	Airborne digital sensor
<b>ANN</b>	.....	Artificial neural network
<b>APIM</b>	.....	Alpine permafrost index map
<b>BAFU</b>	.....	Swiss Federal Office of Environment
<b>BMU</b>	.....	Best matching unit
<b>BTS</b>	.....	Bottom temperature of snow cover
<b>CALM</b>	.....	Circumpolar Active Layer Monitoring Network
<b>DEM</b>	.....	Digital elevation model
<b>EDA</b>	.....	Exploratory data analysis
<b>ERT</b>	.....	Electrical resistivity tomography
<b>ESDA</b>	.....	Exploratory spatial data analysis
<b>FCIR</b>	.....	False-color infra-red
<b>GAM</b>	.....	Generalized additive models
<b>GLM</b>	.....	Generalized linear models
<b>GRNN</b>	.....	General regression neural network
<b>InSAR</b>	.....	Interferometric synthetic aperture radar
<b>LIA</b>	.....	Little Ice Age
<b>MAAT</b>	.....	Mean annual air temperature
<b>MAGST</b>	.....	Mean annual ground surface temperature
<b>MAP</b>	.....	Mean annual precipitation
<b>MAPT</b>	.....	Mean annual permafrost temperature
<b>MARST</b>	.....	Mean annual rock surface temperature
<b>ML</b>	.....	Machine learning
<b>MLP</b>	.....	Multilayer perceptron
<b>NDVI</b>	.....	Normalized difference vegetation index
<b>PACE</b>	.....	Permafrost and Climate in Europe
<b>PERMOS</b>	.....	Swiss permafrost monitoring network
<b>PLL</b>	.....	Permafrost lower limit
<b>PNN</b>	.....	Probabilistic neural network
<b>PISR</b>	.....	Potential incoming solar radiation

<b>RBF</b>	Radial basis function
<b>SAVI</b>	Soil-adjusted vegetation index
<b>SLF</b>	Institute for Snow and Avalanche Research
<b>SOM</b>	Self-organising maps
<b>SVM</b>	Support Vector Machines
<b>VD</b>	Validity domain
<b>WGMS</b>	World Glacier Monitoring Service
<b>AGRNN</b>	Anisotropic general regression neural network
<b>AUROC</b>	Area under the receiver operating characteristics curve
<b>CFS</b>	Correlation-based feature selection
<b>FN</b>	False negative
<b>FNR</b>	False negative rate
<b>FP</b>	False positive
<b>FPR</b>	False positive rate
<b>FR</b>	Feature reduction
<b>FS</b>	Feature selection
<b>ID</b>	Intrinsic dimensionality
<b>ID3</b>	Iterative dichotomiser 3
<b>IG</b>	Information gain
<b>kPCA</b>	Kernel principal component analysis
<b>LDA</b>	Linear discriminant analysis
<b>LLE</b>	Locally linear embedding
<b>LR</b>	Logistic regression
<b>OOB</b>	Out-of-bag error estimate
<b>PC</b>	Parallel coordinates
<b>PCA</b>	Principal component analysis
<b>PERMAL</b>	PERmafrost modelling with MACHine Learning
<b>RBF-SVM</b>	Radial basis function kernel Support Vector Machines
<b>RF</b>	Random forest
<b>RFE</b>	Recursive feature elimination
<b>RFVI</b>	Random forest variable importance
<b>ROC</b>	Receiver operating characteristics curve
<b>SLT</b>	Statistical learning theory
<b>SRM</b>	Structural Risk Minimization
<b>SVM-RFE</b>	Support Vector Machines recursive feature elimination
<b>TN</b>	True negative
<b>TNR</b>	True negative rate
<b>TP</b>	True positive
<b>TPR</b>	True positive rate
<b>t-SNE</b>	t-distributed stochastic neighbour embedding





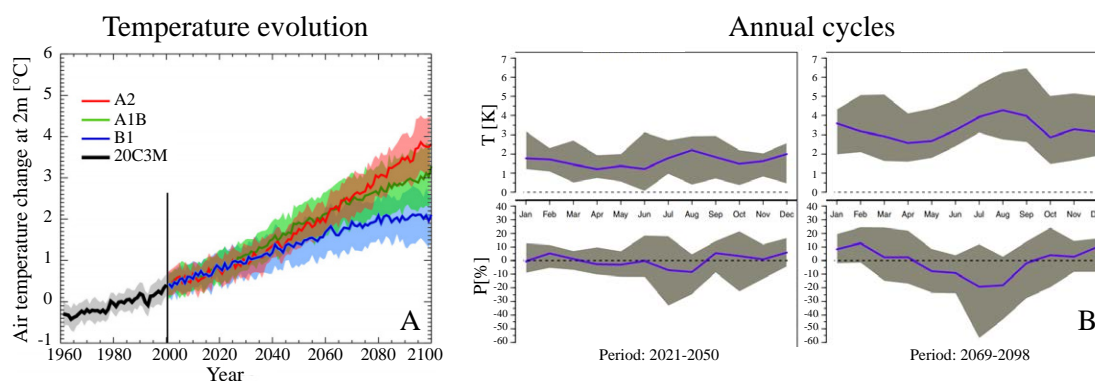
# Introduction

## 1.1 Research context

Assessing the effects of climate change is one of the main challenges of the 21<sup>st</sup> century. Climatic models predict a significant increase in temperatures and precipitations at the global scale, strongly moving away from the values that have characterized the last century (IPCC, 2014, figure 1.1.A). The energy balance and hydrological cycle will be particularly affected by these modifications. Although the increase in the average temperature of the atmosphere is a phenomenon that produces significant changes across the entire planet, they are particularly important for the middle and high latitudes of the Northern hemisphere (IPCC, 2001; Gobiet et al., 2014; EEA, 2017). The cryosphere, which is strongly present at these latitudes and it represents one of the major components of the hydrological cycle, will undergo significant transformations over the next few decades. Mountain environments of the European Alps will be affected as much as the Polar regions (Pepin et al., 2015). For the Swiss Alps, Scherrer et al. (2005) showed a significant decrease in snow precipitation at low altitudes, also attributed to the global warming. This trend is consistent with the increase in temperatures measured at the global and regional level that occurred in the last 150 years (Gobiet et al., 2014; Huss et al., 2017; Beniston et al., 2017a, figure 1.1.A), causing a retreat of alpine glaciers, as evidenced by Zemp et al. (2008) and Fischer et al. (2015). The World Glacier Monitoring Service (WGMS) unravelled that, from 1850 onwards, the majority of the mountain glaciers of the world have undergone considerable retreat or, in some cases, they have even completely disappeared (WGMS, 2005; Zemp et al., 2015).

Analogous correspondence was observed for mountain permafrost (Harris et al., 2003; Gruber and Haeberli, 2009; Etzelmüller, 2013). This phenomenon is defined as a thermal state of the ground corresponding to any lithosphere material that is at or below 0°C for at least two years (Harris et al., 2009). The degradation of alpine permafrost concerns as much the unconsolidated debris materials as the bedrock. Borehole measurements indicate that mountain permafrost is no longer in equilibrium with the climatic conditions, although the response of permafrost to climate change is slightly offset from the warming of the ground surface (Harris et al., 2009; PERMOS, 2016; Salzmann and Gärtner-Roer, 2017). Changing the thermal characteristics of mountain permafrost may indeed lead to a change in the





**Fig. 1.1.:** (A) Temperature evolution over Europe based on the CMIP3 simulations driven by the emission scenarios A2 (red), A1B (green) und B1 (blue). (B) Annual cycle of expected monthly mean change in the Alpine region of temperature (above), precipitation (below). Modified from Gobiet et al. (2014).

mechanical properties of the ground, which may in turn change the frequency, the magnitude and even the location of mass wasting phenomena initiating from the periglacial belt (Harris et al., 2009). In addition, a correlation between a decrease in the stability of frozen rock walls and an increase in temperature due to the melting of the ice that fills and consolidates the cracks in the bedrock has been revealed (i.e. Gruber and Haeberli, 2007; Ravanel et al., 2010). It was observed that a variation in temperature may lead to a weathering by freeze-thaw and to slope failures (Hasler et al., 2012; Kellerer-Pirklbauer, 2017). Data also show that the current permafrost degradation is expressed by a thickening of the active layer and by a warming of the permafrost body (PERMOS, 2016). The increase in rock glacier velocities due to the ground warming has already been measured (Kääb et al., 2007; Delaloye et al., 2008; Roer et al., 2008). The increase of the loose sedimentary stock related to higher rock fall activity above the permafrost lower limit may also enhance ground erodability (Huggel et al., 2010). Furthermore, glacier retreat is exposing large areas of loose materials (Paul et al., 2007; Carrivick and Heckmann, 2017), which may be affected by paraglacial adjustment (Geilhausen et al., 2012; Kellerer-Pirklbauer and Kaufmann, 2017; Zanoner et al., 2017). Sustained melting of ice- and snow-covered sectors will have large impacts on the hydrological cycle (Kobierska et al., 2011). It follows that, locally, the torrential dynamics could change over the next decades because of a more important sedimentary charge transportation (Lane et al., 2007). Lastly, ground cohesion, which increases in permafrost terrain due to particle cementation by ice, strongly decreases when the temperature is rising close to 0°C (Davies et al., 2001; Krautblatter et al., 2013).

To be able to deal with the mentioned permafrost-related hazards, a good representation of the distribution of this phenomenon is nowadays mandatory. Our ability to measure the environment has significantly increased in recent years, helped by technological developments in different disciplines, such as geomorphology, geotech-

nical engineering, geophysics, glaciology, geographical information science, statistics and data mining (Fotheringham et al., 2000; Otto et al., 2017). A rapid growth of computational resources has increased the expectation that the scientific knowledge can be a driver for intelligent decision-making and mountain landscape settlement (Cross and Moscardini, 1985). Accordingly, spatial distribution maps leading to decision guidelines have to be made available for the concerned regions in the near future.

## 1.2 Motivation

Several efforts have been made to improve the knowledge on the potential distribution of alpine permafrost, particularly in mountain regions where infrastructures have been built for tourism, communication, energy supply, etc. (e.g. Delaloye et al., 2001; Scapozza et al., 2011; Deline et al., 2015; Ravelle et al., 2017). The investigations conducted in periglacial environments have unravelled that mountain permafrost is characterised by a high spatial variability since the steep mountain topography has an effect on the strong spatial heterogeneity of the permafrost controlling factors (e.g. Lambiel and Pieracci, 2008; Scapozza et al., 2011; Scapozza, 2015a). Consequently, the extent, the climatic sensitivity and the evolution of the alpine permafrost are less known in comparison to the arctic permafrost (i.e. Gruber and Haeberli, 2009; Etzelmüller, 2013). The controlling factors of the spatial distribution of alpine permafrost can be summarized into three categories, following the spatial scale of reference: *climate*, *topography* and *ground conditions* (Gruber, 2005; Gruber and Haeberli, 2009). These categories are *hierarchical* and not *exclusive*, meaning that at the lower scales, the climatic conditions are less influencing than local parameters but are still important (Sattler, 2016). Permafrost distribution is closely related to air temperature at the regional scale (i.e. the scale of the Alps). In turn, at the local scale, namely the scales ranging from a valley side up to a valley, it depends on the influence of topography on the potential incoming of solar radiation (Hoelzle, 1992). At the scale of a specific landform (defined as the *micro scale*; covering tens to several hundreds of metres), the thermal properties of the surface and sub-surface material control the heat transfer into the ground and, consequently, the permafrost occurrence. At this scale, mountain permafrost is particularly likely present in coarse debris because of the voids between blocks, which make the cooling of the ground possible during the year (Hanson and Hoelzle, 2004). In talus slopes, this phenomenon is enhanced when a ventilation effect occurs (e.g. Delaloye and Lambiel, 2005; Lambiel and Pieracci, 2008). The effect of the scale on controlling factors is also valid for the negative permafrost occurrence. Indeed, this means that large debris-covered areas have been found unfrozen in spite of theoretically favourable topo-climatic conditions, as a consequence of very local parameters such

as material grain size, moisture, snow thickness, etc. (e.g. Reynard et al., 2003; Otto and Sass, 2006; Scapozza et al., 2011).

The permafrost occurrence in rock walls has been successfully modelled in recent years (e.g. Gruber et al., 2004b; Noetzli et al., 2007; Magnin et al., 2015). Compared to what happens on debris-covered surfaces, *climate* and *topography* have a higher influence on the rock temperature than the *ground conditions*. Indeed, steep rock walls are not covered by debris or thick snow layers and thus the rock temperature of rock walls is less dependent on the ground parameters than those controlling the permafrost occurrence in debris materials (Mittaz et al., 2000; Hoelzle et al., 2001; Gruber et al., 2003).

A clear representation on how the phenomenon is spatially distributed in debris-covered surfaces is however still difficult to obtain. The ability of modelling the spatial distribution of mountain permafrost has raised the interest of the scientific community during the last two decades. The first empirical-statistical models were based on simple approaches (such as linear regression) and offered a good overview of the potential distribution of mountain permafrost at the regional scale (i.e. Hoelzle, 1994; Ebohon and Schrott, 2008; Avian and Kellerer-Pirklbauer, 2012, see chapter 2.3). These models are generally thresholding the occurrence of permafrost on the basis of a restricted number of topographical and climatic parameters (i.e. altitude of rock glacier fronts for a given orientation) and they are validated with measurements of the ground surface temperature. The availability of an increasing amount of high resolution data (generally derived from high resolution digital elevation models) opened then the way to new complex statistical models able to deal with a large number of predictors (i.e. Boeckli et al., 2012; Schöner et al., 2012; Azócar et al., 2016; Sattler et al., 2016). However, although they generally offer a good overview of the permafrost distribution at the local scale (i.e. scale of a valley side), these models do not often reflect the great heterogeneity of the phenomenon at the micro scale. As for the Swiss Alps for example, the official permafrost map of the Swiss Alps (BAFU, 2005) and the Alpine permafrost map index (APIM), published by Boeckli et al. (2011), inform on the potential extent of the phenomenon at the regional and local scales, but do not take into account very local complexities proper to the micro scale.

This thesis is a contribution to the research field of the spatial permafrost modelling and responds to the needs of an improved prediction of the permafrost extent at the *micro scale*. By employing classification algorithms belonging to *statistics* and *machine learning* on representative permafrost evidences, this research investigates whether a *data-driven* approach is *able to model the distribution of a so heterogeneous phenomenon* and thus to map its strong spatial discontinuity. Obtaining a reliable model at the micro scale would be important for both scientific interest as well as

the application of this knowledge to concrete problems like the prevention against natural hazards.

### 1.3 Challenges and objectives of the thesis

This research deals with *the analysis of the spatial distribution of mountain permafrost*. Up to now, process-oriented permafrost models have become increasingly complex and have involved many parameters (i.e. Chadburn et al., 2015; Marmy et al., 2015; Wicky and Hauck, 2017). Such models are advanced and physically correct. However, their performance is based on site-specific data that go into the model and thus, they may not be suitable to derive the spatial distribution of permafrost. For these reasons, current permafrost spatial models are simpler than process-oriented models, performing well at the *regional* and *local* scales. At the *micro* scale, specific investigations are always necessary to assess the quality of the result. This gap is due to the complex (non-linear) interrelationship between the atmospheric climatic components (turbulent heat fluxes and radiation balance, the terrain surface and the subsurface properties (topographic parameters, snow cover, vegetation, type of debris/bedrock, soil water, etc.)).

This research directly aims at filling this gap. The potential of a *data-driven* approach for mapping the spatial distribution of alpine permafrost at the micro scale was already investigated in two preliminary studies (Deluigi, 2012; Deluigi and Lambiel, 2013). They unveiled that an exhaustive analysis of available data on mountain permafrost distribution and of their characteristics was necessary and it would provide detailed knowledge on the influence of permafrost controlling factors on its occurrence. Accordingly, four main challenges are identified to develop statistical models, simulating the mountain permafrost occurrence by using a *machine learning approach*. These challenges aims at:

1. investigating and modelling the spatial distribution of mountain permafrost *quantitatively*, without any "a priori" user assumption concerning the relationship between permafrost and its explaining variables;
2. responding to the lack of existing models able to integrate and predict the spatial heterogeneity of mountain permafrost at the micro scale, which represents a case of a *non-linear phenomenon*, as mountain permafrost occurrence does not just increase linearly with altitude at that scale;
3. taking into account the dimensionality of the permafrost distribution space (*N-dimensional geo-feature space*), which is composed of many different control factors;
4. considering the fundamental relevance of the *multi-scale component* of this distribution, increasing the chances to produce an optimal model that effi-

ciently predicts alpine permafrost occurrence both at the regional scale and at the micro-topography scale.

Data-driven algorithms are able to provide distribution maps by learning the relationship between permafrost and environmental predictors proper to both the regional and the micro scale. With the development of a *cutting-edge methodology* for permafrost spatial data analysis and modelling based on machine learning techniques, *new distribution models can be reliable at the regional scale* (e.g. alpine scale), *while maintaining high resolutions* (usefulness at the micro scale) *and being calibrated on local data*, not necessarily collected in the same study area.

Accordingly, the main objective of the present research aims at a systematic and detailed investigation of the potential that a machine learning approach can provide to this cryospheric research field. The main goal of the study is the development of a new modelling methodology for mountain permafrost distribution in mineral surfaces (that include both bedrock outcrops and debris surfaces, other than rock walls) by using machine learning algorithms.

More precisely, the study focuses on the following objectives:

1. to develop a cutting-edge methodology for the spatial analysis, modelling and visualisation of permafrost data by using techniques belonging to machine learning, spatial statistics and geomatics (GIS, remote sensing);
2. to perform a comprehensive analysis of the available data related to the permafrost phenomenon and to learn of their spatial characteristics and features at the micro scale;
3. to provide a statistical identification of which variable plays a major role for the permafrost distribution in order to use it to predict the occurrence of permafrost for unknown sites;
4. to employ multiple data-driven methods to test different techniques and then provide a modelling procedure applicable for other studies;
5. to map the mountain permafrost distribution at the micro scale on the basis of different simulations obtained with employed machine learning tools.

Further in this thesis, the acronym *PERMAL*, standing for *PERmafrost modelling with MACHine Learning* (as it was proposed in Deluigi and Lambiel, 2013), will be used and will refer to the permafrost maps that are obtained using the developed data-driven methodology. For the accomplishment of mentioned objectives, five main tasks are defined:

1. data quality assessment, pre-processing of data and creation of a comprehensive library of variables explaining permafrost presence or absence;
2. exploratory data analysis (EDA) and exploratory spatial data analysis (ESDA);

3. adaptation and application of feature extraction and feature selection contemporary algorithms;
4. application of machine learning classification algorithms;
5. validation and testing of retained models.

The accomplishment of these tasks will require permafrost evidences to be gathered from different sources, such as rock glacier inventories, geophysical data and ground (surface) temperature data. The translation of punctual data into spatial data will be carried out by using interpolation and/or manual delimitation. From this, *maps of the permafrost presence and absence* will provide *positive and negative training examples* to machine learning algorithms. Successively, EDA and ESDA will be performed. The first is an approach of data analysis that carries out a deep understanding of the dataset information. This approach employs different techniques that allow underlying structures to be revealed and detecting the presence of anomalies or outliers into the data. It also unveils the relationships among the explanatory predictors that allow using the selection of appropriate modelling algorithms. The adopted techniques are mostly graphical (histograms, box plots, mean plots, standard deviation plots of raw data). Additional techniques will also be needed for detecting spatial patterns in data and understanding their spatial proprieties (ESDA). By finding structures (e.g. clusters) in large and unnecessarily homogeneous spatial datasets, ESDA will provide the knowledge to assess and employ the correct spatial modelling approach to be adopted successively.

When dealing with large set of high-dimensional data, one of the major problems is the number of input features (predictors) involved. Consequently, the analysis of a large amount of raw data is not always an easy task in terms of computational difficulties. Applying classification algorithms to a large number of variables may also lead to the risk of overfitting, with the consequence of a poor generalization of the produced models. Feature extraction and feature selection techniques help avoid these issues by simplifying the amount of predictors required to the final modelling. On the one hand, this process leads to an improved knowledge about the features used and their relation with the studied phenomenon. On the other hand, taking away irrelevant features from the data provides an effectively improvement of the machine learning classification.

For the mountain permafrost case, the main task of the machine learning is the classification of observations corresponding to two binary classes (*absence* and *presence* of permafrost) with corresponding data modelling uncertainties. Machine learning algorithms usually have the embedded ability to transform the binary result into a probabilistic output, which is easier to interpret. Algorithms should supply to accurate decision boundaries, taking into account non-linear solutions and misclassifications of observed data samples. Techniques providing these requirements and considering the complexity of the permafrost distribution (linked to a high

amount of variables, which lead to a high dimensional feature space of factors related to the permafrost occurrence) are necessary. Nowadays, many different well-known machine learning algorithms exist. In this research, the most widely used algorithms for environmental data mapping are employed: Logistic regression, Support Vector Machines, Multilayer perceptrons and Random forest (see chapter 4.7 for further details). They are known to be efficient and robust methods to solve classification problems. For each of the mentioned machine learning algorithms, permafrost distribution maps will be produced. Although statistical validation and tests are part of the machine learning approach, model results will be compared with published permafrost evidences and presented in the final part of this thesis.

## 1.4 Thesis outline

This thesis is organized in seven chapters. This first part (Chapter 1) introduced the scope of the study and the main challenges that the current state of the research has to overcome. It is followed by Chapter 2, that presents the scientific background on mountain permafrost and the existing approaches developed to model and map this phenomenon. Chapter 3 focuses on the data and the region of interest. It presents the study area, as well as permafrost evidences and environmental variables that were selected to explain and predict the potential distribution of mountain permafrost. Chapter 4 focuses on the modelling approach and it is divided in two parts. The first one gives an overview on the main techniques that were employed to explore permafrost data, followed by a synthesis on the main algorithms used for the analysis of the variable relevance. The second part of the chapter illustrates the main characteristics of the adopted machine learning techniques and how they are evaluated. The results are presented in Chapter 5, followed by their interpretation and discussion (Chapter 6). The best performing permafrost map is analysed and compared in this part of the thesis. Finally, Chapter 7 states the main conclusions.







## Present state of knowledge

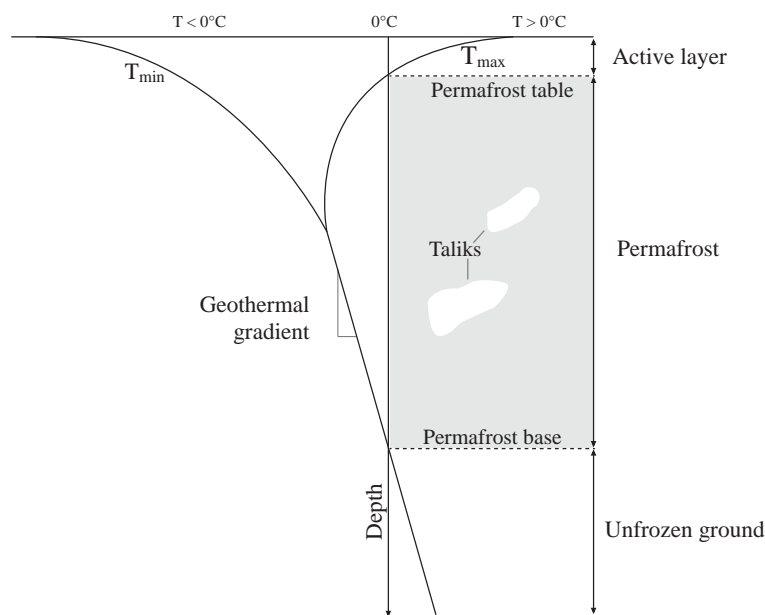
### 2.1 Mountain permafrost

Permafrost is defined as a subsurface material whose temperature never goes above 0°C for at least two years (Washburn, 1979; Harris et al., 2009). This definition refers to a thermal state of the ground, meaning that the phenomenon is independent from the existence of water (or ice) in the substrate. In the presence of a porous soil (movable sediments, cracked rocks), infiltrating water can however turn into ice due to constant negative temperatures. Depending on the area covered by frozen ground, permafrost is distinguished in *continuous* and *discontinuous* permafrost. The former refers to regions where the surface is completely covered by permafrost, while the latter indicates areas where permafrost only covers parts of the territory of interest (Harris et al., 1988).

*Mountain permafrost* (or alpine permafrost) is the term used to denominate this cold thermal state of the ground in mountain environments. Mountain permafrost is different from Arctic permafrost (located at high latitudes) due to a strong spatial variability and it is linked to the spatial heterogeneity of its controlling factors. These parameters are strongly influenced by the mountain topography (Gruber and Haeberli, 2009; Etzelmüller, 2013).

Permafrost *stratigraphy* presents the succession of subsequent layers shown in figure 2.1 (Haeberli et al., 1993). The *active layer*, which is located between the ground surface and the *permafrost table*, is characterized by a thaw and a refreezing during spring and autumn seasons. This layer is strongly subjected to energy exchanges between the atmosphere and the subsurface at depths and it normally varies between 0.5 and 8 metres in temperate mountain chains like the Alps (Gruber and Haeberli, 2009). The *body of permafrost* is in turn characterized by a permanently frozen layer. Its temperature remains at or below 0°C for consecutive years. In the permafrost layer it is possible to find unfrozen layers named *taliks*, as well as massive ice bodies. The presence of massive ice leads to the definition of three types of permafrost: the *sub-saturated permafrost*, when only a few voids are completely frozen; the *saturated permafrost*, when the ice fills all the pores and the *super-saturated permafrost*, corresponding to a soil in which the ice occupies a volume greater than that of the voids. Generally, the geothermal gradient controls the thickness of the permafrost body. In mountain environments, lateral heat fluxes may however have

an influence on the permafrost thickness (Noetzli et al., 2007), that can vary from few metres to tens of metres in fine materials, to several hundred of metres in bedrock (Noetzli and Gruber, 2005; PERMOS, 2016). Nevertheless, a discrepancy does nowadays exist between the thickness of mountain permafrost and the actual climatic conditions, with occurrences of permafrost observed at low altitudes as witnesses of past climatic conditions (from the Little Ice Age; see Etzelmüller et al., 2001). The permafrost body is limited by the so-called *base of the permafrost*. From this threshold, the increase in temperature due to the geothermal gradient no longer allows the maintenance of negative temperatures in the ground.



**Fig. 2.1.:** Permafrost stratigraphy and ground thermal profile (modified from Haeberli et al., 1993).

### 2.1.1 Permafrost controlling factors

In mountainous regions, the spatial distribution of permafrost depends on three main categories of controlling factors, that are *climate*, *topography* and *ground conditions* (Gruber and Haeberli, 2009; Sattler et al., 2016). Their importance and effect on the permafrost extent vary according to the scale of reference to whom the permafrost occurrences are considered. Indeed, at larger scales (i.e. global to continental scales), the existence of permafrost is mainly related to the energy fluxes that reach the planet (French, 2013). These fluxes are responsible for the delimitation of the bioclimatic regions of the Earth and, accordingly, they have an influence on the mean annual air temperature (MAAT). The spatial extent of permafrost is strongly linked to the decrease in temperature with the elevation (Hoelzle et al., 2001). At a regional

scale (i.e. the scale of the Alps), this link reveals to be particularly important and it is considered as one of the main controlling factor of the phenomenon distribution.

At the local scale, corresponding for instance to the scale of a catchment or of a valley side, Hoelzle et al. (2001) indicate that the MAAT alone is no longer able to control the distribution of permafrost. The *permafrost lower limit* (PLL) is a theoretical threshold below which the discontinuous permafrost is generally not observed (Haeberli, 1985; Barsch, 1996). It depends on local topographic settings, such as the terrain slope and the aspect, the presence of rock walls, etc. These parameters strongly influence the solar radiation received by the ground (Gruber et al., 2004a; Lewkowicz and Ednie, 2004), resulting in aspect-dependent variations of the permafrost extent. The presence of permafrost is indeed less probable in sun-exposed slopes than in shaded areas and thus, the PLL varies accordingly. In addition to the terrain aspect, the potential incoming solar radiation (PISR) received by the ground can be minimized by the presence of a snow layer (Hanson and Hoelzle, 2004). Indeed, the latter can insulate and deaden the energy transfer that generally occurs between the ground surface and the atmosphere (see Keller, 1994; Hoelzle et al., 2001; Hanson and Hoelzle, 2004).

At the micro scale, referring to the scale of a specific landform (tens to several hundreds of metres), three filters (the so-called *thermal offsets*), the snow cover, the ground surface characteristics and the active layer characteristics (see Hoelzle et al., 2001; Delaloye, 2004) influence the air exchanges between the atmosphere and the permafrost table. As shown above for the local scale, the snow cover has an influence on the PISR reaching the ground surface. Additionally, at the micro scale, the air content of a thick snow cover lowers its thermal conductivity and thus protects the ground from the atmospheric variations that occur during the winter season. The snow cover that melts in late spring or in early summer protects the ground from the warm temperatures that occur in this period of the year, reducing the duration of the energy exchanges during the hottest months. Similarly, the time of the first snow fall has a strong influence on ground surface temperatures, especially when occurring in late summer or early autumn. The sub-surface thermal state is not going to cool down and warmer temperatures are maintained in the ground during the cold season. Accordingly, the snow cover has a dual cooling/warming effect on the mean annual ground temperature (MAGT), depending on factors, such as thickness, duration, redistribution by wind and avalanches and timing (Hoelzle et al., 2001; Hanson and Hoelzle, 2004).

The ground surface characteristics (vegetation, grain size, albedo, etc.) represent the second thermal offset that can influence the way the radiative exchanges operate between the atmosphere and the ground (Gruber et al., 2003). The vegetation influence on permafrost occurrence has been poorly investigated in the literature. It

is commonly accepted that alpine permafrost is generally absent in sectors covered by vegetation. This observation was firstly proposed by Haerberli (1975) and Haerberli (1996) and it was included in his *rule-of-thumb* guidelines for permafrost presence (see section 2.3.1). Other authors studied the interaction between vegetation and permafrost/rock glacier dynamics (see Johnson, 1966, Furrer and Hutter, 1994, Cannone and Gerdol, 2003, Burga et al., 2004, Colombo et al., 2016). They particularly focused on the disturbances of the permafrost creep on the plant growth. The main findings were obtained by observing the vegetation growth in comparison to the decrease of rock glacier dynamics. As a general rule, permafrost does not influence the vegetation. The latter usually grows on debris-covered surfaces with fine grain size materials, which are often unfavourable to permafrost because their high water content acts as a source of latent heat, hindering the effective cooling of the sub-surface (e.g. Hoelzle et al., 1999; Heggem et al., 2005). As a consequence, vegetation has a marginal effect on the occurrence of alpine permafrost, which is mostly influenced by the grain size of the soil rather than the vegetation-cover.

Lastly, the active layer characteristics (porosity, moisture, thickness, etc.) constitute the third filter between the ground surface and the permafrost table (Burn and Smith, 1988; Hoelzle et al., 2001). The heat transfer into the ground is controlled by the thermal properties of the materials at the surface and sub-surface. Coarse debris are favourable to permafrost because of the presence of air that fills the voids and enhances the cooling of the ground during the whole year (e.g. Hanson and Hoelzle, 2004; Scapozza et al., 2011). In contrast, as mentioned above, fine sediments are unfavourable to permafrost because of the potentially higher moisture compared to coarse debris.

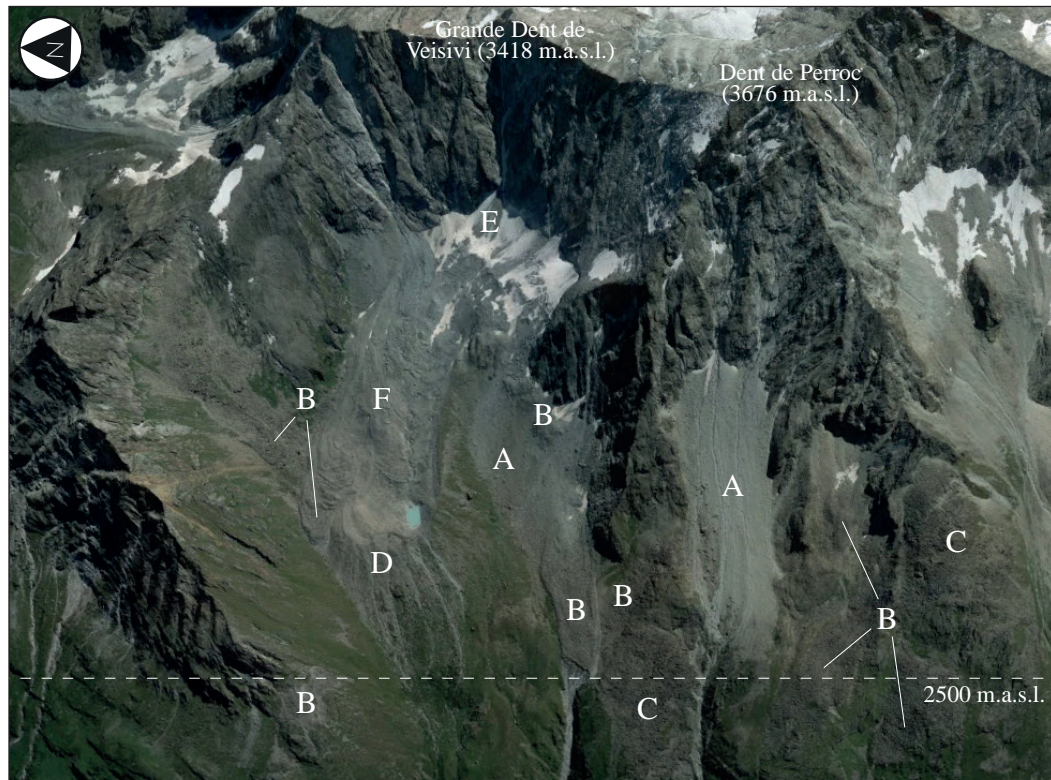
All these characteristics can greatly vary over extremely short distances (metres to few tens of metres), especially in sedimentary deposits, where the ground surface characteristics or the snow cover are generally extremely heterogeneous. Hence, the conditions for permafrost existence may be completely different from one sector to another, located only a few metres nearby.

## 2.2 Mountain permafrost distribution

The *periglacial environments* are often characterized by a strong variability of landforms (see figure 2.2). Within a same landscape, permafrost may be present or absent in a specific landform, as well as being heterogeneously distributed within the same landform. As mentioned in section 2.1, permafrost is a phenomenon defined as a thermal state of the ground and it is generally not observable on the surface. However, some morphological indicators may reveal its presence or absence. Rock glaciers are indeed superficial evidences of the potential permafrost occurrence.

Intact rock glaciers generally reveal its presence, while relict rock glaciers are indicative of its absence. Rock walls and talus slopes are less helpful to visually identify the permafrost occurrence, despite some topographical or morphological characteristics that may be indicative of the potential presence or absence of this phenomenon (i.e. altitude, aspect, grain size, etc.).

In the next sections the main mechanisms controlling the permafrost occurrence in these landforms will be presented.



**Fig. 2.2.:** The Tsarmine sector (Valais Canton), a typical periglacial landscape located above the PLL. Various glacial and periglacial landforms are visible: (A) talus slope (B) rock glacier (C) rockslide (D) morainic complex (E) glacier (F) debris-covered glacier (3D map: Windows Maps ©).

### 2.2.1 Mountain permafrost occurrence and distribution in debris-covered surfaces

In the periglacial environments, the main landforms or areas characterized by permafrost are rock glaciers, talus slopes (or scree slopes), moraines and proglacial margins. The next sections illustrate their characteristics and how they are affected by mountain permafrost.

## Rock glaciers

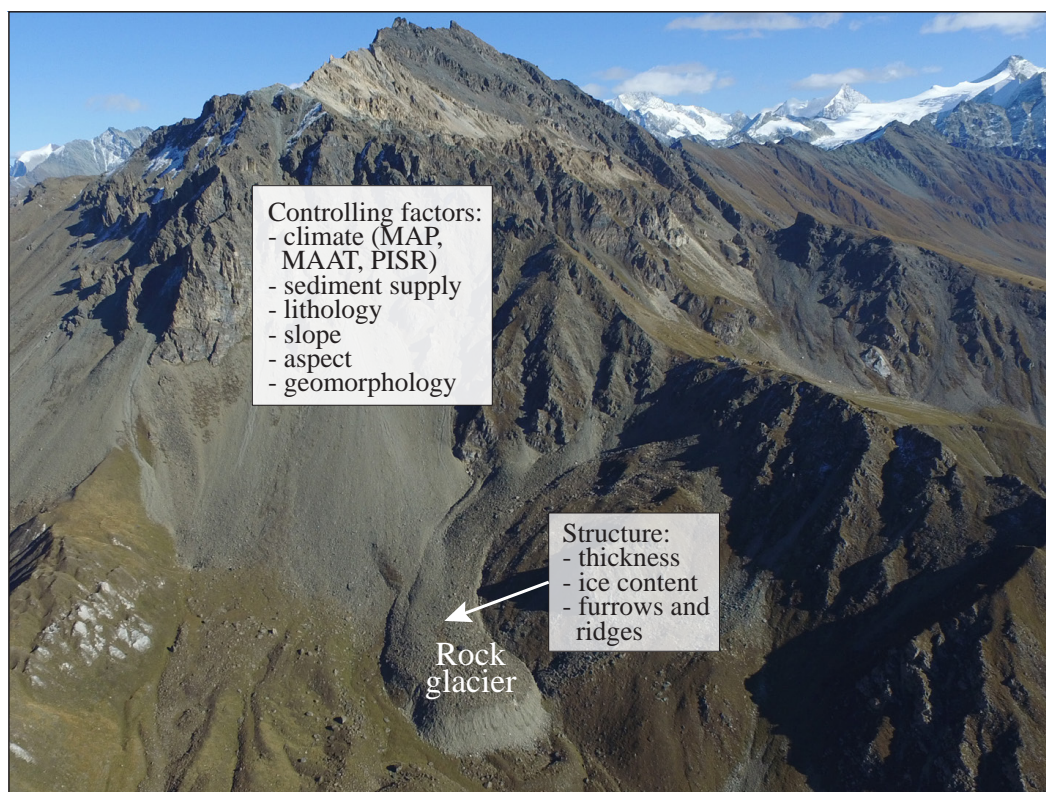
Rock glaciers are the most visible expression of the presence of permafrost (Berthling, 2011). These landforms, generally located above 2300 m.a.s.l. in the Swiss Alps (Haeberli, 1975), are characterized by an internal structure presenting a mixture of inhomogeneous sediment and ice. Creeping occurs in reason of ice-supersaturated sediment deformation (Barsch, 1996). Besides, gravity-induced downslope movements usually result in furrows and surface compression ridges on the rock glacier surface, explained by differences in flow velocities between the basal and surface parts of the landform as well as higher velocities in its central part (Kääb and Weber, 2004).

Classifying rock glaciers according to their dynamics is the most frequent choice (Haeberli, 1985; Barsch, 1996), resulting in the definition of three types of rock glaciers: *relict rock glaciers*, that are motionless and ice-free; *(climatic) inactive rock glaciers*, whose movement is absent even if a frozen body persists; *active rock glaciers*, which are locally super-saturated with ice and present measured displacements that vary from some centimetres to several metres per year. Destabilized rock glacier can eventually reach displacements of ten to tens of metres per year (Krysiecki et al., 2008; Delaloye et al., 2010; Kellerer-Pirklbauer and Kaufmann, 2012; Schoeneich et al., 2015).

<i>Intact rock glacier</i>	<i>Relict rock glacier</i>
<ul style="list-style-type: none"> <li>- Steep front (40-45° for active rock glaciers, 35-40° for inactive rock glaciers)               <ul style="list-style-type: none"> <li>- Convex body</li> </ul> </li> <li>- Water sources at the front of the rock glacier (0-2°C)               <ul style="list-style-type: none"> <li>- Permafrost probable / possible</li> </ul> </li> <li>- Absence of lichens and presence of fine materials at the front of the rock glacier (if active)</li> <li>- Possible presence of lichens and fine materials at the front of the rock glacier (if inactive)</li> </ul>	<ul style="list-style-type: none"> <li>- Front slope lower than 35°</li> <li>- Evident absence of displacements</li> <li>- Water sources at the front of the rock glacier (warmer then 2°C)</li> <li>- Permafrost unlikely</li> <li>- Presence of thermokarst depressions</li> </ul>

**Tab. 2.1.:** Intact and relict rock glacier characteristics permit to interpret the potential permafrost occurrence.

Although it is impossible to assess its activity state without recurring to measurements of their dynamics (Kenner et al., 2014), a rock glacier can be classified thanks to observable characteristics. Morphological evidences, such as the steepness of the front, the convexity of the rock glacier body or the presence of vegetation, can be indicative of the rock glacier state (Frauenfelder, 1998; Gruber and Haeberli, 2009). Indeed, the front of active rock glaciers are generally steep, while they have a gentler slope when the rock glacier activity is reduced. Observable characteristics that help distinguish intact rock glaciers from relict rock ones are summarized in table 2.1 (Haeberli, 1985; Barsch, 1988).



**Fig. 2.3.:** Rock glacier controlling factors and morphological characteristics (structure) (Les Cliossez rock glacier, Hérens valley, Valais Canton; drone photo: S. Rüttimann).

The origin of rock glaciers can vary according to the sediment origin. Barsch (1969) and Haeberli (1985) distinguish *talus-derived rock glaciers*, that are located at the foot of talus slopes, from *moraine-derived rock glaciers*, that develop from ice-cored moraine debris. The morphogenesis of *intact talus-derived rock glaciers* (active and inactive) depends on several controlling factors (figure 2.3; Frauenfelder, 2004). Both the climate and the terrain characteristics of the sectors, where the rock glacier develops, play a fundamental role (Berthling, 2011). The material supply heavily depends on the lithology of the site. The ice content depends on the internal structure of the sedimentary landform where the rock glacier starts developing (Haeberli, 1985). If an ice-rich debris surface become unstable, namely when the shear stress of the ice-supersaturated debris exceeds the critical limit (Barsch, 1996), a first ridge, referred as *protalus rampart* in the literature, may consequently appear (Haeberli, 1985; Scapozza et al., 2011; Scotti et al., 2013). If the MAAT and the annual average precipitation (MAP) are favourable to the maintenance of permafrost conditions and sediment is continuing to be supplied, a rock glacier develops from this embryonic lobe (Janke and Frauenfelder, 2008).

### Permafrost in talus slopes

Talus slopes are detrital landforms that gradually set in place due to the erosion and fall of rock fragments from a rock wall or slopes that are inclined to more than 40-45°

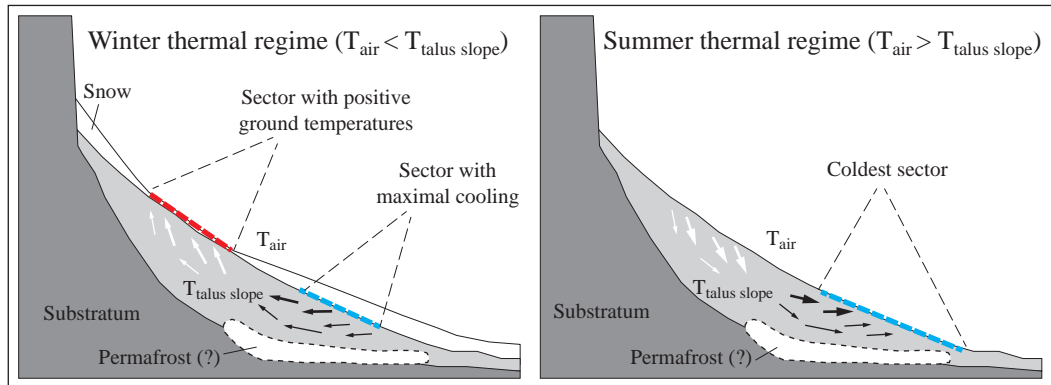


(Francou and Hétu, 1989). In a periglacial environment, sedimentary materials are usually mobilized by gelifraction. This mechanical process is characterized by the breakup and churning of rock as a consequence of repeated freezing and thawing of water within its cracks (see French, 2013). The alternation between water and ice phases leads to a destabilization of the bedrock, resulting in the subsequent fall of blocks once the rock is poorly stable (see Weber et al., 2017).

In talus slopes, the size of the sediments is generally coarse and heterometric. Their distribution is normally graded, with larger blocks located in the distal part of the landform in reason of their higher kinetic energy during the transfer. This is not the same for fine sediments. In its distal part, coarse blocks present empty pores, especially on surface, due to the migration of fine materials to the subsurface, filling the pores between the coarser blocks.

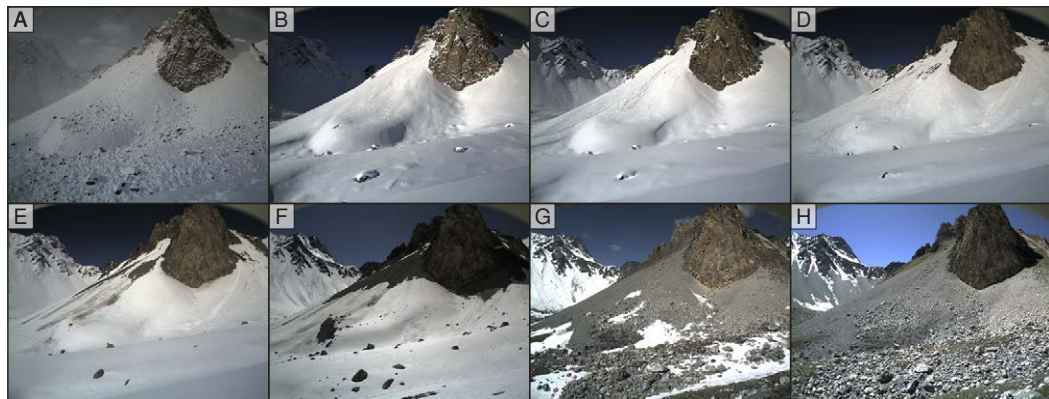
The grading of the sediments has an effect on permafrost occurrence. As mentioned in section 2.1.1, permafrost presence is particularly favourable in coarse debris thanks to existence of voids that are filled by air facilitating the cooling of the ground (Hanson and Hoelzle, 2004). The micro-topography of coarse debris-covered surfaces is not propitious for the installation of an homogeneous snow cover. The air with negative temperatures is free to sink into funnels that are present in the ground, where the warmer air is found. The so-called *Balch effect* takes place, with colder air displacing the lighter warmer air during the colder months (Balch, 1900). During the summer months, the presence of cold air in the ground prevents the infiltration of the warmer summer air and thus, it allows maintaining cold subsurface temperatures. Delaloye and Lambiel (2005) unravelled a similar effect that occurs in talus slopes. It is generally referred as *chimney effect* and it ventilates these landforms. This mechanism allows the migration of the air present within the ground thanks to the porous nature of the subsurface (figure 2.4; Schaeflein, 1962; Wakonigg, 1996; Sawada et al., 2003) During the winter months, air advection occurs in the ground due to a contrasting thermal gradient between the temperature of the external air, which is generally cooler, and the air contained in the talus slope. The lower density of the latter produces an upward circulation and a consequently evacuation of the warmer air in the upper part of the slope. This mechanism generates an air depression in the lower part of the scree, which is filled with a cold air suction from the outside (Morard et al., 2010). During the summer months, the air present within the talus slope is generally colder due to the air stratigraphy mentioned above and therefore it is more dense compared to the external air. Consequently, it flows by gravity allowing a discharge of a colder air mass from the upper part to the lower part of the talus slope. This ventilation effect has therefore consequences on the thermal regime of the landform, creating a positive temperature anomaly, that occurs in the upper part of the talus slope, and a negative annual temperature anomaly that takes place in the lower part of the slope (Morard et al., 2010). The chimney

effect is considered one of the main reason able to explain why permafrost presence is often only restricted to the mid-lower area of talus slopes (Lambiel and Pieracci, 2008; Scapozza et al., 2011).



**Fig. 2.4.:** Schematic yearly thermal regimes of a ventilated talus slope (modified from Morard et al., 2010).

The strength of this ventilation system is strictly related to the thermal gradient: the higher the temperature differences between the internal and external air masses in the scree, the more efficient the mechanism is. In addition, during the winter period, a layer of snow is often present on these sedimentary formations. During the coldest months, the warm air is evacuated in the upper part of the scree through the snow layer by opening *melting chimneys* that can be converted into *melting windows* (figure 2.5).



**Fig. 2.5.:** Snow melting timelapse at the Petit Mont Rouge talus slope (Fontanesses sector, Valais Canton; UNIL webcam images). (A) October 2012: first snow (B) December 2012: thickening of the snow cover (C) February 2013: avalanches thickening the snow cover downslope (D) April 2013: appearance of melting chimneys (E) May 2013: melting windows in the upper part of the talus slope (F) June 2013: distal part starts to melt when the upper part is snow free (G) July 2013: last avalanche deposits start to melt (H) From August 2013: snow free period.

The mechanism is less prone to occur in fine-grained materials because of fewer air-filled pores and a great moisture, hindering a pronounced cooling of the ground (Heggem et al., 2005). The sediment *grain size* is therefore one of the main controlling factors of the permafrost occurrence in talus slopes.

The mechanism promotes the formation of interstitial ice in the distal part of a scree, with a freezing of the percolating water. A creep of the landform is possible if various conditions are met: the debris sub-surface must be ice-supersaturated, while the shear stress, the slope gradient and the thickness of the deposit must reach critical limits (Barsch, 1996). The constant creep and the materials supply from beneath rock walls can lead to the formation of a first ridge, commonly known as *protalus rampart* (Haeberli, 1985; Scapozza et al., 2011; Scotti et al., 2013), which can then evolve into a talus-derived rock glacier (see section 2.2.1).

In fine-grained scree slopes, solifluxion may also occur. This process is defined as *mass movement of soil and regolith affected by alternate freezing and thawing* (Matsuoka, 2001; Matsumoto et al., 2010) and it can take place in sectors with favourable permafrost occurrence. As illustrated by Scapozza (2013), solifluction (intended as *gelifluction*) lobes may however also form in fine-grained materials that are located within the PLL, but not affected entirely by permafrost.

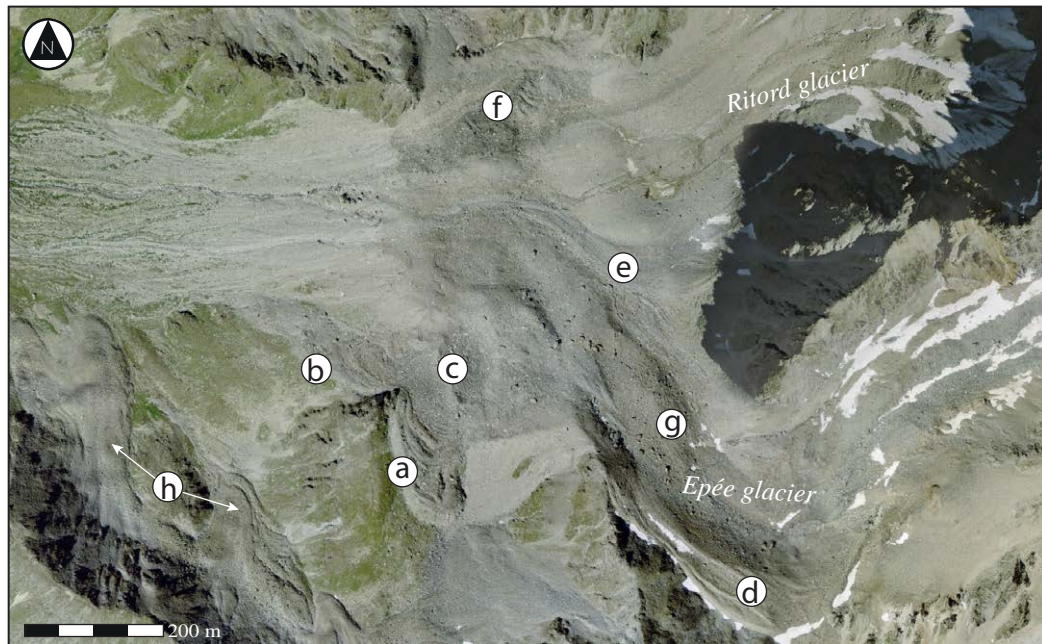
### **Permafrost in morainic terrain**

The permafrost occurrence in the Little Ice Age (LIA) moraine deposits is characterized by an extreme spatial discontinuity. Within LIA proglacial areas of large valley glaciers located below the lower altitude limit of permafrost (the PLL), the absence of permafrost conditions does not allow the creation or the maintenance of ground ice (neither magmatic nor sedimentary) (Etzelmüller and Hagen, 2005; Ribolini et al., 2010). Similarly, in sectors where the Younger Dryas moraine deposits are also located, the climatic conditions are even more unfavourable to permafrost presence, especially when considering the presence of vegetation that hinders an effective sub-surface cooling.

Above the PLL, the maintenance of pre-LIA permafrost in proglacial areas is strongly dependent on the thermal regime of the former glacier (Kneisel, 2003; Delaloye, 2004; Waller and Tuckwell, 2005; Ribolini et al., 2007). Although permafrost conditions can be maintained under cold-based glaciers (Reynard et al., 2003; Waller et al., 2012; Pollard, 2018), alpine temperate glaciers are generally characterized by a water percolation at their base that does not allow negative temperatures to be preserved. In addition, the geometric and thermal disturbance that occurred during the LIA glacier advance were unfavourable to the maintenance of permafrost below the glacier base (Pollard, 2018).

In the proglacial areas of present glaciers, permafrost is generally more likely present very locally in their lateral margins, as it was documented after the investigations carried out in the proglacial areas of Becca of Agè (Reynard et al., 2003), Gruben (Haeberli, 2005), Ritord and Epée (figure 2.6; Delaloye, 2004), Muragl (Kneisel and Käab, 2007) or Col des Gentianes (Lambiel and Schuetz, 2008). The presence of

ice in the sub-surface can have a *periglacial* origin and be induced by the refreezing of percolating water (the so-called *congelation or magmatic ice*) or it can have a *glacial origin*, resulting from sedimentary ice that was left by the glacier in the glacial till or integrated into marginal sediments (Bosson and Lambiel, 2016). Haeberli et al. (2006) suggest that the metamorphosis of the buried snow can also transform into ground ice. In addition, the deformation of frozen sediments present in the sub-surface prior the LIA glacier advance (also known as *push moraines*) can often be observed in recently exposed lateral margins of glacier forefields (see i.e. Delaloye, 2004; Kirkbride and Deline, 2013).



**Fig. 2.6.:** Glacier forefield of the Ritord and Epée glaciers (Valais Canton, Switzerland. (a & b) Marginal rock glaciers, (c) buried ice, (d, e & f) push moraines, (g) debris covered glacier and (h) talus rock glaciers (modified from Delaloye, 2004; orthophoto: swisstopo ©).

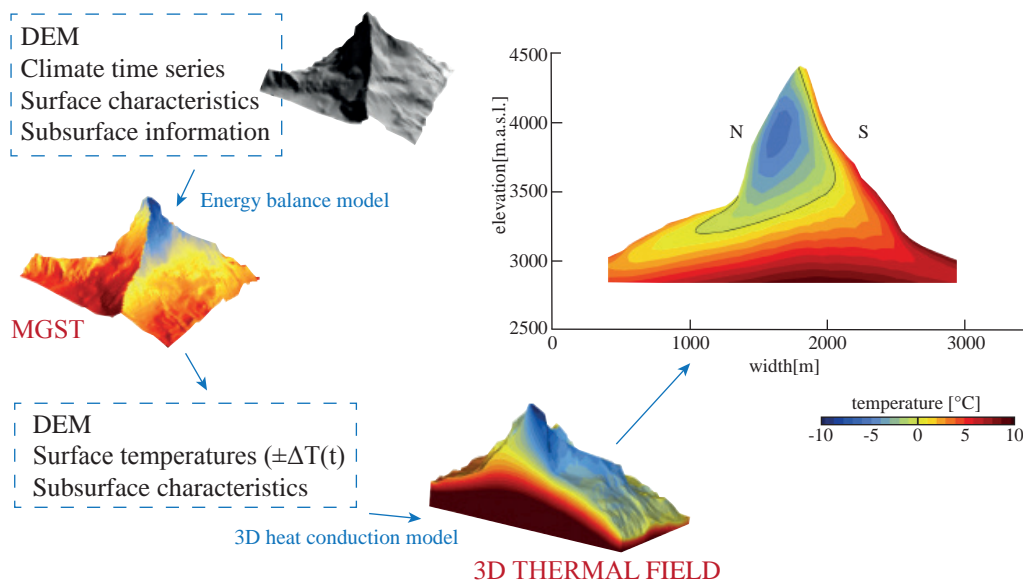
In these environments periglacial creep landforms may also appear. When the thickness of the ice-debris mixture and the slope are sufficient, debris deposits can creep downward (or backward, as it was for example observed for push moraines of the Aget glacier forefield by Delaloye, 2004). Protalus rampart or glacier-derived rock glaciers (marginal rock glaciers) may thus develop (i.e. figure 2.6.a, Berthling, 2011; Monnier and Kinnard, 2015; Bosson and Lambiel, 2016).

In coarse debris located within the proglacial margins, the existence of voids, where cold air can accumulate, could enhance a postglacial permafrost aggradation (neopermafrost). Indeed, in some deglaciated areas recently formed ground ice was evidenced, indicating that permafrost conditions may exist in these sectors (Kneisel, 2003; Etzelmüller and Hagen, 2005; Kneisel and Käab, 2007; Ribolini et al., 2010). In turn, fine-grained sediments hardly induce a thermal rebalancing of the ground with the present atmospheric conditions.

In summary, the distribution of alpine permafrost in moraine deposits is very heterogeneous and highly dependent on micro scale conditions (grain size, water percolation, presence of ground ice, etc.), even in sectors located above the PLL. Despite some small portions of their marginal areas where these conditions are met, glacier forefields are commonly recognized as areas characterized by the absence of permafrost in reason of the strong perturbation that occurred during the LIA glacier advance.

## 2.2.2 Mountain permafrost occurrence in rock walls

The degradation of permafrost has been identified as one of the most important phenomena that lead to the destabilization of rock walls located within periglacial environments (e.g. Kellerer-Pirklbauer et al., 2012; Ravelin et al., 2013; Deline et al., 2015). The total absence of debris coverage due to the steepness of these landforms results in a significantly faster response in time to temperature changes (Mittaz et al., 2000; Hoelzle et al., 2001). Gruber et al. (2003) also recall that the almost total absence of snow cover on rock walls, due to their significant slope and the absence of a complex thermal shift between surface and subsurface, are factors that need to be taken into consideration in order to understand the mechanisms that govern the occurrence of permafrost in rock walls at the local scale. Therefore, the main controlling variables are the altitude and the direct solar radiation, which vary greatly depending on the aspect and the albedo of the rock.



**Fig. 2.7.:** Modelling approach and representation of the subsurface thermal field of a mountainous massif, illustrating the differences between northern and southern exposed slopes. Example of the Matterhorn, Canton of Valais (modified from Noetzli et al., 2007).

These factors were first identified by Hall (1997) by using temperature data logger measurements on rock walls in Antarctica. This study illustrated that in the absence of direct solar radiation, rock temperature is very similar to air temperature. However, during sunshine hours the albedo becomes the factor that regulates the ground temperatures with very contrasting differences between the different types of surfaces and orientations. Noetzli et al. (2007) were able to model the bedrock heat flux that moves by conductive transfer and to represent the sub-surface thermal field inside a mountainous massif (figure 2.7). The North-exposed side of a massif is generally colder compared to South-exposed slopes. The absence of thermal filters (i.e. a debris or snow cover) on the rock wall surfaces enhances the influence of atmospheric conditions and makes bedrock permafrost in high mountains particularly sensitive to changes (Noetzli et al., 2007).

## 2.3 Permafrost distribution modelling: an overview

Among the permafrost models proposed during the past decades, two categories of models can be identified: those based on an empirical-statistical approach, which are suitable to represent spatial permafrost distribution; those based on a physical approach, that are more adapted to represent the energy fluxes between the active layer and the atmosphere. Both require a selection of parameters that depend on the geographical scale that they refer to. This chapter will give a brief overview of the existing permafrost distribution models and their main differences by paying particular attention to those methods that refer to the mountain permafrost extent.

### 2.3.1 First permafrost modelling attempts

The first attempts in the definition of a set of rules that describe the spatial permafrost extent were proposed by Haeberli (1973), Haeberli (1975), and Haeberli (1985). These rules are the result of several field observations made in the Alps between the 1970s and the 1980s. At first, they informed on the topo-climatic conditions favouring the potential permafrost occurrence. Haeberli (1996) completed the so-called *permafrost-Faustregeln* by adding additional criteria, resulting in the following set of rules:

1. In slopes steeper than 5°, the terrain aspect induces a variation of direct solar radiation and thus it has a great influence on the lower limit of discontinuous permafrost.
2. In slope feet, where the snow is maintained until summer, the lower limit of the discontinuous permafrost is generally lowered.
3. In slopes less steep than 5°, the influence of air temperature and snow height is more important than the aspect.

4. Below the forest upper limit, permafrost presence is very sporadic; above, permafrost is potentially present under certain conditions.
5. Intact rock glaciers (active/inactive) are indicators of permafrost presence.
6. Perennial snowfields or snow patches are indicators of the permafrost presence.
7. In alpine meadow areas, the probability of finding permafrost is low (<25%); in non-vegetated talus slopes it is high (>75%).
8. Coarse block-covered surfaces are favourable to maintain permafrost; fine grain size materials are less favourable.
9. Temperatures of water sources below 1°C can only be explained by the contact of water with a permafrost body (unless snow melting water seeps into the ground).
10. Ground water temperatures above 2°C indicate ground thermal conditions unfavourable to permafrost presence.

These empirical observations initiated the development of the first permafrost distribution models. Haeberli (1996) was indeed one of the firsts to propose a topo-climatic model (table 2.2), defining the lower limits of the permafrost belt for different aspect and topographic features of the studied terrain. In addition, this simple model provided a first classification of permafrost occurrence, indicating the altitudes above which the presence of permafrost is possible or probable.

<i>Aspect</i>	<i>Possible occurrence of permafrost above [m]</i>		<i>Probable occurrence of permafrost above [m]</i>	
	<i>Steep slope</i>	<i>Foot slope</i>	<i>Steep slope</i>	<i>Foot slope</i>
N	2400	2100	2600	2550
NE	2450	2300	2600	2700
E	2600	2400	3000 ?	2700
SE	2850	2300	3000 ?	2700
S	3000	2250	?	2600
SW	2700	2150	2900	2550
W	2500	2100	2600	2450
NW	2350	2050	2400	2400
<i>Wind exposed flat slopes</i>	2600		2700	
<i>Wind protected flat slopes</i>	2650		3000 ?	

**Tab. 2.2.:** Topo-climatic model of the PLL (possible or probable) for different terrain orientations and locations (modified from Haeberli, 1996).

### 2.3.2 First empirical models of alpine permafrost distribution

The first numerical representation of permafrost distribution in the Alps was designed by Keller (1992). Developed in an ARC/info environment, named PERMAKART and based on the observations made by Haeberli (1973) and Haeberli (1975), this model was able to predict the potential distribution of mountain permafrost. The model employed the altitude and the terrain aspect as predictors (also see Keller and Hoelzle, 1996). Subsequently, by following the empirical rules proposed by Haeberli (1996), Imhof (1996a) proposed an improvement of this module by

including different types of land cover. A second module, known as PERM, was thus developed and successfully tested in the Bernese Alps. Similarly, Dousse (1992) and Dousse and Lugon (1996) simulated the spatial distribution of permafrost in relation to the aspect and altitude of the slopes and weighted its occurrence in relation to the ground cover (also see Lugon and Delaloye, 2001).

Other distribution models of discontinuous permafrost were successively published (Delaloye and Morand, 1997; Lambiel and Reynard, 2001; Baroni et al., 2004; Seppi et al., 2005), all using the lowest altitude of intact rock glaciers to determine the lower limit of the permafrost belt for a given aspect. With this empirical approach, maps of the potential distribution of permafrost could be established. However, as long as the regions where the rock glaciers were inventoried were different, the altitude limits above which permafrost could be found could slightly vary. For this reason, it was only possible to apply these models at a regional scale and a comparison of the resulting maps was not recommended.

### 2.3.3 Empirical-statistical models of the distribution of permafrost

Not wanting to employ the rock glacier lower limit predictor as the sole environmental variable, purely empirical models left room to empirical-statistical modelling approaches. The best-known model of this type was PERMAMAP (Hoelzle and Haeblerli, 1995). To predict the occurrence of permafrost, this model employed the empirical relationship existing between the snow free-period direct solar radiation and the air temperature of boundary zones, namely those sectors where the bottom temperatures of the snow cover (BTS) measurements indicate ground temperatures between  $-2^{\circ}\text{C}$  and  $-3^{\circ}\text{C}$ . PERMAMAP was efficiently tested in several study sites (i.e. Frauenfelder, 1998; Nyenhuis et al., 2005), confirming that BTS measurements were a valid alternative to rock glacier inventories. As a consequence, temperatures at the base of the snow layer were adopted for models based on different formulations, both empirical-statistical and physical approaches (Keller et al., 1998; Riedlinger and Kneisel, 2000; Gruber and Hoelzle, 2001; Stocker-Mittaz et al., 2002; Ebohon and Schrott, 2008). Similarly, borehole temperatures were also employed to develop similar models (Gruber et al., 2004b; Heggem et al., 2005; Etzelmüller et al., 2006; Etzelmüller et al., 2007; Allen et al., 2009).

### 2.3.4 Physical models of alpine permafrost distribution

No model among those described above took soil characteristics into account, which significantly influence the occurrence of permafrost at the local and micro scales. Indeed, the energy balances are strongly dependent on the type of the soil cover and the duration of the snow cover. Although the influence of energy fluxes is relatively



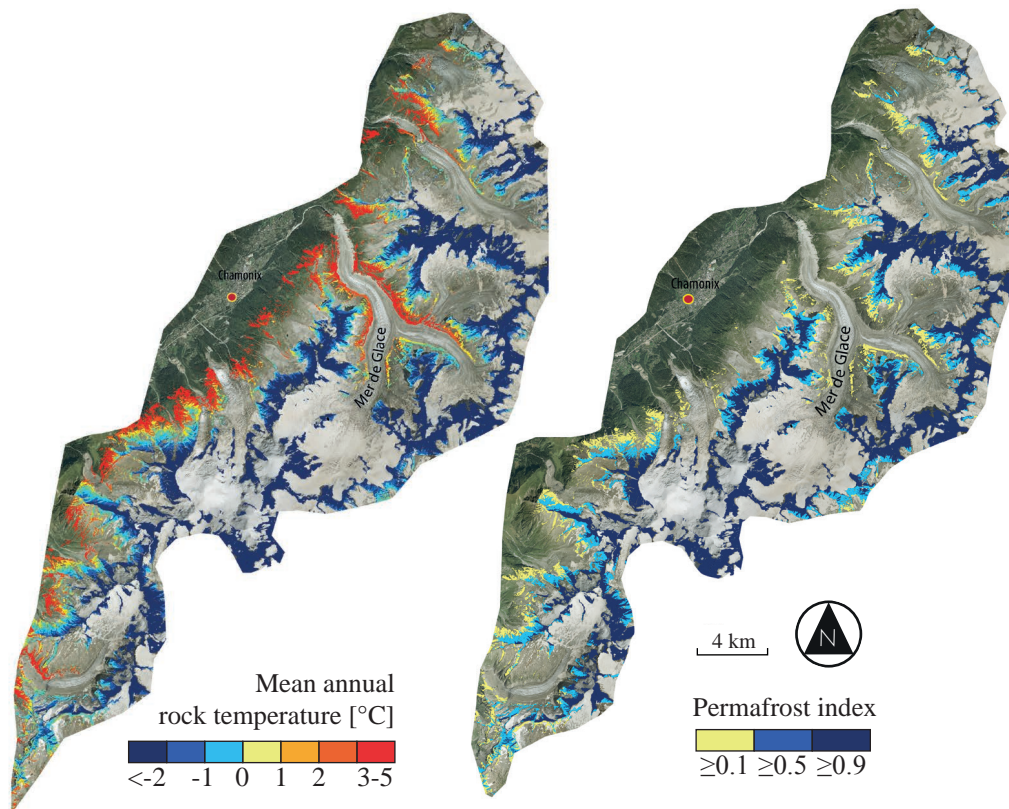
easy to describe for steep rock walls, in less steep slopes the ground temperature does not vary linearly with atmospheric conditions as it reacts in a more complex way. Mittaz et al. (2000) and Hoelzle et al. (2001) proposed thus the first models attempting to describe the energy balances occurring in the active layer. Stocker-Mittaz et al. (2002) developed the *PERMEBAL* model, which integrated the thermal offsets depending on both the snow cover and the characteristics of the active layer. In a first step, the model took the surface energy fluxes into account in order to calculate the temperature at the ground surface. It required meteorological data, a digital elevation model and data concerning soil characteristics, such as the albedo, the emissivity and the roughness. The computed surface temperature, coupled with the thermal offset that resulted from the characteristics of the active layer (porosity, size of blocks, presence of snow, etc.), was then employed to calculate the temperature of the ground (MAGST). Recently, Wicky and Hauck (2017) were able to simulate the thermal regime occurring in a talus slope (see 2.2.1) and to reconstruct the potential permafrost distribution at the scale of this landform.

These models, like any physical ones, are unluckily effective only at a punctual scale (1-dimensional scale) and when many data are available. The spatialization of permafrost occurrence is extremely hard to obtain. This was indeed one of the limits faced by Guglielmin et al. (2003) during the development of the *PERMACLIM* model, which is composed of two modules allowing computing the MAGST, on the basis of the measured MAAT, and calculating the thermal shift between the MAGST and the MAAT induced by the snow layer. This model was one of the first process-based attempts to produce a permafrost distribution map. However, it proved to be severely limited when having poor information available, especially concerning both the characteristics of the snow cover, like the thickness and the density, and its remobilization by the avalanches and wind. Unfortunately, it is very difficult to produce reliable representations of the permafrost spatial extent by only considering information proper to the punctual scale.

### 2.3.5 Permafrost models in rock walls

The first modelling attempts to obtain the mean annual rock surface temperature (MARST) were proposed by Gruber et al. (2003) and Gruber et al. (2004b). They adapted the *PERMEBAL* model to rock wall conditions. It was thus possible to employ temperatures measured with data loggers that were installed in rock walls with slopes greater than 70°. The data collected made it possible to understand that the heat transfer by thermal conduction was one of the most important factors governing the presence of permafrost. This was due to the near-total absence of the snow cover and porous material in such environments (Gruber et al., 2003). The resulting average isotherm of 0°C for a period of 21 years was then modelled

(Gruber et al., 2004b). BAFU (2005) integrated the model of Gruber et al. (2004b) into the official Swiss permafrost map. Lately, rock wall measurements obtained by the PERMOS network were employed by Boeckli et al. (2012) for the Alpine Permafrost Index Map (rock module). The APIM rock wall module was then also used in Magnin et al. (2015) for the development of a new permafrost map index in the rock walls of the Chamonix region (France) (figure 2.8).



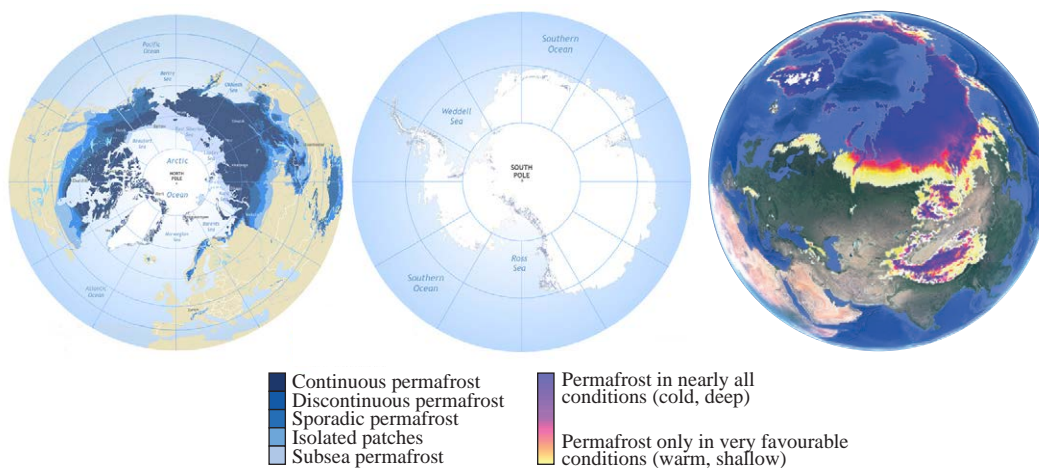
**Fig. 2.8.:** MARST map and permafrost index in rock walls (Chamonix region, France; modified from Magnin et al., 2015).

## 2.4 Permafrost distribution maps

Permafrost distribution maps can be useful supplementary tools for decision-makers, who deal with natural hazards and land-use planning of mountain regions. The impact of the climate change on sectors that are potentially affected by permafrost degradation raised the necessity of having informative permafrost distribution maps. Several models exist and cover different scales and different parts of the Earth. In the next sections, maps of the permafrost extent and their peculiarities will be illustrated.

## 2.4.1 Global scale

At the global scale, the best-known map was proposed by Brown et al. (1997) as part of the international research project *PACE* (figure 2.9). Exclusively based on topoclimatic variables, such as latitude, altitude, air and soil temperature, this map covers the area between the 30°E and 25°W parallels, including Scandinavia, Iceland, Svalbard and the Alps. It did not consider Greenland and Russia. In their analysis of the produced map, Harris et al. (2003) described an important contrast between the Arctic archipelago of Svalbard and the lower latitudes of Scandinavia and the Alps. In the first region, permafrost was mapped as continuous, while in the Alps and in Scandinavia, the elevation and orientation of the slopes played a major role. Permafrost was thus mapped as discontinuous or sporadic (Brown et al., 1997).



**Fig. 2.9.:** Maps illustrating the distribution of permafrost of the North hemisphere (left), Antarctica (center) published by the CALM network and its evolution into the GPZI (right). Modified from Brown et al. (1997) and Gruber (2012).

Other models considering the continental scale were proposed by other permafrost monitoring networks, including polar and subpolar regions. Some examples are the maps published by the *Circumpolar Active layer Monitoring* network (CALM) that, on the basis of borehole temperature measurements, temperature data loggers and meteorological data, simulated the permafrost extension at the low latitudes of the northern and southern Hemisphere (Oelke and Zhang, 2004).

An updated map of the global permafrost extent (except Antarctica) was published by Gruber (2012). The resulting *Global permafrost zonation index map* considered similar parameters, such as the snow cover, the exposition to the solar radiation, the vegetation and the subsurface properties, at a ground resolution of 1 kilometre. This variables were also previously adopted by Brown et al. (1997). As Gruber (2012) indicated, a permafrost map covering the entire planet does not often have a resolution good enough to help decision-makers. Therefore, although these

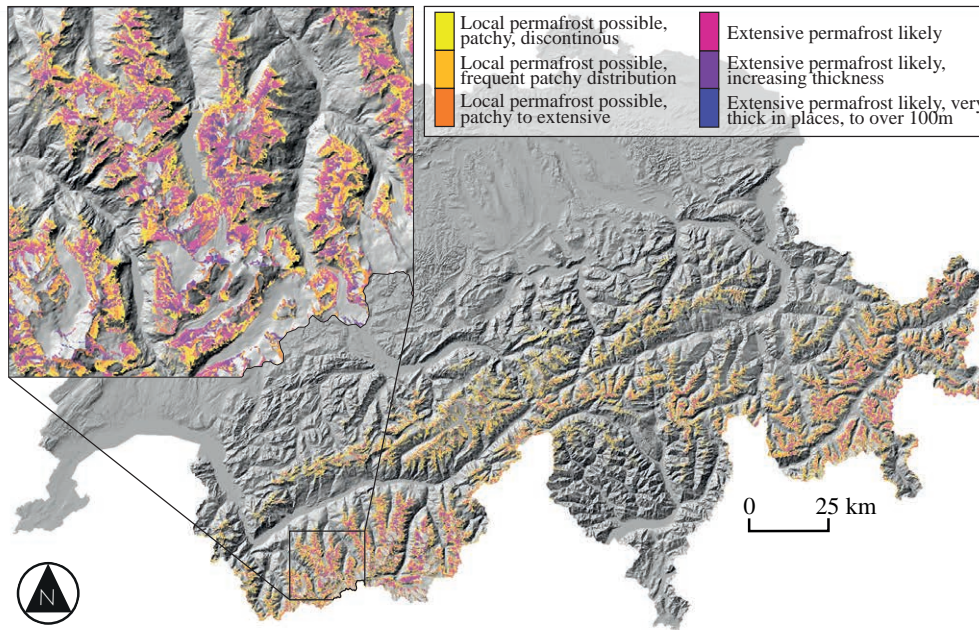
simulations are important tools for understanding this phenomenon and the global relationships between other major issues, they are difficult to exploit for local and regional analyses.

## 2.4.2 Mountain range system scale

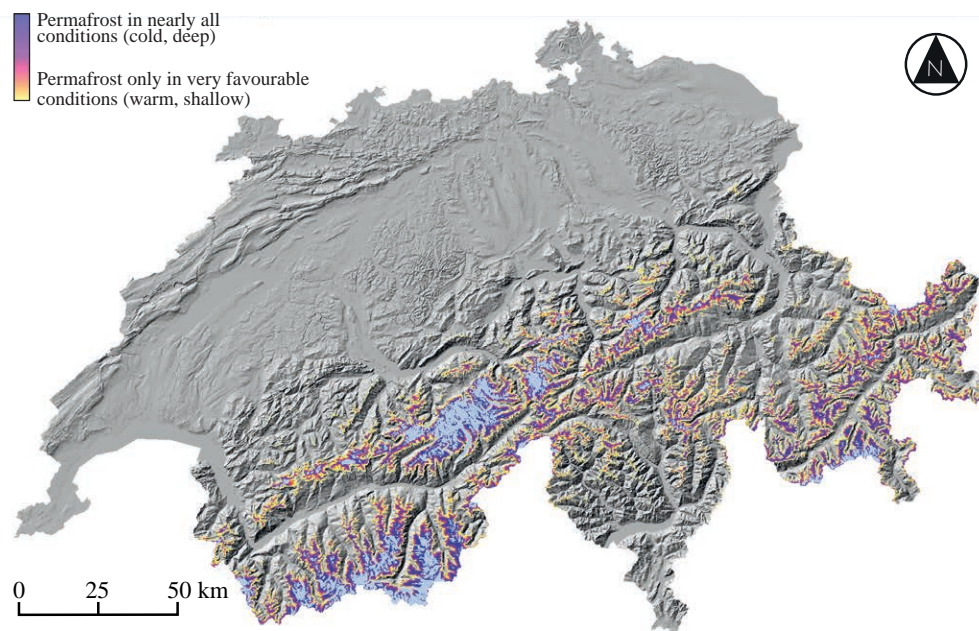
At the beginning of the 2000s, a planning tool designed for areas located in periglacial environments was considered necessary by the Swiss Federal Office of Environment (BAFU). The companies Geotest, Geo7 and Academia Engiadina were thus mandated to create a map of the potential extension of permafrost for the Swiss Alps. By using an empirical approach, this model took into account four types of landcover: glaciers, water surfaces, rocks and debris-covered surface. Permafrost topographic variables were extrapolated from a digital terrain model with a 25 metres of resolution. The altitude limits of permafrost in sedimentary deposits were then empirically obtained by field measurements and observations. The physical model published by Gruber et al. (2004b) was incorporated into the BAFU map to predict permafrost occurrence in rock walls. The resulting map informed on the possible and probable permafrost-covered areas (figure 2.10). However, as the BAFU map was developed by using DEM-extracted variables, no information on the local land-cover and the ground properties (such as lithology, grain size, etc.) were considered. As a consequence, this map does not reflect the high discontinuity of permafrost, showing a general trend of increase of the permafrost occurrence linearly correlated to the increase in elevation.

Boeckli et al. (2012) successively proposed a map of the mountain permafrost distribution at the scale of the European Alps (figure 2.11), known as *Alpine permafrost index map* (APIM). They used an inventory of rock glaciers and the average rock surface temperatures (MARST) as a proxy. The model was divided into two modules, one simulating the permafrost extent in sediment, the other dealing with the distribution of permafrost in rock walls. The debris module exploits environmental variables, such as the MAAT, the potential direct solar radiation and the average annual precipitation, to predict whether the rock glacier of various inventories are intact or relict. An elevation limit, above which permafrost is possible, was then determined from this module. The second module models permafrost in steep slopes and it uses the MAAT and the potential direct solar radiation to set an altitude limit for each orientation. The calibration of this model was performed by using the MARST measured by data loggers installed in rock walls.

Similar maps have been recently published after the emergence of new statistical approaches as an alternative to traditional empirical models, allowing the improvement of the permafrost extent modelling. Generalized Linear Models (GLM) and Generalized Additive models (GAM) were indeed used by several authors to map



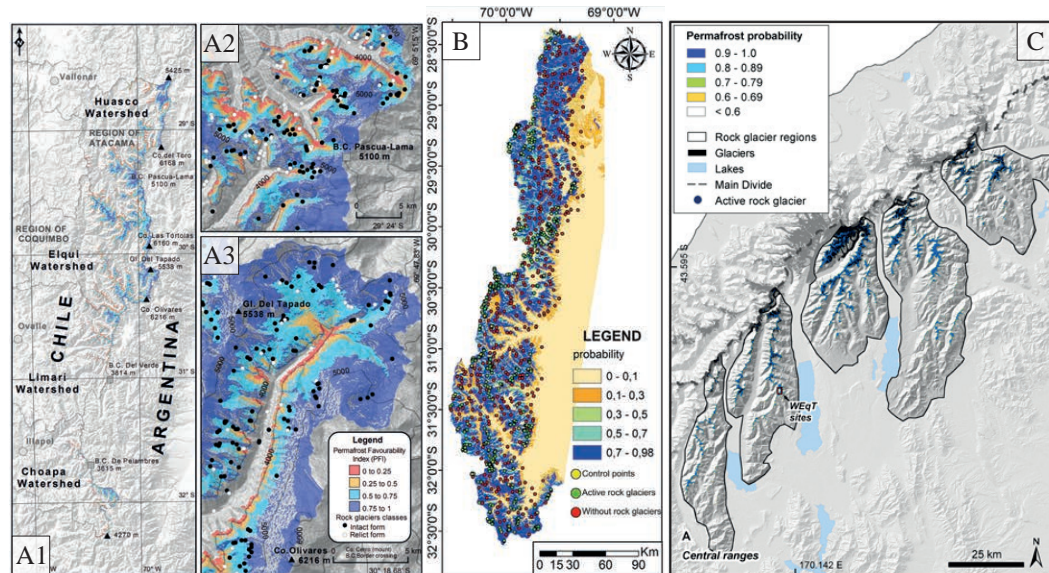
**Fig. 2.10.:** Potential permafrost distribution map published by Swiss Federal Office of Environment (BAFU) and covering the Swiss Alps. Zoom illustrates the region of interest of the present thesis (modified from BAFU, 2005; hillshade: swisstopo).



**Fig. 2.11.:** Extract of the Alpine permafrost index map (APIM) covering the Swiss territory (modified from Boeckli et al., 2012).

the permafrost distribution (i.e. Lewkowicz and Ednie, 2004; Heggem et al., 2005; Brenning and Azócar, 2010; Bonnaventure et al., 2012). The most recent use of this algorithm was published by Azócar et al. (2016) and used to map the permafrost distribution of the semi-arid Chilean Andes (figure 2.12.A). In this region, Brenning and Trombotto (2006) previously modelled the permafrost distribution by using

Logistic regression. This statistical technique was also used to map the permafrost extent of the Maritime Southern Alps (New Zealand, figure 2.12.C; Sattler et al., 2016) and the San Juan Dry Andes (Argentina, figure 2.12.B; Angillieri, 2017). The Regression Spline on permafrost data was also employed as an alternative to Logistic Regression by Zhang et al. (2012).



**Fig. 2.12.:** Overview of recent maps covering the Chilean Andes (A1, A2, A3; modified from Azócar et al., 2016), the Argentinian Andes (B; modified from Angillieri, 2017) and the Maritime Southern Alps of New Zealand (C; modified from Sattler et al., 2016).

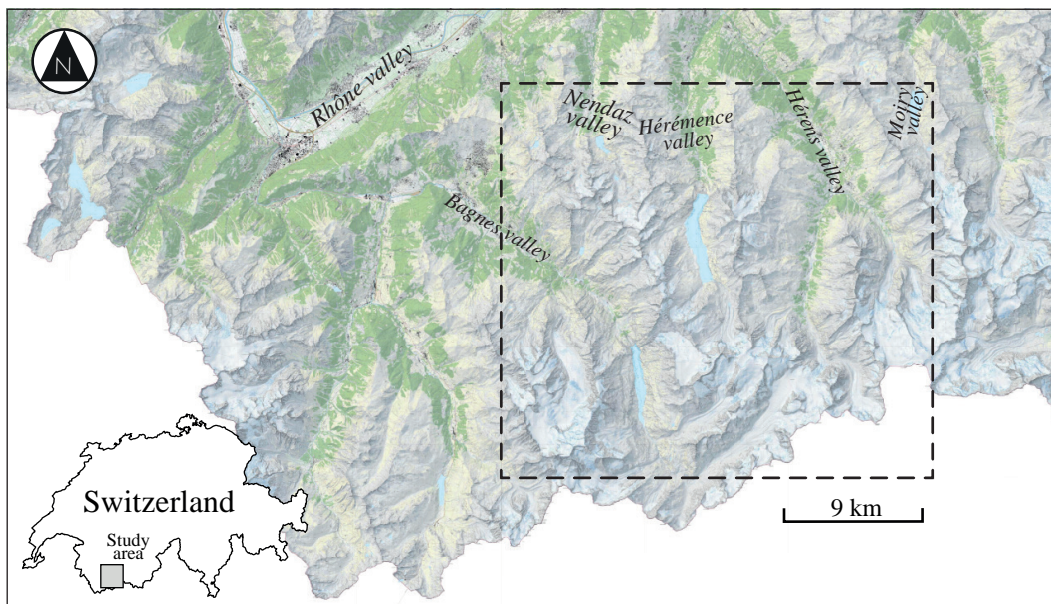
Although the maps proposed by these authors satisfactorily apply to the mountain range scale and the approach used to predict the distribution of the phenomenon is different from classical statistical models, permafrost representation at the local level has the same informative issues as those previously mentioned. None of the employed approaches prove to be appropriate at the micro scale and an interpretation of the corresponding maps is therefore necessary. This is not always obvious for non-expert users. Therefore, permafrost maps at smaller scales, such as the local or the micro scales, are required and the complexity of permafrost distribution has to be reflected. Some of the first analyses of permafrost data aiming at mapping the permafrost distribution at the micro scale were presented in Deluigi (2012) and Deluigi and Lambiel (2013). These attempts used machine learning algorithms and their promising results encouraged the beginning of the present thesis.



# Materials

## 3.1 Region of interest

The region of interest of this study was selected with respect to specific characteristics. The existence of a large set of good-quality data was mandatory to be able to employ a machine learning approach, to model and to map the potential permafrost distribution. Therefore, some criteria had to be taken into account for the success of this research. On the one hand, the study area must have been located at altitudes above the discontinuous permafrost lower limit and characterized by sectors where permafrost evidences have been observed. Indeed, the presence of previously studied periglacial landforms, such as rock glaciers, talus slopes, protalus ramparts, etc., was required in order to take advantage of the existing knowledge on the permafrost presence and absence. On the other hand, the study area must have offered a large spatial diversity and have covered as many geographical and geomorphological settings as possible, such as various aspects, different slopes, a wide geomorphological diversity, etc. The Western Swiss Alps satisfied thus these criteria and seemed to be an appropriate study area. Indeed, the selected portion of the territory contains many field sites that have already been investigated in recent years.

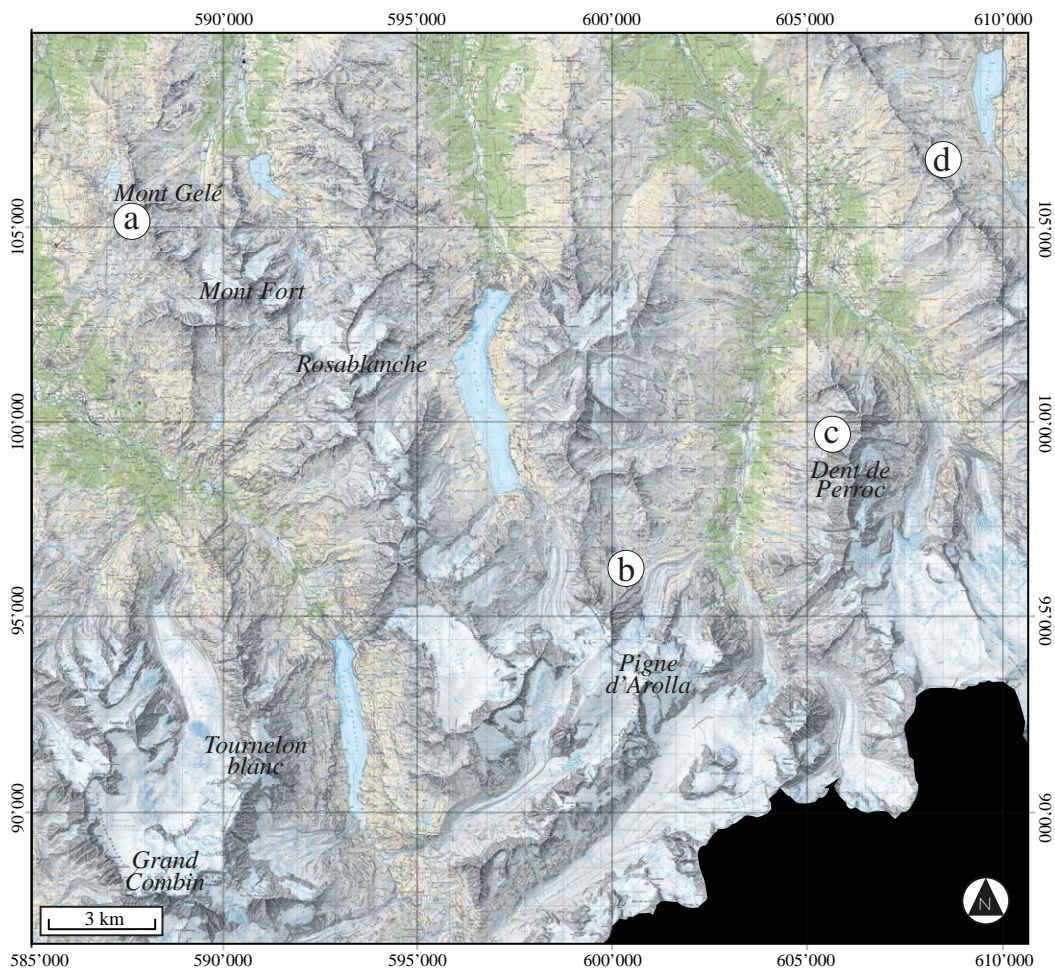


**Fig. 3.1.:** Situation map and extent of the region of interest (topographic map: swisstopo).



The chosen region of interest covers a regular grid of 588 km<sup>2</sup> (figure 3.1). About 360 km<sup>2</sup> are located above the theoretical permafrost lower limit of 2500 m.a.s.l. roughly delimiting the lower boundary of the periglacial belt (Lambiel and Reynard, 2001). As a consequence, about 60% of the territory is potentially affected by permafrost if favourable conditions exist and allow its development and maintenance.

The area is located in the orographic left side of the Rhône valley (Valais Alps) and it is divided by five main valleys: the Bagnes Valley, the Nendaz Valley, the Hérémence Valley, the Hérens Valley and the Moiry Valley (see figure 3.1). The topography of the Eastern and South-Eastern parts of the area is characterized by steep slopes and high elevations. It is indeed in these locations that most of the glaciers are located (Corbassière, Giétro, Brenay, Otemma, Mont Miné, etc.) and dominated by the highest peaks of the region of interest (Combin de Grafeneire, 4314 m.a.s.l.; Mont Blanc de Cheilon, 3870 m.a.s.l.; Pigne d’Arolla, 3772 m.a.s.l.; Mont Collon, 3636 m.a.s.l.; etc.). The Dix lake watershed covers the central part of the area.



**Fig. 3.2.:** Extent of the region of interest and localization of (a) the Mont Fort-Mont Gelé area, (b) the Fontanesses area, (c) the Veisivi ridge and (d) the Tsaté-Moiry sector (topographic map: swisstopo).

The region of interest is characterized by an intra-alpine continental climate, with precipitations between 1200 and 1500 mm per year (Reynard, 2000), which are typical for the central Canton of Valais (Baeriswyl and Rebetez, 1997). According to the Attelas weather station (meteosuisse) installed in the area, the MAAT measured at 2727 m.a.s.l. is approximately -1°C (for the 1993-2016 period).

Since 1992, the Universities of Lausanne and Fribourg have been engaged in various studies on several domains: geomorphological mapping, characterization and mapping of permafrost through geophysical surveys and thermal methods, monitoring of rock glacier kinematics, geophysical prospection of glacial and periglacial landforms, etc. To carry out these researches, four sites have been chosen to be investigated and monitored thanks to their accessibility and to the presence of a wide variety of landforms and geomorphological diversity.

The Mont Fort-Mont Gelé region (figure 3.2.a, Bagnes Valley) has been particularly studied for the identification of the main permafrost controlling factor at the local scale (Reynard et al., 1999; Delaloye et al., 2001; Delaloye et al., 2003a; Lambiel and Reynard, 2003; Marescot et al., 2003; Delaloye, 2004; Lambiel and Delaloye, 2004; Lambiel, 2006; Delaloye et al., 2008; Lambiel and Pieracci, 2008; Hilbich et al., 2008; Delaloye et al., 2010; Hilbich, 2010; Scapozza et al., 2011; Ravanel et al., 2012; Ravanel and Lambiel, 2013; Scapozza, 2015a). It presents a wide heterogeneity of landforms, such as talus slopes (i.e. the Lapires and the Attelas talus slopes), periglacial lobes (i.e. Lac des Vaux site), rock glaciers (i.e. Yettes Conja and Lues Rares rock glaciers) and moraines (i.e. Col des Gentianes).

On the North of the Pigne d'Arolla peak (3772 m.a.s.l., Hérens Valley), the Fontanesses area has been also particularly studied (figure 3.2.b). In this sector, permafrost occurrence has been investigated since the mid-1990s, moving from studying rock glaciers dynamics and their ice content (Gardaz, 1997; Gardaz, 1998; Morand, 2000) to focusing on permafrost distribution and related processes within talus slopes (Delaloye, 2004; Delaloye and Lambiel, 2005; Lambiel, 2006; Lambiel and Pieracci, 2008; Scapozza et al., 2011; Scapozza, 2013; Scapozza, 2015a) and on dating periglacial landforms (Scapozza, 2013). More recently, field investigations have turned to the study of a destabilized rock glacier (Lambiel et al., 2017; Vivero et al., 2017).

Close to the Fontanesses area, the Veisivi ridge is another sector where several studies have also been conducted, engaging in the investigation of periglacial processes (figure 3.2.c; i.e. Lambiel and Delaloye, 2004; Lambiel and Pieracci, 2008; Delaloye et al., 2008; Delaloye et al., 2010) and in the identification and characterization of slope movements (Delaloye et al., 2007; Barboux et al., 2014a) and of sediment transfer processes (Micheletti et al., 2015; Kummert et al., 2017).

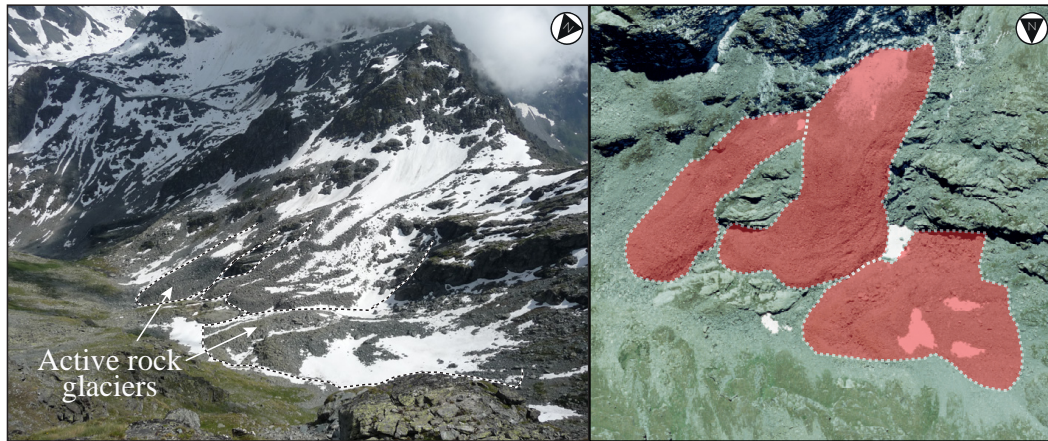
Lastly, in the Tsaté-Moiry sector, located in the Moiry Valley (figure 3.2, the permafrost extent and the associated control factors allowing the development of periglacial landforms have also been intensively investigated. Lambiel (2006) firstly studied the structure and the dynamics of two rock glaciers. Scapozza (2013) extended the investigation to the talus slopes located nearby these landforms, assessing both the occurrence of permafrost in fine debris and the main properties that permit its existence. The main results of these studies are presented in Roer et al. (2008), Lambiel (2011) and Scapozza and Lambiel (2013).

These researches helped collecting data that illustrate the distribution of mountain permafrost at different geographical scales, as well as to understand the mechanisms governing permafrost occurrence, especially in talus slopes and rock glaciers. Gathered data, coupled with other study results, such as rock glacier inventories (Delaloye and Morand, 1998; Morand, 2000; Lambiel and Reynard, 2003) and geomorphological maps (Lambiel et al., 2016), serve as evidence of the permafrost presence and absence. These evidences are presented in section 3.2 and have been employed in this study to model the permafrost distribution of the entire study area.

## 3.2 Mapping permafrost evidences

To map the potential distribution of permafrost with data-driven techniques, evidences of permafrost absence and presence had to be gathered. The main employed sources of data were *rock glacier inventories* and *survey-derived data*. Permafrost presence or absence were indeed derived from *rock glacier maps* based on the activity of these landforms. Active or inactive rock glaciers suggest a potential existence of permafrost conditions, whereas relict ones indicate its absence (see Haeberli, 1985; Humlum, 1996; Barsch, 2012, presented in section 2.2.1). Rock glacier recognition is based on their morphological characteristics and location. Intact rock glaciers are generally located at higher elevations compared to inactive rock glaciers and they can be identified by a steeper front, by the absence of vegetation on its body and by a convex cross-profile. Lastly, the main characteristics presented in table 2.1 (section 2.2.1) were used to distinguish intact rock glaciers from relict ones.

Delaloye and Morand (1998), Morand (2000) and Lambiel and Reynard (2003) mapped directly on the field and inventoried most of the rock glaciers located within the region of interest (figure 3.3). In this research, these inventories were employed and improved: some missing rock glaciers were added to the inventory and new data were also extracted from the Hérens valley geomorphological map, published by Lambiel et al. (2016). All rock glacier limits were corrected by a comparison with recent orthophotos (swisstopo SWISSIMAGE, ground resolution: 50 cm). The maximum extent of the rock glacier surfaces was established at the external border



**Fig. 3.3.:** The Yettes Condja active rock glaciers (Mont Gelé/Mont Fort sector; left image) and their mapping by ortho-image interpretation (right image; orthophoto: swisstopo).

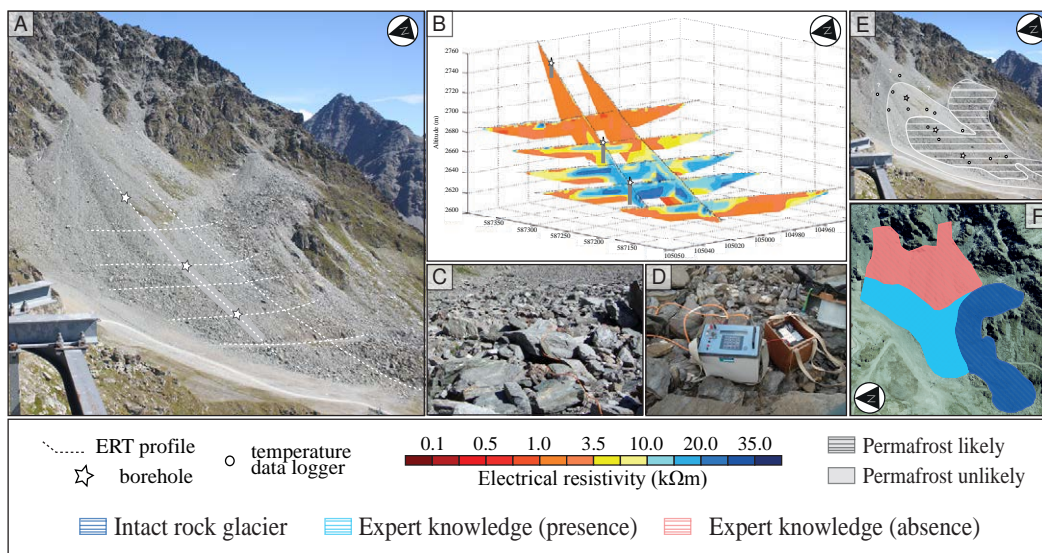
of their bodies in order to take the advantage of curvature predictors (see section 3.3.1). Unlike other studies, where only the punctual characteristics at the roots of a rock glacier were considered (such as the altitude, the MAAT and the PISR at these location, i.e. Azócar et al., 2016; Sattler et al., 2016), in this research the entire surface of rock glaciers was mapped and taken into account. Training data were thus not generated from single points but polygons.



**Fig. 3.4.:** The rock glaciers of the Tsarmine sector (left) and their mapping by ortho-image interpretation (center; orthophoto: swisstopo). Right image illustrates the InSAR signal between 23/07/2009 and 08/10/2009.

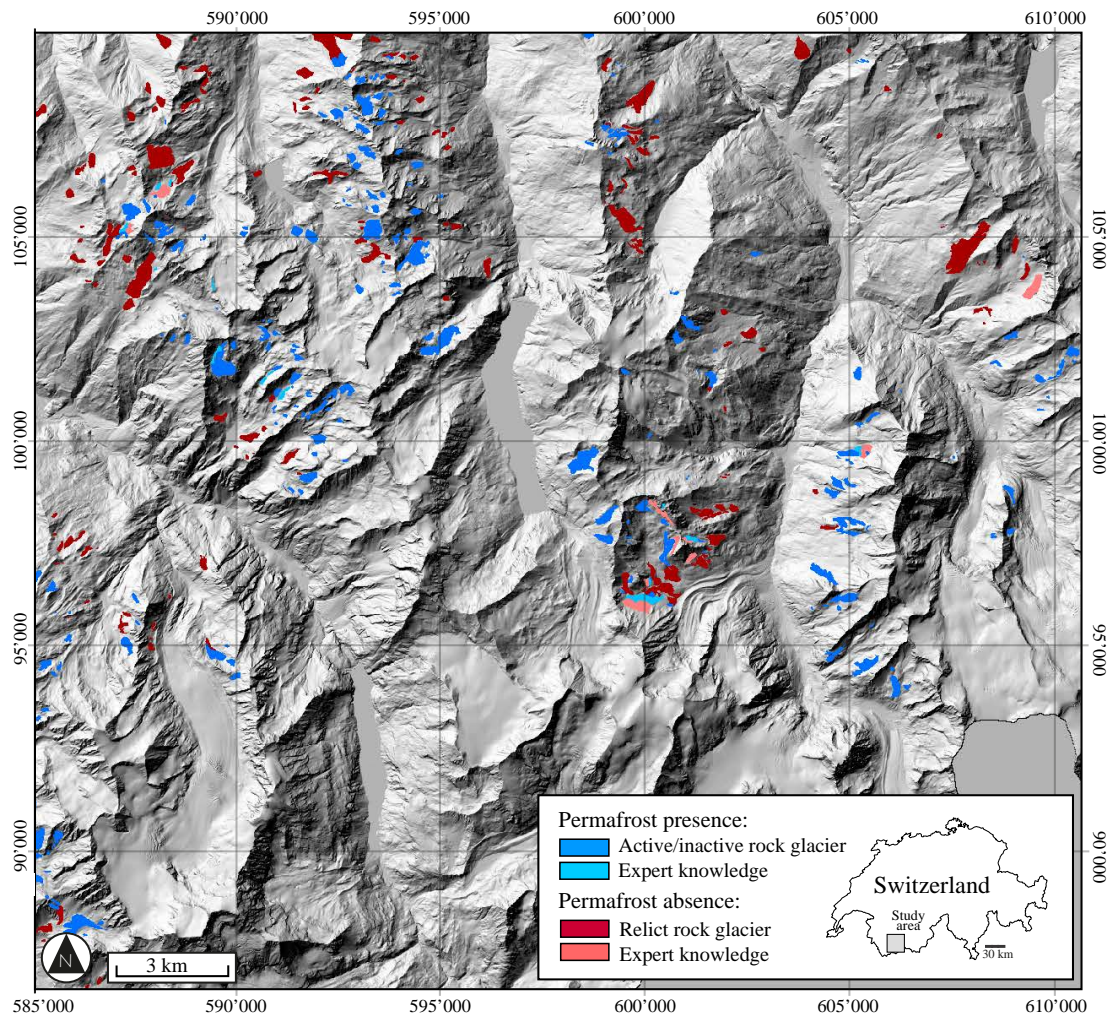
The activity of these rock glaciers was verified with the analysis of geomorphic signatures through ortho-image interpretation (thanks to the mentioned characteristics of these landforms) and *interferometric synthetic aperture radar* (InSAR) signals (Delaloye et al., 2007; Barboux et al., 2014a) (figure 3.4). This technique employs and compares the outgoing radar waves with the known phases to create an *interferogram*. It illustrates the phase difference between two acquisitions, unravelling terrain displacements in the line of sight direction when those are higher than half of the wavelength size ( $\pi$ ). In figure 3.4 (right image), a decorrelated signal indicates that the Tsarmine rock glacier and the marginal rock glacier alimanted by the Tsarmine debris-covered glacier are characterized by movements.

