Research papers

Efficient multi-objective calibration and uncertainty analysis of distributed snow simulations in rugged alpine terrain

J.M. Thornton [a,*], T. Brauchli [b,c], G. Mariethoz [b], P. Brunner [a]

[a] Centre for Hydrogeology and Geothermics, University of Neuchâtel, 2000 Neuchâtel, Switzerland
[b] Institute of Earth Surface Dynamics, University of Lausanne, 1015 Lausanne, Switzerland
[c] Centre de Recherche sur l'Environnement Alpin (CREALP), 1950 Sion, Switzerland

ARTICLE INFO

ABSTRACT

In mountainous terrain, reliable snow simulations are crucial for many applications. However, except in highly instrumented research catchments, meteorological data are usually limited, and so the interpolated spatial fields used to force snow models are uncertain. Moreover, certain potentially important processes cannot presently be simulated at catchment scales using entirely physical algorithms. It is therefore often appropriate to introduce empirical parameters into otherwise physically-based snow models. Many opportunities to incorporate snow observations into the parameter estimation process now exist, but they remain to be fully exploited. In this context, a novel approach to the calibration of an energy balance-based snow model that additionally accounts for gravitational redistribution is presented. Several important parameters were estimated using an efficient, gradient-based method with respect to two complementary observation types – Landsat 8-derived snow extent maps, and reconstructed snow water equivalent (SWE) time-series. When assessed on a per-pixel basis, observed patterns were ultimately reproduced with a mean accuracy of 85%. Spatial performance metrics compared favourably with those previously reported, whilst the temporal evolution of SWE at the stations was also satisfactorily captured. Subsequent uncertainty and data worth analyses revealed that: i) the propensity for model predictions to be erroneous was substantially reduced by calibration, ii) pre-calibration uncertainty was largely associated with two parameters which modify the longwave component of the energy balance, but this uncertainty was greatly diminished by calibration, and iii) a lower elevation SWE series was particularly valuable, despite containing comparatively few observations. Overall, our work demonstrates that contemporary snow models, observation technologies, and inverse approaches can be combined to both constrain and quantify the uncertainty associated with simulations of alpine snow dynamics.

1. Introduction

1.1. The significance of mountainous water resources

Meltwater derived from seasonal snowpacks currently dominates annual groundwater recharge and cumulative streamflow of many mid-elevation temperate mountainous catchments. At higher elevations, the progressive ablation of firn and glacier ice throughout summer periods represent major additional inputs of liquid water to the terrestrial hydrosphere. Globally, these snow and ice-derived meltwaters directly sustain millions of people (Pritchard, 2019) and constitute an ecosystem service of enormous value (Sturm et al., 2017). However, hydrological regimes which have historically been heavily influenced by snow and ice are likely to be greatly affected by ongoing warming (Barnett et al., 2005; Viviroli et al., 2011), with summer low flow magnitudes particularly vulnerable (Jenicek et al., 2016; Dierauer et al., 2018).

Indeed, a wealth of evidence attesting the widespread decline of glaciers and other hydrologically-relevant components of the mountain cryosphere now exists (Klein et al., 2016; Huss et al., 2017; Beniston et al., 2018; Bolch et al., 2012; Bormann et al., 2018; Vuille et al., 2018), and the resultant impacts on stream discharges are increasingly detectable (Casassa et al., 2009; Micheletti & Lane, 2016; Lane & Nienow, 2019). Predictions of the future quantity and timing of mountain runoff accordingly remain in high demand, and the substantial body of literature in which hydrological models are applied to generate such predictions continues to grow (e.g. Fatichi et al., 2015; Huss & Hock,

2018).

Rain-on-snow events, convective thunderstorms, and more sustained episodes of frontal rainfall also have high (and potentially increasing) importance Alpine terrain, including with respect to floods, debris flows, and landslide hazard (Papathoma-Köhle et al., 2011; Rössler et al., 2014; Leonarduzzi et al., 2017). As such, spatio-temporal patterns of liquid precipitation must also be given due consideration. The same applies to evapotransporative losses (Herrnegger et al., 2012; Mutzner et al., 2015; Cochand et al., 2019).

1.2. Progress in incorporating spatial snow information

Although many widely used box-type hydrological models can often consistently reproduce (even independent) discharge observations following calibration against measurements of this variable alone, observed internal spatial dynamics – including those pertaining to the snowpack – may remain poorly captured (Duethmann et al., 2014; Shrestha et al., 2014). This can be attributed to the considerable "freedom" that traditional calibration approaches afford, as well as the fact that discharge measurements provide only indirect, integrated information on internal system functioning.