Geoelectrical and thermal data were also employed to generate training examples. Direct-current (DC) resistivity methods are well established tools for detecting permafrost in sediments (Hauck and Kneisel, 2008). Electrical resistivity tomography (ERT) is especially often used to detect ground ice and to characterize frozen materials in permafrost environments (i.e. Hauck et al., 2003; Hilbich et al., 2009; Otto et al., 2012; Rosset et al., 2013; Mollaret et al., 2017, figure 3.5). This technique consists in injecting electrical current into the ground between a pair of electrodes and in measuring the potential difference between another pair of electrodes. Apparent resistivities characterizing the sub-surface can be calculated and positioned at pseudo-depths according to a geometrical reconstruction (inversion process, see Edwards, 1977), resulting in a pseudo-section that represents an approximate model of true sub-surface resistivity distribution (Hack, 2000). The ERT profiles only provide an indication of the electrical conductivity (or resistivity) of the ground; thus an interpretation of the result is always required. In fact, measured values highly vary depending on the ground properties (lithology, presence of air, water or ice). For instance, frozen soils and ice-filled sediments present resistivity values of several orders of magnitude higher in comparison to non-frozen sediments (e.g. figure 3.5.B). Such data must often be coupled with ground surface temperature measurements in order to infer the permafrost presence or absence (Hoelzle et al., 1999; Staub et al., 2015; Carturan et al., 2015). Coupling geoelectrical and thermal data can thus improve the reliability of permafrost mapping.



**Fig. 3.5.:** Survey of the Attelas talus slope (Mont Gelé-Mont Fort sector). (A) Overview of the talus slope and localization of the ERT profiles and boreholes (B) 3D view of the ERT profiles, illustrating the ground electrical resistivity (C/D) ERT cables and measuring instrument (Syscal Pro), injecting DC electrical current into the ground between one pair of electrodes and measuring the voltage between another pair (E) Interpretation and mapping of the ERT results, coupled with thermal data (boreholes and data loggers) (F) Overview of the mapped permafrost evidences.

To obtain a consistent collection of permafrost evidences to be used, namely a map of the known permafrost *presence* and *absence*, the research outputs published by Marescot et al. (2003), Reynard et al. (2003), Delaloye (2004), Delaloye and Lambiel (2005), Lambiel (2006), Delaloye et al. (2008), Lambiel and Pieracci, 2008, Scapozza et al. (2011), Scapozza (2013) and Staub et al. (2015) were analysed. The focus of these studies was the detection and the mapping of permafrost and/or ground ice in permafrost environments (mainly talus slopes and glacier forefields). Thermal measurements gathered for the Swiss Permafrost Monitoring Network (PERMOS, 2016) and by other unpublished projects (UNIL Master thesis, Bachelor projects, etc.) were also considered. From this compilation and the combination of geoelectrical and thermal data, permafrost evidences were identified following the delimitation procedure employed by Lambiel (2006) and Scapozza et al. (2011). This operation allowed delimiting various sectors where permafrost *presence* and permafrost *absence* were certain. These data completed the collection of permafrost examples by providing permafrost evidences also located outside rock glaciers.



**Fig. 3.6.:** Extent of the study area and localization of known (presumed) permafrost evidences (hillshaded map: swisstopo).

Observations of the permafrost negative occurrence (known permafrost absence) not only resulted from in-situ measurements indicating warm conditions or absence of ground ice, but also from expert knowledge. The conclusions of Lambiel and Pieracci (2008) and Scapozza et al. (2011), showing the general absence of permafrost in the upper half of talus slopes, were particularly followed for this purpose.

Figure 3.6 presents the intact and relict rock glaciers map, as well as the sectors where the permafrost occurrence is presumed (geophysical prospection or expert knowledge). A total of 523 polygons were selected as training examples, that correspond to more than 93'000 evidences (pixels) of the permafrost absence and presence.

### 3.3 Mountain permafrost predictors

One of the goals of this research is the modelling of the potential mountain permafrost distribution in the most reliable way with regards to the micro scale. As a consequence, the identification of environmental predictors able to explain the permafrost occurrence at this scale were required. Various variables were selected in this study. Some of them are commonly used in the field of permafrost modelling and are well known for controlling the permafrost occurrence at different scales, as presented in section 2.1.1. The altitude, the MAAT, the aspect or the PISR are typical examples. Therefore, to be able to consider the strong discontinuity of mountain permafrost at the landform scale, other predictors had to be used. Morphological predictors like curvature indices were thus selected as a representation of the micro-topography of the periglacial landforms. Other environmental predictors and landcover variables, such as NDVI, vegetation or grain size, were also considered.

An explanation of the physical relevance of the adopted DEM-derived predictors and of morphological and land-coved variables are presented in the next sections, along with their sources and methodology that allow their computation.

#### 3.3.1 DEM-derived predictors

*Altitude*: at the regional scale, it is commonly accepted that the permafrost occurrence increases with *altitude*, due to the decrease of the MAAT. This variable was derived from the SwissAlti3D digital elevation model provided by the Swiss Federal Office of Topography (swisstopo). It is a spatial grid with a resolution of 2 metres above 2000 m.a.s.l. (1 metre below 2000 m.a.s.l.) and it is produced by stereo correlation with 1-3 m average error. The density of at least 2 points/m<sup>2</sup> avoids any noise in the data. From this variable, most of the following predictors were derived.

*Mean annual air temperature*: the MAAT was computed on the basis of a linear temperature gradient. The approach used to obtain this environmental variable follows the methodology proposed by Bouët (1978). The MAAT was thus computed as:

$$T = T_0 - \left[ (z - z_0) \cdot \frac{\Delta T}{\Delta h} \right] \quad (3.1)$$

for weather stations located above 1500 m.a.s.l..

In equations 3.1  $T$  is the temperature (in Celsius degrees) calculated at the altitude  $z$ ,  $T_0$  is the measured temperature located at the altitude  $z_0$ ,  $z$  is the altitude of the location where the temperature has to be computed,  $z_0$  is the altitude of the weather station, and  $\frac{\Delta T}{\Delta h}$  is the temperature gradient ( $^{\circ}\text{C}/\text{m}$ ).

$\frac{\Delta T}{\Delta h}$  was obtained from air temperature data collected during the period 1980-2010 at the Evolène-Villa (1825 m.a.s.l.) and the Montana (1427 m.a.s.l.) weather stations (SwissMetNet, meteosuisse). These two stations were preferred from other station located higher because of their weak influence of topographic barrier effect of the southern alpine ridge on the measured temperatures (i.e. compared to the *Col du Grand St-Bernard* meteosuisse weather station). The obtained temperature gradient was  $-0.54^{\circ}\text{C}/100\text{m}$ , which is similar to the one calculated by Magnin et al. (2015) for the Mont Blanc region (France,  $-0.54^{\circ}\text{C}/100\text{m}$ ).

According to Fallot et al. (2012), at the local scale cold air only cumulates in large valley bottoms, such as the Rhône Valley. In the lateral valleys the temperature variability at this scale can be neglected. In turn, at the micro scale, micro-climates resulting from the influence of the relief on local temperature variability exist but are generally hard to model just from the interpolation of data collected by meteorological stations. More advanced modelling strategies should be employed (i.e. Pozdnoukhov et al., 2009) for this purpose. For these reasons, in this study it was decided to compute the MAAT using the methodology proposed by Bouët (1978), even though the resulting predictor ends up being extremely correlated to the altitude.

*Aspect*: the terrain orientation is also considered extremely relevant for the permafrost presence or absence. Surfaces with different aspects have different energy inputs due to a different radiation angle. The amount of energy per unit area can actually vary in a ratio of 1 to 10 between the sunny and the shady sides of a relief. Since terrain orientation is a circular variable (between  $0^{\circ}$  and  $360^{\circ}$ ), two uncorrelated indicators were considered: the *northness*, which corresponds to the cosine of aspect angle, and the *eastness*, equal to the minus sine of the aspect value (see Brenning and Trombotto, 2006). This allows comparing very similar orientations in the geographical space, but very sparse in terms of aspect values, such as for example  $1^{\circ}$  and  $359^{\circ}$ . These values correspond indeed to a very similar aspect (North), but

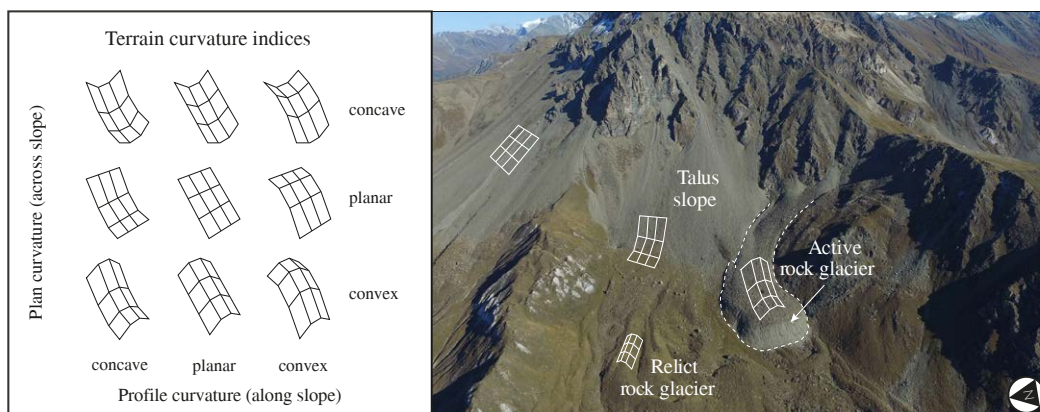


would be treated very differently by the employed machine learning algorithm if not transformed in *northness* and *eastness*.

*Slope*: the terrain slope angle influences the permafrost occurrence by governing the direct solar radiation that reaches the ground surface. In addition, the snow cover may have a different thickness depending on the slope: rock walls are generally snow free, while flat areas and footslopes may be covered by a thick snow cover (Mittaz et al., 2002).

*Potential incoming solar radiation*: the amount of energy reaching the ground was calculated as potential incoming solar radiation (PISR). For this purpose the ArcMap *Area Solar Radiation* module was used on the 2-metre DEM, taking into account the shadowing effect of the relief. The transmissivity and diffusivity were set at 0.4 and 0.5 respectively. The PISR was computed for the snow free period, between July 1<sup>st</sup> and October 31<sup>st</sup> (usually corresponding to the period without the snow cover), because the PISR does not have relevant effects on the ground temperature when a snow layer is present (Hoelzle, 1994).

*Morphological predictors*: as mountain permafrost is a thermal phenomenon, it is only observable with the appearance of geomorphological indices, such as rock glacier lobes or with in-situ measurements. Thus, taking into account *terrain curvature indices* helps recognize the presences of lobes that are potentially occupied by permafrost. The Gaussian terrain curvature (the derivative of the slope angle) was computed at different window sizes (10, 30, 50, 100, 500, 1000 metres) and was included into the dataset as an indicator of various landforms. Large window curvatures (i.e. concavities) are intended to help machine learning algorithms to detect the presence of valley bottoms. Conversely, small window terrain curvatures are indicative of lobes, crests or small depressions (figure 3.7).



**Fig. 3.7.:** Terrain curvature types and associated landforms for Les Cliosses sector, with an active rock glacier indicating the presence of permafrost (Hérens Valley, Valais; photo: S. Rüttimann).

Three types of curvature indices were computed: *planform curvature* (also called *plan curvature*), *profile curvature* and the *combination* of both indices. The *Planform curvature* indicates the curvature of a terrain perpendicularly to the direction of the maximum slope. Positive values refer to sidewardly convex surfaces (e.g. a moraine crest), whilst negative values indicate concave sidewardly surfaces (e.g. a debris flow channel). In other words, they control the flow divergence or convergence. The *Profile curvature* illustrates the curvature of a surface parallel to the direction of the maximum slope. Positive values show an upwardly concave surface (e.g. a talus slope), while negative values illustrate an upwardly convex terrain (e.g. a rock glacier front). These two types of curvature indices can also be expressed by their *combination*. In this study, the planform curvature will often be abbreviated as *pcurv*, the profile curvature as *profcurv* and the combined curvature as *curv*.

### 3.3.2 Land-cover predictors

Land-cover predictors were computed in order to identify areas where permafrost is generally completely absent or potentially present. The delimitation of these sectors allows defining a *validity domain* (see section 3.4), namely the area where the prediction is performed. The following land-cover predictors were considered:

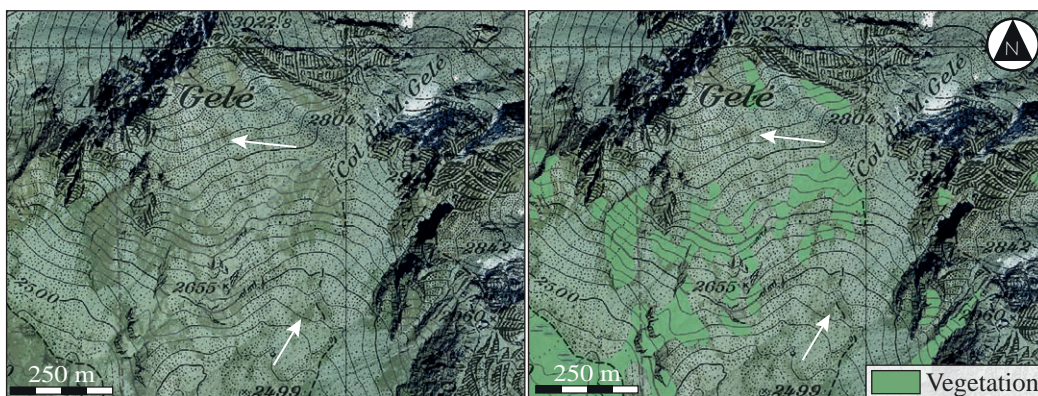
*Glaciers*: glaciated surfaces were derived from the swisstopo primary surface map (VECTOR25, last update: 2006). Unluckily, no information is provided concerning the type of glacier regime. It is thus impossible to know whether the regime of a glacier is polythermal or cold without proper in-situ investigation. The first type of glacier is generally indicative of permafrost absence due to thermal perturbation at its base, whilst the second type may be related to permafrost presence due to negative temperatures that are maintained at its base (see 2.2.1). For need of generalization and simplicity, all glaciers of the region of interest were considered as an indicator of permafrost absence and thus subjected to be excluded from the validity domain. In other words, no permafrost occurrence is predicted for surfaces covered by glacier in the context of this study.

*Glacier forefields*: these surfaces were obtained by subtracting the current glacier extent from the Little Ice Age glacier extent. The latter was first reconstructed by Maisch (2000) who compiled the extension of the 1850 glacial maximum for the entire Swiss Alps. The LIA extent resulted from different sources: on the one hand, the digitalization of the Swiss Dufour map that was produced between 1839 and 1862 and thus representing the LIA maximum of most of the glaciers (Ivy-Ochs et al., 2009); on the other hand, it resulted from the digitalization of the Siegfried maps dating from the 1870s. Paul (2004) lately completed and corrected this inventory. As illustrated in section 2.2.1, in glacier forefields permafrost is often absent in sectors occupied by glaciers during the Little Ice Age. However, permafrost may be present

locally, especially at their margins. This is particularly true for small glaciers that were located above the current PLL and from which some buried ice may still exist in sediments (see Lukas, 2011; Lilleøren et al., 2013; Bosson et al., 2015; Bosson and Lambiel, 2016).

*Hydrological network:* Lakes were extracted from the swisstopo primary surface map and compared to recent orthophotos. River channels were extracted from the swisstopo hydrological network. As observed by Haerberli (1996), permafrost is generally absent where water runs or exists. Indeed, negative ground temperatures cannot be maintained where lakes and permanent rivers are present. In addition to the mapped river channels, a buffer of 10 metres in the nearby areas was considered as permafrost free to account their potential lateral migration.

*Vegetation and Normalized Difference Vegetation Index:* vegetation was not derived from the swisstopo primary surfaces because these are extracted from topographic maps and their accuracy does not always match to reality. Indeed, they do not always represent the exact ground truth, especially when compared to the most recent high resolution orthophotos. Figure 3.8 illustrates the inaccuracies of the delimitation between vegetation and rock surfaces that can occur when employing the primary surfaces (as arrows indicate). In this example, the South side of the Mont Gelé peak is presented. The majority of the slope is characterized by debris-covered surfaces at the exception of some patches of vegetation. The vegetation, resulting from swisstopo primary surface map (right image in figure 3.8), does not follow the real limit between vegetation and debris-covered surfaces. As a result, at the micro scale the cartography of vegetated surfaces presents inaccuracies.



**Fig. 3.8.:** Orthophoto - primary surface map comparison for the Mont Gelé sector (orthophoto, topographic and primary surface maps: swisstopo).

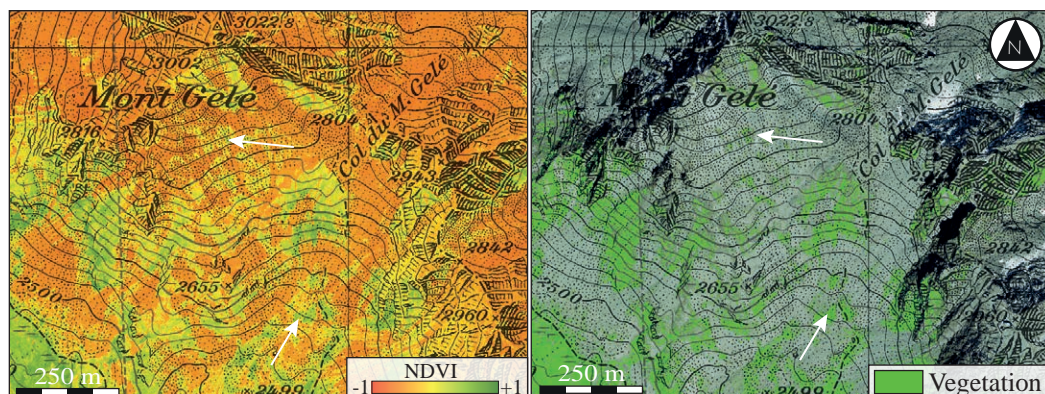
The Normalized Difference Vegetation Index (NDVI) was calculated from false-colour infrared images (FCIR, 0.25 metres of resolution) to detect the correct limit between vegetation and rock surfaces. These images have three channels, namely red, blue and near infra-red (R, B, NIR) and were recorded with the airborne digital sensor

(ADS) cameras by swisstopo. The NDVI is an index calculated on near infra-red and red bands of an image to quantify the active vegetation (Kriegler et al., 1969; Crippen, 1990). Indeed, the NDVI is computed from the reflectance in the red and near infra-red bands of the electromagnetic spectrum and it is widely adopted in remote sensing to measure the photosynthetic activity (Justice et al., 1985; Tucker and Sellers, 1986; Fontana et al., 2008). It is commonly calculated as follows:

$$NDVI = \frac{NIR - R}{NIR + R} \quad (3.2)$$

The index values have a range between -1 and 1. Cited references propose a threshold for different kind of vegetation. Usually, low values of NDVI (0 and below) correspond to barren areas of rock, sand, or snow. Moderate values represent grassland (0 to 0.3), while high values correspond to temperate and tropical rainforests (0.6 to 0.8).

The NDVI variable was included in the dataset as a continuous predictor and employed to differentiate the vegetation from other types of land cover, as well as to characterize the local variability of sediment textures. In this research, NDVI had indeed this dual use and was preferred to the Soil-Adjusted Vegetation Index (SAVI) that would have been more specific to vegetation alone.



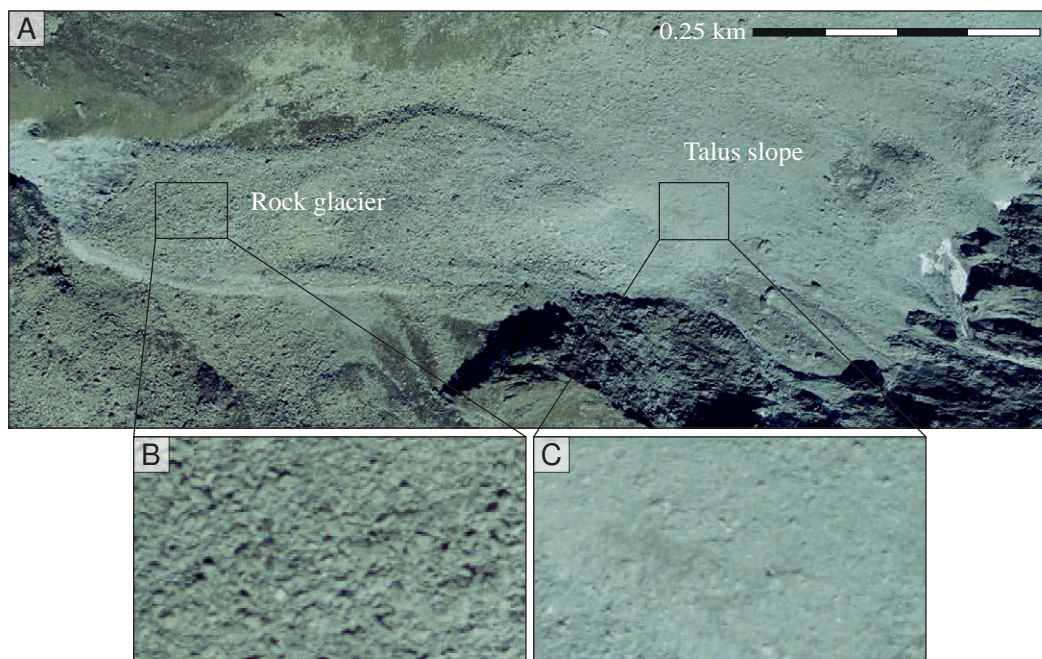
**Fig. 3.9.:** NDVI map of the Mont Gelé sector and delimitation of the vegetation extent (orthophoto, FCIR images and topographic map: swisstopo).

Although the limit between vegetation and rock surfaces was identified from the minimum of the histogram distribution of all the computed NDVI values (around 0.05), a slightly higher value was employed to avoid the bias of a subjective choice or of a wrong thresholding due to inaccuracies in the FCIR values (resulting from the different time of year in which the images were taken). As a consequence, NDVI values less than 0.1 were considered for texture identification (and consequently, values greater than 0.1 for the delimitation of vegetation).

Figure 3.9 illustrates the extent of the vegetation obtained by thresholding the NDVI map. The figure shows all the pixels with a NDVI value greater than 0.1. After a qualitative verification, the extent of the vegetation at the micro scale appeared to be more reliable to the field reality compared to the extent offered by the primary surface map.

### 3.3.3 Grain size

In geomorphology, *grain size* (or granulometry) refers to the diameter of individual blocks in sedimentary materials (Wentworth, 1922; Cailleux and Tricart, 1959). It is one of the main factors that controls the ground (surface) temperature and thus the permafrost occurrence in mountain areas (i.e. Sawada et al., 2003; Gruber and Hoelzle, 2008). As illustrated in section 2.2.1, a coarse blocky terrain can be affected by air advection processes that can provoke strong spatial variations of the ground temperatures (e.g. Morard et al., 2010; Rödder and Kneisel, 2012). The automatic detection of grain size by using remote sensing techniques may therefore help to improve models of the mountain permafrost extent.



**Fig. 3.10.:** (A) Comparison of grain sizes for two sectors with coarse material on a rock glacier (B) and finer material at its roots (C).

The block edge-detection could allow detecting rock boulder contours. However, this technique is hardly applicable to orthophotos because of their coarse resolution (see figure 3.10.C). A single pixel may refer to a single block and its contours would not be correctly recognized. For this reason, alternative approaches had to be investigated. If the reference scale is fine enough, granulometry can be associated to the ruggedness of the terrain and the grain size classes can be obtained by performing a surface roughness analysis (Grohmann et al., 2011). Different techniques can

be applied to a high resolution DEM in order to collect roughness information. The most used one is the Topographic Position Index which is considered as a terrain ruggedness metric and a local elevation index (Jenness, 2002). The standard deviation of elevation is another measure of the topographic roughness (Haneberg et al., 2005; Ascione et al., 2008; Grohmann et al., 2011) and it was successfully used for grain size characterization by Otto et al. (2012). In their study, they quantified the surface roughness by using the standard deviation of the residual topography which results from the difference between the original and a smoothed DEM. Without being influenced by the local slope, the residual topography corresponds to the local variation between the lowest and the highest point in the neighbourhood of each analysed pixel. The topographic roughness may also be expressed as the slope convexity, the variability of plan convexity or the standard deviation of the slope (Ruszkiczay-Rüdiger et al., 2009). However, these indices are strongly dependent on the resolution of the DEM in order to express granulometry as a function of roughness.

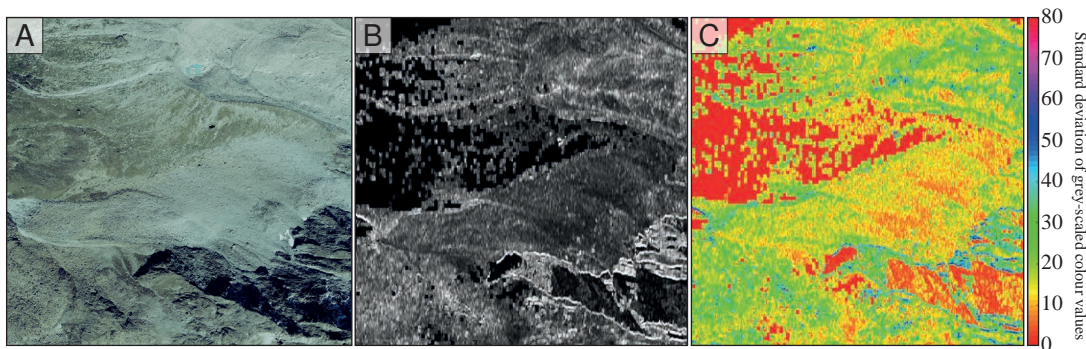
In Deluigi and Lambiel (2014) a different approach was tested to automatically detect the *grain size*. In a first step, a dataset was built around four texture operators (Euclidean mean distance, variance, kurtosis and skewness; see Petrou and Sevilla, 2006) that were calculated on the false-color infra-red (FCIR) images provided by swisstopo (resolution of 0.25 metres). The texture operators were computed for different pixel windows (3x3, 5x5, 7x7, 9x9). These operators provide useful information about the variation of the grey values or the variation of the spectral radiance, which are related to the variation of the ground surface characteristics. Fine materials have similar spectral properties within the chosen pixel window; therefore coarse blocks have a high spectral variability due to different rock properties, shadow, pores, etc. In a second step, two training classes corresponding to zones with small grain size cobbles and high grain size blocks were provided to a Support Vector Machine (SVM) classifier in order to predict the granulometry in sectors where it was unknown. The resulting classification allowed mapping sedimentary materials with regards to their grain size. Nevertheless, this approach showed some limitations because, on the one hand, the results were not accurate for sectors affected by shadow or snow. On the other hand, it required a lot of computational power and time to run and training examples were collected manually on the field or from the interpretation of orthophotos, making it biased or not suitable for granulometry detection at a large geographical scale. The obtained result was indeed satisfactory for small sectors, but hardly applicable to the extent of the region of interest.

As an alternative to the real ground granulometry, the grain size was derived by simply applying a local standard deviation filter to the orthophoto pixel grey-scaled colours. Computing a moving window filter on the image provided a standard deviation value to the central pixel of the chosen window:

$$stdfilt = \sqrt{\frac{\sum(x_i - \bar{x})^2}{r \cdot c - 1}} \quad (3.3)$$

where  $x_i$  were the pixel values of the grey-scaled images,  $\bar{x}$  the mean of the pixel values considered by the filter,  $r$  and  $c$  the size of the filter in rows and columns.

In the resulting texture map, rock surfaces with a fine grain size correspond to low values of the standard deviation. Conversely, coarse blocks are represented by higher standard deviation values. This is linked to the poor difference in colour that fine materials (or bedrock outcrops) have on an orthophoto at 25 centimetres of resolution. Indeed, one single pixel often covers several rocks and the near pixels have similar grey values because of the absence of shadows between the sediments. Therefore, orthophotos showing sectors with coarse blocks generally present heterogeneous values of the grey values because of the presence of shadow between the blocks, consequentially resulting in higher values of the standard deviation filter. To consider blocks with a diameter up to 5 metres, a window size of 21 pixels was selected. Figure 3.11 compares the ground truth with the computed sub-grain size variable. 3.11.C illustrates the distribution of the standard deviation values at a resolution of 25 centimetres. Only pixels corresponding to the rock surfaces were considered and added to the dataset to serve for the classification of permafrost presence and absence.



**Fig. 3.11.:** (A) Orthophoto of the Tsarmine sector (Arolla), with various landforms and debris-covered glacier (swisstopo). (B) Gray-scaled image of the sector (25 centimetres of resolution; swisstopo). (C) Computed grain size of the sector, obtained by applying a standard deviation filter on the grey-scaled image.

## 3.4 Validity domain

The so-called *validity domain*, a delimited area in which predictions were run, was defined based on categorical variables and the NDVI. The prediction of the permafrost occurrence was only computed for sectors where the latter is uncertain (the prediction can be either *permafrost absence* or *permafrost presence*), typically in rock surfaces such as rock glaciers, rock walls, talus slopes, moraine deposits and

other various debris-covered surfaces and bedrock outcrops. Areas where permafrost is generally absent, such as glaciers and glacier forefields, rivers, lakes or vegetated sectors, were excluded from the validity domain (table B.2). As mentioned in section 3.3.2, the NDVI was used both for the validity domain definition as well as for classification. A threshold of 0.1 was selected for the delimitation of the vegetation from other surfaces. As it was shown, vegetation is indeed generally present above this threshold and thus areas with a NDVI greater than this value were excluded from the validity domain. However, values below 0.1 of this predictor inform about the reflectance of the ground and corresponding textures. For these reasons, the NDVI was kept in the dataset as a predictor helping classify the permafrost extent.

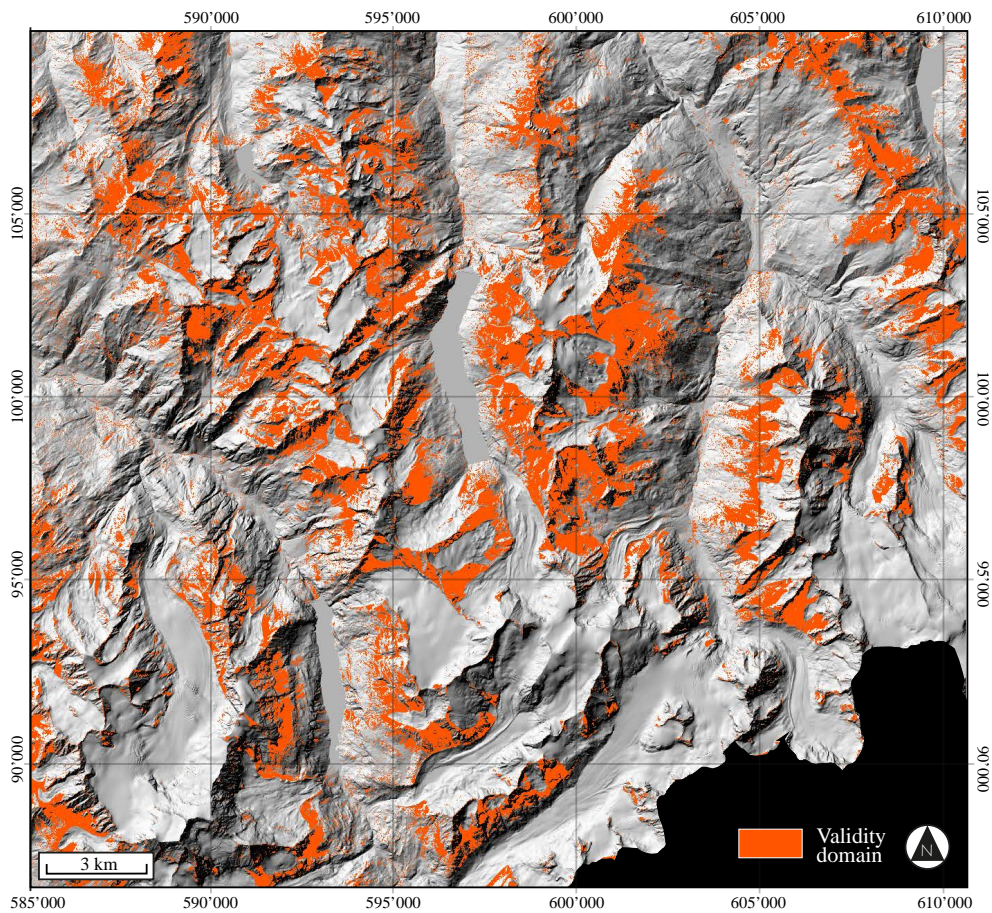
Variable name	Abbreviation	Variable type	Use
Altitude	[altitude]	Continuous	Classification
Northness	[northness]	Continuous	Classification
Eastness	[eastness]	Continuous	Classification
Slope	[slope]	Continuous	Classification
PISR	[radiation]	Continuous	Classification
MAAT	[maat]	Continuous	Classification
Lake	-	Categorical	VD definition
Rock wall	-	Categorical	VD definition
River	-	Categorical	VD definition
NDVI	[NDVI]	Continuous	Classification / VD definition
Grain size	[grain]	Continuous	Classification
Planform curvature 10 m	[pcurv10]	Continuous	Classification
Planform curvature 30 m	[pcurv30]	Continuous	Classification
Planform curvature 50 m	[pcurv50]	Continuous	Classification
Planform curvature 100 m	[pcurv100]	Continuous	Classification
Planform curvature 500 m	[pcurv500]	Continuous	Classification
Planform curvature 1000 m	[pcurv1000]	Continuous	Classification
Profile curvature 10 m	[profcurv10]	Continuous	Classification
Profile curvature 30 m	[profcurv30]	Continuous	Classification
Profile curvature 50 m	[profcurv50]	Continuous	Classification
Profile curvature 100 m	[profcurv100]	Continuous	Classification
Profile curvature 500 m	[profcurv500]	Continuous	Classification
Profile curvature 1000 m	[profcurv1000]	Continuous	Classification
Combined curvature 10 m	[curv10]	Continuous	Classification
Combined curvature 30 m	[curv30]	Continuous	Classification
Combined curvature 50 m	[curv50]	Continuous	Classification
Combined curvature 100 m	[curv100]	Continuous	Classification
Combined curvature 500 m	[curv500]	Continuous	Classification
Combined curvature 1000 m	[curv1000]	Continuous	Classification
Glacier forefield	-	Categorical	VD definition
Glacier	-	Categorical	VD definition

**Tab. 3.1.:** Features contained in the raw dataset and their abbreviation, type and use. Categorical variables served at constructing the validity domain (VD). Continuous variables were used for classifying the permafrost occurrence.

The collected permafrost evidences were mainly derived from investigations made in *debris-covered surfaces*. In comparison, a poor record of permafrost evidences in rock walls unfortunately exist for the region of interest. This lack of permafrost training examples, informing about the known permafrost occurrence and its spatial characteristics within these sectors, did not allow to efficiently employ machine learning algorithms to classify permafrost data in these environments. For this reason, rock wall surfaces, corresponding to bedrock surfaces with a slope greater than 40°, were excluded from the validity domain. As a consequence, the focus of this research was the prediction of the permafrost occurrence in remaining mineral surfaces. As Pogliotti (2011) indicates, a threshold of 40° is the upper slope angle



limit where usually snow-covered areas are located and thus also debris-covered surfaces. No delimitation was however used to distinguish bedrock outcrops with slopes lower than 40° from debris-covered surfaces. In turn, as the grain size is expressed as the standard deviation of the grey values of very detailed orthophotos, bedrock outcrops should fall into the sectors characterized by low granulometry. A coarse debris covered-surface is generally more favourable for permafrost conditions than bedrock outcrops located at the same altitude. Therefore, the distinction between these two types of landcover was implicitly provided by the grain size values and the expected predicted probabilities of permafrost occurrence should be lower for bedrock outcrops.



**Fig. 3.12.:** The extent of the validity domain, delimiting the surfaces where the prediction of the permafrost occurrence was performed (hillshade: swisstopo).

In figure 3.12, an overview of the portion of territory corresponding to the validity domain is presented, covering an area of around 116 km<sup>2</sup>. The permafrost dataset was finally built by considering all environmental predictors and permafrost evidences mentioned in this chapter on a regular grid of 10x10 metres. A spatial resolution of 10 metres was selected according to the computational requirements needed to predict the permafrost distribution for such an extended study area. The minimum size of typical periglacial landforms (such as rock glacier or moraine

ridges) was also considered. Indeed, these landforms could have been represented with at least 2 pixels (a width of 20 metres). A higher resolution of the dataset would have introduced unexpected noise, both in the predictors (i.e. erroneous attribute values) and the class (label noise, i.e. contradictory examples or misclassifications). This would have required an additional preprocessing and a noise filtering analysis to reduce the noise effects. Moreover, due to the complexity of the studied phenomenon, from a physical perspective willing to map the permafrost distribution at a too high resolution was illusory without considering sub-surface characteristics (e.g. porosity; hardly mappable for large areas). Consequently, the original DEM raster and relative extracted predictors controlling the presence and absence of mountain permafrost were resampled by computing the mean value of the 2x2 metre cells (for DEM-derived variables) and 0.25x0.25 metre cells (original resolution of the NDVI and grain size).



## Modelling approach

### 4.1 Machine learning for environmental data analysis and modelling

Machine learning (ML) can be considered as a sub-field of artificial intelligence and it is concerned with the design, the development and the application of algorithms and techniques, allowing computers to learn from data (Vapnik, 1998; Bishop, 2006; Kohonen, 2000; Cherkassky and Mulier, 2007; Kanevski, 2008; Hastie et al., 2009; Haykin, 2009). It is closely connected with both statistics (especially non-parametric and computational statistics) and theoretical computer science. From the mid 20<sup>th</sup> century, machine learning has evolved from the imitation of a simple neuron and artificial neural networks to a solid interdisciplinary field of basic and applied research, having an important impact on many topics: pattern recognition, bio-computing, speech recognition, financial applications, analysis and modelling of high dimensional and multivariate geo- and environmental spatio-temporal data, etc. (Vapnik, 1998; Cherkassky and Mulier, 2007; Hastie et al., 2009; Izenman, 2008; Kanevski, 2008). In recent years, there has been a significant growth in the development of *adaptive and data-driven approaches*. Ensemble models, artificial neural networks (ANN) of different architectures and Support Vector Machines have attracted great attention among the most successfully and widely used techniques of ML. The key feature of the ML algorithms is that they *learn from data and can be used in cases where the modelled phenomena are not very well described*, which is the case in many applications of geospatial data. The ML models are adaptive tools that are currently widely used to solve prediction, characterisation, optimisation and many other problems. Multilayer perceptron (MLP), radial basis function (RBF) networks, general regression neural networks (GRNN), probabilistic neural networks (PNN), Kohonen networks (self-organising maps, SOM) (Cherkassky and Mulier, 2007; Hastie et al., 2009; Izenman, 2008; Haykin, 2009; Kanevski et al., 1996; Kanevski, 1999) are just few examples of algorithms among the most employed ones in geo- and environmental sciences.

In general, geospatial data do not only represent data in a geographical low dimensional (2D, 3D) space, but also data embedded into high dimensional geo-feature spaces, consisting of geographical coordinates and features generated from, for example, digital elevation models, science based models, remote sensing images,

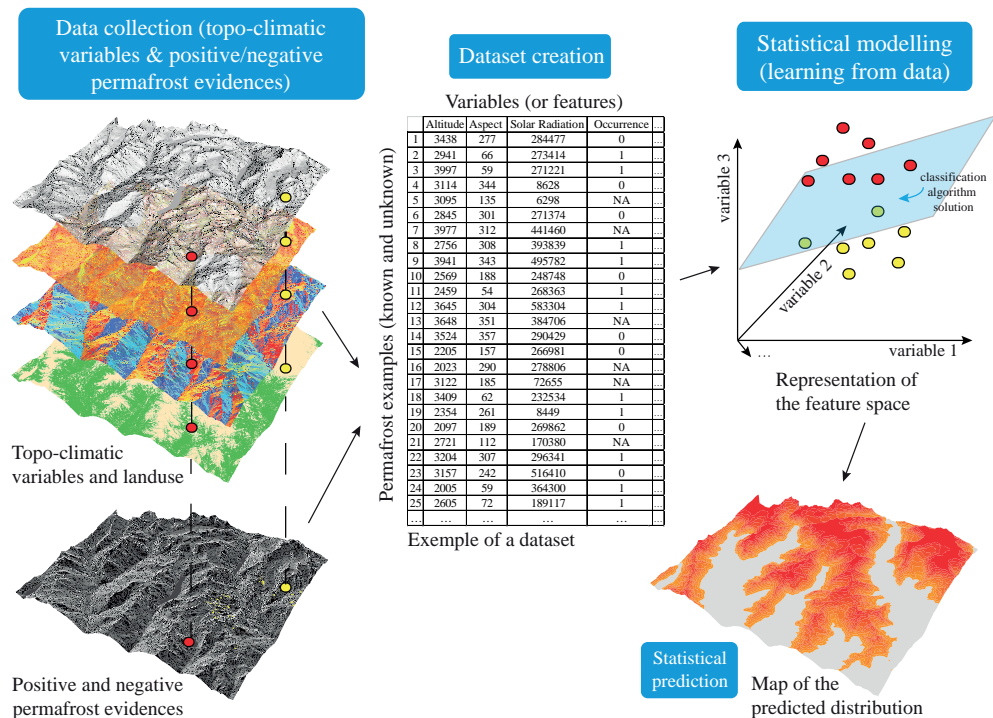
etc. Typical environmental problems can easily reach several dozens (permafrost, avalanches, wind fields, landslides) or even hundreds of dimensions (remote sensing hyperspectral images). In such cases, an important problem is the *curse of dimensionality*, an exponential growth (with the dimension of space) of measurement data necessary to fill the space (Hastie et al., 2009). Therefore, dimensionality reduction methods have gained in popularity and many efficient non-linear dimensionality reduction techniques have been efficiently used in many real life applications (Guyon et al., 2006; Lee and Verleysen, 2007).

#### 4.1.1 General overview of the adopted approach

In order to efficiently map the spatial distribution of mountain permafrost by using machine learning classification algorithms, different interdisciplinary techniques, such as geomorphological field methods, geographical information sciences, remote sensing, statistics and environmental data mining, were combined in this research. Techniques belonging to GIScience, geomorphology, spatial statistics and remote sensing were employed to produce a complete set of data. Figure 4.1 illustrates the general workflow of this study. Selected predictors were coupled with permafrost evidences presented in section 3.2 and constitute the raw dataset. The latter generally takes the form of a matrix containing the spatial information that is associated to the grid pixels covering the region of interest. Instances of the dataset representing permafrost evidences had a known label (typically permafrost *presence* or *absence*) and were employed as *training examples*. In turn, observations with unknown occurrence required to be predicted and constituted a prediction grid. The final permafrost map was obtained by employing a decision function on the prediction grid, function that resulted from the classification of the permafrost occurrence made by the ML algorithms and based on the training examples.

One of the key steps of the experimental setup (figure 4.2) of this research was the pre-processing of the raw dataset. Indeed, data exploration and data reduction techniques were considered before learning the functional dependencies between the permafrost distribution and the adopted features. The first step aimed at understanding the dataset structures through visual exploratory data analysis (EDA). In general, this approach does not include spatial modelling and adopted techniques are mostly graphical. Additional techniques are thus often needed to detect spatial patterns in data and to understand their spatial proprieties. The exploratory spatial data analysis (ESDA) satisfied these needs and allowed formulating hypotheses based on the spatial distribution of the data. By finding structures (e.g. clusters), ESDA provided the knowledge required to assess and employ a representative portion of data to be successively adopted (*sampling*). In a second step, after pre-processing

and the discretization of raw data, the importance of the retained variables was determined. The reduction of the dataset dimensionality was then obtained by using feature extraction and feature selection algorithms, while its quality was assessed with a machine learning classifier.

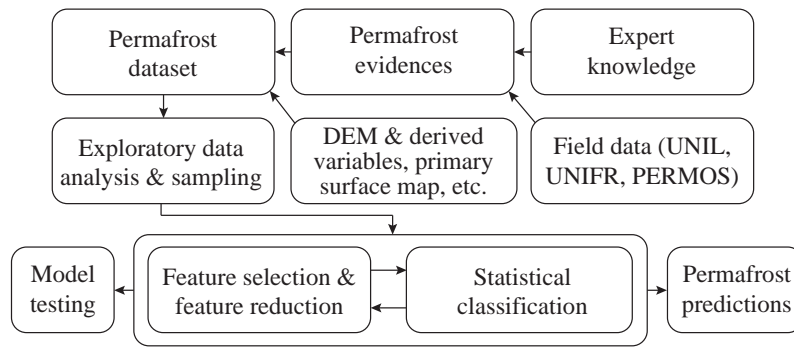


**Fig. 4.1.:** The modelling workflow of this study, from data collection to statistical prediction.

Once the best subset of predictors was selected, permafrost data were classified and the employed techniques were tested and compared. The inferred decision function was finally employed to produce permafrost distribution maps and the corresponding uncertainty graphs and maps (figure 4.2). The following sections will detail the theoretical framework behind mentioned steps by presenting an overview of the approaches used.

## 4.2 Exploratory data analysis

The physical limitation in easily displaying the high number of variables and observations of the employed permafrost dataset (26 variables and thus 26 dimensions) was bypassed by using specific techniques adapted to the visualization of high-dimensional datasets. These approaches generally range from dimension reduction to visual encoding, from interactive exploration to quantitative analysis (see Liu et al., 2017). Two well-used data visualization techniques are presented here to explore the complexity of permafrost data and to highlight unravelled structures: parallel coordinates (PC) and t-Distributed Stochastic Neighbour Embedding (t-SNE).



**Fig. 4.2.:** Overview of the modelling steps: the raw dataset, containing permafrost evidences and environmental predictors, was analysed and resampled. Its dimensionality was then reduced before applying classification algorithms and producing predictions of the permafrost occurrence.

PC are today very popular as they allow exploring high dimensional sets of data efficiently (Inselberg and Dimsdale, 1990; Inselberg, 2009). The  $n$ -dimensions of the dataset are displayed as the  $n$ -vertical axis of a chart, representing each variable. Each observation of the dataset is represented as a polyline with the vertices lying on the parallel  $n$ -vertical axes. The data points with similar values for a given variable are displayed in a similar position within the axis. An interpretation of the relationship between the dimensions can then be made thanks to the visual representation of the different polylines. As Inselberg (1997) indicates, parallel polylines between two axis generally indicate a positive relationship between the two corresponding dimensions. Conversely, x-crossing polylines usually represent a negative relationship, while randomly crossing polylines show no particular relationship and a complex dataset.

The t-SNE technique (Maaten and Hinton, 2008) has been extremely used for the visualization of high-dimensional datasets (i.e. Jamieson et al., 2010; Mohamed et al., 2012; Wilson et al., 2015). It produces 2D embeddings of the original dataset vectors in which observations close to each others in the high-dimensional space are also close in the 2D space. The algorithms convert the  $d_{ij}$  pairwise distances of the original dataset into joint probabilities  $d_{ij} \propto \exp(-d_{ij}^2)$ . The corresponding observations lying in the 2D space are then computed with an iterative search that aims at finding a similar set of joint probabilities.

In addition to data visualization, assessing the complexity of permafrost data allows deciding the best classifier to adopt. Indeed, it is crucial to investigate the non-linearity of the dataset by previously applying simple classifiers to avoid useless computations or poor generalizations.

To determine the potential non-linearity of permafrost data, the latter were classified with both a linear Support Vector Machines (SVM) algorithm and one of its non-

linear extensions, that is the radial basis function kernel SVM (RBF-SVM). If the classification problem is linear, the first technique is generally able to infer, directly in the original dataset feature space, an optimal decision function. Conversely, if the data are not linear, the classification performance of the linear SVM is poor and it increases with the application of a RBF-SVM. In fact, when applying a non-linear SVM, the assumption that data are not linearly separable in the original feature space is legitimately made and data have to be remapped in a high-dimensional space in order to be classified (see Deluigi et al., 2017). By computing the *index of complexity* proposed in Leuenberger and Kanevski (2017), a value of the non-linearity can also be assessed:

$$I_C = \frac{MSE_l - MSE_{nl}}{MSE_l} \quad (4.1)$$

where  $MSE_l$  is the mean squared error of linear SVM and  $MSE_{nl}$  is the mean squared error of the non-linear expansion of the classifier. This index is expressed into the interval  $[0, 1]$  when  $MSE_{nl} \geq MSE_l$ . Values near a  $I_C$  of 0 indicate complexity in the data, whilst a  $I_C$  that tends to 1 shows linearity in the data.

To assess the presence of structure in the data, it is possible to take advantage of the properties of an adaptive (anisotropic) general regression neural network (AGRNN). It is based on non-parametric statistics and it is an extension of the isotropic GRNN (Specht, 1991; Specht, 1994). This algorithm is similar to the probabilistic neural network (Specht, 1994), which was developed for classification problems. A detailed description as well as real data case studies can be found in Kanevski et al. (2009) and Robert et al. (2013).

In the general form, the AGRNN prediction  $f(\vec{x})$  in a high dimensional space ( $m$ ) is described as follows:

$$f(\vec{x}) = \frac{\sum_{i=1}^N f_i \exp[-\sum_{k=1}^m \frac{D_{\vec{x},i,k}^2}{2\sigma_k^2}]}{\sum_{i=1}^N \exp[-\sum_{k=1}^m \frac{D_{\vec{x},i,k}^2}{2\sigma_k^2}]} \quad (4.2)$$

where  $N$  is a number of points,  $D_{\vec{x},i,k}$  is a contribution to the distance between a point of prediction and  $i$ -th measurement from the  $k$  variable and  $\sigma_k$  are the bandwidths for the  $k$ -th variable. In this study, an Euclidean distance was used.

The difference between GRNN and AGRNN is that in case of AGRNN for each independent (input) variable  $m$ , a different  $\sigma$  (kernel bandwidth) is fitted during the training procedure. Adaptive models are very powerful in the modelling of complex phenomena, but they demand more training time and advanced optimization algorithms (Kanevski et al., 2009; Robert et al., 2013). In the present research, both leave-one-out and data splitting training/validation were applied to find correct bandwidths. The optimal values of *sigmas* correspond to the minimum of the vali-



dition error. If the bandwidth is very large (larger than the region of the study, i.e.  $\frac{D^2}{2\sigma_k^2} \ll 1$  for all variables and distances), there is no minimum on validation curve and the solution converges to the mean value. This result indicates that there are no structured patterns in data. In order to verify that structures are present in the dataset, this property can be verified by shuffling the dataset, which destroys structured information in data even if it is available. As a consequence of the shuffling, if the error curve of shuffled data modelling is always above the level of training data variance and it does not reach a minimum, it is generally possible to assess that structures are present in the dataset.

## 4.3 Feature relevance

### 4.3.1 Dataset dimensionality

Dimensionality reduction is a common practice when dealing with machine learning problems, allowing the transformation of data into more compact and predictive ones. Two approaches can be distinguished: feature extraction (FE) and feature selection (FS). FE aims at mapping the original list of features onto a new set of data by combining subsets information that share the same statistical characteristics into synthetic attributes. Typical examples are the Principal Component Analysis (PCA) or the Linear Discriminant Analysis (LDA). Conversely, with FS, a subset of the original attributes is selected in accordance to a quality metric. FS is simpler and less time consuming than FE, as the metric is only evaluated for each feature, that can then be ranked. By thresholding the quality metric, invaluable attributes can successively be discarded or, alternatively, top ranking features can be selected by removing a percentage of the original set variables (e.g. 5%, 10% or 15% are typical numbers). Moreover, by using search algorithms, more time consuming alternatives can be employed to investigate the predictive power of subset of features (e.g. backward feature elimination), rather than using an univariate approach.

The main differences between FE and FS are that, on the one hand, FE produces synthetic features rather than keeping the original ones. As a consequence, experts may have difficulties in interpreting the results because they appear physically meaningless, although these synthetic variables are statistically important for the studied phenomenon. On the other hand, FE belongs to the unsupervised learning domain, which aims at reducing the number of attributes without making use of the class (i.e. permafrost presence and absence). In contrast, FS is a supervised approach and it should be more adapted to lead to a more predictive set of reduced features than the original one.

In the next sections, both dimensionality reduction algorithms (feature extraction and feature selection) will be presented.

### 4.3.2 Feature extraction

The adopted dataset was constituted by a total of 26 predictors (see table B.2) and consequently the input space had a dimension  $d = 26$ . This means that the data are embedded in  $\mathbb{R}^d$  but it may not have a dimension  $d$ . Indeed data points may not occupy the full  $d$ -dimensional space and the dimensionality of the dataset may be represented by the minimum number of variables representing the data without losing information (Camastra, 2003). In other terms, the dataset  $\Omega \subset \mathbb{R}^d$  has an Intrinsic Dimensionality (ID) of  $M$  if the data points lay within a  $M$ -dimensional subspace of  $\mathbb{R}^d$  (Fukunaga, 1982). When trying to use a classifier on a high dimensional dataset, the *curse of dimensionality* can increase computation times or make the algorithm less reliable (Bellman, 1961). Furthermore, in accordance with one of the main principles of the statistical learning theory (Vapnik, 1998), the generalization ability of a classifier depends on the ID. Discovering the ID of a dataset is thus a very important task.

Several techniques of feature extraction exist. Some of the most common ones are Principal component analysis (PCA), Isomap or Linear discriminant analysis (LDA). With such algorithms, the dataset space is reduced to a lower dimensional space whose dimension is equal to the ID of the original set of data. For this study, a Morisita Index was employed to evaluate the ID of the permafrost data (see Golay and Kanevski, 2015; Golay and Kanevski, 2016), implemented in the *R* package IDMining (Golay and Laib, 2016). The resulting ID was used as a target dimension for the reduced dataset that was obtained with feature extraction techniques. In this study, kernel PCA, Locally linear Embedding and Sammon maps were used. The following sections will briefly illustrate their theoretical formulation.

#### Kernel principal component analysis

The Kernel Principal Component Analysis (kPCA) is a non-linear extension of the standard PCA (Mika et al., 1998; Schölkopf and Smola, 2002). It deals with the idea that data can be mapped in a higher dimensional feature space  $F$  before performing a standard PCA. The mapping of  $n$  data points  $x_i \in \mathbb{R}^d$  is given as:

$$x_i \rightarrow \Phi(x_i) \tag{4.3}$$

The standard PCA can be applied in the feature space  $F$  because the vectors  $\Phi(x_i)$  are computed as scalar products (Schölkopf et al., 1998) and, consequently, the mapping (4.3) can be omitted. The kPCA introduces a kernel function  $k(x, y)$  which replaces the scalar product  $(\Phi(x) \cdot \Phi(y))$ . In this extension of the standard PCA, the eigenvectors  $V$  of the covariance matrix in  $F$  are computed as linear combinations of points  $\Phi(x_i)$ , following:

$$V = \sum_{i=1}^n \alpha_i \tilde{\Phi}(x_i) \quad (4.4)$$

$$\tilde{\Phi}(x_i) = \Phi(x_i) - \frac{1}{n} \sum_{r=1}^n \Phi(x_r) \quad (4.5)$$

with  $\tilde{\Phi}(x_i)$  centred around the origin of  $F$  and  $\alpha_i$  the components of a vector  $\alpha$ , which is an eigenvector of the matrix  $\tilde{K}_{ij} = (\tilde{\Phi}(x_i) \cdot \tilde{\Phi}(x_j))$  with a length selected that allows the principal component  $V$  to have a unit length equal to 1:

$$\|V\| = 1 \Leftrightarrow \|\alpha\|^2 = 1/\lambda \quad (4.6)$$

with  $\lambda$  being the eigenvalue of  $\tilde{K}$  corresponding to  $\alpha$ .

The final computation of  $\tilde{K}$  is performed by substituting  $\tilde{\Phi}(x_i)$  and it gives the following function of the kernel matrix  $\tilde{K}_{ij}$ :

$$\tilde{K}_{ij} = k(x_i, x_j) = K_{ij} - \frac{1}{n} \sum_{r=1}^n K_{ir} - \frac{1}{n} \sum_{r=1}^n K_{rj} + \frac{1}{n^2} \sum_{r,s=1}^n K_{rs} \quad (4.7)$$

### Sammon maps

Proposed by Sammon (1969), the Sammon projection is a non-linear algorithm that maps the original dataset of  $N$   $D$ -dimensional vectors  $X_{1,i} = 1, \dots, N$  onto a lower-dimensional  $d$  (whom vectors are defined as  $Y_{1,i} = 1, \dots, N$ ) by preserving the structure of the data. The projection of the original space is considered non-linear because the extracted variables are not represented as a linear combination of the original variables (Ripley, 1996; Venables and Ripley, 2002). In the  $d$ -space, the interpoint distances  $d_{ij} \equiv \text{dist}[Y_i, Y_j]$  approximate the corresponding interpoint distances  $d_{ij}^* \equiv \text{dist}[X_i, X_j]$  in the  $D$ -space. The initial  $d$ -space configuration aims at projecting the  $D$ -dimensional data orthogonally onto a  $d$ -space spanned by the  $d$  original coordinates with the largest variance.

$$Y_1 = \begin{bmatrix} y_{11} \\ \cdot \\ \cdot \\ \cdot \\ y_{1d} \end{bmatrix} \quad Y_2 = \begin{bmatrix} y_{21} \\ \cdot \\ \cdot \\ \cdot \\ y_{2d} \end{bmatrix} \quad \dots \quad Y_N = \begin{bmatrix} y_{N1} \\ \cdot \\ \cdot \\ \cdot \\ y_{Nd} \end{bmatrix} \quad (4.8)$$

Successively, the algorithm searches the minimum of the error function  $E$  by recomputing the distances  $d_{ij}$ .

$$E = \frac{1}{\sum_{i < j} [d_{ij}^*]} \sum_{i < j}^N \frac{[d_{ij}^* - d_{ij}]^2}{d_{ij}^*} \quad (4.9)$$

$E$  is a function of the  $d \times N$  variables with  $y_{pq}$ ,  $p = 1, \dots, N$  and  $q = 1, \dots, N$ . Although its first implementation used gradient descent (Sammon, 1969), in the present work the minimization was performed iteratively.

### Locally Linear Embedding

Unlike classic feature extraction approaches, which try to preserve the pairwise distances between points (i.e. Sammon maps), the Locally Linear Embedding (LLE) algorithm aims at recovering global non-linear structures from locally linear fits (Roweis and Saul, 2000). LLE is based on the geometric intuition that  $N$  vectors  $\vec{X}_i$  of dimension  $D$  can be sampled from some underlying manifold. The algorithm assumes that similar data points lay on or close to a locally linear patch of the manifold. The geometry of these local patches is discovered by some linear coefficients that reconstruct every dataset example from its neighbours. The cost function of the reconstruction error is given as follow:

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2 \quad (4.10)$$

with  $W_{ij}$  representing the weight of the contribution of the  $j$ th point of the dataset to the  $i$ th reconstruction. These weight  $W_{ij}$  are computed by minimizing the cost function under two main constraints: (1) each  $\vec{X}_i$  data point is reconstructed from the neighbour points (selected with nearest neighbours in Euclidean distance), which enforce  $W_{ij}$  to be equal to 0 if  $\vec{X}_j$  is not linked to the neighbours points of  $\vec{X}_i$ ; (2) the sum of the rows of the weight matrix is equal to 1 ( $\sum_j W_{ij} = 1$ ). A least-squares problem has to be solved to obtain the optimal weights  $\sum_j W_{ij}$  subject to these constraints.

The dataset dimension is reduced by using the same weights  $W_{ij}$  of the  $i$ th data point in the  $D$ -dimension to reconstruct the same point in the lower  $d$ -dimension. This is done based on the assumption that the data lay exactly or approximately on a non-linear manifold in the lower dimensions  $d$ -space. Indeed, the original  $X_i$  point is mapped onto a  $Y_i$  point (belonging to the  $d$  dimension) by minimizing the following cost function:

$$\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2 \quad (4.11)$$

Conversely to the equation (4.10), with the cost function  $\Phi(Y)$ , the optimization of the dataset coordinates is made with fixed  $W_i$  weights and a minimization directly performed on the  $Y_i$  points. Complete information of the optimization problem can be find in Roweis and Saul (2000).

### 4.3.3 Feature selection

It is common practice to avoid selecting predictors linked by strong correlations. Dormann et al. (2012) suggest analysing employed variables and removing those variables that have a correlation exceeding  $\pm 0.75$ . Pairs of correlated variables can be explored by computing a correlation matrix. However, a ranking of the feature importance is required to decide which variable has to be removed from the dataset. This can be achieved by using feature selection techniques that provide a measure of the contribution of an environmental predictor with regard to the permafrost occurrence. Once this ranking obtained, the least important variables between a pair of strongly correlated features can be discarded from the dataset.

In the FS domain, three different approaches exist: filters, wrappers and embedded methods. The first approach measures the feature relevance independently from any learning algorithm (Hall and Smith, 1997) and it removes non-important attributes from the original dataset before the learning process. Conversely, wrappers use a target learning algorithm to estimate the quality of the attribute subset selection, assessing the process with a statistical resampling procedure (i.e. cross-validation). Filters are much faster than wrappers because they do not need to call the learning algorithm repeatedly (Saeys et al., 2007). Embedded methods perform the measure of a feature subset usefulness as part of the learning process that is internally done by a machine learning algorithm. The search is guided by the learning process and the result is assessed by using cross-validation. Typical examples are Random forest variable importance (RFVI) or Support Vector Machine recursive feature elimination (SVM-RFE). In terms of computational complexity, these techniques are between filters and wrappers (Saeys et al., 2007).

In this study, the mountain permafrost attributes were analysed with five types of feature selection algorithms: Information Gain (IG), RELIEFF, Correlation-based feature selection (CFS) and RFVI and SVM-RFE. Attributes were ranked based on their relevance to the permafrost distribution class (labelled as absence/presence) and the final variable importance results from a majority vote of all employed techniques.

#### Information Gain

The Information Gain (IG) algorithm is a filter technique that measures the importance of an attribute by computing the amount of bits of information by taking into account the absence or the presence of a feature (predictor) for the class prediction (Mitchell, 1997; Yang and Pedersen, 1997; Roobaert et al., 2006). In other terms, it evaluates the expected reduction in entropy (uncertainty) caused by partitioning the dataset observations according to this feature. IG is computed by subtracting

$H(C|X_i)$ , the expected value of the entropy of original dataset  $C$  partitioned by using the investigated feature  $X$ , from entropy of  $C$ :

$$I(C, X_i) = H(C) - H(C|X_i) \quad (4.12)$$

where  $H(C)$  and  $H(C|X_i)$  are calculated as follows:

$$H(C) = - \sum_C p(c) \log_2 p(c) \quad (4.13)$$

$$H(C|X_i) = - \sum_C \sum_X p(x, c) \log_2 p(c|x) \quad (4.14)$$

An IG greater than zero indicates that the features contained in the subset are relevant. The same technique is adopted in decision trees, which are presented in section 4.7.4.

## RELIEFF

Proven to be accurate with noisy observations and dependent predictors, RELIEF is an algorithm which was developed by Kira and Rendell (1992) and which, in its original form, is limited to two-class problems. Kononenko (1994) modified the algorithm for multi-class problems and renamed it RELIEFF.

The algorithm aims at estimating the importance of a predictor according to how the latter differs from neighbouring observations. If the predictor values differentiate from neighbouring observations (which have a different class), the predictor has a greater importance. Conversely, if the values are similar, then the importance of the predictor decreases. Different values are called *nearest misses*, while similar values are known as *nearest hits*. Kononenko (1994) defines the relevance of a predictor as a difference of probabilities:

$$W[F] = P(\text{diff. value of } F \mid \text{nearest inst. from diff. class}) - P(\text{diff. value of } F \mid \text{nearest inst. from same class}) \quad (4.15)$$

In a first step, the algorithm selects a random observation  $R_n$  of the feature  $F$  and looks for the nearest hits ( $NH_i$ ) or nearest misses ( $NM_i$ ).  $NH_i$  correspond to observations with the same class, while  $NM_i$  distinguish from  $R_n$  as they have different classes. The difference between the values of the predictor  $F$  are computed by the function  $\text{diff}(F, I_1, I_2)$ , which calculates the distance between the two observation  $I_1$  and  $I_2$ :

$$\text{diff}(f, I_1, I_2) = \frac{|\text{value}(F, I_1) - \text{value}(F, I_2)|}{\max(F) - \min(F)} \quad (4.16)$$

The importance of the predictor  $W[F]$  is computed as the difference between  $\sum \text{diff}(F, R_n, NM_i)$  and  $\sum \text{diff}(F, R_n, NH_i)$ , corresponding to the average contribution of the misses and the average contribution of the hits. RELIEFF considers a

predictor relevant, namely that it well distinguishes between two classes, when the average difference for the misses is large. Conversely, if the value of the average difference for the hits is large, the relevance value is low.

### Correlation-based feature selection

The *correlation-based feature selection* algorithm was developed by Hall and Smith (1999). It computes a ranking of the predictor importance firstly by discretizing numeric features, then by using symmetrical uncertainty to estimate the degree of association between them. Feature ranking follows the principle that "good feature subsets contain features highly correlated with the class, but uncorrelated with each other" (Hall and Smith, 1999).

$$M_S = \frac{k\overline{r_{cx}}}{\sqrt{k + k(k-1)\overline{r_{xx}}}} \quad (4.17)$$

$M_S$  is the heuristic merit of a subset  $S$  containing  $X_i$  features,  $\overline{r_{cx}}$  is the mean feature-class correlation and  $\overline{r_{xx}}$  is the average feature-feature correlation. As a consequence, subsets of predictors with higher importance will have a greater merit  $M_S$  than subsets of predictors with lower importance.

### Support vector machines recursive feature elimination

The SVM-RFE is a feature selection algorithm that uses SVM as a wrapper to compute a ranking or relevant predictors. It was first proposed by Guyon et al. (2002) and it is based on both the *recursive feature elimination* (RFE) technique and the ability of the linear SVM to weight the features (see Boser et al., 1992; Vapnik, 1998). For more details on the mechanisms of the SVM algorithm, see section 4.7.3 further in this manuscript.

Initially employed by Kohavi and John (1997) as a backward feature elimination procedure to be applied to wrappers, RFE was used by Guyon et al. (2002) for iteratively training the SVM with the objective of removing the predictors providing the smallest resulting weight. In synthesis, the main steps of the procedure are the following:

1. Train the SVM classifier;
2. Compute the weight of the predictors;
3. Remove the predictor with the smallest weight;
4. Repeat steps 1-3 until all features are ranked.

SVM-RFE allows ranking the features of the dataset according to their weight that was computed during the classification process. However, this procedure does not attribute a value to the relevance of the predictor, requiring the elimination of one feature at a time.

### Random forest variable importance

Developed by Breiman, 2001, Random forest (RF) is an ensemble learning method for classifications, operating by constructing a large collection of decorrelated classification trees and by predicting a class through a majority vote. The RF algorithm has recently gained a great popularity because it also provides a weight of each variable contribution as part of its overall operation. Refer to section 4.7.4 for further theoretical details on this procedure which is strongly linked to the RF classifier mechanisms.

## 4.4 Statistical classification

In machine learning, a classification task is required to discover the class of unlabelled samples with the help of various input values. Among the multiple existing classification algorithms, supervised and unsupervised classifiers can be distinguished.

*Supervised classifiers* require a training dataset and are inferred by known output and input variables (that are permafrost occurrence and environmental variables in the present study). During the supervised learning process, the machine performs a training procedure aiming at learning the relationships that exist in the dataset between the inputs and the outputs. The examples that are provided to the machine are called training samples. Each training example corresponds to an input vector  $x \in R^N$  to which is associated a response variable  $y$  (the output). Each training input can be represented in a  $N$ -dimensional space where the number of dimensions corresponds to the number of selected predictors.

If the outputs are continuous real values ( $y \in R$ ), the learning task is considered a regression problem. Conversely, if the input is discrete, the learning task can be considered as a classification problem, which can be a binary classification task ( $y \in -1, 1$ ) or a multi-class classification task ( $y \in 1, 2, \dots, k$ ).

The learning process is performed by the machine by reviewing all the training data (inputs and corresponding outputs), resulting in the discovery of a decision function that maps inputs  $x$  to the outputs  $y$  domain. Supervised classifiers are susceptible to over-fit (Forsyth and Ponce, 2012). This is a phenomenon that occurs when the algorithm performs worse on test data than on training data, indicating a poor generalization of the classifier on unknown samples.

The goal of *unsupervised algorithms* is the extraction of information from the data just by employing their statistical properties without prior knowledge of the studied phenomenon distribution. The most known types of unsupervised learning algorithms are the clustering techniques, which consist of grouping the data into clusters



with similar properties and of density estimation. This is an approach that models the probability distribution of the studied phenomenon.

Supervised and unsupervised approaches can also be combined. This approach is referred as semi-supervised learning and belonging techniques that simultaneously deals with labelled and unlabelled training examples.

#### 4.4.1 Statistical Learning Theory

To assess the trade-off of the complexity of a model and its learning capability, Vapnik (1998) developed the *Statistical Learning Theory* (SLT). As mentioned in section 4.4, once the training process of a supervised classifier is complete, a decision function is retained. This function is selected among a set of several function  $F = \{f(x, \alpha), \alpha \in \Lambda\}$ , with  $\alpha$  corresponding to a vector of parameters chosen from the set  $\Lambda$ . The *expected risk* is the criterion used in Vapnik (1998) to define the quality of the final decision function  $f(x)$ , describing the dependencies between the input training samples and their corresponding outputs:

$$R(\alpha) = \int Q(y, F(x, \alpha)) dP(x, y) \quad (4.18)$$

In equation 4.18,  $P(x, y)$  is the unknown joint probability distribution of the training examples and  $Q(y, F(x, \alpha))$  is a loss function. The aim of the *expected risk* is to minimize the expected average loss by reducing the risk. In the classification, the most common used loss function generally takes this form:

$$Q(y, f) = \begin{cases} 0 & \text{if } f(x) = y \\ 1 & \text{otherwise.} \end{cases} \quad (4.19)$$

The expected risk of this loss function corresponds to the probability of a classification error.

#### 4.4.2 Empirical Risk Minimization

As already mentioned, in equation 4.18, the probability distribution  $P(x, y)$  is unknown and only the training samples are known. The theoretical risk can thus be approximated to an empirical risk calculated on the training inputs and outputs as:

$$R_{emp} = \frac{1}{N} \sum_{i=1}^N Q(y_i, f(x_i, \alpha)) \quad (4.20)$$

where  $N$  is the total number of training examples. The best set of model parameters  $\alpha$  is obtained by minimizing the empirical risk function, the so-called *Empirical Risk Minimization*. The selection is however extremely dependent on the training

examples that are taken into account. It is common practice to split the original dataset into independent set of data or to use a cross-validation approach (see section 4.5).

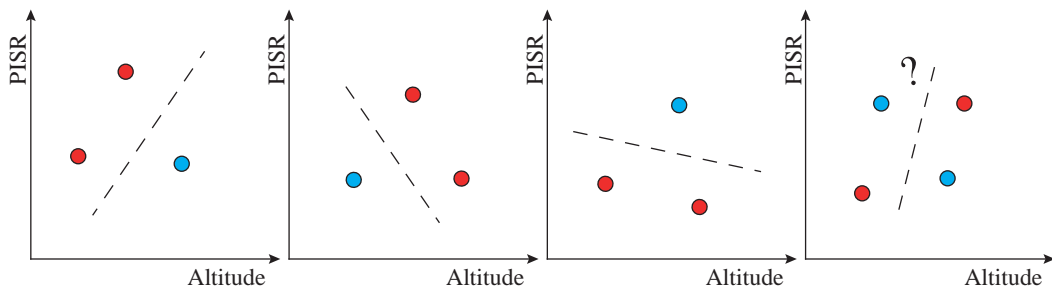
### 4.4.3 Structural Risk Minimization

The concept of the Structural Risk Minimization (SRM) was introduced in the SLT in order to assess the ability of a model to be applicable to new data. Vapnik (1998) suggests using an upper bound for the expected risk presented in equation 4.18. This bound has to vary depending on a confidence interval and the empirical risk. The Structural Risk Minimization is thus computed as:

$$R_\alpha \leq R_{emp} + \sqrt{\frac{h(\log(\frac{2N}{h}) + 1) - \log(\frac{\eta}{4})}{N}} \quad (4.21)$$

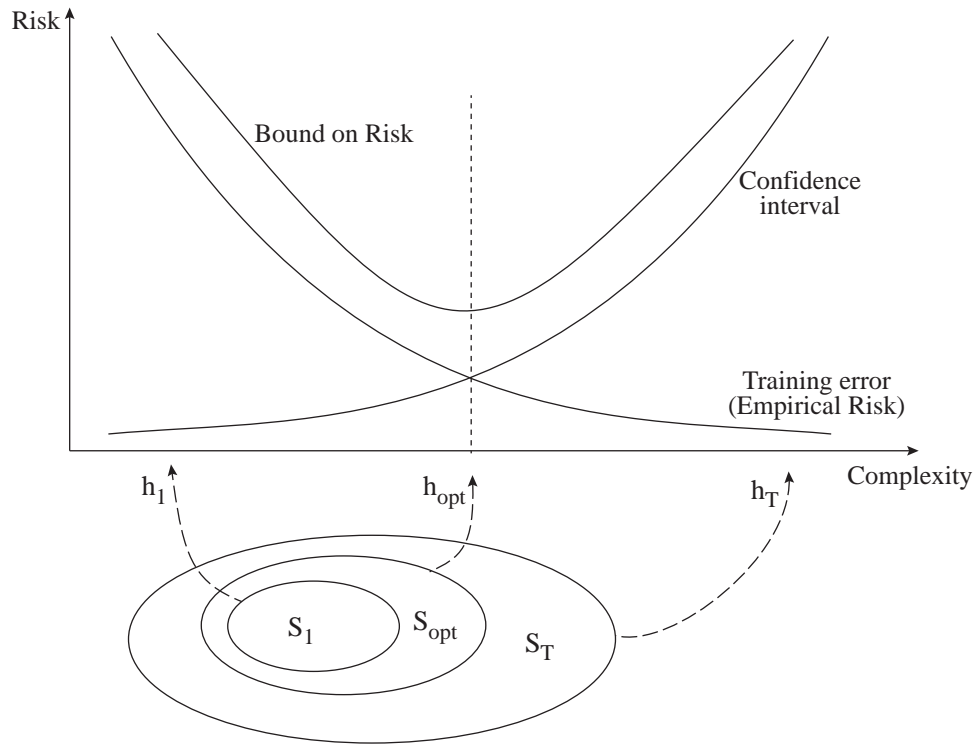
where  $h$  is the so-called Vapnik-Chervonenkis dimension (VC-dimension) of the decision function and  $N$  the number of training examples.

In short, the VC-dimension is a concept defining the cardinality of the largest set of samples that the learning algorithm can shatter. In other terms, the VC-dimension of a model is the maximum number of inputs that can be classified into the given outputs. Figure 4.3 illustrates a simplified example of training data that correspond to permafrost presence and absence inferred by two environmental predictors (altitude and PISR). In this two dimensional data sets consisting of 3 vectors, the two classes can always be separated with a linear function, independently of the label of the training samples. However, if the number of examples is increased to 4, a linear separation is not warranted. As a consequence, SRM suggests that the linear decision function  $f(x) = wx + b$  ( $f(x) \in R^N$ ) always has a VC-dimension of  $N + 1$  (Vapnik, 1998).



**Fig. 4.3.:** Three permafrost observations in 2D can be shattered with a linear decision function. The presented combination of 4 observations cannot be shattered and it requires a classification in a different dimension (modified after Kanevski et al., 2009).

Equation 4.21 indicates that as the confidence interval is small, the complexity of the model is kept small, as well as the risk (figure 4.4). Indeed, a decision function with a VC-dimension  $h$  that perfectly fits the entire training dataset would result in a large expected risk.



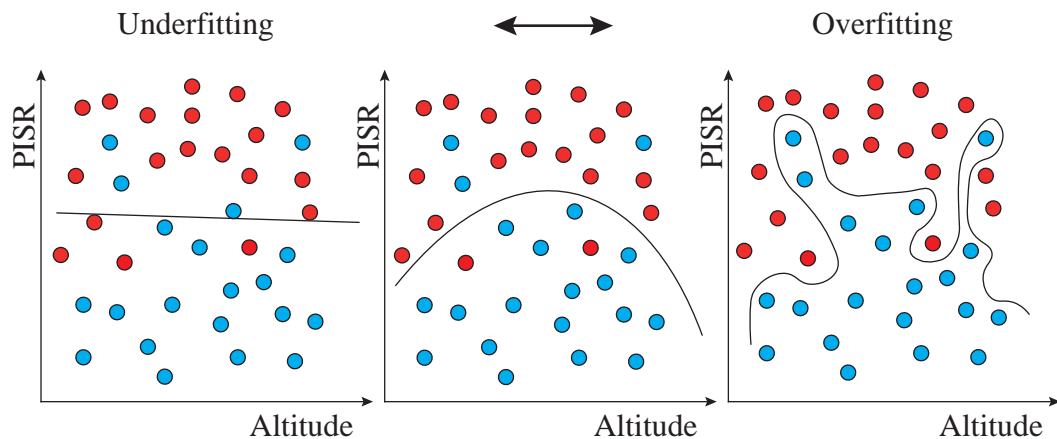
**Fig. 4.4.:** Variation of the bound on risk according to the empirical risk and the confidence interval for an increasing complexity of models (modified after Vapnik, 1998).

## 4.5 Model selection and performance assessment

The SLT provides an evaluation of how good a model is according to its complexity (see 4.4.1). To apply supervised classifiers, two steps must also be followed: the model selection and the model assessment.

When dealing with the selection of the best model, the training error (or empirical error) allows evaluating the quality of the training process performed on the training examples by using the best set of parameters. The term *underfitting* is used when a model poorly fits the training set. Conversely, a model *overfits* when it perfectly or closely fits the totality of the training data, including the noise or the non-representative training samples (see figure 4.5). This results in a model that has a poor generalization ability when it is applied to new data. The selection of the optimal model parameters needs to be performed on an independent set of data, the

so-called validation set. It is generally obtained by splitting the training set allowing computing a validation errors through an optimization process.



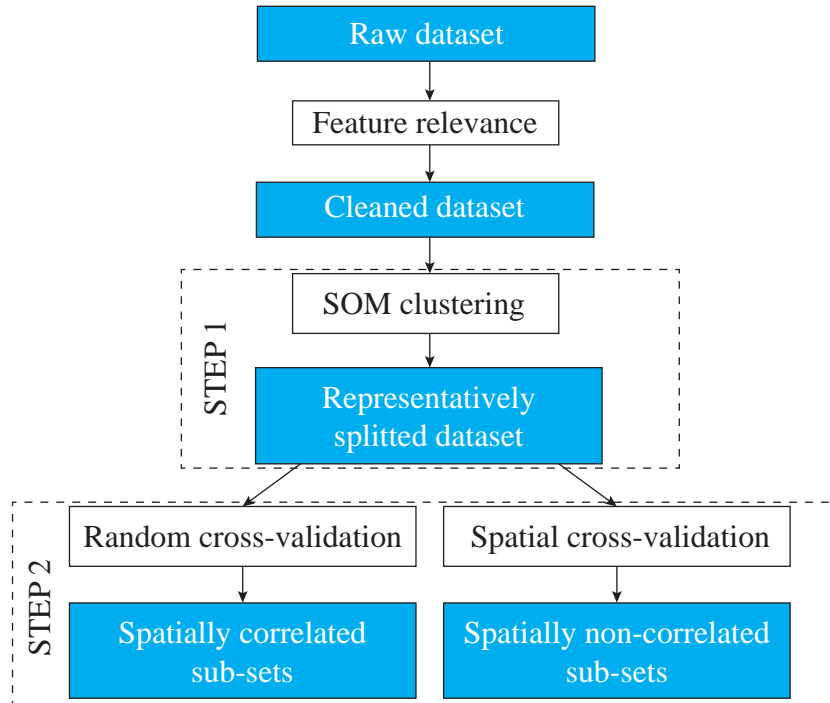
**Fig. 4.5.:** Underfit or overfit of a training dataset. If the model poorly fits (left image) or perfectly fits (right image) the training data, it will have poor generalization ability when applying to new data (modified after Kanevski et al., 2009).

When assessing the quality of the model, the performance of the classification is associated with the generalization error that quantifies the ability of the model to be applied to new data. This is calculated on an independent test set of data that differs from the training and the validation sets. The test error provides a measure of the model performance without using biased data that were already employed during the learning process in order to learn the input-output relationship or to tune the optimal parameters. Several measures of the classification performance exist and are presented in section 4.8 of this manuscript.

## 4.6 Subset sampling strategies

With non-linear data, recent advances in machine learning led to more flexible non-linear techniques, providing excellent performances in comparison to classical linear models. However, when dealing with spatial phenomena, flexibility may translate in a poor generalization and in a high tendency to over-fit the training set of data (e.g. Brenning, 2005). Indeed, it is not always easy to recognize over-fitting when the model performance is computed by using classical random cross-validation on spatial observations located close to each other and thus, share the same spatial characteristics (Koenig, 2002; Hampe, 2004; Hijmans, 2012). In modelling research fields that investigate, for example, species spatial distribution (i.e. spatial ecology or biology), it is common practice to discard observations that are located less than 200 metres one next to each other (see Guisan and Zimmermann, 2000; Schmid et al., 2014). This is generally possible when dealing with punctual training observations. When using polygons or training examples located within the same

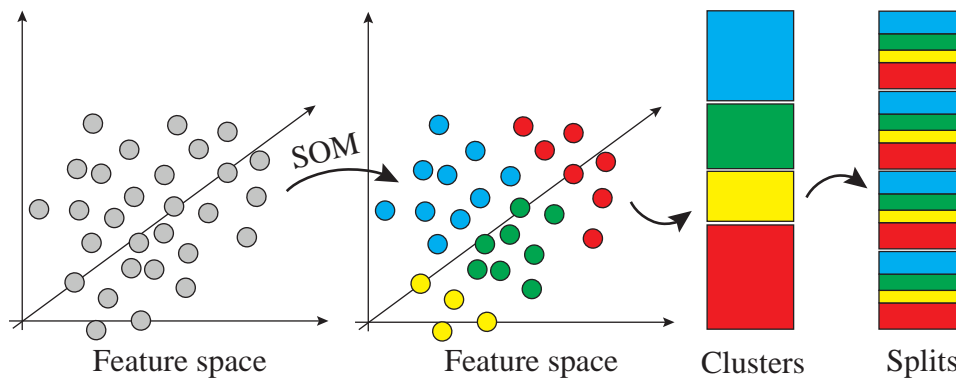
spatial region (i.e cultivated fields, rock glaciers, etc.), the risk of having very similar observations increases. Consequently, Brenning (2012) suggests employing spatial cross-validation to resample the data in order to overcome this problem and to take advantage of their localization within the spatial regions instead of selecting them completely at random.



**Fig. 4.6.:** Overview of the sampling strategy, coupling clustering and spatial/random cross-validation.

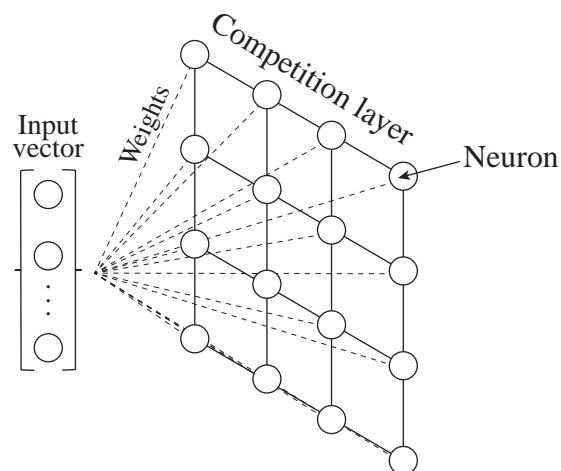
In this research, a hybrid sampling strategy was adopted and investigated by coupling spatial cross-validation and a selection of data according to their characteristics within the feature space (the high-dimensional variable space). As figure 4.6 indicates, after applying feature relevance techniques (mentioned in section 4.3.1), a cleaned dataset was obtained. Due to a consistent number of permafrost observations (more than 90'000), a partition of that cleaned dataset into subsets was necessary to reduce training times and the computational requirements. Data were thus divided into splits (figure 4.6, STEP 1). A purely random splitting may lead to sub-groups of very similar observations, namely training examples with similar coordinates within the high-dimensional feature space. To overcome this issue, a self-organising map (SOM) was computed in order to find similarities in data and to cluster them before the partitioning (see figure 4.7).

SOM is an unsupervised learning procedure that produces a low dimensional representation of a dataset (called map) reproducing the topological structure of the data (Kohonen, 1982; Kohonen, 2000). A map can have any number of dimension, but it



**Fig. 4.7.:** Splitting of the cleaned dataset: SOM mapping allows detecting clusters data that are then split to obtain representative sub-datasets.

is generally represented in 2D. SOM mimics the process of task clustering that takes place in the human brain. It is considered a neural network aiming at visualizing high-dimensional data. It indeed compresses the information contained in a dataset into a low-dimensional map. In SOM, the neurons distinguish in two layers: the input layer and the competition layer (figure 4.8). The number of inputs corresponds to the number of dataset variables, while the competition layer is composed of a grid of neurons (usually 2D) geometrically ordered. These neurons correspond to the outputs and their number depends on the complexity of the problem.



**Fig. 4.8.:** Example of the architecture of a Self-organizing map, with the representation of the input layer connected to the competition layer by weighted vectors (modified after Carrasco Kind and Brunner, 2014).

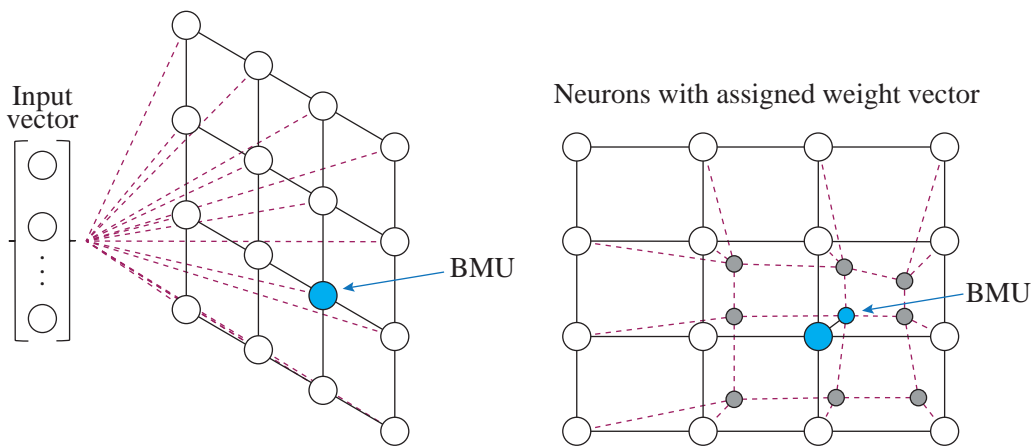
All inputs are connected to each neuron of the competition layer by a weighted vector (figure 4.8). The SOM algorithms compute different models that best describe the input layer on the two-dimensional grids (Kohonen, 2000). The process aims at respecting a *neighborhood relation* between the distribution of the original high-dimensional dataset and its topological low-dimensional representation. The number

of neurons, namely the size of the grid and the type of the map (rectangular or hexagonal grid), has to be optimized depending on the size of the data. Each neuron weight is then calculated through a learning process that starts with random weights and it is optimized iteratively. The distance between randomly sampled inputs and their weights are computed at each step. The neuron of the competition layer with the nearest weight vector to the input is called *Best-Matching Unit* (BMU) and is denoted by  $c$ :

$$\|x - m_c\| = \min_i d(x - m_i) \quad (4.22)$$

where  $d()$  is the distance measure function (an Euclidian distance in this study) and  $x$  is the input observation.

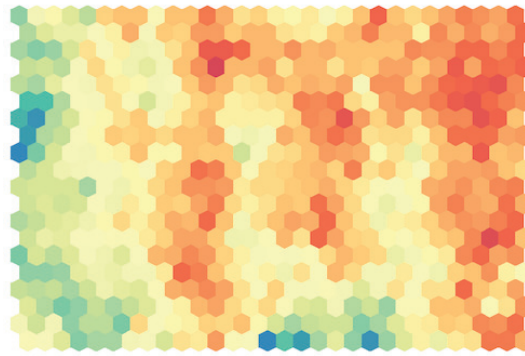
The weight vectors are then updated in order to pull the BMUs and their topological neighbours towards the sampled inputs (figure 4.9). The training process is generally performed in two learning phases that are characterized by different learning rates. A first rough training serves to arrange the neurons on the competition layer and a fine-tuning stage to complete the arrangement of the map.



**Fig. 4.9.:** SOM structure (left) and update of Best Matching Unit (BMU, right; modified after Carrasco Kind and Brunner, 2014).

The visualization of the 2D SOM is performed after the training completion and a so-called *heat map* that indicates the topological structures is obtained (figure 4.10). In other words, it represents the number of times each neuron of the competition layer was the BMU for each input observation, informing on the number of input observations gathered in each neuron. This translates in a grouping of the inputs in clusters.

In this research, the heat map computed on the cleaned dataset was analysed by using a *k-mean* algorithm. The expected number of clusters is generally approximated or extracted by using several indices (e.g. Caliński and Harabasz, 1974; Davies



**Fig. 4.10.:** Example of a heat map illustrating the structures in the topological 2-D representation of the data (generated with the R *kohonen* package; Wehrens, Buydens, et al., 2007).

and Bouldin, 1979; Krzanowski and Lai, 1988). In this study, the optimal number of clusters was computed with the implementation of a wide library of indices contained in the R package *Nbclust*. Once the clusters identified, the cleaned dataset was split into 10 representatively splitted sub-datasets containing observations that were randomly selected from each cluster. The proportion of observations gathered from the same cluster respected the size of the cluster.

This first step allowed reducing the size of the original cleaned dataset into sub-datasets. The adopted approach could be considered very similar to a stratified cross-validation, although the selection of the training observations was performed directly in the high-dimensional feature space based on detected clusters.

The sub-datasets required to be partitioned into training, validation and test sets to be employed by machine learning classifiers and the model quality and their generalization ability to be assessed. Two sampling techniques were thus selected in this research: random cross-validation and spatial cross-validation (see figure (figure 4.6, STEP 2). In its basic approach, random cross-validation has to split the training set into  $k$  smaller sets (folds). The model is then trained by using  $k - 1$  of the folds as training data and the resulting model is validated on the remaining part of the data. The unused fold for the training process serves thus as a test set and it allows assessing the performance of the model. The same process is repeated in loop until all folds are used as test. The final performance of the model is the averages of the values computed during the entire process.

Spatial cross-validation is a very similar technique, but it also takes into account the spatial similarity of the observations. For the permafrost case, two training observations may be located on the same rock glacier being thus spatially correlated. To avoid using potentially similar samples both in training and test, spatial cross-validation randomly selects the training examples by considering their location.



Concretely, it avoids the split of permafrost evidences that belong to the same rock glacier (for example) in both training and test sets.

Overall, from the 10 representatively splitted sub-datasets, training, validation and test sub-sets were obtained by using random and spatial resampling. A 10-fold cross-validation was selected in order to obtain 10 repetitions for each sub-dataset, resulting in a total of 100 subsets (or 100 repetitions) that were employed by machine learning algorithms. Around 30% of a sub-dataset was used as a test set, 15% of the remaining data was employed in validation and the lasting 55% in training.

Both sampling strategies were tested to unravel the main differences in model performances when partitioning a consistent number of observations with respect to their spatial auto-correlation. The main outcomes of this analysis will be presented in section 5.4.2.

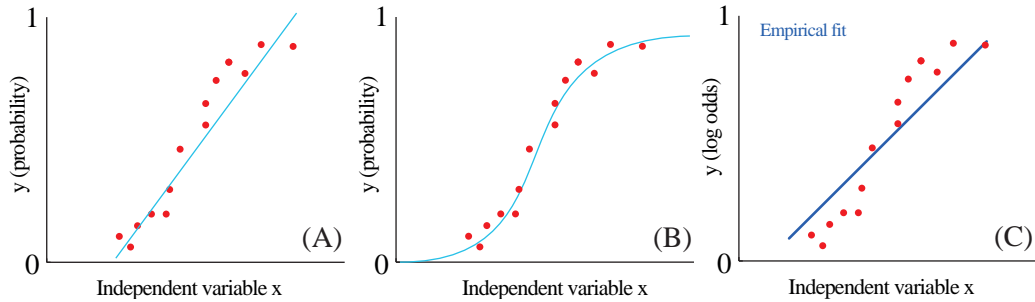
## 4.7 Selected machine learning classifiers

In the following sections, a theoretical overview of Logistic regression (LR), Artificial neural networks (ANN), linear and non-linear Support Vector Machines (SVM) and Random forests (RF) is given. These algorithms were selected because they belong to specific sub-domains of machine learning: LR is a linear parametric classifier and it is commonly used as a benchmark classifier to be employed before using more complex classification algorithms. ANN are supervised learning algorithms that can learn a non-linear functions. They differ from LR because they have one or more non-linear layers between the input and the output layers (called hidden layers). SVM were selected because they represent a non-parametric learning algorithm and they are a member of the so-called kernel methods. Lastly, RF are an ensemble learning method based on bootstrap aggregating.

### 4.7.1 Logistic regression

Logistic regression is one of the most used classification methods in environmental sciences. It is widely used for landslide susceptibility mapping (e.g. Nandi and Shakoor, 2010; Lepore et al., 2012; Trigila et al., 2015), but it was once employed for permafrost distribution modelling by Lewkowicz and Ednie (2004), Brenning and Trombotto (2006), Sattler et al. (2016) and Alm (2017). Testing how it performs compared to the other selected machine learning algorithms is extremely important in the context of this research.

Logistic regression falls into the family of *generalised linear models* and it aims at fitting the best model between independent indicators and dependent variables (Kleinbaum and Klein, 1994). In the case of permafrost, it tries to estimate the best mathematical relationship between the absence and the presence of permafrost (a dependent variable  $y$ ) and a set of explanatory independent environmental variables  $x_1, \dots, x_n$ , which can be both continuous or categorical (Hosmer and Lemeshow, 2000).



**Fig. 4.11.:** Comparison between linear regression (A) and logistic regression (C) of observed probabilities  $y$  on the independent variable  $x$ . (D) Linear regression of the observed log odds  $y$  on the independent variable  $y$ .

Figure 4.11.A illustrates that the linear regression of the observed probabilities  $y$  on the independent variable  $x$  would not provide probability values in the meaningful range of 0 to 1 of the y-axis (by extending the regression line downward and upward along the x-axis). Since the prediction output can only take the form of two values (0 and 1, or *absence* and *presence*), the relationship between the permafrost occurrence and the independent environmental predictor  $x_n$  takes the form of a S-shaped curve (figure 4.11.B). This result is in opposition with linear regression models that assume a linear relationship between the inputs and the output classes. Logistic regression overcomes this problem by calculating conditional probabilities as the natural logarithm of the odds of  $y$  as follows:

$$\ln \left( \frac{P(y = 1)}{1 - P(y = 1)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (4.23)$$

where  $P(y = 1)$  is the probability that the predicted class is 1 (presence of permafrost),  $\beta_0$  is the intercept (the point on the y-axis (log odds) crossed by the regression line when  $x=0$ ), and  $\beta_n$  is the regression coefficient of the environmental variable  $x_n$  (Field, 2009). These coefficients can be exploited as a measure of how each independent environmental variable contributes to the permafrost occurrence.

It results that the s-shaped curve of the relationship between output classes and the input features becomes linear while maintaining the non-linearity of the data (figure 4.11.D). The logistic regression solving for the probability  $P(y = 1)$  is defined as:

$$P(y = 1) = \frac{1}{1 + e^{-A}} \quad (4.24)$$

where  $e$  is the base of the natural logarithm and  $A$  corresponds to

$$A = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (4.25)$$

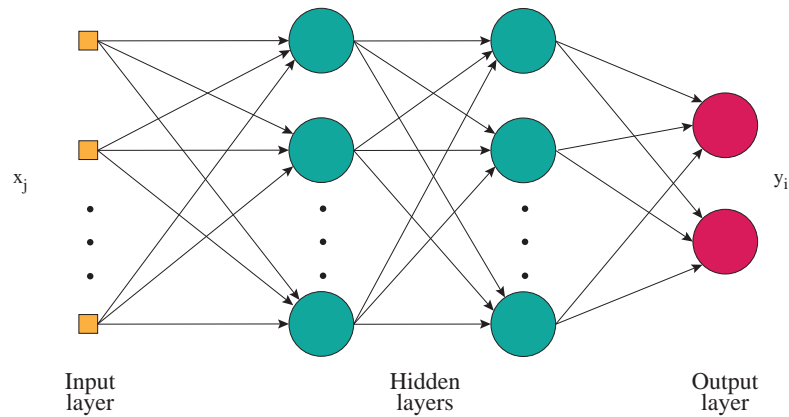
The mathematical background of logistic regression, as well as additional information on the odds interpretation are presented in Peng et al. (2002), Peng and So (2002) and Field (2009). In the present work, the logistic regression is performed by using an improved version of the original algorithm presented in Le Cessie and Van Houwelingen (1994) (implemented in WEKA).

### 4.7.2 Artificial neural networks

Artificial neural networks (ANN) are non-parametric data processing tools that are well suited for modelling problems whose solutions are not explicitly formulated (Kanevski et al., 2009). Their name was inspired by the biological networking of the brain and its ability to treat the information it receives. Indeed, when applied to data analysis, ANN are able to estimate a dependence function without any a priori explicitly given mathematical model that describe the relationship between the input variables and the output classes (Bishop, 2006). ANN were applied for a great variety of environmental data analysis, such as landslide susceptibility mapping (e.g. Neaupane and Achet, 2004; Melchiorre et al., 2008; Kawabata and Bandibas, 2009; Pradhan and Lee, 2010), wind fields (e.g. Fan et al., 2008; Li and Shi, 2010), weather forecasting (e.g. Kuligowski and Barros, 1998; Maqsood et al., 2004), etc.

Compared to classical models, ANN are superior for noisy data, outliers or large errors. In addition, they are particularly robust for data with underlying and complex structures of patterns that are hardly identifiable with standard statistical methods (Masters, 1993). Moreover, their data-driven nature makes them reliable with non-linear and/or multivariate data (thus data described by a high-dimensional input space), which are very common conditions of geo- and environmental processes. These properties make ANN suitable to overcome problems related to the curse of dimensionality (Hastie et al., 2009).

Information processing is performed by artificial neurons, which are mathematical models that simulate the biological neurons. They are the nodes that receive the inputs (the data) and, generally by weighted summation, they combine them to generate an internal activation level. The higher the signal, the stronger the same



**Fig. 4.12.:** An example of a feedforward neural network: multilayer perceptron with input neurons (in yellow), two hidden layers made by hidden neurons (in green) and two output neurons (in red). Connections between neurons represent the  $w_{ij}$  weights (modified after Kanevski et al., 2009).

signal will be transferred to other neurons of the network (see figure 4.12). The mathematical formulation of a neuron takes the form:

$$Z = f\left(\sum_{i=1}^K w_i x^i + b\right) \quad (4.26)$$

where  $x^i$  are the input features (variables),  $w_i$  the summation with weights,  $b$  a bias value and  $f(\cdot)$  the transfer function. In this research, the latter is a S-shaped sigmoid (logistic) function

$$f(x) = \frac{1}{1 + e^x} \quad (4.27)$$

Although they are considered non-parametric models, neural networks require some assumptions on the structure of the model. Two main characteristics of the network have to be defined: the number of neurons and the network topology (structure of connections between neurons). In this study, a *feedforward* architecture was used to build the ANN. This architecture lets the information flow in one direction only to reach the output, resulting in a network that will always provide the same result for the same input.

Multilayer perceptron (MLP) is the most frequently used feedforward neural network that is built with a connected network of neurons organised in layers (see figure 4.12). The network presents hidden layers between the inputs and the output, containing neurons that process the information. The learning process accomplished by MLP is supervised and it uses a *backpropagation* algorithm for training, aiming at iteratively minimizing the error between the output and a precomputed error that was obtained before the re-initialization of the network.

The equation representing the training model of a two-layer perceptron (as it is shown in figure 4.12) is

$$F^m(x^1, x^2, \dots, x^K) = f^{out} \left\{ \sum_{h_2=1}^{H_2} w_{h_2,m}^{(out)} f^{h_2} \left[ \sum_{h_1=1}^{H_1} w_{h_2,h_1}^{(2)} f^{h_1} \left( \sum_{k=1}^K w_{k,h_1}^{(1)} x_k + b_{h_1} \right) + b_{h_2} \right] + b_{h_2} \right\} \quad (4.28)$$

where  $f^{h_s}(\cdot)$  is the transfer function for the hidden layers and  $f^{out}(\cdot)$  the output layer,  $m$  the output index,  $H_1, H_2$  the number of neurons in the two layers,  $K$ , the number of input variables and  $b_k, b_{h_1}, b_{h_2}$  the biases of the layers. Equation 4.28 allows computing new predictors according to the learned weights. The weight  $w_{h_q, h_p}^{(H)}$  is the weight attributed to the link from the neuron  $h_p$  of the first layer to the neuron  $h_q$  of the second layer  $H$ . If this layer is the output layer, then the weights are defined as  $w_{h_q, m}^{(out)}$ , which would be the link between  $h_q$  in the last hidden layer and the output  $m$ .

The weights are optimized with a *backpropagation* training algorithm (short for backward propagation of errors) by using a gradient descent. The algorithm computes the gradient of the error function according to the weights of the MLP. The *backward* procedure indicates that the calculation of the gradient is made backwards through the network. The mathematical formulation of the backpropagation algorithm as well as the gradient-based MSE minimization can be found in Bishop (2006) and Kanevski et al. (2009).

In this study, a MLP was applied to permafrost data. Its implementation in WEKA employs a validation set to automatically build the neural network and optimize the number of layers.

### 4.7.3 Support Vector Machines

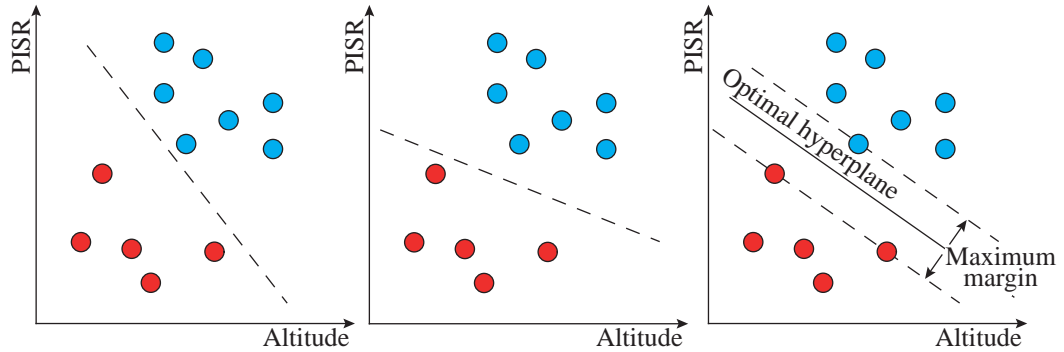
The Support Vector Machines (SVM) is a machine learning algorithm firstly introduced by Boser et al. (1992) and subsequently mathematically developed by Vapnik (1998) on the basis of the statistical learning theory. It is based on the Structural Risk Minimization (section 4.4.3) that minimizes the training error and it controls the complexity of the model in order to improve the generalization ability of a model (Cherkassky and Mulier, 2007). They have been adopted successfully in several fields, such as remote sensing (multispectral images; i.e. Mitra et al., 2004), cancer cell recognition (i.e. Guyon et al., 2002), text mining (i.e. Cortes and Vapnik, 1995) or landslide mapping (i.e. Micheletti et al., 2014).

SVM can be applied to classification tasks and non-linear regression problems. The main principle of this technique presupposes that the set of training vectors

$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , where  $x_i \in R^m, i = 1, \dots, n$  with two classes  $y_i = \{-1, 1\}$  is linearly separable by a hyper-plane:

$$(w \cdot x) + b = 0, w \in R^N, b \in R \quad (4.29)$$

where  $w$  corresponds to the hyper-plane normal,  $(\cdot)$  is a scalar product and  $b$  is a scalar base. The hyper-plane corresponds to the decision boundary and any new sample passing through the algorithm is classified depending on the resulting side of the decision boundary.



**Fig. 4.13.:** The largest margin hyper plane is an intuitively simple linear model with good generalization. Closest training examples to the hyperplane are called *training vectors* (modified after Kanevski et al., 2009).

The SVM algorithm aims at maximizing the largest margin that separates the training vectors (the closest training examples to the hyperplane) and that is expressed as  $\frac{2}{\|w\|}$  after the normalization (figure 4.13). The maximum margin is computed as follows:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (4.30)$$

subjecting to the constrains  $y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, n$ .

The cost function is expressed as:

$$\varnothing(w, b; \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i [w \cdot x_i + b] - 1) \quad (4.31)$$

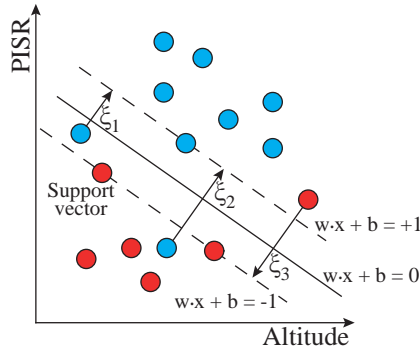
where  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T \in R_+^n$  is a Lagrangian multiplier.

As detailed in Vapnik (1998), the problem has to be solved by a dual minimization of the cost function with respect to  $w$  and  $b$ . Because in most cases the training vectors are rarely linearly separable, a slack variable  $\xi_i$  (figure 4.14) and a penalty term  $C$ , avoiding high values of the latter, are introduced:

$$y_i((w \cdot x_i) + b) \geq 1 - \xi_i, i = 1, 2, \dots, n, \xi_i \geq 0 \quad (4.32)$$

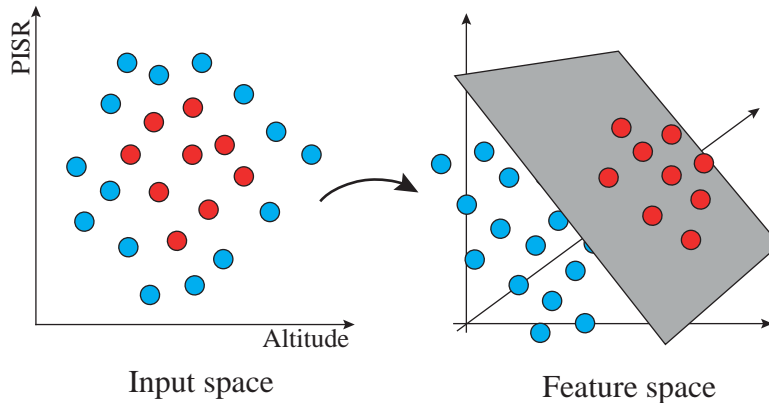
Therefore, the equation, maximizing the margin becomes:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (4.33)$$



**Fig. 4.14.:** Slack variables  $\xi_i$  assigned to noisy observation lying outside their class margin (modified after Kanevski et al., 2009).

Non-linear classification of real data, which is not often linearly separable, can be performed by using the same principles but by applying a kernel function  $K(x_i, x_j)$  (Vapnik, 1998). The kernel maps/transforms the original data into a high dimensional feature space where a linear solution can be obtained (figure 4.15). Depending on the kernel, the linear solution in the feature space corresponds to a non-linear solution in the original space and the kernel controls the complexity of this mapping.



**Fig. 4.15.:** The application of a kernel allows the mapping of the original data into a feature space where a non-linear problem becomes linear and can be easily solved (modified after Kanevski et al., 2009).

In this research, both linear and non-linear SVM are employed and their generalization abilities investigated. The latter employs a radial basis function (RBF) kernel  $K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}$ ,  $\gamma > 0$ , which is one of the most used kernels that provide a good generalization. The kernel parameter  $\gamma$ , as well as the cost  $C$  (the so-called

hyperparameters), had to be optimised. A grid search approach was adopted for this purpose because no analytic function allows a direct selection of the parameters with respect to the changes of the model performances. A quality measure was thus computed for a different sets of  $\gamma$  and  $C$  by using a range of pre-defined values. Lastly, the best  $\gamma$  and  $C$  are the values that produce the minimal error on a validation set.

When used for classification, the SVM decision function produces binary outputs that are difficult to employ for decision making. A probabilistic output is generally more interpretable. Platt (1999) presented an approach able to transform the classification result in probabilities. It is based on Bayes' rule that allows the posterior probability  $P(y = 1 | f(x))$  for an observation  $x$  to belong to the class +1 (presence) to take the following form:

$$P(y = 1 | f) = \frac{p(f | y = 1)P(y = 1)}{p(f)} \quad (4.34)$$

where  $f$  is the SVM decision function,  $p(f)$  its a priori probability,  $p(f | y = c)$  ( $c = -1, +1$ ) the class conditional probability of observing value  $f$  (where  $c = -1, +1$ ) and  $P(y = c)$  the prior class probability (for class  $c$ ).

Lastly, a smooth confident measure  $p : 0 < p(y = \{-1; +1 | x\}) < 1$  is computable by using the following sigmoid function:

$$p = \frac{1}{1 + \exp(A \cdot f(x) + B)} \quad (4.35)$$

where  $A$  and  $B$  are the parameters to tune on the validation set in order to avoid model over-fitting ( $A \in \mathbb{R}^-$  to ensure monotonicity and  $B \in \mathbb{R}$ ).

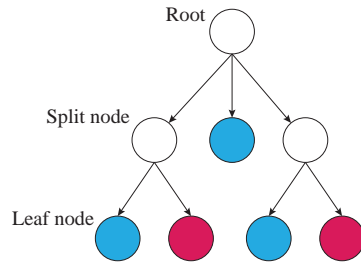
#### 4.7.4 Random forest

Random forest is probably the most well-known algorithm that falls into the category of the ensemble classifiers. Its development derived from decision trees and the ID3 algorithm.

Decision trees are supervised classifiers that were developed in the 1960s for artificial intelligence (see Hunt et al., 1966). They provide decisions at multiple levels giving a generally higher efficiency than single stage classifiers. Indeed, they are constructed as acyclic graphs made of a root node and child nodes. Decisions are performed at each node depending on training predictor values (figure 4.16).

If attributes are binary, the parent node is connected to two child nodes by two directional branches. If the predictors have non-binary values, then the parent node

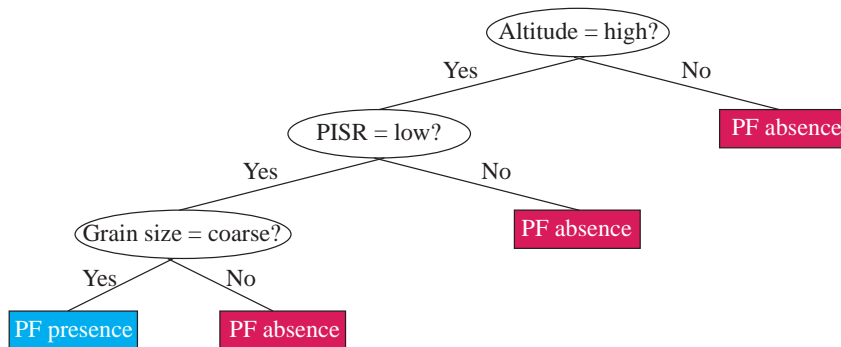




**Fig. 4.16.:** A schematic tree classifiers presenting a root, two split nodes and its terminal leaf nodes.

has one child for each predictor value. The development of the tree is performed until it reaches a terminal node (leaf node) that returns a classification.

A simple example of decision tree could be the classification of permafrost presence and absence of permafrost on the basis of three binary predictors: altitude (coded as *high* and *low*), PISR (coded as *high* and *low*) and grain size (coded as *coarse* and *fine*). Assuming for this example that permafrost is only present at high elevations, low PISR and coarse grain size, the decision tree classifying a simple permafrost dataset would take the form of the graph in figure 4.17.



**Fig. 4.17.:** Development of a decision tree providing the classification of a simple permafrost data example.

A complex problem is divided into smaller problems as the decision tree develops. However, the complexity increases when more predictors are introduced. Indeed, to construct a complex tree the main approach selects specific predictors to partition the data at each child node with the objective of obtaining the most information gain (Pal and Mather, 2001). As a consequence, it is necessary to decide how the dataset has to be split at each child node and when to stop splitting the data in order to obtain a final classification by computing the fewest calculations. The most used split methods are *gini* and *entropy*. The latter has already been defined in section 4.3.3 as

$$entropy(D) = - \sum_{i=1}^m p_i \log(p_i) \quad (4.36)$$

where  $D$  is the observation vector (i.e. a permafrost sample),  $m$  is the number of classes and  $p_i$  is the non-zero probability that  $D$  belongs to the class  $C_i$ . The objective of splitting the dataset by using entropy is to provide the most information gain from the parent node to the child nodes. The splitting condition corresponds to the reaching of  $entropy(D)$  equal to 0, namely when all cases belong to the same class and a classification can be provided (Han et al., 2011).

Entropy is a splitting condition that tries to divide the dataset as evenly as possible. However, the *gini information gain* is another split method that aims at selecting the largest possible homogeneous sub-set of data (Apté and Weiss, 1997):

$$gini(D) = 1 - \sum_{i=1}^m p_i^2 \quad (4.37)$$

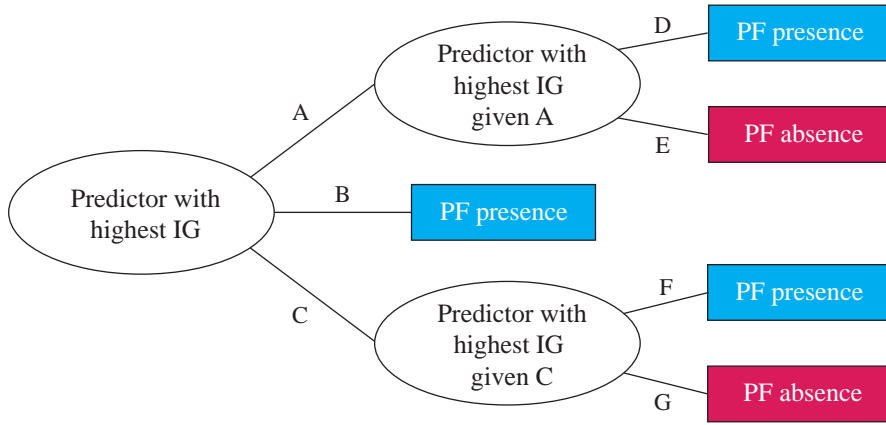
Duda et al. (2001) define the gini information gain (or gini impurity) as the expected error rate obtained when an observation is randomly chosen and randomly labelled according to the distribution of the classes in the subset.

A pure node is a child node containing observation belonging to the same class. A pure node can be defined with impurity indices presented in equations 4.36 and 4.37. Although this would correspond to an ideal node, splitting data excessively may cost too much with noisy data. Conversely, building a tree with a reduced number of splits lowers the number of leaf nodes, with consequently high error rates (Duda et al., 2001). Other splitting methods, such as *twoing*, *node depth* or thresholding the number of the observations at a specific node (see Duda et al., 2001; Venables and Ripley, 2002; Ghose et al., 2010), exist.

The Iterative Dichotomiser 3 algorithm, or ID3, is a variation of decision trees and is well suitable for large datasets. Developed by Quinlan (1986), ID3 employs a portion of the original dataset to grow a tree, an operation that is known as *bootstrapping*. The unused observations are then classified with the same tree. Resulting misclassifications are employed to further develop the tree. The process is then repeated until all training observations are correctly classified. The construction of the decision tree with bootstrapping is an approach that proves to be faster than employing the complete dataset. However, Quinlan (1986) indicates that the convergence to a final tree is not always guaranteed.

Entropy is used by ID3 as the splitting condition and the purity of the node as the stopping condition. The information gain, defined by subtraction of the information required by the tree after being split according to a predictors  $N$  to the information required by the tree itself, is computed as:

$$Gain(N) = Info(D) - Info_N(D) \quad (4.38)$$



**Fig. 4.18.:** Development of an ID3 tree based on the highest IG splitting (entropy or gini).

Depending on the selected split method (entropy or gini),  $Info(D)$  can be computed with 4.36 or 4.37.  $Info_N(D)$  is instead computed as follows:

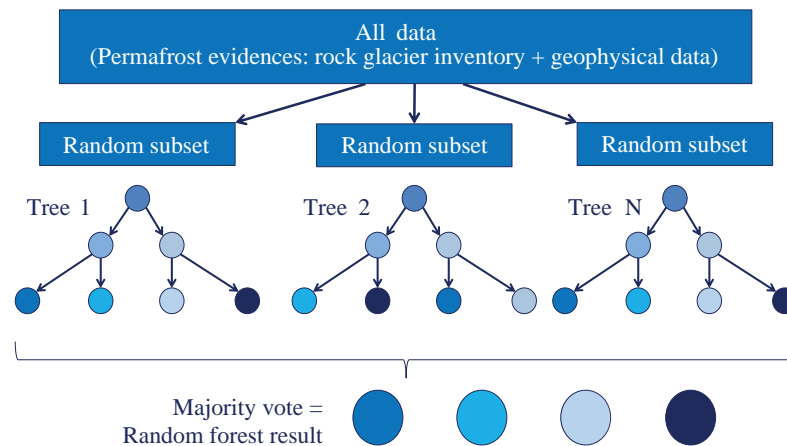
$$Info_N(D) = \sum_{i=1}^n \frac{|D_i|}{|D|} \times Info(D_i) \quad (4.39)$$

where  $\frac{|D_i|}{|D|}$  corresponds to the weight of the  $i^{th}$  split and  $n$  the number of values of the predictor  $N$ . The predictor with the highest information gain is used to split the data at the parent node and the same approach is then repeated recursively for child nodes (figure 4.18).

Based on the ID3 principles, Random forest (RF) was developed by Breiman (2001). This algorithm falls into the family of ensemble algorithms. It computes  $n$  binary classification trees (forest) in order to increase predictive capabilities compared to the classification with a single decision tree (Cutler et al., 2007). Its construction is based on three principles: the manipulation of training sets to gain *ensemble diversity*, the use of these training sets with the same *base classifier* (a random tree), and the *combination of results* by majority voting. Indeed, Breiman (2001) defines a RF as

*a classifier consisting of a collection of tree-structured classifiers  $\{(x, \Theta_k), k = 1, \dots\}$  where the  $\{\Theta_k\}$  are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input  $x$ .*

In other words, RF takes advantage of the employment of several versions of the base classifier to make a final prediction. The membership of a class is selected by a majority vote for the most popular class within the total number of trees (figure 4.19). With the combination of several trees constructed with a random selection of inputs, the classification accuracy usually improve.



**Fig. 4.19.:** Development of a forest of decision trees whose prediction is computed by majority vote.

Breiman (1996) introduced the concept of *bootstrap aggregating*, or *bagging*, based on the idea of *bootstrapping*. The RF algorithm adopts the principles of ID3 and bagging. The trees of the forest are grown by sampling with replacement a random subset, which is used to construct the tree in a similar way to ID3. However, unlike ID3, not all the observations of the bootstrap subset are adopted to grow the tree. Indeed, 2/3 of the bootstrap set is used for this purpose and the remaining 1/3 is used to assess the generalization capacity of the algorithm, the so called out-of-bag (OBB).

Similarly, only a subset of predictors is adopted to construct the tree, which is randomly selected and evaluated. This process avoids that a strong predictor is selected in multiple trees, causing strong correlation among trees. Instead of selecting all the predictors of the dataset to find the optimal split at each node, RF uses this different bagging approach to construct the trees. The number of selected predictors can be defined by the user. However, Breiman (2001) suggests using the greatest integer not greater than  $\log_2 M + 1$ , where  $M$  is the number of predictors.

The classification error is given by the *OOB error estimate*. With RF there is no need for a cross-validation or a separate validation set to obtain unbiased estimate of the test set error. Unselected observations serve indeed to calculate the error of the model (the "out-of-bag" error). Furthermore, because of the random selection of independent variables and observations at each node, it is suggested analysing the correct number of trees to be employed in Random forests in order to obtain a stable model (Catani et al., 2013). Therefore, before applying the classification model to a new prediction, investigating the changes in the OOB error curve is generally a good practice in order to select the correct number of trees.

RF also provides a ranking of the variable importance, as it was mentioned in section 4.3.3. To compute this measure, Breiman (2001) suggests constructing every tree of the forest, applying the tree to the OOB observations and then counting the number

of correct classes. Subsequently, to obtain the importance of the predictor  $m$ , its values must be randomly permuted and the trees applied to these OBB perturbed data. Finally, the variable importance is obtained by subtracting the number of perturbed correct classes from the number of correct classes obtained by applying the trees to non-perturbed observations. The importance of the predictor  $m$  is given by the average of the results of this subtraction over all the trees of the forest. Two evaluators exist: the Mean-Decrease-in-Accuracy, measuring the decrease of the accuracy expressed with the OOB-error when a variable is left out and the Mean-Decrease-in-Gini index, defining how the output at each node is impure (Breiman, 2001). The relevances of the predictors  $m_i$  are calculated as follows:

$$rel(m_i) = \frac{1}{k} \cdot \sum_{j=1}^k \frac{Err_{ij} - Err_{OBB_j}}{Err_{OBB}} \quad (4.40)$$

where  $k$  is the number bootstrap bag,  $Err_{OBB_j}$  the OBB of every decision tree and  $Err_{ij}$  the OBB of the permuted data.

## 4.8 Model performance assessment

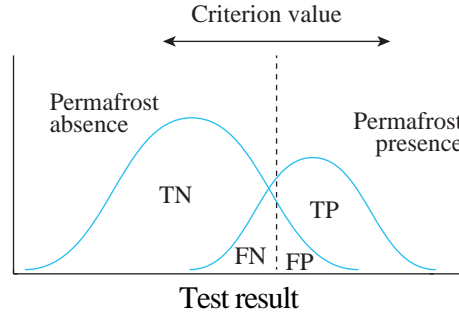
The quality of the prediction of the permafrost presence and absence modelled with supervised learning algorithms can be generally assessed with multiple statistics, such as the root mean square error, kappa statistics, the overall accuracy, etc. (Dowell III et al., 1990). The performances of the presented machine learning algorithms are computed on the test set and the resulting predictions are compared to the observations.

		Predicted		row totals
		Presence	Absence	
Observed	Presence	<i>True positive (hit)</i>	<i>False negative (miss)</i>	observed presence
	Absence	<i>False positive (false alarm)</i>	<i>True negative (correct negatives)</i>	observed absence
column totals		predicted presence	predicted absence	total

**Tab. 4.1.:** Classification confusion matrix for binary predictions of the permafrost occurrence.

As the permafrost problem is binary (classification of two classes), observed permafrost presence and absence are compared to the predicted permafrost presence and absence. For each model, a 2-by-2 contingency table displaying four possible cases was thus computed. It displays the number of true positives (TP), i.e. a pixel with known permafrost presence that was predicted correctly (presence-presence), the number of true negatives (TN), when absence is correctly not forecasted (absence-absence), the number of false negatives (FN), i.e. permafrost presence is not detected (presence-absence), and the number of false positives (FP), for permafrost absence incorrectly forecasted (absence-presence) (table 4.1). When considering the results

of the test for the two classes (permafrost presence and permafrost absence), the perfect separation is rarely observed. The distribution of the test results generally overlaps (figure 4.20). TP, TN, FP and FN fractions change when changing a criterion value that is selected to discriminate the two classes (0.5 for this study).



**Fig. 4.20.:** Schematic representation of the test and the possible prediction of a binary classification.

According to the schematic outcome of a test presented in table 4.1, the following statistic can be defined:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (4.41)$$

$$Sensitivity (True Positive Rate) = \frac{TP}{TP + FN} \quad (4.42)$$

$$Specificity (True Negative Rate) = \frac{TN}{TN + FP} \quad (4.43)$$

$$Positive Predictive value (Precision) = \frac{TP}{TP + FP} \quad (4.44)$$

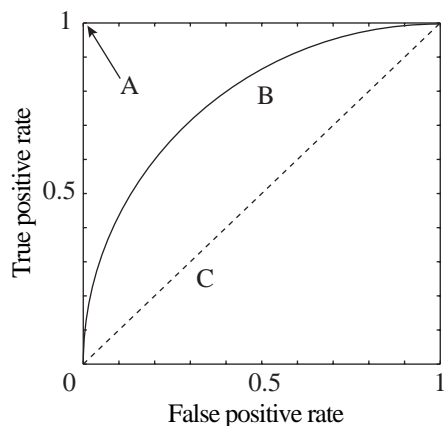
$$Negative Predictive value = \frac{TN}{FP + TN} \quad (4.45)$$

$$False positive rate = 1 - Specificity \quad (4.46)$$

$$False negative rate = 1 - Sensitivity \quad (4.47)$$

*Accuracy* is a measure of the overall quality of the model, but it may be hard to interpret because it takes *FP* and *FN* into account. High values of *accuracy* might be influenced by high number of *FP* and *FN*. *Sensitivity* (also known as *recall*) and *specificity* values are better quality indices when dealing with *TP* and *TN*. The first informs on the probability that the result of a test shows a prediction of permafrost *presence* when the observed permafrost evidence was effectively of *presence*. Similarly, the *specificity* is the probability that the *absence* of permafrost is predicted when the *absence* was also observed. These measures can be coupled with the *Positive Predictive Value* (*PPV* or *precision*), which is the probability that the

positive permafrost occurrence is observed when the test is forecasted as *presence* (TP and FP). A criterion value (threshold) can be set to give more or less importance to TP, TN, FP or FN. When dealing, for example, with permafrost distribution maps, it is preferable to obtain a conservative result that misclassifies permafrost where in reality it is not (FP), rather than the opposite.



**Fig. 4.21.:** The Receiver Operating Characteristics curve, as an indicator of the quality of the classification. (A) corresponds to the perfect result (no errors, AUROC = 1), (B) indicates a typical ROC curve (AUROC=0.85) and (C) refers to an inefficient classification (AUROC = 0.5).

In addition to these rates, a *Receiver Operating Characteristics* curve (ROC) can be built (Fawcett, 2006). This 2-dimensional graph measures the quality of the model prediction by plotting the FP rate (1-specificity) as the horizontal axis and the TP rate (sensitivity) as the vertical axis, according to different susceptibility thresholds (criterion values) (figure 4.21). A test with perfect discrimination (no overlap in the two-class distribution illustrated in figure 4.20) has a ROC curve that reaches the upper left corner (100% sensitivity, 100% specificity), whilst a model not performing well has a ROC curve close to the limit C as shown in figure 4.21 (Fawcett, 2006).

The quantification of the Area Under the ROC curve (or AUROC) provides a measure of the quality of the classification (Swets, 1988). AUROC ranges between 0.5 and 1. Values between 0.5 and 0.7 generally indicate that the learned model has poor predictive capabilities. When AUROC is between 0.7 and 0.8, it means that the classification has moderate predictive abilities. Between 0.8 and 0.9 they are considered good and above 0.9 they become excellent. Nevertheless, recent studies have observed that, for similar AUROCs, the predicted map of a given phenomenon can be particularly different (i.e. Micheletti et al., 2014).

## 4.9 Used software

The analysis modelling of this research was done within several GIS, statistical and mathematical software. Permafrost evidence identification, variable extraction and

dataset building were performed by using ArcMap (ESRI) and Matlab. The pre-processing of the data, as well as the exploratory data analysis and their sampling, were performed in Matlab and R. The feature relevance analysis and the classification of permafrost data were instead carried out in WEKA (Waikato University, Hall et al., 2009). Permafrost occurrence predictions were finally mapped by using ESRI ArcMap and post-produced in Adobe Illustrator.





## Results

### 5.1 Visual exploratory data analysis and structure detection

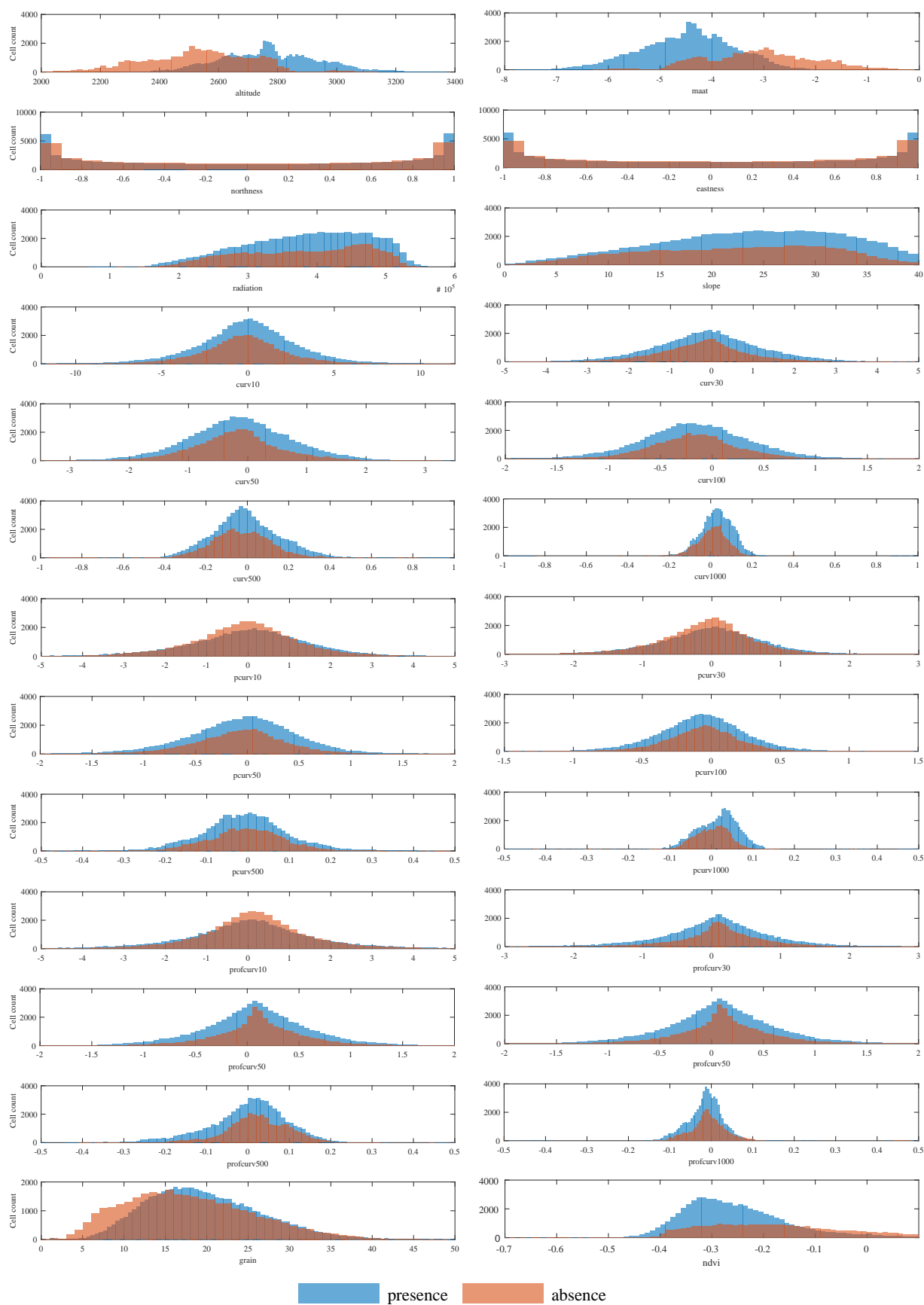
The raw permafrost dataset used in this research contains 523 polygons that were selected as training examples. In total 93'705 permafrost evidences were identified (pixels with 10x10m of resolution), 60'780 of which represent the permafrost presence (64%) and 32'925 the permafrost absence (36%). This corresponds to approximately 8% of the surface covered by the validity domain and 1.6% of the region of interest (see table 5.1).

	# pixels	km <sup>2</sup>
<i>Permafrost presence</i>	60'780	6.08
<i>Permafrost absence</i>	32'925	3.29
<i>(Permafrost evidences)</i>	(93705)	(9.37)
<i>Validity domain</i>	1'167'443	116.74
<i>Region of interest</i>	5'880'000	588

**Tab. 5.1.:** Portion of territory covered by training data, validity domain and region of interest in pixels and km<sup>2</sup>.

The exploratory data analysis was an extremely important first phase of this study. It gives an overview of the distribution of the permafrost evidences with regards to their assigned label (presence or absence) without any *a priori* assumption.

Figure 5.1 presents the environmental predictors considered for this study. The histogram plots illustrate the distribution of variables prior standardization. It can be observed that pixels associated with permafrost absence are generally located below 2850 m.a.s.l., while permafrost presence is more often mapped above 2350/2400 m.a.s.l.. Although the range 2350-2850 m.a.s.l. is represented by both labelled pixels, the altitude presents two distinct peaks. A similar distribution can be described for the temperature (MAAT) variable. The pixels of permafrost presence are more numerous for low values of the MAAT (ranging between -7 and -2°C). Conversely, evidences corresponding to permafrost absence are generally located where the calculated MAAT is above -5°C (up to 0°C).



**Fig. 5.1.:** Histograms of the distribution of the permafrost evidences for each of the retained environmental predictors and relative number of pixels (cell count). For a description of the variable acronyms, see table 3.1.

A less clear distinction can be identified when looking at the value distribution of other predictors. Radiation (PISR) presents a homogeneous repartition for pixels of permafrost presence, as well as for pixels corresponding to its absence. Mapped evidences are reached by a solar radiation greater than 150'000 watt hours per square metre (WH/m<sup>2</sup>). However, no apparent threshold discriminates the permafrost presence from the permafrost absence. The same consideration can be made for the slope variable, where no clear distinction in the label distribution is detected.

All histogram plots of the terrain curvatures indicate a shape corresponding to a normal distribution. Once again, there is not a clear discrimination between the pixels labelled as *presence* or *absence*. Nevertheless, some minor differences are detected when comparing the histograms of the curvatures at different window sizes. Indeed, in the majority of the terrain curvature predictors, the presence and the absence is equally dispersed at the centre of the distribution, with a cell count generally higher for presence-labelled pixels at curvature values of about 0 (flat). The planar curvatures at 10 and 30 metres (pcurv10 and pcurv30) and the profile curvature at 10 metres (profcurv10) indicate however a different dispersion of the values with more pixels of absence-labelled occurrence having such terrain curvature around 0. This indicates that permafrost absence is more likely observed in terrains presenting a flat micro-topography; conversely, the presence of permafrost is more likely to exist where the micro-topography is characterized by more convexities or concavities (at a 10-30 metre scale of resolution).

The distribution of the grain size predictor, which is derived from the standard deviation of grey-values of an orthophoto with 0.25 cm of resolution (see section 3.3.3), illustrates a greater number of observations related to the permafrost absence with low values of the standard deviation compared to evidences of the presence. It results that neighbouring pixels with similar grey-values (corresponding to fine materials of homogeneous surfaces) are more related to mapped permafrost absence, while surfaces with more heterogeneous grey-values (corresponding to coarse materials or surfaces with mixed textures) are indicative of permafrost presence.

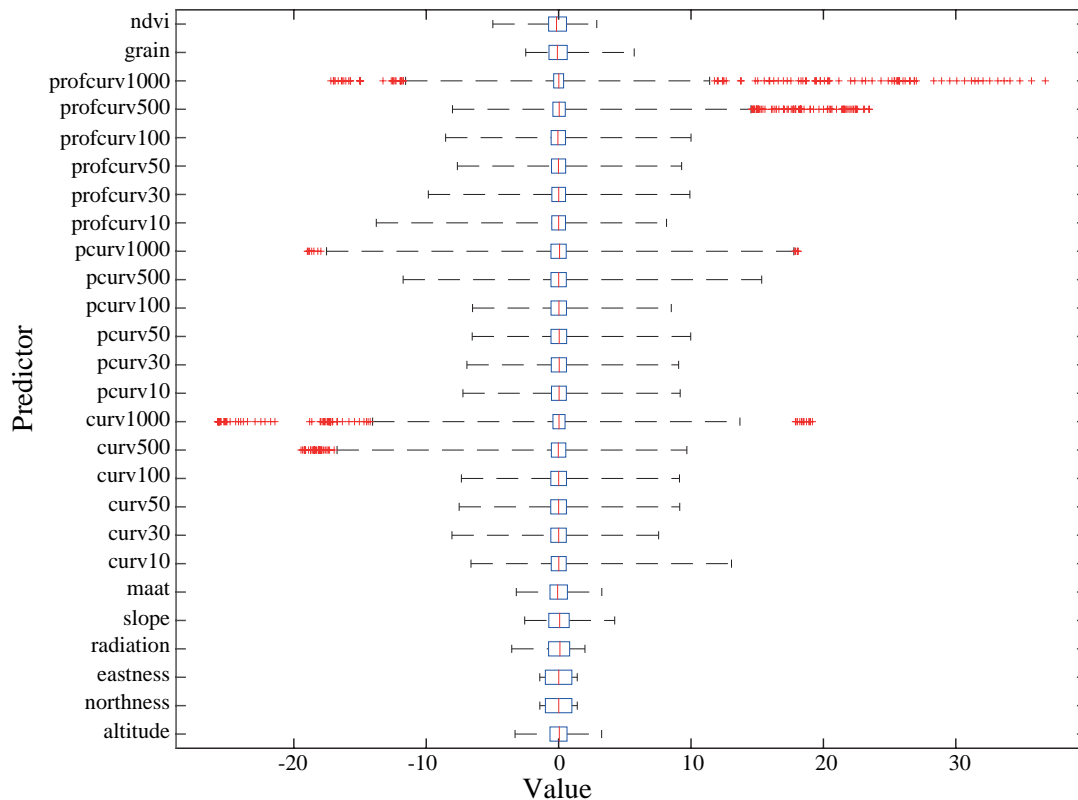
The last histogram plot indicates that pixels of permafrost presence generally have lower values of the NDVI in comparison to the absence-labelled evidences, which in contrast do not have distinctive values to clearly discriminate the permafrost occurrence.

Figure 5.2 illustrates the distribution of the values after standardisation. Values outside whiskers are considered as noisy data and were consequently excluded from the dataset (few hundreds of observations). Permafrost data were standardized (z-score normalized) so that all predictors have the properties of a standard normal distribution, namely  $\mu = 0$  and  $\sigma = 1$ , where  $\mu$  is the mean (average) and  $\sigma$  is the

standard deviation from the mean. Standard scores (also called z-scores) of the data were calculated as

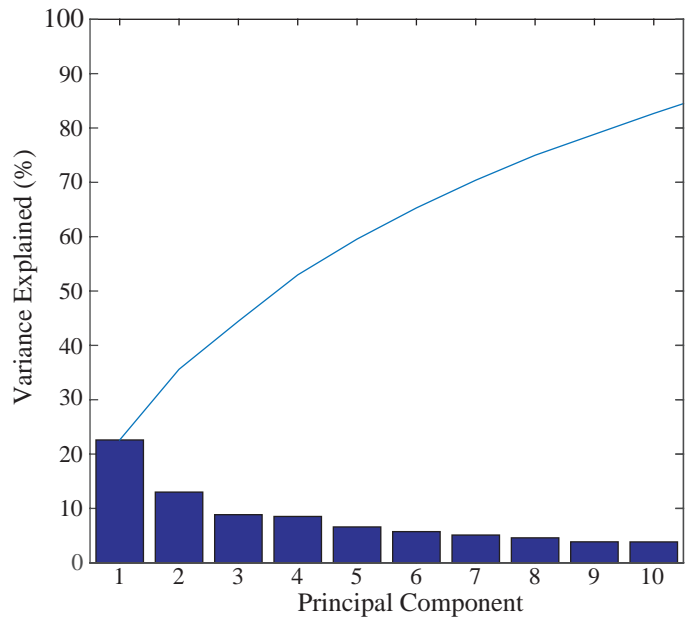
$$z = \frac{x - \mu}{\sigma} \quad (5.1)$$

Standardizing the predictors so that they are centred around 0 with a standard deviation of 1 had a dual importance: on the one hand, it allows computing variables with different units; on the other hand, it is often a requirement for many machine learning algorithms.



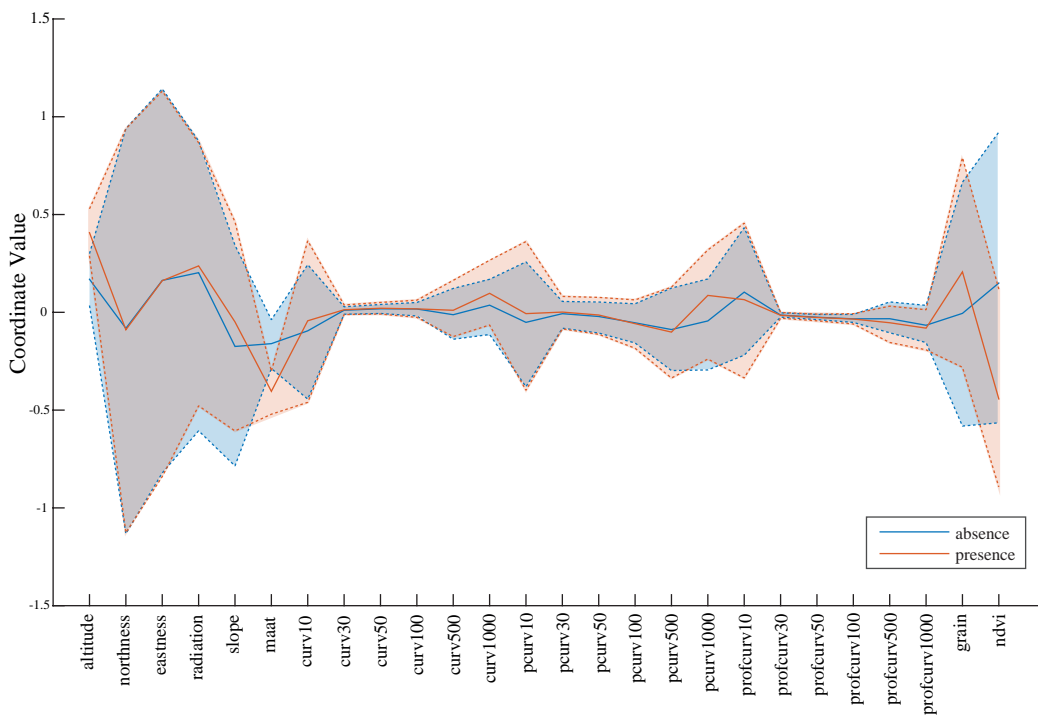
**Fig. 5.2.:** Distribution of standardized values of all the employed environmental predictors (for both positive and negative permafrost evidence). Values outside whiskers are considered as noisy data and were consequently excluded from the dataset. For a description of the variable acronyms, see table 3.1.

A first investigation of the complexity of the data was obtained by performing a principal component analysis (PCA). The plot of the first 10 principal components shows that each component only explains less than 15% of the total variance (at the exception of the first one) (figure 5.3). This is a preliminary insight that informs on the complexity of the dataset and on a potential non-linearity of the phenomenon (from a statistical point of view). Indeed, this indicates that a lot of information is present in the data and that the potentially predictors-class relationship does not depend on just a couple of features.



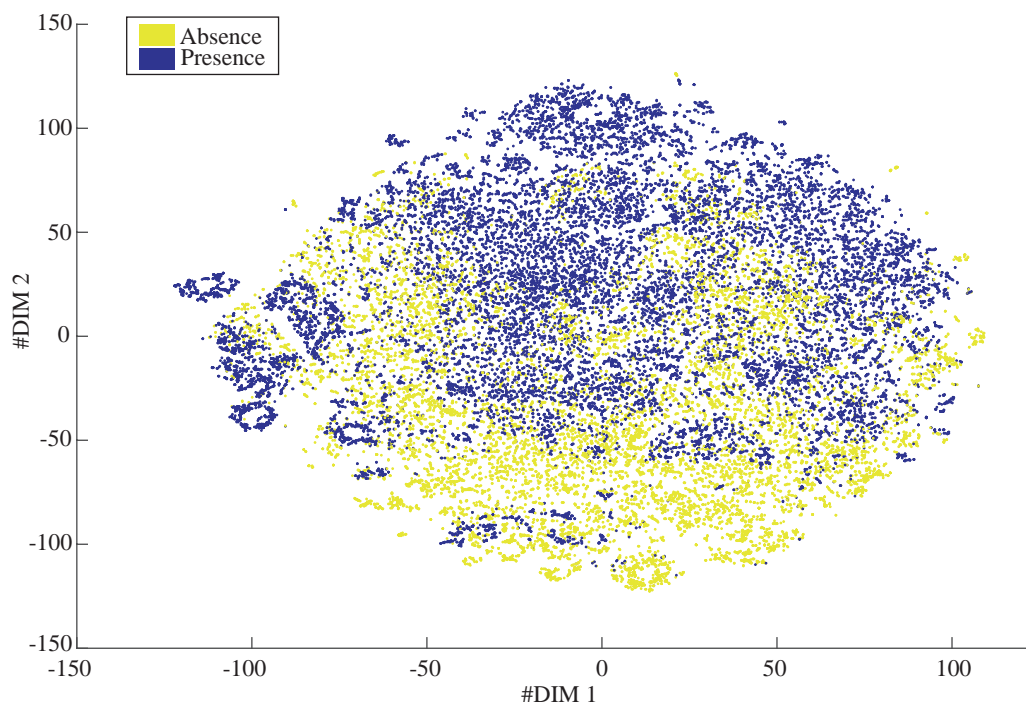
**Fig. 5.3.:** Plot of the PCA principal components illustrating the percentage of explained variance.

An investigation of the data with the parallel coordinate plot (PC plot, presented in section 4.2), allowed exploring the potential relationship between the 26 dimensions of the dataset. As a reminder, with the PC, the predictors are displayed as different vertical axis of a chart and each observation of the dataset is represented as a polyline with the vertices lying on the parallel axes.



**Fig. 5.4.:** Parallel coordinate plot of the permafrost data (median is represented with a bold line, while 25 and 75 percent quartiles are shown by a dashed line).

The computed parallel coordinate plot also indicates that it is not possible to observe a clear relationship between a feature and one of the two predicting classes (figure 5.4). The 93'705 permafrost observations do not allow an interpretation of the plot polylines. In fact, displaying such high number of lines will greatly reduce the readability of the graph. Thus, plot 5.4 only displays the median and the 25 and 75 percent quartile values of each class identified in the dataset. The plot illustrates the parallel coordinates of normalized predictors and shows randomly crossing polylines, which generally indicate no relationship between the dimensions (see section 4.2). This visualisation also shows that no clear structures are clearly observable. In turn, it confirms that *altitude*, *temperature*, *grain size* and *ndvi* better discriminate the permafrost absence from the permafrost presence in comparison to other predictors (as it is illustrated by median parallel polylines that are not parallel).



**Fig. 5.5.:** 2-D manifold of the permafrost data computed with t-SNE, allowing visually unravelling the presence of structures in the dataset.

The visualization of the 26 dimensions of the dataset was performed by applying the t-Distributed Stochastic neighbour Embedding (t-SNE) algorithm to the permafrost data. The 2D embedding of the original dataset allows displaying the presence of structures and thus investigating the complexity of the data. As mentioned in section 4.2, t-SNE is a non-linear dimensionality reduction algorithm that finds patterns in the data and maps the multi-dimensional data to a lower dimensional space and the input features are no longer identifiable. Figure 5.5 shows the detected 2D manifold based on similarity of permafrost observations and their associated features. The algorithm, implemented in the Matlab toolbox for Dimensionality Reduction published by L. van der Maaten (same developer of t-SNE, see Van Der Maaten et al., 2009 for more details), converged to a solution and a stable error

of 0.0245 after 1000 iterations. t-SNE constructed a map in which the separation between permafrost absence and permafrost presence is not well captured (absence of two distinct clusters). As presented by Maaten and Hinton (2008) the algorithm is generally able to efficiently reveal underlying structures in high-dimensional dataset, providing clear discretization between the classes. Nevertheless, in permafrost data it only unravels some structures but does not make the distinction between two clear groups of observations. This visually confirms that the data are complex and possibly not linear.

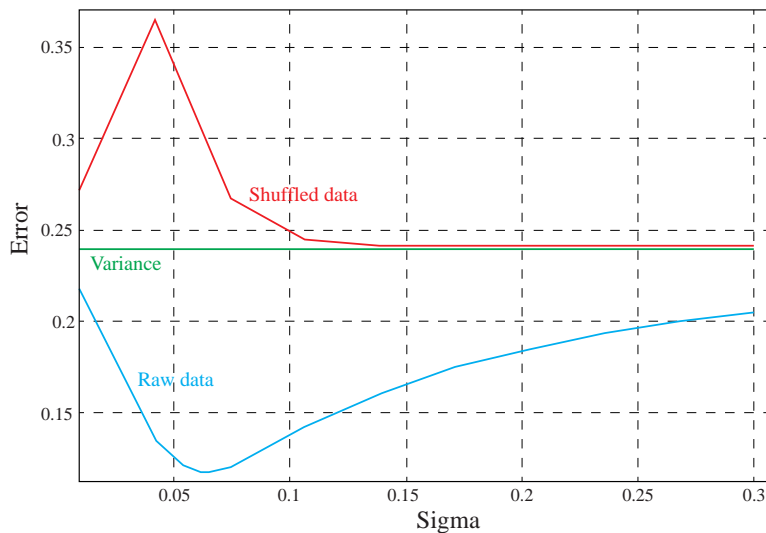
	<i>Linear SVM</i>	<i>RBF-SVM</i>
<b>Precision</b>	0.814	0.859
<b>Recall</b>	0.817	0.86
<b>MSE</b>	0.1834	0.1404
<b>AUROC</b>	0.786	0.83

**Tab. 5.2.:** Summary statistics for a linear SVM and a RBF-SVM, classifying the raw permafrost dataset: precision (positive predictive value), recall (sensitivity), mean squared error (MSE) and area under the ROC curve (AUROC).

The non-linearity of permafrost data was investigated and assessed by classifying the raw dataset with a linear and a non-linear SVM. As explained in section 4.2, data are generally non-linear and complex when the classification performance of a linear classifier is lower than the performance of a non-linear classifier (of the same type). It resulted that RBF-SVM classification measures are generally better than the ones obtained with linear SVM (table 5.2). Indeed, the linear classifier, tested on a 10 cross-validated fold, produced a lower true positive rate and a higher false positive rate than the non-linear variant. The latter also better performed in terms of AUROC. The *complexity index*  $I_C$  (4.1), computed on MSE measures, reported a value of 0.235, identifying a non-linear behaviour between the input predictors and the permafrost occurrence. This shows that the raw dataset was complex and consequently some pre-processing were required. This complexity may indeed be linked to noisy data, redundant environmental predictors or the effective presence of complex structures.

To investigate such presence of structures, the adaptive (anisotropic) general regression neural network (AGRNN) was applied to both permafrost data and a shuffled dataset. By shuffling the class of permafrost evidences, the shuffled dataset does not preserve the original structures. The results of the GRNN training (isotropic version) for the raw and shuffled datasets show that the curve of shuffled data is always above the level of training data variance and does not reach a minimum (figure 5.6). According to the theoretical considerations discussed in section 4.2, these results assessed that structures were present in the permafrost dataset, meaning that patterns exist in the feature space and that non-linear classifiers have to be considered for the success of the study. Indeed, linear classifiers are potentially unable to classify the





**Fig. 5.6.:** Error curves of an AGRNN in regards to selected sigma values.

permafrost dataset with good generalization abilities in comparison to non-linear classifiers, as it has been shown by the quality measures presented in table 5.2.

## 5.2 Statistical feature relevance

The next step of the modelling approach (section 4.1.1) dealt with the study of the relevance of the selected environmental features. The dimensionality of the dataset could be obtained by applying feature extraction and feature selection algorithms. Indeed, the main goal of this step was the extraction of new synthetic datasets or the discard of less statistically relevant predictors, which should provide better generalization of classification models.

### 5.2.1 Feature extraction

The feature extraction techniques presented in section 4.3.2 were used to extract new variables from the original dataset. An intrinsic dimension of 6.77 for the original dataset (rounded to 7) was computed by using the Morisita estimator of ID (MINDID algorithm; see Golay and Kanevski, 2015; implemented in R). The same intrinsic dimensionality was also computed with the maximum likelihood estimation (MLE) implemented into the Dimensionality reduction toolbox. The latter allowed the application of the kernel PCA (using a RBF kernel), the Sammon map and the Local Linear Embedding (using  $k$  optimized at 12) algorithms to permafrost data with the objective to reduce the dimensionality of the dataset to  $d = 7$ . All the algorithms converged to a solution and extracted 3 datasets from the raw dataset. The parallel coordinates of the resulting synthetic dataset are illustrated in figure 5.7. The extracted dimensions were reordered to improve the dataset interpretability.

The order was based on the total Pearson's  $r$  between each two neighbours for all variables and the combination with the highest total correlation was retained. As it was proposed for the raw dataset, only the median, the 25 and 75 percent quartile values are displayed for easier readability and interpretation. Conversely to the plot representing the raw dataset, in the parallel coordinate plot of feature extracted datasets some patterns are more evident. This is particularly true for LLE and Sammon extracted data, where the x-crossing and the parallel polylines are more visible compared to the kPCA, indicating positive and negative relationships between the dimensions (positive and negative correlations between the synthetic variables). The kernel PCA plot illustrates instead more random polylines between the two axes, which are an indication for complexity.

Although the polylines of the three plots present some patterns, it is not possible to distinguish a particular relationship between the dimensions or particular structures that precisely discriminate the permafrost absence or the permafrost presence. Therefore, a further assessment was required to understand how much information was contained in the synthetic datasets.

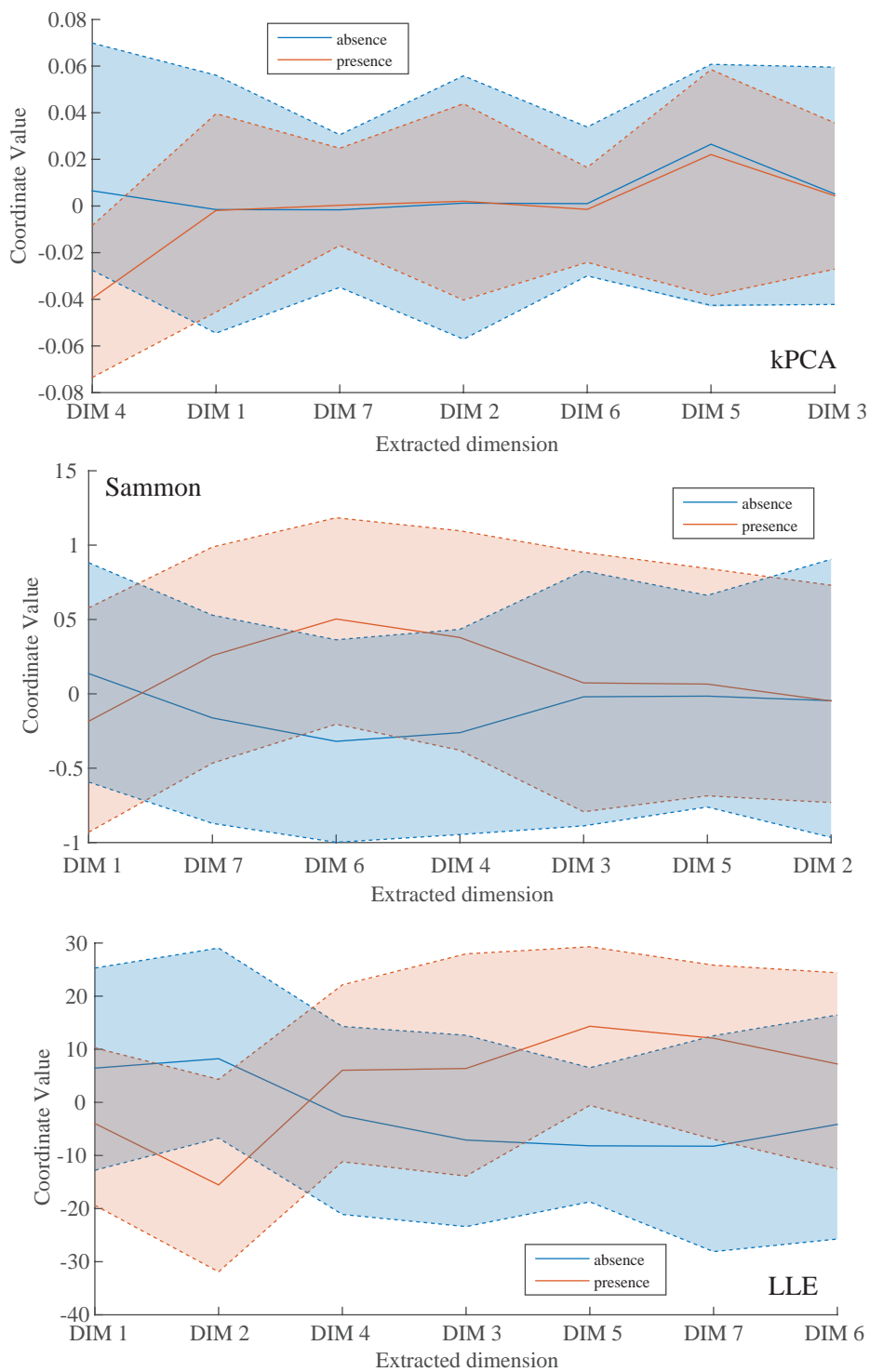
### Effects of feature extraction on the classification of permafrost data

To assess the quality of the employed feature reduction techniques, a linear and a non-linear SVM were also applied to the LLE-, kPCA- and Sammon-extracted datasets. Table 5.3 indicates a decrease in AUCs for the classification on reduced datasets in comparison to the classification of the original raw data. Moreover, the RBF-SVM performances are generally greater than performances obtained for linear SVM, confirming that synthetic datasets remain complex despite being dimensionally reduced by the kPCA, Sammon maps, or LLE. These two observations indicate that, although selected feature extraction techniques were able to reduce the dimensionality of permafrost data, too much noise is still present in the dataset to be able to obtain a good classification, or the data structures were destroyed. The decrease in performances is thus the consequence of such data complexity.

	<i>All features</i>	<i>kPCA</i>	<i>Sammon</i>	<i>LLE</i>
<i>Linear SVM</i>	0.786	0.642	0.508	0.598
<i>RBF-SVM</i>	0.83	0.649	0.51	0.646

**Tab. 5.3.:** Linear and RBF-SVM AUCs for raw and synthetic data.

Since the new extracted dimensions of the three datasets are synthetic and thus physically meaningless (namely not interpretable by an expert of the studied phenomenon), it is impossible to advance any hypothesis on their relevance to the permafrost phenomenon. For this reason, feature selection techniques were also applied to the raw dataset in order to compare the results obtained through feature extraction.

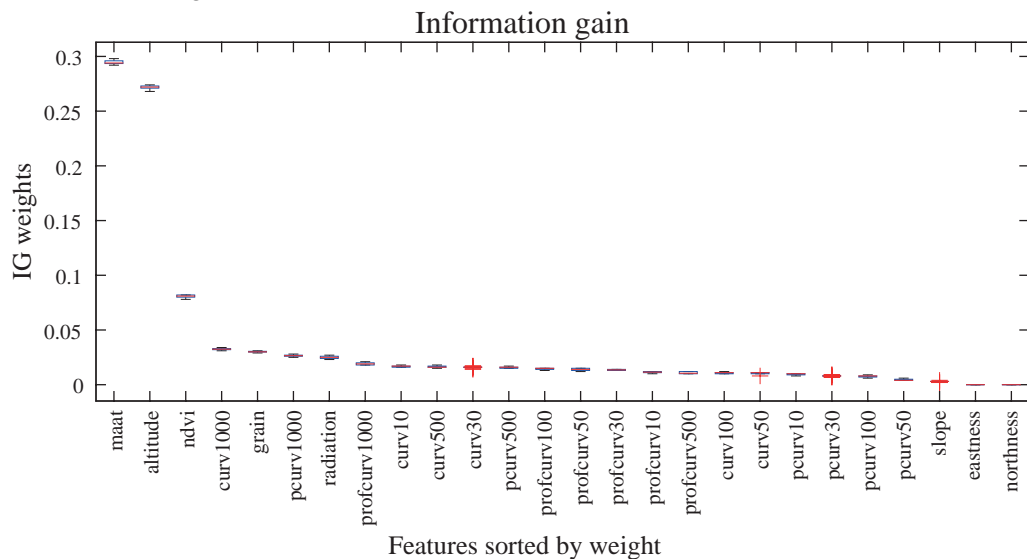


**Fig. 5.7.:** Parallel coordinate plots of the synthetic datasets obtained with kernel PCA, Sammon maps and LLE.

## 5.2.2 Feature selection

Five feature selection algorithms (presented in section 4.3.3) were applied to the raw permafrost dataset: Information Gain, RELIEFF, Correlation-based feature selection, Support Vector Machines recursive feature elimination and Random Forest variable importance. They were computed by partitioning the dataset with a 10-fold cross-validation in WEKA. With the exception of IG, which uses a greedy forward search through the space of attributes (GreedyStepwise search algorithm), the other FS methods evaluate the attributes by their individual evaluations and rank them accordingly.

The ranking sorted by Information Gain (figure 5.8) indicates that *temperature* and *altitude* have a weight that is more than 8 times greater than the one assigned to the other features. In turn, *eastness* and *northness* are ranked as the least useful predictors. Apart for the NDVI, the other features provide almost the same amount of information gain.



**Fig. 5.8.:** Information Gain variable importance, with features ranked according to their IG weights.

A more clear ranking was obtained with RELIEFF (figure 5.9), which also indicated *temperature* and *altitude* as predictors with a consistent average merit and *northness* and *eastness* as features with low merit. Conversely to Information Gain, RELIEFF attributed a greater importance to *radiation*, *grain*, *slope* and *ndvi* in comparison to the curvature predictors.

The same trend is also observable in the ranking obtained by computing a Correlation-based feature selection on the raw dataset (figure 5.10). *Temperature* and *altitude* appear to be two predictors that are very important for the permafrost distribution for the given dataset. *NDVI* and *grain* were again ranked in the first positions, whilst *northness* and *eastness* were again the least useful environmental variables.

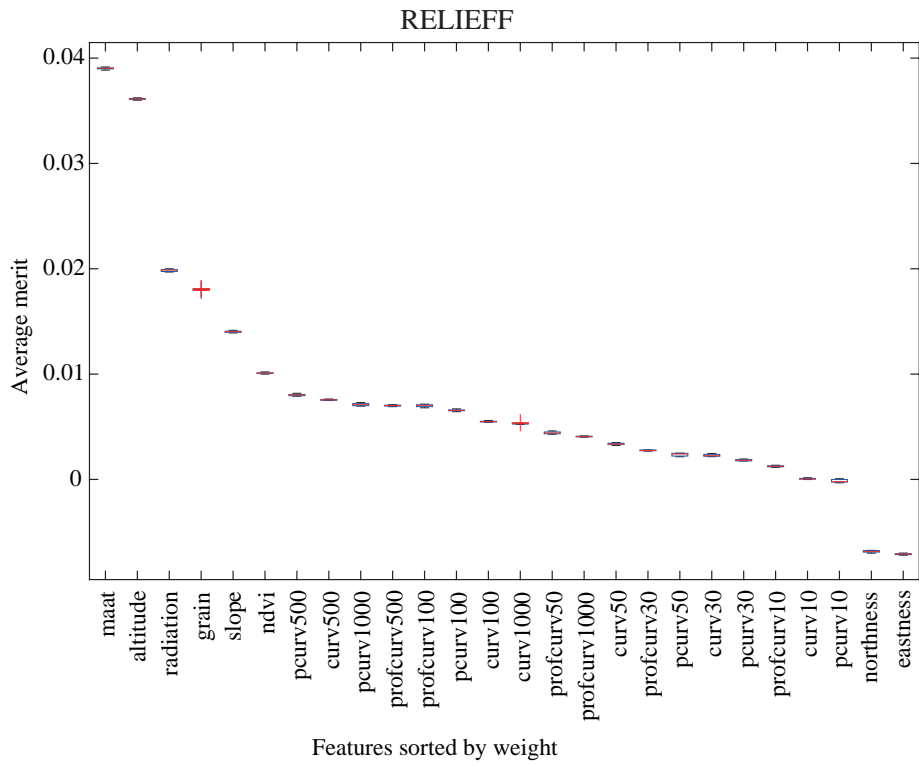


Fig. 5.9.: Variable importance provided by RELIEFF, with features ranked according to their weights.

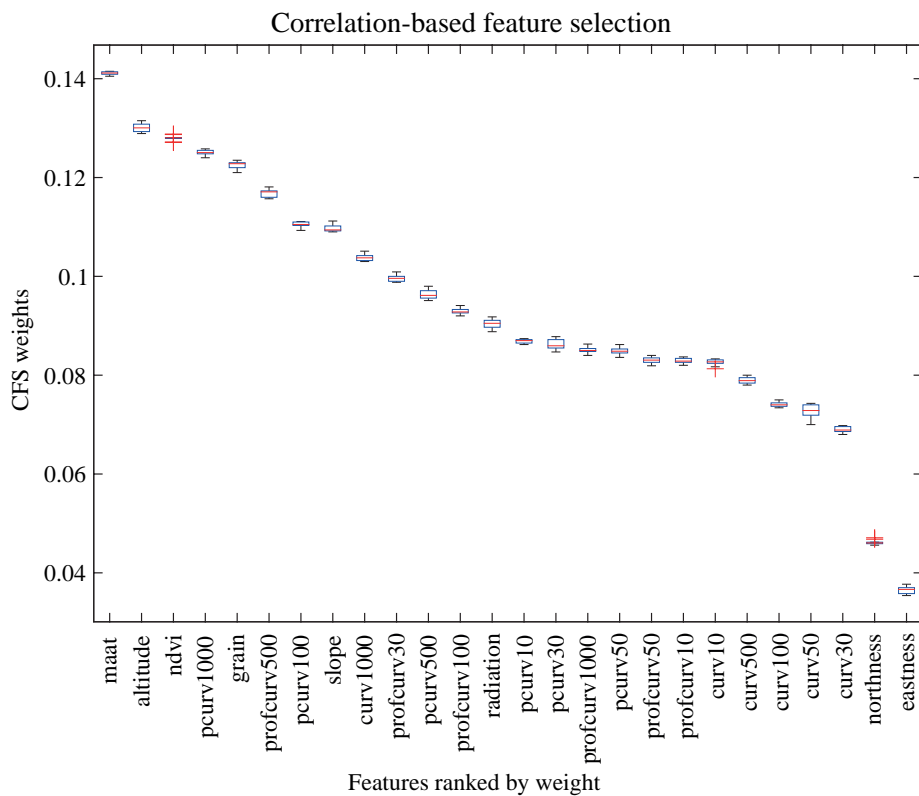


Fig. 5.10.: Correlation-based feature selection for the assessment of variable importance.

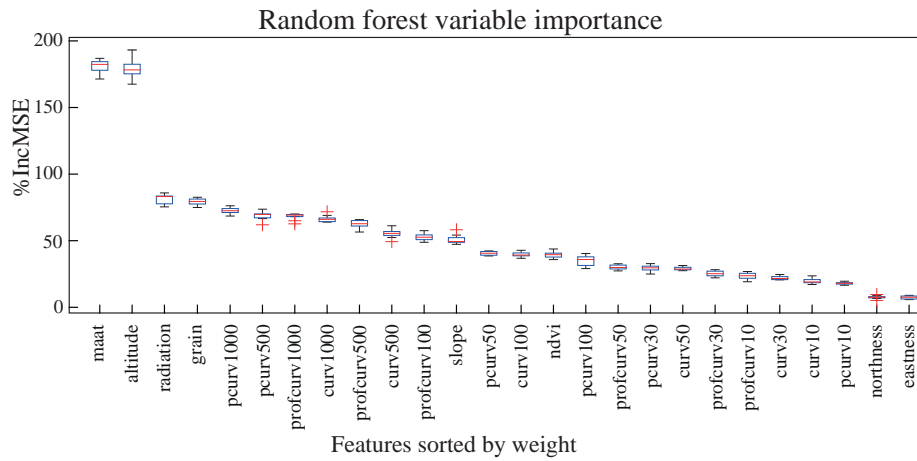


Fig. 5.11.: Variable importance measures embedded in the Random forest algorithm.

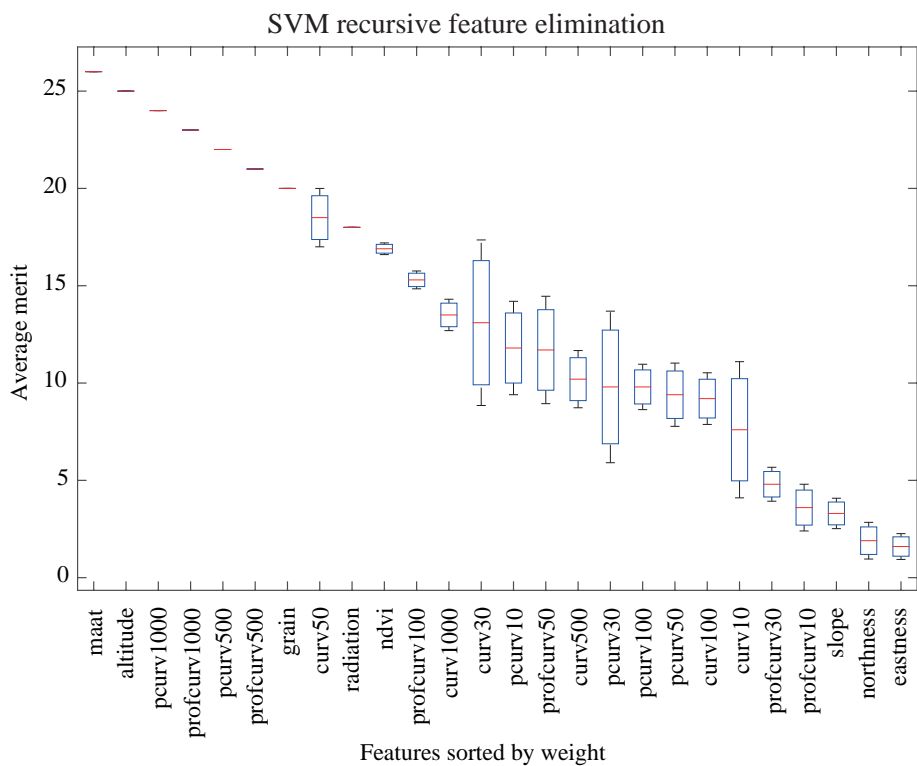


Fig. 5.12.: Feature ranking resulting from a SVM recursive feature elimination procedure.

The statistical relevance computed by using the two FS-embedded techniques (RFVI, figure 5.11, and SVM-RFE, figure 5.12) also confirms that *temperature* and *altitude* were the most relevant predictors. Overall, a large window curvature variable (500-1000 metre window width) appeared to be more relevant than the same variable calculated on a smaller scale.

To sum up, although results slightly differ from an algorithm to another, the same most important variables are generally ranked in the first ten positions. *MAAT* and *altitude* are always considered the most relevant features, while *eastness* and

*northness* the least important ones. It is noteworthy to observe that the other variables have similar weights when compared to each other.

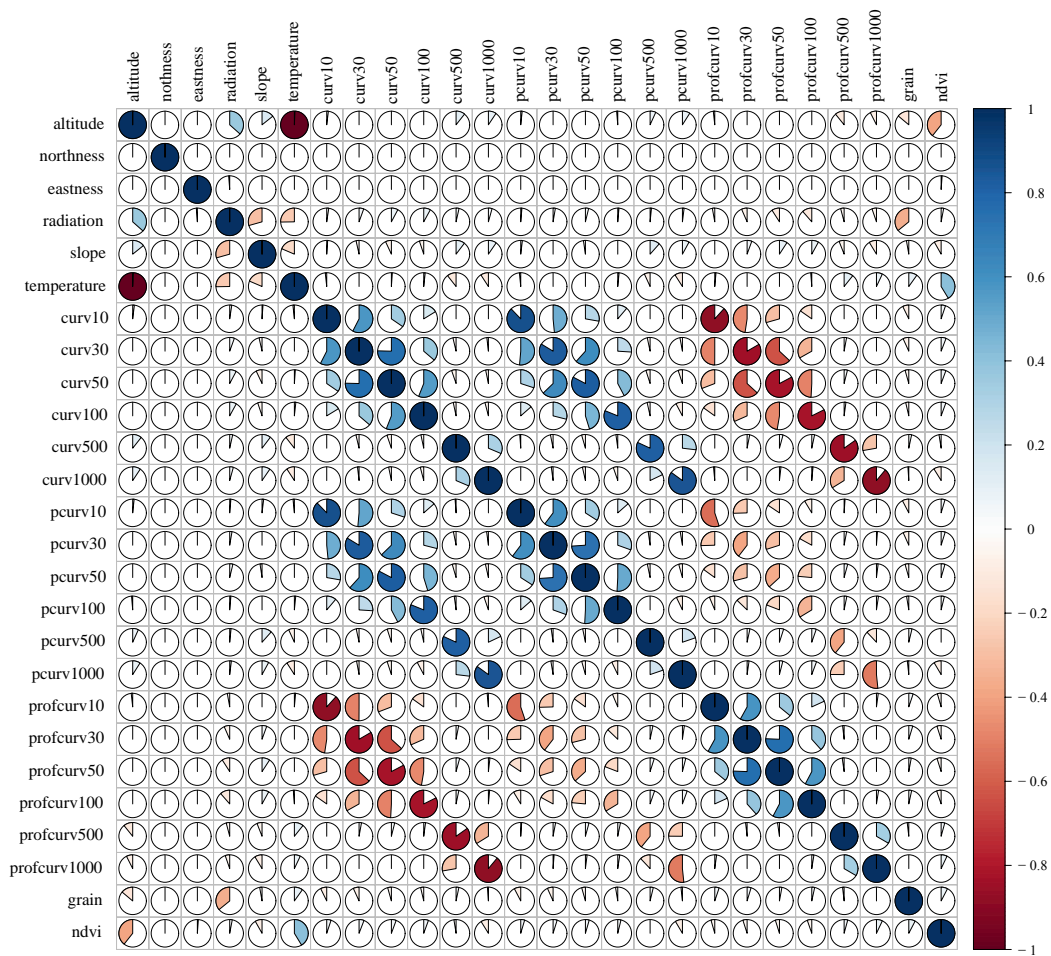
To avoid any subjective choice and the selection of a single feature ranking, a *FS Score* was computed by taking into account the five overall rankings. This measure is the result of a rescaling of the merit of each feature between 0 and 1 and the average of its relevancies. The *FS Score* ranged between 0 (most relevant predictor) to 1 (least relevant predictor) as well. Table 5.4 illustrates the final feature ranking for the adopted environmental variables.

Following the approach proposed by Dormann et al. (2012), correlated predictors (with more than 0.75 or less than  $-0.75$ ; figure 5.13) were discarded from the dataset according to their *FS Score* relevance. *MAAT* and *altitude* resulted to be the most correlated, which was expected since the *MAAT* was computed with a linear gradient and it decreases with the increase of altitude (see section 3.3). The curvature indices (*curv*, *pcurv* and *profcurv* in figure 5.13) also appear to be correlated between each other. Curvature indices (*curv*) are generally positively correlated to the planar curvatures (*pcurv*), whilst they are negatively correlated to the profile curvature indices (*profcurv*).

Pairs of predictors were then analysed: when the correlation between two variables exceeded the  $\pm 0.75$  threshold, the least relevant predictor was removed from the dataset. This procedure was used for all features with the exception of altitude and temperature. Indeed, although the latter had a better *FS Score*, it was calculated on the basis of empirical data collected by two weather stations, thus, prone to errors (see section 3.3.1). Temperature was finally discarded because its value distribution was considered very similar to the altitude one and it was directly derived from the DEM covering the region of interest. Table 5.5 presents the cleaned dataset with the 15 retained environmental predictors to be employed for the classification of permafrost observations.

<i>Feature</i>	<i>FS Score</i>
maat	0
altitude	0.04
pcurv1000	0.16
grain	0.17
ndvi	0.181
profcurv500	0.248
pcurv500	0.347
curv1000	0.387
radiation	0.414
profcurv100	0.429
profcurv1000	0.447
pcurv100	0.479
curv500	0.486
slope	0.507
profcurv50	0.537
curv100	0.567
profcurv30	0.593
curv50	0.615
curv30	0.684
pcurv30	0.703
pcurv50	0.725
pcurv10	0.748
curv10	0.773
profcurv10	0.807
northness	0.968
eastness	0.987

Tab. 5.4.: Feature merit.



**Fig. 5.13.:** Correlation matrix for all environmental variables. Pie plots indicate how positively or negatively correlated each pair of predictor is. Blue pies refer to positively correlated predictors, whilst red pies indicate negatively correlated pairs.

<i>Variable name [label]</i>	<i>Variable type</i>
Altitude [altitude]	Continuous
Northness [northness]	Continuous
Eastness [eastness]	Continuous
Slope [slope]	Continuous
PISR [radiation]	Continuous
NDVI [ndvi]	Continuous
Grain size [grain]	Continuous
Planform curvature 10 m [pcurv10]	Continuous
Planform curvature 30 m [pcurv30]	Continuous
Planform curvature 100 m [pcurv100]	Continuous
Planform curvature 500 m [pcurv500]	Continuous
Planform curvature 1000 m [pcurv1000]	Continuous
Profile curvature 50 m [profcurv50]	Continuous
Profile curvature 100 m [profcurv100]	Continuous
Profile curvature 500 m [profcurv500]	Continuous

**Tab. 5.5.:** List of the 15 environmental predictors retained in the clean dataset, according to their relevance computed with feature relevance techniques. The table also shows the type, as well as the used abbreviations of each variable.



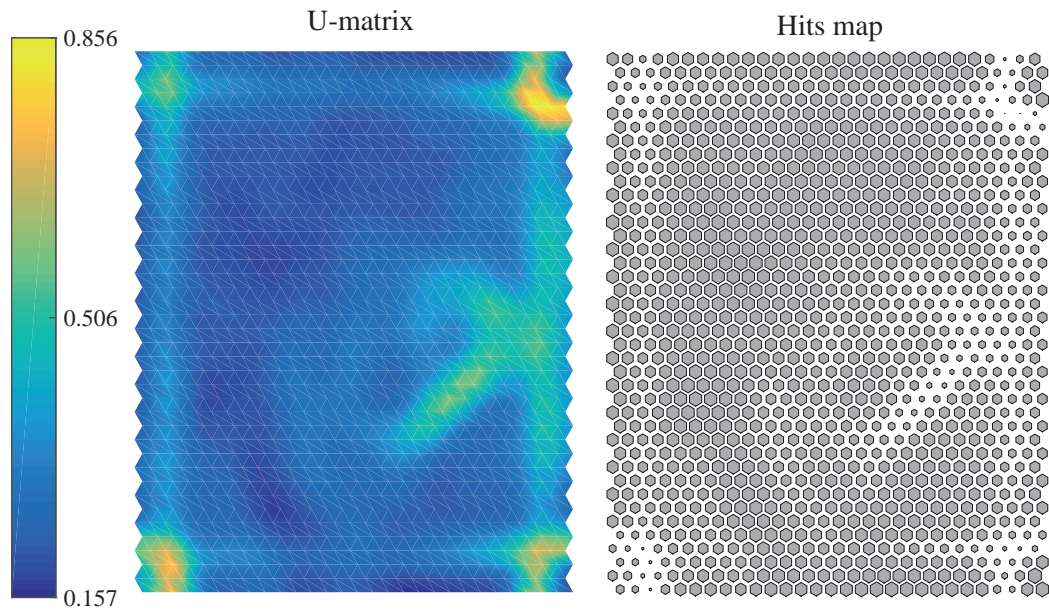
## 5.3 Subset sampling

The cleaned dataset obtained by removing the least relevant and correlated features was sampled according to the strategy presented in section 4.6. The first step mainly involved the clustering of the feature space with SOM (see figure 4.6), in order to split the data in 10 representative pools equally distributed. The computed self-organised map allowed the 2D visualization and the extraction of the topological structure from the permafrost dataset.

To accomplish this task, the Matlab toolbox for Self-Organizing Maps created by Vesanto et al. (2000) was used. This complete library of functions allows creating, displaying and clustering the SOM maps. Prior to the training phase, the topology structure of the map and the number of map units had to be determined. A hexagonal lattice was used and the number of units was calculated according to the heuristic formula  $munits = 5 * \sqrt{\text{number of observations}}$  ( $munits = 5 * \sqrt{93705} \approx 1530$ ). Once the number of map units was determined, the map size was also determined. The two biggest eigenvalues of the training data were calculated and the ratio between sidelengths of the map grid was set to this ratio. The actual sidelengths were then set so that their product was as close to the desired number of map units as possible. The computed map size was selected to be 45x34 units. Then the SOM was initialized. First, a linear initialization along two greatest eigenvectors was tried, but if this could not be done (the eigenvectors could not be calculated), a random initialization could be used instead. After initialization, the SOM was trained in two phases: first, rough training; then, fine-tuning. At the end of these training phases, two error measures were computed: the final quantization error (2.981) and the final topographic error (0.078). The first measured the average distance between each data vector and its best matching unit (BMU). It can be seen as a measure of the map resolution. The second error corresponded to the proportion of all data vectors whom first and second BMUs were not adjacent units. The second error was a measure of the topology preservation. The final combined error of the SOM applied to the cleaned permafrost dataset was 3.184.

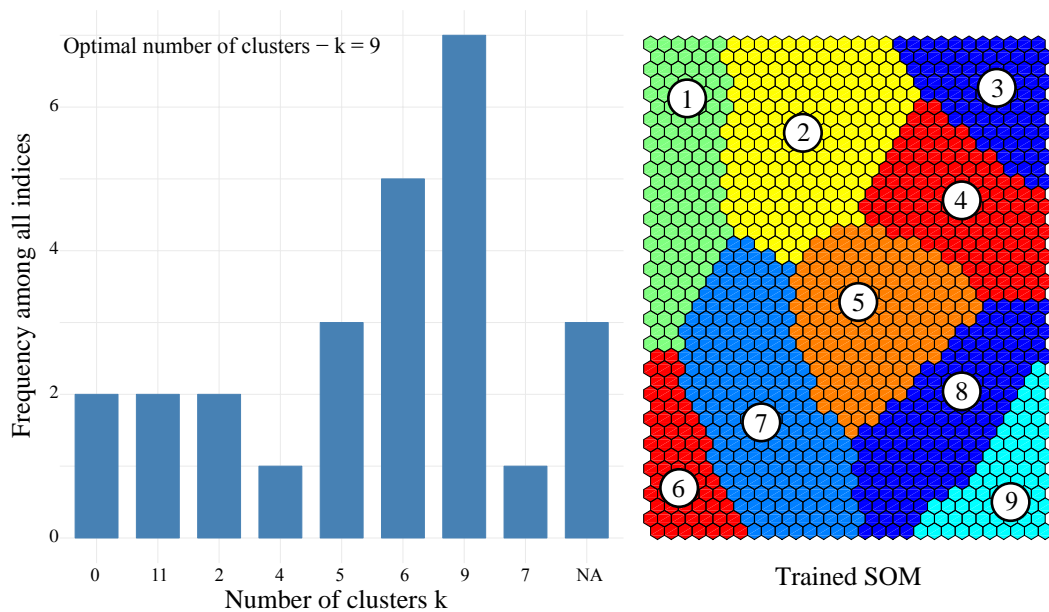
Figure 5.14 illustrates the matrix of distances (U-matrix) between neighbouring map neurons, that had to be clustered in order to find groups of similar observations, and the respective hits map, that represents how many times each neuron won. The larger the hexagon is, the more observations were associated to the neuron.

When clustering data, the choice of the possible number of clusters is never univocal as it can be discovered by applying cluster validity indices to the data vectors. Indeed, in this research, the optimal number of clusters were identified by computing 26 cluster indices that are implemented in the *NbClust* R package. The latter implements the most known and used clusters indices, such as the Krzanowski and Lai Index,



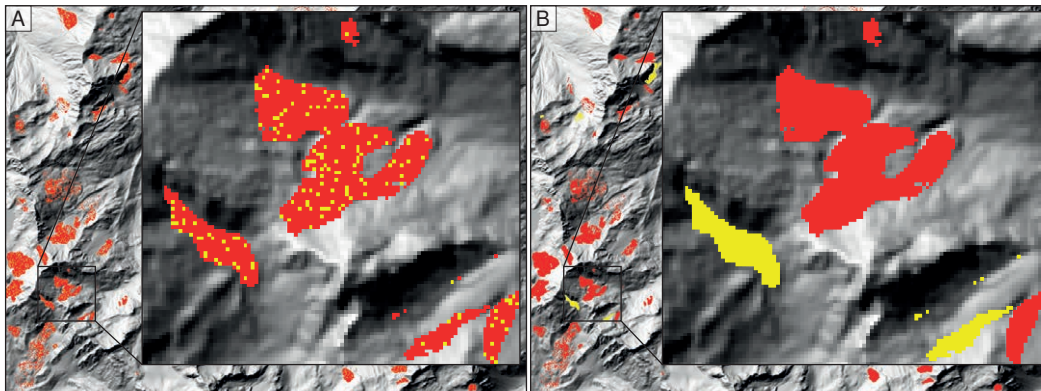
**Fig. 5.14.:** SOM U-matrix and hits map of the permafrost input space.

the Calinski-Harabasz Index, the Davies and Bouldin Index, etc. The complete list of indices and corresponding references can be found in the package documentation (<https://cran.r-project.org/web/packages/NbClust/NbClust.pdf>). This procedure provided the best clustering scheme by varying all combinations of number of clusters, distance measures and clustering methods. Figure 5.15 (left image) shows the selection of the best number of clusters according to a majority rule proposed among all indices.



**Fig. 5.15.:** The majority vote for the assessment of the optimal number of clusters (left); clustering of the SOM neurons according to the NBClust result.

By using a k-means clustering algorithm on the trained SOM, a partition of the 2D map was thus performed (figure 5.15, right image). From the 9 identified clusters, 10 pools of data were then created by proportionally gathering permafrost observations associated to the neurons contained in the clusters.



**Fig. 5.16.:** Difference in the selection of training and test observations according to the random (A) or the spatial sampling strategy (B) for the Yettes-Condja sector (Mont Gelé area).

The second step of the sampling strategy required to partition these 10 pools of data into training, validation and test sets. Spatial and random 10-fold cross-validations were used on the 10 subsets. The latter were thus partitioned into 10 equal sized sub-subsets. This procedure resulted in a total of 100 repetitions for each of the selected partitioning method. Figure 5.16 shows the difference between the two sampling techniques. With random cross-validation, the training set was selected randomly across the geographical space, meaning that a training observation can potentially be located next to a test observation (5.16.A). Conversely, with spatial cross-validation, the partition of the data was performed at the polygon level, meaning that when a training observation is located on a rock glacier (for example), no test observations can be selected within the same polygon (5.16.B). This was meant to avoid the sampling of test samples very similar to the training examples. It is worth mentioning however that figure 5.16 only has an explanatory purpose, because, with SOM clustering, not all pixels of the same polygon are gathered into the same pool of data.

## 5.4 Model performances, interpretation and comparison

### 5.4.1 Classification of permafrost data

This section presents the implementation of *Logistic regression*, *linear* and *RBF-SVM*, *Multilayer perceptron* and *Random forest* applied to permafrost data. It also investigates the performances obtained by the mentioned algorithms by comparing

measures of the classification quality, such as *true positive*, *true negative*, *false positive* and *false negative* rates, *root mean squared error* and *area under the ROC curve* of all the selected machine learning classifiers. In addition to these rates, an analysis of the statistical significance of each model is presented as well as an assessment of the most suitable technique for permafrost mapping.

### Logistic regression

Permafrost data were classified by using the Logistic regression in a WEKA environment. The algorithm provides a probability of a given permafrost observation to belong to the presence or the absence class. In other terms, a prediction equal to 0 indicates that permafrost absence is absolutely certain and, conversely, a model output equal to 1 indicates certainty of permafrost presence. Any value above 0.5 predicts the positive class (presence); likewise, values below 0.5 correspond to the negative class (absence).

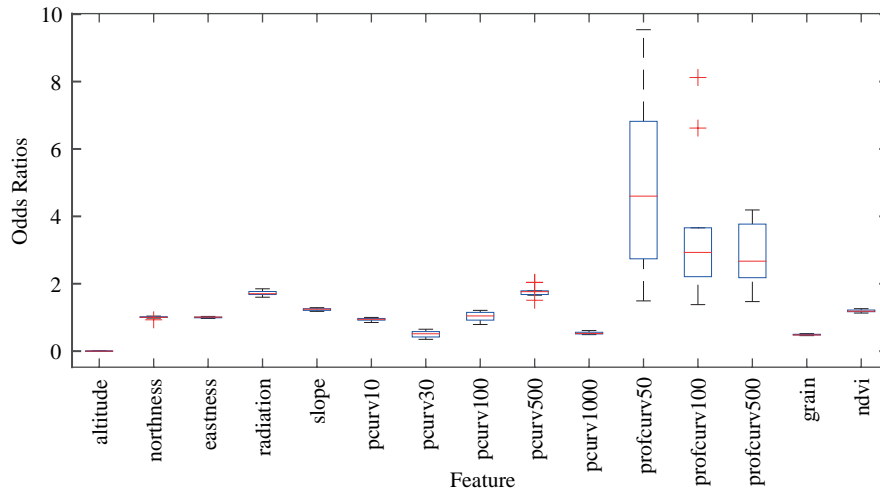
In addition to classic quality measures of the model performance (see sector 4.8), LR provides the so-called *odds ratios*, which are a measure of the effect of each predictor in the respect of the predicted class. For continuous predictors (which is the case for employed permafrost features), odds ratios that are greater than 1 indicate that the occurrence of the class is more likely to occur as the predictor increases. Conversely, odds ratios that are smaller than 1 indicate that the class label is less likely to occur as the predictor increases. Figure 5.17 and table 5.6 show the odds ratios obtained applying LR on permafrost data. The displayed values correspond to the mean odds ratios computed on the 10 repetitions made for each pool of data (100 repetitions in total).

Variable	pool 1	pool 2	pool 3	pool 4	pool 5	pool 6	pool 7	pool 8	pool 9	pool 10
altitude	0	0	0	0	0	0	0	0	0	0
northness	1.04	0.98	1	1.01	1	0.93	1.01	1.02	1.03	1
eastness	0.98	1.01	1.01	0.99	0.99	0.97	1.02	1.01	1.01	1.03
radiation	1.71	1.61	1.77	1.69	1.78	1.85	1.71	1.68	1.6	1.7
slope	1.29	1.24	1.26	1.25	1.21	1.21	1.18	1.18	1.26	1.25
pcurv10	0.97	0.97	0.92	0.85	1	0.96	0.93	0.88	0.93	0.96
pcurv30	0.54	0.35	0.47	0.56	0.58	0.42	0.49	0.62	0.65	0.37
pcurv100	1.02	1.15	1.08	1.07	0.79	1.21	0.92	1.18	0.88	0.94
pcurv500	1.79	2.04	1.68	2.04	1.51	1.74	1.77	1.76	1.77	1.66
pcurv1000	0.52	0.51	0.51	0.49	0.51	0.61	0.56	0.53	0.55	0.57
profcurv50	9.54	1.491	5.49	2.74	3.73	4.78	7.08	4.42	6.82	2.59
profcurv100	8.12	2.21	2.87	6.62	2.64	1.76	3.66	3.05	1.38	2.99
profcurv500	2.64	4.19	2.01	3.82	1.47	3.77	2.7	2.18	3.38	2.25
grain	0.48	0.47	0.48	0.48	0.49	0.48	0.52	0.5	0.46	0.52
ndvi	1.19	1.22	1.19	1.18	1.17	1.15	1.18	1.23	1.26	1.13

Tab. 5.6.: Odds ratios in the respect of the *absence* class label.

Odds ratios were calculated in the respect of the *absence* of the label class. As a consequence, values lower than 1 indicate that the predicted class is more likely to be predicted as permafrost *presence* if the predictor increases. A perfect example is given

by *altitude* that scored an odds ratio of 0, meaning that permafrost presence is more likely to occur as this variable increases. Other variables that were identified having the same behaviour are *planform curvature* indices *pcurv10*, *pcurv30*, *pcurv1000* and *grain* size. In turn, predictors with odds ratios that are greater than 1 are more likely associated to permafrost absence if their value increases. *Radiation*, for example, scored an odds ratio of 1.71, indicating that high values of radiation are an indication of permafrost absence. The *profile curvature* indices also scored high odds ratios, indicating that the class label *absence* is more likely to occur when this predictor has high values, corresponding to concave surfaces (in the slope direction).



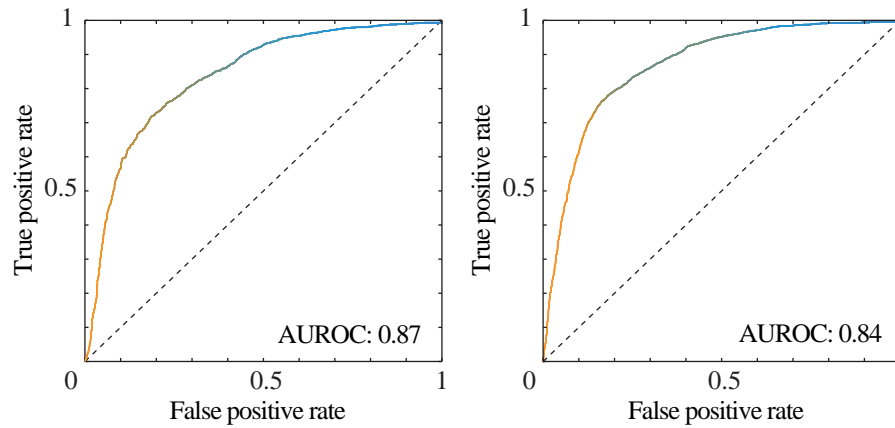
**Fig. 5.17.:** Box-plots of the odds ratios for retained features.

Measure	Random sampling		Spatial sampling	
	Train statistics	Test statistics	Train statistics	Test statistics
Accuracy	81.380	81.348	85.175	79.565
RMSE	0.364	0.366	0.333	0.386
TP rate	0.814	0.675	0.852	0.651
FP rate	0.246	0.112	0.211	0.121
TN rate	0.754	0.888	0.789	0.879
FN rate	0.186	0.325	0.148	0.349
Precision	0.811	0.766	0.850	0.739
Recall	0.814	0.675	0.852	0.651
AUROC	0.874	0.870	0.904	0.844

**Tab. 5.7.:** Quality measures of the LR classification on randomly and spatially sampled data.

The classification of permafrost data converged to a solution, providing an AUROC value of 0.87 (table 5.7) for sub-sets that were obtained with a random sampling strategy and a value of 0.844 for spatially sampled sub-sets. Test statistics, computed on independent test sets with cross-validation, generally indicate that the classification of randomly sampled data offer higher true positive and true negative rates. This translates in a lower precision (sensitivity) and a lower recall (specificity) of spatially sampled sub-sets. In other words, the LR models applied to randomly sampled data have a slightly greater ability to correctly classify permafrost presence (higher sensitivity, or less false positives) as well as to correctly classify permafrost absence (higher specificity, or less false negatives). This observation was also confirmed by

plotting the AUROC curves, which illustrate the trade-off between sensitivity and specificity obtained with the Logistic regression (figure 5.18). For an additional overview of the LR performances, see appendix A.



**Fig. 5.18.:** LR ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.87) and spatially sampled data (right image, AUROC: 0.84).

### Linear SVM

The classification of permafrost data with linear Support Vector Machines was done by using the LIBLINEAR package (LIBSVM library) implemented in WEKA. As a recall, the linear SVM decision function looks for a hyperplane that separates the classes as best as possible by maximising the distance between the two classes. Once the linear SVM fits the data, some feature weights can be obtained. They represent the vector coordinates which are orthogonal to the hyperplane and their direction indicates the predicted class (presence or absence). The absolute size of the coefficients in relation to each other can then be used to determine the feature importance. The magnitude of the feature weight is presented in table 5.8. It shows the mean of computed weights for each of the adopted pool of data and the overall weight of the employed environmental predictors.

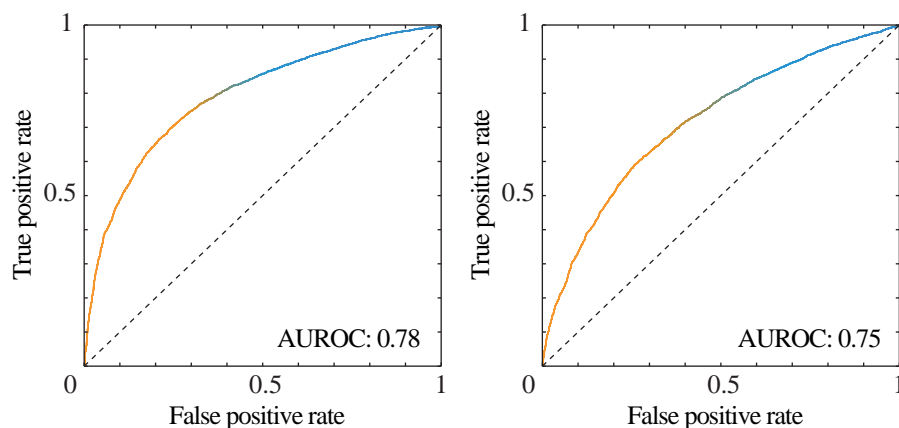
Variable	pool 1	pool 2	pool 3	pool 4	pool 5	pool 6	pool 7	pool 8	pool 9	pool 10	Mean
altitude	3.37 (+)	3.29 (+)	3.31 (-)	3.28 (+)	3.31 (+)	3.39 (+)	3.26 (+)	3.22 (-)	3.46 (-)	3.24 (+)	3.313 (+)
northness	0.01 (-)	0 (-)	0.01 (-)	0.01 (-)	0.01 (-)	0 (-)	0 (-)	0 (+)	0.03 (-)	0.01 (-)	0.008 (-)
eastness	0.01 (-)	0.01 (-)	0 (+)	0 (-)	0 (+)	0 (+)	0.01 (-)	0.01 (+)	0.01 (-)	0 (-)	0.005 (-)
radiation	0.20 (-)	0.21 (-)	0.18 (+)	0.20 (-)	0.20 (-)	0.21 (-)	0.20 (-)	0.02 (+)	0.23 (+)	0.18 (-)	0.201 (-)
slope	0.09 (-)	0.08 (-)	0.07 (+)	0.06 (-)	0.08 (-)	0.06 (-)	0.06 (-)	0.08 (+)	0.06 (+)	0.08 (-)	0.072 (-)
pcurv10	0.01 (+)	0.03 (+)	0.01 (-)	0.04 (+)	0.06 (+)	0 (-)	0.03 (+)	0.02 (-)	0.01 (-)	0.03 (+)	0.024 (+)
pcurv30	0.21 (+)	0.28 (+)	0.37 (-)	0.17 (+)	0.19 (+)	0.2 (+)	0.23 (+)	0.33 (-)	0.31 (-)	0.15 (+)	0.244 (+)
pcurv100	0.01 (+)	0.03 (-)	0.04 (+)	0.05 (-)	0.02 (-)	0.07 (+)	0.04 (+)	0.03 (-)	0.06 (+)	0.05 (+)	0.04 (+)
pcurv500	0.22 (-)	0.20 (-)	0.27 (+)	0.22 (-)	0.27 (-)	0.17 (-)	0.23 (-)	0.21 (+)	0.22 (+)	0.23 (-)	0.224 (-)
pcurv1000	0.22 (+)	0.22 (+)	0.22 (-)	0.20 (+)	0.24 (+)	0.22 (+)	0.19 (+)	0.19 (-)	0.17 (-)	0.2 (+)	0.207 (+)
profcurv50	0.73 (-)	1.17 (-)	1.41 (+)	1.21 (-)	0.36 (-)	1.14 (-)	1.46 (-)	0.31 (+)	1.38 (+)	0.60 (-)	0.977 (-)
profcurv100	0.93 (-)	0.43 (-)	0.32 (+)	0.45 (-)	0.65 (-)	0.38 (-)	0.49 (-)	0.41 (+)	0.22 (+)	0.12 (-)	0.44 (-)
profcurv500	0.36 (-)	0.22 (-)	0.48 (+)	0.29 (-)	0.47 (-)	0.15 (-)	0.34 (-)	0.30 (+)	0.46 (+)	0.44 (-)	0.351 (-)
grain	0.25 (+)	0.25 (+)	0.26 (-)	0.23 (+)	0.25 (+)	0.24 (+)	0.22 (+)	0.23 (-)	0.25 (-)	0.26 (+)	0.244 (+)
ndvi	1.19 (-)	0.06 (-)	0.08 (+)	0.83 (-)	0.06 (-)	0.06 (-)	0.06 (-)	0.05 (+)	0.05 (+)	0.09 (-)	0.066 (-)

**Tab. 5.8.:** Linear SVM feature weights and their position with respect to the margin (positive or negative class).

The highest weight was obtained for *altitude*, with a mean value of 3.313. In comparison with the weight obtained by other predictors, its magnitude is 10-15 times more important. Indeed, computed values for other employed environmental features are generally close to 0 (except for *profcurv50*). This indicates that the linear classification in the feature space is possible thanks to the linear relationship occurring between the permafrost occurrence and *altitude*. Modelling quality measures computed for linear SVM confirm that a classification of retained permafrost data is possible, even though the result is indicative of a moderate predictive ability (mean AUROC of 0.751 and 0.783, depending on the sampling strategy; see figure 5.19). The linear SVM classification performances (table 5.9) indicate the same trend observed for LR, with the test accuracy and the AUROC generally lower for spatially sampled observations. Indeed, although the RMSE and the precision are very similar, the models differ in recall (specificity), indicating that the linear SVM is not generally able to correctly classify permafrost absence. In fact, despite an important TN rate, the classifier predicted a consistent number of false negatives, resulting in a FN rate of 0.405. The classification of randomly sampled sub-sets also presented the same trend, although the precision-recall difference is less pronounced.

Measure	Random sampling		Spatial sampling	
	Train statistics	Test statistics	Train statistics	Test statistics
Accuracy	81.828	81.487	80.223	79.266
RMSE	0.426	0.430	0.445	0.454
TP rate	0.818	0.674	0.802	0.595
FP rate	0.241	0.109	0.259	0.094
TN rate	0.759	0.891	0.741	0.906
FN rate	0.182	0.326	0.198	0.405
Precision	0.816	0.769	0.799	0.768
Recall	0.818	0.674	0.802	0.595
AUROC	0.789	0.783	0.771	0.751

**Tab. 5.9.:** Quality measures of the linear SVM classification on randomly and spatially sampled data.



**Fig. 5.19.:** Linear SVM ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.78) and spatially sampled data (right image, AUROC: 0.75).

Lastly, the probability estimates were done internally in WEKA, instead of obtaining the classical -1/+1 SVM classification, which is hard to interpret for new permafrost predictions. A complete overview of linear SVM performances is presented in appendix A.

### Radial Basis Function (RBF) SVM

Permafrost data were also classified by using the non-linear expansion of SVM, which employs a Gaussian radial basis function (RBF) kernel to map the feature space into a higher dimensional space and generally provides good classification performances. An independent validation set was required in order to select the best model parameters  $C$  and  $\sigma$ . This sub-set was created by taking a portion of the training set (66% in training, 33% in validation). The optimization of the kernel bandwidth  $\gamma$  and the cost  $C$  was done by adopting a grid search algorithm, which sequentially looked for the best couple of parameters by assessing the resulting classification with a quality measure. Both RBF-SVM and grid search algorithm are implemented in WEKA that integrates the *LIBSVM* library.

For this study, the grid parameter selection was assessed on the ROC curve rather than on the overall accuracy, which is often misleading depending on the number of false positives and negatives. The size of the selected grid space ranged between  $10^{-3}$  to  $10^3$  for the cost  $C$  and 1 to 4 for  $\gamma$ . The maximum value of the latter was selected on the basis of the dimensionality of the permafrost cleaned dataset (15-dimensional feature space). Indeed, the upper bound kernel bandwidth corresponds to the square root of the 15 dimensions of the dataset ( $\sqrt{15} = 3.8730$ , rounded to 4). The grid search hyperparameter optimization were recomputed for all repetitions on the sub-sets of each pool of data.

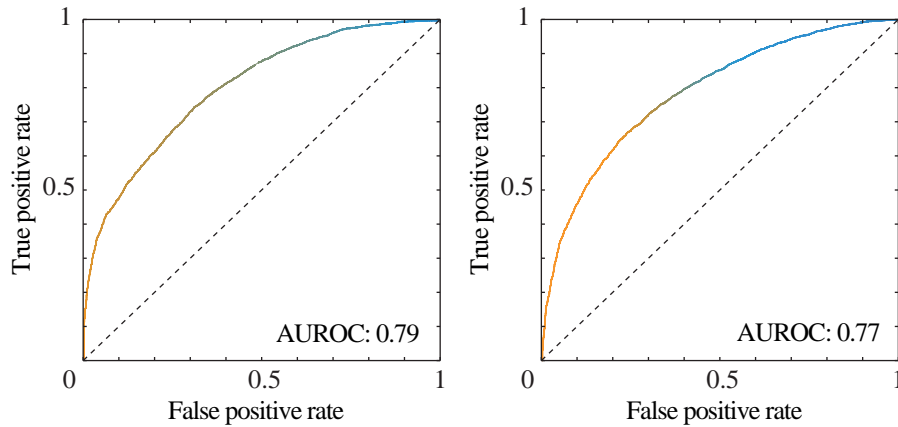
Measure	Random sampling		Spatial sampling	
	Train statistics	Test statistics	Train statistics	Test statistics
Accuracy	86.142	83.310	84.741	79.696
RMSE	0.389	0.408	0.419	0.449
TP rate	0.741	0.651	0.736	0.650
FP rate	0.135	0.069	0.197	0.118
TN rate	0.865	0.931	0.803	0.882
FN rate	0.259	0.349	0.264	0.350
Precision	0.821	0.837	0.801	0.744
Recall	0.741	0.651	0.736	0.650
AUROC	0.827	0.791	0.806	0.766

**Tab. 5.10.:** Quality measures of the RBF-SVM classification on randomly and spatially sampled data.

The performances obtained by classifying permafrost data with RBF-SVM are slightly better than performances obtained with linear SVM (table 5.10). Although the non-linear SVM was able to produce a high number of a *true negative* rate, its ability of predicting *true positives* is moderate, as confirmed by a *false negative* rate of 35%. The AUROC on the test sets is similar, with 0.791, for model obtained on



randomly sampled data, and 0.766 for spatially sampled sub-sets (figure 5.20). Refer to RBF-SVM statistics in appendix A for further details.



**Fig. 5.20.:** RBF-SVM ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.79) and spatially sampled data (right image, AUROC: 0.77).

### Multilayer Perceptron

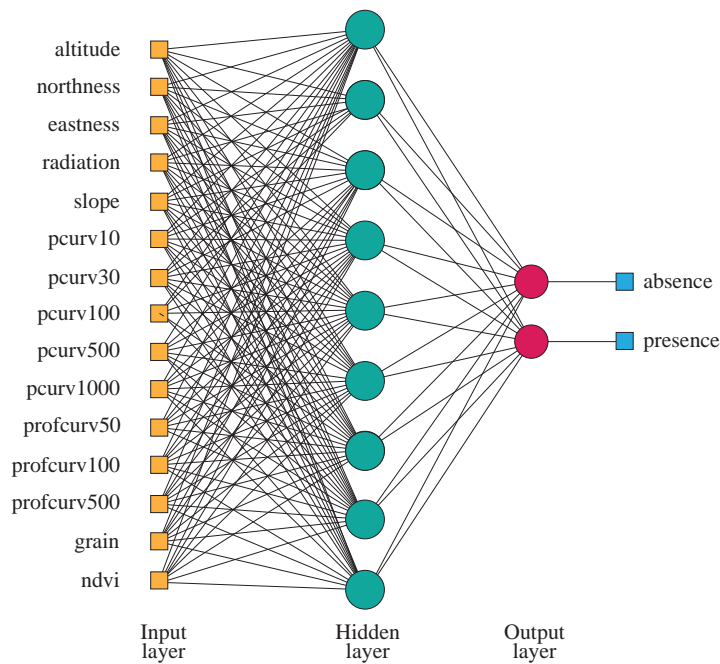
The Neural network were implemented to classify permafrost observations and their associated environmental variables. A Multilayer perceptron was built in WEKA, which uses an algorithm to automatically build the network (manual testing was possible though). After experimenting with different numbers of hidden layers and hidden nodes, it was found that a single hidden layer performed better. Indeed, adding extra hidden layers did not increase the performances of the model enough to justify that choice. The optimal network topology was selected according to the specified learning rate and momentum. The best structure of the MLP network were selected having one hidden layer with 9 nodes (figure 5.21), which are all sigmoid functions used in backpropagation. The optimal learning rate and momentum were selected at 0.25 and 0.15. A validation set was employed to validate the weights and it was created by taking a portion of the training set (66% in training, 33% in validation). In addition, a validation threshold of 20 was set, corresponding to the number of consecutive times the validation set error could deteriorate before the training was stopped (to avoid local minima during gradient descent).

Accuracies obtained by using MLP on permafrost data were 84.29% for randomly sampled observations and 79.1% for spatially sampled data (table 5.11). AUROCs of 0.9 and 0.84 respectively are indicative of a good classification ability of the algorithm (figure 5.22). These values are associated to good true negative rates (especially for randomly sampled observations). Indeed, MLP was able to correctly classify permafrost absence for almost 90% of the observations of the dataset, indicating the negative occurrence of permafrost. In turn, the ability of this technique to correctly classify the permafrost presence was just moderate, with a TP rate of 72% for randomly selected data and of 63% for spatially sampled permafrost training

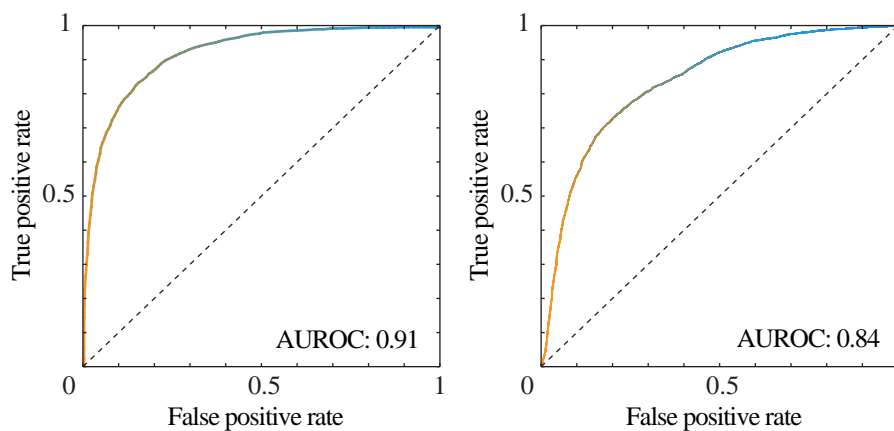
examples. This means that MLP is generally predicting the negative permafrost occurrence in sectors where the presence of permafrost is instead observed.

Measure	Random sampling		Spatial sampling	
	Train statistics	Test statistics	Train statistics	Test statistics
Accuracy	85.175	84.293	85.623	79.103
RMSE	0.326	0.338	0.322	0.396
TP rate	0.852	0.722	0.741	0.633
FP rate	0.176	0.092	0.081	0.118
TN rate	0.824	0.908	0.919	0.882
FN rate	0.148	0.278	0.259	0.367
Precision	0.852	0.812	0.832	0.741
Recall	0.852	0.722	0.741	0.633
AUROC	0.918	0.907	0.920	0.844

**Tab. 5.11.:** Quality measures of the MLP classification on randomly and spatially sampled data.



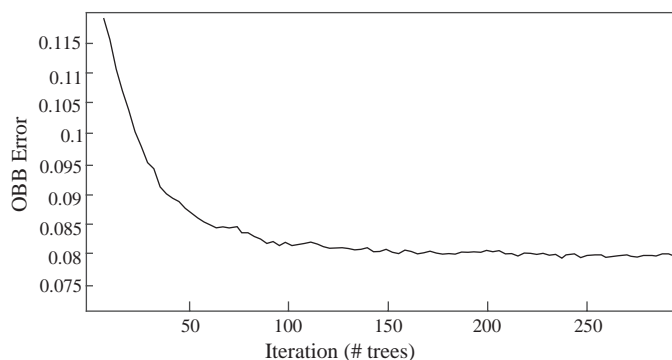
**Fig. 5.21.:** MLP network topology with one hidden layer and 9 non-linear nodes.



**Fig. 5.22.:** MLP ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.91) and spatially sampled data (right image, AUROC: 0.84).

## Random forest

The Random forest algorithm was also employed on permafrost data. It belongs to the so-called group of ensemble learning classifiers and it differs from previous algorithms because it grows decision trees as base learners and it combines their predictions by averaging. Three main parameters have to be tuned in RF: the *ntree* parameter, which is the number of trees required to be grown; the *mtry* parameter, which refers to the number of predictors the algorithm should consider at each node split; and the *nodesize*, which is the number of observations accepted in the terminal nodes. Generally, the larger the number of trees, the more computationally expensive the classification will be. The selection of *ntree* is generally deduced from the out-of-bag error curve, which decreases and becomes stable as soon as the correct number of trees is reached. By trial and error, it was identified that, for employed permafrost data, the OOB error curve generally converges starting from 250 grown trees. It was thus decided to set *ntree* to 300 (figure 5.23). A rule of thumb usually used for selecting the number of tried attributes (*mtry*) is the square root of the dataset dimensions. The use of smaller values of *mtry* is often associated to overfit. As a consequence, in this study the number of variables selected at a node split followed the approach proposed in the original paper on Random Forest (Breiman, 2001), which suggests using the first integer less than  $\log_2(M) + 1$ , where  $M$  is the number of variables of the dataset. A *mtry* value of 4 was thus used ( $\text{int}(\log_2(15) + 1) = 4$ ). As for the last parameter, the *nodesize* corresponds to the so-called *maxDepth* in the RF implementation in WEKA and it controls the maximum depth of the tree. In the current study, no limitation were used and therefore trees were developed to the maximum to be able to recognize all useful signals from the data.