Assessing simulated patterns of model state variables against spatially distributed observations provides a more stringent test of model capabilities, and so represents a means by which the internal consistency of hydrological models can be enhanced. Indeed, ensuring that the spatio-temporal dynamics of all potentially relevant hydrological processes can be acceptably reproduced (i.e. that the "right answers" in terms of discharge are obtained for the "right reasons"; Kirchner, 2006) is likely to lead to more reliable future predictions.

Much progress has already been made in incorporating snow information into hydrological models. For example, Finger et al. (2011) used snow cover images alongside glacier mass balance and discharge measurements in a snow-and ice dominated catchment in Switzerland to identify the best performing distributed model parameters from a large, randomly generated set. Duethmann et al. (2014) also employed a Monte Carlo approach, here in an attempt to quantify the relationship between the information content and the number of snow cover images included in the calibration of a model covering several mountainous catchments in central Asia. In comparing two alternative strategies for simulating hydrological processes the high-elevation Andes, Ragettli et al. (2014) likewise considered snowcover data, although purely for evaluative purposes. More recently, Costa et al. (2018) calibrated a simple snow model using distributed snow observations to investigate the mechanisms responsible for increases in suspended sediment concentrations that were observed in the upper Rhône in the 1980s.

Developments in snow remote sensing and modelling have also been made recently in the western United States thorough the Airbourne Snow Observatory (ASO) and NASA's SnowEx campaign (Behrangi et al., 2018; McGrath et al., 2019; Hedrick et al., 2018). However, these efforts are geographically and temporally limited, and generally rely on observation technologies that are unavailable in other global mountain regions. Finally, some studies have employed distributed snow images not as calibration or evaluation criteria, but as model inputs (Berezowski et al., 2015; Wulf et al., 2016).

1.3. Outstanding snow modelling challenges

Despite these advancements, given the considerable spatio-temporal variability and complexity that many of the factors and processes influencing snow dynamics in rugged terrain exhibit (Clark et al., 2011), many of the simulation approaches that currently prevail in the hydrological literature may be somewhat limited in their ability to represent snow dynamics reliably – especially in moderately large, topographically complex, and data limited headwater regions.

For instance, despite the now widespread availability of relevant spatial data, spatially lumped (e.g. Wagner et al., 2017) or only partially

distributed (Duethmann et al., 2014; Staudinger et al., 2017) hydrological models remain common. Although such lumped models have their uses, they cannot represent heterogeneity below the aggregation unit, and so provide little information on spatial snow patterns. This is unfortunate because spatial patterns of snowcover are imperative for winter tourism (Grünewald et al., 2010) and predicting vegetation species distributions (Randin et al., 2015), amongst other applications. More specifically, integrating distributed observations with lumped models is somewhat complicated; one must resort to comparing spatially-averaged snow covered areas (SCAs) (Ragettli et al., 2014), or else somehow re-impose spatial variability in the simulations (Parajka & Blöschl, 2008).

Irrespective of their spatial discretisation, most hydrological modelling studies that have incorporated distributed snow observations have relied on products from the Moderate Resolution Imaging Spectroradiometer (MODIS) (Clark et al., 2006; Duethmann et al., 2014; Ragettli et al., 2014; Engel et al., 2017; Costa et al., 2018). However, the 500 m pixel resolution at which binary (snow or no-snow) and/or snow covered fraction (f_{SCA}) data are provided (Rittger et al., 2013) is simply too coarse for certain applications. For instance, both Ragettli et al. (2014) and Hanzer et al. (2016) report difficulties capturing the complex snow patterns that are commonly observed in rugged terrain, such as small patches and snow-free ridges, using MODIS imagery. Such relatively fine-scale processes can substantially influence the internal hydrological functioning of steep mountain catchments. Much higher resolution (30 m) and long-term global snow maps can be derived from Landsat imagery, but have been mostly applied for model corroboration/evaluation rather than calibration (but see Schattan et al. 2020). Additionally, with the notable exceptions of Hanzer et al. (2016) and Wayand et al. (2018), previous studies involved only a handful of images (Bernhardt et al., 2012; Warscher et al., 2013; Schöber et al., 2010).