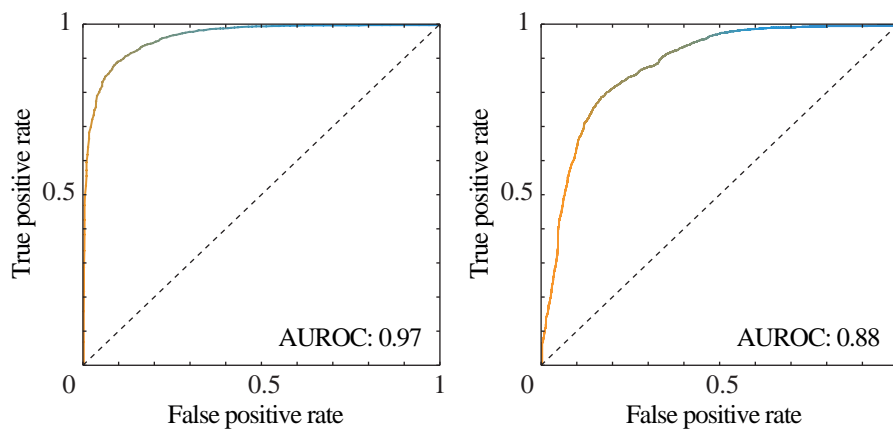
**Fig. 5.23.:** OOB curve used to assess the number of trees required for the dataset (applied to a randomly sampled subset).

Classification statistics indicate that the forest builds on the training set almost perfectly (95% of accuracy, table 5.12). In RF, the accuracy is often wildly optimistic "by design" and it is considered as meaningless. Indeed, as reported by Liaw (the developer of the *randomForest* R package, through relative support forums), when using the *maxDepth* (or *nodesize*) setting mentioned above, the trees are grown to

Measure	Random sampling		Spatial sampling	
	Train statistics	Test statistics	Train statistics	Test statistics
Accuracy	95.09	90.24	95.453	80.5
RMSE	0.205	0.274	0.198	0.368
TP rate	0.951	0.809	0.955	0.635
FP rate	0.071	0.047	0.061	0.097
TN rate	0.929	0.952	0.939	0.902
FN rate	0.049	0.190	0.045	0.364
Precision	0.951	0.902	0.954	0.775
Recall	0.951	0.809	0.955	0.635
AUROC	0.991	0.966	0.992	0.875
OOB error	0.0794		0.0987	

**Tab. 5.12.:** Quality measures of the RF classification on randomly and spatially sampled data.

the maximum size, meaning that it is very likely to reach the terminal nodes with only one data point and the prediction at the terminal nodes are determined by the majority class in the node. If this happens for all the forest trees and in all terminal nodes, only one data point is kept. Then, that specific data point would be "in-bag" for about 64% of the trees of the forest. As a result, the prediction for that data point will be correct for each one of those trees. Moreover, even though all the trees where that data points are out-of-bag gave the wrong prediction, by majority vote of all trees, the prediction would still be correct, providing a perfect or almost perfect prediction on the training set. For this reason, RF is designed to remember the *out-of-bag* (OOB) prediction error, which is a more reliable indicator of the model performance.



**Fig. 5.24.:** RF ROC curves of classification obtained from randomly sampled data (left image, AUROC: 0.97) and spatially sampled data (right image, AUROC: 0.88).

The classification on randomly sampled subsets was computed with an OOB error of 7.94%, whilst the inferred model performed on spatially sampled data had an error of 9.87%. The area under the ROC curves differ by 0.091, with a value of 0.966 for randomly sampled training examples and 0.875 for spatially sampled observations (figure 5.24). This translates in greater precision values (0.902/0.80 and 0.775/0.63). True positives and true negatives were also better classified with randomly sampled permafrost evidences. In fact, a lowest number of true negatives were identified for data splitted with respect to the polygon they belonged to (and thus more false

negatives were predicted). For a more complete overview of classification statistics, see appendix A.

## 5.4.2 The effect of sampling on classification performances

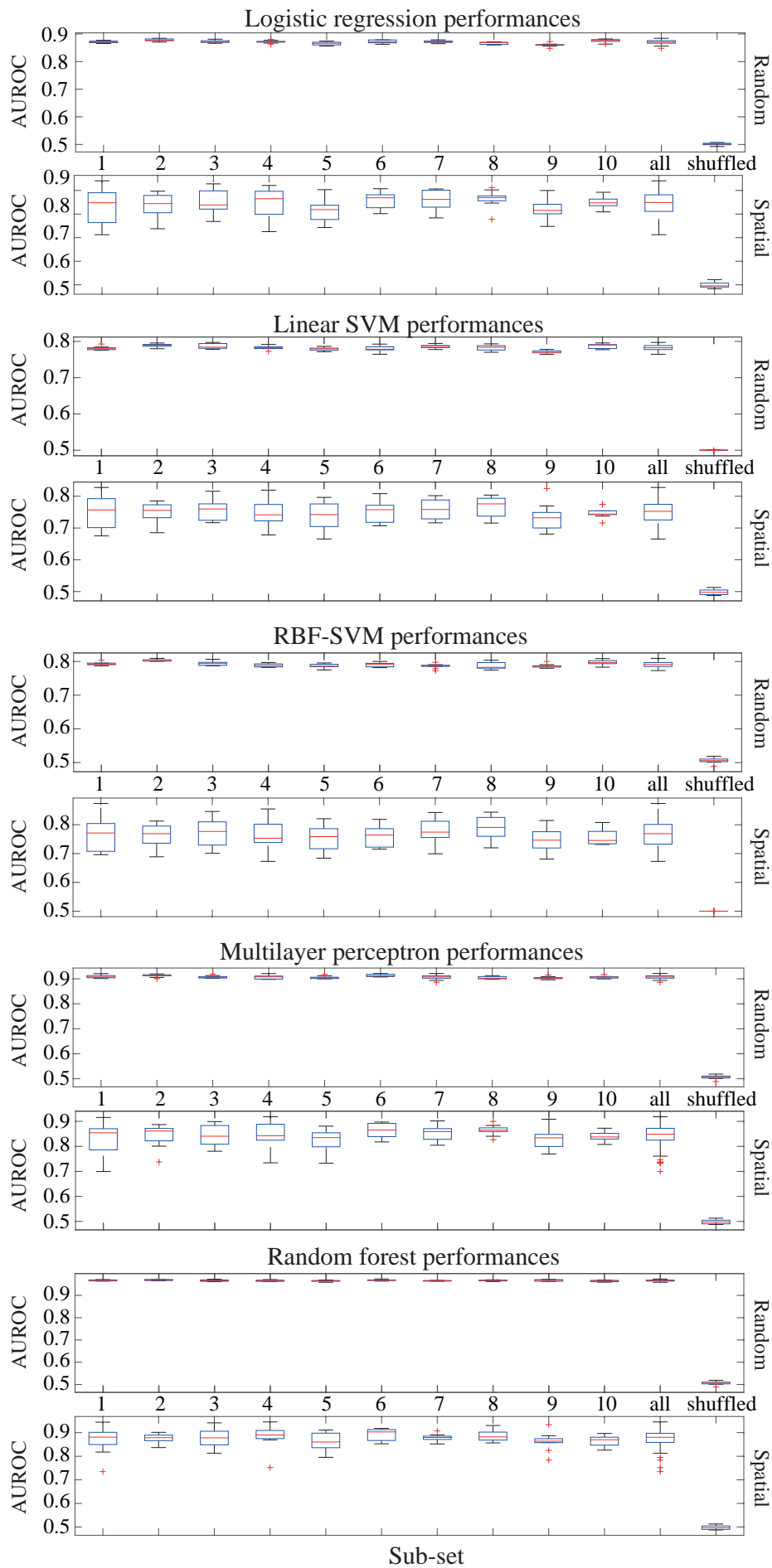
The sub-sampling strategies that were applied to the cleaned dataset aimed at perturbing the data in order to catch the maximum of the contained information. As mentioned, completely random sampling of the training instances from the original dataset could have led to an over-representation of examples with the same spatial characteristics and thus poor classification performances or over-fitting. Therefore, a comparison of the model performances obtained by applying *Logistic regression*, *linear* and *RBF-SVM*, *Multilayer perceptron* and *Random forest* to the different pools of data was conducted. Figure 5.25 shows a comparison of the performances of the employed ML algorithms for the two chosen sampling strategies. Classification performances are presented with relative box-plots of the AUROC values computed at each repetition. The graph illustrates the box-plots for the extracted ten sub-sets of data (1 to 10 in the x-axis), the overall performance (represented as *all* in the x-axis) and the classification performances of shuffled data (to destroy the structures contained in the data).

Sampling strategy	LR		L-SVM		RBF-SVM		MLP		RF	
	random	spatial	random	spatial	random	spatial	random	spatial	random	spatial
AUC	0.87	0.84	0.78	0.77	0.79	0.75	0.90	0.84	0.97 (OOB: 0.86)	0.87 (OOB: 0.89)

**Tab. 5.13.:** Comparison of the AUROC values obtained with the employed ML algorithms for randomly and spatially sampled data.

This analysis shows the quality of the partition of the original clean dataset into the 10 pools of data. Indeed, the model performances obtained for the sub-sets 1 to 10 are generally comparable on average. This indicates that, despite the different sampling strategy selected after the partition in 10 sub-sets, the permafrost evidences were equally represented.

In general, the AUROC of models inferred from randomly sampled data are slightly higher than the values obtained by classifying spatially sampled sub-sets. A clear difference is observed in the variability of the classification statistics among the 100 repetitions computed for each classifier. Indeed, the box-plots of randomly sampled observations have a lower degree of dispersion, indicating that the repetitions made for the same sub-set generally provide very similar results. Performances of models inferred from spatially correlated train-test-validation sub-sets present less variability than the ones made on non-spatially correlated sets. Nevertheless, the overall median AUROC is comparable for all selected ML algorithms with a loss in the AUROC ranging between 0.01 and 0.06 (table 5.13). In addition, the variability in the model performances also serves as a measure of the uncertainties. By computing



**Fig. 5.25.:** Comparison of the AUROC values obtained with the employed ML algorithms for randomly and spatially sampled data sub-sets.

the standard deviation of the AUROC obtained at each repetition, a more complete analysis of the reliability of the ML techniques can be conducted. The investigation of the model uncertainties is presented in the next section (5.4.3).

### 5.4.3 Model comparison and modelling uncertainties

It is a common practice in machine learning to estimate the classification error by using cross-validation and to choose the algorithm that provides the lowest estimate. However, it is also important to investigate which model is statistically the best when performing a specific classification task. As mentioned above, ROC curves were evaluated on independent test sets for comparing the generalization performances of LR, linear SVM, RBF-SVM and RF. The resulting AUROC values were compared by using the Model Evaluator module of WEKA (see Witten and Frank, 2005) that employs a *paired t-test* (confidence: 0.05, two tailed). The test was performed on the results obtained by classifying randomly sampled data (table 5.14) and spatially sampled data (table 5.15).

<i>Classifier</i>	<i>LR</i>	<i>L-SVM</i>	<i>RBF-SVM</i>	<i>MLP</i>	<i>RF</i>
<i>LR</i>	0.87	0.78 *	0.79 *	0.91 v	0.97 v
<i>L-SVM</i>	0.87 v	0.78	0.79 v	0.91 v	0.97 v
<i>RBF-SVM</i>	0.87 v	0.78 *	0.79	0.91 v	0.97 v
<i>MLP</i>	0.87 *	0.78 *	0.79 *	0.91	0.97 v
<i>RF</i>	0.87 *	0.78 *	0.79 *	0.91 *	0.97

**Tab. 5.14.:** Performances and paired t-test outcomes (v: significantly better, \*: significantly worse) for models obtained with random sampling.

<i>Classifier</i>	<i>LR</i>	<i>L-SVM</i>	<i>RBF-SVM</i>	<i>MLP</i>	<i>RF</i>
<i>LR</i>	0.84	0.75 *	0.77 *	0.84 v	0.88 v
<i>L-SVM</i>	0.84 v	0.75	0.77 v	0.84 v	0.88 v
<i>RBF-SVM</i>	0.84 v	0.75 *	0.77	0.84 v	0.88 v
<i>MLP</i>	0.84 *	0.75 *	0.77 *	0.84	0.88 v
<i>RF</i>	0.84 *	0.75 *	0.77 *	0.84 *	0.88

**Tab. 5.15.:** Performances and paired t-test outcomes (v: significantly better, \*: significantly worse) for models obtained with spatial sampling.

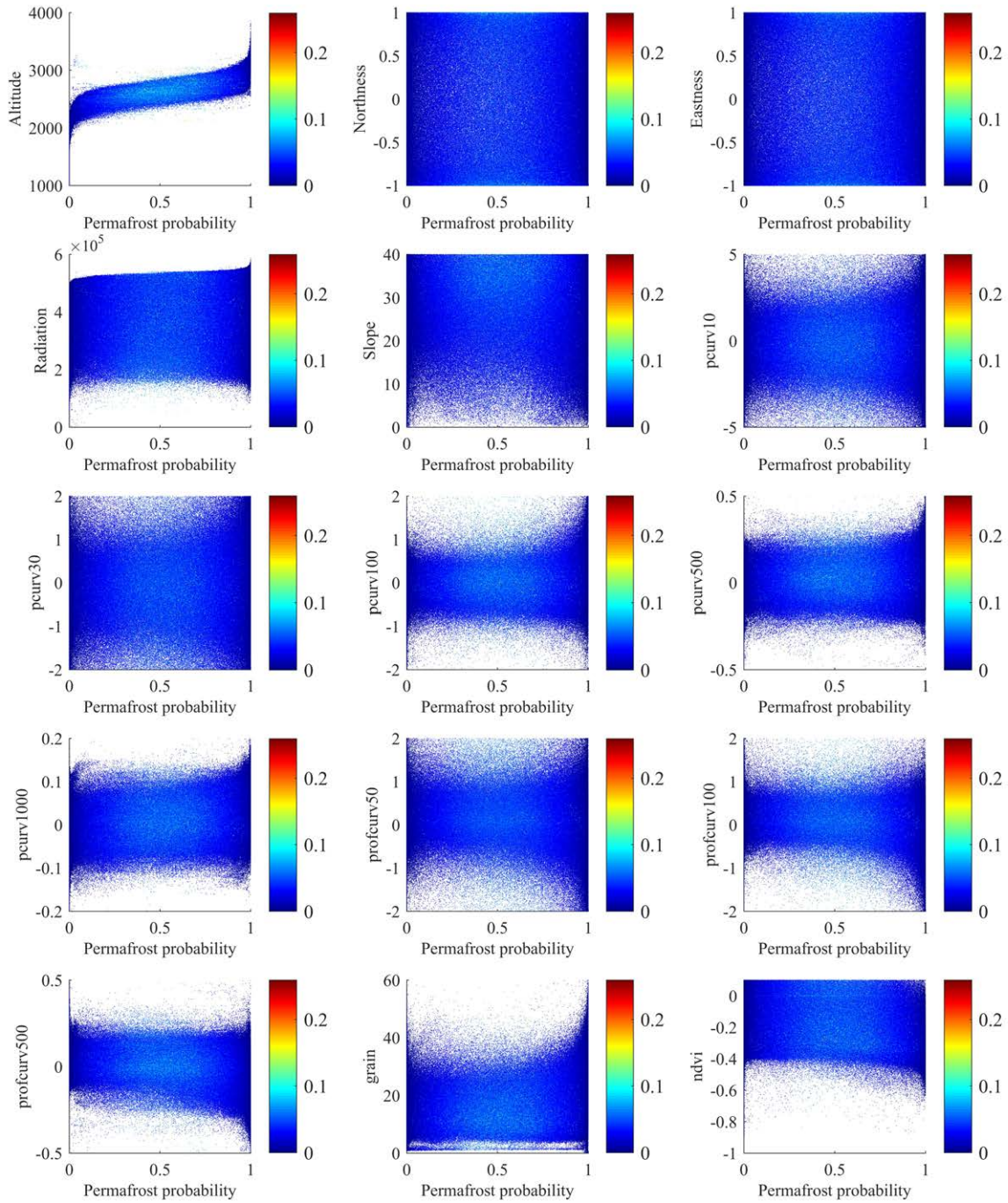
The *t-tests* indicate the same trends for both sampling strategies. The LR models are significantly better than the linear and the RBF SVM models, whereas they are significantly worse than RF. Although a statistically significant difference exists between MLP and LR, the small difference in AUC ( $<0.01$ ) suggests that LR might perform similarly to MLP. The tests also show that the linear SVM results are significantly worse than all other algorithm results, whilst RBF-SVM only performs better than its linear version. Finally, the t-test outcomes indicate that RF is the algorithm whose models are significantly better performing, with MLP also offering good results second only to RF.

A major concern when modelling and predicting the potential distribution of a natural phenomenon like the occurrence of permafrost is the quantification of the uncertainties. By taking advantage of the stochasticity of machine learning algorithms, it was possible to assess the certainty of a prediction for a given pixel

by observing the variability of the resulting classification probabilities. Indeed, at each repetition (as a recall, 100 in total given that a 10-fold cross-validation was used on 10 different sub-sets), the classifier inferred a different probability of membership to the *presence* and *absence* classes. By measuring the standard deviation of these probabilities, not only was it possible to assess how reliable the classification was for each pixel of the validity domain, but also at what range of values of the predictors the model is less uncertain. Uncertainty plots were thus obtained by showing the model uncertainties for all the employed algorithms and sampling strategies. Each pixel of the plots refers to a *pixel of the validity domain* to which a *probability of permafrost occurrence* was attributed (x-axis), as well as a value of the standard deviation of the computed probabilities (right y-axis colorbar). The colour thus informs whether, for that given pixel, the 100 repetitions provided similar or very diverse probabilities. The plots inform on the relevance of the prediction in accordance to the values of each environmental predictor (left y-axis values). The *density of points* shows thus for what *range of values the algorithms predicted a specific range of probabilities*, whereas the *colour of the points* informs on the *degree of certainty*. A total of 15 plots are presented, corresponding to the uncertainties associated to the values of the 15 predictors that were retained in the *cleaner dataset* after performing the *reduction of the dimensionality* of the raw permafrost dataset. Figures 5.26 and 5.27 illustrate the uncertainties obtained by the LR. The models inferred from randomly sampled permafrost examples have low uncertainty, meaning that most of the repetitions produced very similar results. Low probabilities were associated to low *altitudes*, while high probabilities were predicted for pixels located at high *altitude* values. The prediction of permafrost is homogeneously distributed at all values of the *northness* and the *eastness* predictors that express the aspect. The same homogeneous distribution is observed for the PISR (*radiation*) values, with no particular probability patterns that are unravelled by the uncertainty plot. According to the algorithm result, this means that the probabilities of permafrost occurrence are equally distributed for all the PISR values and thus this variable did not help LR for the classification of the presence and absence classes. The *slope* predictor shares the same tendency with *eastness* and *northness*, with the exception of low slope values less represented by probabilities ranging between 0.05 and 0.75. The *curvature* features almost present the same trend with most of the lowest and highest probabilities predicted for pixels with low and high values of curvature. This indicates that LR predicted the absence and the presence of permafrost in convex and concave sectors, whilst medium probabilities were attributed to flat surfaces. Lastly, the *grain* and the *NDVI* predictors present an homogeneous pattern of probabilities for values ranging from 0 to 35 and from 0.1 to -0.4 respectively. However, a higher density of points can be observed for high values of *grain size* and high probabilities of occurrence. This shows that sectors characterized by coarse blocks (expressed as high standard deviation of the grey-scaled orthophoto colour) were predicted as favourable for the presence of permafrost.

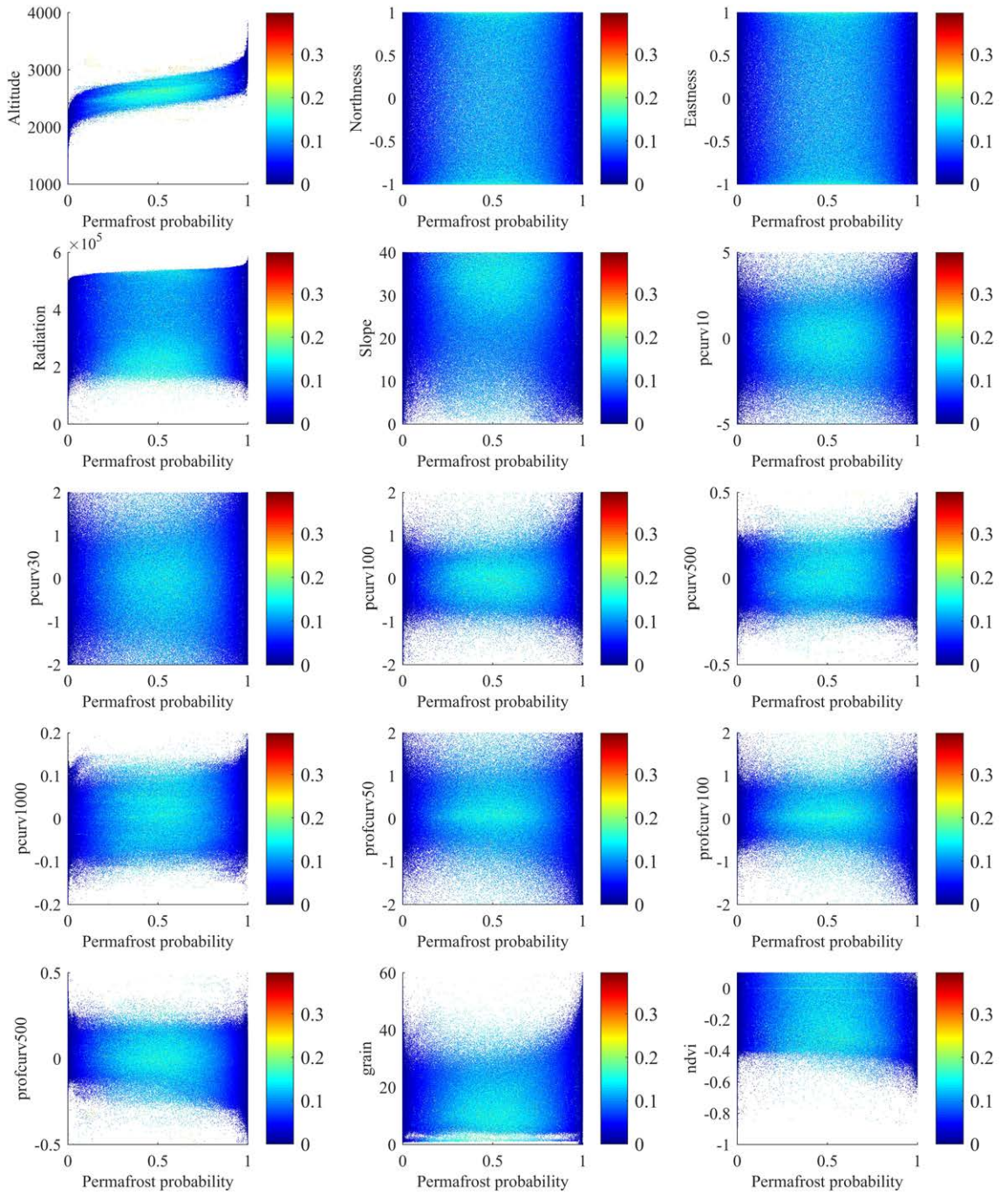


LR uncertainties (random sampling)



**Fig. 5.26.:** Uncertainty plots for Logistic regression classification on randomly sampled sub-sets.

LR uncertainties (spatial sampling)



**Fig. 5.27.:** Uncertainty plots for Logistic regression classification on spatially sampled subsets.

The same density patterns were obtained by the LR models inferred from spatially sampling data, meaning that both sampled strategies provide almost the same classification result (which is confirmed by similar mean of probabilities among all repetitions, as shown in section B). In turn, the standard deviation of all repetition results indicate that, in general, the highest variability was obtained for sectors where the probability of occurrence ranges between 0.2 and 0.8 (figure 5.27). LR permafrost models are less certain for sectors located between 2450 and 2850 metres of *altitude*, where a mix of permafrost evidences corresponding to the presence and absence classes were also identified (as presented in figure 5.1, section 5.1 of this chapter). Similarly, most uncertainties were computed for flat surfaces where the permafrost prediction ranges between 25% and 75% (see curvature uncertainty plots in figure 5.27). For the same interval of probabilities, the LR models are characterized by a higher variability at lower *PISR* values, higher *slopes* and slow *grain* values.

The linear SVM uncertainty plots show the same patterns observed in the plots discussed above (figures 5.28 and 5.29). Indeed, the classification with this algorithm produced almost the same probability distribution as LR in relation to the values of each environmental predictor. Therefore, the comparison between the two adopted sampling strategies unveils that random sampled sub-sets are classified with less divergence from the average prediction than the classification performed on spatially sampled sub-sets (as it was also presented in figure 5.25).

The models computed with the non-linear expansion of SVM, which employs a Radial basis function (RBF) kernel, present less certain results for both sampling strategies (figures 5.30 and 5.31). Indeed, in general, the highest values of the standard deviation between the repetitions were obtained for probabilities ranging between 0.25 and 0.75. This is particularly true for spatially sampled sub-sets. By analysing the distribution of probabilities in relation to the predictor values, the first sampling strategy produced a RBF-SVM model that is less uncertain at *altitudes* between 2400 and 2800 metres, whereas the second strategy presents more uncertain predictions, especially between 2500 and 3000 metres of *altitude*. The same observation is also valid for all other predictors.

From the density of point it is possible to note that lowest probabilities were predicted for the lowest *altitude*, whereas the highest probabilities were attributed to pixels located at the highest elevations. No clear pattern is observable from these plots for the *northness*, *eastness*, *radiation*, *slope*, *pcurv10*, *pcurv30* or *profcurv50*. However, the *pcurv100*, *pcurv500* and *pcurv1000* uncertainty plots show that low probabilities were predicted for low values of these *curvature* indices. This indicates that the permafrost absence is more linked to the negative planform curvatures, namely concavities perpendicular to the direction of the maximum slope at a large scale

Linear SVM uncertainties (random sampling)

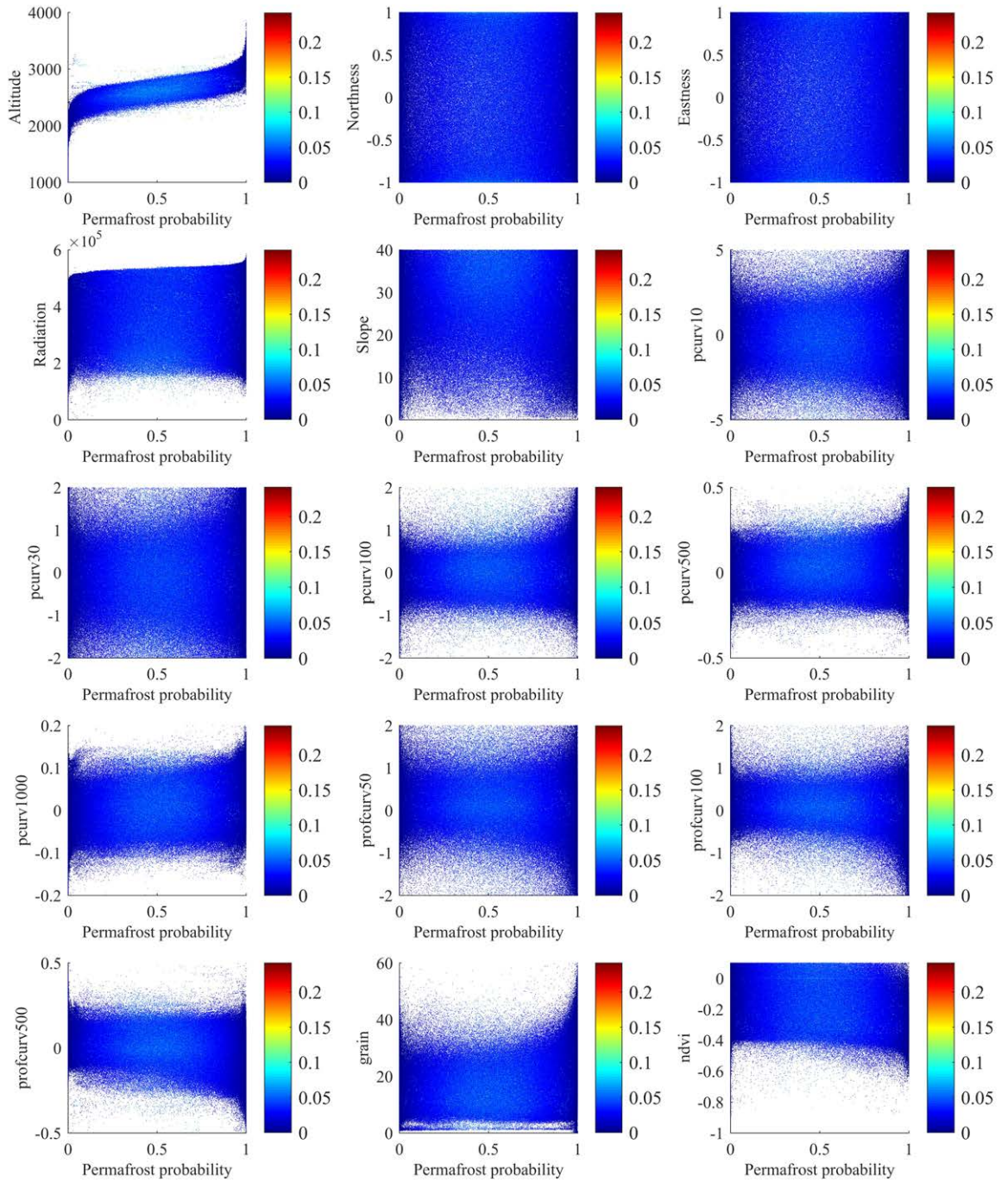


Fig. 5.28.: Uncertainty plots for linear SVM classification on randomly sampled sub-sets.

Linear SVM uncertainties (spatial sampling)

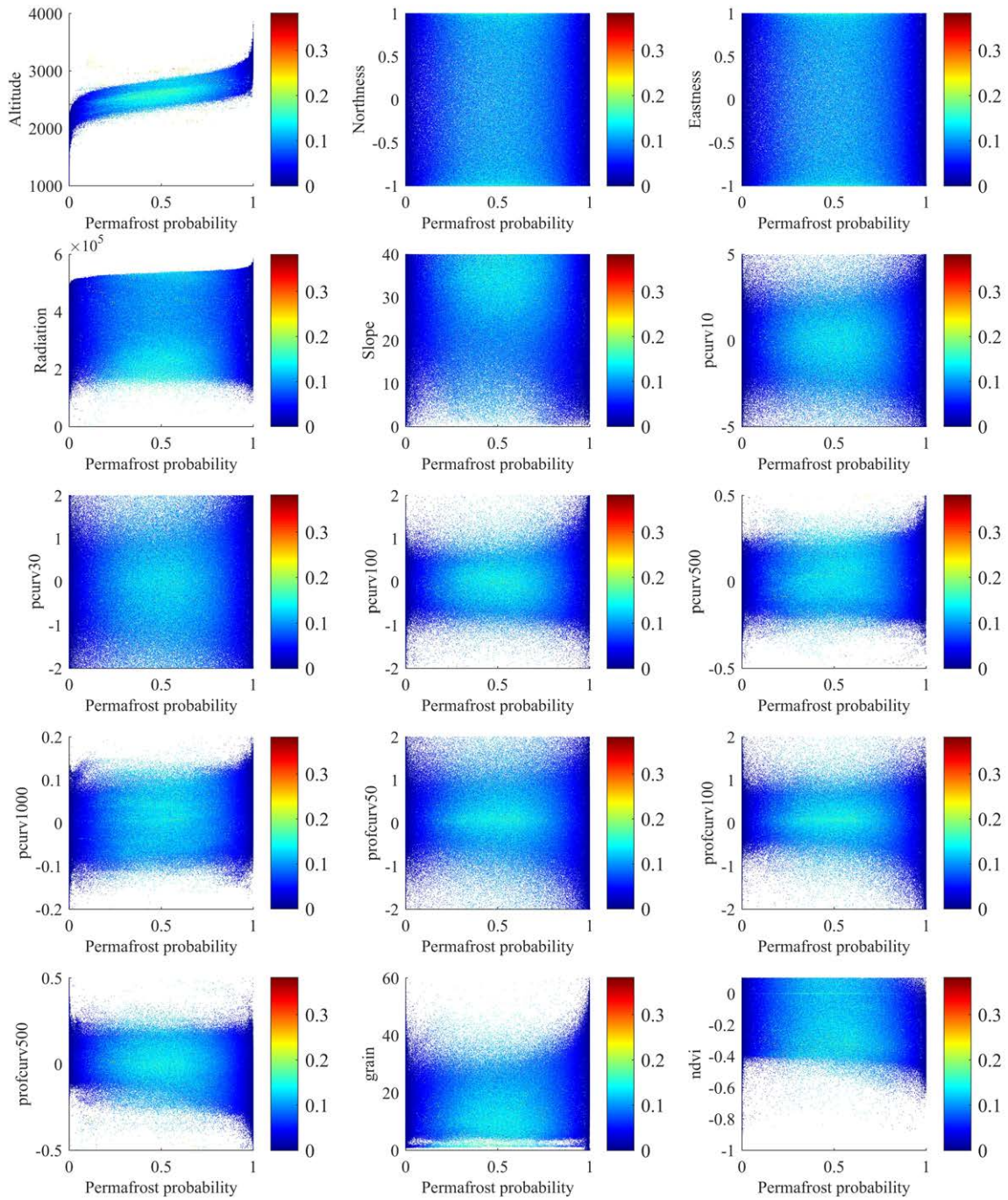


Fig. 5.29.: Uncertainty plots for linear SVM classification on spatially sampled sub-sets.

(100, 500 and 1000 metres). Conversely, convexities perpendicular to the direction of the maximum slope at a large scale were identified by RBF-SVM as controlling the permafrost presence. Similarly, concavities parallel to the direction of the maximum slope at a scale of 100 and 500 metres (expressed as positive values of *profcurv100* and *profcurv500*) are indicative of permafrost presence, whilst convexities in this direction result in being associated to permafrost absence by the algorithm.

The Multilayer perceptron is the ML algorithm that produced the highest divergence in terms of permafrost probabilities (figures 5.32 and 5.33). Indeed, from 10 to 90% of probabilities of permafrost occurrence, the models present the highest standard deviations for all predictors. The results are particularly in contrast with the adopted sampling strategies. On the one hand, the prediction is less certain between 2200 and 2500 metres of *altitude* for the classification on randomly splitted sub-sets for pixels to which a probability between 0.25 and 0.75 was computed. On the other hand, the altitude uncertainty plot of the classification made on sets selected following the spatial strategy illustrates higher uncertainties at altitudes between 2200 m.a.s.l. and 2800 m.a.s.l. for the same range of probabilities. A similar behaviour is unveiled for other environmental predictors, with resulting variability of predictions more accentuated for the models obtained with spatial sampling.

The prediction of the presence and absence of permafrost (according to the prediction values) made by MLP follows the same trend mentioned above for RBF-SVM, with *altitude*, *pcurv100*, *pcurv500*, *pcurv1000*, *profcurv100* and *profcurv500* being the predictors presenting a density of points with a different pattern between the classes.

Uncertainty plots computed for the Random forest result present a lower prediction variability compared to RBF-SVM and MLP (figures 5.34 and 5.35). However, the points are distributed differently, with fewer pixels classified as *permafrost absence* or *permafrost presence* with values close to 0 or 1. This is particularly true for the point cloud of the *curvature* predictors, to whom probabilities of 0 and 1 were generally associated to values close to 0 (corresponding to flat areas). As for LR, linear SVM, RBF-SVM and MLP, the difference between random sampling and spatial sampling is associated to the increased number of uncertainties at medium probabilities. Therefore, particular patterns were observed for the models inferred from spatially selected data. On the one hand, highest discrepancies in the predictions are computed for pixels located between 3000 and 3150 m.a.s.l. to whom a probability of occurrence between 0.45 and 0.6 were associated (see altitude uncertainty plot in figure 5.35). On the other hand, the same range of probabilities also presents a less certain result for *radiation* values above 500'000 WH/m<sup>2</sup>, as well as *slopes* above 30°. The same trend is also unravelled for low *grain* size values and a *NDVI* index close to -0.35.

RBF-SVM uncertainties (random sampling)

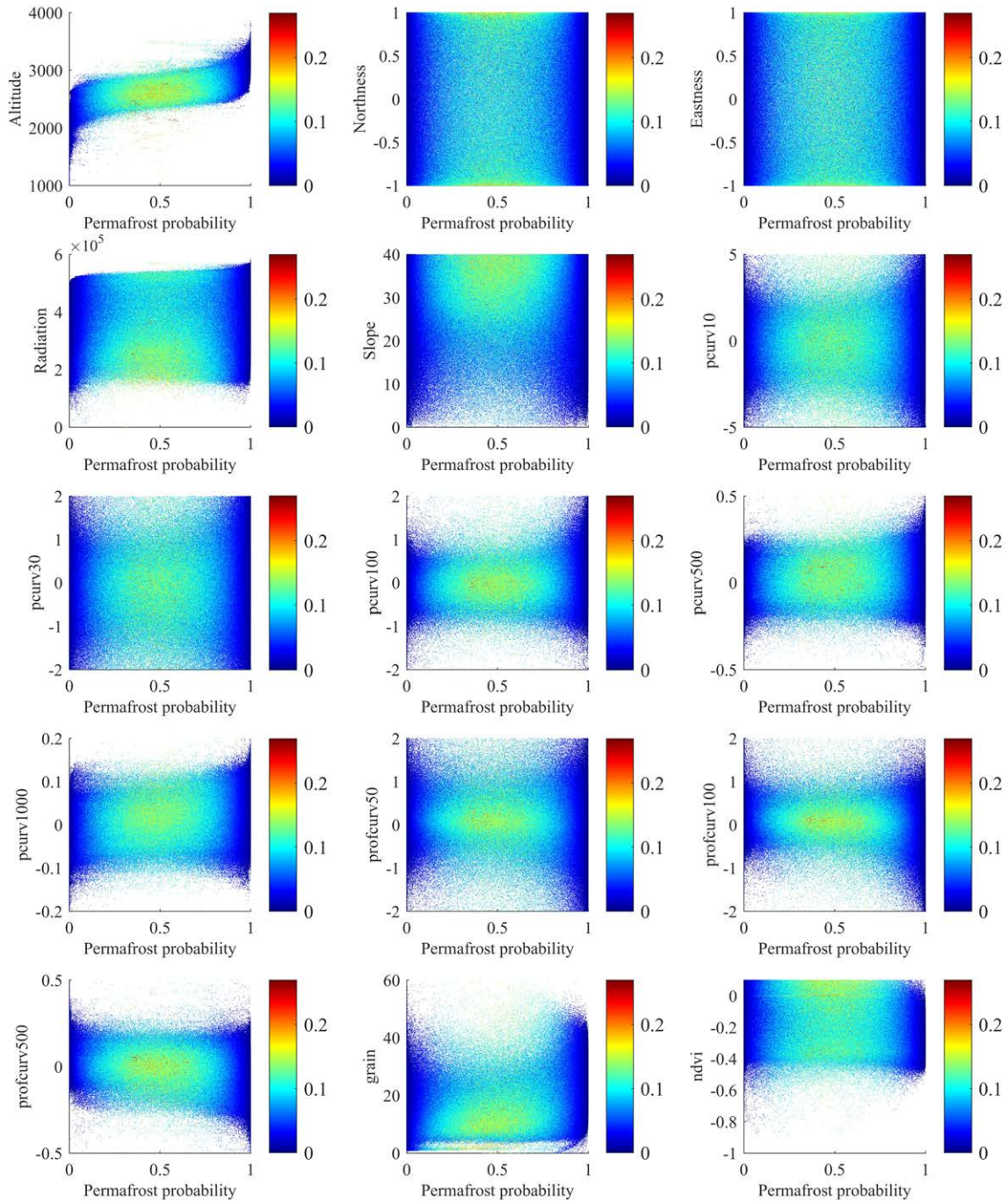


Fig. 5.30.: Uncertainty plots for Radial basis function SVM classification on randomly sampled sub-sets.

RBF-SVM uncertainties (spatial sampling)

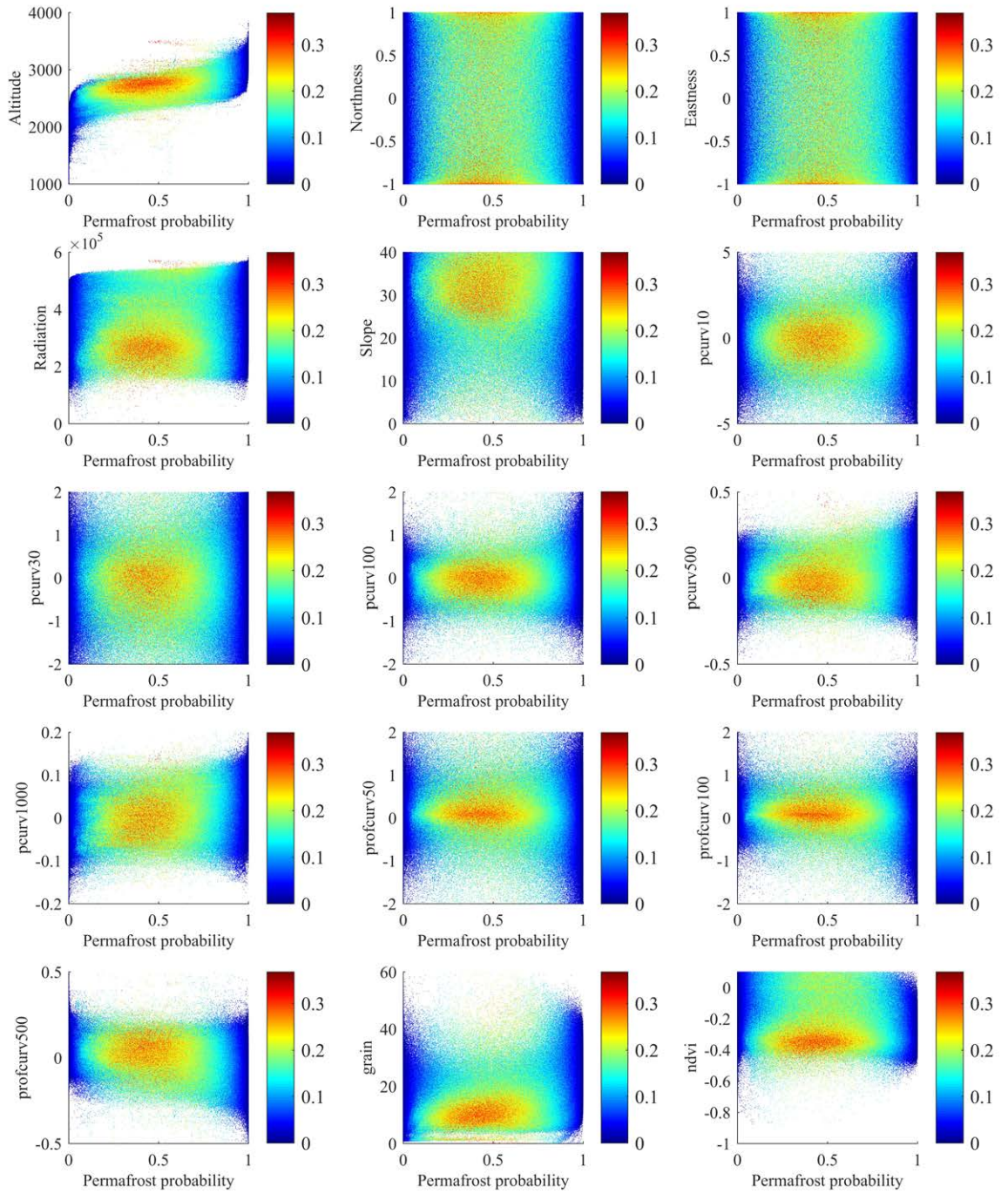


Fig. 5.31.: Uncertainty plots for Radial basis function SVM classification on spatially sampled sub-sets.



MLP uncertainties (random sampling)

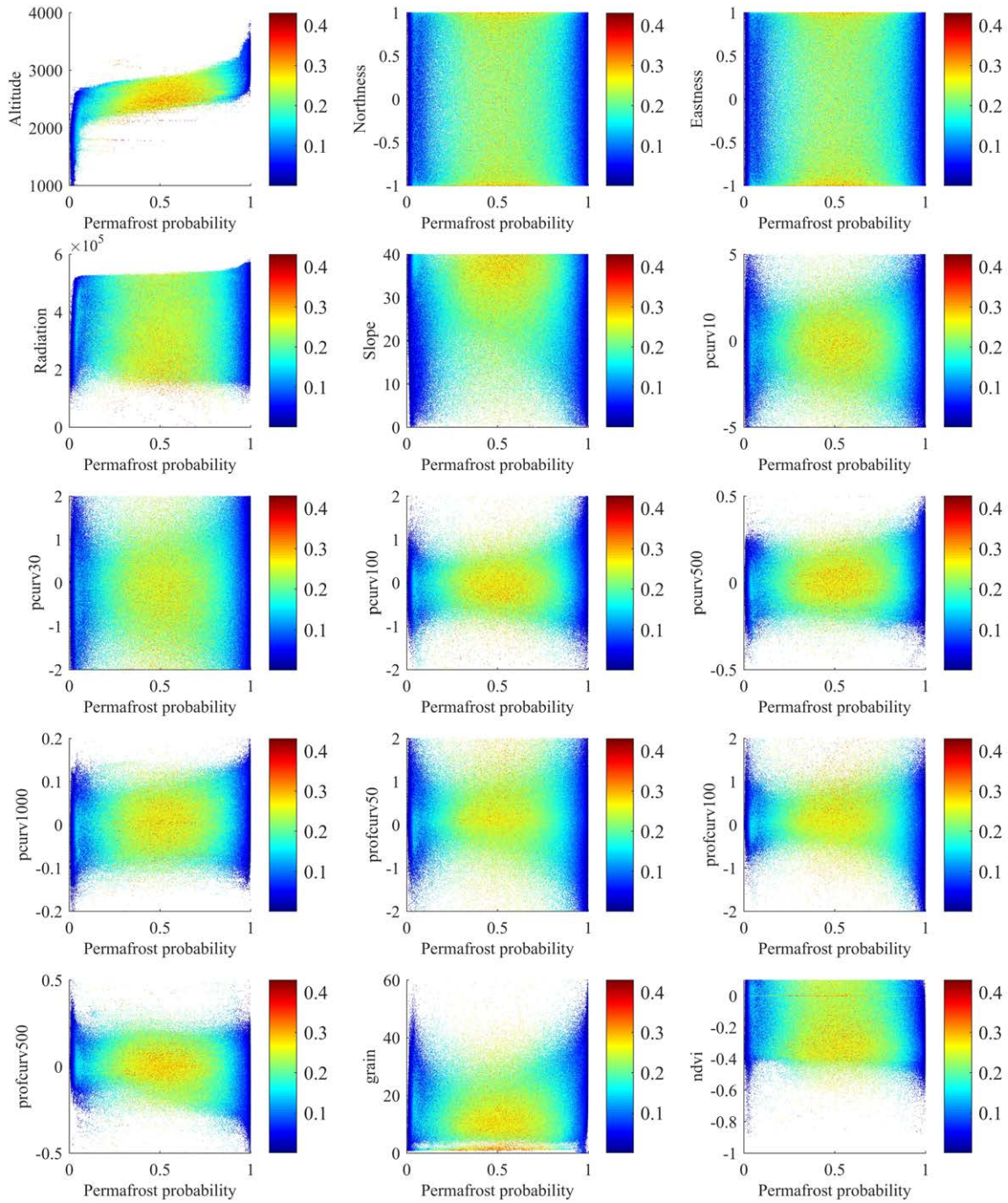
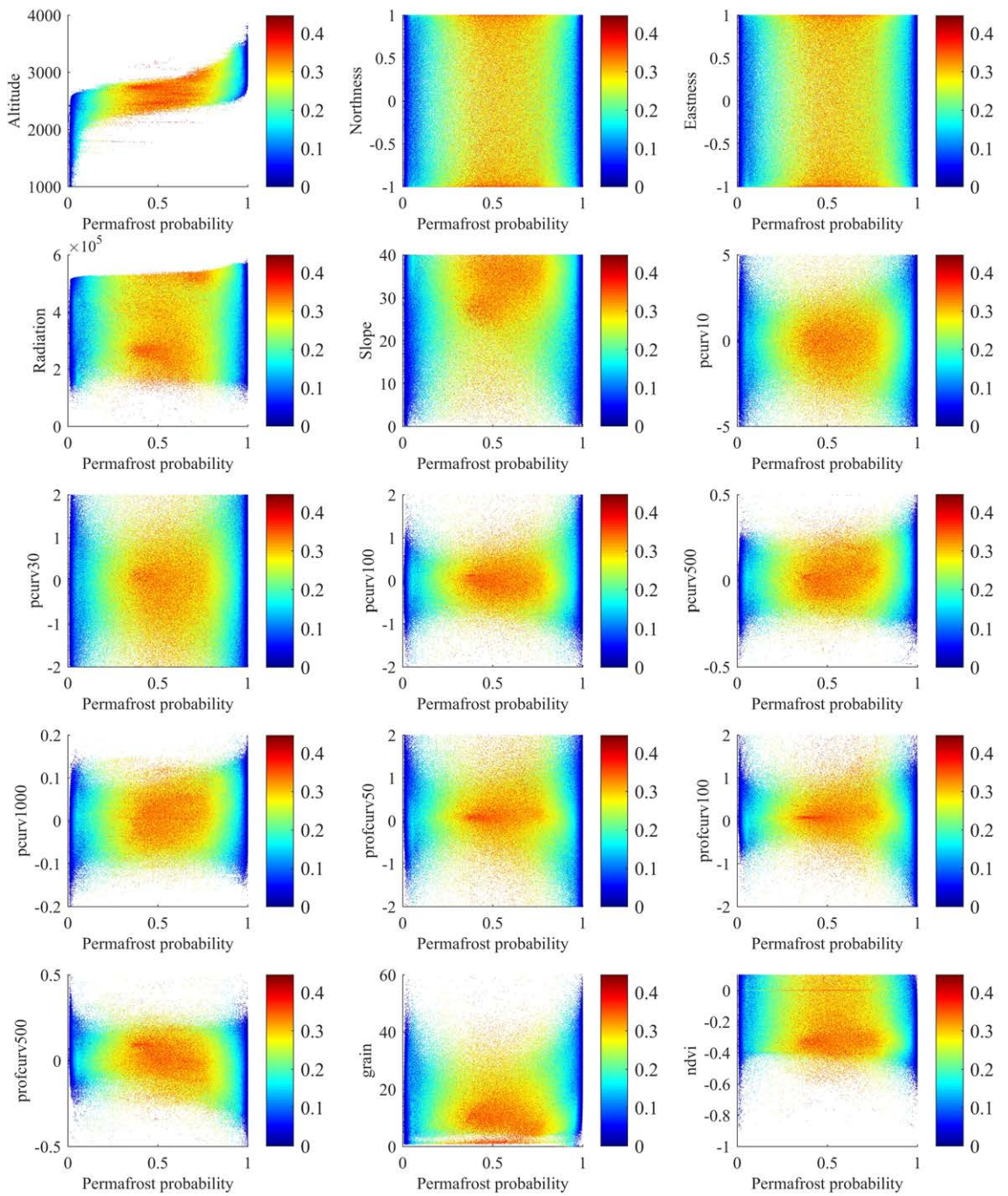


Fig. 5.32.: Uncertainty plots for Multilayer perceptron classification on randomly sampled sub-sets.

MLP uncertainties (spatial sampling)



**Fig. 5.33.:** Uncertainty plots for Multilayer perceptron classification on spatially sampled sub-sets.

As for RBF-SVM and MLP, *altitude*, *pcurv100*, *pcurv500*, *pcurv1000*, *profcurv100* and *profcurv500* are the predictors presenting a density of point with a different pattern between the highest and the lowest values. However, conversely to the other data-driven techniques, RF seems to not predict probabilities close to 0% for curvature indices close to their maximum range (i.e.  $-2$  and  $+2$  for *pcurv500*). This indicates that RF generally does not predict the permafrost absence in sectors characterized by convex or concave surfaces (across all directions).

In summary, the analysis of the model uncertainties illustrated the variability of the results according to the used classification techniques. In general, the classification of randomly sampled datasets produced less variable results in comparison to the classification of spatially sampled data. When comparing the uncertainties obtained by applying each classifiers, some similarities can be observed at the variable level, especially for the curvature predictors. Indeed, the most uncertain predictions were computed for permafrost probabilities between 25% and 75% and curvature values close to 0, indicating that the probabilities of permafrost occurrence predicted for flat surfaces are less certain. High uncertainties were particularly computed by the SVM-RBF and the MLP algorithms.

RF uncertainties (random sampling)

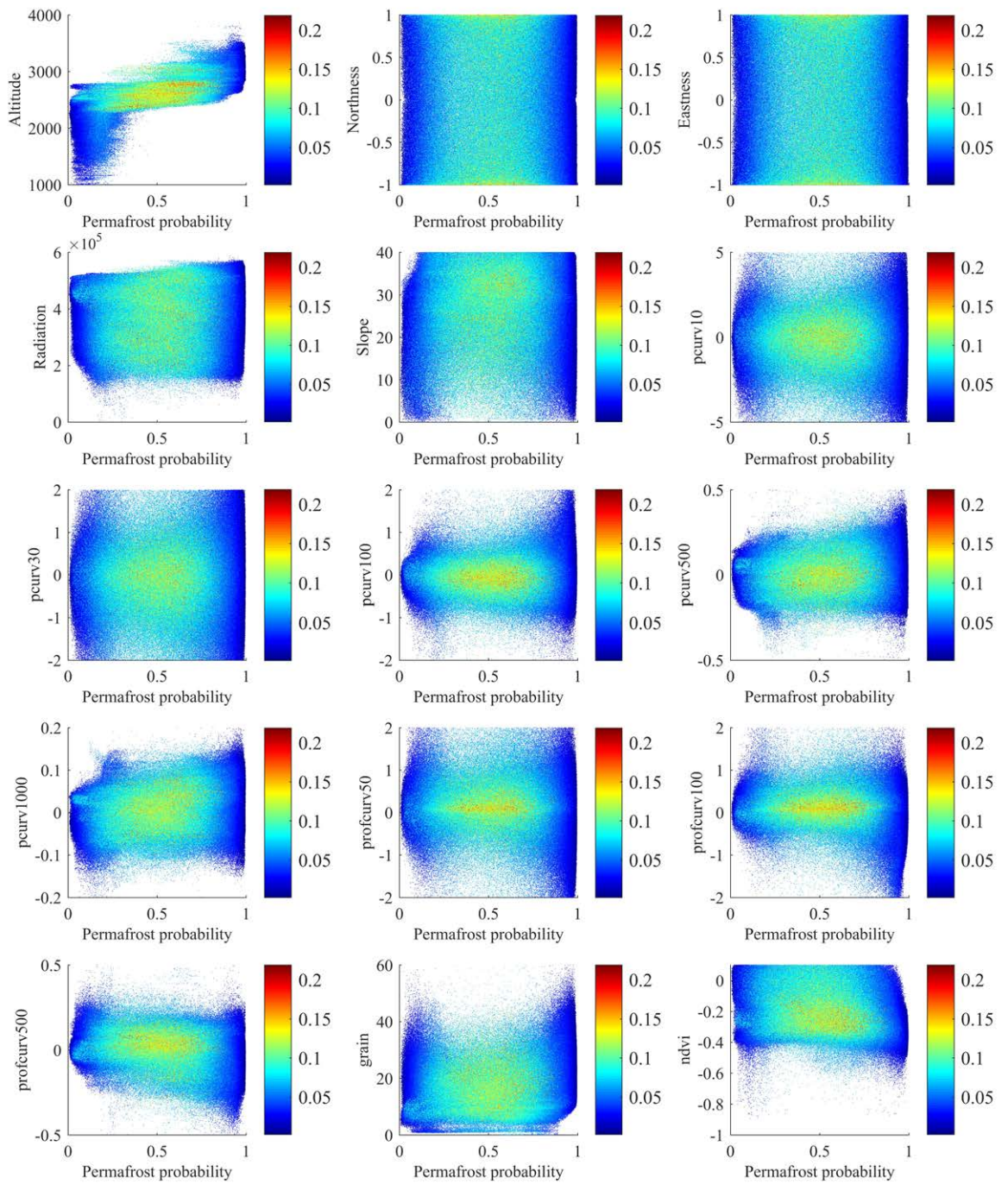


Fig. 5.34.: Uncertainty plots for Random forest classification on randomly sampled sub-sets.

RF uncertainties (spatial sampling)

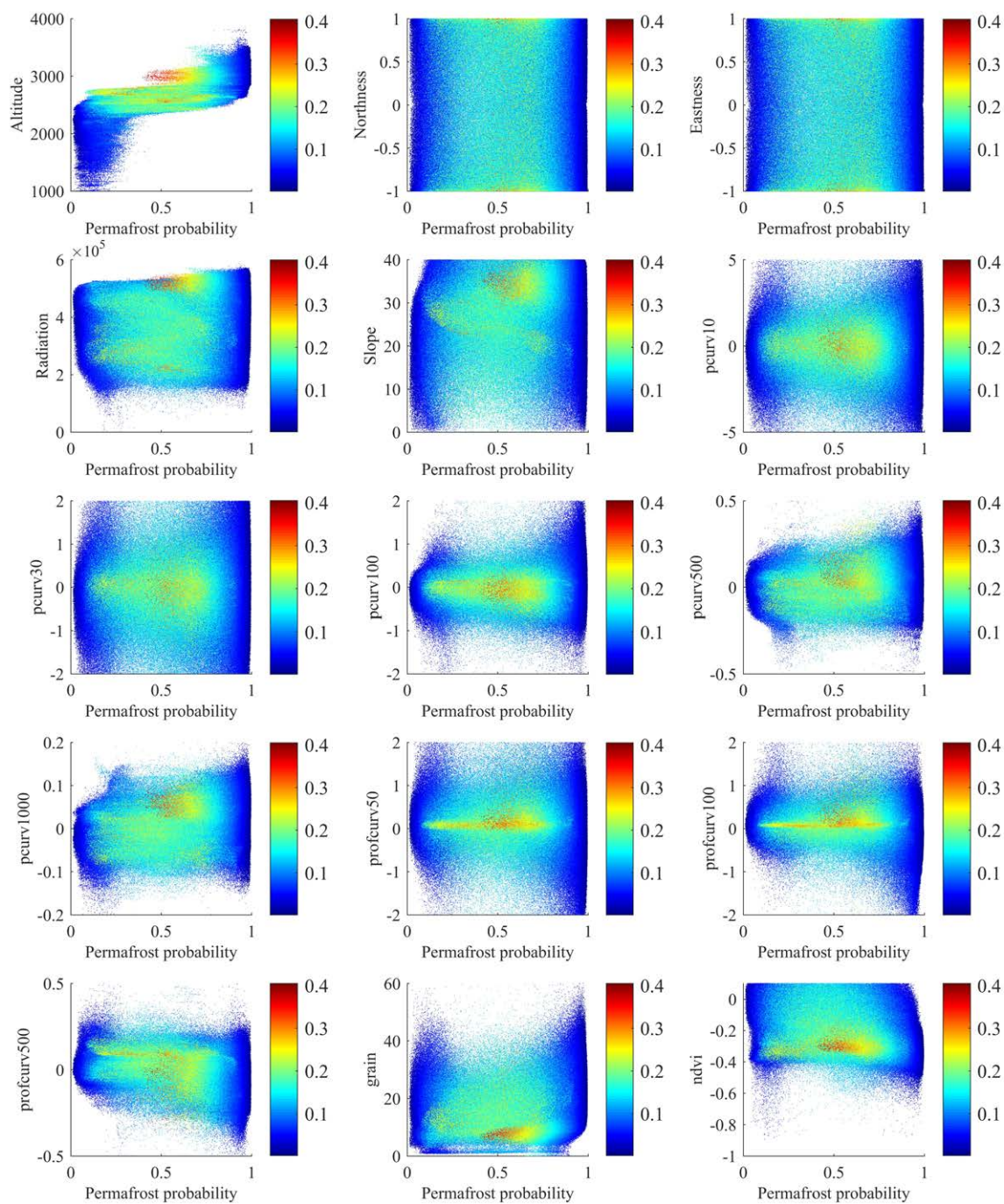


Fig. 5.35.: Uncertainty plots for Random forest classification on spatially sampled sub-sets.

## 5.5 Permafrost mapping

Permafrost mapping is the last step of the modelling approach adopted in this research. It results from the classification of the permafrost training observations with the machine learning algorithms mentioned in the previous sections. Indeed, all models inferred at each repetition were successively re-applied to a prediction grid corresponding to the portion of territory covered by the validity domain (see section 3.4). The extent of the permafrost distribution was thus predicted for the entire region of interest.

For each of the employed machine learning techniques, two distinct maps were obtained: one presenting the prediction of permafrost from randomly sampled permafrost data; one resulting from the classification of spatially sampled observations. Additional maps presenting the classification uncertainties were also associated to the permafrost maps. They describe the variability of the probability of permafrost occurrence for each pixel of the prediction grid, illustrating where the model is more uncertain and thus additional investigations or additional permafrost examples are potentially required. This analysis completes the study of the model uncertainties presented in section 5.4.3, where the reliability of the predictions were only investigated at a predictors- and probability- level, with no consideration of their spatial distribution.

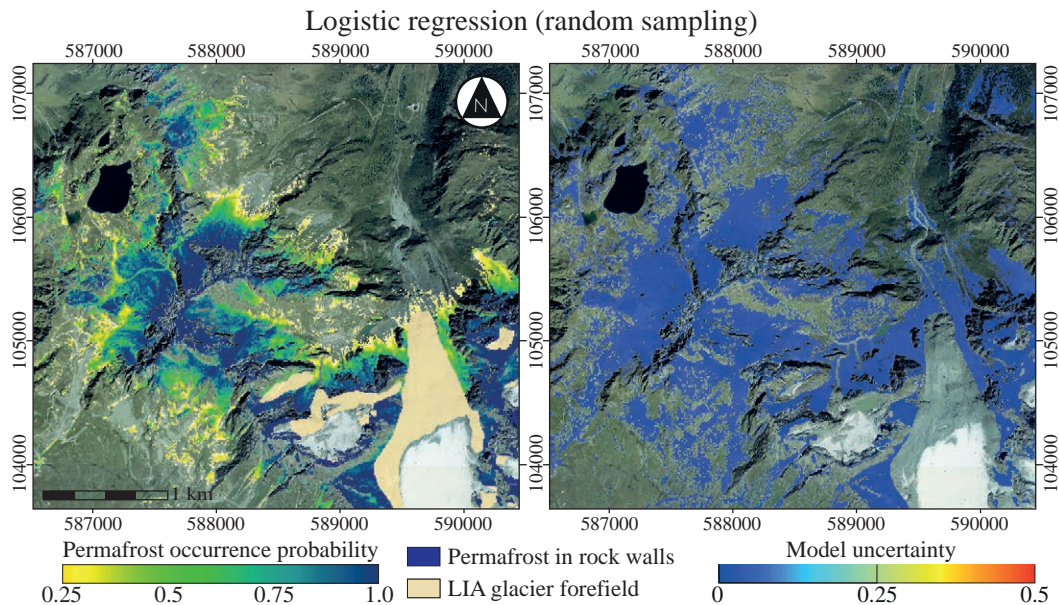
As a reminder, permafrost evidences were only considered for mineral surfaces being part of the validity domain. Rock walls were excluded from the validity domain because of a lack of sufficient permafrost evidences collected in these landforms. Therefore, the prediction of the permafrost occurrence was only performed for mineral surfaces with slopes lower than 40°.

The permafrost occurrence in rock walls was added to the final permafrost maps thanks to the contribution of Dr. Florence Magnin (University of Oslo), who kindly provided a map of the mean annual rock surface temperatures (MARST) for the region of interest of the present study. This map was produced according to the approach presented in Magnin et al. (2015), which uses the *rock model* developed by Boeckli et al. (2012). This is one of the two modules of the APIM model and computes the mean annual temperature at the rock wall surface with a multiple linear regression by employing the PISR and the MAAT (for the 1980-2010 period). The model was thus obtained with the PISR computed for the region of interest and the same temperature gradient of  $-0.54^{\circ}\text{C} / 100\text{m}$ , obtained from air temperature data collected during the period 1980-2010 by the Evolène-Villa (1427 m.a.s.l.) and the Montana (1825 m.a.s.l.) weather stations (SwissMetNet, meteosuisse) (see section 3.3.1). A reclassification of the model values was done in order to integrate the result to the final permafrost maps. In fact, since the *rock model* offers a map

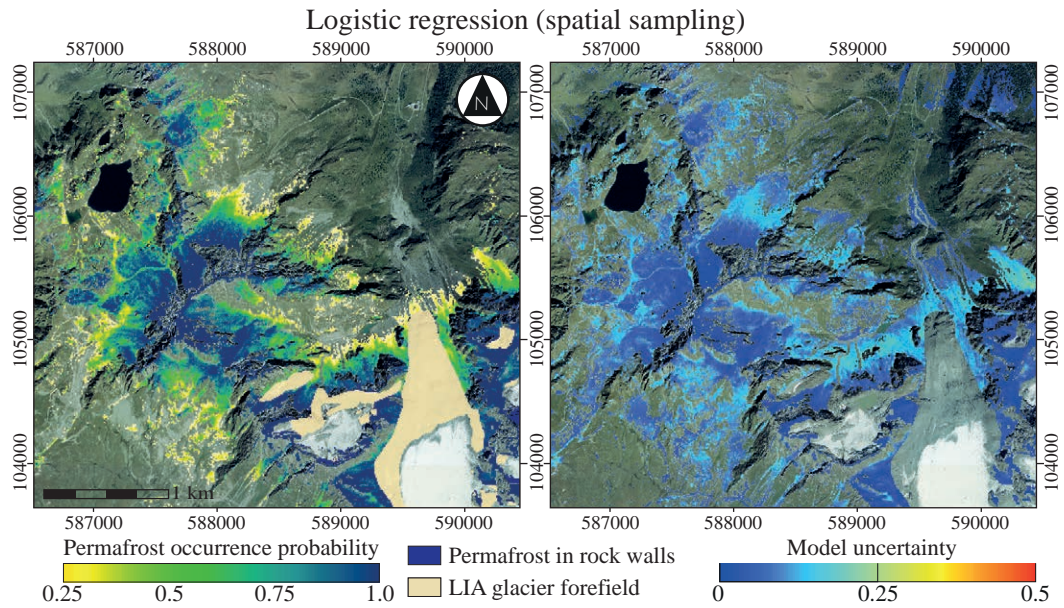
of the rock temperature, the permafrost presence *in rock walls* was extracted by considering all the sectors with rock temperatures lower than 0°C.

In addition to the map of the probabilities of permafrost occurrence and the permafrost map in rock walls, the delimitation of the glacier forefields was added to the map. As shown in section 3.3.2, permafrost is often absent in sectors occupied by glaciers during the Little Ice Age. However, permafrost may be present very locally, especially at their lateral margins.

In the following maps, the potential permafrost distribution for a sector of the region of interest is presented (see Appendix A for complete maps). They give an overview of the permafrost extent at the local scale as well as the main differences in the variability of the ML predictions. Although the complete range of probabilities of occurrence were computed (between 0% and 100%), the probabilities below 25% were purposely masked to enhance the visualization of potentially frozen surfaces. It is however worth mentioning that this threshold was subjectively selected. In addition to probabilities, in the next sections the following notations are going to be used: *permafrost very likely absent* (for probabilities between 0% and 25%), *permafrost likely absent* (for probabilities between 25% and 50%), *permafrost likely present* (for probabilities between 50% and 75%) and *permafrost very likely present* (for probabilities between 75% and 100%). They do not aim at referring to specific permafrost classes of occurrence (i.e. possible or probable permafrost, as it is the case for the BAFU map, see section 2.4.2), but they only indicate the probability range that was assigned to each pixels of the region of interest.



**Fig. 5.36.:** Potential permafrost distribution map obtained with Logistic regression (random sampling) and corresponding uncertainties.



**Fig. 5.37.:** Potential permafrost distribution map obtained with Logistic regression (spatial sampling) and corresponding uncertainties.

The potential permafrost maps obtained by using the Logistic regression are very similar at the local scale, with a general permafrost occurrence that linearly increases with *altitude* (figures 5.36 and 5.37). Although the two maps differ in terms of model uncertainties, with a map obtained with a random sampling less uncertain than the one predicted from spatially sampled data, the overall permafrost distribution is very similar in both models.

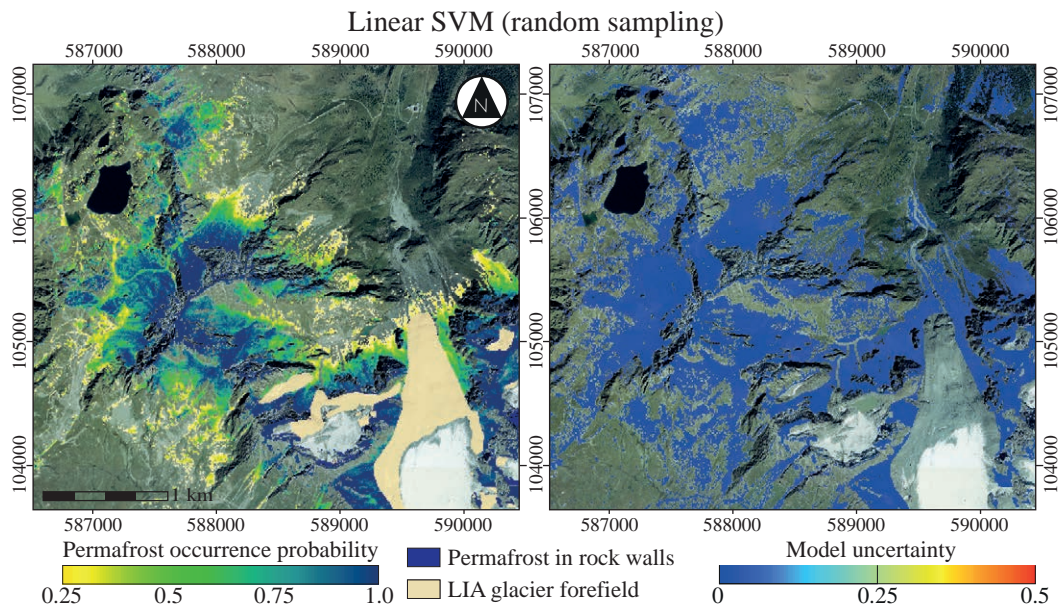
	<i>Probability of occurrence</i>			
	<i>0 - 0.25</i>	<i>0.25 - 0.5</i>	<i>0.5 - 0.75</i>	<i>0.75 - 1</i>
<i>Random sampling</i>	17.59%	12.13%	15.96%	54.32%
<i>Spatial sampling</i>	17.09%	12.09%	16.27%	54.55%

**Tab. 5.16.:** Parts of the validity domain predicted as *permafrost very likely absent* (0-0.25), *permafrost likely absent* (0.25-0.5), *permafrost likely present* (0.5-0.75) and *permafrost very likely present* (0.75-1) by Logistic regression (% of total grid cells).

Table 5.16 confirms this observation, with the percentage of pixels that are predicted almost equally for all four categories of probability mentioned above. At this scale, no substantial differences can be observed in terms of permafrost absence or presence between the two maps. A finer analysis will be presented further in this manuscript, with a comparison between the two models at the micro scale (see section 6.1.3).

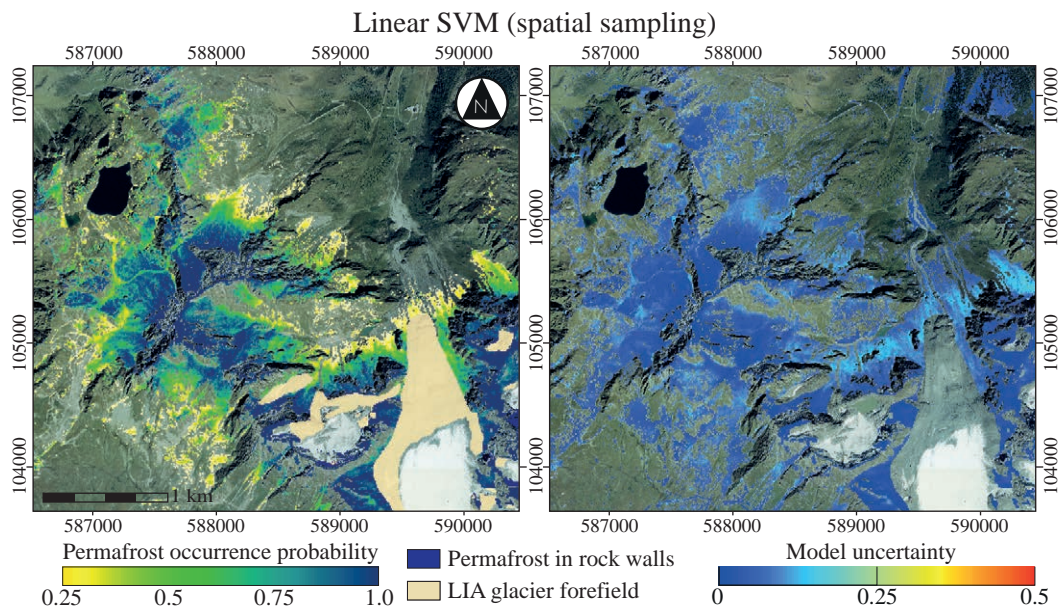
The extent of permafrost predicted by the linear SVM algorithm also presents a general increase of the probabilities with the increase of *altitude* at the local scale. The differences in model uncertainties are spatially less evident when comparing the two models. The most uncertain sectors are located at the lowest *altitude*, especially





**Fig. 5.38.:** Potential permafrost distribution map obtained with Linear SVM (random sampling) and corresponding uncertainties.

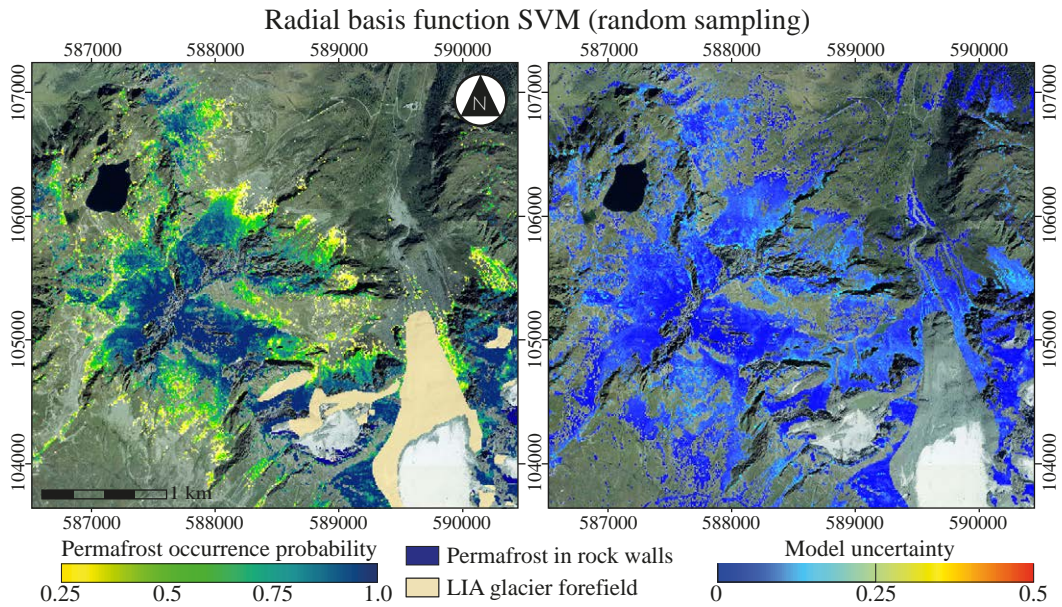
for the linear SVM map resulting from spatially sampled data (figures 5.38 and 5.39). No significant trend is however unveiled by the count of pixels belonging to the four selected categories, which does not differ much in terms of number of pixels classified in the four categories of occurrence (table 5.17). The map obtained with the Radial basis function SVM offers a different repartition of the permafrost distribution, especially at the scale of a valley side (figures 5.40 and 5.41), even though the portion of the map predicted as *permafrost very likely present*



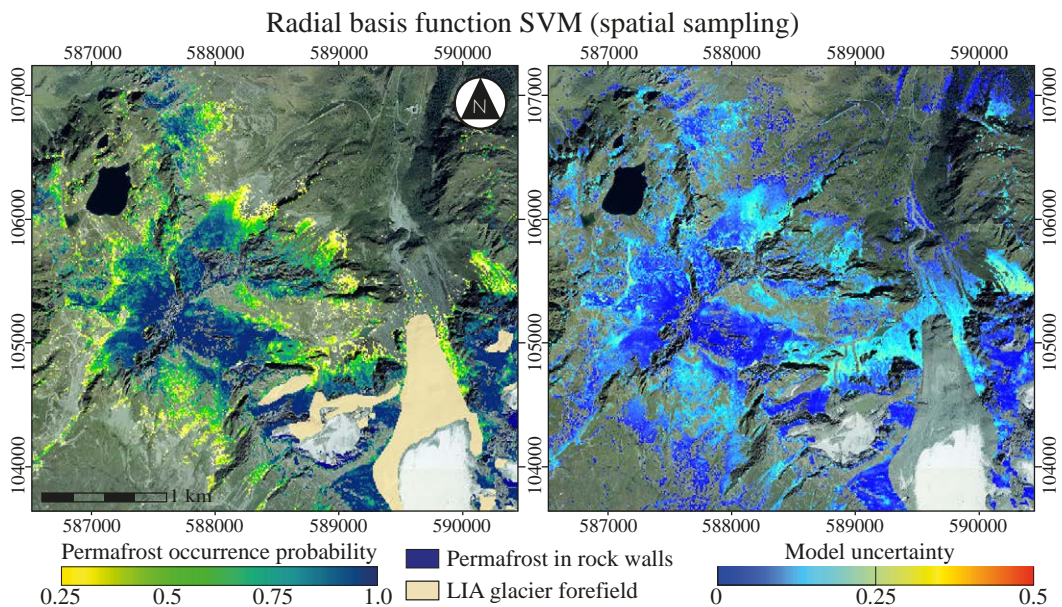
**Fig. 5.39.:** Potential permafrost distribution map obtained with Linear SVM (spatial sampling) and corresponding uncertainties.

	Probability of occurrence			
	0 - 0.25	0.25 - 0.5	0.5 - 0.75	0.75 - 1
Random sampling	17.06%	12.68%	16.86%	53.39%
Spatial sampling	16.52%	12.64%	17.25%	53.59%

**Tab. 5.17.:** Parts of the validity domain predicted as *permafrost very likely absent* (0-0.25), *permafrost likely absent* (0.25-0.5), *permafrost likely present* (0.5-0.75) and *permafrost very likely present* (0.75-1) by linear SVM (% of total grid cells).



**Fig. 5.40.:** Potential permafrost distribution map obtained with RBF-SVM (random sampling) and corresponding uncertainties.



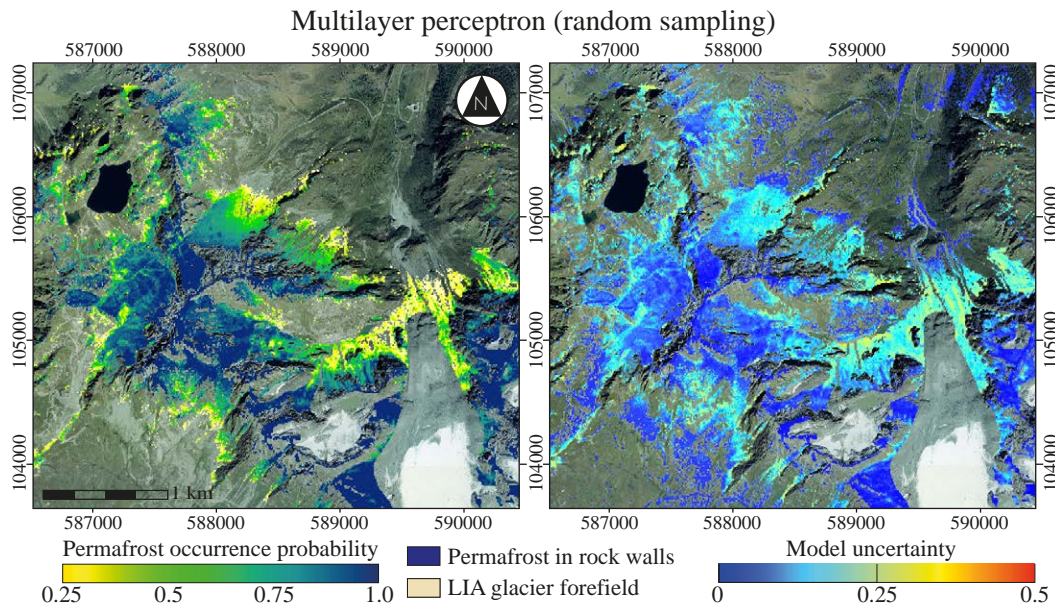
**Fig. 5.41.:** Potential permafrost distribution map obtained with RBF-SVM (spatial sampling) and corresponding uncertainties.

and *permafrost likely present* is almost equally extended as the results produced by LR and linear SVM (around 55% and 15% of the validity domain, table 5.18).

	Probability of occurrence			
	0 - 0.25	0.25 - 0.5	0.5 - 0.75	0.75 - 1
Random sampling	19.49%	12.10%	15.84%	52.57%
Spatial sampling	18.51%	12.39%	16.31%	52.79%

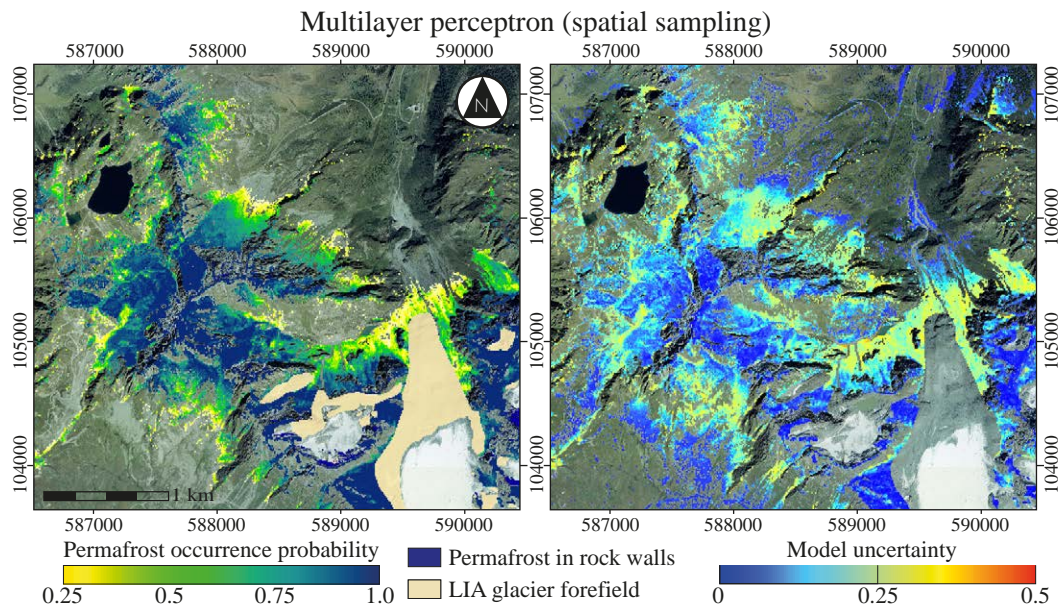
**Tab. 5.18.:** Parts of the validity domain predicted as *permafrost very likely absent* (0-0.25), *permafrost likely absent* (0.25-0.5), *permafrost likely present* (0.5-0.75) and *permafrost very likely present* (0.75-1) by Radial basis function SVM (% of total grid cells).

Indeed, the probabilities of permafrost occurrence produced by RBF-SVM do not present the same linear trend that prevails with the other mentioned ML algorithms (see appendix A for a general overview of the map). Conversely, the permafrost distribution is more spatially heterogeneous with distinct patches that are unravelled. The model inferred from spatially sampled data presents more spatially distributed uncertainties in comparison to the one inferred from randomly sampled observations, with a more important variability of the predictions occurring both at low altitudes (as it was observed for the linear SVM) and at higher located sectors.



**Fig. 5.42.:** Potential permafrost distribution map obtained with Multilayer perceptron (random sampling) and corresponding uncertainties.

Multilayer perceptron also produced a permafrost map where probabilities of occurrences are not linearly increasing with altitude (figures 5.42 and 5.43). While local variability of occurrence is more observable in comparison to the result obtained with LR and linear SVM, at the local scale the MLP maps share more similarities with the RBF-SVM maps. Compared to the latter, MLP produced a higher number of predictions falling in the category of *permafrost very likely present* (56-58% of the total prediction grid, table 5.19) and fewer predictions representing its absence



**Fig. 5.43.:** Potential permafrost distribution map obtained with Multilayer perceptron (spatial sampling) and corresponding uncertainties.

(probabilities between 0 and 0.5). This ML algorithm is thus more optimistic when predicting a positive permafrost occurrence.

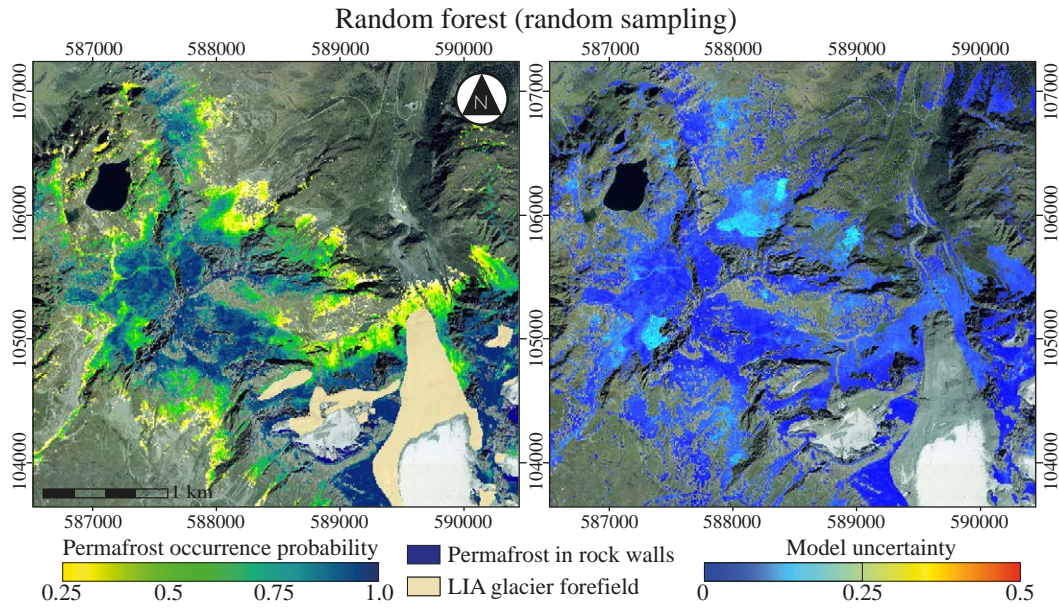
Model uncertainties (figures 5.42 and 5.43, right side images) are more remarkable in comparison to other algorithms. When comparing the MLP results inferred from randomly and spatially sampled sub-sets, the difference is less pronounced with a slightly less uncertain model obtained with randomly selected permafrost observations (see appendix A for complete maps).

	<i>Probability of occurrence</i>			
	<i>0 - 0.25</i>	<i>0.25 - 0.5</i>	<i>0.5 - 0.75</i>	<i>0.75 - 1</i>
<b><i>Random sampling</i></b>	17.71%	10.22%	15.53%	56.54%
<b><i>Spatial sampling</i></b>	17.51%	9.06%	15.07%	58.36%

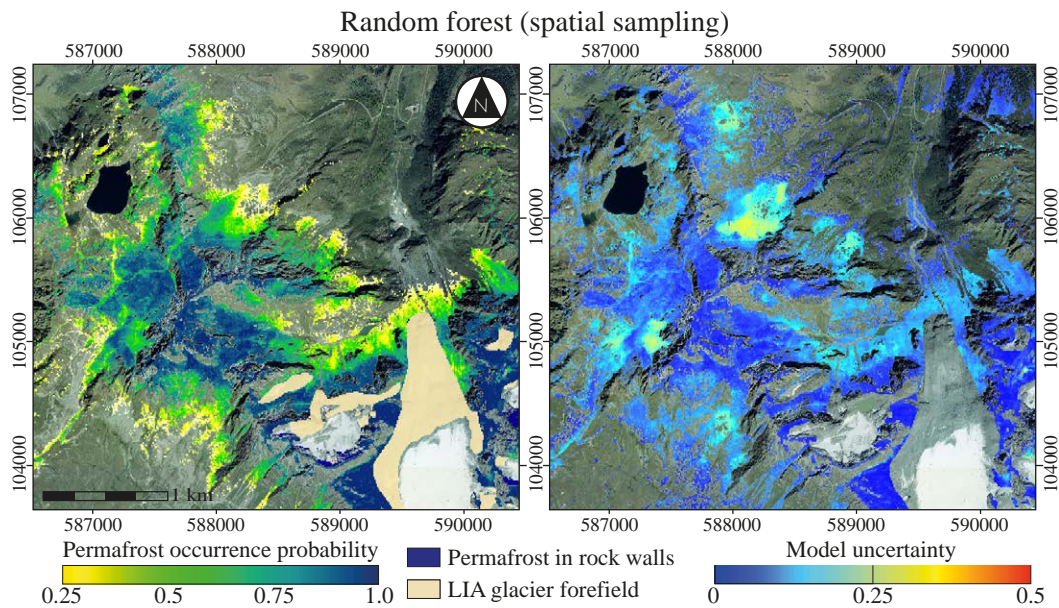
**Tab. 5.19.:** Parts of the validity domain predicted as *permafrost very likely absent* (0-0.25), *permafrost likely absent* (0.25-0.5), *permafrost likely present* (0.5-0.75) and *permafrost very likely present* (0.75-1) by Multilayer perceptron (% of total grid cells).

Lastly, the potential permafrost distribution maps predicted with Random forest (figures 5.44 and 5.45) also reflect a strong spatial discontinuity. Even though the highest probabilities were in fact mapped at the highest altitudes, permafrost distribution between 2400 and 2800 metres of altitude does not linearly increase and it presents heterogeneous spatial patterns. The effect of spatial and random sampling on the final permafrost maps is weak with very similar results (at least at the local scale). The differences are more remarkable when comparing the corresponding uncertainty maps with a less uncertain prediction obtained with a random sampling

approach in comparison to spatially sampled data. Indeed, the latter reveals greater uncertainty, especially in larger debris-covered areas.



**Fig. 5.44.:** Potential permafrost distribution map obtained with Random forest (random sampling) and corresponding uncertainties.



**Fig. 5.45.:** Potential permafrost distribution map obtained with Random forest (spatial sampling) and corresponding uncertainties.

According to table 5.20, in the Random forest maps, the *potential absence of permafrost* is over-represented in comparison to the maps obtained with the other adopted ML techniques, with more than 25% of the prediction grid having probabilities of occurrence ranging between 0 and 0.25. In turn, the *potential presence of permafrost* was predicted fewer times than other algorithms, with only 45% of the prediction grid having a probability of occurrence greater than 0.75.

	<i>Probability of occurrence</i>			
	<i>0 - 0.25</i>	<i>0.25 - 0.5</i>	<i>0.5 - 0.75</i>	<i>0.75 - 1</i>
<i>Random sampling</i>	26.11%	10.42%	18.21%	45.26%
<i>Spatial sampling</i>	25.46%	10.08%	17.70%	46.75%

**Tab. 5.20.:** Parts of the validity domain predicted as *permafrost very likely absent* (0-0.25), *permafrost likely absent* (0.25-0.5), *permafrost likely present* (0.5-0.75) and *permafrost very likely present* (0.75-1) by Random forest (% of total grid cells).

## 5.6 Synthesis

In this chapter the main properties of the collected permafrost data have been visualized and explored with different techniques. In the first steps of the study, the complexity of the raw dataset were unveiled with the employment of exploratory data analysis. The investigation of structures in the data and the presence of a complex and non-linear phenomenon was assessed, opening the way to the application of feature relevance approaches aiming at reducing the dimensionality of the dataset and at removing less relevant or redundant environmental predictors. These procedures allowed the creation of a new dataset to be employed with machine learning algorithms in further steps of the research.

Prior to the classification of permafrost presence and absence, this *cleaned* dataset was sampled according to two different strategies: one randomly selected the training, validation and test sub-sets from the permafrost observations and their corresponding predictors; the other spatially sampled permafrost data according to their location within the geographical space. More precisely, this strategy aimed at avoiding selecting similar pixels (belonging to the same mapped polygon, i.e. same rock glacier) in the train, validation and test sets, and thus potentially occurring in spatial auto-correlation.

The resulting sub-sets were classified with five machine learning algorithms: Logistic regression, linear and RBF- Support Vector Machines, Multilayer perceptron and Random forest. By computing classical quality measures, such as *accuracy*, *true positive*, *true negative*, *false positive* and *false negative rates*, *precision*, *recall* or *area under the ROC curve*, the quality of the classification was assessed for all mentioned techniques. It was thus possible to statistically test the model performances with a paired t-test. It unveiled that the significantly best model resulted to be RF, followed by MLP.

From the inferred classification models, the prediction of the potential occurrence of permafrost was finally possible within the entire region of study. The classification results were expressed in terms of probability of permafrost occurrence, allowing a more useful indication of the potential presence or absence of this phenomenon.

Given the stochasticity of a machine learning approach, a different set of probabilities was obtained at each repetition of the algorithms employed. This provided a measure of the prediction variability illustrating the uncertainties of the results. An overview of the permafrost extent at the local scale was thus presented with the corresponding uncertainty maps.

In the next chapter, a deeper interpretation of these results will be presented and the main research outcomes will be discussed. A particular attention will be given to the main characteristics and differences of the produced permafrost maps at the *micro scale*, including a comparison with the field evidences and other existing models.







## Discussion

In the following pages the results presented in chapter 5 will be discussed. The objectives and the research tasks of the present thesis will be reviewed according to the main outcomes that were obtained by analysing the collected permafrost data. The first part of this chapter will focus on the complexity of permafrost data as well as on the study of the feature relevance, whilst the second part will discuss the permafrost mapping obtained with machine learning classification techniques.

### 6.1 Permafrost data analysis and classification

#### 6.1.1 The complexity of the permafrost data

The first analysis performed on permafrost data unravelled a non-linear relationship between the permafrost evidences and the retained environmental variables. According to the preliminary overview of the distribution of permafrost presence and absence in relation to the selected variables (figure 5.1), no clear relationships were identified between the permafrost occurrence and a specific predictor. The only exception found were the *altitude* and the *MAAT* being extremely correlated as shown in figure 5.13. They are strongly linked to the presence and absence of permafrost occurrence. Indeed, evidences of the permafrost presence were more likely to be found at the highest altitudes and at the lowest MAAT values.

Neither parallel coordinates (figure 5.4) nor 2D manifold t-SNE mapping (figure 5.5) unveiled clear structures in the data or the existence of variables that discriminate the permafrost presence from the permafrost absence. The complexity and the non-linearity were thus assessed by computing the *complexity index* on the MSE measures obtained by applying linear and RBF-SVM classifiers to the original permafrost dataset (see section 5, table 5.2). The calculated value of 0.235 identified the non-linear relationship between permafrost occurrence and the input predictors. The complexity was then confirmed by applying an AGRNN to both raw and shuffled permafrost data, showing the existence of structures despite no clear evidences were discovered with exploratory data analysis (figure 5.6).

The presence of complex structures and the non-linearity of the raw permafrost dataset had an important impact on the classification of synthetic datasets that were obtained with *feature extraction* techniques (see section 5, figure 5.7). Indeed,

the application of linear and RBF-SVM algorithms on LLE-, kPCA-, and Sammon-extracted datasets poorly performed in comparison to the classification of the original data. This can be explained by the loss of information that occurred after the reduction of the data dimensionality. As a result, the structures that were present in the original set of data were completely lost during the process, resulting in classification performances that were similar to the ones obtained with completely shuffled data (process that destroys the structures in the data, see figure 5.25). AUROC values obtained with both classifiers on Sammon-extracted data were in fact close to 0.5, indicating a worthless model. In turn, some structures were preserved in LLE- and kPCA-extracted data, but the predictive capabilities of the linear and RBF-SVM resulted to be poor (AUROC ranging between 0.598 and 0.649, table 5.3) and not satisfactory enough to justify the employment of mentioned feature extraction techniques on permafrost data.

The study on the non-linearity of permafrost data allowed showing that the permafrost distribution is not, at least at the micro scale, a linear phenomenon that exclusively depends on the *altitude*. Even though this trend may be true at the regional scale, permafrost occurrence at the micro scale is more complex and it depends on predictors, such as the surface *granulometry*, *NDVI* or the *potential incoming solar radiation*, that are specific to that scale and that vary depending on the micro-topography. With the proposed data analysis approach, the non-linearity of the phenomenon, often assessed by field observations, was statistically confirmed.

These dependencies were highlighted by performed *feature selection* analyses carried out on the raw dataset. Indeed, Information Gain, RELIEFF, Correlation-based feature selection, SVM recursive feature elimination and Random forest variable importance had all identified *MAAT* and *altitude* as the most important predictors, while the others were generally identified as being relevant but having similar weights (with some differences between the employed techniques). This indicates that no particular feature was able to discriminate the permafrost occurrence at the micro scale. Aiming at predicting the absence and the presence of permafrost at the local scale, *MAAT* and *altitude* (and *PISR*) would probably have been enough to discriminate the two classes with good classification performances. However, when looking at a finer scale (i.e. at the scale of a landform), these two variables were predominant but not enough informative to help predict the strong discontinuity of permafrost, requiring a combination of the remaining features.

A surprising result of the *feature relevance* analysis was the FS score attributed to *eastness* and *northness*. When investigating the potential presence of permafrost of a specific sector, it is generally accepted that the orientation plays a major role (i.e. Lewkowicz and Ednie, 2004, see section 2.1.1). It was for example observed that rock glacier fronts are located at higher altitudes in South-exposed slopes compared

to North exposed sides. Although this common consideration remains valid for the description of the permafrost phenomenon at the local scale, the value of the *eastness* and *northness* predictors did not provide any useful information for the improvement of the permafrost data classification. From a statistical point of view, a specific value of the *aspect* variable (even though it is expressed as the cosine and the minus sine of aspect angle in this particular case) was not relevant and it did not explain the permafrost occurrence. The *potential incoming solar radiation* predictor was in turn more important for this purpose and, since permafrost phenomenon reaches lower PLL at different aspects thanks to the influence of PISR, the latter was expressed with different values according to the *aspect* of the terrain. PISR was thus identified as an important predictor by IG, RELIEFF and RF-VI, and it obtained a final *FS* score of 0.414 (see table 5.4).

### 6.1.2 The relevance of dimensionality reduction

At first glance, the dimensionality reduction (feature selection and feature extraction) might be considered an avoidable step of the modelling process since some features were removed and some data were consequently lost in the process. Moreover, the accomplishment of this stage required some computational time and the employment of non-linear algorithms that cope well with dependent predictors could overcome the presence of redundant variables in the data. MLP, for example, could theoretically approximate any function; however the use of such redundancies could have introduced unnecessary complexities, risking an increasing number of local minima (see e.g. Witten and Frank, 2005).

Some benefits were thus identified when performing this step. On the one hand, not removing redundant or non relevant predictors from the dataset would have resulted in the maintenance of noisy data that might have been classified with good modelling performances. Despite a false feeling of good performance, the latter generally results in model over-fitting rather than in good generalization abilities of the classification algorithm when applied to new permafrost data in prediction. On the other hand, reducing the dimensionality of the input feature space generally allows both using algorithms of higher complexity and performing a finer search of the optimal model hyper parameters. This search is generally the most time consuming part when using machine learning algorithms (i.e. the grid search optimization performed in this study aiming at the optimisation of the RBF-SVM parameters).

One of the main objectives of this research was the investigation of the optimal algorithm allowing the analysis and the mapping of the potential permafrost distribution. The use of dimensionality reduction have made possible the comparison of different ML algorithms. Indeed, the algorithms employed cope differently with correlated

features. For example, if the Gini index is used as the splitting rule, Random forest performances suffer from correlation between predictors (Nicodemus and Malley, 2009). Similarly, Logistic regression, Multilayer perceptrons and SVMs are also dependent upon correlated features (e.g. Toloşi and Lengauer, 2011). Therefore, the removal of redundant and useless features was a required step to adopt in order to test and select the most relevant machine learning techniques for permafrost data (see Dormann et al., 2012; Dormann et al., 2013).

A smaller set of environmental predictors is also more comprehensible for non-expert users. The input-output relationship between permafrost occurrence and the controlling factors is more interpretable when analysing and comparing the modelling results and the corresponding uncertainties that were obtained with the reduced 15-dimensional permafrost dataset, rather than the original 26-dimensional feature space. In addition, the application of *feature selection* instead of *feature extraction* techniques did not change the physical meaning of the adopted variables. In fact, if the synthetic permafrost data extracted with LLE, kPCA and Sammon maps would have provided useful results, they would not have been easily interpretable since they were obtained from the combination of original inputs mapped to a lower manifold.

### 6.1.3 The effect of random and spatial sampling on permafrost spatial modelling

The selection of an appropriate sampling procedure was a very important step of the modelling approach (see figure 4.2 in section 4.1.1). Indeed, the availability of a large number of permafrost observations (93'705 pixels in total), which were the consequence of the employment of polygon-derived training data, would have required higher computational times without a proper pre-treatment of the data. For this thesis, it was thus decided not to subjectively select punctual data, as other researches for permafrost distribution modelling successfully employed (e.g. Azócar et al., 2016; Sattler et al., 2016). In these studies, climatic conditions were approximated at the rock glacier initiation line only (roots, i.e. in Sattler et al., 2016) or at the end of the rock glacier front (i.e. Azócar et al., 2016). On the contrary, as show in section 3.2, the permafrost evidences collected in this research were derived from mapped polygons indicating the known permafrost absence and presence. The final permafrost dataset was created by all polygon pixels to which environmental variable values were associated.

Performing a random cross-validation directly on the high number of evidences contained in the dataset would have potentially led to the selection of very similar or potentially unrepresentative observations. Searching *independent and identically dis-*

*tributed evidences* (the so-called *i.i.d.* or *IID* data; Witten and Frank, 2005; Le Boudec, 2010) was behind the idea of the used sampling strategy, involving a clustering of data that uses self-organizing maps (SOM; Kohonen, 1982; Kohonen, 2000; see section 4.6). The first step of the procedure focused this on splitting the permafrost dataset into 10 sub-sets aiming at the selection of *i.i.d.* data. Permafrost evidences contained in these pools required to be *representative* and *equally distributed*, meaning that the largest palette of variable/permafrost occurrence combinations had to be well represented in each of the sub-sets. Basically, finding similar clusters within the variable space (15D), made it possible to select the observations from all the retained clusters and to populate each sub-set with evidences that were not too similar to each other. This step was performed with the assumption that similar permafrost evidences were gathered in each cluster. By taking a balanced portion of evidences from the identified clusters to populate the sub-sets, the latter ended to being *representative* of the original permafrost dataset (in terms of data distribution), while being reduced in size (around 9370 evidences, 10% of the original size).

The second step of the sampling strategy involved splitting the 10 newly selected sub-sets into training, validation and test sets by using *random* and *spatial* 10-fold cross-validation. This operation provided the creation of a total of 100 (sub-)sub-sets (repetitions). With random cross-validation, the training sets were randomly selected across the geographical space, meaning that training observations could potentially be located in the neighbourhood of a testing observation (as shown in figure 5.16.A). Conversely, with spatial cross-validation, the partition of the data were performed at the polygon level, meaning that training observations were selected from polygons that are different than the ones chosen as testing examples (as illustrated in figure 5.16.B). This two sampling strategies avoided inferring a model from a training set that was very similar to the test set, resulting in meaningless modelling performances. In other words, assuming that two neighbouring pixels are (almost) identical, the main goal of spatial sampling was to avoid the use of a permafrost evidence for training and an identical evidence (in terms of geo-spatial characteristics) for testing the model inferred during the training phase. The quality of the classification was thus more *honest* because it was computed on spatially different data.

As shown in section 5.4.2, classification algorithms were on both random and spatial sampled data. By analysing the classification performances, three main considerations can be made:

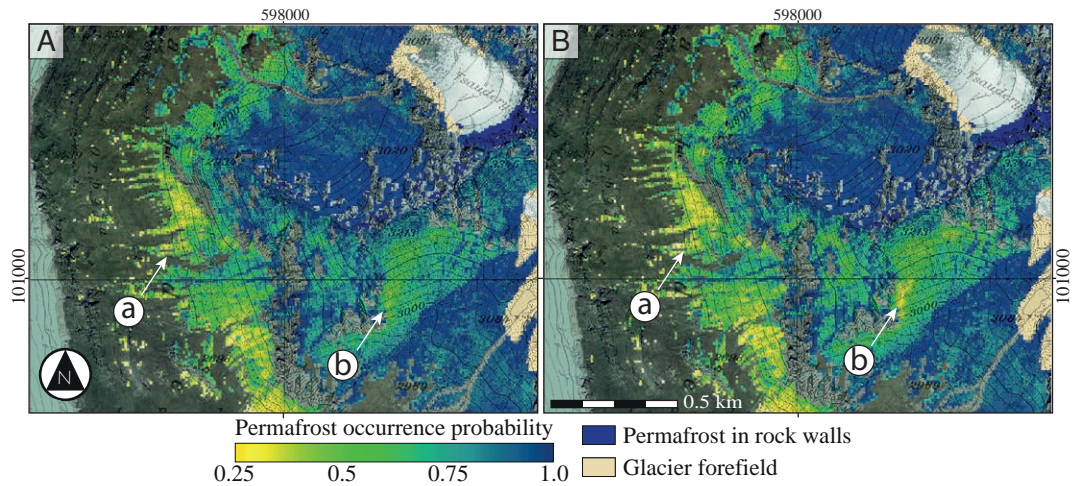
1. Classification performances obtained for each one of the 10 pools of data resulted in very similar AUROC values (figure 5.25), indicating that the first step of the sampling strategy (involving the clustering with SOM) provided very similar sub-sets of data (respecting the "*independent*" and "*equally distributed*" properties proper to *i.i.d.* data). This confirmed that a dataset containing a

high number of observations can be reduced in size if the original distribution of data is respected;

2. The analysis of the modelling performances showed that the generalization ability of the employed techniques were very similar for both randomly and spatially sampled pools of data. However, as a general trend, AUROC values of the classifications, made on spatially sampled permafrost evidences, also resulted in being slightly lower than the ones obtained with randomly sampled data (figure 5.25). This can be explained by the existence of very similar permafrost observations in both randomly selected training and test sets, as it was mentioned above. Consequently, this similarity results in higher measures of the classification quality;
3. The evident difference between the two sampling strategies was the result of the variability obtained among the 100 repetitions that was computed by applying classification algorithms on spatially sampled data (as illustrated in figure 5.13). AUROC values differed the most from the average AUROC because different permafrost evidences were used during the training and the test phases. Pixels of a rock glacier, for example, could be selected to train the model and, because with spatial sampling training and testing observations could not be selected within the same polygon by design, the classification test could only be made on pixels belonging to another rock glacier. Consequently, assuming that the two rock glaciers had different topo-climatic characteristics (e.g. contrastive orientation and altitude), the permafrost evidences contained in both sets were more likely to differ.

Thanks to the availability of a high number of polygons (523), the computation of the 100 repetitions resulted in an average model performance that converged to a similar value obtained with randomly sampled data. In the end, the potential permafrost occurrence predicted on the entire validity domain resulted in being very similar regardless the chosen sampling strategy. Indeed, as shown in figure 6.1 that illustrates the RF permafrost map obtained for the sector located at the South of the Tsaudery glacier, no clear differences can be observed between predicted occurrences. The West-exposed slopes that dominate the Lac des Dix present some slight variations (around 10%) in probability (*a* in figure 6.1), as well as a small portion of the South-exposed talus slope located at a higher altitude (*b* in figure 6.1). The same behaviour can be observed in all the other maps as a result of the employed ML algorithm. However, the variation in probability are not strong enough to declare that random sampling is significantly better or worse than spatial sampling in terms of quality of the final mapping.

According to what was mentioned above, if the number of evidences are enough and the largest palette of variable/permafrost occurrence combinations are well represented in each of the sub-sets, a spatial sampling strategy is recommended



**Fig. 6.1.:** Differences of the prediction of permafrost occurrence according to a random sampling (A) and spatial sampling (B) strategy, with (a) and (b) representing two sectors with different permafrost probability of occurrence.

because it is more suitable for the selection of independent training and test sets, especially when using polygons to map the evidences of a phenomenon. The resulting classifications are indeed less prone to over-performing because they are tested on data that are less similar to the data used to infer the classification models. According to these observations, in the present study it was decided to retain only the maps of the potential permafrost distribution that were obtained with the spatial sampling strategy (see section 6.2).

#### 6.1.4 Analysis and discussion of permafrost data classification

Logistic regression, linear and RBF-SVM, Multilayer perceptron and Random forest were applied to permafrost data with the aim of inferring a classification model able to predict the permafrost occurrence in locations where its presence was unknown. Classification performances, presented in section 5.4, revealed different behaviours depending on the characteristics of the algorithm. AUROC values ranged between 0.78 (linear SVM) and 0.97 (RF) for randomly sampled data and between 0.75 (linear SVM) and 0.88 (RF) for spatially sampled data. According to what was discussed in section 6.1.3, the retained models were those resulting from the classification made on spatially sampled permafrost data.

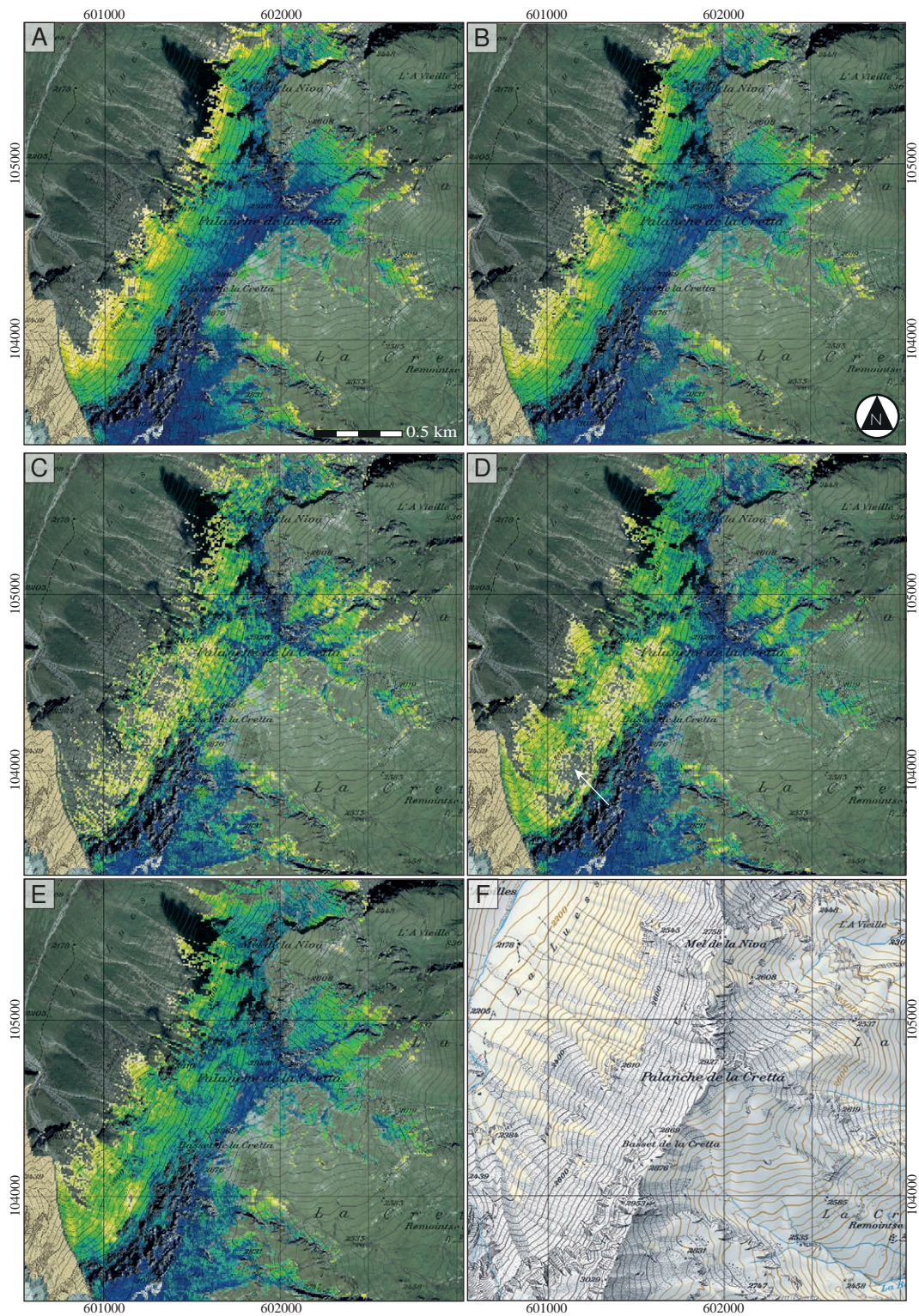
Although each employed algorithm produced a good classification on the test sets, predictions differ according to the type of techniques used to map the potential permafrost occurrence. Figure 6.2 illustrates the differences between the predicted probabilities of permafrost occurrence obtained with LR (figure 6.2.A), linear SVM (figure 6.2.B), RBF-SVM (figure 6.2.C), MLP (figure 6.2.D) and RF (figure 6.2.E). The map shows the potential permafrost distribution around the Palanche de la



Cretta peak (Hérens Valley). The overall predictions are very similar for both models, with the probabilities that only slightly differ in very small sectors. At the local scale, it is in fact barely impossible to see differences. The results present a common trend that emerged with both techniques: the probabilities increase with the increase of the altitude. This translates in two maps that do not consider the non-linearity of the permafrost distribution and that are thus not correct at the micro scale. This result is explained by the important weight given to the *altitude* variable by the two algorithms. Indeed, as shown in tables 5.6 and 5.8, this environmental predictor was considered as extremely important for the classification of permafrost data. The prediction of the permafrost occurrence was thus computed being strongly influenced by this variable, while other features were less considered.

Conversely, since Multilayer perceptrons, Radial basis function SVM and Random forest are non-linear techniques, they were able to take into account the non-linearity of the permafrost distribution and the complexity evidenced in permafrost data during the first data analysis discussed in this chapter (section 6.1.1). As a result, permafrost maps do not present the same behaviour manifested by LR and linear SVM (figure 6.2). Indeed, the *altitude* predictor does not seem to influence the permafrost probabilities as much as observed in figure 6.2.A. For the North-West exposed talus slope, the RBF-SVM result is less optimistic than MLP and RF (figure 6.2.C), presenting a greater surface predicted with low probabilities of permafrost occurrence. In turn, RF map (figure 6.2.E) shows more homogeneous surfaces with talus slopes generally characterized by probabilities of occurrence between 50% and 75%. It is worth noting that, especially at the East of the glacier forefield, the probabilities are greater in the distal part and less important in the upper part (see location indicated by an arrow in figure 6.2.D). This trend is even more remarkable for the MLP result, which predicted *permafrost likely absent* upslope (white arrow in figure 6.2.D). These simulations are in accordance with the considerations discussed in section 2.2.1, indicating that in talus slopes permafrost is less likely to occur in their upper part.

The *paired t-test* computed on AUROC revealed that RF and MLP results were significantly better in comparison to results obtained with LR, linear and RBF-SVM (see section 5.4.3). Both algorithm AUROC values (MLP: 0.84; RF: 0.88) were indicative of an excellent generalization ability and thus, both techniques were able to produce reliable permafrost distribution maps. However, a selection of the best technique could be made by considering other computed measures of the classification quality. Computed *precision* values were particularly interesting for assessing the best model to be employed. This index provided indeed an information about the positive predictive value (PPV) of a model. This measure illustrated the proportion of positive predictions that were *true positives*. In other words, the PPV is the probability that a pixel with a positive permafrost occurrence truly represents a



**Fig. 6.2.:** Comparison of the probability of permafrost occurrence for the talus slope of the Palanche de la Cretta. Maps carried out with LR (A), linear SVM (B), RBF-SVM (C), MLP (D) and RF (E). The swisstopo topographic map is shown in (F) for an overview of the ground truth.

pixel with the presence of permafrost and it is not a *false positive*. MLP classifications scored a precision of 0.741 (table 5.11), while the precision obtained with RF was 0.775 (table 5.12). As a consequence, RF produced a better PPV (77.5%), compared to MLP (74.1%), when applied to permafrost data.

The analysis conducted on uncertainties also evidenced a less uncertain result obtained with RF, especially when studying the distribution of the uncertainties in relation to the predictor values. Indeed, uncertainty plots 5.33 and 5.35 showed that, despite similar generalization abilities, the MLP map is much more uncertain than the RF one. The latter only predicted high variability among the 100 repetitions (in terms of probability of occurrence) at *altitudes* above 3000 m.a.s.l. and, in general, only for sectors with flat curvature indices. In terms of probabilities, most uncertain RF results were obtained for pixels with 50 to 75% of probability of permafrost occurrence, while MLP predictions were more variable for a larger range of values (25 to 75%). These considerations, in addition to what was discussed above, make of RF the best algorithm that provides the most reliable representation of the potential permafrost distribution within the selected region of interest.

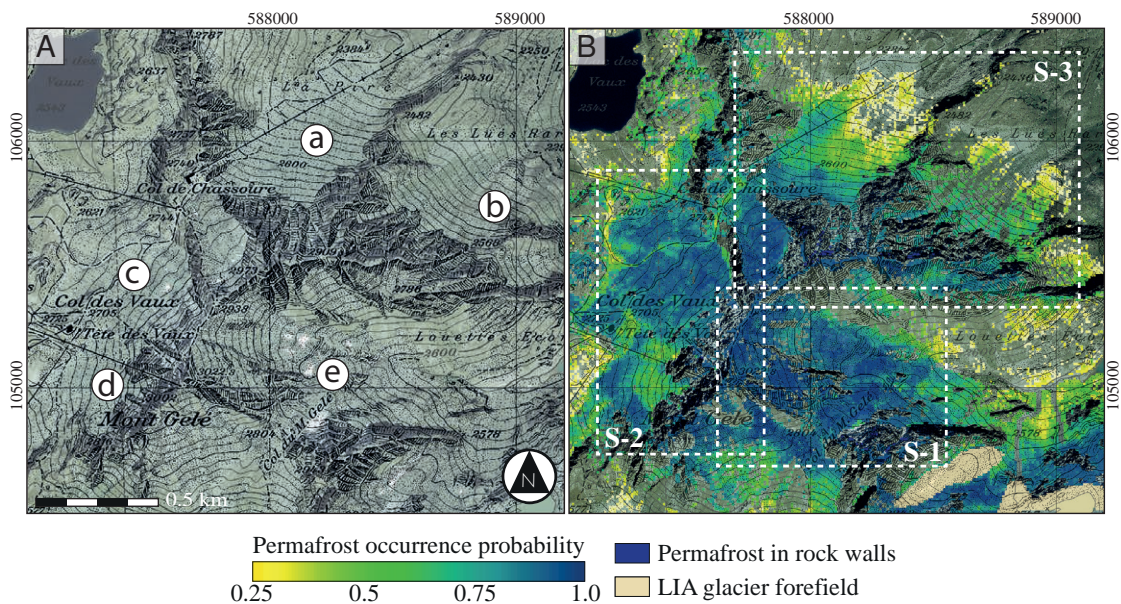
## 6.2 Potential permafrost occurrence maps: interpretation and model comparison

In this section, a sectoral analysis will be conducted to unravel the main features of several areas where theoretical permafrost conditions were expected. The selected sites offer different topographic settings as well as a geomorphological diversity. The main characteristics of the retained RF permafrost map can thus be illustrated and the main advantages of the employment of a data-driven approach for permafrost mapping can be discussed. The acronym *PERMAL*, standing for *PERmafrost modelling with MACHine Learning* (as it was proposed in Deluigi and Lambiel, 2013), will be used further in this chapter and it will refer to the permafrost maps obtained with the employed data-driven approach.

The field studies carried out in recent years by the Universities of Lausanne and Fribourg have provided a deep knowledge on the distribution of permafrost at the micro scale. The investigation of the principal peculiarities of PERMAL will thus also be supported by expert knowledge as well as by the comparison with the BAFU map (see section 2.4.2; BAFU, 2005), which is still a widely used representation of the potential permafrost extent in Switzerland.

## The Mont Gelé region

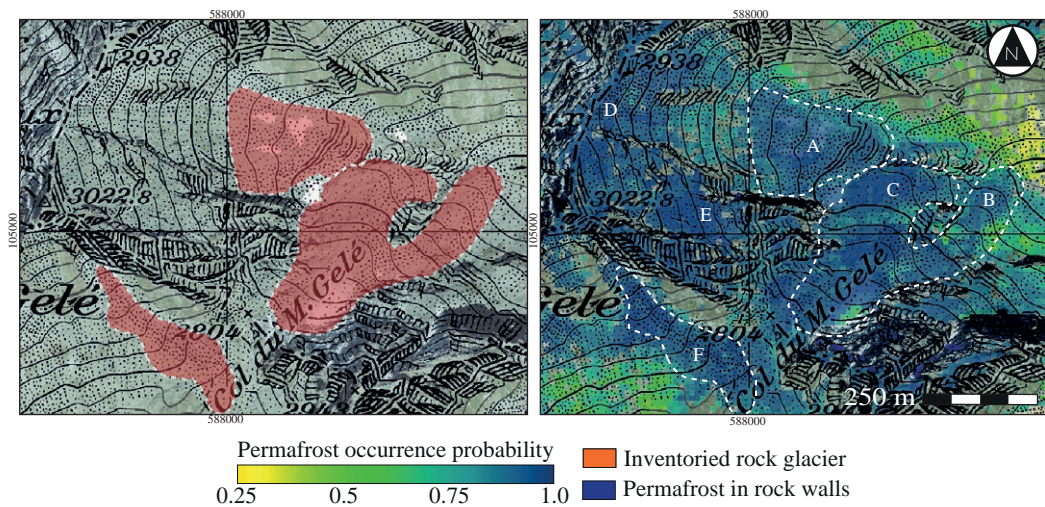
Figure 6.3.A shows the extent of the Mont Gelé region (WGS84: 46.09694, 7.27910; CH1903: 587'662, 105'061), a sector that presents a high variability of periglacial landforms and where field investigations have been carried out during recent years (see section 3.1). According to PERMAL (figure 6.3.B), this region is characterized by large areas covered by permafrost. Its presence is more important in the North, North-West and North-East-oriented sectors, while the predicted probabilities are lower for the other orientations. The PLL was automatically recognized and the limits between the mineral surfaces and other land-cover types were respected due to the validity domain delimitation, performed before the prediction of the permafrost occurrence. Figure 6.3.B also presents the three sectors where the principal studied periglacial landforms of the region are located: the Yettes Condja rock glaciers, which develop on the East side of the Mont Gelé (S-1 in figure 6.3.B); the Les Attelas talus slope, the Lac des Vaux area with a rock glacier and three periglacial lobes, located on the West side of the Mont Gelé (S-2 in figure 6.3.B); the Chassoure and the Lues Rares rock glaciers, as well as the Lapires talus slope, found on the North side of Mont Gelé (S-3 in figure 6.3.B).



**Fig. 6.3.:** (A) Presentation of the Mont Gelé sector: (a) *Lapires* talus slope (b) *Lues Rares* rock glacier (c) *Lac des Vaux* periglacial lobes (d) *Attelas* talus slope (e) *Yettes Condja* rock glaciers; (B) Potential permafrost occurrence map for the Mont Gelé area, based on the RF classification and localization of the the *Yettes Condja* sector (S-1), the *Les Attelas/Lac des Vaux* sectors (S-2) and the *Lapires/Lues Rares* sites (S-3).

In the *Yettes Condja* sector (figure 6.3.A/e), the permafrost probabilities computed for the three intact rock glaciers are indicative of permafrost presence (figure 6.4). Indeed, the rock glacier lobes were well recognized with higher permafrost occurrences. This result is in accordance with both field data (Lambiel, 2006; PERMOS,

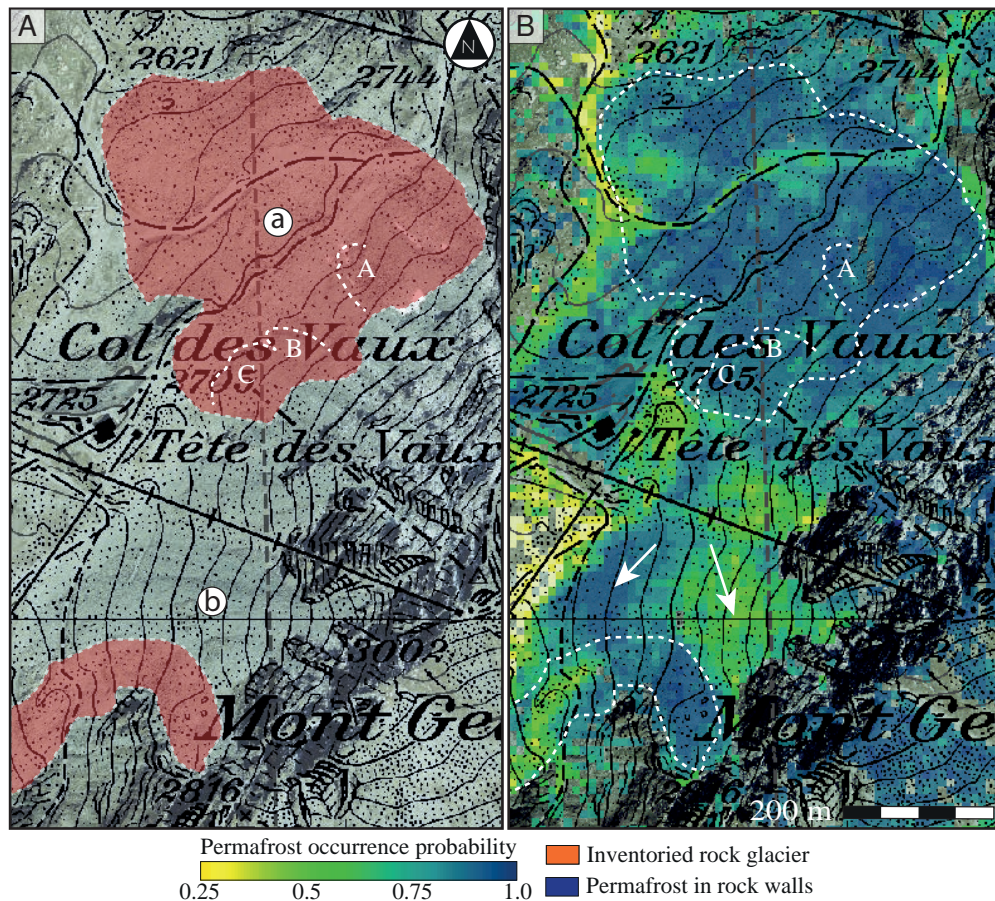
2016) and morphological characteristics observed by ortho-image interpretations (steep fronts and ridges on the rock glacier lobes). South-West of the Col du Mont Gelé, another rock glacier was mapped in the inventory (*F* in figure 6.4). Also for this landform, PERMAL predicted high probabilities of permafrost occurrence, although it does not present a well developed lobe like the one on the Yettes Condja rock glaciers (*B* and *C* lobes in figure 6.4). RF probably considered the coarse grain size of this sector as the most contributing predictor, rather than using the values of the curvature indices (due to the local flat surface of these sites). The same variable probably helped RF predict high probabilities in the talus slopes on the East side of the Mont Gelé (*D* and *E* in figure 6.4).



**Fig. 6.4.:** (A) S-1 sector, characterized by three active rock glaciers on the East side of the Mont Gelé peak and (B) PERMAL probability of permafrost occurrence based on RF classification. (A, B, C & F) Rock glaciers; (D & E) Coarse debris-covered surfaces.

A significant portion of the western slope was also mapped as affected by the presence of permafrost (figure 6.5). In the Les Attelas talus slope, PERMAL predicted high probabilities of occurrence in the mid-lower part of the slope (sectors indicated with white arrows in figure 6.5). This result is in accordance with the borehole and geoelectrical data (Lambiel and Pieracci, 2008; Scapozza et al., 2011; Scapozza, 2015b; Scapozza et al., 2015; PERMOS, 2016). The upper sector (*D* in figure 6.4), characterized by a coarser surface granulometry and located above 2800 m.a.s.l., was also predicted by PERMAL as potentially frozen.

Features of the rock glacier located within the same slope were also considered by PERMAL as potentially favourable for permafrost. Indeed, the rock glacier surface was predicted with high probabilities, especially in the top part of the landform. This outcome revealed to be correct in comparison to field observations (see figure 3.5 in chapter 3).



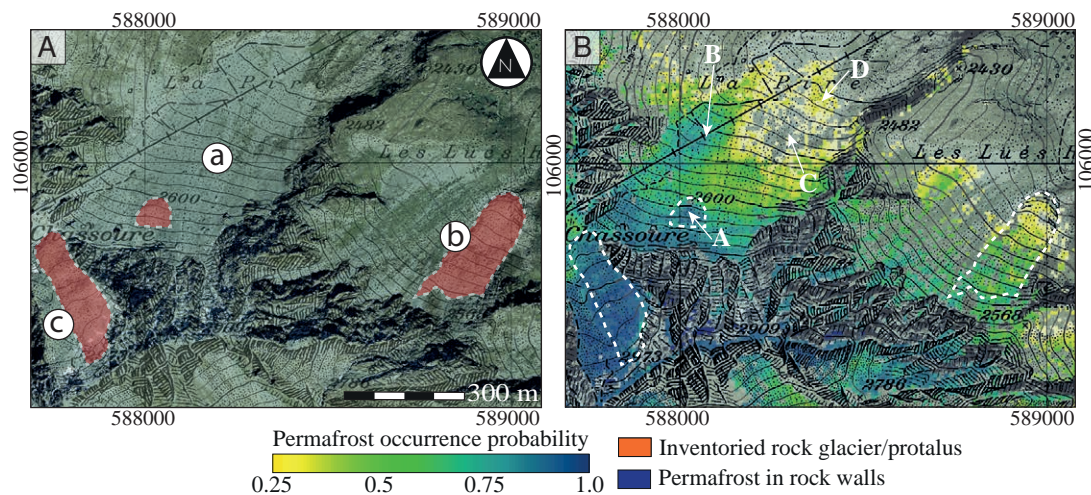
**Fig. 6.5.:** (A) the Mont Gelé western slope (S-2), with the localization of the Lac des Vaux sector (a) and the Les Attelas talus slope (b). (B) PERMAL map of the permafrost extent and localization of the Lac des Vaux periglacial lobes A, B and C (see i.e. Scapozza, 2015a). The white arrows indicate a difference in permafrost probabilities between the mid-upper part and the mid-lower part of the Attelas talus slope.

On the North, the Lac des Vaux sector presents topoclimatic and morphological characteristics that RF recognized as being favourable for permafrost presence. Indeed, PERMAL shows high probabilities of permafrost occurrence, especially within the landform mapped as an intact rock glacier in the inventories used for this study (figure 6.5.a). The Lac des Vaux area was studied by Lambiel (2006), Scapozza (2013) and Scapozza (2015b), who investigated the complex environment characterized by a talus slope in the upper part, with protalus rampart/periglacial lobes developing, and by a rock glacier in the lowest part. Scapozza (2015b) indicated that the three lobes are affected by permafrost. The result of PERMAL seems thus correct, and offering a good overview of the potential permafrost extent at the micro scale.

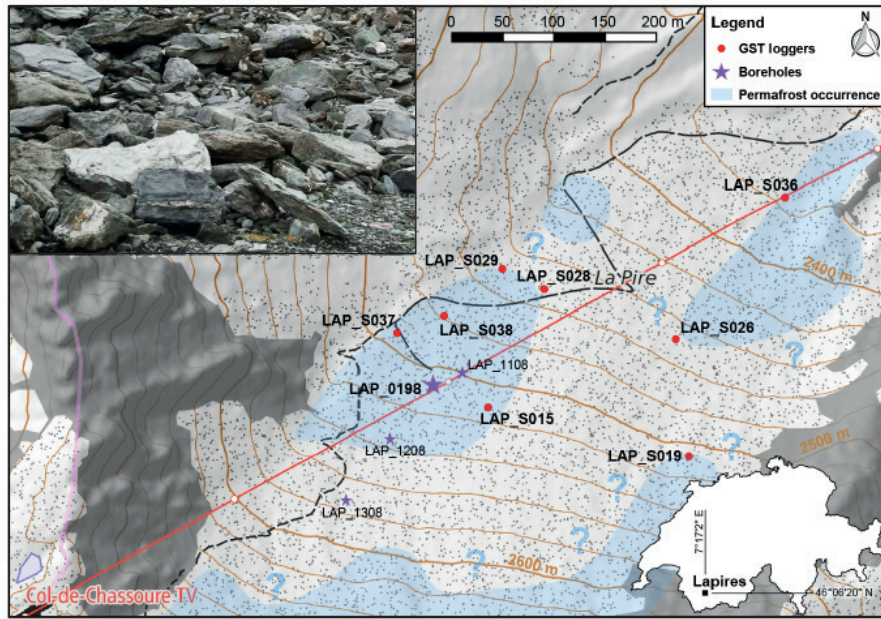
Above the Col de Chassoure (figure 6.6.c), in the area occupied by an intact rock glacier, high probabilities of permafrost occurrence were predicted. Conversely, the Lues Rares rock glacier (figure 6.6.b) is characterized by lower probabilities of

permafrost occurrence. The outcome is probably due to the altitude of the sector at the rock glacier front. Its altitude of 2330 m.a.s.l. is indeed lower than the limit of the periglacial belt reported by Lambiel and Reynard (2003). They observed that the lowest limit of (possible) permafrost for North-East-exposed sites is 2380 m.a.s.l.. Since RF learnt the PLL directly from data (permafrost evidences), the lowest altitude for this orientation is generally located at higher elevations. Although the permafrost probability is under-estimated for the Lues Rares rock glacier (at least at its front), PERMAL was still able to predict favourable permafrost conditions.

The discontinuous permafrost distribution of the Lapires talus slope (figure 6.3.A/a) was also properly modelled, with potential permafrost presence particularly predicted in the western part of the landform (figure 6.6.B). Permafrost conditions are indicated by PERMAL at the top as well as in the mid-lower part of the slope (B, white arrow in figure 6.6.B). The map is in accordance with Delaloye (2004), Delaloye and Lambiel (2005), Lambiel (2006), Scapozza (2013) and Staub et al. (2015), who indicate that permafrost is more likely to occur in this portion of the talus slope, rather than in its Eastern part. Since vegetation or fine debris exist in this sector, PERMAL indicated unfavourable permafrost conditions and the absence of permafrost in the central part of the slope (C in figure 6.6). Low probabilities of occurrence are computed for the distal part of the landform (D in figure 6.6), where permafrost conditions were identified by mentioned studies. This result can be explained by the altitude of this sector, which is located at the limit of the detected PLL. Accordingly, PERMAL under-estimated the permafrost probability, similarly to the prediction made for the Lués Rares rock glacier.



**Fig. 6.6.:** Presentation of the S-3 sector with the localization of the (a) Lapires talus slope, (b) the Lues Rares and (c) the Chassoure rock glaciers (left image). Right image shows the probabilities of permafrost occurrence predicted by PERMAL and (A) the localisation of a protalus rampart on the top part of the Lapires. (B) Portion of the landform characterized by greater probabilities in its mid-lower part.



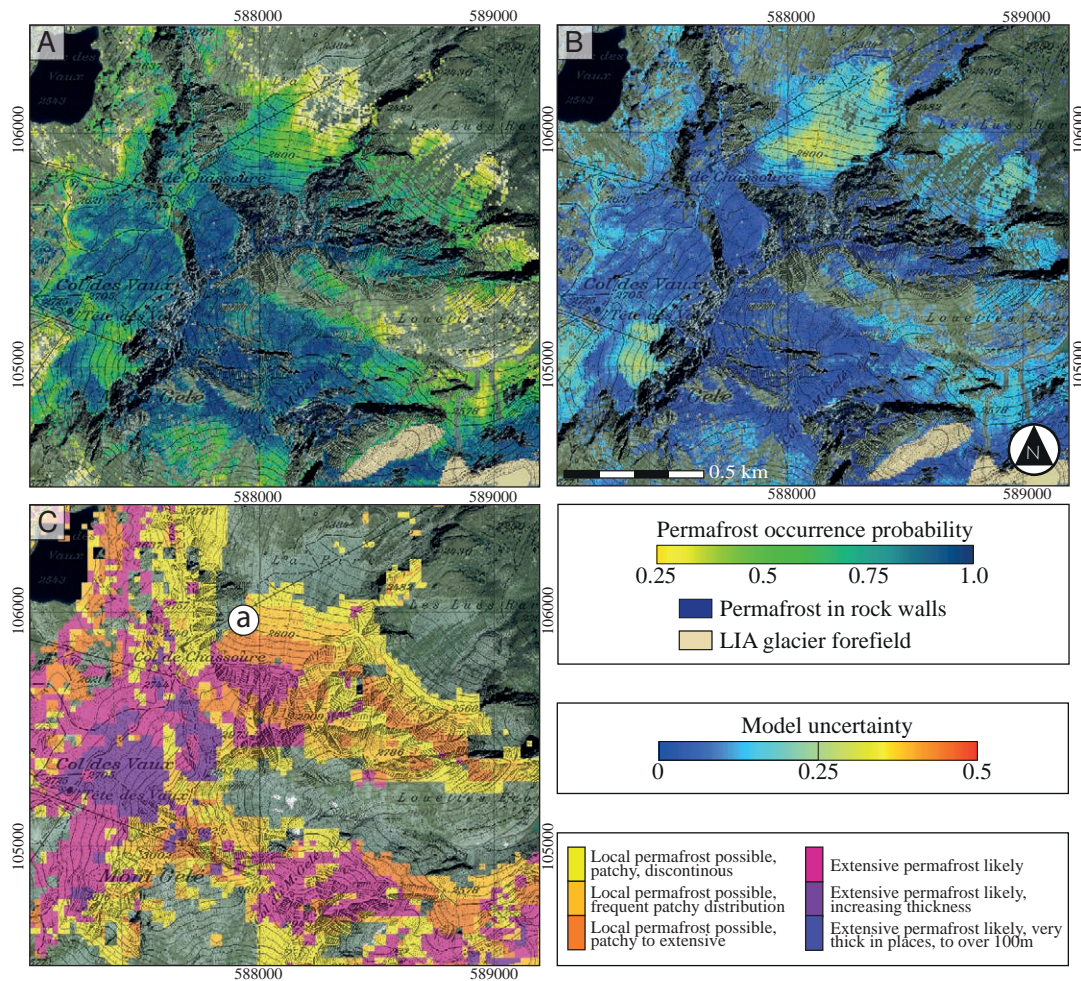
**Fig. 6.7.:** Qualitative indication of the spatial extent of permafrost occurrence according to Delaloye (2004), Lambiel et al. (2016) and Scapozza (2013), as well as the location of boreholes and GST loggers and an illustration of the coarse grain size of the Lapires talus slope. Figure reproduced from Staub et al. (2015).

The small protalus rampart located in the upper part of the sector was considered by PERMAL as potentially affected by permafrost (A in figure 6.6.B), probably because of the coarser grain size and its convex form. The comparison between the RF map of the Lapires talus slope and the delimitation of the affected permafrost sectors in Staub et al. (2015), gathering all field observations conducted in this landform (figure 6.7), shows that the two maps are very similar and, thus, confirms the reliability of PERMAL at the micro-scale.

The uncertainty map presented in figure 6.8.B illustrates moderate RF uncertainties computed at the lowest altitudes, corresponding to the PLL for different orientations. Besides, the greatest uncertainties are generally located in the mid upper part of the talus slopes. This is particularly valid for Les Attelas and Lapires talus slopes and can be explained by the flat curvatures of these sectors, where the most uncertain result was shown by the RF uncertainty plot (see figure 5.35).

By comparing the PERMAL classification to the BAFU map (figures 6.8.A and 6.8.C), RF predicted more reliable probability patterns at the micro scale. PERMAL results are more in accordance with the field data and the field observations than with the representation of the permafrost extent offered by the BAFU map. Permafrost occurrences obtained with the RF prediction are, for example, widely less extended. This means that the ML approach produced a less optimistic result than the empirical-statistical one being, in many cases, more consistent with the reality. The differences between the two models can be explained, on the one hand, by the design of the



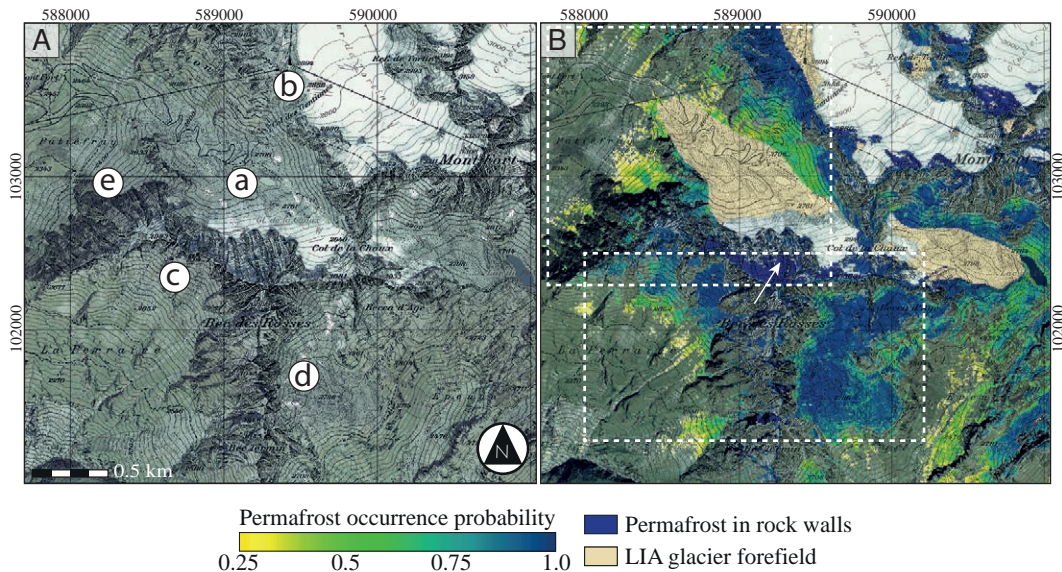


**Fig. 6.8.:** (A) PERMAL map of the permafrost extent of the Mont Gelé sector; (B) Uncertainty map for the Mont Gelé sector, illustrating the probability variability obtained during the RF classification. (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005). (a) shows the altitude thresholds used by this empirical-statistical approach for the delimitation of the permafrost categories indicated in the BAFU legend.

BAFU model, that was primarily based on Haeberli (1996)'s rules, which were developed for the Fluela region (Graubünden Canton). Compared to Valais intact rock glacier lower limits, the BAFU extent is too much optimistic in West/North-West orientations and too much pessimistic in North-East/East aspects (according to Lambiel and Reynard, 2001). On the other hand, an important difference of using a data-driven approach is that, unlike the BAFU map, it did not use any altitude thresholds. As illustrated in figure 6.8.B (colored altitude bands) using fixed threshold values produces abruptly changes between the permafrost classes. This is particularly noticeable for the Lapires talus slope (*a* in figure 6.8.C). Moreover, with the empirical-statistical approach used by BAFU, some landforms, such as the Lues Rares rock glacier, were not taken into account because located below the threshold limits.

## Bec des Rosses sector

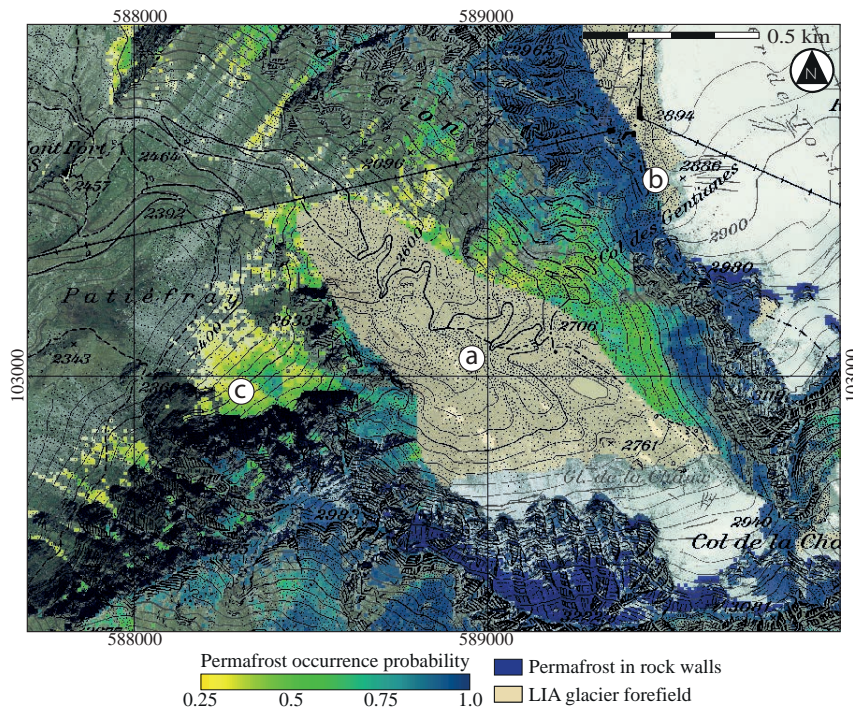
The Bec des Rosses sector (WGS84: 46.07194, 7.29998; CH1903: 589'272, 102'278; figure 6.9) is characterized by various glacial and periglacial landforms.



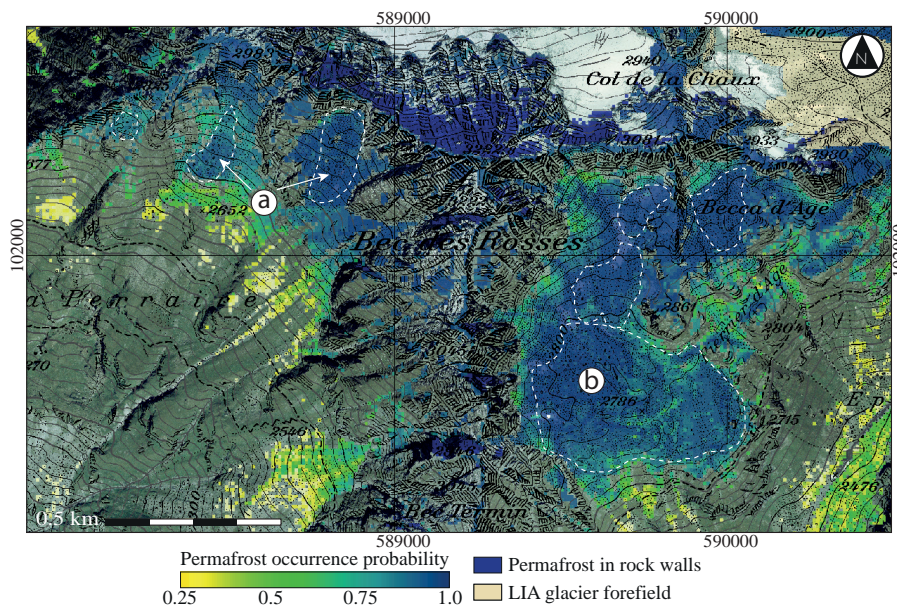
**Fig. 6.9.:** (A) The Bec des Rosses sector: (a) La Chaux glacier forefield (b) the Gentianes moraine (c) talus slope and two inventoried rock glaciers (d) rock glacier (e) Patiéfray talus slope; (B) Potential permafrost occurrence map for the Bec des Rosses area, based on the RF classification, and delimitation of two sectors presented in figures 6.10 and 6.11.

In the La Chaux glacier forefield, geoelectrical surveys conducted by Reynard et al. (2003) showed that the sub-surface ice was absent in most of the glacier forefield (figure 6.9.A/a). However, the debris-covered ice is still present in the slope located on the northern face of the Bec des Rosses. According to the map of the permafrost occurrence in rock walls, some parts of the northern slopes are frozen (white arrow in figure 6.9.B).

The Gentianes moraine (figure 6.10.b) was also mapped as frozen. PERMAL only predicted half of the moraine as being occupied by permafrost. This is due to the delimitation imposed by the validity domain that did not allow the prediction within these surfaces. However, the Gentianes moraine was widely investigated with geoelectrical surveys, LIDAR scans and borehole measurements (see Lambiel and Schuetz, 2008; Lambiel and Baron, 2008; Ravanel and Lambiel, 2013; PERMOS, 2016), unravelling the presence of permafrost conditions in the moraine. This example clearly shows that permafrost can be present in the margins of the glacier forefields. The latter not being mapped "by design" by PERMAL represents a limit of the adopted modelling approach. Consequently, the permafrost map requires careful interpretations when dealing with this type of environment.



**Fig. 6.10.:** PERMAL map of the northern sector of the Bec des Rosses. (a) La Chaux glacier forefield (b) Gentianes moraine (c) Patiéfray talus slope.

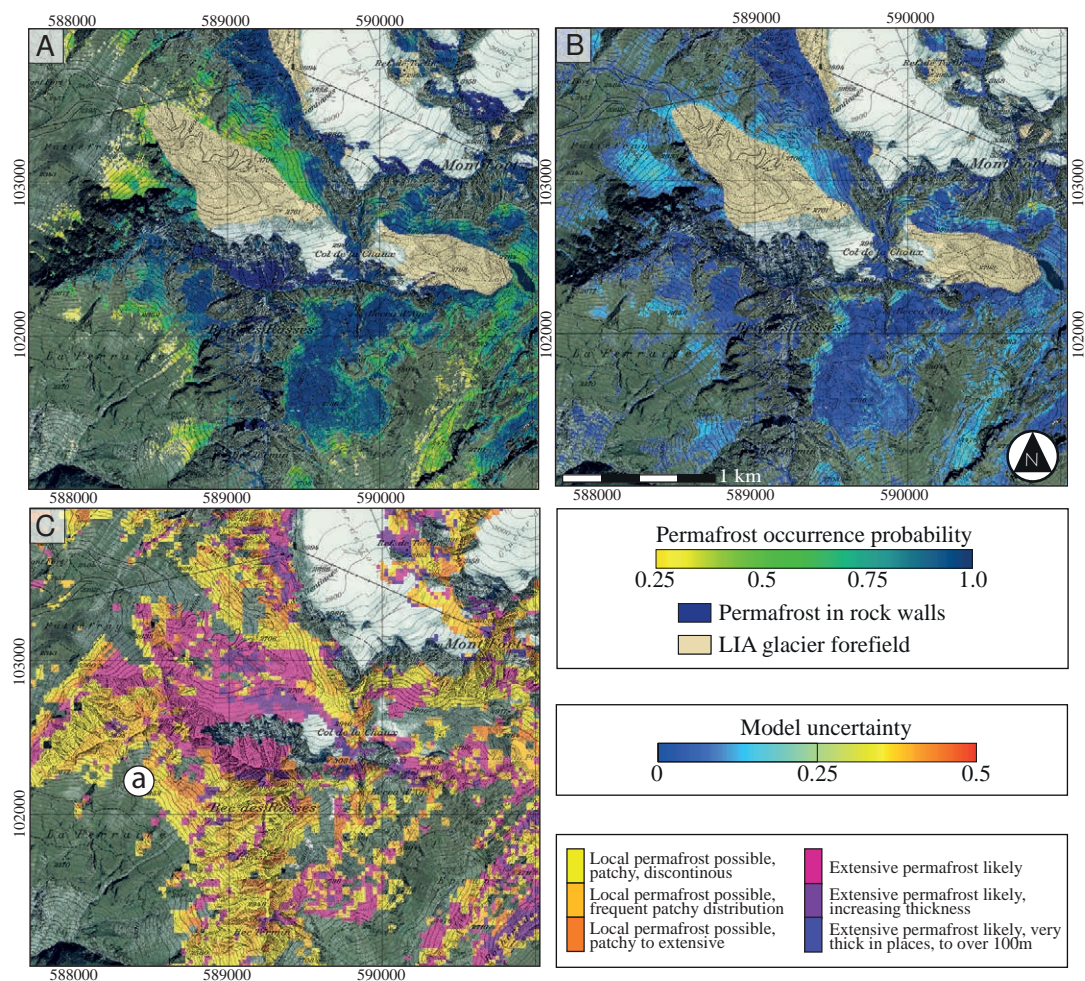


**Fig. 6.11.:** PERMAL map of the southern sector of the Bec des Rosses peak; (a) talus slopes and La Perraire rock glacier 1, 2 & 3 (b) Pierrier d'Aget rock glaciers inventoried by Lambiel (1999).

PERMAL also predicted positive permafrost occurrence in the upper part of the Patiéfray talus slope (figure 6.10.c) above the altitude threshold of 2450 m.a.s.l.. This elevation corresponds to the observed PLL of the region as reported in Lambiel and Reynard (2003). At the micro scale, this sector presents a convex topography due to the deposition of torrential debris at the bottom of a couloir (narrow gully).

The high probabilities of occurrence attributed by PERMAL can thus be explained by the morphological characteristics of the area. On the South-West-exposed slopes of the Bec des Rosses (figure 6.11), high permafrost occurrences were also predicted in debris surfaces. The inventoried rock glaciers that are located in these areas (6.11 a and b) were correctly recognised with probabilities of permafrost occurrence greater than 85%.

As it was shown for the Mont Gelé region, the uncertainties of the RF classification in talus slopes are greater when their slopes are regular and their curvatures are flat. Similarly to the previous sectoral analysis, the most uncertain sectors were located at the lowest altitudes, in correspondence with the PLL (figure 6.12.B).



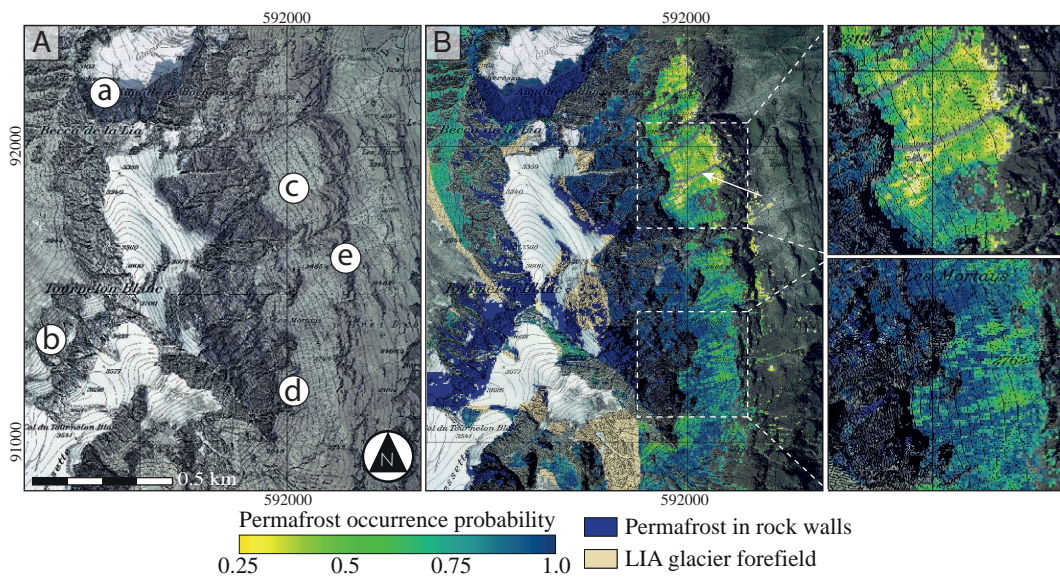
**Fig. 6.12.:** (A) PERMAL map of the Bec des Rosses area; (B) Uncertainty map of the RF classification; (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005).

The main difference between PERMAL and the BAFU map can be observed in the south-western slopes of the Bec the Ross peak (figure 6.12.C). The BAFU map indicates an important permafrost occurrence even in vegetation-covered surfaces (a

in figure 6.12.C), while the prediction obtained from the RF classification seems to be more optimistic as it shows a higher probability of permafrost occurrence in debris-covered surfaces. Unlike the BAFU map, PERMAL does not present any altitude threshold. Moreover, the Pierrier d’Aget rock glacier, located on the South-eastern part of the Bec des Rosses summit (figure 6.11.b), was also mapped differently. The BAFU map does not show continuous permafrost within the landform. PERMAL was in turn able to better recognize the rock glacier limits. However, according to InSAR measurements indicating almost no movements in a large portion of the landform (Delaloye et al., 2007), the PERMAL map presumably over-estimated the permafrost occurrence (figure 6.11). This outcome was probably achieved by taking into account the curvature indices and the coarse surface grain size of the rock glacier, predictors that are not considered by BAFU.

### Tournelon Blanc glacier sector

On the left side of the Mauvoisin lake, permafrost presence was mapped in the rock walls situated on the northern sides of the Aiguille de Bocheresse and the Becca de la Lia (WGS84: 45.97993, 7.31869; CH1903: 590’704, 92’047; location *a* in figure 6.13.A). Similarly, the rock walls below the Tournelon Blanc glacier are largely affected by permafrost (location *b* in figure 6.13.B).



**Fig. 6.13.:** The Tournelon Blanc glacier sector: (a) Becca de la Lia rock walls (b) Tournelon Blanc rock walls (c & d) talus slopes (e) torrential cones; (B) Potential permafrost occurrence map for the Tournelon Blanc glacier area, based on the RF classification.

According to PERMAL, the debris-covered surfaces located on the East of the sector are also partially occupied by permafrost. The vegetation and the mineral surfaces were well delimited by the validity domain, resulting in a permafrost map that

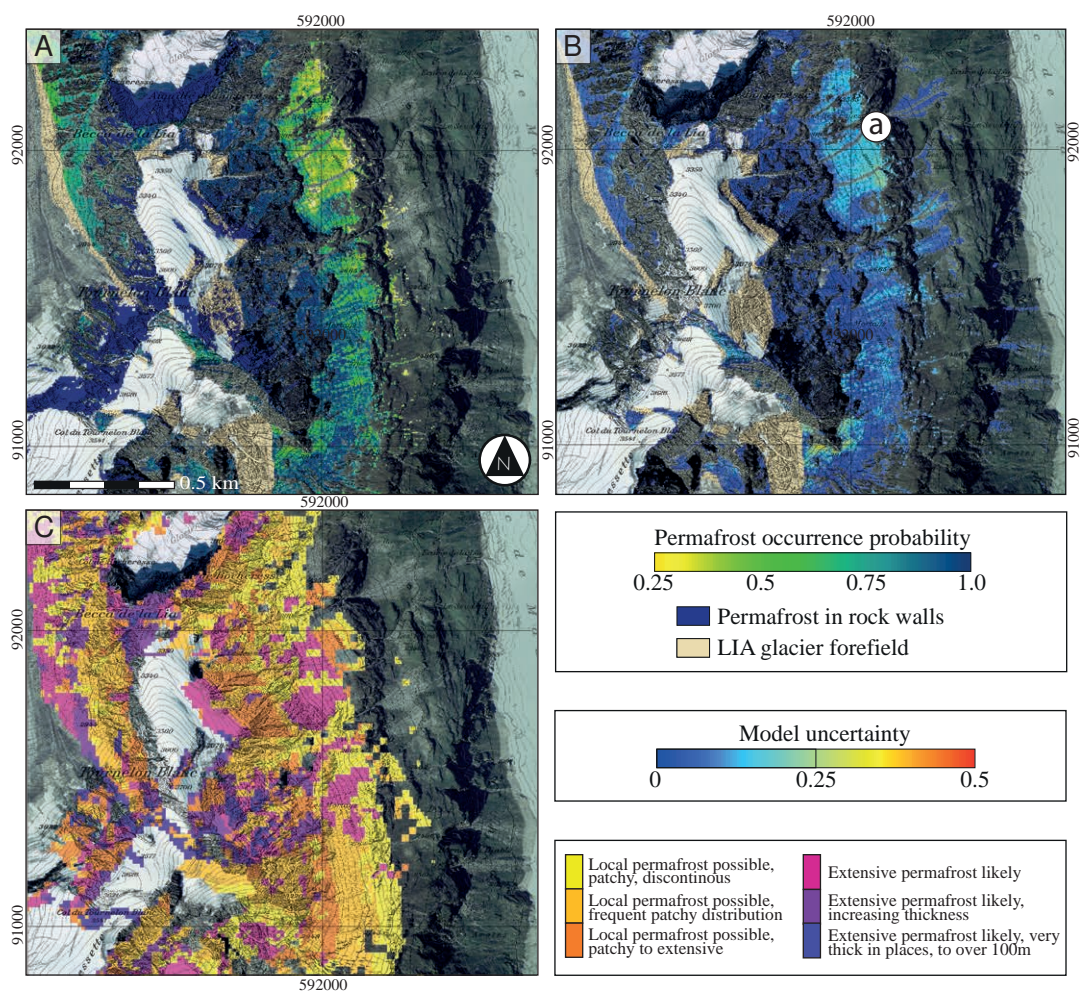
only indicates the probabilities of occurrence in restricted areas of the slope (figure 6.13.B). The talus slopes below the Becca de la Lia are good examples that show how the validity domain delimitation allowed excluding the hydrological network from the predicting grid and, thus, being more reliable at the micro scale (white arrow in figure 6.13.B). Moreover, below 2500 m.a.s.l., no permafrost was predicted on the torrential cones below the small rock walls located midslope (figure 6.13.A/e).

For the sector 6.13.A/c, PERMAL indicates a probability of permafrost occurrence of around 50%, with a PLL located at 2550-2600 m.a.s.l.. This limit is 150 metres higher than the elevation observed by Lambiel and Reynard (2001) for a North-East exposed sector.

Although the general trend indicates an increase in probabilities when there is an increase of in altitude, PERMAL was able to predict a permafrost distribution that presents some spatial patterns visible at the micro scale, especially in the talus slopes located on the East. This result was achieved by exploiting the curvature indices that illustrate the micro-topography. The slope is indeed characterized by an important torrential activity that remobilizes the sediments. Stable sectors with a coarse surface grain size are alternated with sectors presenting finer materials and an incised topography.

The uncertainty map shows that the most uncertain sector is the talus slope c (figure 6.14.A). Similarly to other talus slopes presented in the previous sectoral analysis, it seems that RF produces the greatest variabilities in talus with regular slopes and flat curvatures. The uncertainty plot presented in figure 5.35 unveiled indeed that the RF results are more uncertain when the curvature values are equal or close to 0.

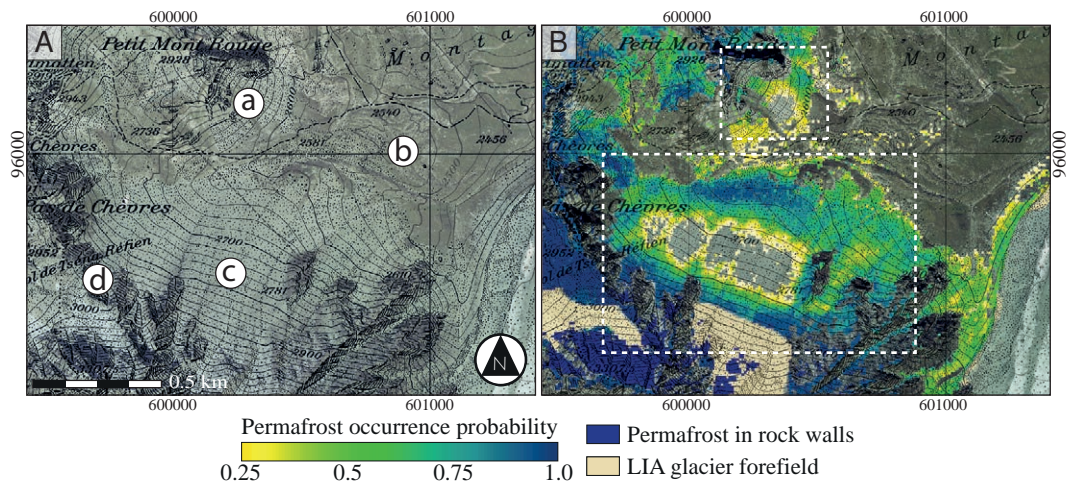
Moreover, the talus slopes located on the East slope of the Tournelon Blanc sector are good examples to illustrate that, unlike the empirical-statistical approach adopted by the BAFU map (figure 6.14.C), a data-driven approach does not make use of altitude thresholds for predicting the permafrost occurrence. In the BAFU map, the boundaries between the vegetation of the mineral surfaces are indeed not well respected and the bands resulting from the employed altitude thresholds are clearly visible, especially in the southern taluses. Compared to the PERMAL map, the potential permafrost distribution illustrated in the BAFU map is more extended. This may be linked to the lack of delimitation between the mineral surfaces and the vegetation.



**Fig. 6.14.:** (A) PERMAL map of the Tournelon Blanc glacier area; (B) Uncertainty map of the RF classification; (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005).

## Petit Mont Rouge and Tsena Réfien sector

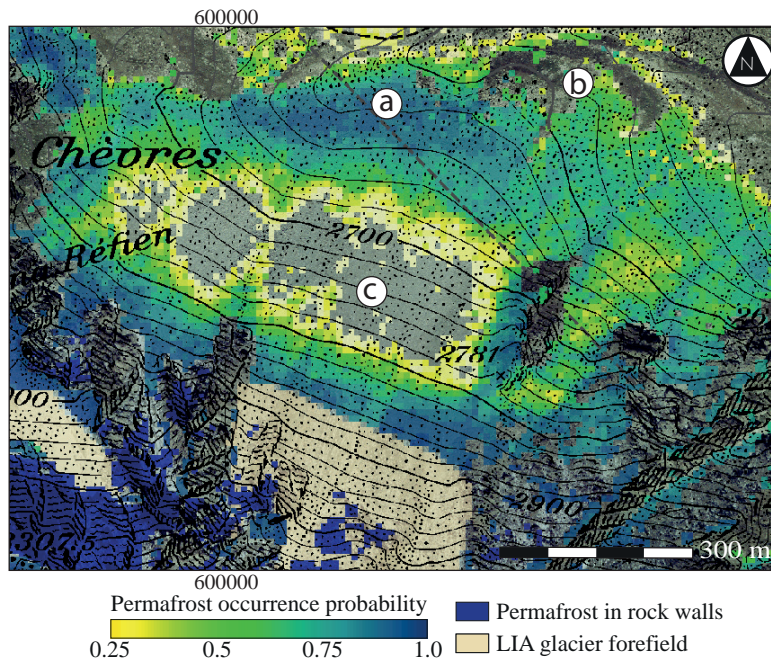
Located on the North of the Pigne d'Arolla peak, this sector presents various periglacial landforms (WGS84: 46.01360, 7.44348; CH1903: 600'375, 95'783; figure 6.15.A/a, b & c). Permafrost has been studied since the mid-1990s in the area and a recent focus turned on its distribution and related processes within the Petit Mont Rouge and Tsena Réfien talus slopes (e.g. Gardaz, 1997; Gardaz, 1998; Lambiel, 2006; Scapozza, 2013; Scapozza, 2015b; see chapter 3.1 for more details).



**Fig. 6.15.:** (A) The Petit Mont Rouge and Tsena Réfien area: (a) the Petit Mont Rouge talus slope, with the presence of a small protalus rampart (b) relict rock glacier (c) the Tsena Réfien talus slope (d) rock walls; (B) Potential permafrost occurrence map for the Petit Mont Rouge and Tsena Réfien sector, based on the RF classification, and delimitation of the sectors presented in figures 6.16 and 6.17.

PERMAL indicates that the rock walls located nearby the Col de Tsena Réfien (6.15.d) are affected by permanent negative temperatures, while the steepest slopes of the Petit Mont Rouge peak are generally permafrost free (6.15.a). Moreover, the map shows a probability of permafrost occurrence in talus slopes that is generally higher at the foot of the landforms, while it is lower in its proximal part. This trend was well mapped in the Tsena Réfien talus slope, where the predicted presence of permafrost is particularly important in the mid-lower part of the landform (figure 6.16), while the probabilities of occurrence decrease in the mid-upper part. These results are particularly in accordance with field data. Indeed, Lambiel (2006) reported the existence of a permafrost body in the distal part of the Tsena Réfien talus slope (a in figure 6.16) and the absence of permafrost conditions midslope (6.16.c). The flat curvature indices influenced the uncertainty of the RF classification. This is similar to what was unveiled by the uncertainty maps presented for other sectoral analyses. RF produced a less certain map at locations occupied by taluses with regular and flat slopes (no convexity nor concavity) (figure 6.18.B). The reason behind this



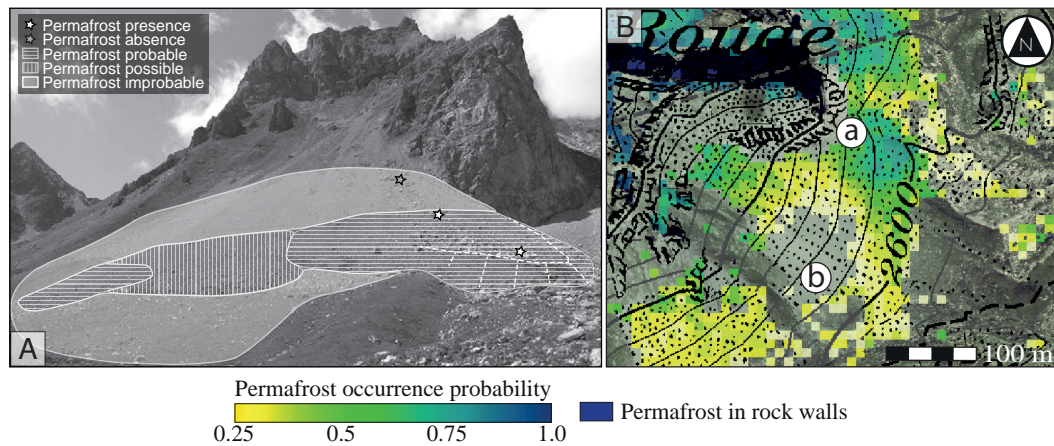


**Fig. 6.16.:** The Tsena Réfien talus slope with (a) a permafrost body identified by Lambiel (2006), (a) two vegetated rock glacier fronts and (c) the mid-upper part of the landform where permafrost is absent.

strong variability in prediction is probably linked to the lower number of permafrost evidences located in the talus slopes than in rock glaciers.

The delimitation of the validity domain excluded the vegetated fronts of the two rock glaciers located at the foot of the Tsena Réfien talus slope (*b* in figure 6.16). Lambiel (2006) reported however that the two landforms were characterized by frozen sediments. PERMAL thus correctly predicted the permafrost presence for the two lobes (except for the fronts that were not considered), despite moderate probabilities of permafrost occurrence compared to the sector located on the West (6.16.a).

In the Petit Mont Rouge talus slope located on the North (figure 6.15.A/a), permafrost presence was assessed by Scapozza (2013) and Scapozza (2015a) thanks to borehole thermal data and geoelectrical surveys. The air mechanism occurring there controls the distribution of permafrost, which is present at the top of the slope and in its distal part, where a protalus developed (figure 6.17.A). PERMAL predicted permafrost occurrence for this site and it revealed being particularly reliable with the field investigations (figure 6.17.B). The protalus rampart was in fact correctly mapped with higher probabilities predicted at this location (*a* in figure 6.17.B). Above the protalus roots, PERMAL indicates a decrease in permafrost occurrence, as it was observed with temperature measurements. The upper part of the talus, located above 2700 m.a.s.l., was predicted by PERMAL as potentially frozen. The lowest southern part of the landform was in turn mapped with permafrost absence (*b*

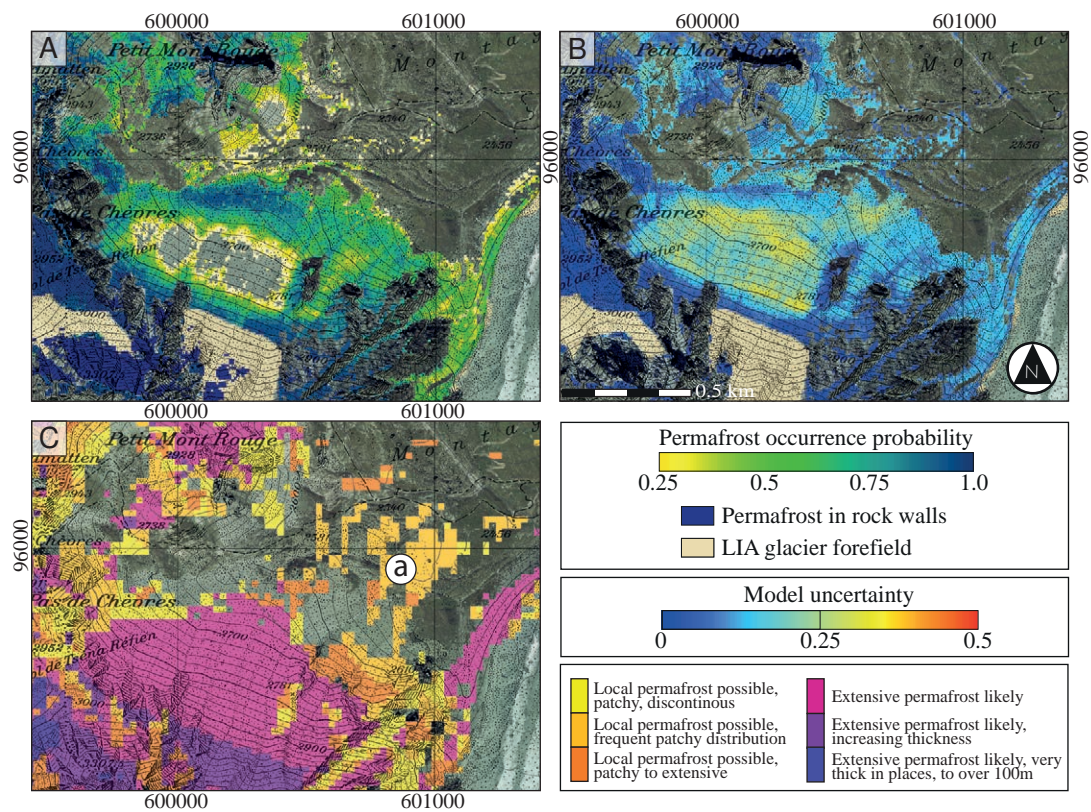


**Fig. 6.17.:** (A) Qualitative indication of the spatial extent of permafrost occurrence of the Petit Mont Rouge talus slope (reproduced from Scapozza, 2013); (B) PERMAL probabilities of permafrost occurrence and localization of (a) the protalus rampart and (b) the portion of the talus slope where permafrost absence was assessed with field data and observations (Scapozza, 2013).

in figure 6.17.B). ERT measurements showed that permafrost is effectively absent in this part of the slope (Scapozza, 2013). The fine grain size probably helped PERMAL obtaining this low prediction of permafrost occurrence.

A relict rock glacier is located on the East of the Petit Mont Rouge sector. PERMAL predicted it as being permafrost free. This result is in contrast with the BAFU map that predicted a local possible occurrence of permafrost (a in figure 6.18.C).

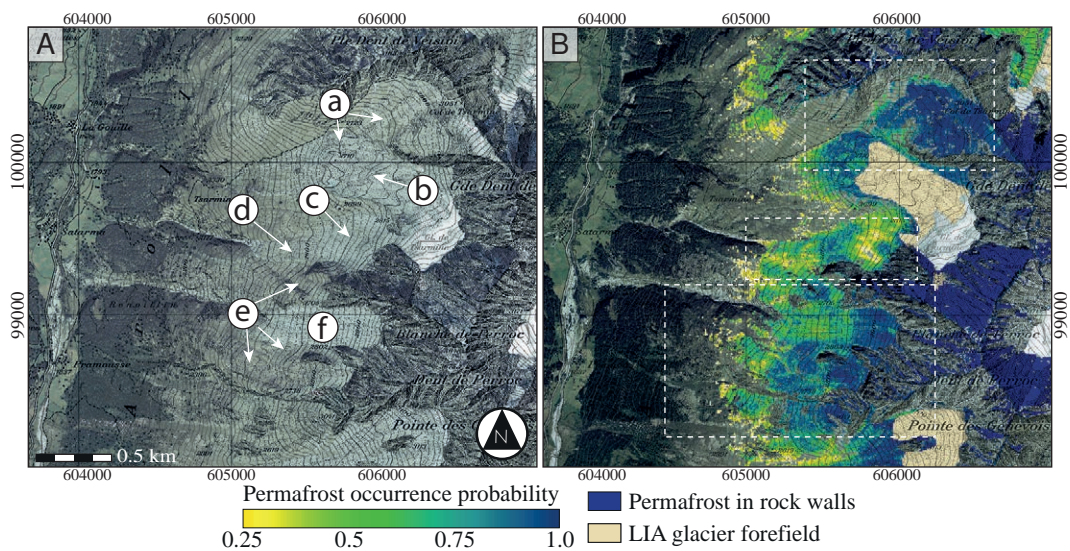
By comparing the two maps (6.18.A and 6.18.C), BAFU tends to be more optimistic than PERMAL, especially in talus slopes. The RF classification seems to have produced a more reliable result at the micro scale thanks to the finer resolution and to the delimitation of the validity domain that excluded vegetated areas. Moreover, the BAFU map resulted from an empirical-statistical approach that did not consider many local conditions, such as the grain size or the curvature indices. For this reason, the result is less prone to identify the strong discontinuity of the mountain permafrost because it does not consider environmental predictors as being proper to the micro-scale. PERMAL, in turn, seems to better translate the field observations by learning them from the permafrost evidences that were provided to the RF algorithm.



**Fig. 6.18.:** (A) PERMAL map of the Tsena Réfien-Petit Mont Rouge area; (B) Uncertainty map of the RF classification; (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005).

## The Veisivi ridge

This sector is located on the East of the Satarma village (WGS84: 46.04622, 7.51288; CH1903: 605'747, 99'412) and presents steep slopes all across the mountain side (figure 6.19.A). An important portion of the rock walls are affected by permafrost, especially those exposed to the North. The last observed event occurred in September 2015, with a rock fall of around 80'000 m<sup>3</sup> that triggered in the West face of the Grande Dent de Veisivi (above *b* in figure 6.19.A). The high elevation of the Veisivi ridge leads to a low MAAT and, thus, to the presence and maintenance of permafrost in debris-covered surfaces, as well as the presence of several intact rock glaciers (*a*, *d* and *e* in figure 6.19.A).

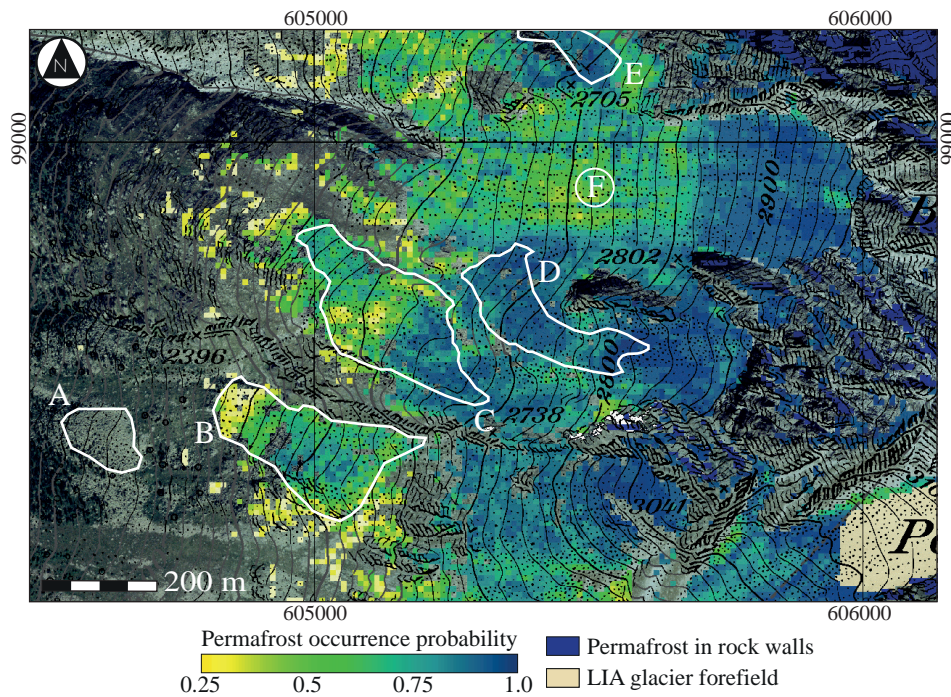


**Fig. 6.19.:** (A) The Veisivi ridge: (a) rock glaciers (b) Tсарmine debris-covered glacier (c) Tсарmine talus slope with protalus rampart (d) Tсарmine rock glacier (e) rock glaciers (f) Perroc talus slope; (B) Potential permafrost occurrence map for the Veisivi area, based on the RF classification.

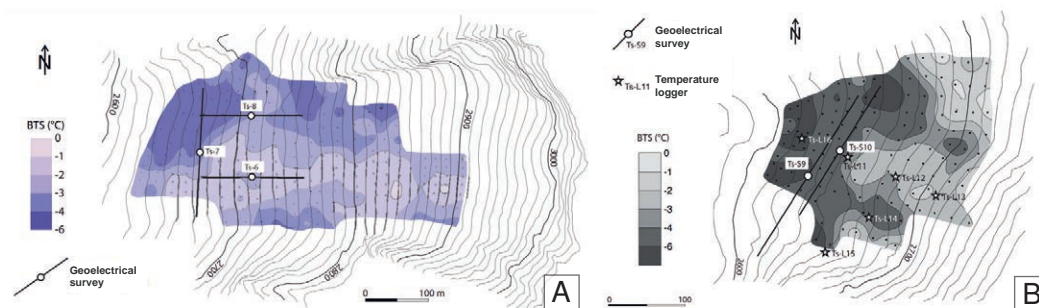
The potential permafrost extent obtained with PERMAL illustrates high probabilities of occurrence in debris-covered surfaces from 2600 m.a.s.l.. The permafrost extent of the Perroc talus slope (*f* in figure 6.19.A) presents some spatial patterns. Above around 2800 m.a.s.l., RF predicted continuous permafrost. The lower part of the talus seems to be less affected by permafrost conditions, especially in its centre (*F* in figure 6.20). On the contrary, the upper part of the landform was mapped with higher probabilities of occurrence. Lambiel (2006) investigated the talus slope with thermal and electrical surveys and he evidenced that permafrost conditions effectively exist at these locations (see figure 6.21.A). PERMAL probably exploited the surface grain size that differs between the lowest part of the talus (coarse granulometry) and the upper part, where fine materials are visible.

Moreover, high probabilities were assigned to some rock glaciers that were mapped in the rock glacier inventories. The PLL recognized by PERMAL allowed excluding

the relict rock glacier A (figure 6.20), while the other four active rock glacier surfaces were mapped as potentially frozen (B, C, D and E in figure 6.20).



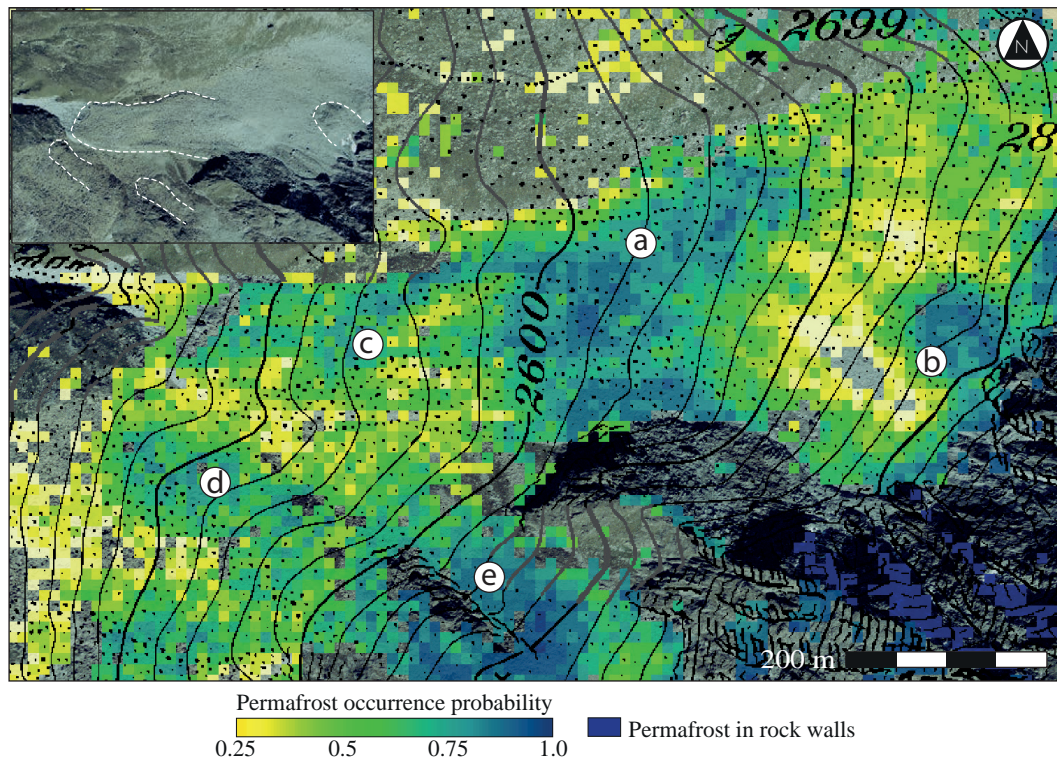
**Fig. 6.20.:** PERMAL map for the Perroc talus slope sector (F) and localization of the interpolation map of BTS measurements presented in figure 6.21 (dashed lines), of four active rock glaciers (B, C, D & E) and a relict rock glacier (A).



**Fig. 6.21.:** Interpolation map of BTS measurements of (A) the Perroc talus slope and (B) the Tsarmine talus slope (reproduced from Lambiel, 2006).

On the North, the Tsarmine talus slope presents a discontinuous permafrost distribution, with the highest probabilities mapped at the distal part of the landform (*a* in figure 6.22). A decrease in probability with the increase of the altitude can be observed, with the exception of the upper sector where higher probabilities were attributed to a small protalus rampart (*b* in figure 6.22). This result was achieved by RF by considering the difference in the surface granulometry between the two portions of the talus. Indeed, the lower part of the landform is characterized by coarse debris, while the mid-upper part presents a finer grain size. The result is in

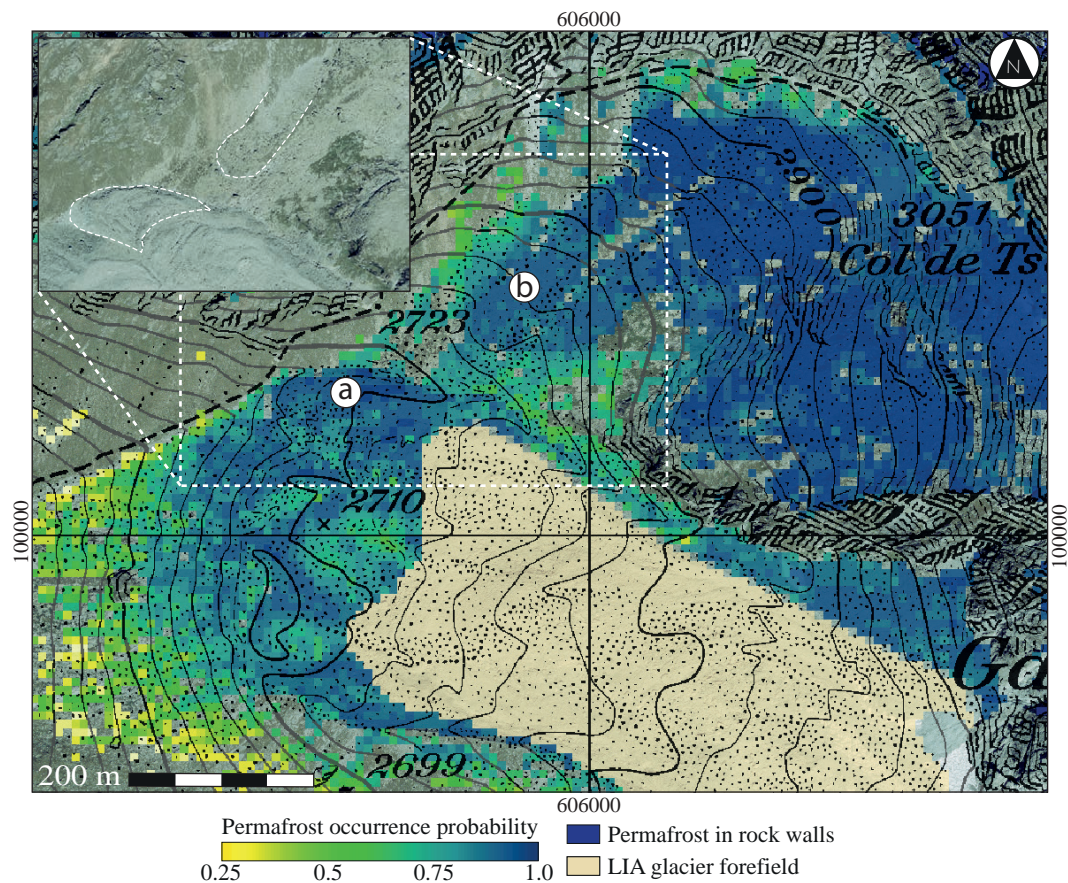
accordance with the thermal and the geoelectrical surveys carried out by Lambiel (2006), evidencing favourable permafrost conditions in almost the same sectors than the ones identified by PERMAL (figure 6.21.B).



**Fig. 6.22.:** PERMAL probabilities of occurrence for the Tsarmine talus slope sector (a), a small protalus rampart (b), the Tsarmine rock glacier (c) and another active rock glaciers (d). Dashed lines indicates the location of the interpolation map of BTS measurements presented in figure 6.21. The limits of the protalus rampart and the two rock glaciers are reported on the orthophoto (top-left image).

The Tsarmine rock glacier (c in figure 6.22) have been extensively studied in recent years (e.g. Lambiel et al., 2008; Delaloye et al., 2010; Barboux et al., 2014b; PERMOS, 2016; Kummert et al., 2017) and its activity state indicates the presence of permafrost. However, the RF permafrost map only predicted a probability of occurrence slightly higher than 50% for this landform. This can be explained by the low altitude reached by the rock glacier (less than 2500 m.a.s.l.). According to the empirical model presented by Lambiel and Reynard (2003), this altitude is at the very limit of the regional PLL for the West-exposed slopes. As a consequence, PERMAL might have under-estimated the probability of permafrost occurrence for the Tsarmine rock glacier. The employment of other predictors, such as the grain size (that illustrates the coarse blocks covering the rock glacier), the NDVI and the curvature indices, were in turn useful to still produce a reliable result. The same consideration can be made for the other rock glacier located South of the Tsarmine rock glacier (d in figure 6.22), whose probabilities of occurrence are just slightly greater than 50-60%.

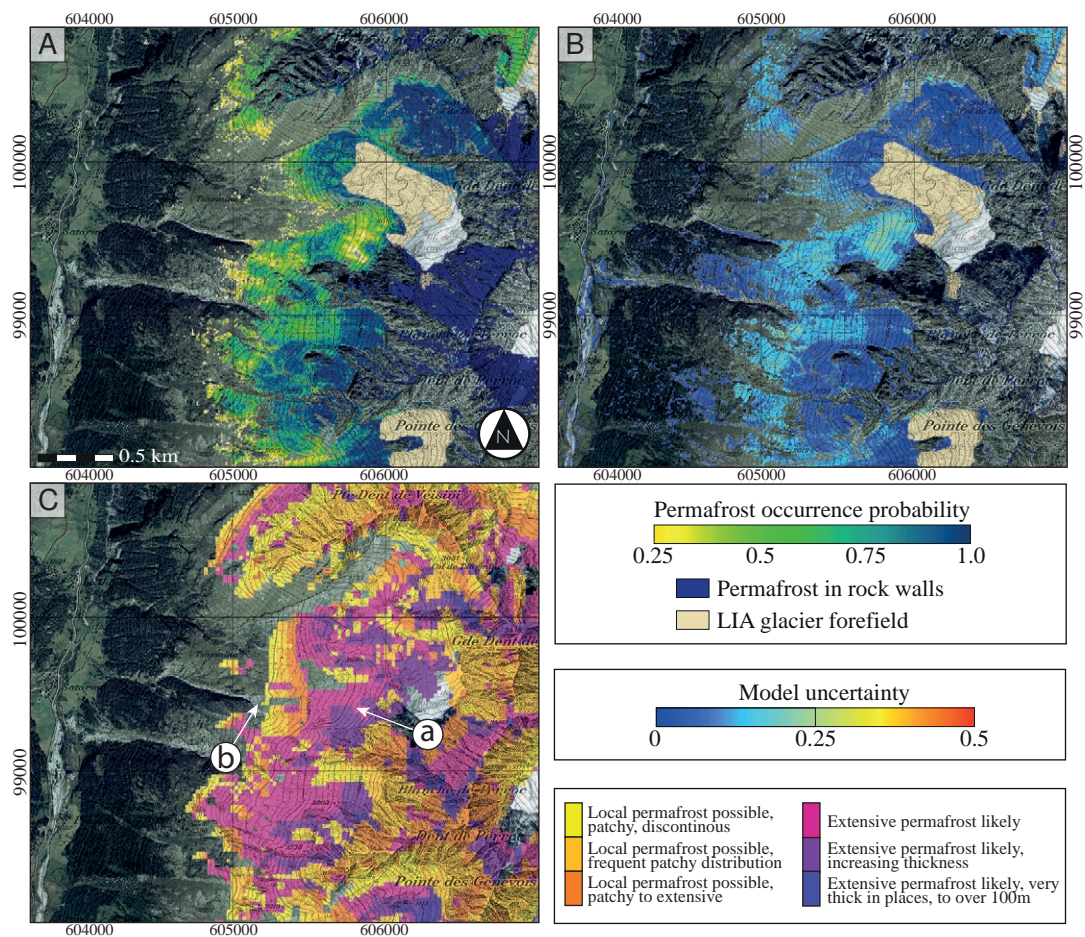
The Veisivi ridge is also characterized by the presence of the Tsarmine debris-covered glacier and recently studied by Fischer et al. (2014), Capt et al. (2016), and Bosson and Lambiel (2016). A marginal rock glacier (*a* in figure 6.23) developed at the northern distal part of the system and is characterized by an internal structure of ice-cored or ice-cemented debris. PERMAL was able to detect favourable permafrost conditions at this location with strong mapped probabilities of occurrence. Similarly, an active rock glacier located on the slopes above the Tsarmine debris-covered glacier, as well as the debris-covered surfaces situated on the East, were also identified as potentially frozen (*b* in figure 6.23).



**Fig. 6.23.:** PERMAL probabilities of occurrence for the Tsarmine debris-covered glacier, with a marginal rock glacier (*a*) that develops from the glacier system and an active rock glacier located on the northern slope. The limits of the rock glaciers are reported on the orthophoto (top-left image).

Similarly to the examples presented for the other sectoral analyses, the permafrost extent obtained with the RF algorithm is less optimistic than the BAFU model. The vegetation/mineral surface boundary was better respected (figure 6.24.A) and no altitude thresholds were used to achieve the presented result. The higher resolution of the obtained map also allowed being more accurate when investigating the occurrence of permafrost at the scale of a landform.

Moreover, the permafrost distribution represented by the two maps is slightly in discordance. The Tsarmine talus slope, for example, was mapped by BAFU as totally covered by permafrost (*a* in figure 6.24.C). Conversely, PERMAL mapped a permafrost distribution more similar to the field observations, namely with lower probabilities of occurrence in the mid-upper part of the Tsarmine talus slope. In addition, a portion of the Tsarmine rock glacier surface was predicted as permafrost free by BAFU (*b* in figure 6.24.C). Despite probable under-estimated probabilities of occurrence, PERMAL was able to map the permafrost presence for the entire rock glacier surface. In conclusion, the comparison between the two maps unravels that the result obtained by PERMAL is more in accordance with field data and observations, with the mapping of probability patterns that better reflected the high discontinuity of the permafrost distribution.

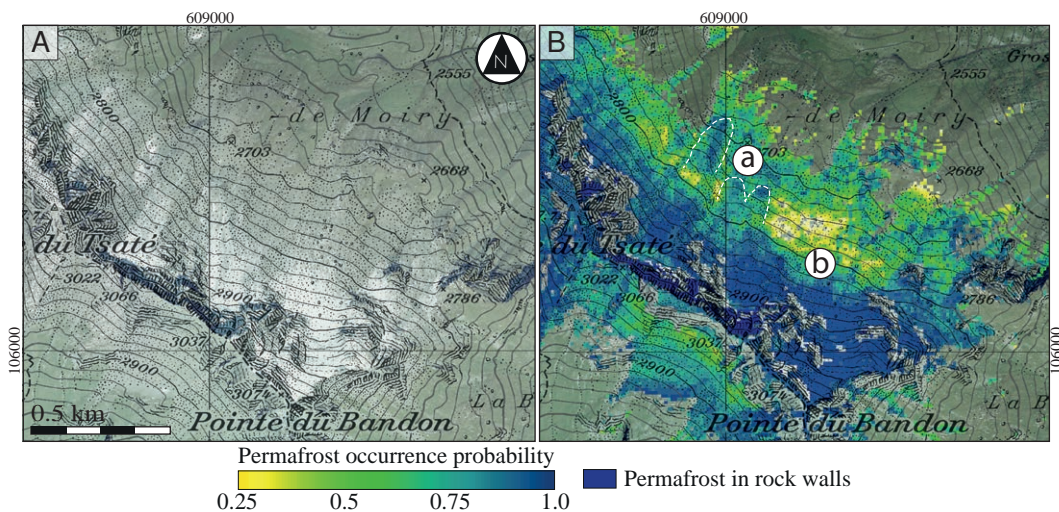


**Fig. 6.24.:** (A) PERMAL map of the Veisivi ridge; (B) Uncertainty map of the RF classification; (C) Potential permafrost extent offered by the BAFU map (BAFU, 2005) and localization of the Tsarmine talus slope (a) and the Tsarmine rock glacier (b).



## Tsaté-Moiry sector

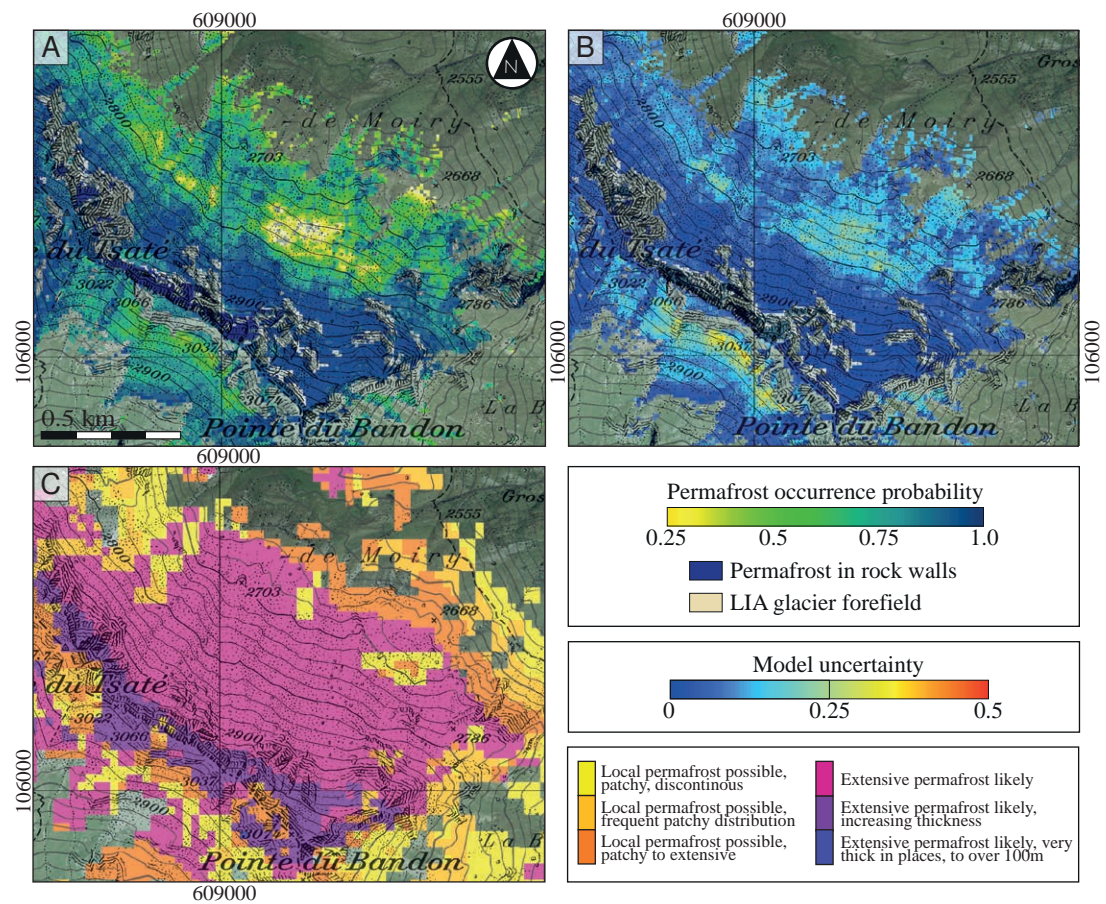
The Tsaté-Moiry sector (WGS84: 46.10885, 7.55554; CH1903: 609'039, 106'379) is located on the South-West of the Moiry dam and on the North of the Pointe de Tsaté peak (3077 m.a.s.l.). Its sectoral analysis is particularly interesting due to the presence of debris-covered surfaces that are potentially frozen despite the fine lithology of the area. Two intact rock glaciers developed from the talus slopes located on the Eastern side of the Pointe de Tsaté (*a* in figure 6.25.B), as well as some solifluxion lobes particularly developed South-East of the rock glaciers (*b* in figure 6.25.B). Lambiel (2006) and Lambiel (2011) investigated the structure and the dynamics of the two rock glaciers, while Scapozza (2013) extended the investigation by also taking into account the talus slopes located nearby these landforms, assessing both the occurrence of permafrost in fine debris and the main properties that allow its existence.



**Fig. 6.25.:** (A) Presentation of the Tsaté-Moiry sector; (B) Potential permafrost occurrence map for the Tsaté-Moiry area, based on the RF classification, and delimitation of two intact rock glaciers.

PERMAL identified and predicted high probabilities of occurrence at the location where the two rock glaciers exist (*a* in figure 6.25.B). However, at the foot of the talus slope, the permafrost occurrence is probably over-estimated. Indeed, according to the geoelectrical surveys carried out in the area (Scapozza and Lambiel, 2013), permafrost is extremely patchy outside the rock glacier limits because of the fine lithology. On the eastern part of the talus, permafrost is discontinuous and it is only restricted to the mid-upper part of the slope and to parts of the solifluxion lobes (*b* in figure 6.25.B). As a consequence, the prediction made by PERMAL in the central part of the slope, that indicates the permafrost absence, revealed to be correct. According to the field observations of Scapozza and Lambiel (2013), the sectors with high probabilities of occurrence, located in the upper part of the landform, also correspond to the sectors where permafrost was found.

The uncertainty map presented in figure 6.26.B illustrates that the mid-lower part of the talus slope was predicted with the highest uncertainties. The RF result must thus be carefully evaluated. In comparison to the BAFU map (figure 6.25.C), PERMAL is much less extended and the probabilities are much lower although this data-driven technique is over-predicting the overall permafrost occurrence. This example clearly shows the differences between two different modelling approaches and resolutions. BAFU shows a high occurrence of permafrost within the entire talus slope because of the coarser resolution and the use of Haeberli (1996)'s rules. In comparison, a finer resolution, the consideration of the grain size, as well as the small scale curvature indices able to detect small lobes, made the produced PERMAL permafrost map more reliable to the micro scale.



**Fig. 6.26.:** (A) Potential permafrost occurrence map for the Tsaté-Moiry area, based on RF classification; (C) RF model uncertainty map; (D) Potential permafrost extent (BAFU, 2005).

## 6.3 Advantages and limitations of the employed data-driven approach

The presented data-driven modelling procedure revealed to be reliable for mountain permafrost distribution modelling. As it was illustrated in the sectoral analyses, some advantages can be distinguished from traditional empirical or empirico-statistical approaches, such as the ones used in PERMAKART (Keller, 1992), PERMAMAP (Hoelzle and Haeberli, 1995), PERMACLIM (Guglielmin et al., 2003) or other simple regional models (i.e. Delaloye and Morand, 1997; Lambiel and Reynard, 2001; Baroni et al., 2004; Seppi et al., 2005).

On the one hand, the procedure developed in this research employed stochastic techniques and, thus, the final map was not produced with a deterministic approach. This means that the mapped permafrost distribution did not result from a fixed set of parameters, like for example some pre-determined altitude thresholds or BTS temperature thresholds that are employed to abruptly represent the permafrost extent in categories (possible and probable are the most common classes). The probabilistic output of the modelling procedure of this thesis provided instead a more transient and realistic visualization of the potential permafrost distribution and the avoidance of maps characterized by coloured bands (as it was shown in the BAFU, 2005 examples). An improvement in this direction was proposed with PERMAKART 3.0 published by Schrott et al. (2012), who adapted the PERMAKART model by transforming the traditional permafrost classes into probabilities, with a consequent enhancement of the map readability.

Conversely to mentioned traditional models, with the proposed machine learning approach, PLL are automatically learnt from data without the need of defining *a priori* rules that discriminate the permafrost presence from the permafrost absence. Data themselves provide such information to the classification algorithms if they are well collected and gathered in a representative dataset. It is worth mentioning that both approaches are not able to predict, without representative data, sporadic permafrost below the theoretical threshold of the PLL, such as turf-banked periglacial solifluction lobes or periglacial processes and landforms in vegetated areas.

Another advantage of a data-driven approach is that there is no need of supplementary test data, because the model quality assessment is part of the modelling procedure. Since hundreds or even thousands of permafrost evidences build the dataset, the split of the data in training, validation and test sets (with cross-validation) can easily be performed without losing useful information. Conversely, the mentioned traditional models are often built with few tens/hundreds of permafrost evidences

and thus required to be validated with additional field data or with the use of part of the observations, resulting in a loss of information (if not cross-validated).

More advanced statistical models, such as those published by Lewkowicz and Ednie (2004), Brenning and Trombotto (2006), Brenning and Azócar (2010), Bonnaventure et al. (2012), Boeckli et al. (2012), Azócar et al. (2016), Sattler et al. (2016), Angillieri (2017) or Alm (2017), overcame the mentioned limitations of the traditional approaches. However, they only usually employ few predictors, such as for example MAAT, altitude or PISR. This is often enough to map the permafrost occurrence at the regional or at the local scale, but it is insufficient for finer scales.

The employment of a larger number of environmental variables represents the originality of the data-driven approach proposed in this thesis. Indeed, as confirmed by the reliable maps presented in section 6.2, by integrating morphological predictors that represent the ground characteristics (curvature indices, NDVI and surface grain size), a correct representation of the potential permafrost extent at the micro scale can be obtained. Classification algorithms, belonging to statistics or machine learning, can thus be employed to produce permafrost maps at this scale, if they are able to consider the non-linearity of the permafrost data. As it was shown in section 5.4, RBF-SVM, MLP and RF were prone to consider the contribution of less influencing predictors other than just the altitude or the PISR.

Therefore, the presented approach also revealed to have some limitations. Since the inputs-output relationship was directly learnt from data, data-driven outcomes are often less interpretable than deterministic ones. The classifications vary indeed according to the data provided to the algorithm. In addition, the creation of a validity domain allows producing permafrost maps that are restricted to mineral surfaces. However, its delimitation depends on the quality of orthophotos, of primary surface maps and of FCIR images used to detect various types of landcover. Not disposing of good quality data or data with a low spatial resolution would result in areas where the permafrost occurrence would not be predicted. Moreover, the choice of excluding glacier forefields from the validity domain translated in large sectors that were not considered by the PERMAL approach, although they were located above the PLL and they were potentially affected by permafrost. This limitation demands a geomorphological interpretation of the map that can be hard to be made by non-expert readers.

Uncertainties must also be mentioned and distinguished into two categories, namely *uncertainties in data* and *modelling uncertainties*. Most of the permafrost evidences were obtained from rock glacier inventories, as these landforms are considered the most evident expression of the permafrost presence or absence. Except for InSAR-detected movements and very few geophysical investigation data, an accurate

validation of the ground truth was impossible without *in-situ* investigations. For this reason, polygons totally representing the presence or the absence of permafrost were used, without considering the differences that may exist at the *landform scale* (fast-moving, slow-moving, stagnant portion of the rock glacier, etc.). This procedure indirectly injected some bias in the data, since they were collected based on the own expert knowledge and the interpretation of field data and observations.

PERMAL produced modelling uncertainties for prediction performed for landforms other than rock glaciers, since only a small portion of the training examples corresponded to investigated talus slopes or other periglacial landforms. An increasing number of permafrost evidences outside rock glaciers or the use of over-sampling techniques (which also add a bias in data) would have been required in order to reduce the modelling uncertainty.

Another limitation of the proposed methodology was the exclusion of the steep bedrock areas ( $>40^\circ$ ) from the validity domain due to the lack of permafrost evidences that could be used to train the model. Similarly, very few permafrost evidences were available for training the model for bedrock outcrops with slopes below  $40^\circ$  (some GST data only exist). Since a data-driven approach takes advantage of the employment of hundreds (or thousands) of training examples, producing a permafrost map for these landforms was not possible. PERMAL also indirectly predicted lower probabilities of permafrost occurrence in mineral surfaces thanks to the low grain size values attributed to this type of landcover. Even though coarse debris surfaces are generally more favourable for permafrost conditions than bedrock outcrops located at the same altitude, at high elevations ( $>2800$  m.a.s.l.) flat bedrock outcrops are generally also frozen. Improvements for a better distinction between bedrock surfaces ( $<40^\circ$ ) and debris areas are thus required to obtain a more accurate map.

## 6.4 Significance of permafrost maps and their usability

In this research, the term of *potential permafrost distribution map* was used to refer to the maps obtained by adopted machine learning techniques. This concept has often been employed by researchers working in the field of permafrost modelling with a general consensus, accepting that permafrost spatial modelling produces permafrost maps. As mentioned in Chapter 2, permafrost is however a thermal and invisible phenomenon if no geomorphological evidences attest its presence. As a consequence, most of the existing permafrost maps are only an abstraction of the field observations or the morphological evidences that appear consequently to the presence of this phenomenon. Permafrost maps should thus be considered more as some helpful tools indicating that the terrain meets the requirements of being

potentially frozen, namely that the specific sector of interest has proper topo-climatic characteristics allowing the maintenance of negative temperatures in the ground, rather than maps indicating the effective permafrost occurrence. This would indeed avoid that non-expert map readers make a misleading interpretation of the map.

For the APIM model, Boeckli et al. (2011) proposed an interpretation key that allows an easier reading of their map and that can also be adopted for the interpretation of PERMAL. According to the characteristics of the sector someone is interested in, Boeckli et al. (2011) suggested some considerations that are required to interpret the reliability of the map in comparison to the ground truth. Four land-cover surfaces were taken into account by the APIM interpretation key: *talus slopes*, *partially vegetated slopes*, *rock glaciers* and *rock walls*.

In *talus slopes*, Boeckli et al. (2011) indicate that different thermal conditions may exist according to the debris size, the possible existence of an air convection/advection mechanism or the presence of a thick snow layer. Often, the distal part of the landform has the coldest MAGST in reason of the presence of coarser materials and a layer of snow that last longer in late spring/early summer, brought downstream by avalanches. Permanent snow patches may also be indicative of permafrost conditions. Since the upstream part is generally composed of more fine materials, it is often affected by warmer thermal conditions, which may translate in the absence of permafrost.

In *partially vegetated slopes*, the size of the materials, the soil properties and the abundance of vegetation are the main factors controlling the presence or the absence of permafrost. Indeed, porous terrains not filled with fine materials may be indicative of a negative thermal state of the ground, while the absence of permafrost is linked to the existence of fine sediments or coarse blocks saturated by fine materials. For these reasons, the interpretation key reminds that vegetated or partially vegetated surfaces are also generally indicative of the absence of permafrost.

*Intact rock glacier* are witnesses of the presence of permafrost when a steep front and deformation wrinkles can be observed. In turn, Boeckli et al. (2011) indicate that the assessment of the thermal conditions in the neighbourhood of these landforms is difficult without additional investigations.