Empirical temperature and other index-based methods for estimating snow (and ice) melt rates (Hock, 2003) also remain standard (Ragettli & Pellicciotti, 2012; Addor et al., 2014; Etter et al., 2017), despite their abilty to satisfactorily reproduce snow dynamics in complex alpine terrain being questionable (Warscher et al., 2013). Provided additional meteorological data are available, more sophisticated, distributed energy balance approaches (both full physics, multiple snow-layer configurations as well as simplified alternatives) have been recommended (Magnusson et al., 2015; Meeks et al., 2017). One attraction of such models in steep, complex terrain is that they explicitly represent most of the fluxes influencing melt, including the (often pronounced) spatio-temporal variability thereof (e.g. the effects of slope aspect and topographic shading effects on incoming radiation). Another advantage is that they can simulate melt during critical events (e.g. rain-on-snow events), which are mainly driven by turbulent fluxes, better than their simpler counterparts (Würzer et al., 2017). Finally, energy balance models are more likely to perform reliably under forcing conditions that exceed the range of historical observations, as is typical in climate change impact assessments (Mas et al., 2018).

The most sophisticated energy balance models (e.g. Alpine3D; Lehning et al. 2006) include full-physics, multi-layered snowpack representations and therefore theoretically provide the most comprehensive representation of the complex mass and energy exchange processes that affect mountain snowpacks. However, for the purposes of hydrological predictions, they are coupled with highly simplified conceptual or "bucket-type" representations of subsurface processes and flow routing (Gallice et al., 2016). In geologically complex settings, which many Alpine regions inherently are, such simplifications may be unsuitable. Moreover, the simulation of wind and gravitational snow redistribution processes at catchment or larger scales using physical algorithms remains computationally prohibitive (Mott & Lehning, 2010; Musselman et al., 2015; Brauchli et al., 2017). Yet in very steep terrain in particular, accounting for gravitational snow redistribution is paramount to produce hydrologically realistic simulations of the evolution of snow water equivalent (SWE), and thus patterns of meltwater

generation (Bernhardt et al., 2012; Kerr et al., 2013; Sommer et al., 2015); *in extremis*, failure to do so can lead to "snow towers", which are undesirable model artifacts (Freudiger et al. 2017). Gravitational redistribution is also often critical to glacier accumulation (Mott et al., 2019). Various pragmatic empirical correction methods and algorithms have been developed enable such processes to still be represented (e.g. Bernhardt et al., 2012; Vögeli et al., 2016; Marshall et al., 2019).

A persistent fundamental challenge associated with modelling mountain hydrological systems is that the meteorological inputs are often very poorly constrained. Due to wind-induced gauge undercatch, precipitation measurements at stations are typically underestimated (Pan et al., 2016; Kochendorfer et al., 2017). This bias is most pronounced when the precipitation phase is solid, and at higher wind speeds. In addition, even in comparatively well-instrumented regions like the European Alps, meteorological station density decreases substantially with elevation (Pepin et al., 2015). Coupled with the high spatial variability that characterises mountain meteorology (Mott et al., 2014), this means that even if the original ground measurements could be made perfectly, subsequent interpolated spatial fields would still be highly uncertainty. As such, irrespective of its complexity, snow model performance will always be intimately related to forcing data quality.

1.4. Inverse approaches and uncertainty quantification

Assessing the error characteristics of common instruments (see e.g. the WMO-SPICE project; Kochendorfer et al., 2017) and systematically testing various spatial interpolation methods (Tobin et al., 2011) have both been pursued to address the aforementioned deficiencies in forcing datasets. "Inverse" methods, whereby distributed models and snow observations are combined to estimate the values of important but uncertain "correction" parameters, are also beginning to be applied to this end.

For instance, in a snow model analysis involving both MODIS and Landsat-derived snow observations for evaluation, Engel et al. (2017) found modifying a "snow correction factor" to be necessary to compensate for biased winter precipitation measurement. Shrestha et al. (2014) actually calibrated a distributed, multi-layer water and energy balance model (WEB-DHM-S) in order to minimise the cumulative error in snow cover pattern (again according to MODIS) and discharge simulations. In so doing, an elevation-dependent snowfall correction factor was optimised. A particular novelty of this study was that correspondence between simulated and observed patterns was expressed at the pixel level. Naseer et al. (2019) applied the same code but avoided traditional linear, elevation-dependent lapse rates for meteorological data interpolation (which may break down in complex terrain) by integrating 3D temperature profiles derived from climate model reanalysis data. The calibration undertaken had no spatial component, however. Most recently, Ruelland (2020) sought to infer uncertain mountain precipitation and temperature gradients in the French Alps via inversion using a very simple snow and hydrological model alongside discharge and MODIS snow cover data.