Finally, the micro-topography of a *rock wall*, particularly related to rock fracturing, makes the permafrost distribution in this type of periglacial environment very discontinuous. A very heterogeneous surface often allows the deposit of a snow layer on fractured areas, favouring cold thermal conditions. The vertical rock walls are, on the contrary, often indicative of a warmer thermal state, especially at low elevations. The assessment of the presence of permafrost in such environments requires additional *in situ* observations, for example in relation to the exposure of the study site in the respect of the Sun.

In PERMAL, the LIA glacier forefields were excluded because considered as an indicator of the permafrost absence (see section 2.2.1). However, it is worth adding these alpine environments to the Boeckli et al. (2011)'s key of interpretation. Indeed, in the lateral margins of recently exposed proglacial areas, permafrost is generally more likely present than in their centre, with the potential existence of frozen sediments in the moraine. When the ice-debris mixture thickness and the slope are sufficient, periglacial creep landforms (such as protalus rampart or small marginal rock glaciers) could also develop. As a consequence, if the produced permafrost map indicates high probabilities of occurrence in the neighbourhood of a periglacial forefield, it is possible that permafrost is also present in its lateral margins.

In synthesis, the predictions of the permafrost occurrence inferred with the data-driven approach adopted in this research must be interpreted as the probability that a specific location meets all the topo-climatic requirements, allowing the ground being permanently frozen, rather than indicating a certain probability of finding permanent negative temperatures in the ground. The map is thus indicative of the *potential* permafrost presence or absence and therefore must be considered as a *permafrost favorability index* (as proposed by Azócar et al., 2016) instead of a tool showing the exact ground truth. Accordingly, field observations and expert interpretations must be coupled with PERMAL in order to assess the effective permafrost occurrence.

## 6.5 Synthesis

In this study, a comprehensive analysis of the available data concerning the permafrost phenomenon was conducted. A collection of permafrost evidences was gathered from different sources (see chapter 3) in order to build a complete and exhaustive dataset of the known permafrost presence and absence occurring in the study area. Permafrost occurrences were associated with environmental predictors proper of the sectors where the evidences were collected. The retained dataset not only informed about the localization of the sectors where known permafrost absence or presence was observed, but also about the climatic parameters (such as the MAAT and the PISR), topographic/morphological characteristics (such as slope, aspect and curvature indices) and the ground properties (terrain grain size, NDVI).

The first part of this research unravelled that, despite the presence of underlying and not well expressed structures, permafrost data are complex. It confirms that the permafrost phenomenon is non-linear and follows the hierarchical dominance of environmental variables, which become more important with the decrease of the geographical scales. The statistical identification of the predictor relevance, presented in section 5.2.2, illustrated that the altitude and the MAAT are two predominant variables controlling the permafrost occurrence. However, other features

also revealed to be relevant for the permafrost phenomenon. Ground characteristics, expressed by the surface grain size and the NDVI variables, as well as large curvature indices, were indeed identified for their relevance and their contribution in terms of feature importance. This result confirmed that, at lower scales than the regional scale, the decrease of the MAAT with the altitude alone (and the PISR variations according to the topography) are insufficient to explain the permafrost occurrence, especially at the micro scale.

The analysis of the complexity of permafrost data and the relevance of retained predictors was followed by the application of the machine learning classifiers. Five algorithms were used: Logistic regression (LR), Artificial neural networks (ANN), linear and non-linear Support Vector Machines (SVM) and Random forests (RF). These algorithms were selected because they belong to specific sub-domains of machine learning: Logistic regression is a linear parametric classifier and it was used as a benchmark classifier to be employed before using more complex classification algorithms. ANN are supervised learning algorithms that can learn a non-linear functions (Multilayer Perceptron, MLP, in the context of this study). They differ from Logistic regression because they have one or more non-linear layers between the input and the output layers (called hidden layers). Support Vector Machines were selected because they represent a non-parametric learning algorithm and they are members of the so-called kernel methods. Lastly, Random forests are an ensemble learning method based on bootstrap aggregating.

With these machine learning algorithms, classification models were inferred without recurring to variable value thresholds or fixed and user-pre-determined rules. Indeed, they learned the statistical relationship between the training permafrost evidences and the permafrost explanatory variables directly from data. Modelling performances, assessed with the AUROC curves, ranged between a minimum of 0.75 (for linear Support Vector Machines) to a maximum of 0.88 (for Random forests). These values generally indicate moderate to good classification performances. Random forests and Multilayer Perceptron were assessed as being significantly better than other tested data-driven methods, according to the results of a paired t-test that compared the modelling performances.

Each inferred model was used to predict the permafrost occurrence for sectors where the presence or absence of the phenomenon was unknown. Maps of the potential extent of mountain permafrost were obtained with their corresponding uncertainties. The latter informed about which range of values for each predictor contributed the most to the permafrost presence or absence. In addition, uncertainties also indicated in which areas of the region of interest the most uncertain results were obtained.



Logistic regression was successfully employed by Lewkowicz and Ednie (2004), Brenning and Trombotto (2006), Sattler et al. (2016), Angillieri (2017) and Alm (2017) to map the permafrost occurrence at the regional and local scales. In the present study, Logistic regression predicted a smooth map with a linear increase of the permafrost probability with the altitude, which is correct at the mentioned scales, but it does not reflect the strong spatial discontinuity of the phenomenon at the micro scale. The correct representation of the mountain permafrost distribution at this scale was indeed one of the main objectives of this research. A similar result was obtained with linear Support Vector Machines, with a mapping generally correct at the local scale, but not reliable for the description of the phenomenon at a smaller scale. It was shown that the altitude predictor had a too important weight to allow other retained variables having a stronger influence on the classification, which explains the strong tendency of the increase in permafrost probabilities with the increase in elevation. With Radial Basis Function Support Vector Machines, Multilayer Perceptron and Random forests, the potential permafrost extent appeared to be more reliable to the field observations and to better respect the high discontinuity of mountain permafrost. Among the three non-linear ML algorithms, the permafrost discontinuity was best reproduced with Random forests, as shown in the sectoral analyses presented in this chapter. The occurrence of permafrost did not tend to increase linearly with altitude and it was indeed possible to observe higher permafrost probabilities in the lower half of some talus slopes, which is in agreement with field data. Results revealed thus a good reliability with the field observations and a good representation of the permafrost heterogeneity. The delimitation of the validity domain allowed the exclusion of sectors where the permafrost absence is well-known (such as the vegetation, lakes, etc.) and thus no prediction was necessary in these areas. The discrimination between mineral and vegetation surfaces prior to the application of data-driven techniques also offered a map of the permafrost extent that respected the boundaries between these two land cover types.

Compared to traditional modelling approaches, no altitude thresholds were employed with a data-driven approach to predict the permafrost occurrence. It was shown that, unlike the BAFU map, the automatic detection of the PLL, as well as the delimitation of the validity domain, permitted to obtain a map that was less optimistic and, thus, more in accordance with the field reality. In addition, a greater number of environmental predictors (than those used by empirical-statistical models), especially those related to the ground characteristics, like the NDVI and the surface grain size, improved the quality of permafrost map. Indeed, at the micro scale, lower probabilities were attributed to sectors where the granulometry appears to be fine, even for sites located where permafrost conditions potentially exist. Therefore, the importance of the grain size and the NDVI, evidenced with the feature relevance analysis, also had an influence on the prediction of permafrost occurrence and its reliability at the micro scale.

Nevertheless, the presented study showed some limitations. Firstly, machine learning results are less interpretable when trying to understand the combination of predictor values that lead to a particular result. As a reminder, data-driven techniques are non-deterministic methods and thus it is hard to explain why a particular result is obtained at a particular location. Therefore, uncertainty plots and uncertainty maps were used as supplementary tools trying to shed light on the internal mechanisms of the employed machine learning techniques. Personal expert judgement is however still required when trying to explain and to interpret the obtained results.

The results cannot definitely be employed alone like a trustworthy tool that assesses the permafrost occurrence without recurring to additional interrogations on the field (measurements, expert knowledge, interpretation key, etc.). This is due to the fact that permafrost is a thermal phenomenon and thus invisible if no morphological evidences attest its presence. However, the overall result of this data-driven approach is, at its present-state, satisfactory. Some improvements were indeed obtained in terms of representation of the strong discontinuity of the mountain permafrost, especially when mapping its extent at the micro scale.



## Conclusion and perspectives

This Ph.D. thesis provides new knowledge about the analysis, classification and mapping of mountain permafrost data. A new modelling procedure was elaborated to study existing permafrost evidences and to map the potential permafrost occurrence at the micro scale. In the next sections the final research considerations will be presented, as well as some recommendations for further modelling researches.

### 7.1 Main achievements

The application of machine learning to the field of permafrost modelling showed promising results, especially at the micro scale. The main objective of this thesis was the *systematic and detailed investigation of the potential of data-driven techniques applied to permafrost data* (see section 1.2). The final permafrost map presented in this research would not have been possible without the integration of several modelling steps. The first step involved an *exhaustive analysis of available permafrost evidences*. This study provided detailed knowledge on the complexity of permafrost data at the micro scale, shedding light on the best data-driven classifier to be employed in order to obtain reliable results. The analysis unveiled that permafrost data are not linear and that their classification consequently required the selection of some machine learning algorithms able to incorporate the non-linear relationship between the environmental predictors and the permafrost occurrence. This relationship was studied *without any "a priori" assumption on the importance* that retained environmental variables had for the permafrost presence or absence at the micro scale. The analysis confirmed the hierarchical influence of *climate, topography and ground conditions* on the permafrost occurrence, as described by several authors on the basis of their field observations (i.e. Haeberli, 1996; Gruber and Haeberli, 2009). Altitude and MAAT were identified as the most relevant predictors to the employed feature selection algorithms. However, other variables, such as surface grain size, terrain curvatures and NDVI, selected for their ability to represent the ground characteristics at the micro scale, revealed to be important and to have an effect on the strong discontinuity of the permafrost distribution. Accordingly, the dataset built for this study considered a greater number of predictors compared to one of traditional approaches (e.g. BAFU, 2005; Schrott et al., 2012; Boeckli et al., 2012; Azócar et al., 2016; Sattler et al., 2016), aiming at producing the most reliable potential permafrost extent at the micro scale. The second step of the modelling approach

involved the classification of the most informative permafrost data with five algorithms that belong to statistics and machine learning: Logistic regression (LR), linear and non-linear SVM, Multilayer perceptrons (MLP) and Random forests (RF). These techniques inferred a classification function from the labelled training data (pixels of permafrost absence and presence) to predict the permafrost occurrence where this was unknown. The classification performances were assessed with AUROC curves, obtaining values of 0.75 for the linear SVM, 0.77 for RBF-SVM, 0.84 for LR and MLP and 0.88 for RF. These values indicated good overall modelling performances. With a t-test analysis, the RF algorithm resulted to be the best performer. As machine learning is a non-deterministic approach, an overview of the model uncertainties was also offered by the data-driven approach. It informed about the location of the most uncertain sectors where further field investigations will be required to be conducted in order to improve the reliability of permafrost maps.

The last step of this thesis involved the *mapping of the permafrost occurrence at the micro scale*. The RF algorithm demonstrated to be efficient thanks to consistent results that were comparable to the field observations. The employment of environmental variables illustrating the micro-topography and the ground characteristics favoured the prediction of the permafrost distribution at the micro scale. These maps presented variations of the probability of permafrost occurrence within distances of few tens of metres. In some taluses, for example, a lower probability of occurrence in the mid-upper part of the slope was predicted. In addition, permafrost lower limits were automatically recognized from permafrost evidences. Lastly, the high resolution of the input dataset (10 metres) allowed elaborating maps at the micro scale with a modelled permafrost spatial distribution, which was less optimistic than traditional spatial models. Therefore, the permafrost prediction was computed without recurring to altitude thresholds (above which permafrost may be found) and the representation of the strong discontinuity of mountain permafrost at the micro scale was better respected.

## 7.2 Research perspectives

During the 2000s, the emergence of sub-decimeter DEMs and their corresponding extracted geomorphological indicators (i.e. aspect, slope, solar radiation, terrain curvatures, etc.) allowed new applications in geomorphology (e.g Cavalli et al., 2008; Kalbermatten et al., 2012; Heckmann et al., 2012). Aerial and satellite image acquisitions, LiDAR scans and drone flights are just few examples of sources that provide extremely accurate spatial data.

New data structures or phenomena can be studied thanks to the finer resolution of these sources of spatial information, opening the way to new challenges in data

analysis. The field of the mountain permafrost spatial modelling experienced the same trend, witnessing a significant increase in the knowledge on the distribution of this phenomenon within the last 15-20 years (see Gruber and Haeberli, 2009; Etzelmüller, 2013; PERMOS, 2016). The observed complexity of the mountain permafrost distribution, both in terms of spatial repartition and sensitivity to current warming, required however a clearer and more reliable representation on how the phenomenon is spatially distributed, especially at the micro scale.

The modelling procedure presented in this thesis proved to be adapted for the purpose, enhancing the modelling and the mapping of permafrost data at this scale. However, the limitations evidenced in section 6.3 required further improvements and additional field investigations in order to take into account all types of land-cover surfaces that exist within the periglacial environments. At the micro scale, the knowledge of the permafrost presence in glacier forefields is insufficient to be correctly integrated in this data-driven approach. Treating glacier forefields in the same way as other debris-covered surfaces would produce permafrost maps that overestimate the effective permafrost extent, without reflecting the observations proposed for these types of environments, which indicate the presence of the phenomenon only restricted to the lateral margins of recently deglaciated areas (i.e. Reynard et al., 2003; Kneisel, 2003; Käab and Kneisel, 2006; Ribolini et al., 2010; Bosson and Lambiel, 2016).

Similarly, known evidences of the permafrost occurrence in talus slopes are, at the present-state, just a small portion of the overall spatial knowledge on this phenomenon in debris surfaces, because rock glacier inventories represent most of the permafrost spatial data. PERMAL showed that the majority of the modelling uncertainties occurred in talus areas. As a consequence, additional efforts in gathering a complete dataset of the existing data are necessary. A new joint inventory of permafrost evidences outside rock glaciers could for instance be planned in coordination with different Universities and Research institutes, similarly to the Alpine Space project *Permafrost long-term monitoring network* (PermaNET) that proposed an alpine wide polygon-based rock glacier inventory (Cremonese et al., 2011). Gathering together field data and interpreting the permafrost extent of investigated areas could provide new permafrost evidences that would potentially improve the permafrost distribution models and their reliability, especially in talus slopes.

New environmental predictors might also be included in the permafrost dataset. Boeckli et al. (2012) used the mean annual precipitation (MAP) as a predictor that influences the permafrost presence. To consider this feature at the micro scale, a downscaling of existing spatial models would be required in order to account local variabilities in precipitation. A snow cover predictor could also be added to the dataset by employing glacio-hydrological models, such as for example the

SM-SOCONT model (Glacier and SnowMelt - SOil CONTRibution model; see Schaepli et al., 2005) or the ALPINE3D model (Raderschall et al., 2008; Bavay et al., 2009). As for MAP, the its spatial resolution should however reflect the micro scale to be effectively employed.

Lastly, advanced statistical models (e.g. Bonnaventure et al., 2012; Boeckli et al., 2012; Sattler et al., 2016; Angillieri, 2017) and studies using more advanced machine learning techniques (e.g. Azócar et al., 2016; Deluigi et al., 2017) showed that they are going in the right direction to obtain a clearer and more reliable representation of the heterogeneous distribution of mountain permafrost. In fact, greater improvements are expected in the next few years.



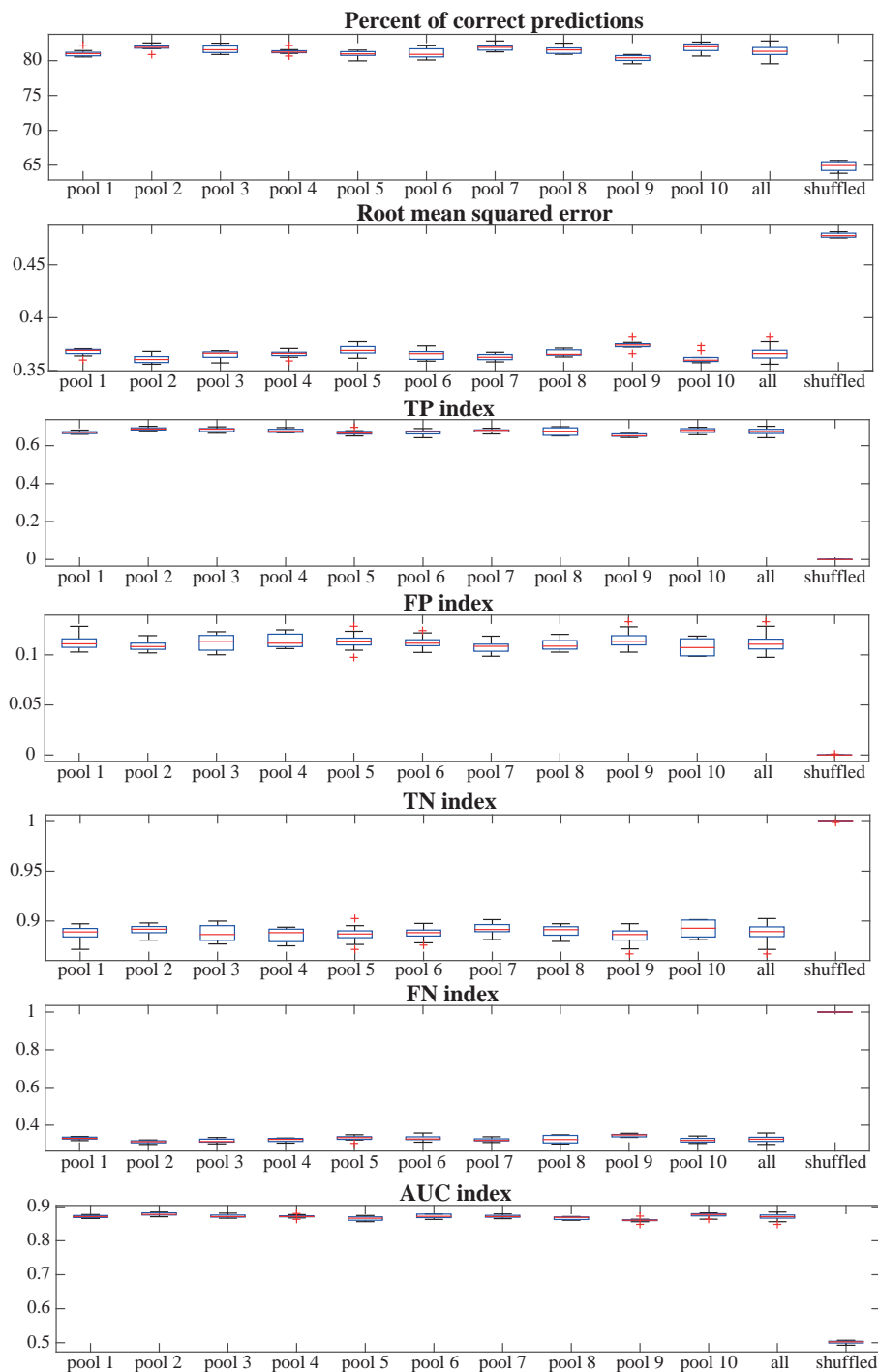




# Appendix

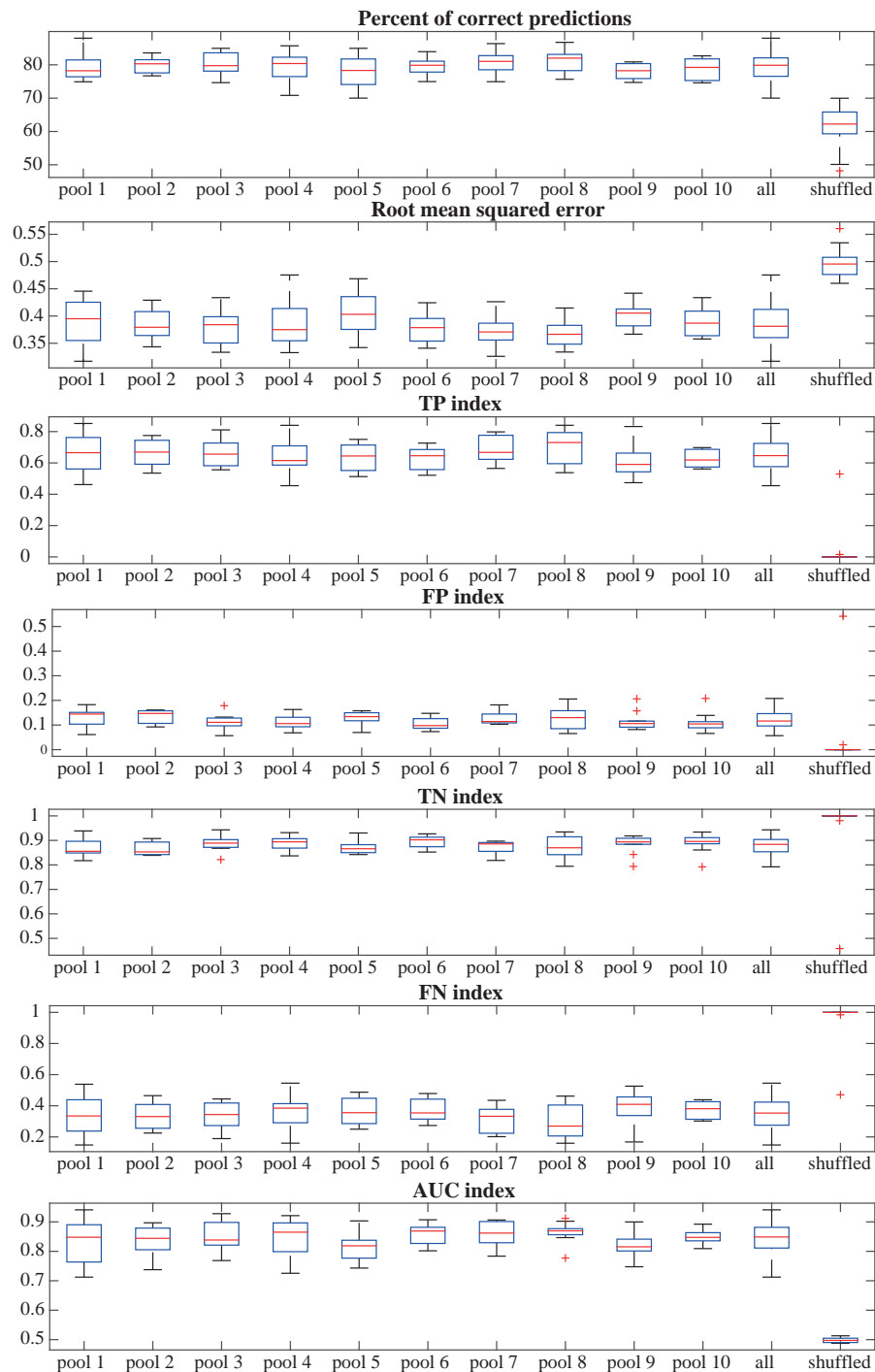
# A

## Logistic regression modelling performances (random sampling)



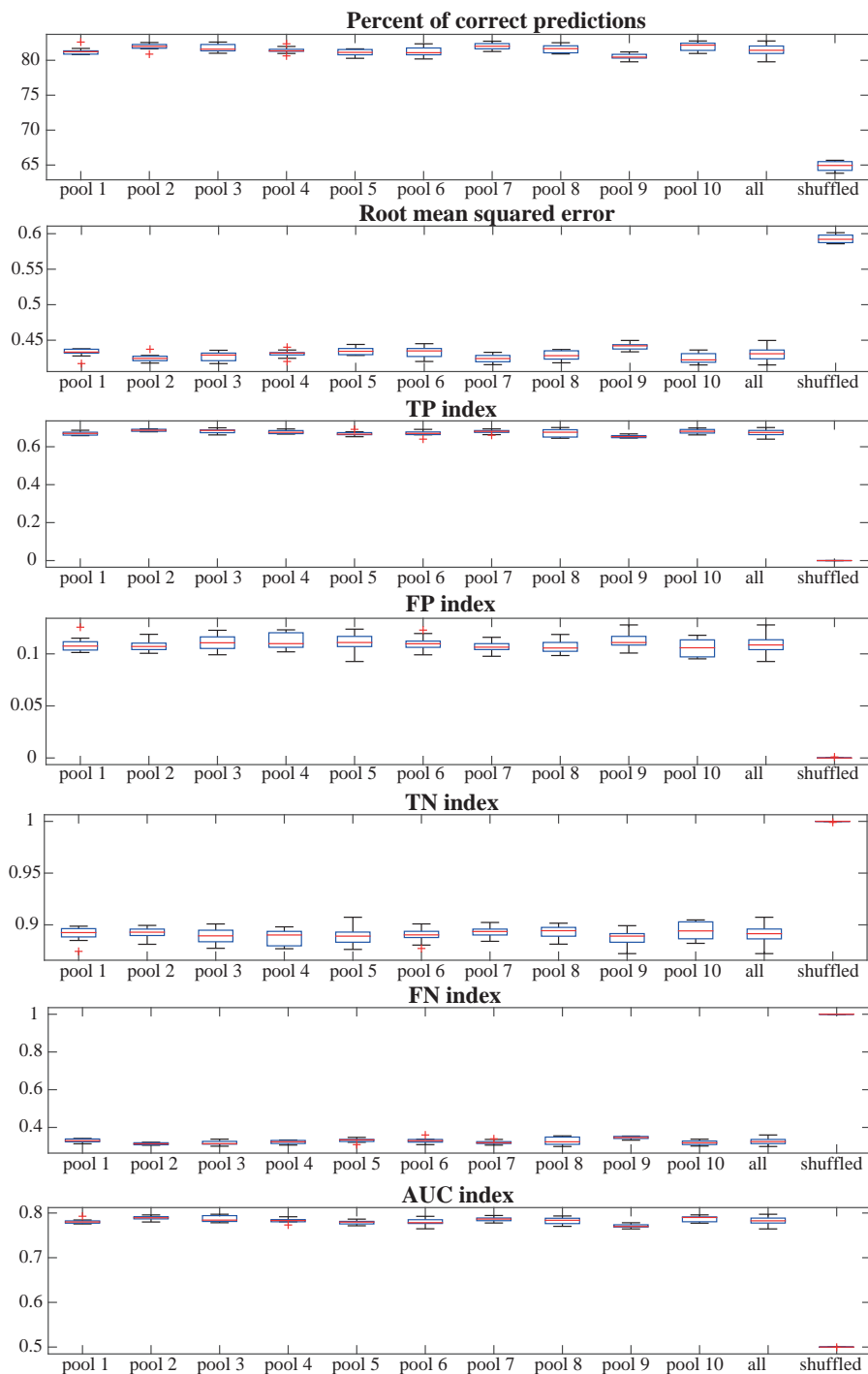
**Fig. A.1.:** Summary statistics for a the Logistic regression classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

## Logistic regression modelling performances (spatial sampling)



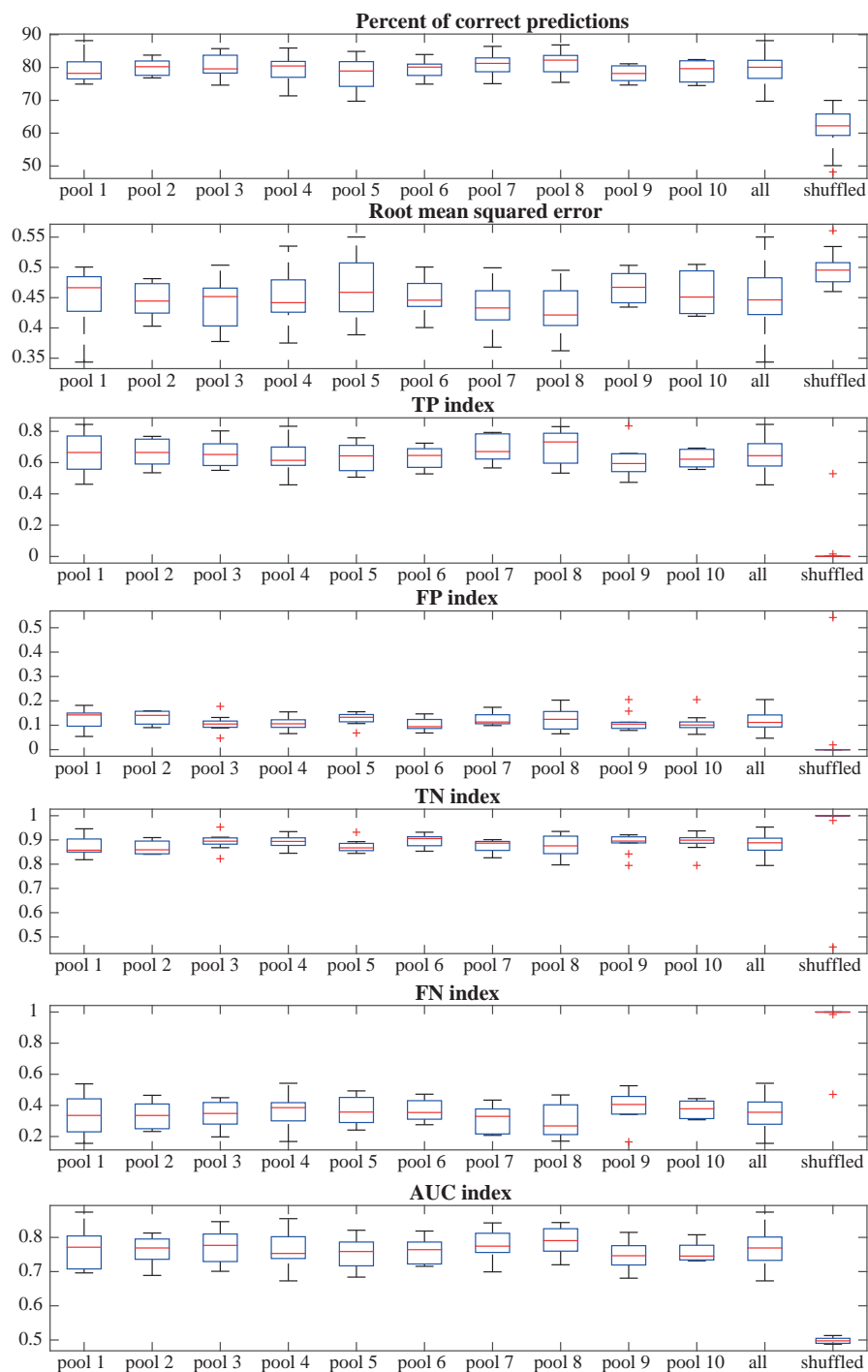
**Fig. A.2.:** Summary statistics for a the Logistic regression classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

## Linear SVM modelling performances (random sampling)



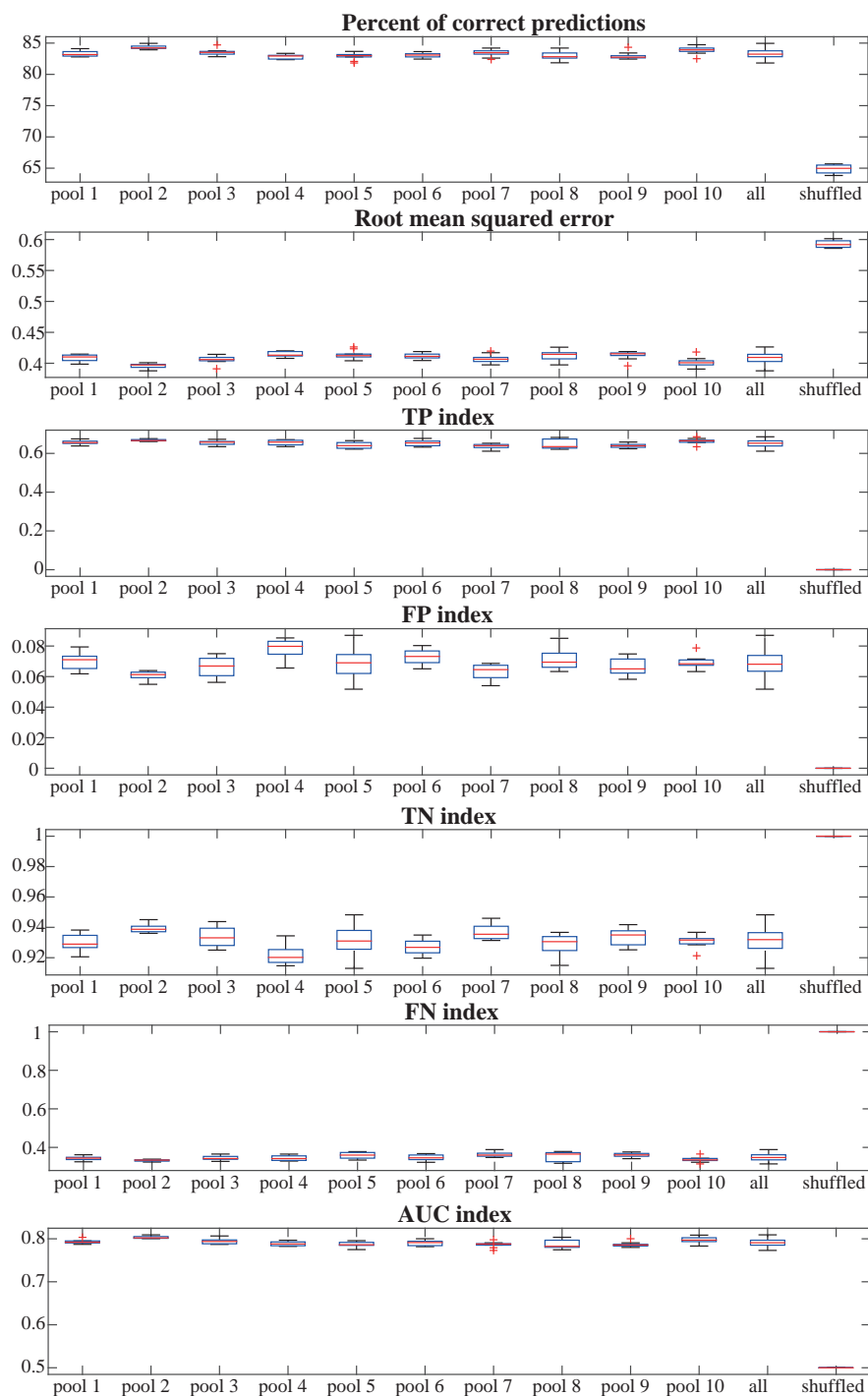
**Fig. A.3.:** Summary statistics for a the linear SVM classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

## Linear SVM modelling performances (spatial sampling)



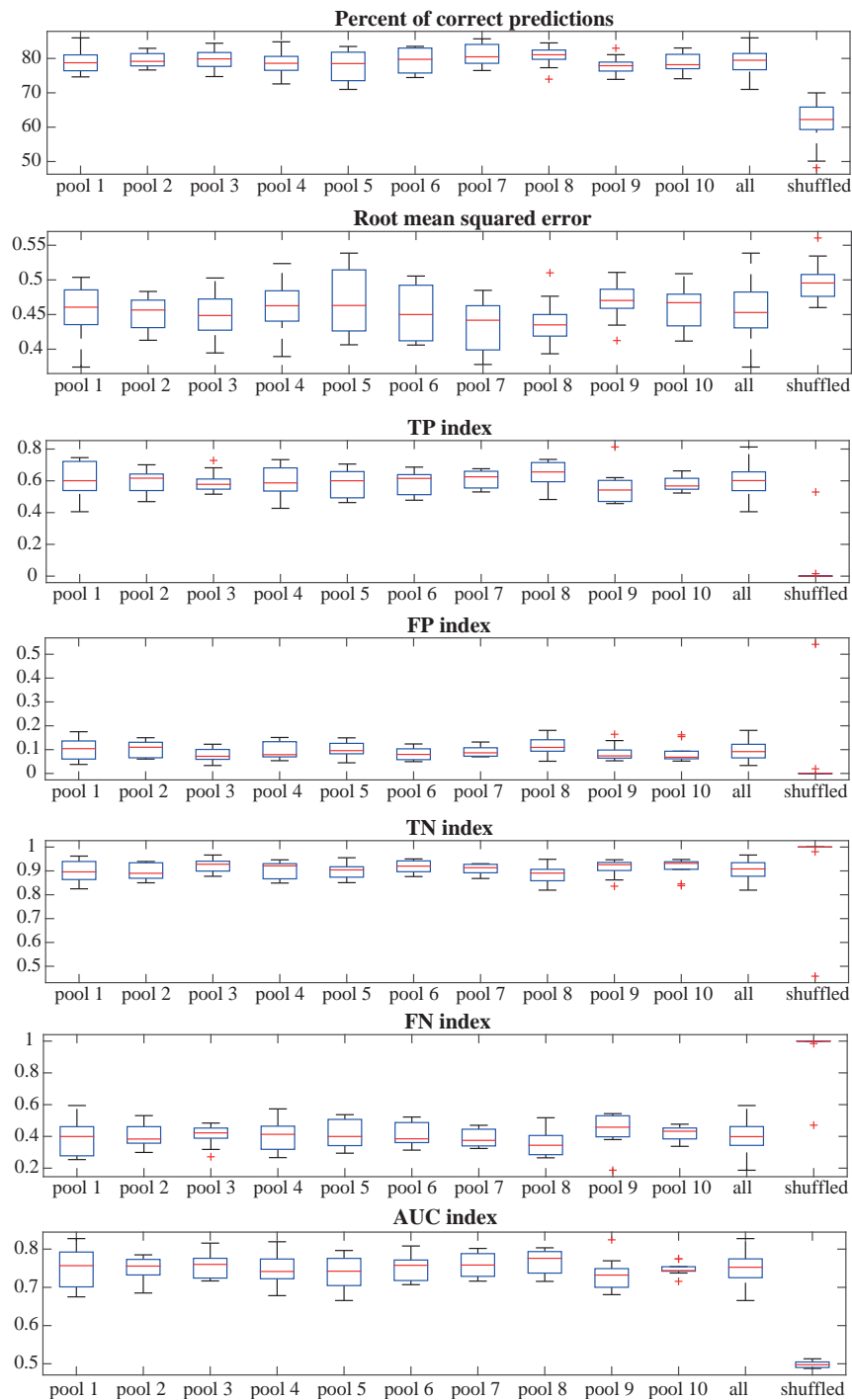
**Fig. A.4.:** Summary statistics for a the linear SVM classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

## Radial basis function SVM modelling performances (random sampling)



**Fig. A.5.:** Summary statistics for a the Radial basis function SVM classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

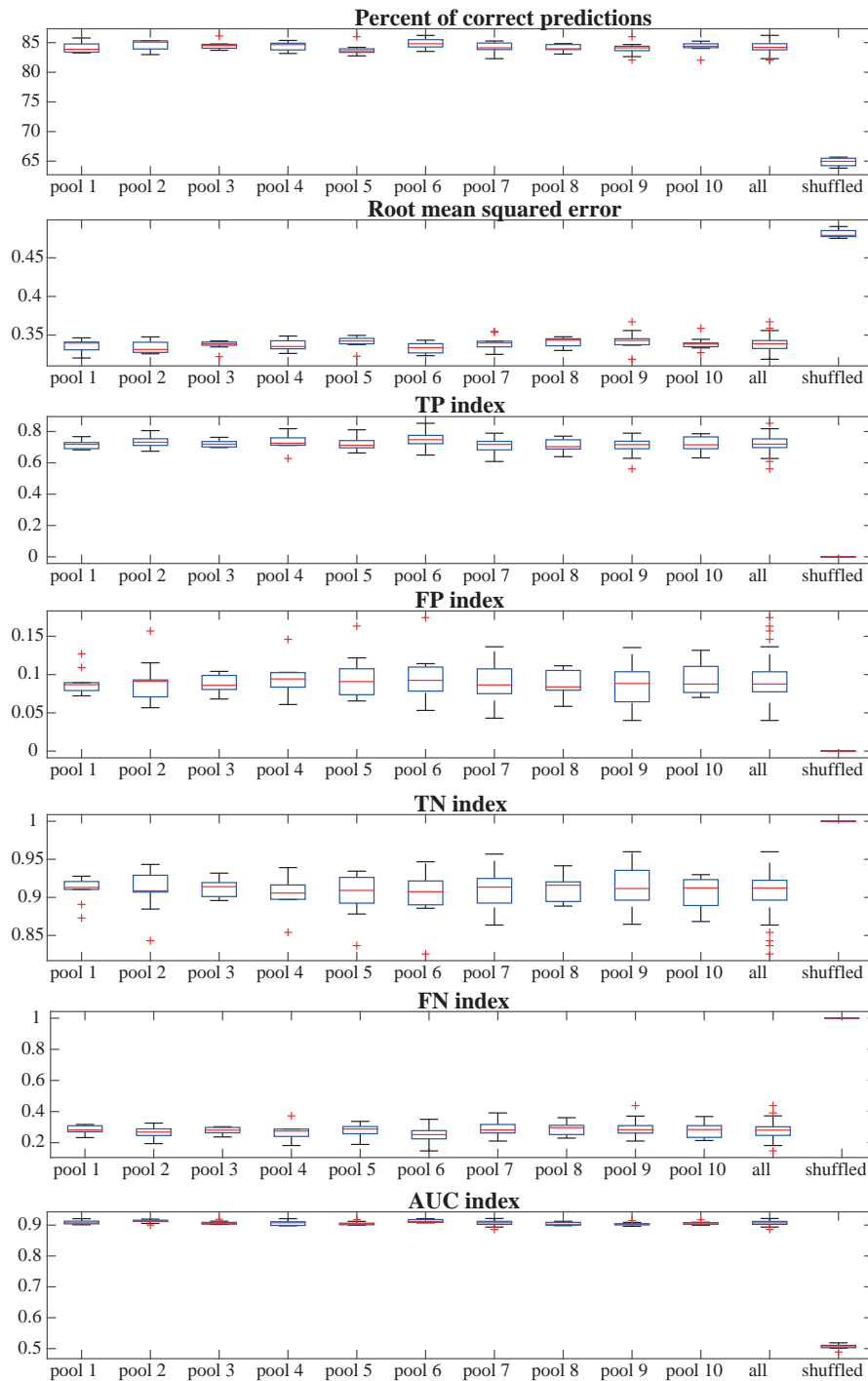
## Radial basis function SVM modelling performances (spatial sampling)



**Fig. A.6.:** Summary statistics for a the Radial basis function SVM classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

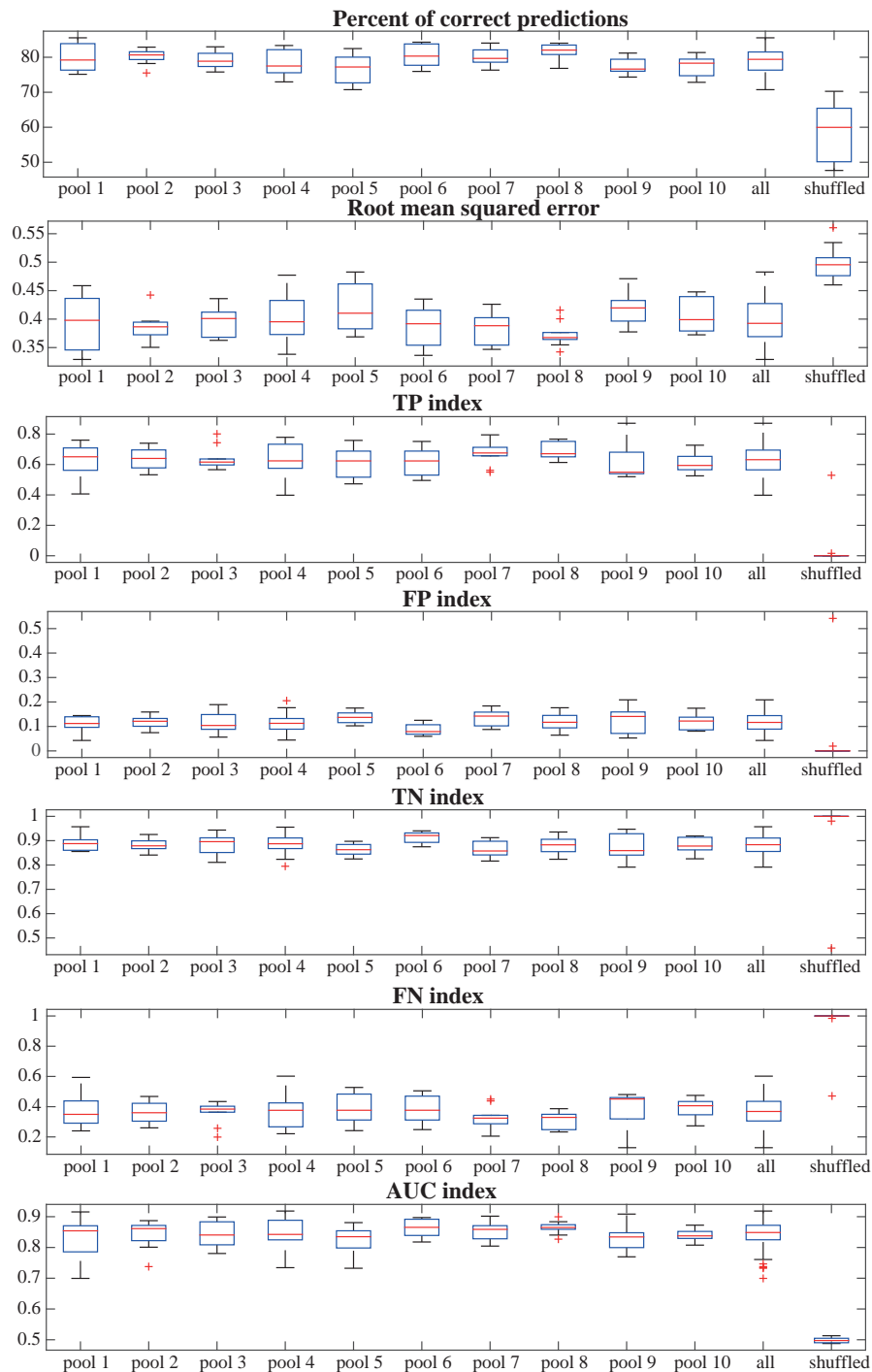


## Multilayer perceptron modelling performances (random sampling)



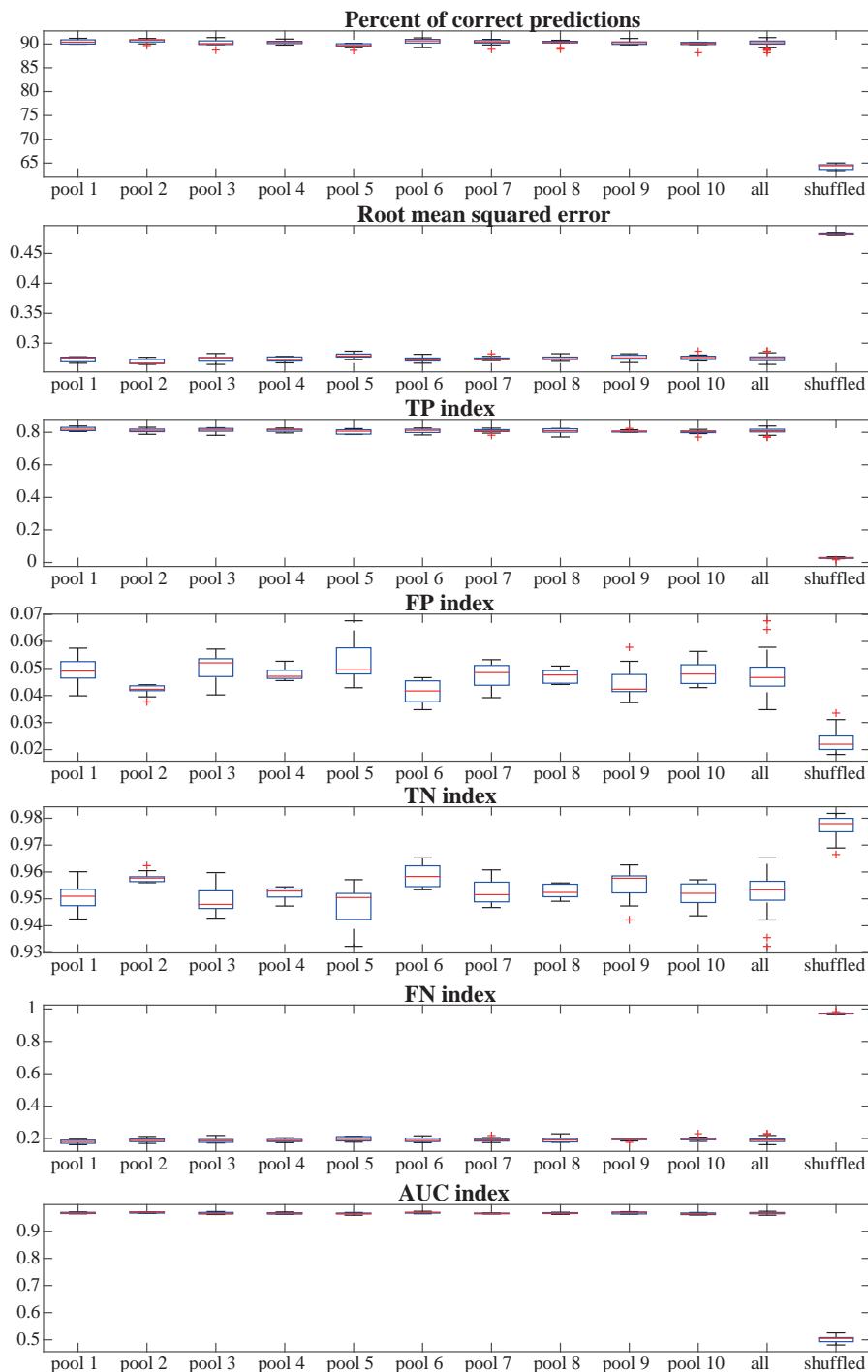
**Fig. A.7.:** Summary statistics for a the Multilayer perceptron classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

## Multilayer perceptron modelling performances (spatial sampling)



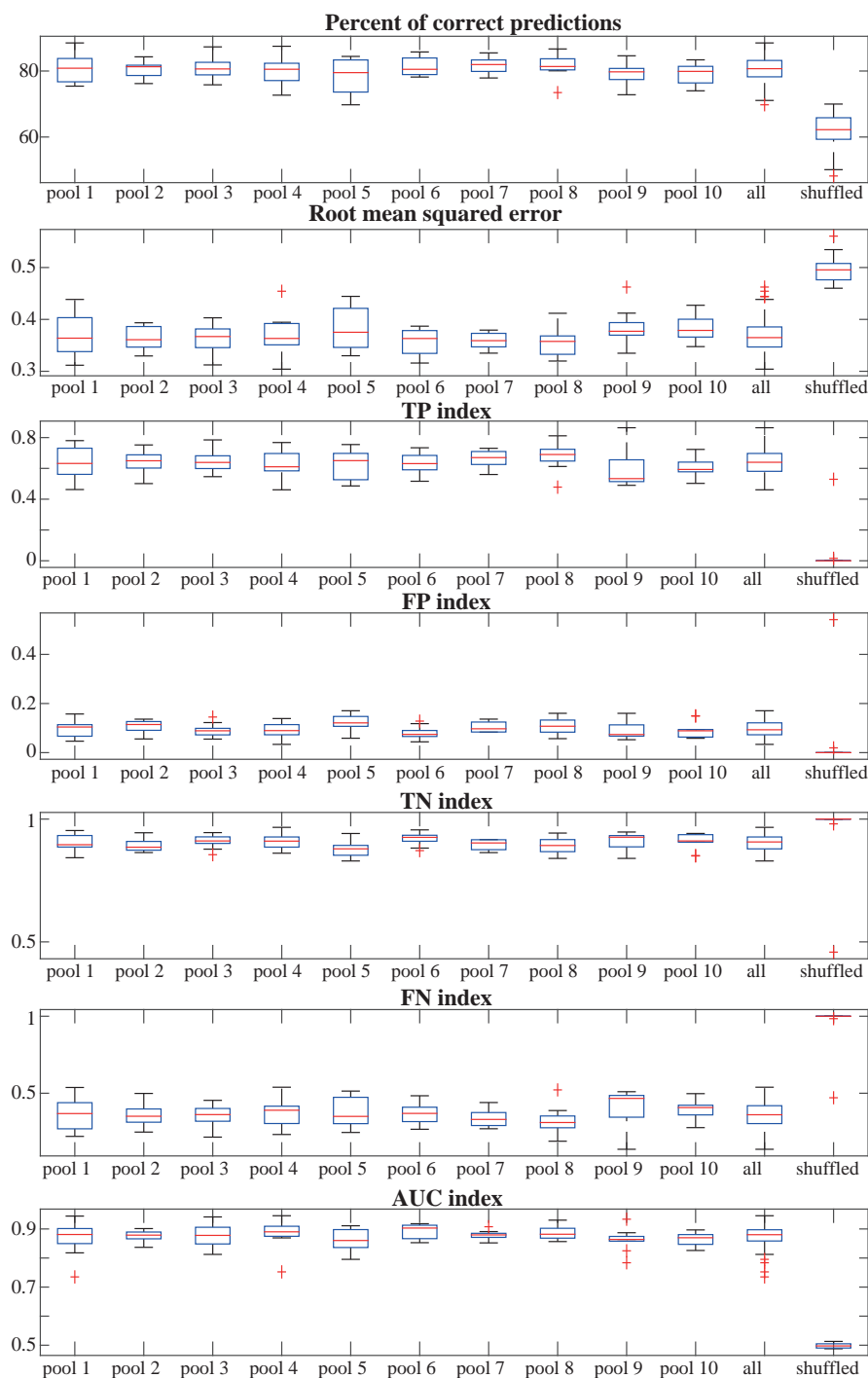
**Fig. A.8.:** Summary statistics for a the Multilayer perceptron classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

## Random forest modelling performances (random sampling)



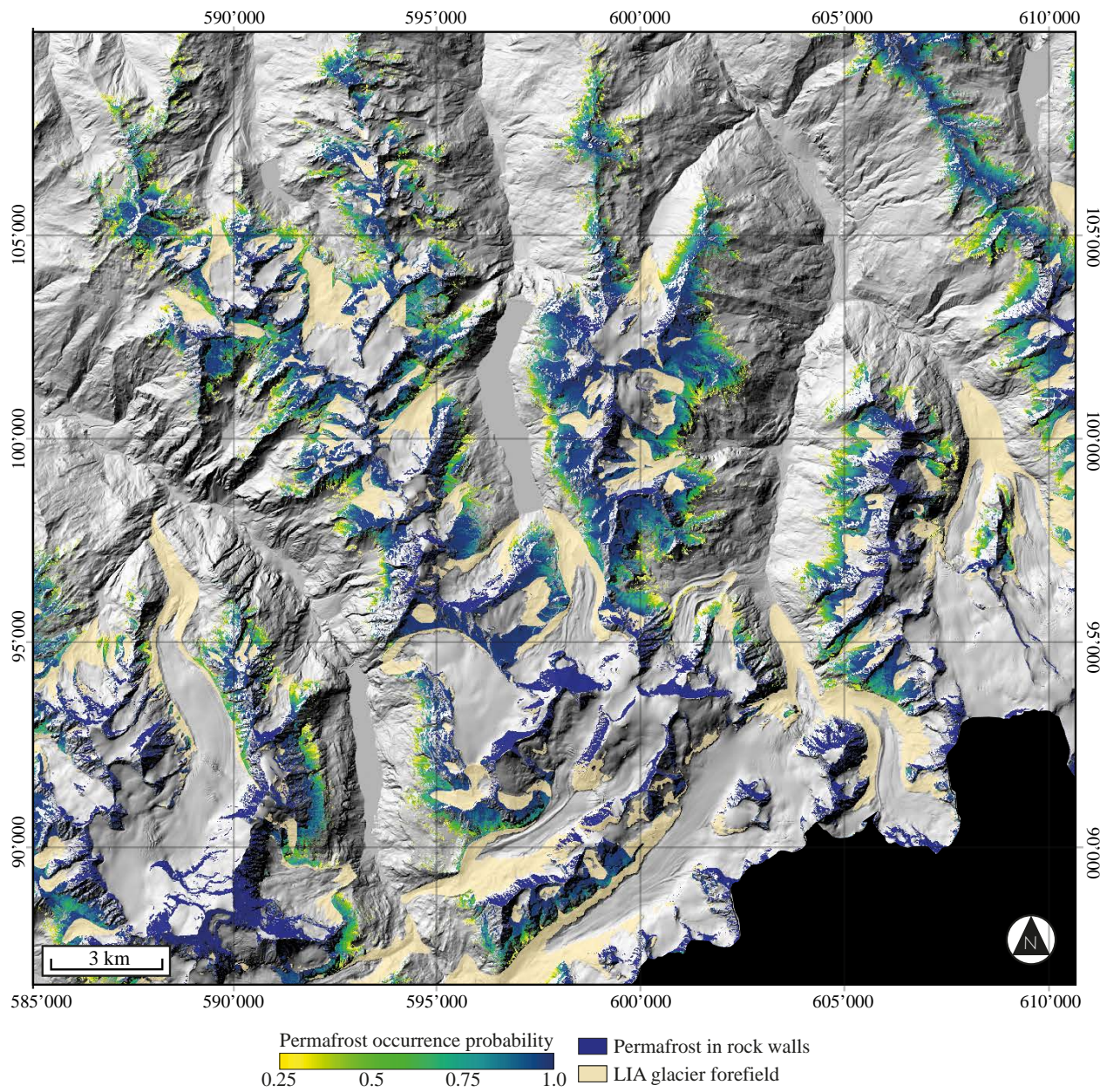
**Fig. A.9.:** Summary statistics for a the Random forest classification of 10 pools of permafrost data (random sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

## Random forest modelling performances (spatial sampling)

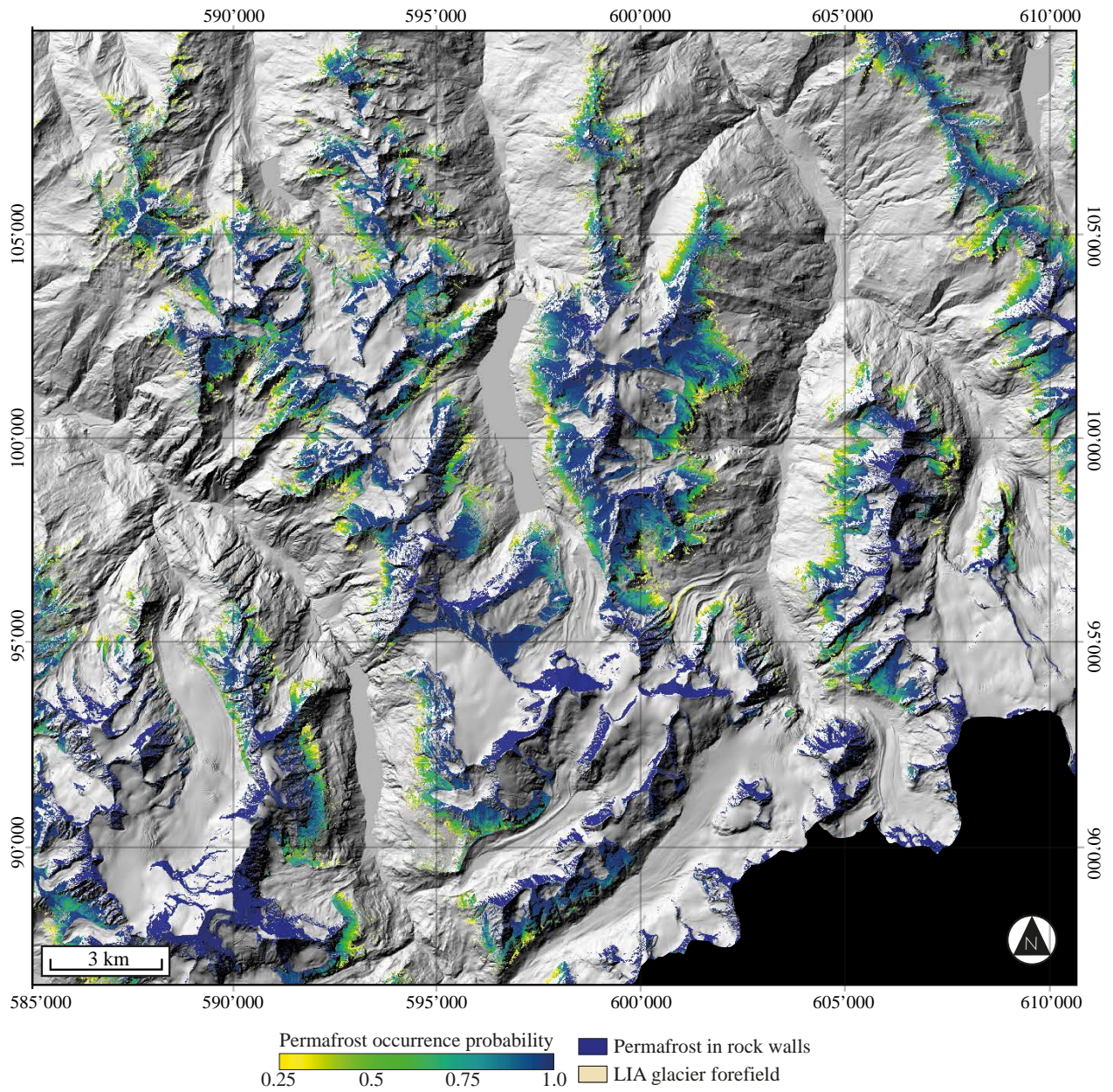


**Fig. A.10.:** Summary statistics for a the Random forest classification of 10 pools of permafrost data (spatial sampling): accuracy (percentage of correct predictions), root mean squared error (RMSE), true positive rate (TP index), false positive rate (FP index), true negative rate (TN index), false negative rate (FN index) and area under the ROC curve (AUC index).

## Logistic regression permafrost maps

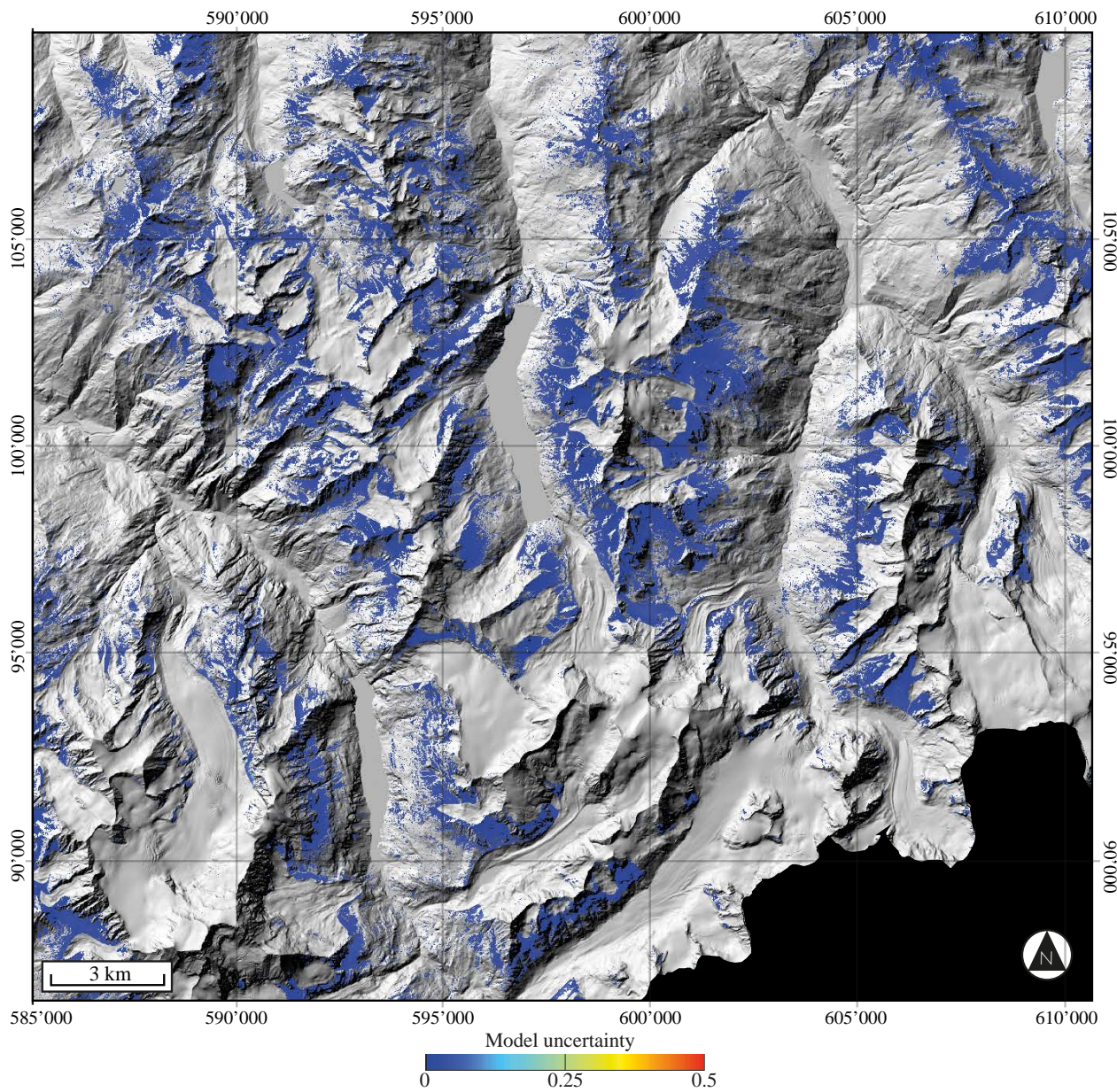


**Fig. A.11.:** Potential permafrost extent obtained with LR classification on randomly sampled data.

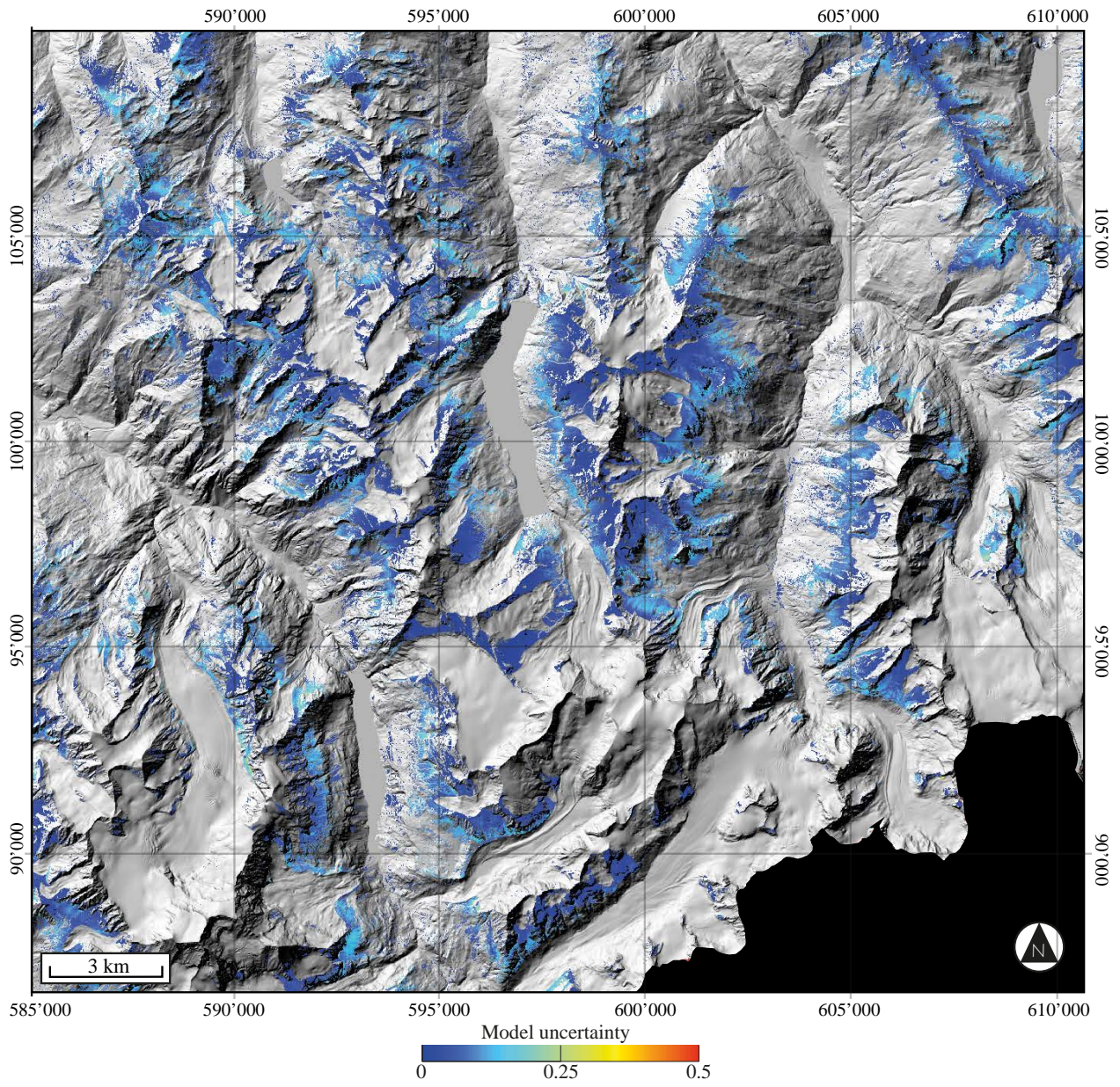


**Fig. A.12.:** Potential permafrost extent obtained with LR classification on spatially sampled data.

## Logistic regression uncertainty maps



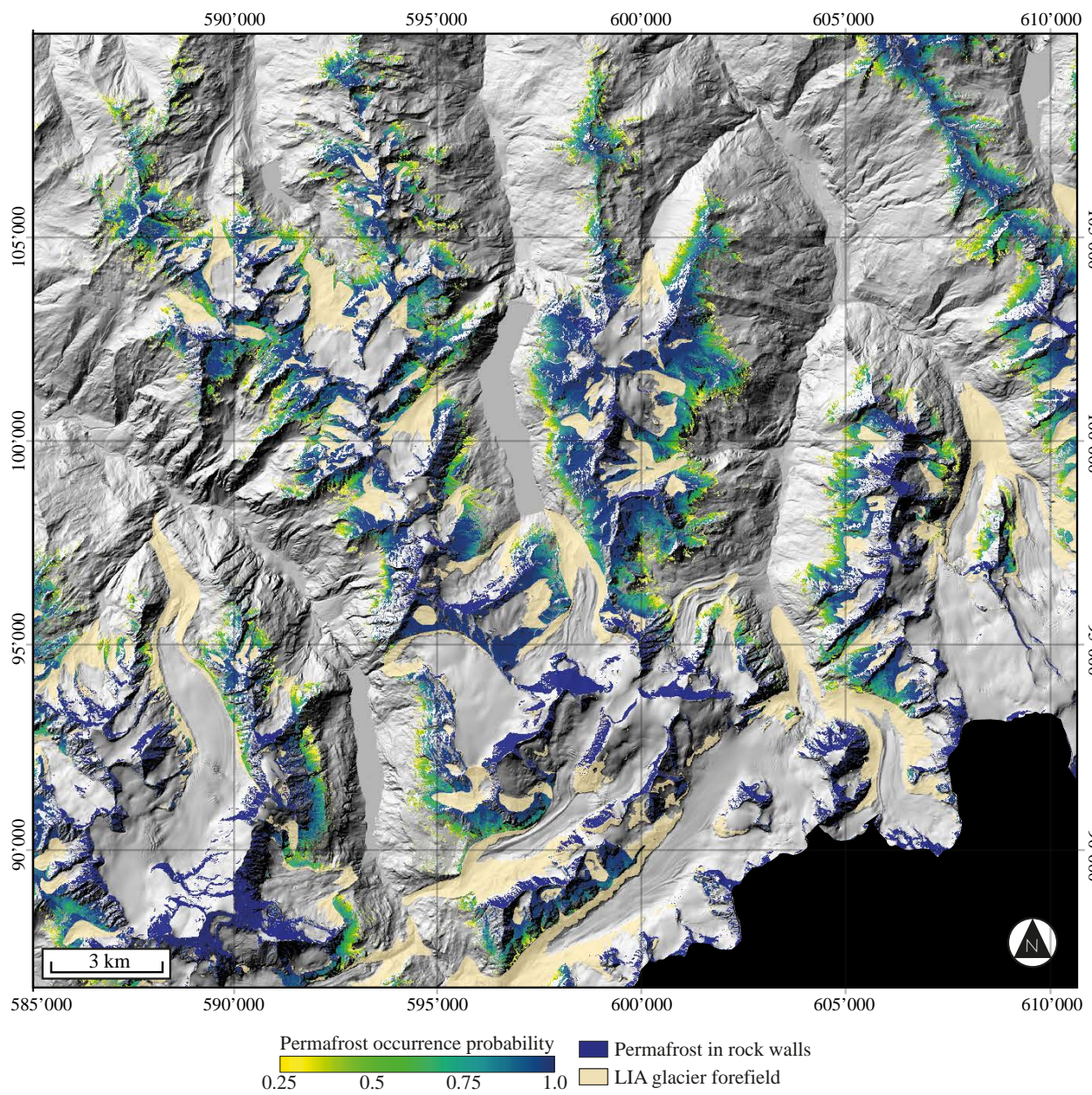
**Fig. A.13.:** Uncertainty map obtained with LR classification on randomly sampled data.



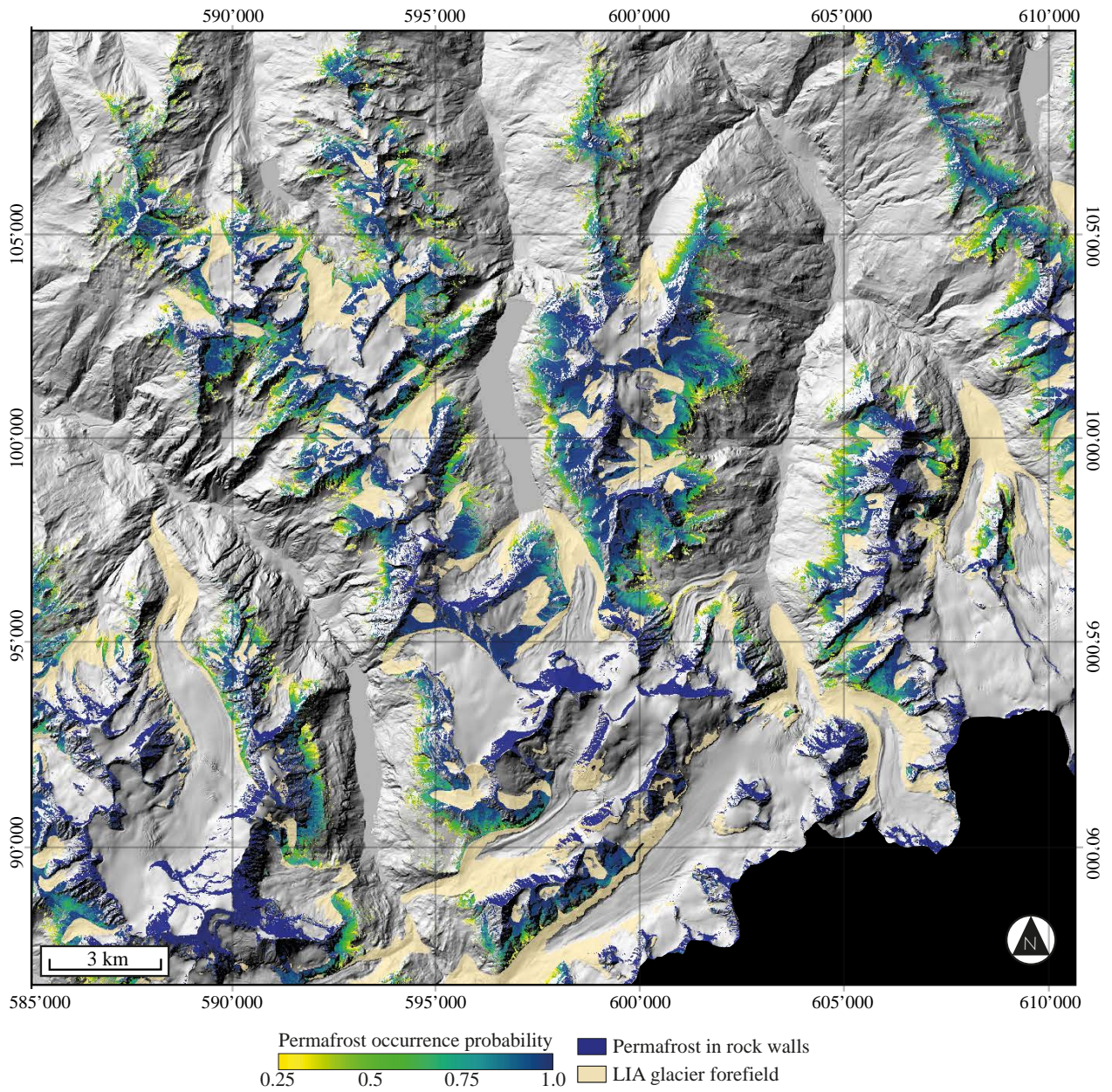
**Fig. A.14.:** Uncertainty map extent obtained with LR classification on spatially sampled data.



## Linear SVM permafrost maps

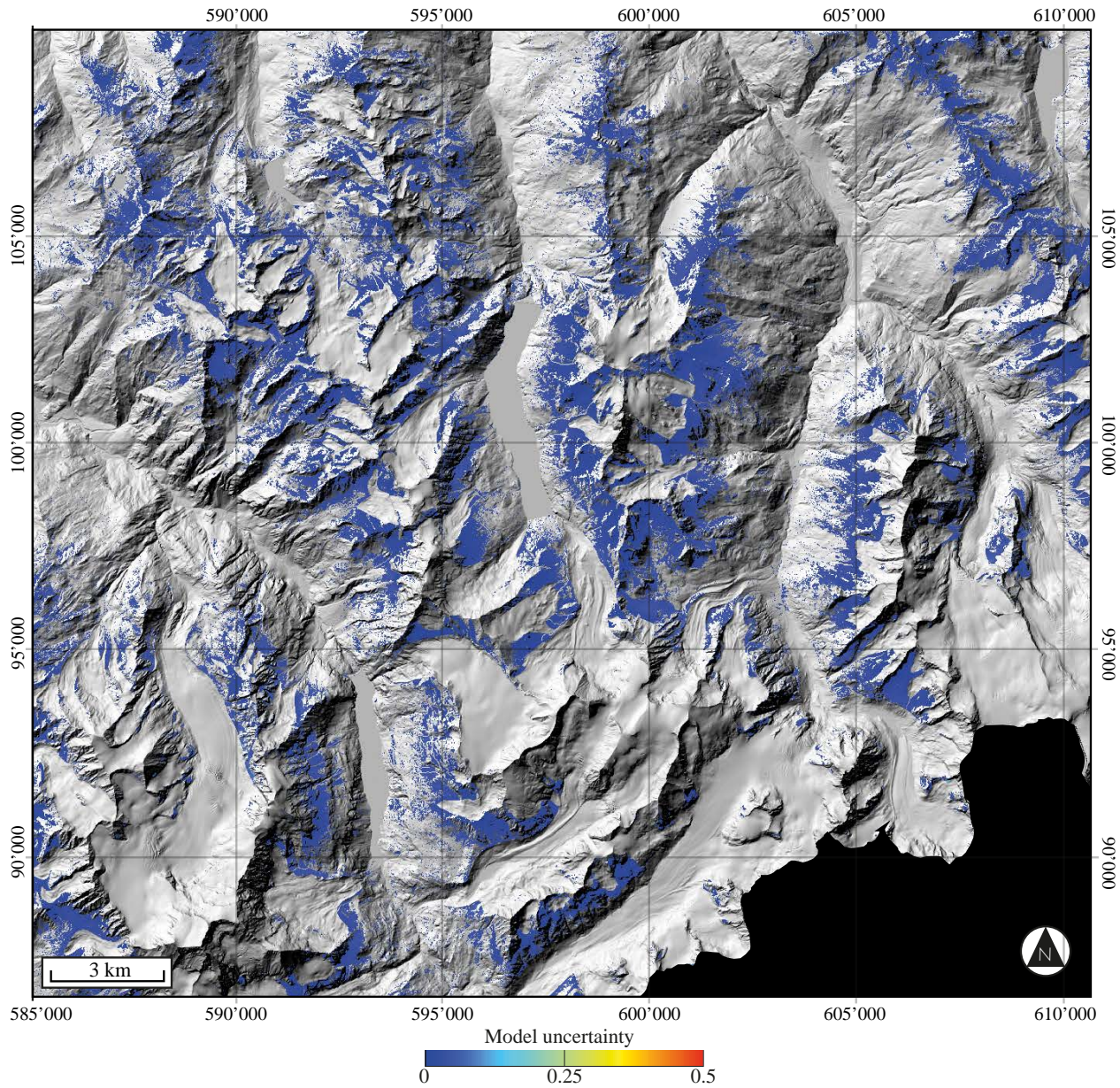


**Fig. A.15.:** Potential permafrost extent obtained with linear SVM classification on randomly sampled data.

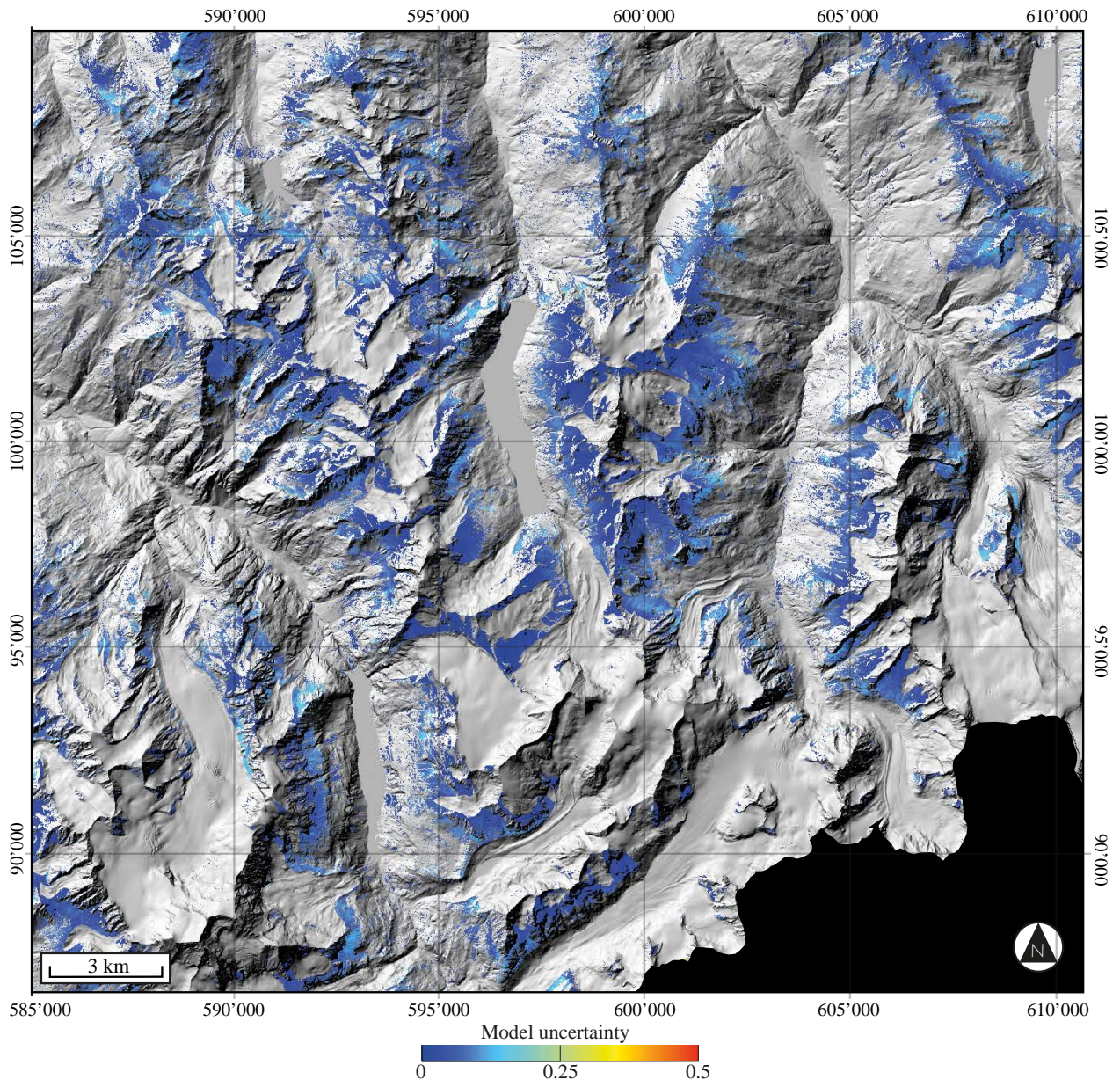


**Fig. A.16.:** Potential permafrost extent obtained with linear SVM classification on spatially sampled data.

## Linear SVM uncertainty maps

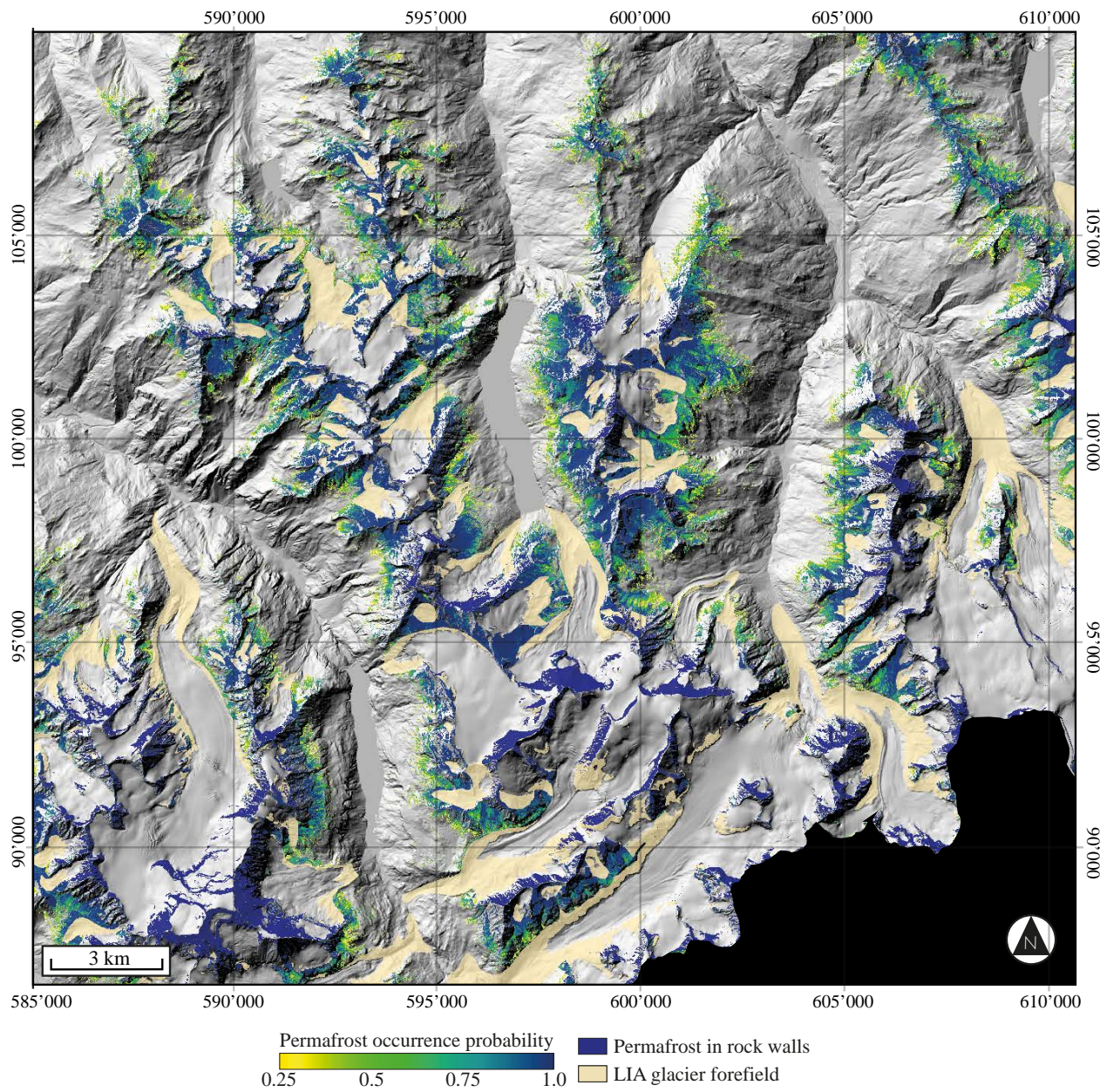


**Fig. A.17.:** Uncertainty map obtained with linear SVM classification on randomly sampled data.

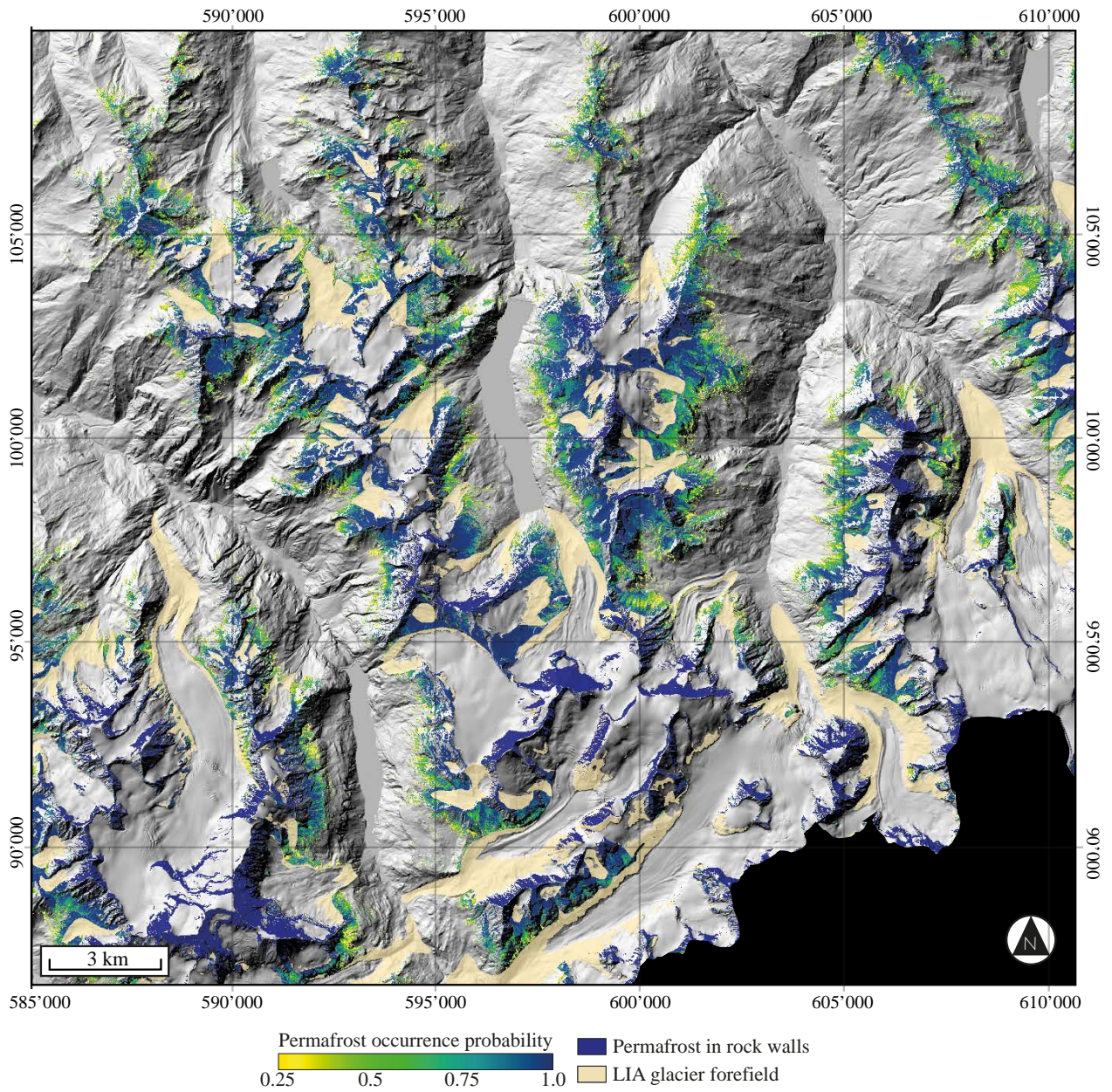


**Fig. A.18.:** Uncertainty map extent obtained with linear SVM classification on spatially sampled data.

## Radial basis function SVM permafrost maps

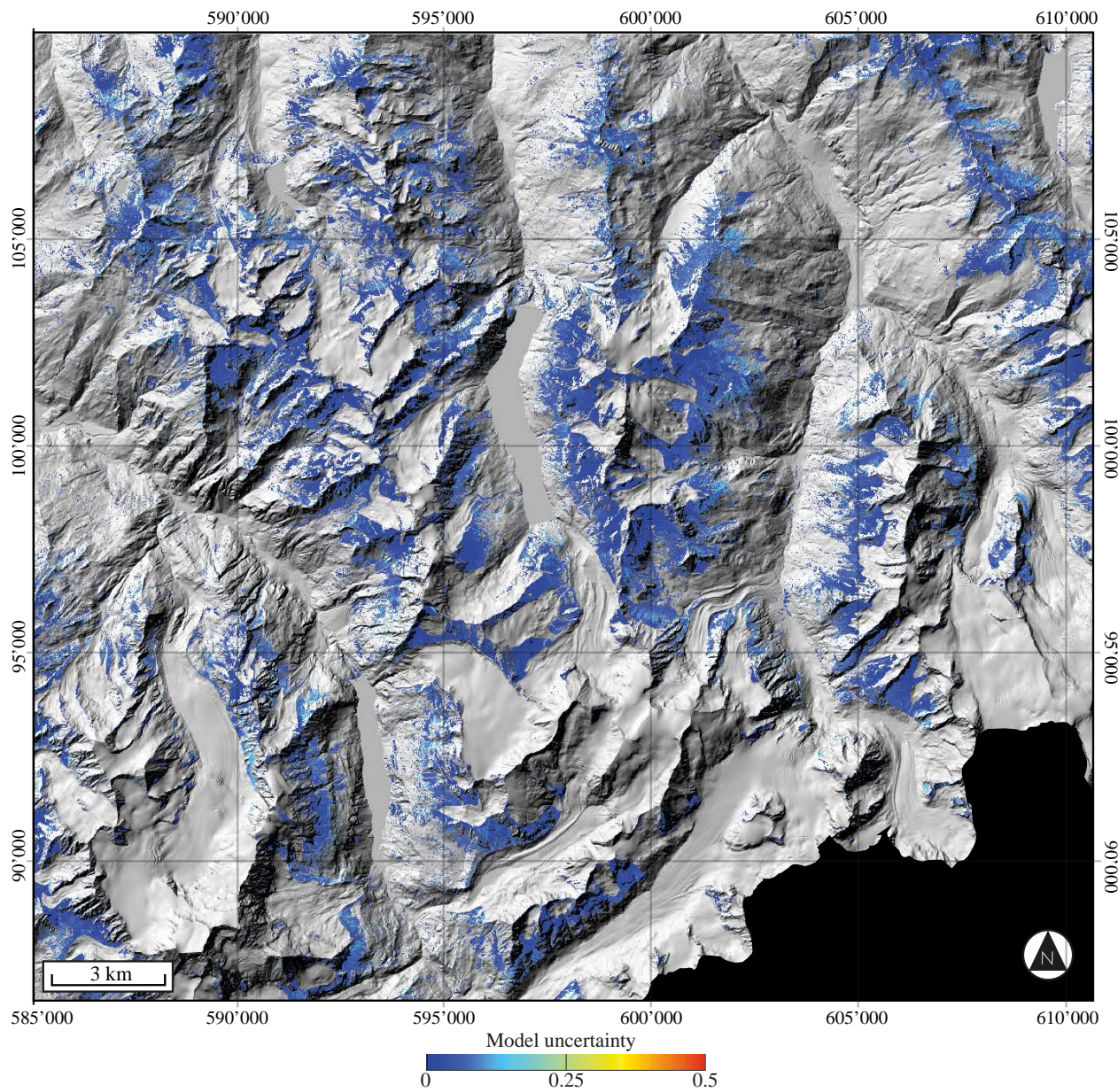


**Fig. A.19.:** Potential permafrost extent obtained with RBF-SVM classification on randomly sampled data.

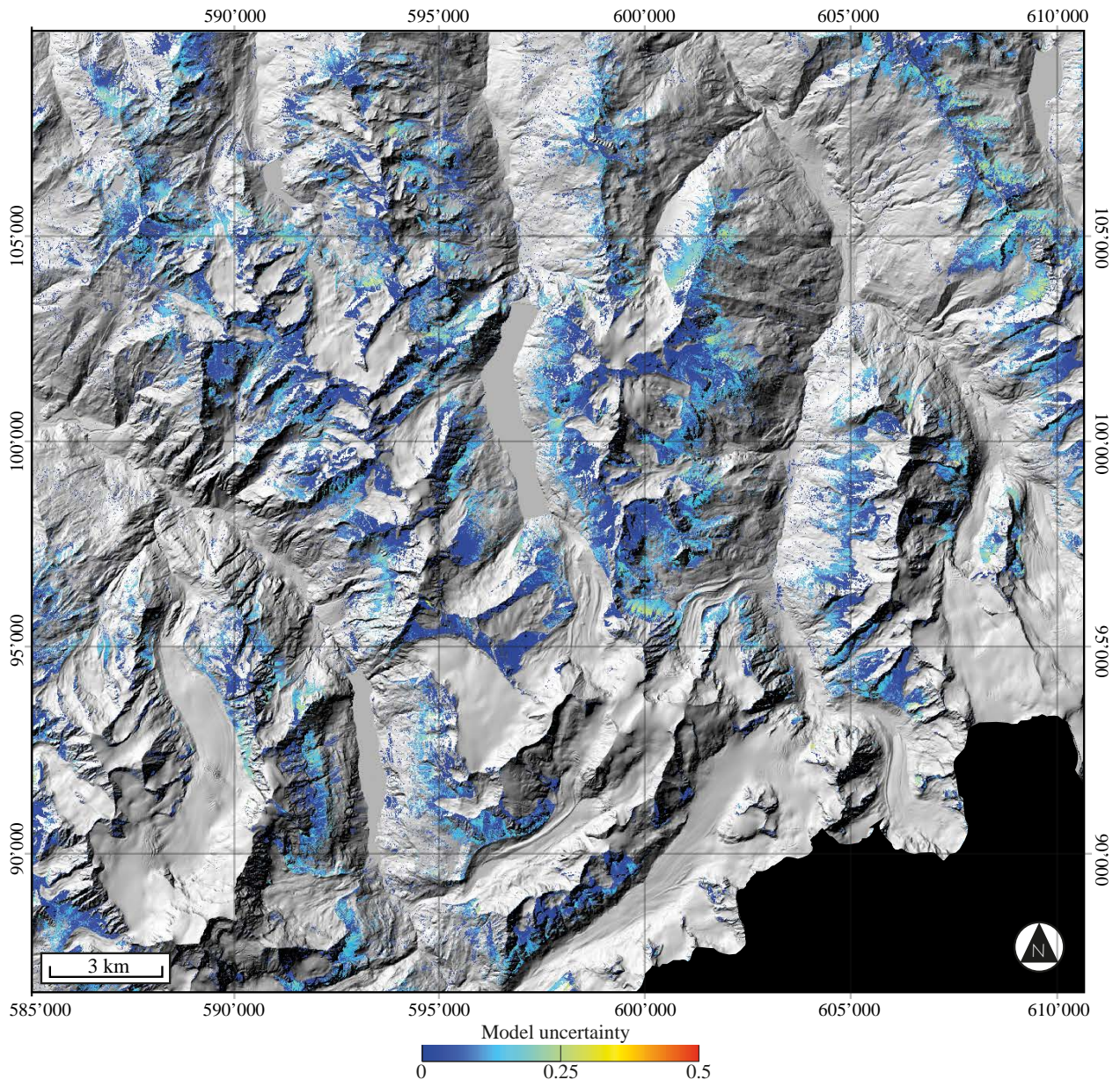


**Fig. A.20.:** Potential permafrost extent obtained with RBF-SVM classification on spatially sampled data.

## Radial basis function SVM uncertainty maps



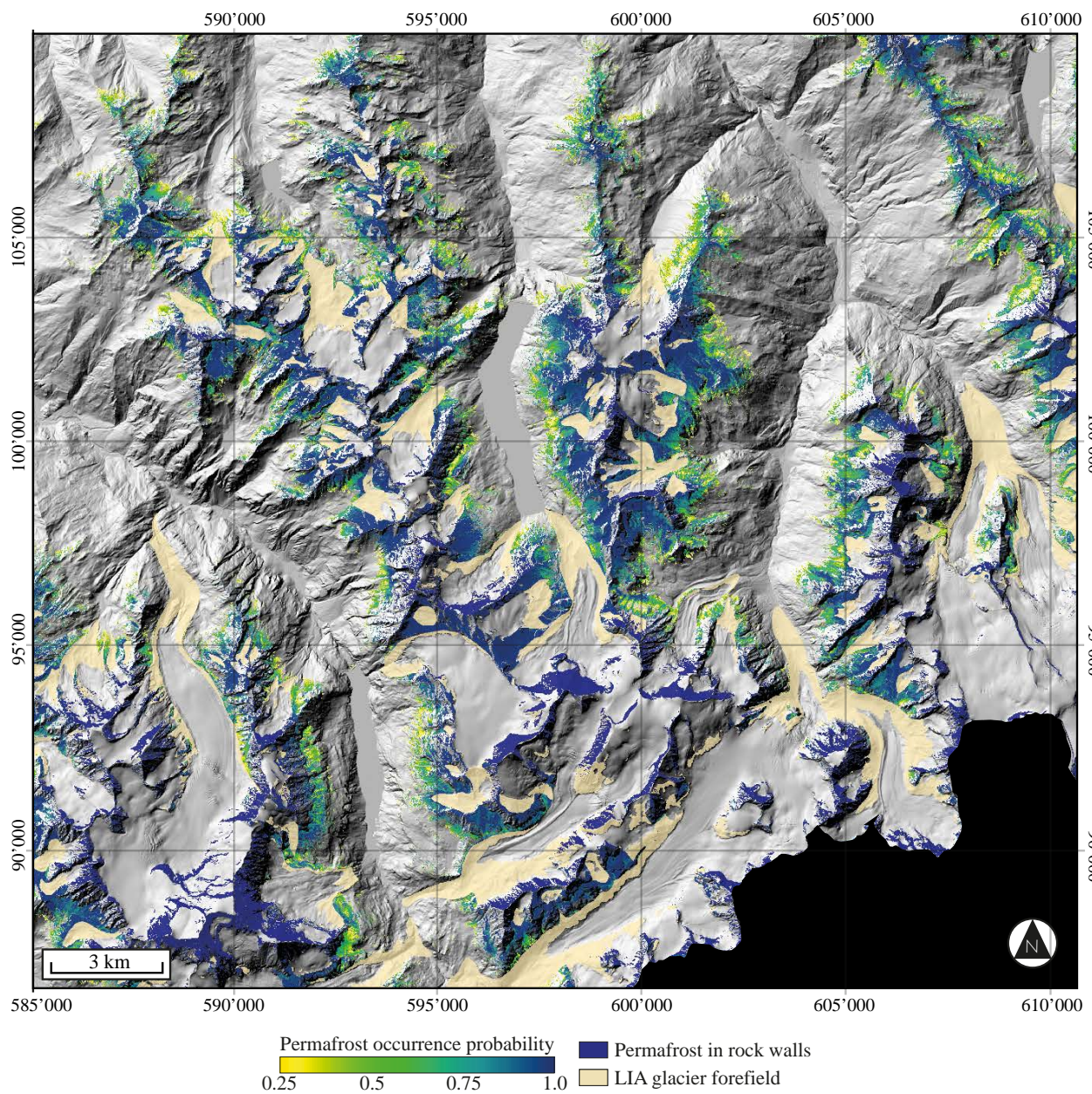
**Fig. A.21.:** Uncertainty map obtained with RBF-SVM classification on randomly sampled data.



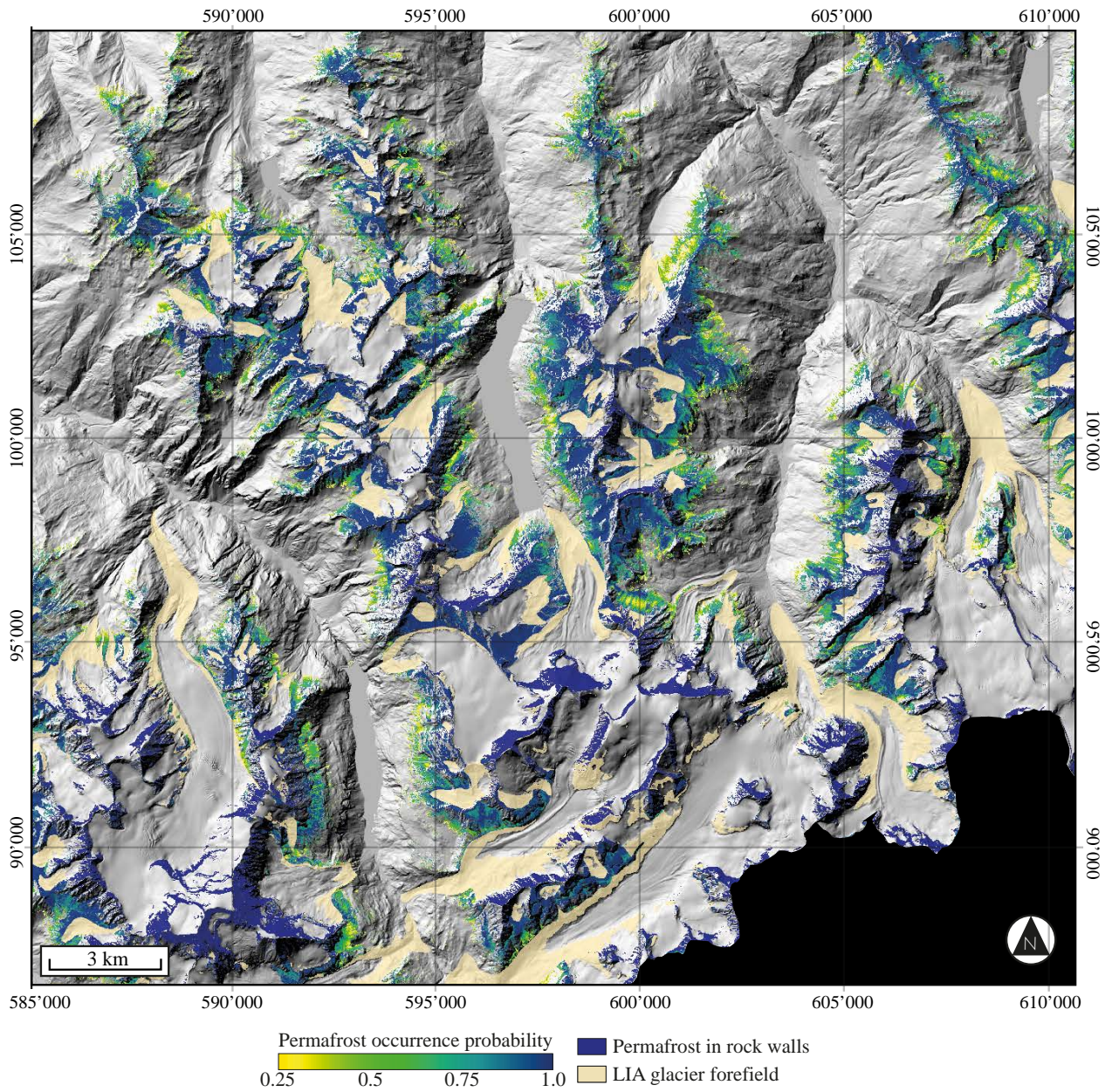
**Fig. A.22.:** Uncertainty map extent obtained RBF-SVM classification on spatially sampled data.



## Multilayer perceptron permafrost maps

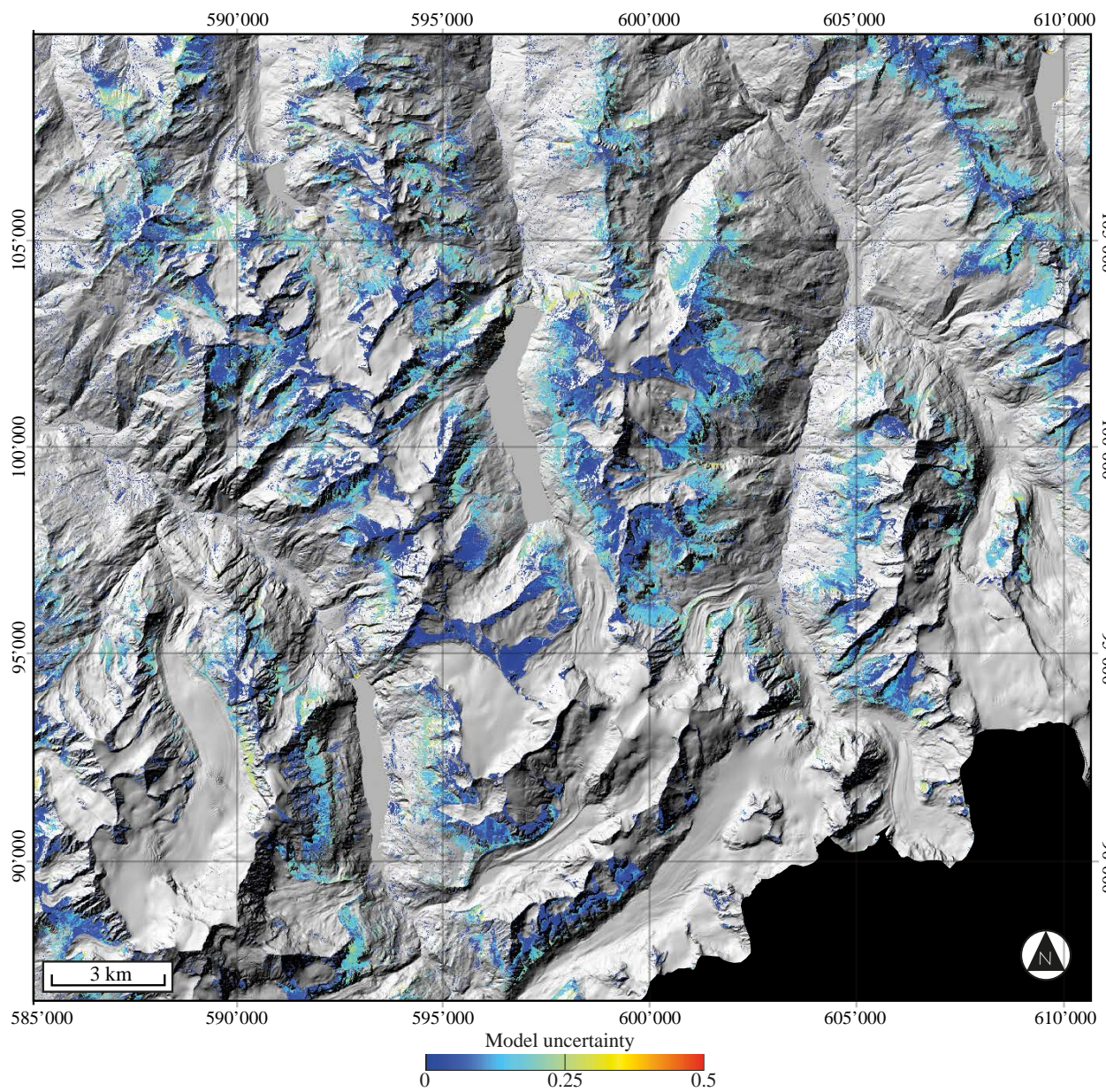


**Fig. A.23.:** Potential permafrost extent obtained with MLP classification on randomly sampled data.

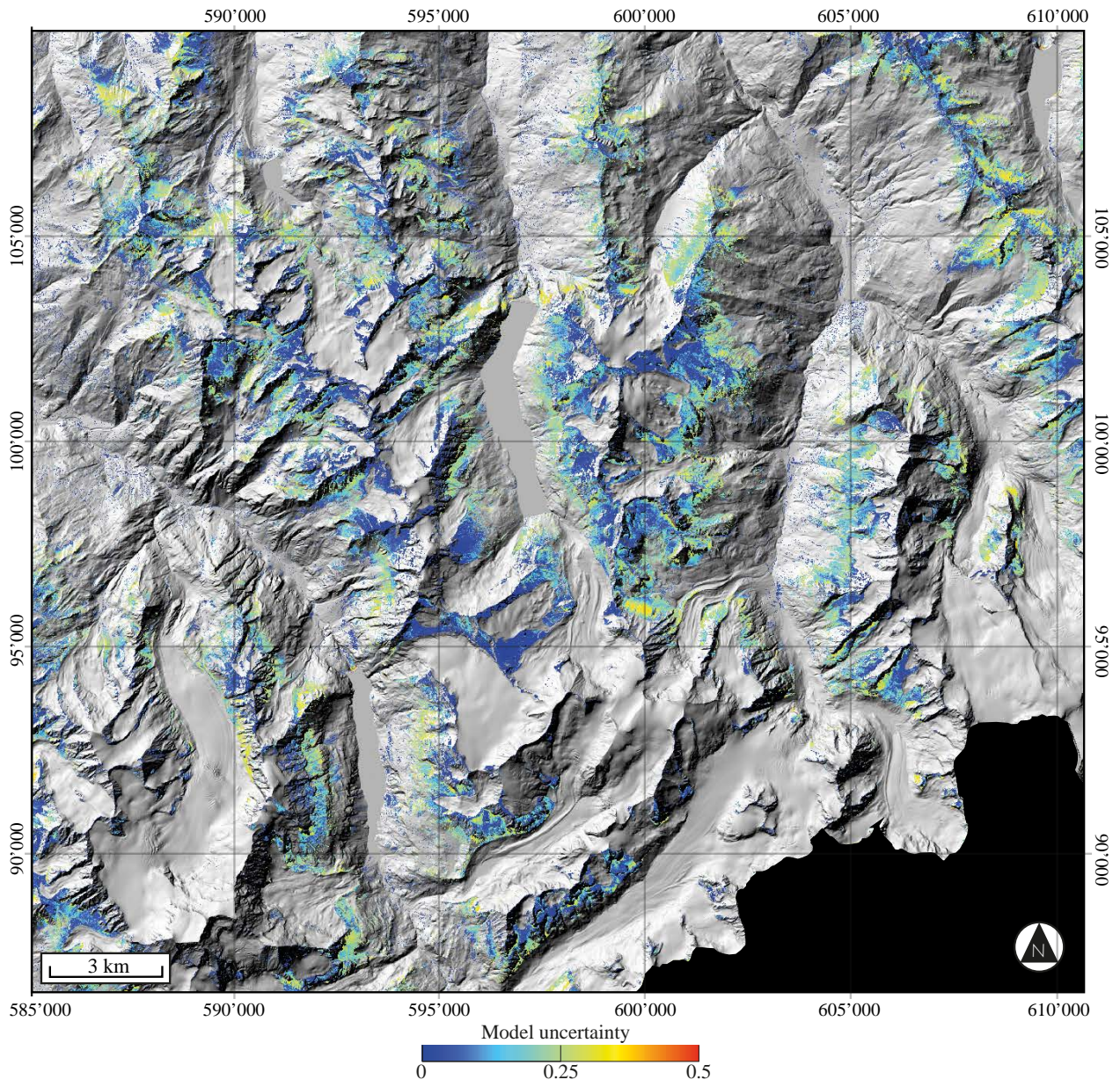


**Fig. A.24.:** Potential permafrost extent obtained with MLP classification on spatially sampled data.

## Multilayer perceptron uncertainty maps

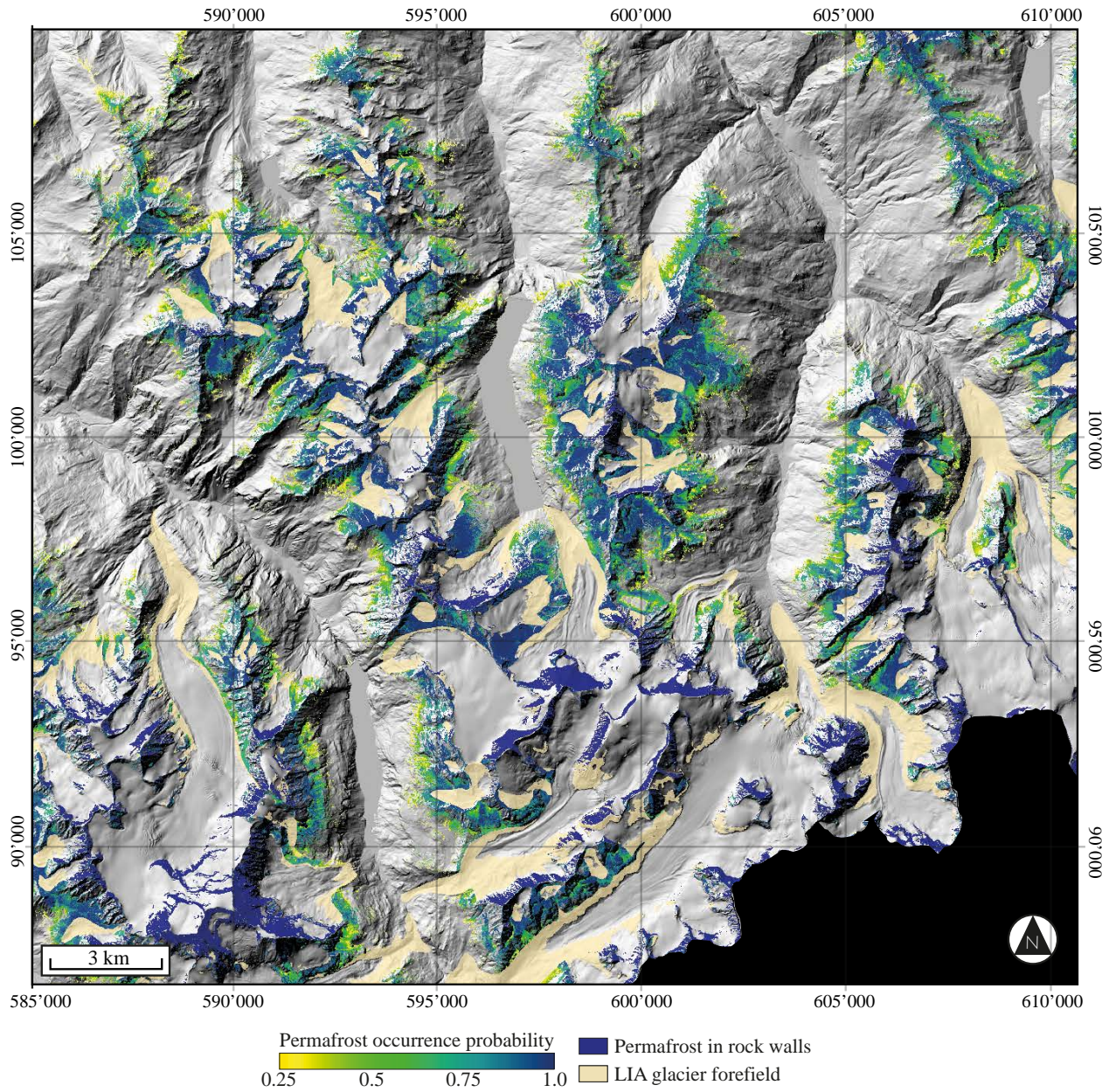


**Fig. A.25.:** Uncertainty map obtained with MLP classification on randomly sampled data.

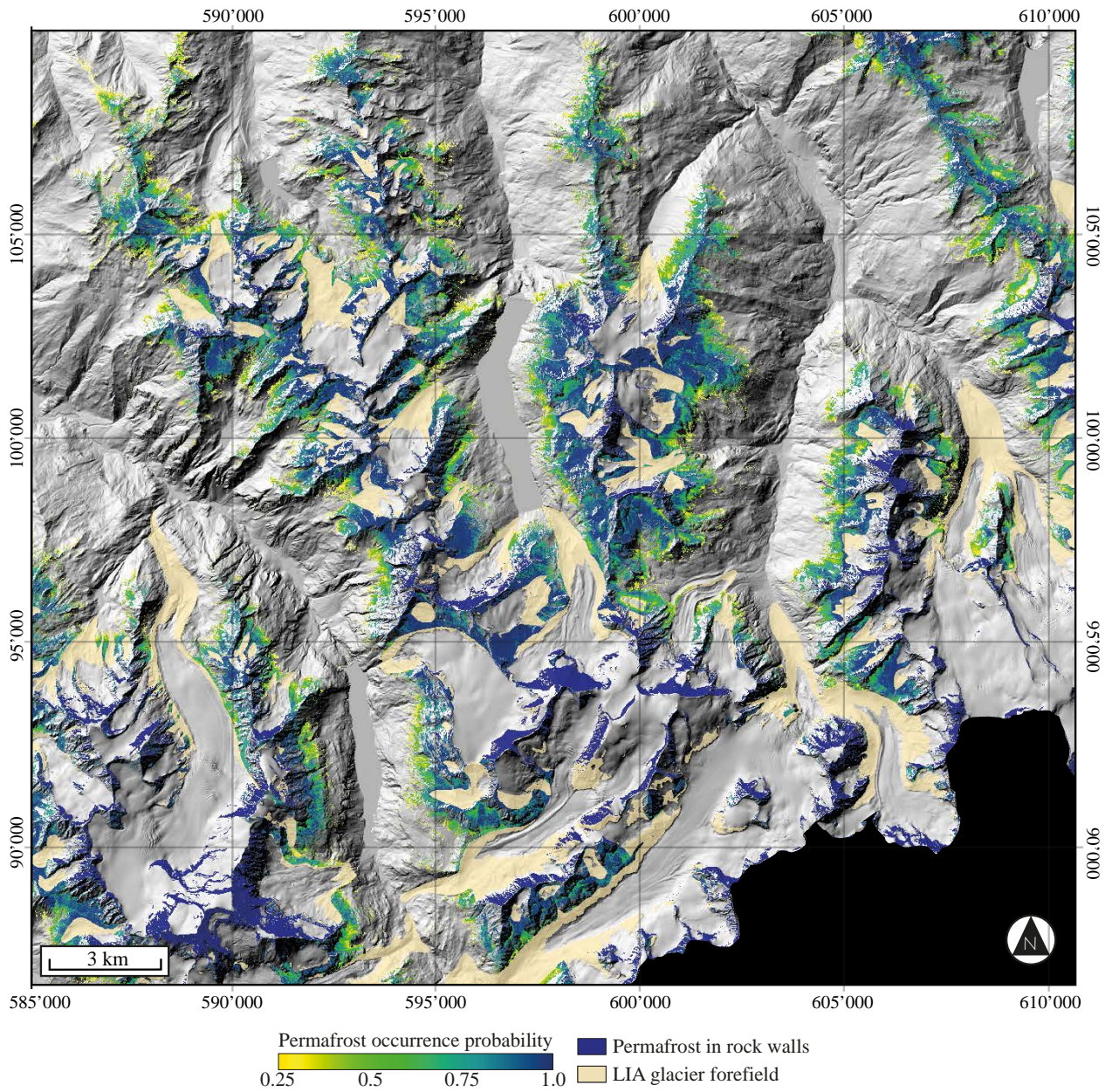


**Fig. A.26.:** Uncertainty map extent obtained MLP classification on spatially sampled data.

## Random forest permafrost maps

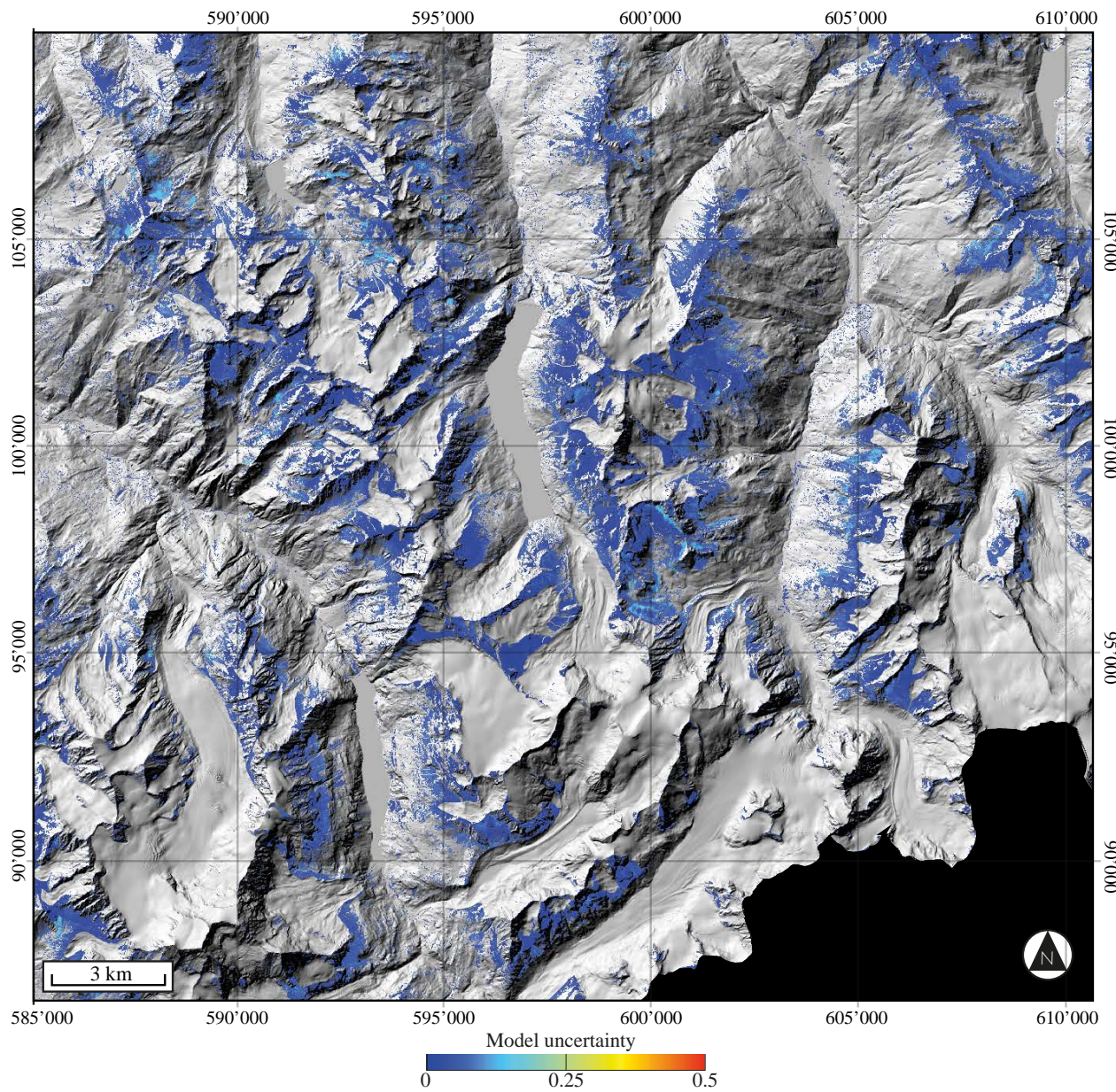


**Fig. A.27.:** Potential permafrost extent obtained with RF classification on randomly sampled data.

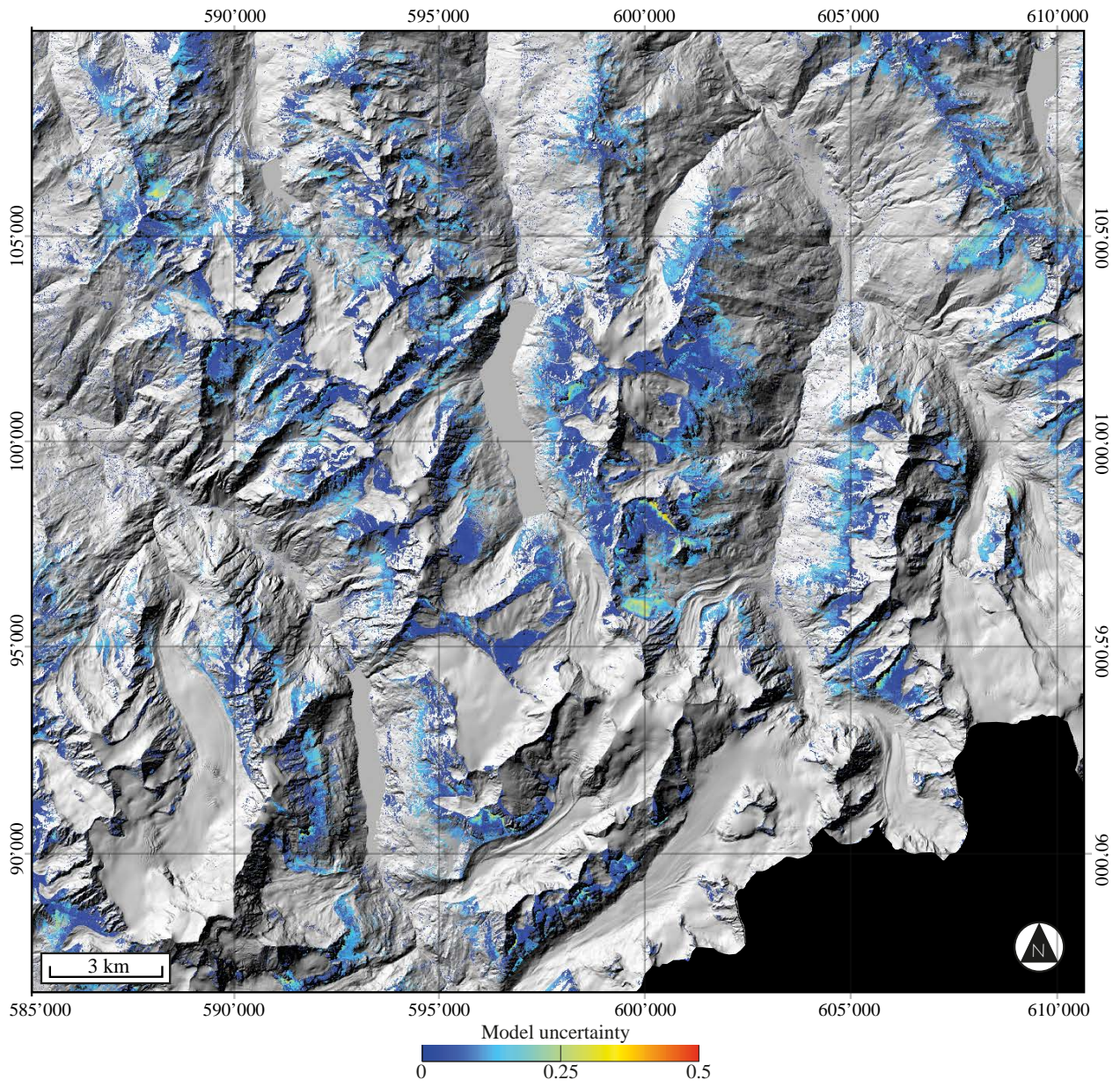


**Fig. A.28.:** Potential permafrost extent obtained with RF classification on spatially sampled data.

## Random forest uncertainty maps



**Fig. A.29.:** Uncertainty map obtained with RF classification on randomly sampled data.



**Fig. A.30.:** Uncertainty map extent obtained RF classification on spatially sampled data.



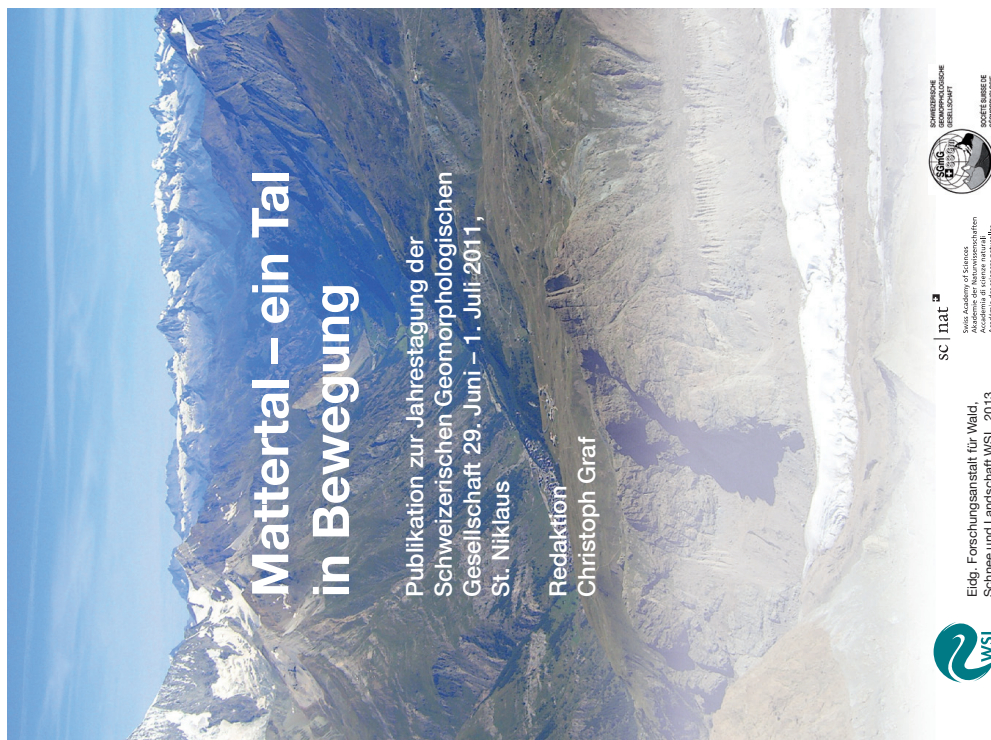


## Published papers

# B

### Publication:

Deluigi, N., & Lambiel, C. (2013). PERMAL: a machine learning approach for alpine permafrost distribution modeling. Schweizerischen Geomorphologischen Gesellschaft pp. 47-62, in Graf C. (eds.) *Mattertal - ein Tal in Bewegung. Jahrestagung der Schweizerischen Geomorphologischen Gesellschaft 29. Juni - 1. Juli 2011, St. Niklaus*, Eidg. Forschungsanstalt WSL.



### Context

The potential of the employment of a machine learning approach for permafrost distribution modelling was investigated in this first paper and a permafrost map that covered a small portion of the region of interest of this thesis was produced. This achievement was obtained by classifying permafrost data with Radial Basis Function Support Vector Machines.

This paper was published in the Proceedings of the Swiss Geomorphological Society conference held in 2012. It presents the main results of Deluigi (2012), a Master thesis presented at the University of Lausanne (Switzerland) by the author of the

present manuscript. The acronym PERMAL (*PERmafrost modelling with MACHine Learning*) was proposed for the first time and successively employed in the present research. PERMAL appeared to be reliable for the prediction of the high discontinuity of mountain permafrost, with a permafrost extent in accordance with the field observations. This preliminary study proven the potential of a data-driven approach and set the basis of further investigations, favouring the submission of a proposal to the Swiss National Science Foundation.

## Abstract

Most of the existing models of alpine permafrost distribution show a direct correlation between the permafrost occurrence and the increase in altitude. This may be correct at a regional scale, but it is often not valid at a more local scale, because of the high spatial discontinuity of alpine permafrost. For instance, the fact that permafrost is usually present only in the lower part of talus slopes has never been modeled. This paper presents a new model of alpine permafrost distribution that includes data obtained from field investigations carried out on various alpine landforms. The main goal of the study was to develop a model as reliable as possible at the local scale and to test the potential of an innovative approach in the field of permafrost modeling: Support Vector Machines (SVMs). This method is based on machine learning and provides a classification of samples produced by learning statistical dependencies between the studied phenomenon and other variables. This technique was used to model the spatial permafrost distribution in sedimentary landforms. In a second step, the lower limits of permafrost in rock walls, obtained by field measurements by PERMOS, were added to the model.

## Keywords

Mountain permafrost, modeling, machine learning, Support Vector Machines, Swiss Alps.

## Introduction

Nowadays, permafrost mapping has a fundamental relevance for natural hazard prevention as for the design and the maintenance of infrastructures in high Alpine regions. As permafrost is invisible at the ground surface, its spatial modeling is one of the most important tasks of alpine permafrost research. Initial efforts were proposed by Haeberli (1975) with the so-called empirical "rules of thumb", which considered the relationship between some topographic parameters and permafrost occurrence and were successively implemented in a GIS environment by Keller (1992). Later, other spatial models were proposed by using rock glacier inventories (e.g. Imhof,

1996b; Delaloye and Morand, 1997; Lambiel and Reynard, 2001; Baroni et al., 2004; Seppi et al., 2005), potential direct solar radiation (e.g. Funk and Hoelzle, 1992; Hoelzle and Haeberli, 1995; Frauenfelder, 1998; Nyenhuis et al., 2005), near ground surface and borehole temperatures (e.g. Gruber et al., 2004b; Etzelmüller et al., 2006; Etzelmüller et al., 2007; Allen et al., 2009) or snow basal temperature (BTS) (e.g. Keller et al., 1998; Gruber and Hoelzle, 2001; Ebohon and Schrott, 2008). These models generally offer a good overview of the potential spatial distribution of mountain permafrost at the regional scale. Other more process oriented models (e.g. Stocker-Mittaz et al., 2002) succeeded to simulate with good accuracy the ground temperature at the point scale, but the high number of input parameters needed prevent a regionalization of the information. All in all, there is a lack of models able to simulate the strong heterogeneity of mountain permafrost at the local scale (e.g. Lambiel and Pieracci, 2008; Scapozza and Lambiel, 2013, this volume). The majority of existing models indicate a general correlation between permafrost occurrence and the increase of altitude (Keller, 1992; Imhof, 1996b; Gruber and Hoelzle, 2001; BAFU, 2005; Boeckli et al., 2011). If this may appear correct at a regional scale, it is not always valid at a more local scale, because of the high spatial discontinuity of the permafrost extension. Effectively, recent studies have shown that alpine permafrost is usually present only in the lower half of talus slopes in relation to the so-called chimney effect, allowing air circulation through coarse blocky surfaces (Delaloye et al., 2003b; Morard et al., 2010; Scapozza et al., 2011).

The first goal of this study is to propose a new model aiming at integrating and simulating the heterogeneity of mountain permafrost, and thus to propose a model as reliable as possible at the local scale. Accordingly, a machine learning approach was adopted (cf. 3.1). Thus, the second objective of the study is to test the potential of this innovative approach in the field of permafrost modeling.

## Study area and known permafrost distribution

For this study, we choose the entire Swisstopo 1:25'000 "Rosablanche" topographic map to develop and to test the model. The research area is located in the Western Swiss Alps and it covers three main valleys: Bagnes, Nendaz and Hérémece. Since 1998, the Mont Gelé – Mont Fort region, situated in the western part of the map (Verbier/Nendaz area), has been studied by several field campaigns conducted by the geography institutes of the universities of Lausanne and Fribourg. The first attempts to estimate the potential permafrost distribution were proposed by Reynard (1996), Reynard et al. (1999), Lambiel and Reynard (2001) and Lambiel and Delaloye (2004), thanks to geomorphological mapping, rock glacier inventoring and 1-D resistivity prospecting. During the last decade, various methods such as electrical resistivity tomography, seismic refraction, ground (surface) temperature

measurements and borehole logging were used to map the permafrost distribution and to characterize the ground stratigraphy in different landforms such as rock glaciers, talus slopes and moraine deposits (e.g. Delaloye et al., 2003a; Reynard et al., 2003; Marescot et al., 2003; Delaloye, 2004; Delaloye and Lambiel, 2005; Lambiel, 2006; Lambiel and Pieracci, 2008; Lambiel and Schuetz, 2008; Hilbich et al., 2009; Scapozza et al., 2011; Scapozza, 2013). As we will describe in the next chapter, the modeling approach chosen for this study requires a calibration based on examples that indicate the permafrost occurrence. The availability of data for the Mont Gelé-Mont Fort region made the Rosablanche topographic map the best choice for this purpose.

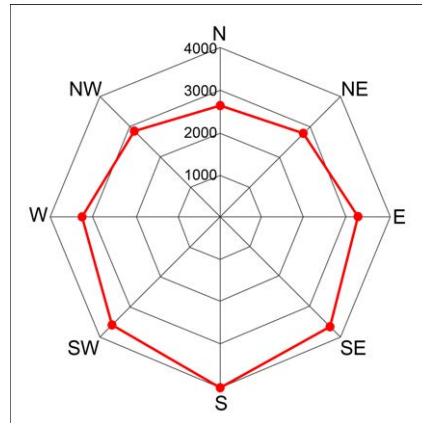
## Methodical background

### Learning from data

The last decades were characterized by the development of numerical models based on physical or statistical approaches. The relevance of these tools, as help for decision-making, became crucial since the availability of a wide number of spatial data and measurements, which have increased due to technological improvements in automated environmental monitoring. Rapidly, the research field of statistical machine learning grew, permitting the analysis of huge volumes of data and to discovery of the dependencies hidden inside them (Kanevski et al., 2009). The application of the binary classification to machine learning was one of most important tasks introduced by the Statistical Learning Theory (Vapnik, 1998). Without, a priori, any users' assumptions about phenomena distribution, data-driven algorithms should search for a decision function discriminating all samples of a dataset in two binary classes (such as "absence" or "presence" of a phenomenon). The classification is produced by learning statistical dependencies between the studied phenomenon and other variables in a so-called training samples dataset. These techniques should supply to accurate decision boundaries, taking into account non-linear solutions and misclassifications of observed data samples.

For this study, a technique providing these requirements and considering the complexity of the permafrost distribution was necessary. This complexity is linked to a high amount of variables (predictors), which lead to a high dimensional feature space of factors related to the permafrost occurrence. Therefore, Support Vectors Machines (SVMs) were chosen because of their suitability for high-dimensional datasets, avoiding model overfitting and allowing a probabilistic interpretation of the outputs through a continuous decision function. However, this method requires abundant examples that indicate the presence of permafrost. Since a good set of training samples indicating the permafrost distribution was unavailable for rock walls, SVMs were only used to predict the permafrost distribution in sedimentary landforms such

as rock glaciers and talus slopes. The lower limits of permafrost in rock walls were obtained by field measurements from the Swiss Permafrost Monitoring Network (PERMOS, 2009) (Fig. B.1) and added to the model in a second step.

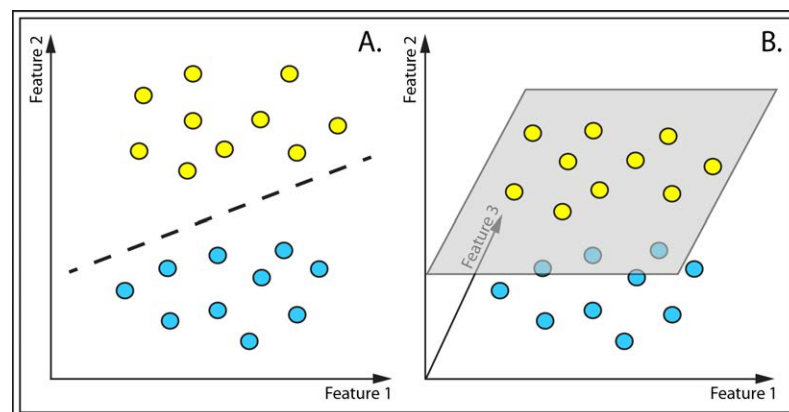


**Fig. B.1.:** Permafrost lower limits in rockwalls, resulting from PERMOS (2009) field measurements.

## Support Vector Machines

### Main concepts

The main concept of SVMs introduced by Vapnik (1998) indicates that, when dealing with a problem in which different objects have to be divided in two categories by placing a discriminating boundary, the most intuitive option is to draw a separating line. In an N-dimensional space, the line becomes a hyperplane (Fig. B.2).



**Fig. B.2.:** Linear separable samples in 2-D (A) and 3-D (B) features space.

With a SVMs approach, we must firstly suppose that the data samples  $(x_1, y_1), \dots, (x_N, y_n)$ , where  $x$  are the input features (e.g. see Table B.1) and  $y \in \{-1; +1\}$  the binary labels (e.g. absence and presence of permafrost), are linearly separable. This allows samples to be classified into two binary classes. The SVMs algorithm aims to deter-

mine the function that maximizes the distance between the training points and the hyperplane (the so-called margin  $\rho$ ). The linear decision function is defined as

$$f(x) = wx + b \quad (\text{B.1})$$

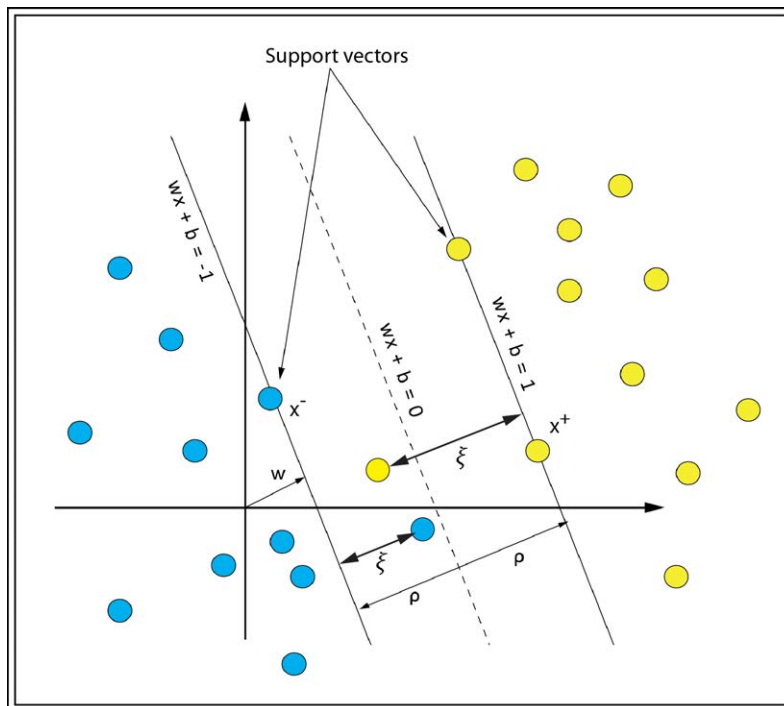
where  $w$  is a weighing vector which needs to be optimized along with the scalar  $b$  in order to maximize the margin  $\rho$ . The maximal margin is obtained by minimizing  $\frac{1}{2} \|w\|^2$ . Moreover, the sign of the function  $f(x)$  determines the class in which a predicted sample belongs. The optimization of the parameters  $w$  and  $b$  is a quadratic programming problem with linear constraints and unique solution. Furthermore,  $w$  is defined as a linear combination of the training samples  $x_i$ , in which almost the total of them have a weight  $\alpha_i$  equal to zero:

$$w = \sum_{i=1}^N \alpha_i y_i x_i \quad (\text{B.2})$$

The solution of the optimization problem allows the SVMs decision function to be formulated as

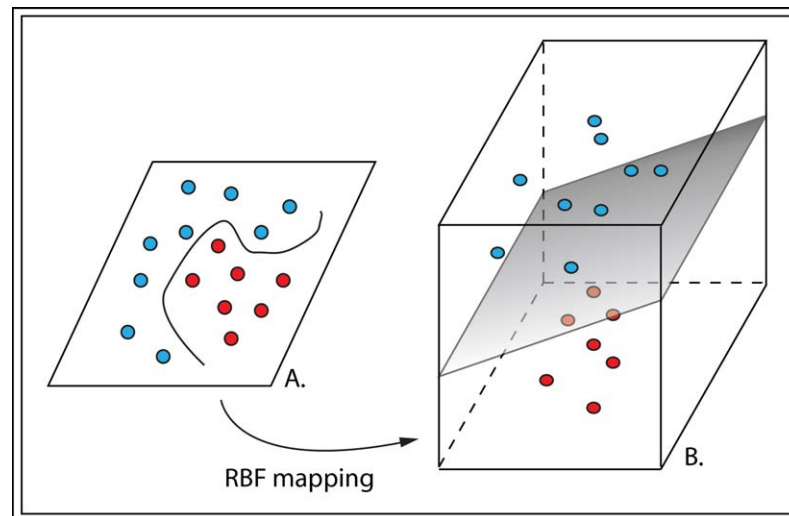
$$f(x) = \sum_{i=1}^N \alpha_i y_i x x_i + b \quad (\text{B.3})$$

All nearest samples to the decision boundary are the only ones that contribute to the maximum margin solution. They have a non-zero weight and they are called Support Vectors (Fig. B.3).



**Fig. B.3.:** binary classification of two groups of samples by maximizing the margin  $\rho$  and allowing misclassifications.

A linearly separable data samples case, where the two classes are not overlapping, is an ideal situation one will rarely be dealing with. In fact, data could be noisy and it is not always possible to avoid training errors when drawing a separating line. The SVMs classifier accounts for overlapping data with a soft margin adaptation: slack variables  $\xi_i$  are assigned to noisy samples lying outside their class margin. Therefore, at this point, a coefficient that affects the trade-off between complexity and proportion of non-separable samples is required and it must be selected by the user (Cherkassky and Mulier, 2007). For this reason, a so-called cost hyperparameter  $C$  is added to the optimization in order to keep a balance between empirical error and the minimization of  $\frac{1}{2} \|w\|^2$  for the finding of the largest possible margin  $\rho$ .



**Fig. B.4.:** RBF mapping allowing non-linear classification in a linear space to be solved.

Up to this point, we have seen how a linear decision function can be optimally applied in order to classify our examples with binary labels. When dealing with datasets where the input/output relationships are non-linear (Fig. B.4.A), we need to find a more clever way to discriminate the two classes. The tricky idea is to map the dataset into a space of higher dimensions and then to perform the well-known linear separation on the transformed data (Fig. B.4.B), rather than applying complex decision functions directly on the initial dataset. In this case, a “kernel trick” is used in equation B.3 to substitute the dot product between the input vector of a sample  $x$  and all training samples  $x_i$  with a kernel function  $K(\cdot, \cdot)$ , allowing non-linear decision boundaries to be determined. Accordingly, the final formulation of the decision function for a classification task takes the form:

$$f(x) = \sum_{i=1}^N \alpha_i y_i K(x, x_i) + b \quad (\text{B.4})$$

Among the available kernel functions, a user’s choice often falls on the Gaussian Radial Basis Function (Gaussian RBF) kernel  $K(x, x_i) = e^{-\frac{(\|x-x_i\|)^2}{2\sigma^2}}$  because of the



simple geometrical interpretation it provides. The parameter  $\sigma$ , controlling the bandwidth of the Gaussian surface centered on vector  $x$ , needs to be tuned using a validation dataset.

## Data preparation

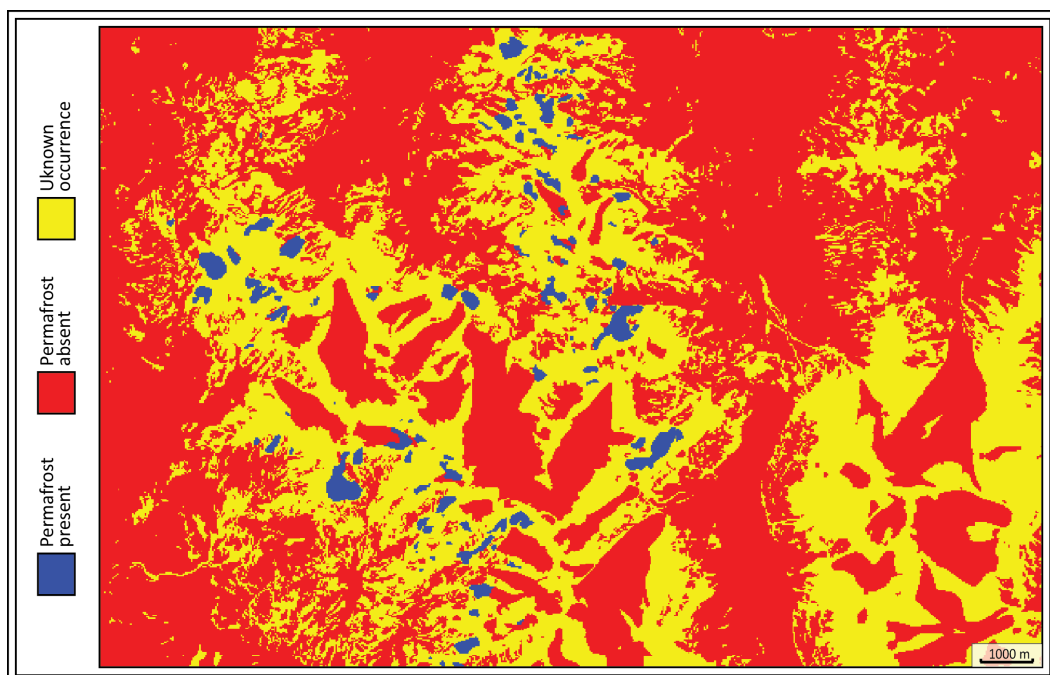
For this case study a pool of 15 features was retained (Table B.1). Some variables are strictly related to permafrost presence or absence, but for some of them a generalization was necessary. This was, for example, the case for glacier forefields, in which permafrost is often restricted to the lateral and frontal margin (e.g. Reynard et al., 2003). For simplicity, we decided to label this variable as an indicator of permafrost absence.

Altitude, potential direct solar radiation, aspect, slope and curvature were simply extrapolated from the Swisstopo's digital elevation model with a 25x25 meter resolution. Lakes, human infrastructures, current glacier extension and mineral/vegetation-covered surfaces were extracted from the Swisstopo's primary surfaces map. From these variables, other features were calculated. The mean annual air temperature (MAAT) was obtained for the entire Rosablanche map with a linear regression involving daily temperatures measured at Les Attelas ENET meteorological station, according to Bouët (1978)'s formulas. Moreover, glacier forefields were extrapolated by subtracting the current glacier extension from the Little Ice Age glacier extension (Maisch, 2000). Talus slopes and rock wall mapping was produced by a DEM-based geomorphometric approach proposed by Loye et al. (2009), which permitted the identification of these features according to a slope angle distribution analysis. Intact rock glaciers (actives and inactives) were extracted from the region's rock glacier inventory (Lambiel, 1999; Reynard et al., 2003). The dataset was completed by empirical data from field campaigns, which show permafrost presence in talus slopes.

Variable	Permafrost occurrence	Description
Altitude	-	Elevation in meters
Aspect	-	Exposition in degrees
Slope	-	Slope angle in degrees
Potential direct solar radiation	-	Direct solar radiation for the snow-free period (1 <sup>st</sup> July-31 <sup>th</sup> October)
MAAT	-	Mean annual air temperature (in degrees) for the 1996-2009 period
Curvature	-	Topographic curvature in degrees
Glacier	Unlikely/Rare	Binary variable indicating glacier presence/absence ('1' or '0')
Intact rock glacier	Probable	Binary variable indicating rock glacier presence/absence ('1' or '0')
Human infrastructure	Unlikely	Binary variable indicating human infrastructure presence/absence ('1' or '0')
Lake	Unlikely	Binary variable indicating lake presence/absence ('1' or '0')
Glacier forefield	Unlikely/Rare	Binary variable indicating glacier forefield presence/absence ('1' or '0')
Mineral-covered surface	Possible	Binary variable indicating mineral-covered surface presence/absence ('1' or '0')
Rock wall	Possible	Binary variable indicating rock wall presence/absence ('1' or '0')
Vegetation	Unlikely	Binary variable indicating vegetation presence/absence ('1' or '0')
Talus slope	Possible	Binary variable indicating talus slope presence/absence ('1' or '0')

**Tab. B.1.:** The list of features used for the permafrost distribution modelling in sedimentary landforms.

The input vectors were composed of several groups of features for each location. In order to reduce the dimensionality, all values were standardized (z-score). The goal of the SVMs training process was to generate discriminative predictants related to permafrost occurrence (presence, absence or uncertainty). On a total of 295'680 samples (pixels), 6'193 of them indicated presence of permafrost (2%) and 182'173 indicated absence of permafrost (62%). The last 107'314 samples (36%) were unlabelled and needed to be classified (Fig. B.5). Only a small portion of samples were randomly extracted from the two first categories and used for the training process. This means that SVMs predicted the permafrost occurrence for the entire study area. Remaining samples from these categories were used to validate the model. In fact, the modeling process involved a random partitioning of the original dataset into 3 sub-datasets: a training set to fit the model and maximize margins, a validation set to estimate the prediction error for model selection (tuning hyperparameters) and a test set, used for assessment of the generalization error of the final chosen model. This process was repeated 30 times using different random sample pools and different proportions between the sub-datasets.



**Fig. B.5.:** Labeled samples (pixels) used by Support Vector Machines (SVMs) for the simulation.

## Prediction process and probabilistic SVMs output interpretation for decision support and model accuracy

By using the training set to find the final decision function, the classification was possible. The so-called *prediction process* was then executed, in which each unlabelled pixel was placed under or above the margin  $\rho$  (hyperplane). During this simulation phase, the test sets were also used to validate the final model and to calculate its error.

In the Statistical Learning Theory, Vapnik (1998) shows that the SVMs approach usually provides an excellent generalization ability and accurate predictions in high dimensional spaces. Moreover, this technique usually avoids over-fitting and accounts auto-correlation between similar input features. However, decision function output labels  $y = \{-1; +1\}$  are difficult to interpret and employ for decision making. A smooth confident measure  $p : 0 < p(y = \{-1; +1\} | x) < 1$  is computable by using the Platt (1999) sigmoid function:

$$p = \frac{1}{1 + \exp(A \cdot f(x) + B)} \quad (\text{B.5})$$

where A and B are the parameters to tune on the validation set in order to avoid model over-fitting.

To evaluate the quality of this modeling, two approaches were chosen: the overall accuracy ratio (OAR) (equation B.6) and the area under the receiver-operating characteristics (AUROC). The first ratio, which ranges between 0 and 1, reports the number of correct *predictions* over the total number of points, indicating the reliability of the model. The receiver-operating characteristic (ROC) plots in a 2-dimensional graph the false positive rate (horizontal axis) (equation B.7) and the true positive rate (the vertical axes) (equation B.8). It results in an interesting quality measure because a SVMs binary classification is executed according to a defined threshold resulting in a positive class label if the score is above the threshold  $t(f(x) > t)$ , or in a negative one if the value is lower than  $t(f(x) > t)$ . When computing these TP/FP rates for the classifications obtained with thresholds varying from their minimal to maximal values, we will be able to plot a point (FP rate, TP rate) associated with each selected threshold (Fawcett, 2006; Hamel, 2009). AUROC ranges between 0.5, which means a random model behavior, and 1 (perfect model).

$$\text{Overall accuracy ratio} = \frac{\text{hits} + \text{correct negatives}}{\text{total}} \quad (\text{B.6})$$

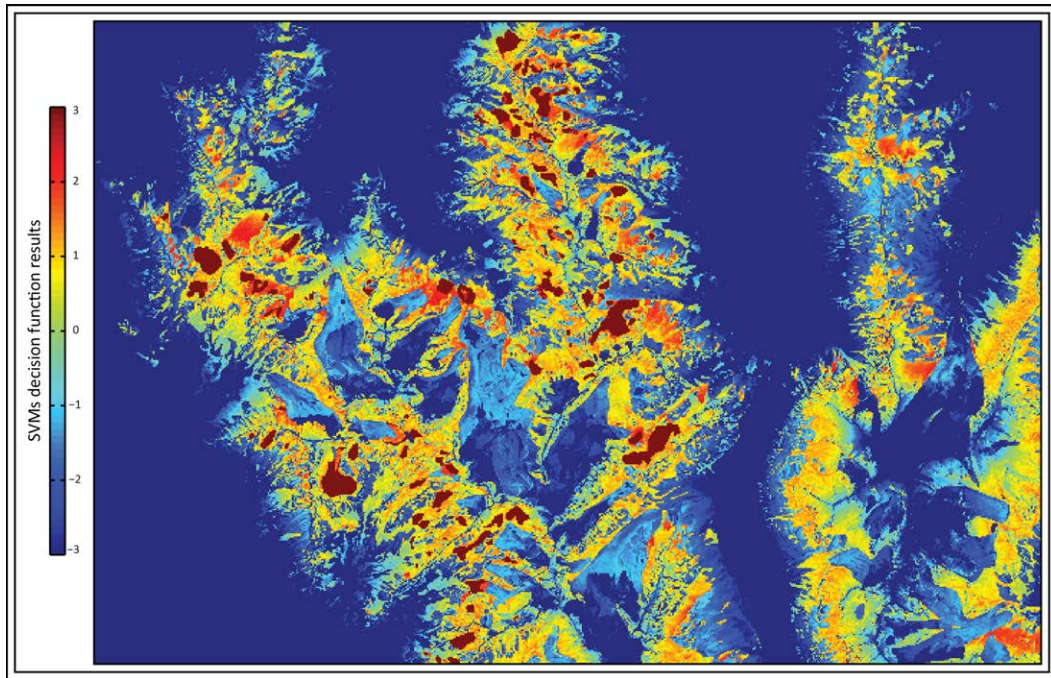
$$\text{False positive rate} = \frac{\text{false alarm}}{\text{false alarms} + \text{correct negatives}} \quad (\text{B.7})$$

$$\text{True positive rate} = \frac{\text{hits}}{\text{hits} + \text{misses}} \quad (\text{B.8})$$

## Results and discussion

### The Rosablanca permafrost extension map

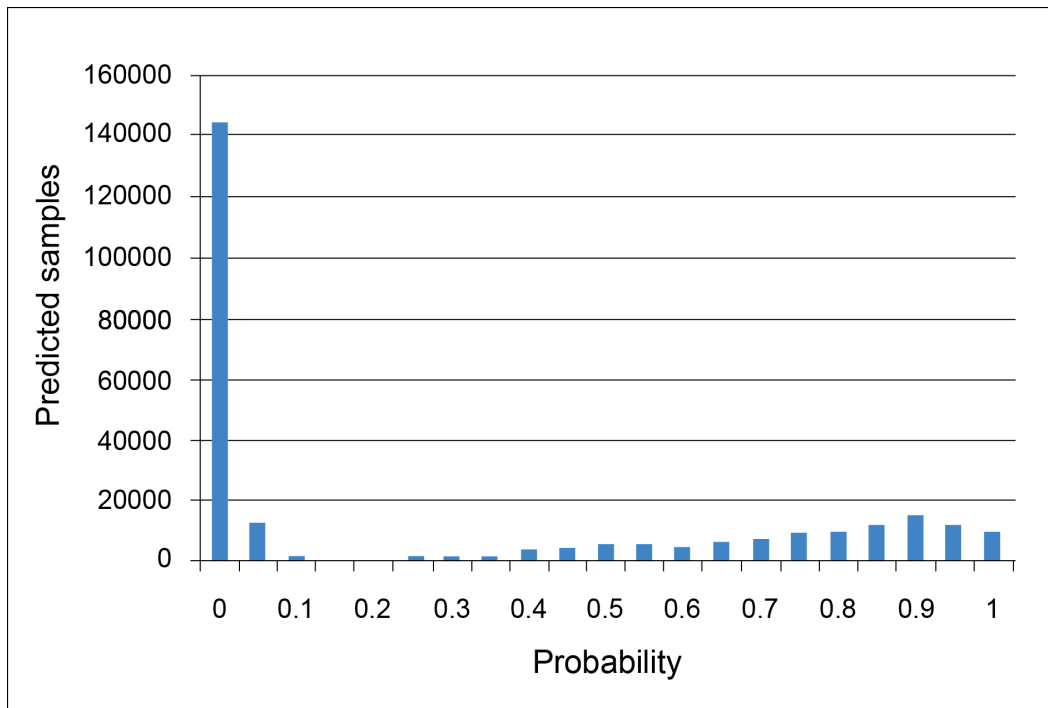
The SVMs final model indicates all values calculated by the decision function (Fig. B.6). Values  $f(x) < -1$  or  $f(x) > +1$  correspond to samples that are labeled as a negative or a positive class, or more precisely as a pixel in which permafrost is absent or present. However, all values predicted in the  $[-1; +1]$  range (between margins) cannot be classified with certitude. The Platt sigmoid function B.5 is then required in order to allow a probabilistic interpretation of these samples.



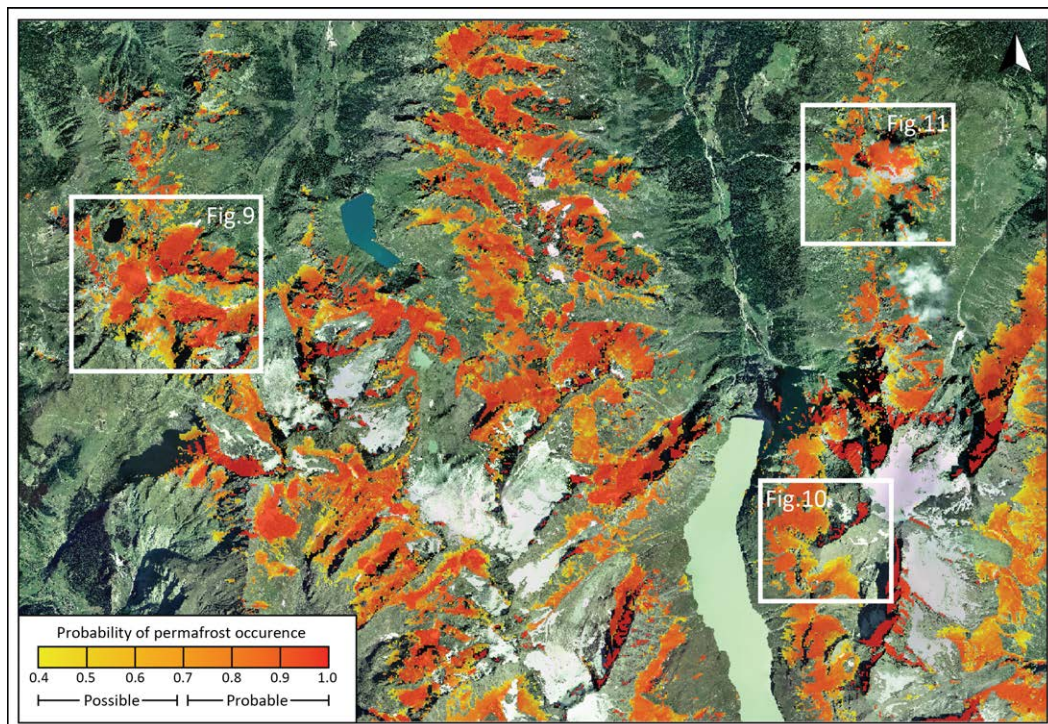
**Fig. B.6.:** Support Vector Machines decision function results. Pixel values require a probabilistic transformation to be interpreted.

Resulting occurrence probabilities (Fig. B.7) were classified into three categories. Predictions below 0.4 were classified as "permafrost absent". The remaining 0.6 was divided into two equal classes: from 0.4 to 0.7 predictions were labelled as "permafrost possible"; from 0.7 to 1.0 as "permafrost probable". Figure B.8 illustrates the final model of the potential permafrost extent, called PERMAL (for "Permafrost Machine Learning"), which was produced by merging the probabilistic SVMs output

with the permafrost map for rock walls resulting from PERMOS (2009) measured lower limits.



**Fig. B.7.:** Distribution of permafrost occurrence's probabilities resulting from the Platt sigmoid function.

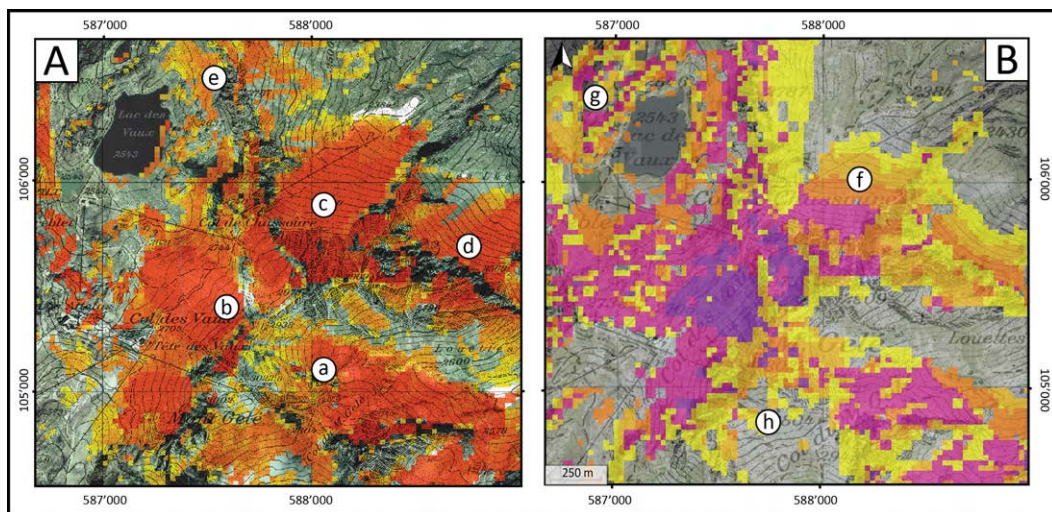


**Fig. B.8.:** The final potential alpine permafrost distribution map for the entire Rosablanche topographic map area.

## PERMAL result analysis

In order to provide a quantitative analysis, the OAR and the AUROC were calculated, providing values of 0.967, respectively 0.975. These rates indicate an almost perfect model, which is usually synonym of a model over-estimation. Unluckily, SVMs are a "black-box" type modeling, a disadvantage the machine learning methods are often criticized for. This means that it is impossible to understand in what way predictors influenced the permafrost occurrence without the embedding of additional algorithms. As a result, we must suppose that the permafrost extent is probably less important than the modeled one.

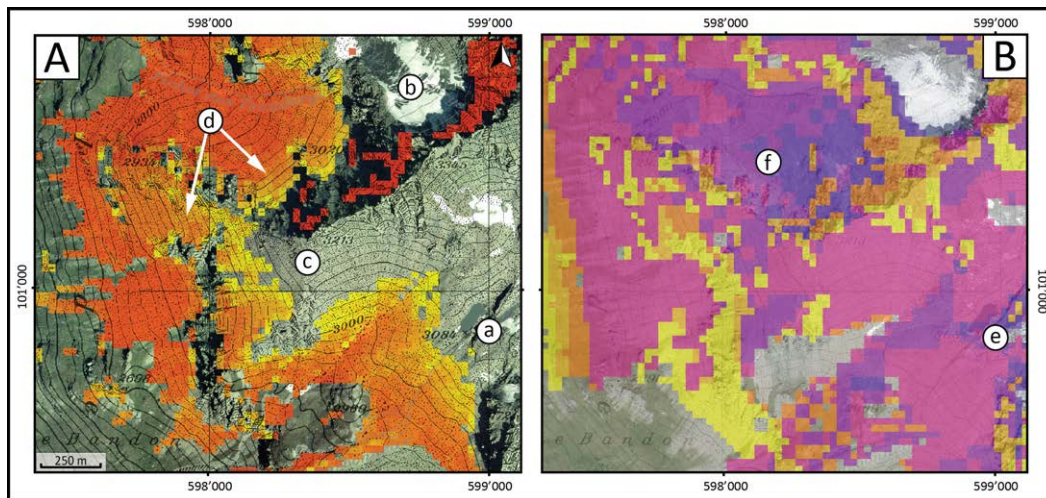
A qualitative analysis was carried out on selected regions of interest. The Mont Gelé sector (Fig. B.9.A) provides good examples of how SVMs predicted the permafrost occurrence for different landforms. Training samples, such as rock glaciers of the Yettes Condja valley (a), periglacial lobes at Lac des Vaux (b) and the Lapires talus slope (c) were correctly recognized. On the other hand the decrease of permafrost occurrence with altitude in the talus slopes was not modeled correctly (c, d). Another result is that permafrost boundaries were automatically limited to mineral-covered surfaces, according to the Swisstopo's primary surfaces mapping (e).



**Fig. B.9.:** PERMAL simulation (A) versus the BAFU one (B) for the Mont Gelé sector. Refer to text for the meaning of the small letters.

Darbonneire and Tsauderys glacier cirques (Fig. B.10.A) supply supplementary examples that illustrate the way SVMs modeled the permafrost extension. These sectors are characterized by extended mineral-covered surfaces and are particularly interesting because they do not contain any training samples. For the entire area, the permafrost distribution was simulated by the SVMs calculated decision function. The PERMAL result appears correct for all eight aspects. Northern exposed slopes are characterized by a more important permafrost occurrence than the south-facing

ones. Furthermore, permafrost was predicted as absent for the Darbonneire glacier forefield (a) and for the Tsauderys glacier surface (b). For most talus slopes, permafrost occurrence was simulated as probable. However, some contrasts can be observed between the lower and the upper parts of some slopes. It is especially the case in the center of the map (c), where PERMAL predicted a total absence of permafrost for the upper part of these regular slopes. In sector d, even though the lower part was not clearly separated from the upper part of the slope, a decrease in the probability of permafrost occurrence with altitude was modeled. Thus, for these sectors, SVMs appears to be modeling correctly the atypical permafrost distribution for talus slopes.

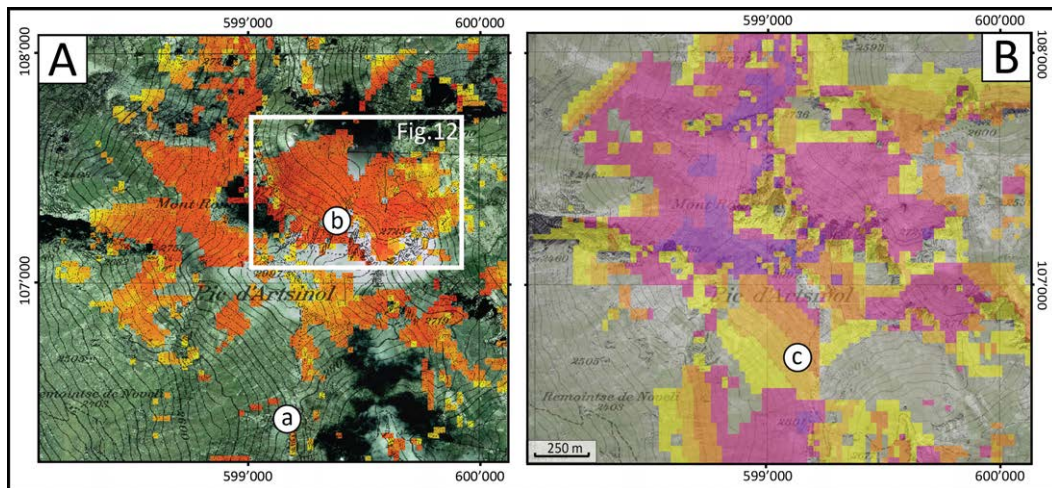


**Fig. B.10.:** PERMAL simulation (A) versus the BAFU one (B) for the Darbonneire and Tsauderys glacier cirques sector. Refer to text for the meaning of the small letters.

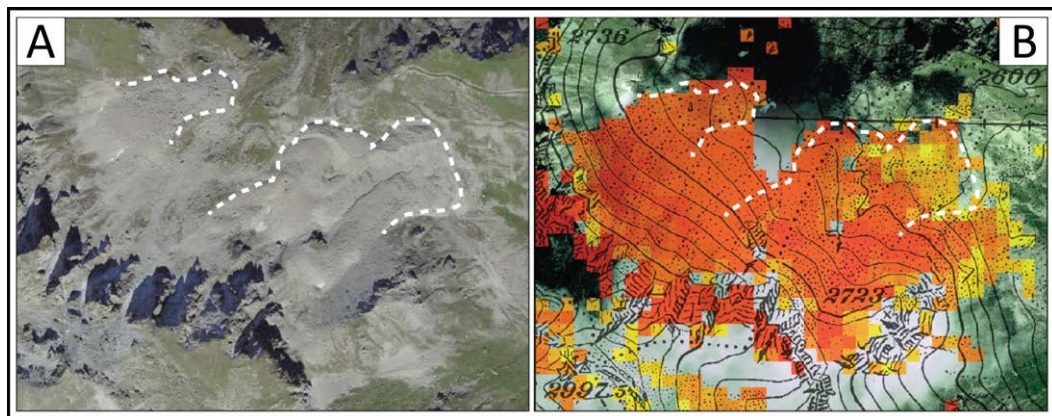
The last example illustrates the permafrost distribution for the Pic d'Artsinol area (Fig. B.11.A and B.12). PERMAL simulated correctly a low occurrence of permafrost in vegetation-covered slopes and probabilities above 40% for mineral-covered surfaces. The two rock glaciers located on the Northeastern side of the Pic d'Artsinol (b) were not provided to the machine because they were not included in the rock glacier inventory used for the training process. As indicated in figure B.12, SVMs were capable of predicting the boundaries of these landforms with accuracy.

## PERMAL – BAFU comparison

In order to show the PERMAL qualities and limitations, a comparison with the BAFU model is proposed. In this model, two different approaches were used for sedimentary landforms and for bedrock. The modeling of permafrost in sediment was based on the "rules of thumb" of Haeberli (1975) and implemented in the



**Fig. B.11.:** PERMAL simulation (A) versus the BAFU one (B) for the Pic d'Artsinol sector. For the significance of the small letters see the text. The highlighted zone corresponds to the zoom in Fig. B.12.



**Fig. B.12.:** Rock glacier limits (A) and PERMAL simulation (B) in the northern part of the Pic d'Artsinol.

PERMAKART model by Keller (1992). For bedrock, the modeling was based on a physical model developed by Gruber et al. (2004b). In the end, the BAFU map is built from different altitude thresholds above which permafrost may be found.

For the Mont Gelé sector (Fig. B.9), PERMAL and the BAFU model show quite different results. For example, in the Lapires talus slope, the BAFU map shows a direct correlation between permafrost occurrence and the increase of altitude (f), contrary to PERMAL, which presents the same probability for the entire slope. Despite the fact that the contrast between the upper part and the lower part has not been modeled by PERMAL, this result is nevertheless more in accordance with the field data (Scapozza, 2013)). Moreover, PERMAL considers the boundaries between vegetation and mineral-covered surfaces in a better way than the BAFU model. On the one hand, this provides an extension map in which permafrost is



not present in vegetation-covered surfaces, as for slopes around the Lac des Vaux (g). On the other hand, PERMAL sometimes indicates doubtful permafrost presence for sites characterized by mineral cover. The south-facing slope of the Mont Gelé (h) is a good example in which we observe that the BAFU map, because of altitude thresholds, does not show any permafrost occurrence, contrary to PERMAL that simulates permafrost presence, probably because it recognizes a mineral-covered slope.

For both Darbonneire-Tsauderys glaciers cirques and Pic d'Artsinol sectors (Fig. B.10, B.11 and B.12), the BAFU model is more optimistic and the permafrost extension is more important than what PERMAL simulates. It's interesting to observe that samples labeled as "permafrost probable" are almost the same for both models. Nevertheless, some important differences need to be evoked. For the first site, the BAFU model points out permafrost presence in the Darbonneire's glacier forefield (e). Also, contrary to PERMAL, this model does not consider a lower permafrost probability for the upper half of talus slopes (f). This sector clearly illustrates the consequence of the altitude thresholds used by the BAFU model for the permafrost simulation: gradually as the altitude increases, the permafrost probability becomes higher. In a similar way, these thresholds are evidently shown in the Pic d'Artsinol area (Fig. B.11). South of the summit, different bands of probability are simulated and their altitudes vary according to the aspect of the slope (c). Moreover, for this site, we observe that the BAFU model predicts permafrost presence also for vegetation-covered slopes, which is generally not the case with PERMAL (a).

## Conclusion

PERMAL appears to be thorough in the simulation of the high discontinuity of mountain permafrost. Results revealed a rather good reliability with the field data and a good discrimination between mineral and vegetation surfaces, where permafrost is often absent. An important advantage of PERMAL is that it does not use any altitude thresholds, conversely to the BAFU model. In the end, comparisons between both models showed that PERMAL is less optimistic than the BAFU model. Presented examples indicate that results of PERMAL may be more consistent with the field data than those of the BAFU model.

Nevertheless, the presented analysis indicated some limitations. Firstly, the "black-box" model characterization of SVMs does not allow the weights of the variables and their importance for the permafrost occurrence to be understood. These predictors were chosen in accordance to field data and consecutively included in the dataset. However, more variables could be added. For example, in order to improve the results and to better model the contrast of permafrost occurrence between the lower and the upper part of talus slopes, two new predictors should be included

in the dataset ("lower part" and "upper part" of the slope). These discriminating variables could help SVMs to recognize the pixel localization in the slope. Thus, a more significant occurrence of permafrost in the lower part of the slopes could be simulated. In fact, this contrast was modeled only for a couple of examples. Besides, boundaries that describe vegetation and mineral-covered surfaces are precisely taken into account by PERMAL. Accordingly, Swisstopo's primary surfaces must be accurate to obtain a reliable result. Thus, next improvements should include more accurate data extracted from satellite and aerial images, and high resolution DEMs. Further development would also require more detailed information about grain size of mineral-covered surfaces, in order to predict a higher permafrost occurrence in coarse blocky slopes and to better simulate the strong heterogeneity of mountain permafrost at the local scale.



## Publication:

Deluigi, N., Lambiel, C., & Kanevski, M. (2017). Data-driven mapping of the potential mountain permafrost distribution. *Science of The Total Environment*, 590, 370-380.



Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: [www.elsevier.com/locate/scitotenv](http://www.elsevier.com/locate/scitotenv)



## Data-driven mapping of the potential mountain permafrost distribution

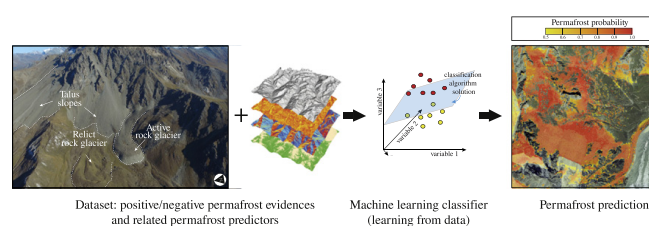
Nicola Deluigi\*, Christophe Lambiel, Mikhail Kanevski

*Institute of Earth Surface Dynamics, University of Lausanne, Lausanne, Switzerland*

### HIGHLIGHTS

- Investigation of an innovative permafrost distribution modelling approach
- Three classifiers belonging to statistics and machine learning were applied.
- Machine learning algorithms provide precise distribution maps at the micro-scale.
- Predicted permafrost distribution is in accordance with the field reality.

### GRAPHICAL ABSTRACT



## Context

This paper, published in the special issue *Mapping the Environment* (Science of the Total Environment, Pereira et al., 2018), assessed the quality of three machine learning algorithms (Logistic Regression, Radial Basis Function Support Vector Machines and Random forest) and compared the resulting permafrost distribution maps with the BAFU model. One of the main goals of this study was the investigation of the improvements brought by morphological predictors to the prediction of mountain permafrost distribution *at the micro scale*. A new dataset was thus built by including curvature indices that cover several resolutions, with the purpose of representing the topographical settings of convex and concave landforms, such as rock glaciers, moraines, small depressions, etc.

It is worth noting that model performances differ from the ones of the final version of PERMAL (presented in Chapter 5), since the dataset used did not include all the permafrost evidences presented in Chapter 3 and the validity domain did not considered glacier forefield and the hydrological network (see table 3.1).

## Abstract

Existing mountain permafrost distribution models generally offer a good overview of the potential extent of this phenomenon at a regional scale. They are however not always able to reproduce the high spatial discontinuity of permafrost at the micro scale (scale of a specific landform; ten to several hundreds of meters). To overcome this lack, we tested an alternative modelling approach using three classification algorithms belonging to statistics and machine learning: Logistic regression, Support Vector Machines and Random forests. These supervised learning techniques infer a classification function from labelled training data (pixels of permafrost absence and presence) with the aim of predicting the permafrost occurrence where it is unknown. The research was carried out in a 588 km<sup>2</sup> area of the Western Swiss Alps. Permafrost evidences were mapped from ortho-image interpretation (rock glacier inventoring) and field data (mainly geoelectrical and thermal data). The relationship between selected permafrost evidences and permafrost controlling factors was computed with the mentioned techniques. Classification performances, assessed with AUROC, range between 0.81 for Logistic regression, 0.85 with Support Vector Machines and 0.88 with Random forests. The adopted machine learning algorithms have demonstrated to be efficient for permafrost distribution modelling thanks to consistent results compared to the field reality. The high resolution of the input dataset (10 meters) allows elaborating maps at the micro scale with a modelled permafrost spatial distribution less optimistic than classic spatial models. Moreover, the probability output of adopted algorithms offers a more precise overview of the potential distribution of mountain permafrost than proposing simple indexes of the permafrost favorability. These encouraging results also open the way to new possibilities of permafrost data analysis and mapping.

## Keywords

Mountain permafrost, classification algorithms, machine learning, mapping.

## Introduction

Within the 21<sup>st</sup> century the Alpine environment is going to experience deep modifications of the cryosphere as a consequence of the increase in air temperatures and the modifications of precipitation regimes. Among the cryospheric components, mountain permafrost describes a ground with temperatures at or below 0°C for two consecutive years (Harris et al., 2009; Beniston et al., 2017b). Permafrost in rock walls and sedimentary accumulations may degrade as a consequence of the

climate change (Etzelmüller and Frauenfelder, 2009). A thickening of the active layer and a warming of the permafrost body can have various effects on mountain slope stabilities, such as an increasing rock fall activity (Gruber and Haeberli, 2007; Ravanel et al., 2010) or a rock glacier acceleration (Kääb et al., 2007; Roer et al., 2008; Delaloye et al., 2010; PERMOS, 2016), leading to an increase of the sediment transfer rates (Lane et al., 2007; Kobierska et al., 2011).

In the European Alps, the periglacial belt is generally marked by the absence of trees, a reduced vegetation cover (where existing essentially made of meadow, mosses and lichens), large volumes of sediment debris, steep slopes and rock faces. Although permafrost may affect all these different types of terrains, its unambiguous morphological manifestation only occurs in active rock glaciers, which are considered as the visible expression of mountain permafrost creep (Haeberli, 1985). Other permafrost indicators are thrust- or push-moraines, corresponding to frozen sediments deformed by the glacier advance during the Little Ice Age, whereas large areas in glacier forefields located in the periglacial belt appear to be unfrozen (Reynard et al., 2003; Harris and Murton, 2005; Kneisel and Kääb, 2007; Bosson et al., 2015). Talus slopes constitute other major landforms of alpine environments where permafrost is generally restricted to the lower half of the slope (e.g. Lambiel and Pieracci, 2008; Otto and Sass, 2006; Scapozza et al., 2011). It is also well established that terrains covered by alpine meadow are generally permafrost free (Haeberli, 1975). The distribution of mountain permafrost is thus extremely discontinuous in mountain areas (see also Ribolini et al., 2010; Otto et al., 2012).

The ability of modelling the spatial distribution of such a complex phenomenon became an important task for the alpine permafrost research during the last two decades. First empirico-statistical models were based on simple approaches (such as linear regression) and offered a good overview of the potential distribution of mountain permafrost at the regional scale (i.e. Hoelzle, 1994; Ebohon and Schrott, 2008; Avian and Kellerer-Pirklbauer, 2012). These models are generally thresholding the occurrence of permafrost on the basis of a restricted number of topographical and climatic parameters (i.e. altitude of rock glacier fronts for a given orientation) and are validated with measurements of the ground surface temperature, which may be subject to bias. The availability of an increasing amount of high resolution data (generally derived from high resolution digital elevation models) opened then the way to new complex statistical models able to deal with a large number of predictors (i.e. Boeckli et al., 2012; Schöner et al., 2012; Magnin et al., 2015; Azócar et al., 2016; Sattler et al., 2016). Although they offer a good overview of the permafrost distribution at local scale (i.e. scale of a valley side), these models do not reflect the great heterogeneity of the phenomenon at the scale of a specific landform (the micro scale; covering ten to several hundreds of meters).

To address the need of an improved prediction of the permafrost extent at the micro scale, we propose an alternative approach, which employs classification algorithms belonging to statistics and machine learning, namely Logistic regression, Support Vector Machines and Random forests. These algorithms can deal with complex and high dimensional datasets (Bishop, 2006) and they derive functional dependencies directly from data without appealing to physical models (Hastie et al., 2009). They have successfully been adopted for mapping the spatial distribution of several natural phenomena (i.e. Amatulli et al., 2013; Varley et al., 2016). In the periglacial research such techniques have been already used for geomorphological mapping (Luoto and Hjort, 2005), landform characterization (Marmion et al., 2008) or permafrost mapping using satellite images (Ou et al., 2016). Accordingly, we collected field observations indicating the known presence or the known absence of mountain permafrost and related topo-climatic data for a specific area of the Western Swiss Alps. The dataset built was analyzed and used to investigate the potential of machine learning techniques for mapping the high spatial discontinuity of mountain permafrost. Furthermore, as the potential permafrost distribution in rockwalls had already been successfully modeled in other studies (i.e. Gruber et al., 2004a; Noetzli et al., 2007; Magnin et al., 2015), the present work focuses only on sedimentary accumulations.

## Materials and methods

### Permafrost evidences and explanatory variables

This study was carried out in a sector of the Western Valais Alps (Switzerland) covering a regular grid of 588 km<sup>2</sup>, with more than 60% above the theoretical permafrost lower limit of 2500 m.a.s.l., delimiting the lower boundary of the periglacial belt in the area (Lambiel and Reynard, 2001).

We used evidences of known permafrost presence or absence collected since the mid 1990's by the Universities of Lausanne and Fribourg as training data for employed machine learning algorithms (figure B.13). These evidences have been obtained from two distinct sources:

- *Rock glacier inventories*. Permafrost presence or absence can be derived from rock glacier maps, based on their activity. Indeed, active or inactive rock glaciers suggest the existence of permafrost conditions, whereas relict ones indicate its absence (see Haeberli, 1985; Humlum, 1996; Barsch, 2012). For this study, we employed some existing inventories (Delaloye and Morand, 1998; Morand, 2000; Lambiel and Reynard, 2003), for which rock glaciers were mapped directly in the field. Some additional rock glaciers located within

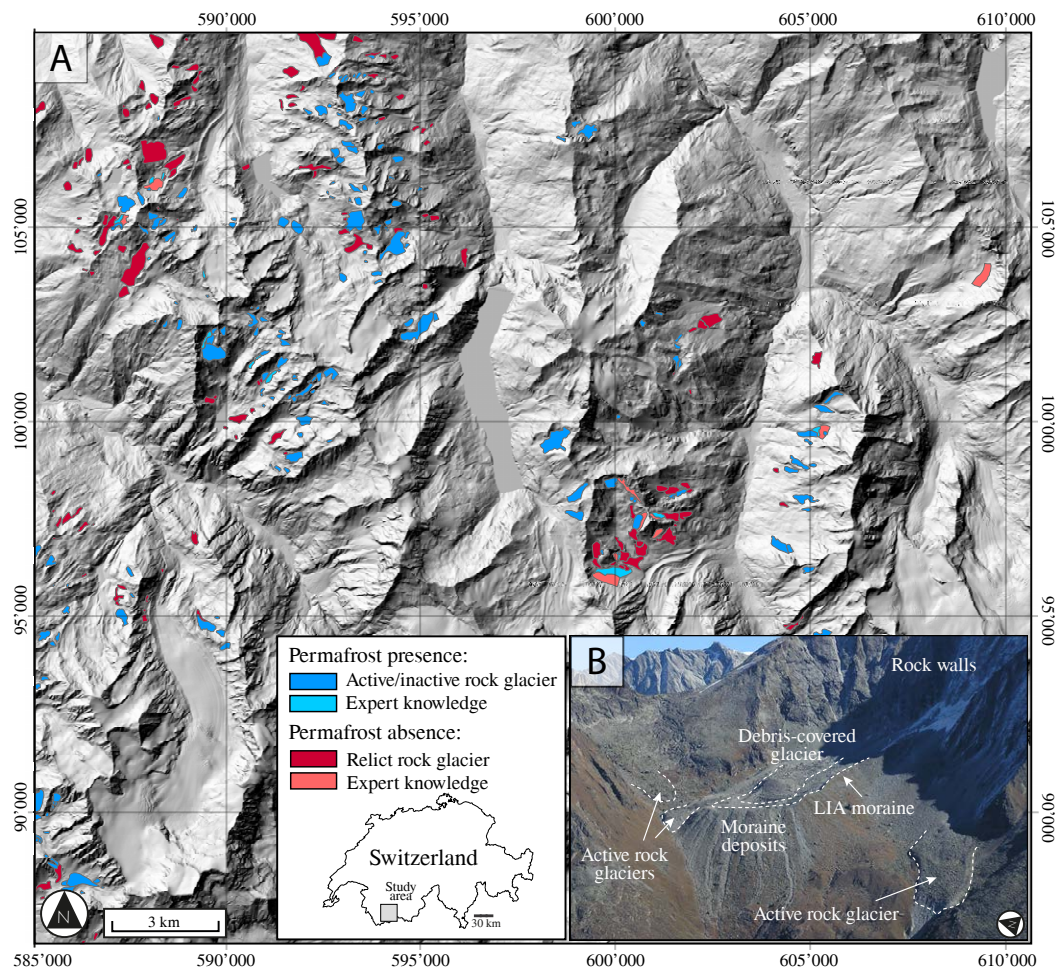
the study area were also added through ortho-image interpretation. All rock glacier limits were then corrected with a comparison with recent orthophotos (Swissimage, from swisstopo) and their activity was verified with the analysis of geomorphic signatures and InSAR signals (Delaloye et al., 2007; Barboux et al., 2014a).

- *Geoelectrical and thermal data.* Direct-current (DC) resistivity methods are well established tools for detecting permafrost in sediments (Hauck and Kneisel, 2008). Electrical resistivity tomography (ERT) is especially often utilized to detect ground ice and characterize frozen materials in permafrost environments (e.g. Hauck et al., 2003; Hilbich et al., 2009; Otto et al., 2012). In addition, permafrost can also be inferred from ground surface temperature measurements (Hoelzle et al., 1999; Carturan et al., 2015). Coupling geoelectrical and thermal data can thus improve the reliability of permafrost mapping. Following the procedure employed by Lambiel (2006, p. 95) and Scapozza et al., 2011, we compiled and combined geoelectrical and thermal data collected in the framework of different studies aiming at detecting and mapping ground ice in permafrost environments - mainly talus slopes and glacier forefields - of our study area (Marescot et al., 2003; Reynard et al., 2003; Delaloye, 2004; Delaloye and Lambiel, 2005; Lambiel, 2006; Delaloye et al., 2008; Lambiel and Pieracci, 2008; Scapozza et al., 2011; Scapozza, 2013; Staub et al., 2015). Completed by thermal measurements gathered for the Swiss Permafrost Monitoring Network (PERMOS, 2016) and by other unpublished projects, these data were used to map the permafrost extension in the prospected landforms. This provided to the classification algorithms additional training examples also located outside rock glaciers. Negative training observations (known permafrost absence) resulted not only from in-situ measurements indicating warm conditions or absence of ground ice, but also from expert knowledge. We particularly used the conclusions of Lambiel and Pieracci, 2008 and Scapozza et al., 2011 that showed the general absence of permafrost in the upper half of talus slopes.

For this study, we selected environmental variables that are commonly used in the field of permafrost modelling (e.g. Etzelmüller et al., 2001; Guglielmin et al., 2003; Boeckli et al., 2012) such as altitude, mean annual air temperature, aspect, terrain slope angle and potential incoming solar radiation. In addition, we computed the NDVI and the planar, profile and combined terrain curvature indices, which are morphometric predictors important in characterizing specific periglacial landforms such as rock glaciers or moraines. The relevance of selected predictions is presented in the following sub-sections:

- *Altitude:* Permafrost occurrence increases with the altitude at the regional scale, due to the decrease of the mean annual air temperature (MAAT) at





**Fig. B.13.:** (A) Extent of the study area (relief map) and localization of known permafrost evidences. (B) Typical Alpine periglacial landscape characterized by active rock glaciers, debris-covered glacier, talus slopes, moraine deposits and rock walls (Arolla valley, Valais; photo: R. Delaloye).

higher altitudes. As MAAT is calculated on the basis on a linear temperature gradient ( $-0.59^{\circ}\text{C} / 100 \text{ m}$  above 1500 m, for methodology see Bouët, 1978), these two predictors were extremely correlated (see figure B.14). Accordingly, to avoid redundancy, the adopted dataset excluded the MAAT variable and only considered the altitude. The latter is derived from the SwissAlti3D digital elevation model from the Swiss Federal Office of Topography (swisstopo). It is a spatial grid with an aperture width of 2 m above 2000 m.a.s.l. and it is produced by stereo correlation with 1 - 3 m average error. The density of at least 2 points/m<sup>2</sup> avoids noise in the data.

- Aspect: The terrain orientation is also considered extremely relevant for the permafrost presence / absence. Terrains with different aspects have different energy inputs due to a different radiation angle. The amount of energy per unit area can actually vary in a ratio of 1 to 10 between the sunny side and the shady side of a relief. Since terrain orientation is a circular variable (between

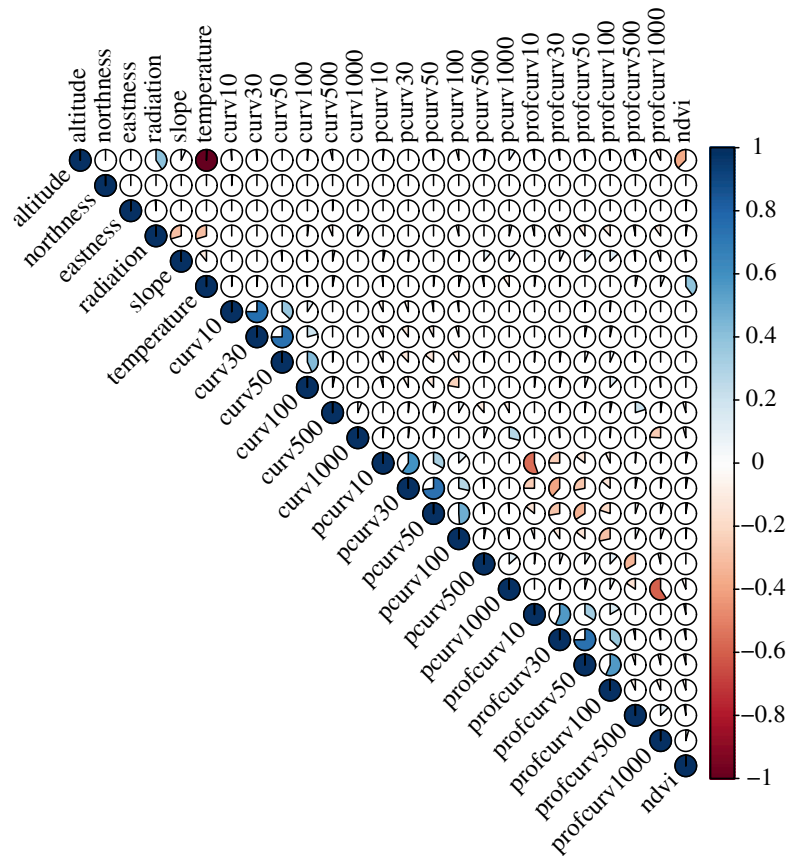
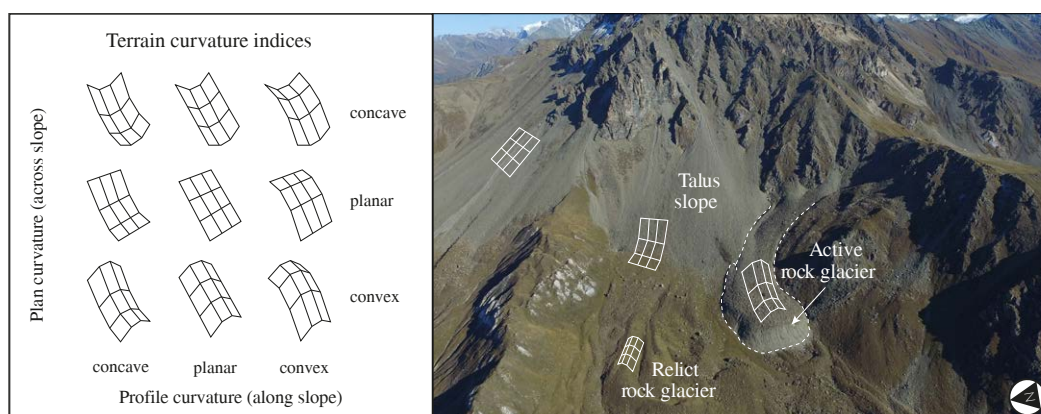


Fig. B.14.: Correlation plot of selected environmental variables within the validity domain.

0° and 360°), we built two uncorrelated indicators: the "northness", which corresponds to the cosine of aspect angle, and the "eastness", equal to the minus sine of the aspect value (see Brenning and Trombotto, 2006).

- Slope: The terrain slope angle influences the permafrost occurrence by governing the direct solar radiation reaching the ground surface. In addition, snow cover may have different depths depending on the slope: rock walls are generally snow free while flat areas of footslopes may be covered by an important amount of snow (Mittaz et al., 2002).
- Potential incoming solar radiation: The amount of energy reaching the ground was calculated as potential incoming solar radiation (PSIR). For this we used the ArcMap "Area Solar Radiation" tool on the 2 m DEM, taking into account shadowing effect of the relief. PSIR was computed for the snow free period, between July 1<sup>st</sup> and October 31<sup>st</sup> (which usually corresponds to the period without snow cover), because PSIR does not have relevant effects on the ground temperature if the snow is present (Hoelzle, 1994).
- Terrain curvature: As mountain permafrost is a thermal phenomenon, it is only observable with the appearance of geomorphological indices such as

rock glaciers lobes or with in-situ measurements. Thus, adding a curvature indicator helps recognizing the presences of lobes that are potentially occupied by permafrost. The Gaussian terrain curvature (the derivative of the slope angle) was computed at different window sizes (10, 30, 50, 100, 500, 1000m) and was included to the dataset as an indicator of various landforms. Large window curvatures (i.e. concavities) help machine learning algorithms to detect the presence of a valley bottom. Conversely, small window terrain curvatures indicate lobes, crests or small depressions (figure B.15).



**Fig. B.15.:** Terrain curvature types and associated landforms for Les Cliosses sector, with an active rock glacier indicating the presence of permafrost (Hérens Valley, Valais; photo: S. Rüttimann).

- Normalized Difference Vegetation Index: A Normalized Difference Vegetation Index (NDVI) was also calculated from false-color infrared images (swisstopo) and included in the dataset. This continuous variable is used for differentiating the vegetation from other land types, as well as characterize local variability in sediment textures.

## Adopted dataset

A delimited area in which proposed simulations were run (the so-called validity domain) was defined based on categorical variables. The prediction of the permafrost occurrence was only computed for sectors where the latter is uncertain (the prediction can be either "permafrost absence" or "permafrost presence"), typically in sedimentary accumulations such as talus slopes, moraine deposits or other debris surfaces. Areas where permafrost is generally absent, such as temperate glaciers, rivers, lakes or vegetation areas, were excluded from the validity domain (table B.2).

The dataset was built on a regular grid of 10 x 10 m. Consequently, the original DEM raster and relative extracted predictors controlling the presence and absence

Variable name	Variable type	Use
Altitude	Continuous	Classification
Northness	Continuous	Classification
Eastness	Continuous	Classification
Slope	Continuous	Classification
PSIR	Continuous	Classification
Air temperature	Continuous	Classification
Lake	Categorical	VD definition
Rockwall	Categorical	VD definition
River	Categorical	VD definition
NDVI	Continuous	Classification / VD definition
Plan curvature 10 m	Continuous	Classification
Plan curvature 30 m	Continuous	Classification
Plan curvature 50 m	Continuous	Classification
Plan curvature 100 m	Continuous	Classification
Plan curvature 500 m	Continuous	Classification
Plan curvature 1000 m	Continuous	Classification
Profile curvature 10 m	Continuous	Classification
Profile curvature 30 m	Continuous	Classification
Profile curvature 50 m	Continuous	Classification
Profile curvature 100 m	Continuous	Classification
Profile curvature 500 m	Continuous	Classification
Profile curvature 1000 m	Continuous	Classification
Surface curvature 10 m	Continuous	Classification
Surface curvature 30 m	Continuous	Classification
Surface curvature 50 m	Continuous	Classification
Surface curvature 100 m	Continuous	Classification
Surface curvature 500 m	Continuous	Classification
Surface curvature 1000 m	Continuous	Classification
Glacier	Categorical	VD definition

**Tab. B.2.:** Features contained in the raw dataset. Categorical variables serve constructing the validity domain (VD), continuous variables are used for classifying the permafrost occurrence.

of mountain permafrost were resampled by computing the mean value of the 2 x 2 m cells. A spatial resolution of 10 m was selected in reason of the computational requirements needed to predict the permafrost distribution for such an extended study area and the minimum size of typical periglacial landforms, such as rock glacier or moraine ridges, that can be represented with at least 2 pixels (a width of 20 m). A higher resolution of the dataset would introduce unexpected noise, both in the attributes (i.e. erroneous attribute values) and the class (label noise, i.e. contradictory examples or missclassifications). This would require additional preprocessing and noise filtering analysis to reduce the noise's effects. Moreover, due to the complexity of the studied phenomenon, from a physical perspective willing to map the permafrost distribution at a too high resolution is illusory.

## Training, validation and test sets

In the present research, classification algorithms were trained, validated and tested with separate independent sub-sets obtained from the original permafrost dataset. The latter can be considered a benchmark data bank to be used for the analysis of the influence of environmental predictors on the permafrost distribution as well as for the modelling of the potential permafrost distribution with machine learning

(both are parts of ongoing research). These sub-sets were selected by sampling individual observations from group of data with different characteristics, previously identified by clustering the original input space (the variable space) with the help of a Self-Organizing Map (SOM, Kohonen, 1982; Kohonen and Honkela, 2007). This type of artificial neural network was trained using unsupervised learning and provided an ordered mapping of the data observations onto a two-dimensional grid. The SOM algorithm computed different models associated with each node of the grid. Observations were mapped into the node whose model is similar to data observation itself. The 2-dimensional map was then clusterized by using the k-means algorithm, while the number of clusters was selected by computing the Davies–Bouldin index (Davies and Bouldin, 1979). By grouping observations with similar characteristics within the input space, a balanced selection from pools of samples with similar characteristics was made. Rather than randomly sampling the permafrost observations, this sampling strategy avoided the presence of highly auto-correlated training samples within the separate sub-sets and produced less classification overfitting.

## Classifiers

In this work, three classifiers were applied to permafrost data. At present, a large palette of classification methods, belonging to statistics and machine learning, exists (Kanevski et al., 2009). For instance, some of these common statistical techniques are the k-nearest neighbor (e.g. Altman, 1992; Everitt et al., 2011), the linear discriminant analysis (e.g. Friedman, 1989) and the logistic regression (see McLachlan, 2004). More recently, other complex algorithms have been developed in order to deal with the emergence of complex set of data and the lack of non-linear solutions. Artificial neural networks, Support Vector Machines and Random forests are just a few examples (Cherkassky and Mulier, 2007; Izenman, 2008; Haykin, 2009; Hastie et al., 2009). These algorithms aim at assigning the class of an observation (here a pixel of the study grid) based on related environmental variables. In order to work, these techniques require a set of training data that are used to fit the classification decision function.

In the following sections, a theoretical overview of Logistic regression (LR), Support Vector Machines (SVM) and Random forests (RF) is given. We selected these algorithms because they belong to three specific sub-domain of machine learning: the former is a linear parametric classifier and it is commonly used as a benchmark classifier to be employed before using more complex classification algorithms. Non-linear SVM is a non-parametric learning algorithm and it is a member of the so-called kernel methods. Finally, RF are an ensemble learning method based on bootstrap aggregating.

### Logistic regression

Logistic regression is one of the most used method for susceptibility mapping in geosciences (see Brenning, 2005; Trigila et al., 2015). This technique fits the best model between independent indicators to dependent variables (Kleinbaum and Klein, 1994). In our case, it tries to estimate the best mathematical relationship between the absence and the presence of permafrost and a set of explanatory independent environmental variables  $x_1, \dots, x_n$ , which can be both continuous or categorical (Hosmer and Lemeshow, 2000). The model output for each grid cell represents the probability  $p$  to belong to the permafrost presence. It is based on the logistic function  $p_i$ , which is defined as follows:

$$\text{logit}(p_i) = \log \left[ \frac{p_i}{1 - p_i} \right] \quad (\text{B.9})$$

where the likelihood ratio corresponds to the ratio between the probability  $p$  that the predicted class is 1 (presence of permafrost) and the probability  $1 - p$  that the class is 0 (absence of permafrost). The final linear logistic model takes this form:

$$\text{logit}(p_i) = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (\text{B.10})$$

with  $\beta_0, \beta_1, \dots, \beta_n$  the coefficients measuring how each independent environmental variable contributes to the permafrost occurrence.

In the present work, the logistic regression is performed using its implementation in WEKA, which is an improved version of the original algorithm presented in Le Cessie and Van Houwelingen, 1994.

### Support Vector Machines

Support Vector Machines (SVM) is a machine learning algorithm based on the statistical learning theory developed by Vapnik (1998). It is based on Structural Risk Minimization that minimizes the training error and controls the complexity of the model in order to improve the generalization ability of a model (Cherkassky and Mulier, 2007).

SVM can be applied to classification tasks and non-linear regression problems. The main principle of this technique presupposes that the set of training vectors  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  where  $x_i \in R^m, i = 1, \dots, n$  with two classes  $y_i = \{-1, 1\}$  is linearly separable by a hyper-plane:

$$(w \cdot x) + b = 0, w \in R^N, b \in R \quad (\text{B.11})$$

where  $w$  corresponds to the hyper-plane normal,  $(\cdot)$  is a scalar product and  $b$  is a scalar base.

The SVM algorithm aims at maximizing the largest margin that separates the training vectors and that is expressed as  $\frac{2}{\|w\|}$  after the normalization. The maximum margin is computed as follows:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (\text{B.12})$$

subjecting to the constrains  $y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, n$ .

The cost function is expressed as:

$$\varnothing(w, b; \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i [w \cdot x_i + b] - 1) \quad (\text{B.13})$$

where  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)^T \in R_+^n$  is a Lagrangian multiplier.

As detailed in Vapnik (1998), the problem has to be solved by dual minimization of the cost function with the respect to  $w$  and  $b$ . Because in most cases the training vectors are rarely linearly separable, a slack variable  $\xi_i$  and a penalty term  $C$ , avoiding high values of the latter, are introduced:

$$y_i((w \cdot x_i) + b) \geq 1 - \xi_i, i = 1, 2, \dots, n, \xi_i \geq 0 \quad (\text{B.14})$$

Therefore, the equation, maximizing the margin becomes:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \quad (\text{B.15})$$

Non-linear classification of real data is performed using the same principles, but only by applying a kernel trick (Vapnik, 1998) when a kernel function  $K(x_i, x_j)$  maps/transforms the original data into a high dimensional feature space where, again, we are looking for a linear solution. Depending on the kernel, the linear solution in the feature space corresponds to a non-linear solution in the original space and the kernel controls the complexity of this mapping.

In this work, we used a radial basis function (RBF) kernel  $K(x_i, x_j) = e^{(-\gamma \|x_i - x_j\|^2)}$ ,  $\gamma > 0$  which is one of the most used kernels providing a good generalization. Moreover, we employed the libSVM library implemented in WEKA. For additional information on the application of this algorithm, please see Deluigi and Lambiel, 2013.

### Random forests

Developed by Breiman (2001), Random Forests is an ensemble algorithm that computes  $n$  binary classification trees (forest) with the purpose of having higher predictive capabilities compared to the classification with a single decision tree (Cutler et al., 2007). With the combination of several trees constructed with a random selection of the inputs, the classification accuracy usually improves. Implemented in

WEKA, this technique is suitable when dealing with both categorical and numerical predictors. It adopts bagging to randomly select permafrost observations and their relative variables to train the model. The membership of a class is selected by a majority vote for the most popular class within the total number of trees. In this study, we used by default 2/3 bootstrap of the training set to construct each tree. The remaining 1/3 bootstrap of the training data was used to assess the generalization capacity of the algorithm.

With this technique there is thus no need for a cross-validation or a separate validation set to obtain unbiased estimate of the test set error. Unselected observations serve indeed to calculate the error of the model (the "out-of-bag" error, or OOB). The error estimate also allows measuring the contribution of each predictor by computing the average decrease in model accuracy on the OOB samples when the values of the respective feature are randomly permuted. Two evaluators of the variable importance exist with Random forests: the Mean-Decrease-in-Accuracy, measuring the decrease of the accuracy expressed with the OOB-error when a variable is left out and the Mean-Decrease-in-Gini index, defining how the output at each node is impure (Breiman, 2001).

Because of the random selection of independent variables and observations at each node, it is suggested analysing the correct number of trees to be employed in Random forests in order to obtain a stable model (Catani et al., 2013). Therefore, before applying the classification model on a new prediction, we firstly investigated the changes in the OOB error curve, aiming at selecting the correct number of trees (see section B).