Lastly, uncertainty quantification should ideally form a central pillar of any environmental modelling exercise. Although some previous studies have directly assessed uncertainty in SWE reconstructions (Franz et al., 2010; He et al., 2011; Slater et al., 2013; Meeks et al., 2017), this has largely been undertaken only at discrete station locations (i.e. using non-distributed models). As one seeks to progress beyond this situation, the efficiency of calibration and uncertainty quantification algorithms becomes a crucial consideration, especially as the sophistication, scale, and resolution a given "forward" model increases; despite ever-increasing computational power, "brute force" approaches involving thousands of Monte Carlo simulations can still quickly become impractical.

1.5. The present study

In this context, and with the intention of improving the representation of meltwater dynamics in relatively data-scarce and rugged mountain settings, this study proposes a model-independent framework for integrating high-resolution snow observations in distributed snowpack simulations. It is not our intention to focus here on the development of combinstion of improved physically-based algorithms for representing the various complex individual processes and phenomena that can influence snow variability in alpine terrain, but rather to present a novel means by which complementary snow data can be used to constrain a series of generally important but highly uncertain parameters within any distributed snow model. The moderate-complexity snow code selected for our exemplification of the framework aligns well with our expectation that, in this specific setting as well as likely many others (especially at larger spatial scales), any uncertainties associated with internal snow model structures will often be overshadowed by those from other sources (e.g. uncertainties in forcing data fields). In our approach, not only are the major uncertainties involved in a typical simulation chain explicitly acknowledged, but they are overcome as far as is possible

More specifically, a fully-distributed energy balance-based snow model that additionally represents gravitational redistribution is initially established at high spatio-temporal resolution (25 m, hourly). Then, an objective function incorporating both high-resolution snow cover maps derived from satellite imagery and reconstructed SWE time-series at two locations is developed and minimised using an efficient, iterative calibration algorithm.

A single layer snowpack configuration is applied in WaSiM (v10.04.01; Schulla, 2017). This code provides an appropriate balance between snow model simplicity and complexity, and also enabled several additional steps to be incorporated relatively easily, namely: i) the correction and spatial interpolation of meteorological station data, ii) the representation of gravitational snow redistribution (which is important given the steepness of the study catchment in questions), iii) the representation of glacier mass balance/melt, and iv) the generation of commensurate potential evapotranspiration (ET_p) estimates for subsequent hydrological modelling.

This is one of the first instances in which Landsat-derived snow cover images are included in the calibration of a distributed snow model in a spatially-explicit (i.e. per-pixel) fashion. Other distinguishing features of this study are that spatial fit metrics are computed for a much larger catalogue of images than previously, and that the uncertainties associated with selected key predictions, as well as the contributions of different parameters and groups of observations to uncertainty reduction, are elucidated. By including high-resolution distributed snow observations, SWE time-series, and a model that accounts for gravitational redistribution, this approach builds somewhat upon that of Shrestha et al. (2014), providing a framework that is more suited to steep and rugged terrain. Finally, given the open source software used throughout and the long temporal coverage and global availability of the Landsat archive, our approach has great potential to improve the representation of snow dynamics in many mountain regions globally. The input files and model-data linking code are provided with the intention of helping to facilitate this. Furthermore, because the methodology is fundamentally code and data agnostic, alternative distributed codes and snow-cover products can easily be substituted.

The model is not extended here to generate streamflow outputs. This decision was taken because the study catchment under consideration is extremely geologically complex, replete with folded, faulted sequences of limestones, shales, and marls, whilst the representation of subsurface processes in WaSiM – as in most popular hydrological models – is rather simplistic. Instead, the resultant high-resolution, spatio-temporal gridded datasets representing the arrival of snowmelt, firn melt, ice melt, and liquid precipitation at the land surface (collectively, "all liquid water") and ET_p have been applied as boundary conditions for a fully

integrated (i.e. surface–subsurface-evapotranspiration) model, presented separately (Thornton et al., under review). Such integrated models can ingest 3D representations of geology and simulate the mechanistic interactions between various components of hydrological systems in a physically-based, spatially explicit fashion. As such, they offer considerable potential to better understand and predict the nuances of such systems, including their possible responses to external change, in a more comprehensive and robust fashion than hitherto possible. However, most integrated models also rely on simplified empirical snow processes representations. Hence, our broader effort seeks to leverage the respective benefits of alternative simulation approaches whilst mitigating their respective limitations.

2. Study area

The 36.7 km² study area includes two adjacent headwater catchments – the Vallon de Nant and the Vallon de La Vare – of the western Swiss Alps (Fig. 1). The elevational range is considerable, extending from 950 to over 3050 m a.s.l., and slopes are accordingly extremely steep (mean 35°). The topography is rugged, and lower parts of the study area are forested. At the Last Glacial Maximum, only the highest peaks protruded above the ice (Bini et al., 2009). An array of Quaternary unconsolidated sedimentary features with glacial, fluvial, and mass movement origins overly the complex Mesozoic bedrock, several of

which likely act as aquifers.

Approximately 45% of annual precipitation (≥1400 mm) falls as snow. Due to the catchment's steepness, gravitational snow redistribution occurs frequently, as evidenced by snow-free slopes and cliffs in the winter months. The photographs of Fig. S1, taken in the Vallon de Nant on the 31 January 2018 following a period of exceptional snowfall (Bründl et al., 2019), illustrate the considerable redistribution that can occur under more extreme conditions.

Intense summer thunderstorms are a further noteworthy feature of the area's meteorological regime. The surficial hydrology of the Vallon de Nant is characterised by numerous temporary torrents whose discharge responds rapidly to rainfall and snowmelt. Being shaded by surrounding cliffs, several small glaciers persist at relatively low elevations in the north-facing upper reaches of both sub-catchments. The region remains in a highly natural state, making it rather unusual in the context of the European Alps. Reflecting this, several recent environmental investigations have focused upon the area, including those of Vittoz et al. (2009), Grand et al. (2016), Lane et al. (2016), Benoit et al. (2018), and Giaccone et al. (2019).
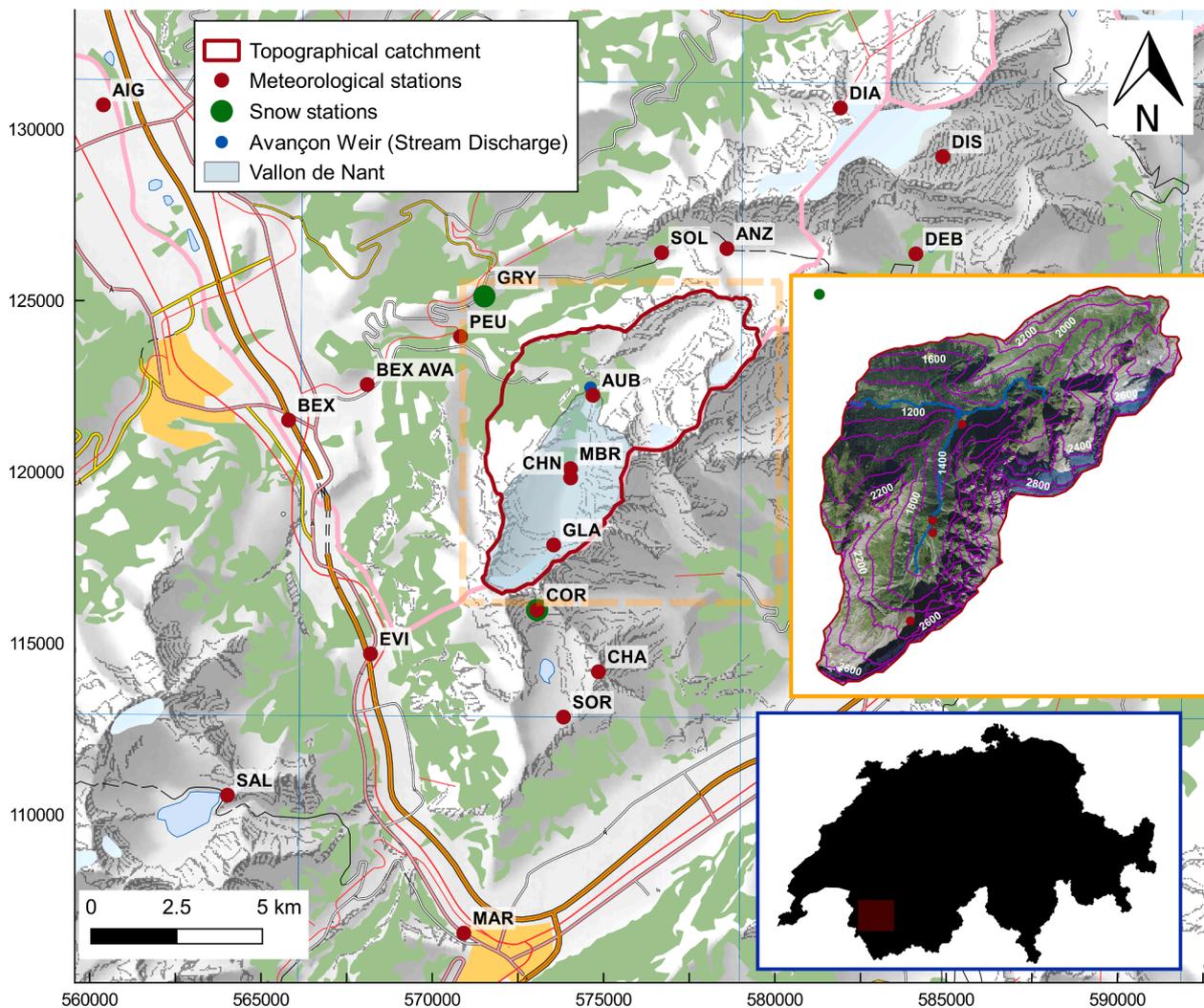
Fig. 1. The study catchment and locations of stations that provided data to the present study. The precipitation data at the SOR station were ultimately removed from the inputs because the cumulative totals measured were considered unrealistically low compared with other nearby stations at similar elevations, indicating a probable station issue. Background data © swisstopo.