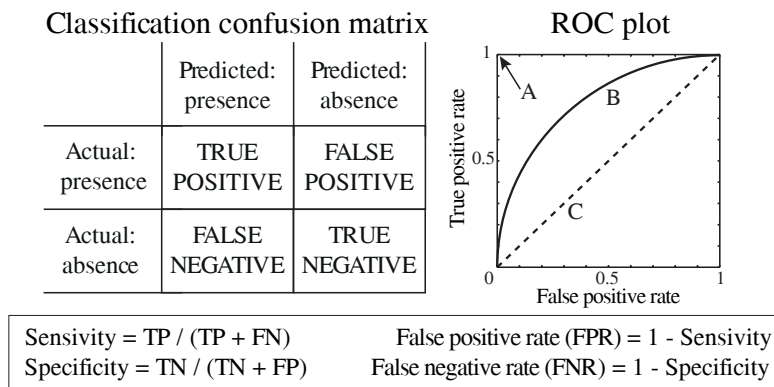
## Classification quality measures

The known permafrost absence and presence was binary coded as [-1, 1]. Nevertheless, with the chosen classifiers it is possible to obtain the probability of belonging to the class as model output. This result is in fact most convenient when dealing with a natural phenomenon.

The quality of the prediction of the presence and absence of permafrost modeled with supervised learning algorithms can be generally assessed with multiple statistics (Doswell III et al., 1990). As we deal with binary classification, in this study we computed 2-by-2 contingency tables that display four possible cases (see figure B.16): number of true positives (TP), i.e. a pixel with known permafrost presence that was predicted correctly (presence-presence); true negatives (TN), when absence is correctly not encountered (absence-absence); false negatives (FN), i.e. permafrost presence is not detected (presence-absence) or false positives (FP) for permafrost absence incorrectly forecasted (absence-presence). Indeed, when dealing with the



management of natural hazards related to periglacial source areas, it is preferable to obtain a conservative result that miss-classifies permafrost where in reality it is not, rather than the opposite.



**Fig. B.16.:** Classification confusion matrix (left) and evaluation rates (bottom) used to assess the model quality. The Receiver Operating Characteristics curve also indicates the quality of the classification: (A) corresponds to the perfect result (no errors, AUROC = 1), (B) indicates a typical ROC curve (AUROC=0.85) and (C) refers to an inefficient classification (AUROC = 0.5).

In addition to these rates, a Receiver Operating Characteristics curve (ROC) can be built (Fawcett, 2006). This 2-dimensional graph measures the quality of the model prediction by plotting the FP rate as the horizontal axis and the TP rate as the vertical axis, according to different susceptibility threshold values. Swets, 1988 indicates that the Area Under the ROC curve (or AUROC) ranges between 0.5 and to 1. Values between 0.5 and 0.7 generally indicate that the learned model has poor predictive capabilities. When AUROC is between 0.7 and 0.8, it means that the classification have moderate predictive abilities. Between 0.8 and 0.9 they are considered good and above 0.9 they become excellent. Nevertheless, recent studies have observed that, for similar AUROCs, the predicted map of a given phenomenon can be particularly different (i.e. Micheletti et al., 2014).

## Used softwares

In this research we have used algorithms that are already implemented in WEKA (Waikato University, Hall et al., 2009). The LibSVM library (Chang and Lin, 2011) is implemented in this software as a package. Permafrost evidence identification, variables extraction and dataset building were performed by using ArcGIS (ESRI) and Matlab softwares.

## Results and discussion

In order to select the best map of the potential distribution of mountain permafrost for the Western part of the Valais Alps, modelling performances of standard Logistic regression were compared to the ones of Support Vector Machines and Random forests. The model setup was characterized by the selection of training and test data including observations of the permafrost presence and permafrost absence coupled together with environmental variable grids. As mentioned, we decided to present the algorithm predictions in the form of probabilities, which provide easiness when comparing different model outputs. Moreover, for coherence with existing models and ease of map comparability, presented maps visualize probabilities greater than 0.5, corresponding to the possible presence of permafrost.

### Model assessment

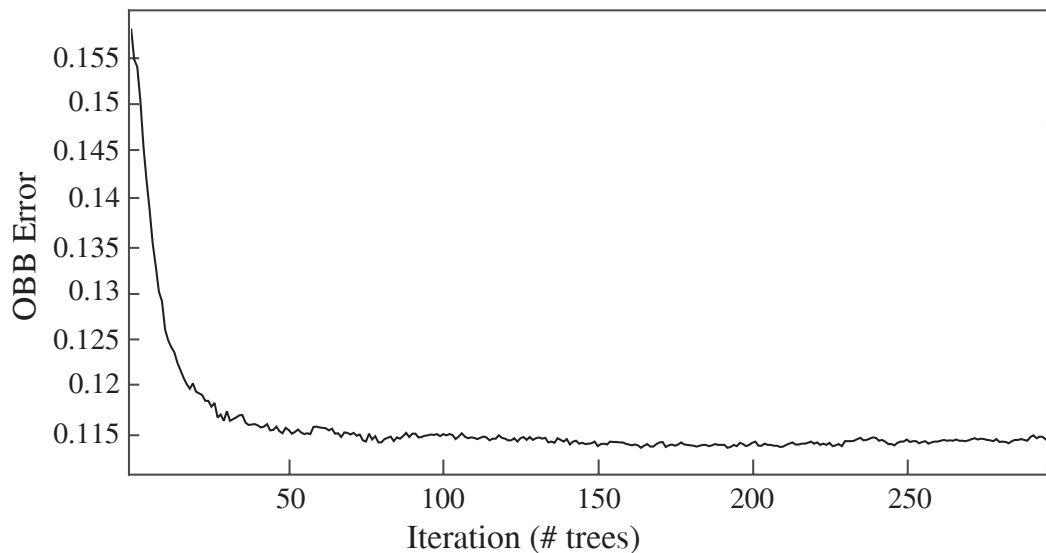
By removing auto-correlated variables from the dataset (figure B.14) before applying the Logistic regression, the algorithm converged to a solution, providing an AUROC value of 0.807 (table B.3). The sum of percentage of false positives and false negatives is greater than 25% of the total prediction. Besides, 11.5% of the study area is simulated as potentially frozen (726'397 pixel of the prediction grid with  $p > 0.5$ ).

Statistics	LR	SVM	RF
Precision	0.743	0.858	0.797
Recall	0.746	0.859	0.798
AUROC	0.807	0.848	0.884
FP (%)	10.72	5.97	8.78
FN (%)	14.70	8.17	11.37
# cells with $p > 0.5$	726'397	1'079'745	793'117

**Tab. B.3.:** Summary statistics of the Logistic regression, Support Vector Machines and Random forests models.

The prediction of the permafrost occurrence with RF was performed with the same sub-set used for LR. The variability of the OOB error was analyzed in order to select the correct number of trees to use. As figure B.17 illustrates, the out-of-bag error becomes stable between 200 and 250 trees. We decided thus to use 300 trees for the RF modelling configuration.

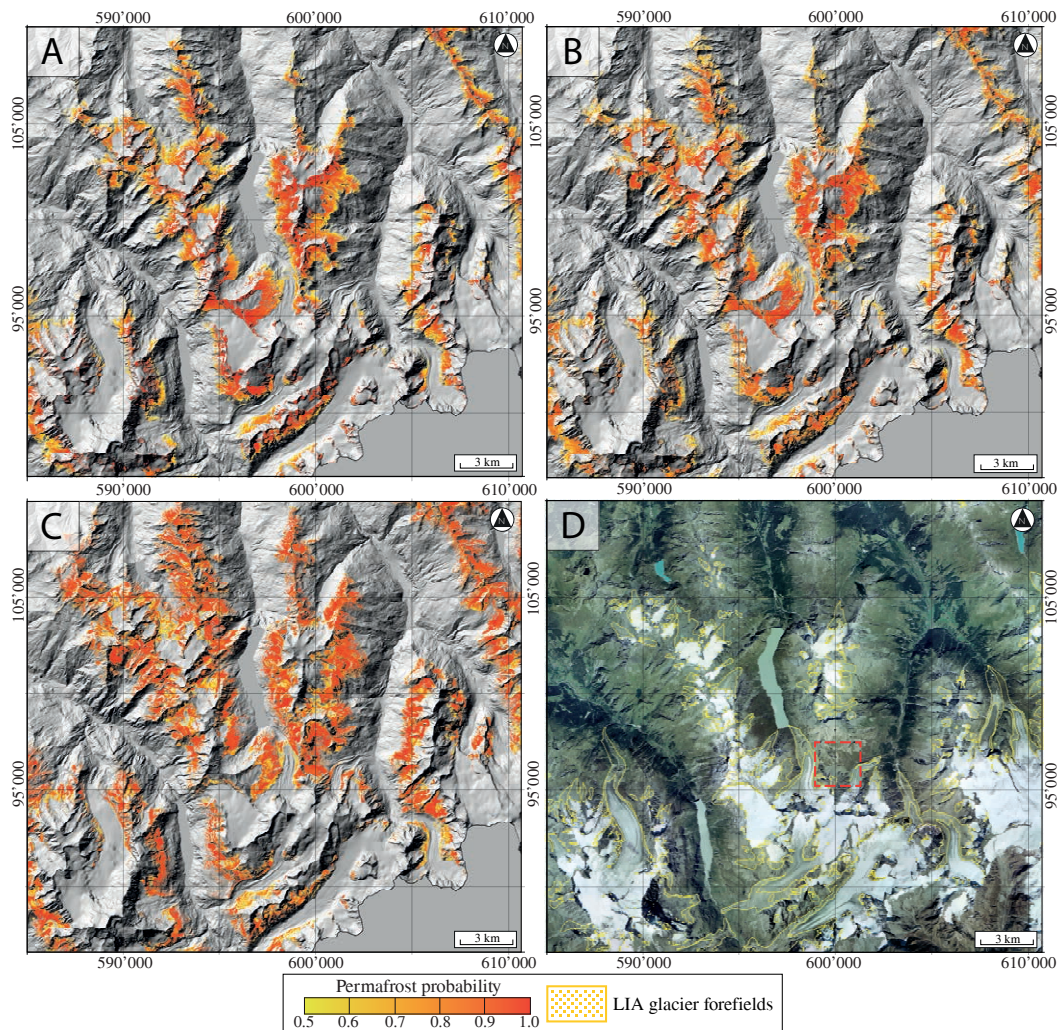
Obtained classification AUROC is 0.884 and around 20% of the total number of observations correspond to false positives and false negatives. A total of 793'117 pixel was predicted with a positive occurrence of permafrost, that is 12.6% of the area of interest. The extent of the permafrost presence ( $p > 0.5$ ) is thus slightly



**Fig. B.17.:** OBB error as a function of the number of trees.

greater than the one obtained with Logistic regression. The number of expected false positives also decreases with this algorithm. Compared to the map of the permafrost distribution produced with LR, RF map is less smooth, with the presence of some spatial artefacts, which is a common issue encountered when using this technique (Brenning, 2005). However, it is worth mentioning that permafrost lower limits are visually located at the same altitude to the ones obtained with LR (see figures B.18.A and B.18.B).

A RBF kernel was selected as a kernel function for the non-linear SVM classification, accordingly to results obtained in a preliminary test study of Deluigi and Lambiel, 2013. After the cross-validation step, the kernel parameters, as well as the regularization parameter  $C$  and the threshold  $\xi$ , were optimized via grid search. The number of support vectors is close to the 39% of the training observations (around 36'000 samples). This value is not only dependent on how much slack is allowed, but also on the complexity of the model. To allow users interpreting the binary result provided by SVM in a meaningful way, the classification of presence and absence of permafrost was transformed by post-processing to yield a posteriori probability (from categorical to probabilistic prediction). The technique, presented by Platt (1999), applies a maximum likelihood-optimized logistic transformation on the SVM decision function in order to obtain permafrost probabilities. Classification AUROC of the RBF-SVM is equal to 0.848, with around 14.2% of the total prediction classified as false positives and false negatives. The number of pixels indicating a probability of permafrost occurrence greater than 0.5 is 1'079'745, which corresponds to 17.13% of the total region extent. This is reflected by a map with more extensive surface potentially occupied by permafrost.

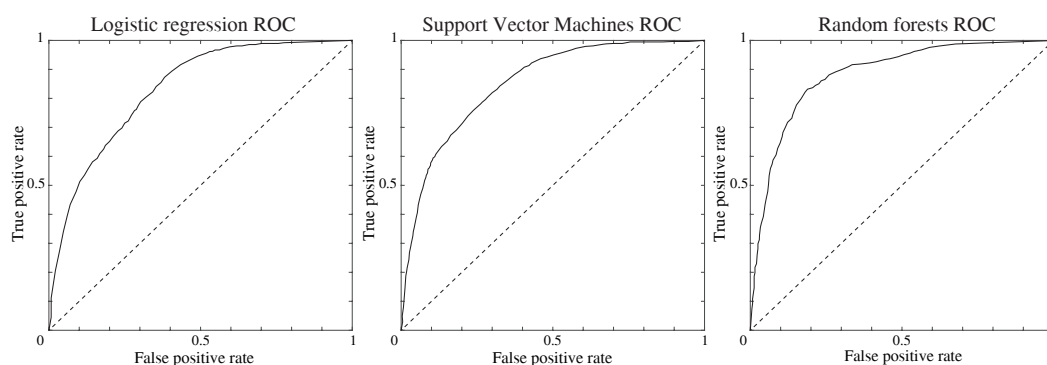


**Fig. B.18.:** Permafrost distribution maps obtained with (A) Logistic regression, (B) Random Forests and (C) Support Vector Machines. (D) Ground-truth of the study area (aerial images from swisstopo) and LIA glacier forefield extent for ease of interpretation. Red dashed square indicates the location the overviews of figure B.20.

## Model comparison

It is common practice in machine learning to estimate the classification error by using cross-validation and to choose the algorithm that provides the lowest estimate. However, it is important to investigate which model is statistically the best when performing a specific classification task. ROC curves (figure B.19) were evaluated on independent test sets for comparing the generalization performances of LR, SVM and RF.

Resulting performances were compared by using the Model Evaluator module of WEKA (see Witten and Frank, 2005) that employs a paired t-test (confidence: 0.05, two tailed) (table B.4). Results of the test indicate that SVM is not significantly



**Fig. B.19.:** ROC curves for classification by Logistic regression (0.807, left), Support Vector Machines (0.848, center) and Random forests (0.884, right).

worse than RF. However, the paired t-test outcome suggests that RF is significantly better than LR.

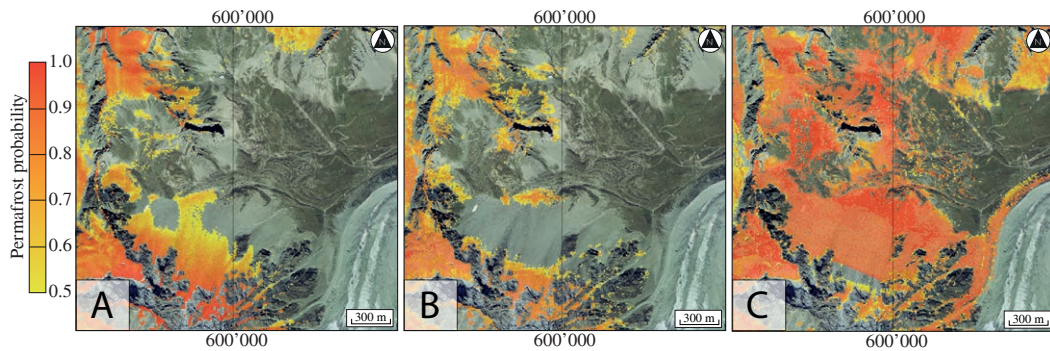
Classifier	SVM	RF	LR
Performance (AUROC)	0.848	0.884	0.807*
	(v/ /*)	(0/1/0)	(0/0/1)

**Tab. B.4.:** Model performances and paired t-test outcomes (v: significantly better, \*: significantly worse).

## Expert domain quality evaluation

In addition to the statistical evaluation of the presented models, the three maps were also analysed according to their geomorphological relevance. The lower limit of permafrost given by SVM is 100 - 150 m lower than the one given by LR and RF, being more in accordance with field observations (see figures B.20.A, B.20.B and B.20.C). SVM probabilities appear to be generally higher than the ones given by LR and RF. In comparison with the permafrost map of the Swiss Federal Office for the Environment (BAFU, 2005) or the Alpine Permafrost Index Map (APIM) of Boeckli et al., 2012, the proposed simulations provide less optimistic results. Indeed, especially for LR and RF, the prediction of the presence of permafrost in sediments is restricted to smaller surfaces. In addition, maps produced by using machine learning do not present the typical altitude thresholds that indicate higher permafrost occurrence with the increase in altitude (see Lambiel and Reynard, 2001; BAFU, 2005; Schrott et al., 2012; Boeckli et al., 2012). As the relationship between permafrost evidences and environmental predictors is learnt directly from data without recurring to physical models, probabilities are predicted for each independent pixel. Machine learning algorithms appear thus more suitable to simulate the high spatial discontinuity of the phenomenon at the micro scale.

Despite the conservative result in terms of surfaces potentially frozen and a slightly lower AUROC compared to RF, SVM granted a classification with the lowest number



**Fig. B.20.:** Potential permafrost distribution in the Fontanesses sector (Arolla valley) obtained with (A) Logistic regression, (B) Random Forests and (C) Support Vector Machines.

of miss-classifications. Obtained AUROCs range mid-way between existing studies such as Boeckli et al., 2012, Azócar et al., 2016 and Sattler et al., 2016. This may be explained by the addition to the adopted dataset of indicators of know permafrost absence not only restricted to relict rock glacier, but also in talus slopes and other sediment accumulations.

When observing the interpretability of the proposed maps, the RF map is less smooth than the LR and SVM ones and presents a higher number of artefacts. Conversely, SVM gives less variability of probability when observing a given pixel and the one in the neighborhood. Furthermore, the potential distribution of permafrost is more conservative with this approach. Its great heterogeneity is also better respected with SVM. In fact, LR produces probabilities higher than 0.5 that increase linearly with the altitude. If the permafrost probability increases effectively with the altitude at the regional scale, permafrost occurrence at the local scale can change within distances of few tens of meters due to the high variability of the local characteristics (see Otto et al., 2012; Rödder and Kneisel, 2012). In some talus slopes, the SVM map illustrates a lower probability in the mid-upper part of the slope (figure B.20.C). As showed by Lambiel and Pieracci (2008) or Scapozza et al. (2011) this distribution is more in accordance with the field reality. The same behaviour is also partially simulated with LR and RF, but their result is less conservative and it presents an increased number of false positives and false negatives. In glacier forefields, highlighted with a yellow mask in figure B.18.D), the permafrost distribution may be even more complex due to thermal and mechanical perturbations by the glacier advance during the Little Ice Age (Reynard et al., 2003; Kneisel and Käab, 2007). These studies illustrated that permafrost is often restricted to the lateral and frontal margin of these environments, where the former glacier was the thinnest. Hence, one has to be careful when looking at the potential permafrost extent within these areas.

## Conclusions

In this study we presented the results of the classification and mapping of mountain permafrost data using Logistic regression (LR), Random forests (RF) and Support Vector Machines (SVM) in a large area of the Western Swiss Alps. Machine learning algorithms provided precise distribution maps at the micro scale by learning the statistical relationship between training permafrost evidences and permafrost explanatory variables. LR predicted a smooth map with an increase of the permafrost probability linearly with the altitude, which does not reflect the strong spatial discontinuity of the phenomenon. RF provided excellent classification performances despite the similar permafrost extent of the LR map, but it differs in terms of the result smoothness. SVM model performance resides between the two other applied classifiers. It is characterized by a lower number of miss-classifications and the potential permafrost map obtained tends to be more conservative in comparison to LR and RF. In addition, the permafrost discontinuity was best reproduced with SVM. Conversely to LR, the occurrence of permafrost does not tend to increase linearly with the altitude and it is indeed possible to observe higher permafrost probabilities in the lower half of some talus slopes, which is in agreement with field data.

The internal mechanisms of the three employed classification algorithms open the way to further analysis on permafrost data such as active learning, allowing choosing training data from selected support vectors, and featuring ranking, with embedded measures of the variable importance. It follows that the approach presented not only helps in mapping the mountain permafrost distribution but can also be employed to understand related data properties. Much research remains thus to be done to improve the prediction at the micro scale. For example, model uncertainties characterization could reduce misclassification of known permafrost evidences and thus improve the model robustness. The extraction of new environmental variables controlling permafrost conditions at the micro scale, such as the grain size of sedimentary deposits, could as well help refining obtained results.

## Acknowledgements

We thank two anonymous reviewers for their constructive comments. We would also like to acknowledge the Swiss National Science Foundation for the funding of this research (project no. 200021\_152924) as well as Prof. R. Delaloye and colleagues from the University of Fribourg for having provided permafrost evidences.







# Bibliography

- Allen, S. K., Gruber, S., and Owens, I. F. (2009). Exploring steep bedrock permafrost and its relationship with recent slope failures in the Southern Alps of New Zealand. *Permafrost and Periglacial Processes* 20.4, pp. 345–356 (cit. on pp. 23, 223).
- Alm, M. (2017). Probability Modelling of Alpine Permafrost Distribution in Tarfala Valley, Sweden (cit. on pp. 72, 177, 182).
- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* 46.3, pp. 175–185 (cit. on p. 248).
- Amatulli, G., Camia, A., and San-Miguel-Ayanz, J. (2013). Estimating future burned areas under changing climate in the EU-Mediterranean countries. *Science of The Total Environment* 450, pp. 209–222 (cit. on p. 242).
- Angillieri, M. Y. E. (2017). Permafrost distribution map of San Juan Dry Andes (Argentina) based on rock glacier sites. *Journal of South American Earth Sciences* 73, pp. 42–49 (cit. on pp. 29, 177, 182, 188).
- Apté, C. and Weiss, S. (1997). Data mining with decision trees and decision rules. *Future generation computer systems* 13.2-3, pp. 197–210 (cit. on p. 81).
- Ascione, A., Cinque, A., Miccadei, E., Villani, F., and Berti, C. (2008). The Plio-Quaternary uplift of the Apennine chain: new data from the analysis of topography and river valleys in Central Italy. *Geomorphology* 102.1, pp. 105–118 (cit. on p. 45).
- Avian, M. and Kellerer-Pirklbauer, A. (2012). Modelling of potential permafrost distribution during the Younger Dryas, the Little Ice Age and at present in the Reisseck Mountains, Hohe Tauern Range, Austria. *Austrian Journal of Earth Sciences* 105.2, pp. 140–153 (cit. on pp. 4, 241).
- Azócar, G. F., Brenning, A., and Bodin, X. (2016). Permafrost Distribution Modeling in the Semi-Arid Chilean Andes. *The Cryosphere* 11.2, p. 877 (cit. on pp. 4, 28, 29, 35, 146, 177, 180, 185, 188, 241, 257).
- Baeriswyl, P. A. and Rebetez, M. (1997). Regionalization of precipitation in Switzerland by means of principal component analysis. *Theoretical and Applied Climatology* 58.1-2, pp. 31–41 (cit. on p. 33).
- BAFU (2005). Hinweiskarte der potenziellen Permafrostverbreitung der Schweiz. Bundesamt für Umwelt (cit. on pp. 4, 25, 28, 152, 158, 161, 164, 168, 173, 175, 176, 185, 223, 256).
- Balch, E. S. (1900). *Glaciers, or Freezing Caverns*. Allen Lane & Scott (cit. on p. 16).

- Barboux, C., Delaloye, R., and Lambiel, C. (2014a). Inventorying slope movements in an Alpine environment using DInSAR. *Earth Surface Processes and Landforms* 39.15, pp. 2087–2099 (cit. on pp. 33, 35, 243).
- Barboux, C., Delaloye, R., and Lambiel, C. (2014b). Inventorying slope movements in an Alpine environment using DInSAR. *Earth Surface Processes and Landforms* 39.15, pp. 2087–2099 (cit. on p. 171).
- Baroni, C., Carton, A., and Seppi, R. (2004). Distribution and behaviour of rock glaciers in the Adamello–Presanella Massif (Italian Alps). *Permafrost and Periglacial Processes* 15.3, pp. 243–259 (cit. on pp. 23, 176, 223).
- Barsch, D. (1969). Permafrost in der oberen subnivalen Stufe der Alpen. *Geographica Helvetica* 24.1, pp. 10–12 (cit. on p. 15).
- Barsch, D. (1988). Advances in periglacial geomorphology. Ed. by M. J. Clark. John Wiley & Sons. Chap. Rockglaciers, pp. 69–90 (cit. on p. 14).
- Barsch, D. (1996). Rockglaciers: indicators for the present and former geocology in high mountain environments. Vol. 16. Springer Science & Business Media. (cit. on pp. 11, 14, 15, 18).
- Barsch, D. (2012). Rockglaciers: indicators for the present and former geocology in high mountain environments. Vol. 16. Springer Science & Business Media (cit. on pp. 34, 242).
- Bavay, M., Lehning, M., Jonas, T., and Löwe, H. (2009). Simulations of future snow cover and discharge in Alpine headwater catchments. *Hydrological Processes* 23.1, pp. 95–108 (cit. on p. 188).
- Bellman, R. E. (1961). Adaptive control processes: a guided tour. Princeton university press (cit. on p. 57).
- Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., Fantini, A., Giacona, F., Hauck, C., Huss, M., Huwald, H., Lehning, M., López-Moreno, J.-I., Magnusson, J., Marty, C., Moran-Tejéda, E., Morin, S., Naaim, M., Provenzale, A., Rabatel, A., Six, D., Stötter, J., Strasser, U., Terzago, S., and Vincent, C. (2017a). The European mountain cryosphere: A review of past, current and future issues. *The Cryosphere Discussions*, pp. 1–60 (cit. on p. 1).
- Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., Fantini, A., Giacona, F., Hauck, C., Huss, M., Huwald, H., Lehning, M., Lopez-Moreno, J. I., Magnusson, J., Marty, C., Moran-Tejéda, E., Morin, S., Naaim, M., Provenzale, A., Rabatel, A., Six, D., Stötter, J., Strasser, U., Terzago, S., and Vincent, C. (2017b). The European mountain cryosphere: A review of past, current and future issues. *The Cryosphere Discussions*, pp. 1–60 (cit. on p. 240).
- Berthling, I. (2011). Beyond confusion: Rock glaciers as cryo-conditioned landforms. *Geomorphology* 131.3, pp. 98–106 (cit. on pp. 14, 15, 19).
- Bishop, C. M. (2006). Pattern recognition. *Machine Learning* 128 (cit. on pp. 51, 74, 76, 242).
- Boeckli, L., Brenning, A., Gruber, S., and Noetzli, J. (2011). A statistical permafrost distribution model for the European Alps. *The Cryosphere Discussions* 5, pp. 1419–1459 (cit. on pp. 4, 179, 180, 223).

- Boeckli, L., Brenning, A., Gruber, S., and Noetzli, J. (2012). A statistical approach to modelling permafrost distribution in the European Alps or similar mountain ranges. *The Cryosphere* 6.1, pp. 125–140 (cit. on pp. [4](#), [25](#), [27](#), [28](#), [133](#), [177](#), [185](#), [187](#), [188](#), [241](#), [243](#), [256](#), [257](#)).
- Bonnaventure, P. P., Lewkowicz, A. G., Kremer, M., and Sawada, M. C. (2012). A permafrost probability model for the southern Yukon and northern British Columbia, Canada. *Permafrost and Periglacial Processes* 23.1, pp. 52–68 (cit. on pp. [28](#), [177](#), [188](#)).
- Boser, B. E., Guyon, I. M., and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the 5th annual workshop on Computational learning theory*. ACM, pp. 144–152 (cit. on pp. [62](#), [76](#)).
- Bosson, J.-B. and Lambiel, C. (2016). Internal structure and current evolution of very small debris-covered glacier systems located in alpine permafrost environments. *Frontiers in Earth Science* 4, p. 39 (cit. on pp. [19](#), [42](#), [172](#), [187](#)).
- Bosson, J.-B., Deline, P., Bodin, X., Schoeneich, P., Baron, L., Gardent, M., and Lambiel, C. (2015). The influence of ground ice distribution on geomorphic dynamics since the Little Ice Age in proglacial areas of two cirque glacier systems. *Earth Surface Processes and Landforms* 40.5, pp. 666–680 (cit. on pp. [42](#), [241](#)).
- Bouët, M. (1978). Le Valais. Climatologie der Schweiz 2. Régionale Beschreibungen. 1. Teil. *Beih. Ann. Schweiz. Meteorolog. Zentralanst* 1977, pp. 88–114 (cit. on pp. [39](#), [228](#), [244](#)).
- Breiman, L. (1996). Bagging predictors. *Machine learning* 24.2, pp. 123–140 (cit. on p. [83](#)).
- Breiman, L. (2001). Random forests. *Machine learning* 45.1, pp. 5–32 (cit. on pp. [63](#), [82–84](#), [114](#), [250](#), [251](#)).
- Brenning, A. (2005). Spatial prediction models for landslide hazards: review, comparison and evaluation. *Natural Hazards and Earth System Science* 5.6, pp. 853–862 (cit. on pp. [67](#), [249](#), [254](#)).
- Brenning, A. (2012). Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package *sperrorest*. *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*. IEEE, pp. 5372–5375 (cit. on p. [68](#)).
- Brenning, A. and Azócar, G. F. (2010). Statistical analysis of topographic and climatic controls and multispectral signatures of rock glaciers in the dry Andes, Chile (27–33 S). *Permafrost and Periglacial Processes* 21.1, pp. 54–66 (cit. on pp. [28](#), [177](#)).
- Brenning, A. and Trombotto, D. (2006). Logistic regression modeling of rock glacier and glacier distribution: topographic and climatic controls in the semi-arid Andes. *Geomorphology* 81.1, pp. 141–154 (cit. on pp. [28](#), [39](#), [72](#), [177](#), [182](#), [245](#)).
- Brown, J., Ferrians Jr, O. J., Heginbottom, J. A., and Melnikov, E.S. (1997). Circum-Arctic map of permafrost and ground-ice conditions. US Geological Survey Reston, VA (cit. on p. [26](#)).
- Burga, C. A., Frauenfelder, R., Ruffet, J., Hoelzle, M., and Käab, A. (2004). Vegetation on Alpine rock glacier surfaces: a contribution to abundance and dynamics on extreme plant habitats. *Flora-Morphology, Distribution, Functional Ecology of Plants* 199.6, pp. 505–515 (cit. on p. [12](#)).

- Burn, C. R. and Smith, C. A. S. (1988). Observations of the "thermal offset" in near-surface mean annual ground temperatures at several sites near Mayo, Yukon Territory, Canada. *Arctic*, pp. 99–104 (cit. on p. 12).
- Cailleux, A. and Tricart, J. (1959). Initiation à l'étude des sables et des galets. (Cit. on p. 44).
- Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods* 3.1, pp. 1–27 (cit. on p. 70).
- Camstra, F. (2003). Data dimensionality estimation methods: a survey. *Pattern recognition* 36.12, pp. 2945–2954 (cit. on p. 57).
- Cannone, N. and Gerdol, R. (2003). Vegetation as an ecological indicator of surface instability in rock glaciers. *Arctic, Antarctic, and Alpine Research* 35.3, pp. 384–390 (cit. on p. 12).
- Capt, M., Bosson, J.-B., Fischer, M., Micheletti, N., and Lambiel, C. (2016). Decadal evolution of a very small heavily debris-covered glacier in an Alpine permafrost environment. *Journal of Glaciology* 62.233, pp. 535–551 (cit. on p. 172).
- Carrasco Kind, M. and Brunner, R. J. (2014). SOM z: photometric redshift PDFs with self-organizing maps and random atlas. *Monthly Notices of the Royal Astronomical Society* 438.4, pp. 3409–3421 (cit. on pp. 69, 70).
- Carrivick, J. L. and Heckmann, T. (2017). Short-term geomorphological evolution of proglacial systems. *Geomorphology* (cit. on p. 2).
- Carturan, L., Zuecco, G., Seppi, R., Zanoner, T., Borga, M., Carton, A., and Dalla Fontana, G. (2015). Catchment-Scale Permafrost Mapping using Spring Water Characteristics. *Permafrost and Periglacial Processes* 27.3, pp. 253–270 (cit. on pp. 36, 243).
- Catani, F., Lagomarsino, D., Segoni, S., and Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences* 13.11, pp. 2815–2831 (cit. on pp. 83, 251).
- Cavalli, M., Tarolli, P., Marchi, L., and Dalla Fontana, G. (2008). The effectiveness of airborne LiDAR data in the recognition of channel-bed morphology. *Catena* 73.3, pp. 249–260 (cit. on p. 186).
- Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P. (2015). An improved representation of physical permafrost dynamics in the JULES land-surface model. *Geoscientific Model Development* 8.5, pp. 1493–1508 (cit. on p. 5).
- Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2.3, p. 27 (cit. on p. 252).
- Cherkassky, V. and Mulier, F.M. (2007). Learning from data: concepts, theory, and methods. John Wiley & Sons (cit. on pp. 51, 76, 227, 248, 249).
- Colombo, N., Giaccone, E., Paro, L., Buffa, G., and Fratianni, S. (2016). The recent transition from glacial to periglacial environment in a high altitude alpine basin (Sabbione basin, North-western Italian Alps). Preliminary outcomes from a multidisciplinary approach. *Geografia Fisica e Dinamica Quaternaria* 39.1, pp. 21–36 (cit. on p. 12).
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning* 20.3, pp. 273–297 (cit. on p. 76).

- Cremonese, E., Gruber, S., Phillips, M., Pogliotti, P., Böckli, L., Noetzli, J., Suter, C., Bodin, X., Crepaz, A., Kellerer-Pirklbauer, A., et al. (2011). An inventory of permafrost evidence for the European Alps. *The Cryosphere* 5, pp. 651–657 (cit. on p. 187).
- Crippen, R. E. (1990). Calculating the vegetation index faster. *Remote Sensing of Environment* 34.1, pp. 71–73 (cit. on p. 43).
- Cross, M. and Moscardini, A. O. (1985). Learning the art of mathematical modelling. John Wiley & Sons, Inc. (cit. on p. 3).
- Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology* 88.11, pp. 2783–2792 (cit. on pp. 82, 250).
- Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2, pp. 224–227 (cit. on pp. 70, 248).
- Davies, M. C. R., Hamza, O., and Harris, C. (2001). The effect of rise in mean annual temperature on the stability of rock slopes containing ice-filled discontinuities. *Permafrost and Periglacial Processes* 12.1, pp. 137–144 (cit. on p. 2).
- Delaloye, R. (2004). „Contribution à l'étude du pergélisol de montagne en zone marginale“. PhD thesis. Université de Fribourg (cit. on pp. 11, 18, 19, 33, 37, 156, 157, 224, 243).
- Delaloye, R. and Lambiel, C. (2005). Evidence of winter ascending air circulation throughout talus slopes and rock glaciers situated in the lower belt of alpine discontinuous permafrost (Swiss Alps). *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography* 59.2, pp. 194–203 (cit. on pp. 3, 16, 33, 37, 156, 224, 243).
- Delaloye, R. and Morand, S. (1997). Du Val Ferret au Grand-Combin (Alpes Valaisannes): Inventaire des glaciers rocheux et analyse spatiale du pergélisol à l'aide d'un système d'information géographique (IDRISI) (cit. on pp. 23, 176, 223).
- Delaloye, R. and Morand, S. (1998). Les glaciers rocheux de la région d'Entremont (Alpes valaisannes): inventaire et analyse spatiale à l'aide d'un SIG. *Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie*. Vol. 158. ETH Zürich, pp. 75–86 (cit. on pp. 34, 242).
- Delaloye, R., Reynard, E., and Lambiel, C. (2001). Pergélisol et construction de remontées mécaniques: l'exemple des Lapires (Mont-Gelé, Valais). Vol. 141, 103-113. Société suisse de mécanique des sols et des roches (cit. on pp. 3, 33).
- Delaloye, R., Métrailler, S., and Lugon, R. (2003a). Évolution du pergélisol dans les complexes glaciers/glaciers rocheux des Beccs-de-Bosson et de Lona (Nax, St-Martin, VS). *Bulletin de la Murithienne* 121, pp. 7–20 (cit. on pp. 33, 224).
- Delaloye, R., Reynard, E., Lambiel, C., Marescot, L., and Monnet, R. (2003b). Thermal anomaly in a cold scree slope (Creux du Van, Switzerland). *Proceedings of the 8th International Conference on Permafrost*. Vol. 2125. Balkema: Zurich (cit. on p. 223).
- Delaloye, R., Lambiel, C., Lugon, R., Raetzo, H., and Strozzi, T. (2007). Typical ERS InSAR signature of slope movements in a periglacial mountain environment (Swiss Alps). *Proceedings Envisat Symposium*. Vol. ESA SP-636. Montreux, Switzerland (cit. on pp. 33, 35, 162, 243).

- Delaloye, R., Perruchoud, E., Avian, M., Kaufmann, V., Bodin, X., Hausmann, H., Ikeda, A., Kääh, A., Kellerer-Pirklbauer, A., Krainer, K., et al. (2008). Recent interannual variations of rock glacier creep in the European Alps. *Proceedings of the 9th International Conference on Permafrost, Fairbanks, Alaska*. Vol. 29, pp. 343–348 (cit. on pp. [2](#), [33](#), [37](#), [243](#)).
- Delaloye, R., Lambiel, C., and Gärtner-Roer, I. (2010). Overview of rock glacier kinematics research in the Swiss Alps: seasonal rhythm, interannual variations and trends over several decades. *Geographica Helvetica* 65.2, pp. 135–145 (cit. on pp. [14](#), [33](#), [171](#), [241](#)).
- Deline, P., Gruber, S., Delaloye, R., Fischer, L., Geertsema, M., Giardino, M., Hasler, A., Kirkbride, M., Krautblatter, M., Magnin, F., et al. (2015). Ice loss and slope stability in high-mountain regions. *Snow and Ice-Related Hazards, Risks, and Disasters*, edited by: Haeberli, W., Whiteman, C., and Shroder, JF, Elsevier Science, Saint Louis, pp. 521–561 (cit. on pp. [3](#), [20](#)).
- Deluigi, N. (2012). „Modélisation de la répartition du pergélisol alpin à l’aide de l’apprentissage automatique“. MSc thesis. University of Lausanne (cit. on pp. [5](#), [29](#), [221](#)).
- Deluigi, N. and Lambiel, C. (2013). PERMAL: a machine learning approach for alpine permafrost distribution modeling. *Jahrestagung der Schweizerischen Geomorphologischen Gesellschaft*. Vol. 4, pp. 47–62 (cit. on pp. [5](#), [6](#), [29](#), [152](#), [250](#), [254](#)).
- Deluigi, N. and Lambiel, C. (2014). An attempt to detect sedimentary materials grain size using texture analysis of FCIR orthophotos. *EGU General Assembly Conference Abstracts*. Vol. 16 (cit. on p. [45](#)).
- Deluigi, N., Lambiel, C., and Kanevski, M. (2017). Data-driven mapping of the potential mountain permafrost distribution. *Science of The Total Environment* 590, pp. 370–380 (cit. on pp. [55](#), [188](#)).
- Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B., et al. (2012). Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography* 39.12, pp. 2119–2131 (cit. on pp. [60](#), [102](#), [146](#)).
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36.1, pp. 27–46 (cit. on p. [146](#)).
- Doswell III, C. A., Davies-Jones, R., and Keller, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. *Weather and Forecasting* 5.4, pp. 576–585 (cit. on pp. [84](#), [251](#)).
- Dousse, J.P. (1992). Systeme d’information géographique et télédétection satellitaire: deux nouveaux instruments dans la recherche du pergélisol alpin. *Travail de diplôme, Institut de Géographie de l’Université de Fribourg* (cit. on p. [23](#)).
- Dousse, J.P. and Lugon, R. (1996). Analyse spatiale numérique du pergélisol alpin dans les Alpes Valaisannes. *Simulation der Permafrostverbreitung in den Alpen mit geographischen Informationssystemen*, pp. 19–24 (cit. on p. [23](#)).
- Duda, R. O., Hart, P. E., and Stork, D. G. (2001). Pattern classification. 2nd. John Wiley & Sons., p. 55 (cit. on p. [81](#)).

- Ebohon, B. and Schrott, L. (2008). Modeling Mountain Permafrost Distribution: A New Permafrost Map of Austria. *Proceedings of the 9th International Conference on Permafrost*. Ed. by D. Kane and K. Hinkel. Fairbanks, Alaska, pp. 397–402 (cit. on pp. 4, 23, 223, 241).
- Edwards, L. S. (1977). A modified pseudosection for resistivity and IP. *Geophysics* 42.5, pp. 1020–1036 (cit. on p. 36).
- EEA (2017). *Climate change, impacts and vulnerability in Europe 2016. An indicator-based report*. Tech. rep. European Environment Agency (EEA) Report No 1 (cit. on p. 1).
- Etzelmüller, B. (2013). Recent advances in mountain permafrost research. *Permafrost and Periglacial Processes* 24.2, pp. 99–107 (cit. on pp. 1, 3, 9, 187).
- Etzelmüller, B. and Frauenfelder, R. (2009). Factors controlling the distribution of mountain permafrost in the Northern Hemisphere and their influence on sediment transfer. *Arctic, Antarctic, and Alpine Research* 41.1, pp. 48–58 (cit. on p. 241).
- Etzelmüller, B. and Hagen, J. O. (2005). Glacier-permafrost interaction in Arctic and alpine mountain environments with examples from southern Norway and Svalbard. *Geological Society, London, Special Publications* 242.1, pp. 11–27 (cit. on pp. 18, 19).
- Etzelmüller, B., Hoelzle, M., Flo Heggem, E. S., Isaksen, K., Mittaz, C., Mühlh, D. V., Ødegård, R. S., Haerberli, W., and Sollid, J. L. (2001). Mapping and modelling the occurrence and distribution of mountain permafrost. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography* 55.4, pp. 186–194 (cit. on pp. 10, 243).
- Etzelmüller, B., Heggem, E. S., Sharkhuu, N., Frauenfelder, R., Kääb, A., and Goulden, C. (2006). Mountain permafrost distribution modelling using a multi-criteria approach in the Hövsgöl area, northern Mongolia. *Permafrost and Periglacial Processes* 17.2, pp. 91–104 (cit. on pp. 23, 223).
- Etzelmüller, B., Farbro, H., Gudmundsson, A., Humlum, O., Tveito, O. E., and Björnsson, H. (2007). The regional distribution of mountain permafrost in Iceland. *Permafrost and Periglacial Processes* 18.2, pp. 185–199 (cit. on pp. 23, 223).
- Everitt, B. S., Landau, S., Leese, M., and Stahl, D. (2011). Miscellaneous clustering methods. *Cluster Analysis, 5th Edition*, pp. 215–255 (cit. on p. 248).
- Fallot, J.-M., Rome, S., and Bigot, S. (2012). Influence de la topographie et des accumulations d'air froid sur les températures moyennes mensuelles et annuelles en Suisse. Laboratoire d'étude des Transferts en Hydrologie et Environnement, Université Joseph Fourier (Grenoble 1), Domaine Universitaire, BP53, F-38041 Grenoble cedex 9, France: Service IRIS, Grenoble (cit. on p. 39).
- Fan, G.-F., Wang, W.-S., Liu, C., and DAI, H.-Z. (2008). Wind power prediction based on artificial neural network [J]. *Proceedings of the CSEE* 34, pp. 118–123 (cit. on p. 74).
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters* 27.8, pp. 861–874 (cit. on pp. 86, 230, 252).
- Field, A. (2009). *Discovering statistics using SPSS*. Sage publications (cit. on pp. 73, 74).
- Fischer, M., Huss, M., Barboux, C., and Hoelzle, M. (2014). The new Swiss Glacier Inventory SGI2010: relevance of using high-resolution source data in areas dominated by very small glaciers. *Arctic, Antarctic, and Alpine Research* 46.4, pp. 933–945 (cit. on p. 172).
- Fischer, M., Huss, M., and Hoelzle, M. (2015). Surface elevation and mass changes of all Swiss glaciers 1980–2010. *The Cryosphere* 9.2, pp. 525–540 (cit. on p. 1).



- Fontana, F., Rixen, C., Jonas, T., Aberegg, G., and Wunderle, S. (2008). Alpine grassland phenology as seen in AVHRR, VEGETATION, and MODIS NDVI time series—a comparison with in situ measurements. *Sensors* 8.4, pp. 2833–2853 (cit. on p. 43).
- Forsyth, D. A. and Ponce, J. (2012). *Computer vision: a modern approach*. Upper Saddle River, NJ; London: Prentice Hall., pp. 519–539 (cit. on p. 63).
- Fotheringham, A. S., Brunsdon, C., and Charlton, M. (2000). *Quantitative geography: perspectives on spatial data analysis*. Sage (cit. on p. 3).
- Franco, B. and Héty, B. (1989). Eboulis et autres formations de pente hétérométriques: contribution à une terminologie géomorphologique (cit. on p. 16).
- Frauenfelder, R. (1998). Rock Glaciers, Fletschhorn Area, Valais, Switzerland. *International Permafrost Association, Data and Information Working Group, NSIDC, University of Colorado at Boulder* (cit. on pp. 14, 23, 223).
- Frauenfelder, R. (2004). „Regional-scale modelling of the occurrence and dynamics of rockglaciers and the distribution of paleopermafrost“. PhD thesis. University of Zurich (cit. on p. 15).
- French, H. M. (2013). *The periglacial environment*. John Wiley & Sons (cit. on pp. 10, 16).
- Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American statistical association* 84.405, pp. 165–175 (cit. on p. 248).
- Fukunaga, K. (1982). 15 Intrinsic dimensionality extraction. *Handbook of statistics* 2, pp. 347–360 (cit. on p. 57).
- Funk, M. and Hoelzle, M. (1992). A model of potential direct solar radiation for investigating occurrences of mountain permafrost. *Permafrost and Periglacial Processes* 3.2, pp. 139–142 (cit. on p. 223).
- Furrer, H. and Hutter, J. (1994). Die Vegetation im Corvatsch-Gebiet und ihre Beziehung zum Permafrost. Unveroff. Diplomarbeit, ETH Zurich. (cit. on p. 12).
- Gardaz, J.-M. (1997). Distribution of mountain permafrost, Fontanesses Basin, Valaisian Alps, Switzerland. *Permafrost and Periglacial Processes* 8.1, pp. 101–105 (cit. on pp. 33, 165).
- Gardaz, J.-M. (1998). „Permafrost Prospecting, Periglacial and Rock Glacier Hydrology in Mountain Areas: Case Studies in the Valais Alps, Switzerland“. PhD thesis. University of Fribourg (cit. on pp. 33, 165).
- Geilhausen, M., Otto, J.-C., and Schrott, L. (2012). Spatial distribution of sediment storage types in two glacier landsystems (Pasterze & Obersulzbachkees, Hohe Tauern, Austria). *Journal of Maps* 8.3, pp. 242–259 (cit. on p. 2).
- Ghose, M. K., Pradhan, R., and Ghose, S. S. (2010). Decision tree classification of remotely sensed satellite data using spectral separability matrix. *International Journal of Advanced Computer Science and Applications* 1.5, pp. 93–101 (cit. on p. 81).
- Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., and Stoffel, M. (2014). 21st century climate change in the European Alps - a review. *Science of The Total Environment* 493, pp. 1138–1151 (cit. on pp. 1, 2).
- Golay, J. and Kanevski, M. (2015). A new estimator of intrinsic dimension based on the multipoint Morisita index. *Pattern recognition* 48.12, pp. 4070–4081 (cit. on pp. 57, 96).

- Golay, J. and Kanevski, M. (2016). Unsupervised Feature Selection Based on the Morisita Estimator of Intrinsic Dimension. *Knowledge-Based Systems* 135, pp. 125–134 (cit. on p. 57).
- Golay, J. and Laib, M. (2016). Intrinsic Dimension for Data Mining, R Package (cit. on p. 57).
- Grohmann, C. H., Smith, M. J., and Riccomini, C. (2011). Multiscale analysis of topographic surface roughness in the Midland Valley, Scotland. *IEEE Transactions on Geoscience and Remote Sensing* 49.4, pp. 1200–1213 (cit. on pp. 44, 45).
- Gruber, S. (2005). Mountain permafrost: Transient spatial modelling, model verification and the use of remote sensing. *Doktorarbeit, Universität Zürich, Zürich* (cit. on p. 3).
- Gruber, S. (2012). Derivation and analysis of a high-resolution estimate of global permafrost zonation. *The Cryosphere* 6.1, p. 221 (cit. on p. 26).
- Gruber, S. and Haeberli, W. (2007). Permafrost in steep bedrock slopes and its temperature-related destabilization following climate change. *Journal of Geophysical Research: Earth Surface (2003–2012)* 112.F2 (cit. on pp. 2, 241).
- Gruber, S. and Haeberli, W. (2009). Mountain permafrost. *Permafrost soils*. Springer, pp. 33–44 (cit. on pp. 1, 3, 9, 10, 14, 185, 187).
- Gruber, S. and Hoelzle, M. (2001). Statistical modelling of mountain permafrost distribution: local calibration and incorporation of remotely sensed data. *Permafrost and Periglacial Processes* 12.1, pp. 69–77 (cit. on pp. 23, 223).
- Gruber, S. and Hoelzle, M. (2008). The cooling effect of coarse blocks revisited: a modeling study of a purely conductive mechanism. *Proceedings of the 9th International Conference on Permafrost*. Vol. 29, pp. 557–561 (cit. on p. 44).
- Gruber, S., Peter, M., Hoelzle, M., Woodhatch, I., and Haeberli, W. (2003). Surface temperatures in steep alpine rock faces - a strategy for regional-scale measurement and modelling. *Proceedings of the 8th International Conference on Permafrost*. Vol. 1. Swets & Zeitlinger Lisse, pp. 325–330 (cit. on pp. 4, 11, 20, 24).
- Gruber, S., Hoelzle, M., and Haeberli, W. (2004a). Permafrost thaw and destabilization of Alpine rock walls in the hot summer of 2003. *Geophysical Research Letters* 31.13 (cit. on pp. 11, 242).
- Gruber, S., Hoelzle, M., and Haeberli, W. (2004b). Rock-wall temperatures in the Alps: modelling their topographic distribution and regional differences. *Permafrost and Periglacial Processes* 15.3, pp. 299–307 (cit. on pp. 4, 23–25, 27, 223, 235).
- Guglielmin, M., Aldighieri, B., and Testa, B. (2003). PERMACLIM: a model for the distribution of mountain permafrost, based on climatic observations. *Geomorphology* 51.4, pp. 245–257 (cit. on pp. 24, 176, 243).
- Guisan, A. and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling* 135.2, pp. 147–186 (cit. on p. 67).
- Guyon, I., Weston, J., Barnhill, S., and Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine learning* 46.1, pp. 389–422 (cit. on pp. 62, 76).
- Guyon, I., Gunn, S., Nikravesh, M., and Zadeh, L. A. (2006). Feature Extraction: Foundations and Applications (Studies in Fuzziness and Soft Computing). Springer (cit. on p. 52).

- Hack, R. (2000). Geophysics for slope stability. *Surveys in geophysics* 21.4, pp. 423–448 (cit. on p. 36).
- Haeberli, W. (1973). Die Basis-Temperatur der winterlichen Schneedecke als möglicher Indikator für die Verbreitung von Permafrost in den Alpen. *Zeitschrift für Gletscherkunde und Glazialgeologie* 9, pp. 221–227 (cit. on pp. 21, 22).
- Haeberli, W. (1975). Untersuchungen zur Verbreitung von Permafrost zwischen Flüelapass und Piz Grialetsch (Graubünden). Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie (cit. on pp. 12, 14, 21, 22, 222, 234, 241).
- Haeberli, W. (1985). Creep of mountain permafrost: internal structure and flow of alpine rock glaciers. *Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie an der ETH Zurich* 77, pp. 5–142 (cit. on pp. 11, 14, 15, 18, 21, 34, 241, 242).
- Haeberli, W. (1996). Die Permafrost-Faustregeln der VAW/ETHZ-einige grundsätzliche Bemerkungen. *Simulation der Permafrostverbreitung in den Alpen mit geographischen Informationssystemen*, pp. 13–18 (cit. on pp. 12, 21, 22, 42, 158, 175, 185).
- Haeberli, W. (2005). Investigating glacier-permafrost relationships in high-mountain areas: historical background, selected examples and research needs. *Geological Society, London, Special Publications* 242.1, pp. 29–37 (cit. on p. 18).
- Haeberli, W., Hoelzle, M., Keller, F., Schmid, W., Vonder Mühl, D., and Wagner, S. (1993). Monitoring the long-term evolution of mountain permafrost in the Swiss Alps. *Proceedings of the 6th International Conference on Permafrost*. Vol. 1, pp. 214–219 (cit. on pp. 9, 10).
- Haeberli, W., Hallet, B., Arenson, L., Elconin, R., Humlum, O., Käab, A., Kaufmann, V., Ladanyi, B., Matsuoka, N., Springman, S., et al. (2006). Permafrost creep and rock glacier dynamics. *Permafrost and Periglacial Processes* 17.3, pp. 189–214 (cit. on p. 19).
- Hall, K. (1997). Rock temperatures and implications for cold region weathering. I: New data from Viking Valley, Alexander Island, Antarctica. *Permafrost and Periglacial Processes* 8.1, pp. 69–90 (cit. on p. 21).
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 11.1, pp. 10–18 (cit. on pp. 87, 252).
- Hall, M. A. and Smith, L. A. (1997). Feature subset selection: a correlation based filter approach. *International Conference on Neural Information Processing and Intelligent Information Systems*, pp. 855–858 (cit. on p. 60).
- Hall, M. A. and Smith, L. A. (1999). Feature Selection for Machine Learning: Comparing a Correlation-Based Filter Approach to the Wrapper. *FLAIRS conference*. Vol. 1999, pp. 235–239 (cit. on p. 62).
- Hamel, L. (2009). Model assessment with ROC curves. *Encyclopedia of Data Warehousing and Mining, Second Edition*. IGI Global, pp. 1316–1323 (cit. on p. 230).
- Hampe, A. (2004). Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography* 13.5, pp. 469–471 (cit. on p. 67).
- Han, J., Pei, J., and Kamber, M. (2011). Data mining: concepts and techniques. Elsevier (cit. on p. 81).

- Haneberg, W. C., Creighton, A. L., Medley, E. W., and Jonas, D. A. (2005). Use of LiDAR to assess slope hazards at the Lihir gold mine, Papua New Guinea. *Proceedings, International Conference on Landslide Risk Management, Vancouver, British Columbia* (cit. on p. 45).
- Hanson, S. and Hoelzle, M. (2004). The thermal regime of the active layer at the Murtel rock glacier based on data from 2002. *Permafrost and Periglacial Processes* 15.3, pp. 273–282 (cit. on pp. 3, 11, 12, 16).
- Harris, C. and Murton, J. B. (2005). Interactions between glaciers and permafrost: An introduction. *Geological Society, London, Special Publication* 242.1, pp. 1–9 (cit. on p. 241).
- Harris, C., Mühll, D. V., Isaksen, K., Haeberli, W., Sollid, J. L., King, L., Holmlund, P., Dramis, F., Guglielmin, M., and Palacios, D. (2003). Warming permafrost in European mountains. *Global and Planetary Change* 39.3, pp. 215–225 (cit. on pp. 1, 26).
- Harris, C., Arenson, L., Christiansen, H. H., Etzelmüller, B., Frauenfelder, R., Gruber, S., Haeberli, W., Hauck, C., Hoelzle, M., Humlum, O., Isaksen, K., Kääb, A., Kern-Luetsch, M.A., Lehning, M., Matsuoka, N., Murton, J. B., Noezli, J., Phillips, M., Ross, N., Seppala, M., Springman, S. M., and Vonder Muehll, D. (2009). Permafrost and climate in Europe: Monitoring and modelling thermal, geomorphological and geotechnical responses. *Earth-Science Reviews* 92.3, pp. 117–171 (cit. on pp. 1, 2, 9, 240).
- Harris, S. A., French, H. M., Johnston, G. H., Ladanyi, B., Segó, D. C., and Everdingen, R. O. van (1988). La terminologie du pergélisol et notions connexes (cit. on p. 9).
- Hasler, A., Gruber, S., and Beutel, J. (2012). Kinematics of steep bedrock permafrost. *Journal of Geophysical Research: Earth Surface* 117.F1 (cit. on p. 2).
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). Unsupervised learning. *The elements of statistical learning*. Springer, pp. 485–585 (cit. on pp. 51, 52, 74, 242, 248).
- Hauck, C. and Kneisel, C. (2008). Electrical methods. *Applied geophysics in periglacial environments*, pp. 3–27 (cit. on pp. 36, 243).
- Hauck, C., Vonder Mühll, D., and Maurer, H. (2003). Using DC resistivity tomography to detect and characterize mountain permafrost. *Geophysical prospecting* 51.4, pp. 273–284 (cit. on pp. 36, 243).
- Haykin, S. (2009). *Neural networks and learning machines*. Vol. 3. Pearson Education Upper Saddle River (cit. on pp. 51, 248).
- Heckmann, T., Bimböse, M., Krautblatter, M., Haas, F., Becht, M., and Morche, D. (2012). From geotechnical analysis to quantification and modelling using LiDAR data: a study on rockfall in the Reintal catchment, Bavarian Alps, Germany. *Earth Surface Processes and Landforms* 37.1, pp. 119–133 (cit. on p. 186).
- Heggem, E. S. F., Juliussen, H., and Etzelmüller, B. (2005). Mountain permafrost in central-eastern Norway. *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography* 59.2, pp. 94–108 (cit. on pp. 12, 17, 23, 28).
- Hijmans, R. J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93.3, pp. 679–688 (cit. on p. 67).
- Hilbich, C. (2010). Time-lapse refraction seismic tomography for the detection of ground ice degradation. *The Cryosphere* 4.3, p. 243 (cit. on p. 33).

- Hilbich, C., Hauck, C., Delaloye, R., and Hoelzle, M. (2008). A geoelectric monitoring network and resistivity-temperature relationships of different mountain permafrost sites in the Swiss Alps. *Proceedings 9th International Conference on Permafrost, June*, pp. 699–704 (cit. on p. 33).
- Hilbich, C., Marescot, L., Hauck, C., Loke, M. H., and Mäusbacher, R. (2009). Applicability of electrical resistivity tomography monitoring to coarse blocky and ice-rich permafrost landforms. *Permafrost and Periglacial Processes* 20.3, pp. 269–284 (cit. on pp. 36, 224, 243).
- Hoelzle, M. (1992). Permafrost occurrence from BTS measurements and climatic parameters in the Eastern Swiss Alps. *Permafrost and Periglacial Processes* 3.2, pp. 143–147 (cit. on p. 3).
- Hoelzle, M. (1994). Permafrost und Gletscher im Oberengadin: Grundlagen und Anwendungsbeispiele für automatisierte Schätzverfahren. Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie an der Eidgenössischen Technischen Hochschule Zürich. VAW (cit. on pp. 4, 40, 241, 245).
- Hoelzle, M. and Haeberli, W. (1995). Simulating the effects of mean annual air-temperature changes on permafrost distribution and glacier size: an example from the Upper Engadin, Swiss Alps. *Annals of Glaciology* 21.1, pp. 399–405 (cit. on pp. 23, 176, 223).
- Hoelzle, M., Wegmann, M., and Krummenacher, B. (1999). Miniature temperature dataloggers for mapping and monitoring of permafrost in high mountain areas: first experience from the Swiss Alps. *Permafrost and Periglacial Processes* 10.2, pp. 113–124 (cit. on pp. 12, 36, 243).
- Hoelzle, M., Mittaz, C., Etzelmüller, B., and Haeberli, W. (2001). Surface energy fluxes and distribution models of permafrost in European mountain areas: an overview of current developments. *Permafrost and Periglacial Processes* 12.1, pp. 53–68 (cit. on pp. 4, 10–12, 20, 24).
- Hosmer, D. W. and Lemeshow, S. (2000). Introduction to the logistic regression model. Wiley Online Library, pp. 1–30 (cit. on pp. 73, 249).
- Huggel, C., Fischer, L., Schneider, D., and Haeberli, W. (2010). Research advances on climate-induced slope instability in glacier and permafrost high-mountain environments. *Geographica Helvetica* 65.2, pp. 146–156 (cit. on p. 2).
- Humlum, O. (1996). Origin of rock glaciers: observations from Mellemfjord, Disko Island, central West Greenland. *Permafrost and Periglacial Processes* 7.4, pp. 361–380 (cit. on pp. 34, 242).
- Hunt, E. B., Marin, J., and Stone, P. J. (1966). Experiments in induction. (Cit. on p. 79).
- Huss, M., Bookhagen, B., Huggel, C., Jacobsen, D., Bradley, R. S., Clague, J. J., Vuille, M., Buytaert, W., Cayan, D. R., Greenwood, G., et al. (2017). Toward mountains without permanent snow and ice. *Earth's Future* 5.5, pp. 418–435 (cit. on p. 1).
- Imhof, M. (1996a). Modelling and verification of the permafrost distribution in the Bernese Alps (Western Switzerland). *Permafrost and Periglacial Processes* 7.3, pp. 267–280 (cit. on p. 22).
- Imhof, M. (1996b). PERM: ein Programm für die automatisierte Kartierung von Permafrost in den Schweizer Alpen. *Simulation der Permafrostverbreitung in den Alpen mit geographischen Informationssystemen*, pp. 25–33 (cit. on pp. 222, 223).

- Inselberg, A. (1997). Multidimensional detective. *Information Visualization, 1997. Proceedings., IEEE Symposium on.* IEEE, pp. 100–107 (cit. on p. 54).
- Inselberg, A. (2009). *Parallel coordinates.* Springer (cit. on p. 54).
- Inselberg, A. and Dimsdale, B. (1990). Parallel coordinates: a tool for visualizing multi-dimensional geometry. *Proceedings of the 1st conference on Visualization'90.* IEEE Computer Society Press, pp. 361–378 (cit. on p. 54).
- IPCC (2001). *Climate Change 2001: The scientific basis. Contribution of working group 1 to the third assessment report of the intergovernmental panel on climate change.* Cambridge University Press, Cambridge, UK and NY, USA.: IPCC (cit. on p. 1).
- IPCC (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change.* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva (Switzerland): IPCC (cit. on p. 1).
- Ivy-Ochs, S., Kerschner, H., Maisch, M., Christl, M., Kubik, P. W., and Schlüchter, C. (2009). Latest Pleistocene and Holocene glacier variations in the European Alps. *Quaternary Science Reviews* 28.21, pp. 2137–2149 (cit. on p. 41).
- Izenman, A. J. (2008). *Modern multivariate statistical techniques: regression, classification, and Manifold Learning.* Springer Texts in Statistics, New York (cit. on pp. 51, 248).
- Jamieson, A. R., Giger, M. L., Drukker, K., Li, H., Yuan, Y., and Bhooshan, N. (2010). Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE. *Medical physics* 37.1, pp. 339–351 (cit. on p. 54).
- Janke, J. and Frauenfelder, R. (2008). The relationship between rock glacier and contributing area parameters in the Front Range of Colorado. *Journal of Quaternary Science* 23.2, pp. 153–163 (cit. on p. 15).
- Jenness, J. (2002). *Surface Areas and Ratios from Elevation Grid (surfgrids. avx) extension for ArcView 3. x-version 1.2.* Jenness Enterprises. Tech. rep. (cit. on p. 45).
- Johnson, A. W. (1966). Plant ecology in permafrost areas. *Permafrost international conference: proceedings 11-15 November 1963 Lafayette, Indiana.* 1287. National Academies, p. 25 (cit. on p. 12).
- Justice, C. O., Townshend, J. R. G., Holben, B. N., and Tucker, C. J. (1985). Analysis of the phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing* 6.8, pp. 1271–1318 (cit. on p. 43).
- Kääb, A. and Kneisel, C. (2006). Permafrost creep within a recently deglaciated glacier forefield: Muragl, Swiss Alps. *Permafrost and Periglacial Processes* 17.1, pp. 79–85 (cit. on p. 187).
- Kääb, A. and Weber, M. (2004). Development of transverse ridges on rock glaciers: field measurements and laboratory experiments. *Permafrost and Periglacial Processes* 15.4, pp. 379–391 (cit. on p. 14).
- Kääb, A., Frauenfelder, R., and Roer, I. (2007). On the response of rockglacier creep to surface temperature increase. *Global and Planetary Change* 56.1, pp. 172–187 (cit. on pp. 2, 241).

- Kalbermatten, M., Van De Ville, D., Turberg, P., Tuia, D., and Joost, S. (2012). Multiscale analysis of geomorphological and geological features in high resolution digital elevation models using the wavelet transform. *Geomorphology* 138.1, pp. 352–363 (cit. on p. 186).
- Kanevski, M. (2008). Advanced Mapping of Environmental Data: Geostatistics. *Machine Learning and Bayesian Maximum Entropy*. iSTE and Wiley, London (cit. on p. 51).
- Kanevski, M., Arutyunyan, R., Bolshov, L., Demyanov, V., and Maignan, M. (1996). Artificial neural networks and spatial estimation of Chernobyl fallout. *Geoinformatics* 7.1-2, pp. 5–11 (cit. on p. 51).
- Kanevski, M., Pozdnoukhov, A., and Timonin, V. (2009). Machine learning for spatial environment data. Theory, applications and software. EPFL Press (cit. on pp. 55, 65, 67, 74–78, 224, 248).
- Kanevski, M. F. (1999). Spatial predictions of soil contamination using general regression neural networks. *SYSTEMS RESEARCH AND INFORMATION SCIENCE* 8, pp. 241–256 (cit. on p. 51).
- Kawabata, D. and Bandibas, J. (2009). Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). *Geomorphology* 113.1, pp. 97–109 (cit. on p. 74).
- Keller, F. (1992). Automated mapping of mountain permafrost using the program PERMAKART within the geographical information system ARC/INFO. *Permafrost and Periglacial Processes* 3.2, pp. 133–138 (cit. on pp. 22, 176, 222, 223, 235).
- Keller, F. (1994). Interaktionen zwischen Schnee und Permafrost, eine Grundlagenstudie im Oberengadin. Centre d'études nordiques-Université Laval (cit. on p. 11).
- Keller, F. and Hoelzle, M. (1996). PERMAKART und PERMAMAP. *Haeblerli, W., Hölzle, M., Dousse, JP, Gardaz, JM, Imhof, M., Keller, F., Kunz P., Lugon, R., Reynard, E., Simulation der Permafrostverbreitung in den Alpen mit geographischen Informationssystemen, Arbeitsbericht FP 31*, pp. 35–46 (cit. on p. 22).
- Keller, F., Lugon, R., and Reynard, E. (1998). Permafrost map of Switzerland. Centre d'études nordiques-Université Laval (cit. on pp. 23, 223).
- Kellerer-Pirklbauer, A. (2017). Potential weathering by freeze-thaw action in alpine rocks in the European Alps during a nine year monitoring period. *Geomorphology* 296, pp. 113–131 (cit. on p. 2).
- Kellerer-Pirklbauer, A. and Kaufmann, V. (2012). About the relationship between rock glacier velocity and climate parameters in central Austria. *Austrian Journal of Earth Sciences* 105.2 (cit. on p. 14).
- Kellerer-Pirklbauer, A. and Kaufmann, V. (2017). Deglaciation and its impact on permafrost and rock glacier evolution: New insight from two adjacent cirques in Austria. *Science of The Total Environment* (cit. on p. 2).
- Kellerer-Pirklbauer, A., Lieb, G. K., Avian, M., and Carrivick, J. (2012). Climate change and rock fall events in high mountain areas: Numerous and extensive rock falls in 2007 at Mittlerer Burgstall, Central Austria. *Geografiska Annaler: Series A, Physical Geography* 94.1, pp. 59–78 (cit. on p. 20).

- Kenner, R., Bühler, Y., Delaloye, R., Ginzler, C., and Phillips, M. (2014). Monitoring of high alpine mass movements combining laser scanning with digital airborne photogrammetry. *Geomorphology* 206, pp. 492–504 (cit. on p. 14).
- Kira, K. and Rendell, L. A. (1992). The feature selection problem: Traditional methods and a new algorithm. *Aaai*. Vol. 2, pp. 129–134 (cit. on p. 61).
- Kirkbride, M. P. and Deline, P. (2013). The formation of supraglacial debris covers by primary dispersal from transverse englacial debris bands. *Earth Surface Processes and Landforms* 38.15, pp. 1779–1792 (cit. on p. 19).
- Kleinbaum, D. G. and Klein, M. (1994). Logistic regression: a self-learning text. Springer-Verlag (cit. on pp. 73, 249).
- Kneisel, C. (2003). Permafrost in recently deglaciated glacier forefields—measurements and observations in the eastern Swiss Alps and northern Sweden. *Zeitschrift für Geomorphologie, NF*, pp. 289–305 (cit. on pp. 18, 19, 187).
- Kneisel, C. and Kääh, A. (2007). Mountain permafrost dynamics within a recently exposed glacier forefield inferred by a combined geomorphological, geophysical and photogrammetrical approach. *Earth Surface Processes and Landforms* 32.12, pp. 1797–1810 (cit. on pp. 18, 19, 241, 257).
- Kobierska, F., Jonas, T., Magnusson, J., Zappa, M., Bavay, M., Bosshard, T., Paul, F., and Bernasconi, S. M. (2011). Climate change effects on snow melt and discharge of a partly glacierized watershed in Central Switzerland (SoilTrec Critical Zone Observatory). *Applied geochemistry* 26, S60–S62 (cit. on pp. 2, 241).
- Koenig, W. D. (2002). Global patterns of environmental synchrony and the Moran effect. *Ecography* 25.3, pp. 283–288 (cit. on p. 67).
- Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence* 97.1-2, pp. 273–324 (cit. on p. 62).
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics* 43.1, pp. 59–69 (cit. on pp. 68, 147, 248).
- Kohonen, T. (2000). Self-organizing maps. Springer (cit. on pp. 51, 68, 69, 147).
- Kohonen, T. and Honkela, T. (2007). Kohonen network. *Scholarpedia* 2.1, p. 1568 (cit. on p. 248).
- Kononenko, I. (1994). Estimating attributes: analysis and extensions of RELIEF. *European conference on machine learning*. Springer, pp. 171–182 (cit. on p. 61).
- Krautblatter, M., Funk, D., and Günzel, F. K. (2013). Why permafrost rocks become unstable: a rock–ice–mechanical model in time and space. *Earth Surface Processes and Landforms* 38.8, pp. 876–887 (cit. on p. 2).
- Kriegler, F. J., Malila, W. A., Nalepka, R. F., and Richardson, W. (1969). Preprocessing transformations and their effects on multispectral recognition. *Remote Sensing of Environment*, VI, p. 97 (cit. on p. 43).
- Krysiecki, J.-M., Bodin, X., and Schoeneich, P. (2008). Collapse of the Bérard rock glacier (southern French Alps). *9th International Conference on Permafrost: Institute of Northern Engineering, University of Alaska Fairbanks*, pp. 153–154 (cit. on p. 14).



- Krzanowski, W. J. and Lai, Y. T. (1988). A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, pp. 23–34 (cit. on p. 71).
- Kuligowski, R. J. and Barros, A. P. (1998). Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks. *Weather and Forecasting* 13.4, pp. 1194–1204 (cit. on p. 74).
- Kummert, M., Delaloye, R., and Braillard, L. (2017). Erosion and sediment transfer processes at the front of rapidly moving rock glaciers: Systematic observations with automatic cameras in the western Swiss Alps. *Permafrost and Periglacial Processes* (cit. on pp. 33, 171).
- Lambiel, C. (1999). Inventaire des glaciers rocheux entre le val de Bagnes et le val d'Hérémence (Valais) (cit. on pp. 160, 228).
- Lambiel, C. (2006). Le pergélisol dans les terrains sédimentaires à forte déclivité: distribution, régime thermique et instabilités. UNIL-Faculté des géosciences et de l'environnement-Institut de géographie (cit. on pp. 33, 34, 37, 153, 155, 156, 165, 166, 169–171, 174, 224, 243).
- Lambiel, C. (2011). Le glacier rocheux déstabilisé de Tsaté-Moiry (VS): caractéristiques morphologiques et vitesses de déplacement. *Géovisions* 36, pp. 211–224 (cit. on pp. 34, 174).
- Lambiel, C. and Baron, L. (2008). Two-dimensional geoelectrical monitoring in an alpine frozen moraine. *Proceed. 9th Int. Conf. Permafrost, Fairbanks 2008*. Vol. Extended Abstracts Volume, pp. 161–162 (cit. on p. 159).
- Lambiel, C. and Delaloye, R. (2004). Contribution of real-time kinematic GPS in the study of creeping mountain permafrost: Examples from the Western Swiss Alps. *Permafrost and Periglacial Processes* 15.3, pp. 229–241 (cit. on pp. 33, 223).
- Lambiel, C. and Pieracci, K. (2008). Permafrost distribution in talus slopes located within the alpine periglacial belt, Swiss Alps. *Permafrost and Periglacial Processes* 19.3, pp. 293–304 (cit. on pp. 3, 17, 33, 37, 38, 154, 223, 224, 241, 243, 257).
- Lambiel, C. and Reynard, E. (2001). Regional modelling of present, past and future potential distribution of discontinuous permafrost based on a rock glacier inventory in the Bagnes-Hérémence area (Western Swiss Alps). *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography* 55.4, pp. 219–223 (cit. on pp. 23, 32, 158, 163, 176, 223, 242, 256).
- Lambiel, C. and Reynard, E. (2003). Cartographie de la distribution du pergélisol et datation des glaciers rocheux dans la région du Mont Gelé (Valais). *Physische Geographie* 41, pp. 91–104 (cit. on pp. 33, 34, 156, 160, 171, 242).
- Lambiel, C. and Schuetz, P. (2008). Ground characteristics and deformation of a frozen moraine affected by tourist infrastructures (Col des Gentianes, Valais). *Klimaveränderungen auf der Spur. Studien des Europäischen Tourismus Instituts an der Academia Engiadina, Samedan*. (Cit. on pp. 18, 159, 224).
- Lambiel, C., Delaloye, R., Strozzi, T., Lugon, R., and Raetzo, H. (2008). ERS InSAR for assessing rock glacier activity. *Ninth International Conference on Permafrost, edited by: Kane, DL, Hinkel, KM, Fairbanks: University of Alaska Fairbanks*, pp. 1019–1024 (cit. on p. 171).
- Lambiel, C., Maillard, B., Kummert, M., and Reynard, E. (2016). Geomorphology of the Hérens valley (Swiss Alps). *Journal of Maps* 12.1, pp. 160–172 (cit. on pp. 34, 157).

- Lambiel, C., Rüttimann, S., Meyrat, R., and Vivero, S. (2017). Capturing the crisis of an active rock glacier with UAV survey. *EGU General Assembly Conference Abstracts*. Vol. 19, p. 7014 (cit. on p. 33).
- Lane, S. N., Tayefi, V., Reid, S. C., Yu, D., and Hardy, R. J. (2007). Interactions between sediment delivery, channel change, climate change and flood risk in a temperate upland environment. *Earth Surface Processes and Landforms* 32.3, pp. 429–446 (cit. on pp. 2, 241).
- Le Boudec, J.-Y. (2010). Performance evaluation of computer and communication systems. Epfl Press (cit. on p. 147).
- Le Cessie, S. and Van Houwelingen, J. C. (1994). Logistic regression for correlated binary data. *Applied Statistics* 43, pp. 95–108 (cit. on pp. 74, 249).
- Lee, J. A. and Verleysen, M. (2007). Nonlinear dimensionality reduction. Springer Science & Business Media (cit. on p. 52).
- Lepore, C., Kamal, S. A., Shanahan, P., and Bras, R. L. (2012). Rainfall-induced landslide susceptibility zonation of Puerto Rico. *Environmental Earth Sciences* 66.6, pp. 1667–1681 (cit. on p. 72).
- Leuenberger, M. and Kanevski, M. (2017). Study of Environmental Data Complexity using Extreme Learning Machine. *EGU General Assembly Conference Abstracts*. Vol. 19, p. 14015 (cit. on p. 55).
- Lewkowicz, A. G. and Ednie, M. (2004). Probability mapping of mountain permafrost using the BTS method, Wolf Creek, Yukon Territory, Canada. *Permafrost and Periglacial Processes* 15.1, pp. 67–80 (cit. on pp. 11, 28, 72, 144, 177, 182).
- Li, G. and Shi, J. (2010). On comparing three artificial neural networks for wind speed forecasting. *Applied Energy* 87.7, pp. 2313–2320 (cit. on p. 74).
- Lilleøren, K. S., Etzelmüller, B., Gärtner-Roer, I., Käab, A., Westermann, S., and Guðmundsson, Á. (2013). The Distribution, Thermal Characteristics and Dynamics of Permafrost in Tröllaskagi, Northern Iceland, as Inferred from the Distribution of Rock Glaciers and Ice-Cored Moraines. *Permafrost and Periglacial Processes* 24.4, pp. 322–335 (cit. on p. 42).
- Liu, S., Maljovec, D., Wang, B., Bremer, P.-T., and Pascucci, V. (2017). Visualizing high-dimensional data: Advances in the past decade. *IEEE Transactions on Visualization and Computer Graphics* 23.3, pp. 1249–1268 (cit. on p. 53).
- Loye, A., Jaboyedoff, M., and Pedrazzini, A. (2009). Identification of potential rockfall source areas at a regional scale using a DEM-based geomorphometric analysis. *Natural Hazards and Earth System Sciences* 9.5, pp. 1643–1653 (cit. on p. 228).
- Lugon, R. and Delaloye, R. (2001). Modelling alpine permafrost distribution, Val de Re'chy, Valais Alps (Switzerland). *Norsk Geografisk Tidsskrift-Norwegian Journal of Geography* 55.4, pp. 224–229 (cit. on p. 23).
- Lukas, S. (2011). Ice-cored moraines. *Encyclopedia of Snow, Ice and Glaciers*. Springer, pp. 616–619 (cit. on p. 42).
- Luoto, M. and Hjort, J. (2005). Evaluation of current statistical approaches for predictive geomorphological mapping. *Geomorphology* 67.3, pp. 299–315 (cit. on p. 242).
- Maaten, L. Van der and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research* 9.Nov, pp. 2579–2605 (cit. on pp. 54, 95).

- Magnin, F., Brenning, A., Bodin, X., Deline, P., and Ravanel, L. (2015). Statistical modelling of rock wall permafrost distribution: application to the Mont Blanc massif. *Geomorphologie: relief, processus, environnement*, p. 20 (cit. on pp. 4, 25, 39, 133, 241, 242).
- Maisch, M. (2000). Die Gletscher der Schweizer Alpen: Gletscherhochstand 1850, aktuelle Vergletscherung, Gletscherschwund-Szenarien. Vdf, Hochschulverlag AG an der ETH (cit. on pp. 41, 228).
- Maqsood, I., Khan, M. R., and Abraham, A. (2004). An ensemble of neural networks for weather forecasting. *Neural Computing & Applications* 13.2, pp. 112–122 (cit. on p. 74).
- Marescot, L., Loke, M. H., Chapellier, D., Delaloye, R., Lambiel, C., and Reynard, E. (2003). Assessing reliability of 2D resistivity imaging in mountain permafrost studies using the depth of investigation index method. *Near Surface Geophysics* 1.2, pp. 57–67 (cit. on pp. 33, 37, 224, 243).
- Marmion, M., Hjort, J., Thuiller, W., and Luoto, M. (2008). A comparison of predictive methods in modelling the distribution of periglacial landforms in Finnish Lapland. *Earth Surface Processes and Landforms* 33.14, pp. 2241–2254 (cit. on p. 242).
- Marmy, A., Rajczak, J., Delaloye, R., Hilbich, C., Hoelzle, M., Kotlarski, S., Lambiel, C., Noetzli, J., Phillips, M., Salzmann, N., et al. (2015). Semi-automated calibration method for modelling of mountain permafrost evolution in Switzerland. *The Cryosphere Discussions* 9.5, pp. 4787–4843 (cit. on p. 5).
- Masters, T. (1993). Practical neural network recipes in C++. Morgan Kaufmann (cit. on p. 74).
- Matsumoto, H., Yamada, S., and Hirakawa, K. (2010). Relationship between ground ice and solifluction: Field measurements in the Daisetsu Mountains, northern Japan. *Permafrost and Periglacial Processes* 21.1, pp. 78–89 (cit. on p. 18).
- Matsuoka, N. (2001). Solifluction rates, processes and landforms: a global review. *Earth-Science Reviews* 55.1, pp. 107–134 (cit. on p. 18).
- McLachlan, G. (2004). Discriminant analysis and statistical pattern recognition. Vol. 544. John Wiley & Sons (cit. on p. 248).
- Melchiorre, C., Matteucci, M., Azzoni, A., and Zanchi, A. (2008). Artificial neural networks and cluster analysis in landslide susceptibility zonation. *Geomorphology* 94.3, pp. 379–400 (cit. on p. 74).
- Micheletti, N., Foresti, L., Robert, S., Leuenberger, M., Pedrazzini, A., Jaboyedoff, M., and Kanevski, M. (2014). Machine learning feature selection methods for landslide susceptibility mapping. *Mathematical Geosciences* 46.1, pp. 33–57 (cit. on pp. 76, 86, 252).
- Micheletti, N., Lambiel, C., and Lane, S. N. (2015). Investigating decadal-scale geomorphic dynamics in an alpine mountain setting. *Journal of Geophysical Research: Earth Surface* 120.10, pp. 2155–2175 (cit. on p. 33).
- Mika, S., Schölkopf, B., Smola, A. J., Müller, K.-R., Scholz, M., and Rätsch, G. (1998). Kernel PCA and De-noising in feature spaces. *Advances in neural information processing systems*. Vol. 11, pp. 536–542 (cit. on p. 57).
- Mitchell, T. M. (1997). Machine learning. 1997. *Burr Ridge, IL: McGraw Hill* 45.37, pp. 870–877 (cit. on p. 60).

- Mitra, P., Shankar, B. U., and Pal, S. K. (2004). Segmentation of multispectral remote sensing images using active support vector machines. *Pattern recognition letters* 25.9, pp. 1067–1074 (cit. on p. 76).
- Mittaz, C., Hoelzle, M., and Haeberli, W. (2000). First results and interpretation of energy-flux measurements over Alpine permafrost. *Annals of Glaciology* 31.1, pp. 275–280 (cit. on pp. 4, 20, 24).
- Mittaz, C., Imhof, M., Hoelzle, M., and Haeberli, W. (2002). Snowmelt Evolution Mapping Using an Energy Balance Approach over an Alpine Terrain. *Arctic, Antarctic, and Alpine Research* 34.3, pp. 274–281 (cit. on pp. 40, 245).
- Mohamed, A.-R., Hinton, G., and Penn, G. (2012). Understanding how deep belief networks perform acoustic modelling. *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, pp. 4273–4276 (cit. on p. 54).
- Mollaret, C., Hilbich, C., and Hauck, C. (2017). Analysis Procedures of an ERT Monitoring Network to Assess Mountain Permafrost Degradation Rate. *23rd European Meeting of Environmental and Engineering Geophysics* (cit. on p. 36).
- Monnier, S. and Kinnard, C. (2015). Geomorphology, internal structure, and successive development of a glacier foreland in the semiarid Chilean Andes (Cerro Tapado, upper Elqui Valley). *Geomorphology* 250, pp. 461–463 (cit. on p. 19).
- Morand, S. (2000). *Inventaire des glaciers rocheux du Val d'Arolla*. Tech. rep. Rapport interne, Service des forêts et du paysage du Canton du Valais, Sion (cit. on pp. 33, 34, 242).
- Morard, S., Delaloye, R., and Lambiel, C. (2010). Pluriannual thermal behaviour of low elevation cold talus slopes in western Switzerland. *Geographica Helvetica* 65.2, pp. 124–134 (cit. on pp. 16, 17, 44, 223).
- Nandi, A. and Shakoor, A. (2010). A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology* 110.1, pp. 11–20 (cit. on p. 72).
- Neaupane, K. M. and Achet, S. H. (2004). Use of backpropagation neural network for landslide monitoring: a case study in the higher Himalaya. *Engineering Geology* 74.3, pp. 213–226 (cit. on p. 74).
- Nicodemus, K. K. and Malley, J. D. (2009). Predictor correlation impacts machine learning algorithms: implications for genomic studies. *Bioinformatics* 25.15, pp. 1884–1890 (cit. on p. 146).
- Noetzli, J. and Gruber, S. (2005). Alpiner Permafrost—ein Überblick. *Jahrbuch des Vereins zum Schutz der Bergwelt* 70, pp. 111–121 (cit. on p. 10).
- Noetzli, J., Gruber, S., Kohl, T., Salzmann, N., and Haeberli, W. (2007). Three-dimensional distribution and evolution of permafrost temperatures in idealized high-mountain topography. *Journal of Geophysical Research: Earth Surface (2003 - 2012)* 112.F2 (cit. on pp. 4, 10, 20, 21, 242).
- Nyenhuis, M., Hoelzle, M., and Dikau, R. (2005). Rock glacier mapping and permafrost distribution modelling in the Turtmanntal, Valais, Switzerland. *Zeitschrift für Geomorphologie, NF*, pp. 275–292 (cit. on pp. 23, 223).
- Oelke, C. and Zhang, T. (2004). A model study of circum-Arctic soil temperatures. *Permafrost and Periglacial Processes* 15.2, pp. 103–121 (cit. on p. 26).

- Otto, J.-C. and Sass, O. (2006). Comparing geophysical methods for talus slope investigations in the Turtmann valley (Swiss Alps). *Geomorphology* 76.3, pp. 257–272 (cit. on pp. 4, 241).
- Otto, J.-C., Keuschnig, M., Goetz, J., Marbach, M., and Schrott, L. (2012). Detection of mountain permafrost by combining high resolution surface and subsurface information—an example from the Glatzbach catchment, Austrian Alps. *Geografiska Annaler: Series A, Physical Geography* 94.1, pp. 43–57 (cit. on pp. 36, 45, 241, 243, 257).
- Otto, J.-C., Prasicek, G., Blöthe, J., and Schrott, L. (2017). GIS Applications in Geomorphology. Reference Module in Earth Systems and Environmental Sciences (cit. on p. 3).
- Ou, C., LaRocque, A., Leblon, B., Zhang, Y., Webster, K., and McLaughlin, J. (2016). Modelling and mapping permafrost at high spatial resolution using Landsat and Radarsat-2 images in Northern Ontario, Canada: Part 2—regional mapping. *International Journal of Remote Sensing* (cit. on p. 242).
- Pal, M. and Mather, P. M. (2001). Decision tree based classification of remotely sensed data. *Proceedings of 22nd Asian Conference on Remote Sensing*. Vol. 5, p. 9 (cit. on p. 80).
- Paul, F. (2004). The new Swiss glacier inventory 2000: application of remote sensing and GIS. *Geographisches Inst.* (cit. on p. 41).
- Paul, F., Maisch, M., Rothenbühler, C., Hoelzle, M., and Haeberli, W. (2007). Calculation and visualisation of future glacier extent in the Swiss Alps by means of hypsographic modelling. *Global and Planetary Change* 55.4, pp. 343–357 (cit. on p. 2).
- Peng, C.-Y. J. and So, T.-S. Harry. (2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences* 1.1, pp. 31–70 (cit. on p. 74).
- Peng, C.-Y. J., Lee, K. L., and Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The journal of educational research* 96.1, pp. 3–14 (cit. on p. 74).
- Pepin, N., Bradley, R. S., Diaz, H. F., Baraër, M., Caceres, E. B., Forsythe, N., Fowler, H., Greenwood, G., Hashmi, M. Z., Liu, X. D., et al. (2015). Elevation-dependent warming in mountain regions of the world. *Nature Climate Change* 5.5, pp. 424–430 (cit. on p. 1).
- Pereira, P., Brevik, E., and Trevisani, S. (2018). Mapping the environment. *Science of The Total Environment* 610-611, pp. 17–23 (cit. on p. 239).
- PERMOS (2009). Permafrost in Switzerland 2004/2005 and 2005/2006. Noetzli, J., Naegeli, B., and Vonder Muehll, D. (eds.) *Glaciological Report (Permafrost) No. 6/7 of the Cryospheric Commission of the Swiss Academy of Sciences* (cit. on pp. 225, 232).
- PERMOS (2016). Permafrost in Switzerland 2010/2011 to 2013/2014. Noetzli, J., Luethi, R., and Staub, B. (eds.) *Glaciological Report (Permafrost) No. 12-15 of the Cryospheric Commission of the Swiss Academy of Sciences* (cit. on pp. 1, 2, 10, 37, 153, 154, 159, 171, 187, 241, 243).
- Petrou, Maria and Sevilla, Pedro Garcia (2006). Image processing: dealing with texture. Vol. 1. Wiley Online Library (cit. on p. 45).
- Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers* 10.3, pp. 61–74 (cit. on pp. 79, 230, 254).

- Pogliotti, P. (2011). „Influence of snow cover on MAGST over complex morphologies in mountain permafrost regions“. PhD thesis. PhD Thesis, University of Torino, Torino, 79 pp (cit. on p. 47).
- Pollard, W. (2018). Periglacial Processes in Glacial Environments. *Past Glacial Environments*. Elsevier, pp. 537–564 (cit. on p. 18).
- Pozdnoukhov, A., Foresti, L., and Kanevski, M. (2009). Data-driven topo-climatic mapping with machine learning methods. *Natural hazards* 50.3, pp. 497–518 (cit. on p. 39).
- Pradhan, B. and Lee, S. (2010). Regional landslide susceptibility analysis using back-propagation neural network model at Cameron Highland, Malaysia. *Landslides* 7.1, pp. 13–30 (cit. on p. 74).
- Quinlan, J. R. (1986). Induction of decision trees. *Machine learning* 1.1, pp. 81–106 (cit. on p. 81).
- Raderschall, N., Lehning, M., and Schär, C. (2008). Fine-scale modeling of the boundary layer wind field over steep topography. *Water Resources Research* 44.9 (cit. on p. 188).
- Ravel, L. and Lambiel, C. (2013). Evolution récente de la moraine des Gentianes (2894 m, Valais, Suisse): un cas de réajustement paraglaciale? *Environnements périglaciaires* 18-19, 8p (cit. on pp. 33, 159).
- Ravel, L., Allignol, F., Deline, P., Gruber, S., and Ravello, M. (2010). Rock falls in the Mont Blanc Massif in 2007 and 2008. *Landslides* 7.4, pp. 493–501 (cit. on pp. 2, 241).
- Ravel, L., Lambiel, C., Jaboyedoff, M., and Oppikofer, T. (2012). Mesure de l'évolution de la moraine des Gentianes (Valais, Suisse) par balayage laser terrestre entre 2007 et 2011. *Journées de la Section Glaciologie-nivologie de la Société Hydrotechnique de France* (cit. on p. 33).
- Ravel, L., Deline, P., Lambiel, C., and Vincent, C. (2013). Instability of a high alpine rock ridge: The lower Arête des Cosmiques, Mont Blanc Massif, France. *Geografiska Annaler: Series A, Physical Geography* 95.1, pp. 51–66 (cit. on p. 20).
- Ravel, L., Magnin, F., and Deline, P. (2017). Impacts of the 2003 and 2015 summer heatwaves on permafrost-affected rock-walls in the Mont Blanc massif. *Science of The Total Environment* 609, pp. 132–143 (cit. on p. 3).
- Reynard, E. (1996). Glaciers rocheux et limite inférieure du pergélisol discontinu dans le vallon de Tortin (Nendaz-Valais). *Bulletin de la Murithienne* 114, pp. 135–150 (cit. on p. 223).
- Reynard, E. (2000). „Gestion patrimoniale et intégrée des ressources en eau dans les stations touristiques de montagne: Les cas de Crans-Montana-Aminona et Nendaz (Valais)“. PhD thesis. UNIL Institut de géographie (cit. on p. 33).
- Reynard, E., Delaloye, R., and Lambiel, C. (1999). Prospection géoélectrique du pergélisol alpin dans le massif des Diablerets (VD) et au Mont Gelé (Nendaz, VS). *Bulletin de la Murithienne* 117, pp. 89–104 (cit. on pp. 33, 223).
- Reynard, E., Lambiel, C., Delaloye, R., Devaud, G., Baron, L., Chapellier, D., Marescot, L., and Monnet, R. (2003). Glacier/permafrost relationships in forefields of small glaciers (Swiss Alps). *Proceedings 8th International Conference on Permafrost, Zurich, Switzerland*. Vol. 2, pp. 947–952 (cit. on pp. 4, 18, 37, 159, 187, 224, 228, 241, 243, 257).

- Ribolini, A., Chelli, A., Guglielmin, M., and Pappalardo, M. (2007). Relationships between glacier and rock glacier in the Maritime Alps, Schiantala Valley, Italy. *Quaternary Research* 68.3, pp. 353–363 (cit. on p. 18).
- Ribolini, A., Guglielmin, M., Fabre, D., Bodin, X., Marchisio, M., Sartini, S., Spagnolo, M., and Schoeneich, P. (2010). The internal structure of rock glaciers and recently deglaciated slopes as revealed by geoelectrical tomography: insights on permafrost and recent glacial evolution in the Central and Western Alps (Italy–France). *Quaternary Science Reviews* 29.3, pp. 507–521 (cit. on pp. 18, 19, 187, 241).
- Riedlinger, T. and Kneisel, C. (2000). Interaktionen von Permafrost und Ausaperung im Gletschervorfeld des Vadret da Rosatsch. *Oberengadin, Schweiz., Trierer Geographische Studien* 23, pp. 147–164 (cit. on p. 23).
- Ripley, B. D. (1996). Pattern recognition via neural networks. *a volume of Oxford Graduate Lectures on Neural Networks, title to be decided*. Oxford University Press. [See <http://www.stats.ox.ac.uk/ripley/papers.html>.] (Cit. on p. 58).
- Robert, S., Foresti, L., and Kanevski, M. (2013). Spatial prediction of monthly wind speeds in complex terrain with adaptive general regression neural networks. *International Journal of Climatology* 33.7, pp. 1793–1804 (cit. on p. 55).
- Rödder, T. and Kneisel, C. (2012). Influence of snow cover and grain size on the ground thermal regime in the discontinuous permafrost zone, Swiss Alps. *Geomorphology* 175, pp. 176–189 (cit. on pp. 44, 257).
- Roer, I., Haeberli, W., Avian, M., Kaufmann, V., Delaloye, R., Lambiel, C., and Käab, A. (2008). Observations and considerations on destabilizing active rock glaciers in the European Alps. *Ninth International Conference on Permafrost*. Vol. 2, pp. 1505–1510 (cit. on pp. 2, 34, 241).
- Roobaert, D., Karakoulas, G., and Chawla, N. V. (2006). Information gain, correlation and support vector machines. *Feature Extraction*. Springer, pp. 463–470 (cit. on p. 60).
- Rosset, E., Hilbich, C., Schneider, S., and Hauck, C. (2013). Automatic filtering of ERT monitoring data in mountain permafrost. *Near Surface Geophysics* 11.4, pp. 423–433 (cit. on p. 36).
- Roweis, S. T. and Saul, L. K. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science* 290.5500, pp. 2323–2326. eprint: <http://science.sciencemag.org/content/290/5500/2323.full.pdf> (cit. on p. 59).
- Ruszkiczay-Rüdiger, Z., Fodor, L., Horváth, E., and Telbisz, T. (2009). Discrimination of fluvial, eolian and neotectonic features in a low hilly landscape: A DEM-based morphotectonic analysis in the Central Pannonian Basin, Hungary. *Geomorphology* 104.3, pp. 203–217 (cit. on p. 45).
- Saeys, Y., Inza, I., and Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics* 23.19, pp. 2507–2517 (cit. on p. 60).
- Salzmann, N. and Gärtner-Roer, I. (2017). Climate Change and Permafrost. *The International Encyclopedia of Geography* (cit. on p. 1).
- Sammon, J. W. (1969). A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers* 100.5, pp. 401–409 (cit. on pp. 58, 59).

- Sattler, K. (2016). Periglacial preconditioning of debris flows in the Southern Alps, New Zealand. Springer (cit. on p. 3).
- Sattler, K., Anderson, B., Mackintosh, A., Norton, K., and Róiste, M. de (2016). Estimating Permafrost Distribution in the Maritime Southern Alps, New Zealand, Based on Climatic Conditions at Rock Glacier Sites. *Frontiers in Earth Science* 4, p. 4 (cit. on pp. 4, 10, 29, 35, 72, 146, 177, 182, 185, 188, 241, 257).
- Sawada, Y., Ishikawa, M., and Ono, Y. (2003). Thermal regime of sporadic permafrost in a block slope on Mt. Nishi-Nupukaushinupuri, Hokkaido Island, Northern Japan. *Geomorphology* 52.1, pp. 121–130 (cit. on pp. 16, 44).
- Scapozza, C. (2013). Stratigraphie, morphodynamique, paléoenvironnements des terrains sédimentaires meubles à forte déclivité du domaine périglaciaire alpin. Institut de géographie- Université de Lausanne (cit. on pp. 18, 33, 34, 37, 155–157, 165–167, 174, 224, 235, 243).
- Scapozza, C. (2015a). Evolution des glaciers et du pergélisol depuis le dernier maximum glaciaire dans la région du mont Gelé-Mont Fort (Alpes Valaisannes, Suisse): chronologie, modalités de la dernière déglaciation et datations des âges d'exposition à l'aide du marteau de Schmidt. *Quaternaire. Revue de l'Association française pour l'étude du Quaternaire* 26.2, pp. 141–173 (cit. on pp. 3, 33, 155, 166).
- Scapozza, C. (2015b). Investigation on protalus ramps in the Swiss Alps. *Geographica Helvetica* 70.2, pp. 135–139 (cit. on pp. 154, 155, 165).
- Scapozza, C. and Lambiel, C. (2013). Structure interne et répartition du pergélisol dans l'éboulis «à galets» de Tsaté-Moiry (VS). *Mattertal - ein Tal in Bewegung. Publikation zur Jahrestagung des Schweizerischen Geomorphologischen Gesellschaft (SGmG), St. Niklaus (Schweiz)* 29, pp. 33–45 (cit. on pp. 34, 174, 223).
- Scapozza, C., Lambiel, C., Baron, L., Marescot, L., and Reynard, E. (2011). Internal structure and permafrost distribution in two alpine periglacial talus slopes, Valais, Swiss Alps. *Geomorphology* 132.3, pp. 208–221 (cit. on pp. 3, 4, 12, 15, 17, 18, 33, 37, 38, 154, 223, 224, 241, 243, 257).
- Scapozza, C., Baron, L., and Lambiel, C. (2015). Borehole logging in Alpine periglacial talus slopes (Valais, Swiss Alps). *Permafrost and Periglacial Processes* 26.1, pp. 67–83 (cit. on p. 154).
- Schaepli, B., Hingray, B., Niggli, M., and Musy, A. (2005). A conceptual glacio-hydrological model for high mountainous catchments. *Hydrology and Earth System Sciences Discussions* 9.1/2, pp. 95–109 (cit. on p. 188).
- Schaeftlein, H. (1962). Ein eigenartiges Hochmoor in den Schladminger Tauern (cit. on p. 16).
- Scherrer, S. C., Appenzeller, C., Liniger, M. A., and Schär, C. (2005). European temperature distribution changes in observations and climate change scenarios. *Geophysical Research Letters* 32.19 (cit. on p. 1).
- Schmid, S., Guisan, A., and Di Cola, V. (2014). „Assessing spatial autocorrelation in biodiversity data in the Western Swiss Alps“. MA thesis. Switzerland: Lausanne University (cit. on p. 67).



- Schoeneich, P., Bodin, X., Echelard, T., Kaufmann, V., Kellerer-Pirklbauer, A., Krysiacki, J.-M., and Lieb, G. K. (2015). Velocity changes of rock glaciers and induced hazards. *Engineering Geology for Society and Territory-Volume 1*. Springer, pp. 223–227 (cit. on p. 14).
- Schölkopf, B. and Smola, A. J. (2002). Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press (cit. on p. 57).
- Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation* 10.5, pp. 1299–1319 (cit. on p. 57).
- Schöner, W., Boeckli, L., Hausmann, H., Otto, J.-C., Reisenhofer, S., Riedl, C., and Seren, S. (2012). Spatial patterns of permafrost at Hoher Sonnblick (Austrian Alps)-extensive field-measurements and modelling approaches. *Austrian Journal of Earth Sciences* 105.2, pp. 154–168 (cit. on pp. 4, 241).
- Schrott, L., Otto, J.-C., and Keller, F. (2012). Modelling alpine permafrost distribution in the Hohe Tauern region, Austria. *Austrian Journal of Earth Sciences* 105.2, pp. 169–183 (cit. on pp. 176, 185, 256).
- Scotti, R., Brardinoni, F., Alberti, S., Frattini, P., and Crosta, G. (2013). A regional inventory of rock glaciers and protalus ramparts in the central Italian Alps. *Geomorphology* 186, pp. 136–149 (cit. on pp. 15, 18).
- Seppi, R., Carton, A., and Baroni, C. (2005). Proposta di una nuova scheda per il censimento dei rock glaciers da fotografie aeree: applicazione sull'Alta Val Ultimo (Gruppo Ortles-Cevedale). *Geografia Fisica e Dinamica Quaternaria* 7, pp. 329–338 (cit. on pp. 23, 176, 223).
- Specht, D. (1991). A general regression neural network. *IEEE Transaction on Neural Networks* 2.6, pp. 568–576. (Cit. on p. 55).
- Specht, D. (1994). Experience with adaptive probabilistic neural networks and adaptive general regression neural networks. *IEEE World Congress on Computational Intelligence* (cit. on p. 55).
- Staub, B., Marmy, A., Hauck, C., Hilbich, C., and Delaloye, R. (2015). Ground temperature variations in a talus slope influenced by permafrost: a comparison of field observations and model simulations. *Geographica Helvetica* 70.1, pp. 45–62 (cit. on pp. 36, 37, 156, 157, 243).
- Stocker-Mittaz, C., Hoelzle, M., and Haeberli, W. (2002). Modelling alpine permafrost distribution based on energy-balance data: a first step. *Permafrost and Periglacial Processes* 13.4, pp. 271–282 (cit. on pp. 23, 24, 223).
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science* 240.4857, pp. 1285–1293 (cit. on pp. 86, 252).
- Toloşi, L. and Lengauer, T. (2011). Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics* 27.14, pp. 1986–1994 (cit. on p. 146).
- Trigila, A., Iadanza, C., Esposito, C., and Scarascia-Mugnozza, G. (2015). Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampileri (NE Sicily, Italy). *Geomorphology* 249, pp. 119–136 (cit. on pp. 72, 249).
- Tucker, C. J. and Sellers, P. J. (1986). Satellite remote sensing of primary production. *International journal of remote sensing* 7.11, pp. 1395–1416 (cit. on p. 43).

- Van Der Maaten, L., Postma, E., and Herik, J. Van den (2009). Dimensionality reduction: a comparative. *J Mach Learn Res* 10, pp. 66–71 (cit. on p. 94).
- Vapnik, V. (1998). *Statistical learning theory*. Wiley, New York (cit. on pp. 51, 57, 62, 64–66, 76–78, 224, 225, 230, 249, 250).
- Varley, A., Tyler, A., Smith, L., Dale, P., and Davies, M. (2016). Mapping the spatial distribution and activity of 226 Ra at legacy sites through Machine Learning interpretation of gamma-ray spectrometry data. *Science of The Total Environment* 545, pp. 654–661 (cit. on p. 242).
- Venables, W. N. and Ripley, B. D. (2002). Random and mixed effects. *Modern applied statistics with S*. Springer, pp. 271–300 (cit. on pp. 58, 81).
- Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). *SOM toolbox for Matlab 5*. Tech. rep. (cit. on p. 104).
- Vivero, S., Deluigi, N., and Lambiel, C. (2017). Monitoring high rates of permafrost creep in the Western Swiss Alps. *Swiss Polar Institute* (cit. on p. 33).
- Wakonigg, H. (1996). Unterkühlte schutthalden. na (cit. on p. 16).
- Waller, R. I. and Tuckwell, G. W. (2005). Glacier-permafrost interactions and glaciotectonic landform generation at the margin of the Leverett Glacier, West Greenland. *Geological Society, London, Special Publications* 242.1, pp. 39–50 (cit. on p. 18).
- Waller, R. I., Murton, J. B., and Kristensen, L. (2012). Glacier–permafrost interactions: Processes, products and glaciological implications. *Sedimentary Geology* 255-256, pp. 1–28 (cit. on p. 18).
- Washburn, A. L. (1979). *Geocryology: a survey of periglacial processes and environments*. Ed. by London: Edward Arnold (cit. on p. 9).
- Weber, S., Beutel, J., Faillettaz, J., Hasler, A., Krautblatter, M., and Vieli, A. (2017). Quantifying irreversible movement in steep, fractured bedrock permafrost on Matterhorn (CH). *The Cryosphere* 11.1, p. 567 (cit. on p. 16).
- Wehrens, R., Buydens, L., et al. (2007). Self-and super-organizing maps in R: the Kohonen package. *J Stat Softw* 21.5, pp. 1–19 (cit. on p. 71).
- Wentworth, C. K. (1922). A scale of grade and class terms for clastic sediments. *The Journal of Geology* 30.5, pp. 377–392 (cit. on p. 44).
- WGMS (2005). *Fluctuations of glaciers, 1995-2000*. VIII, Paris (cit. on p. 1).
- Wicky, J. and Hauck, C. (2017). Numerical modelling of convective heat transport by air flow in permafrost talus slopes. *The Cryosphere* 11.3, p. 1311 (cit. on pp. 5, 24).
- Wilson, N. K., Kent, D. G., Buettner, F., Shehata, M., Macaulay, I. C., Calero-Nieto, F. J., Castillo, M. S., Oedekoven, C. A., Diamanti, E., Schulte, R., et al. (2015). Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell stem cell* 16.6, pp. 712–724 (cit. on p. 54).
- Witten, I. H. and Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann (cit. on pp. 118, 145, 147, 255).
- Yang, Y. and Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. *Icml*. Vol. 97, pp. 412–420 (cit. on p. 60).

- Zanoner, T., Carton, A., Seppi, R., Carturan, L., Baroni, C., Salvatore, M. C., and Zumiani, M. (2017). Little Ice Age mapping as a tool for identifying hazard in the paraglacial environment: The case study of Trentino (Eastern Italian Alps). *Geomorphology* 295, pp. 551–562 (cit. on p. 2).
- Zemp, M., Paul, F., Hoelze, M., and Haeberli, W. (2008). Glacier fluctuations in the European Alps, 1850–2000. *Darkening Peaks Glacier Retreat Sci. Soc* (cit. on p. 1).
- Zemp, M., Frey, H., Gärtner-Roer, I., Nussbaumer, S. U., Hoelzle, M., Paul, F., Haeberli, W., Denzinger, F., Ahlstrøm, A. P., Anderson, B., et al. (2015). Historically unprecedented global glacier decline in the early 21st century. *Journal of Glaciology* 61.228, pp. 745–762 (cit. on p. 1).
- Zhang, X., Nan, Z., Wu, J., Du, E., Wang, T., and You, Y. (2012). Mountain permafrost distribution modeling using Multivariate Adaptive Regression Spline (MARS) in the Wenquan area over the Qinghai-Tibet Plateau. *Sciences in Cold and Arid Regions* 4.5, pp. 361–370 (cit. on p. 29).