3. Data availability and processing

3.1. Meteorological forcing

The model required gridded estimates of incoming shortwave radiation, precipitation, relative humidity, sunshine duration, air temperature, vapour pressure, and wind speed. No third-party meteorological datasets with the desired spatio-temporal resolution and which were contemporaneous with the available field measurements, the earliest of which began in 2016, existed at the outset. Forcing datasets were therefore developed from meteorological station measurements (Fig. 1; Table S1). Most of the stations used belonged to the official networks of MeteoSwiss and the WSL Institute for Snow and Avalanche Research (SLF), and were located several kilometres from the study catchment and, crucially, at lower elevations. At these stations, hourly sums for precipitation and hourly means for all other variables were downloaded from IDAWEB – the online data portal of MeteoSwiss (MeteoSwiss, 2019) – for the hydrological years 2015–2018 (i.e. 1 October 2014 to 30 September 2018).

These data were complemented by observations from stations belonging to a local network operated by the University of Lausanne (S. Hiscox, *pers. comm*; Michelon et al. (2017) and subsequent updates via personal communication), some of which are located within the study catchment itself. Unsurprisingly given the harsh environment and limited access (especially in winter), the local stations had a higher proportion of missing data than the nationally operated ones, meaning intensive processing and quality assurance efforts were required.

Irrespective of the station operator, precipitation data from unheated gauges were not considered. Also, the precipitation data at SOR were eventually removed from the input dataset as the cululative totals were deemed unrepresentative. Specifically, they were unduly low compared with nearby stations at similar elevations (Brauchli et al., 2018). The processed time-series were plotted and inspected interactively using niVis (SLF, 2019a); the hourly time-series themselves are presented in Fig. S14, whilst the temporal coverage of variables between stations and overall missing data percentages are shown in Figures S3 to S9.

The simulation period was limited to the four hydrological years 2015–2018 by lack of reliable local meteorological data prior to this. Although relatively short, the simulation period contains a reasonable diversity of snow conditions, including relatively snow-rich and snow-poor winters.

The challenge of obtaining accurate spatial fields of meteorological variables in mountainous regions hastwo direct implications for modelling. The first is that precipitation measurements made using traditional instruments must be corrected for wind-induced undercatch and any other factors that induce systematic bias towards underestimation. Here, different corrections were applied depending on the incident precipitation phase (Equation (1); Schulla, 2017):

$$
\begin{aligned}
P_{corr} &= P(snoa{\cdot}WS + snob), \quad TA < rstt \\
P_{corr} &= P(liqa{\cdot}WS + liqb), \quad TA > rstt
\end{aligned}
\tag{1}
$$

where P is station measured precipitation (mm), P_{corr} is corrected precipitation (mm), *snoa* (-) and *snob* (-) are global correction factors for solid precipitation, *liqa* (-) and *liqb* (-) are global correction factors for solid precipitation, *WS* is wind speed (m·s^{-1}), and *rstt* is the rain-snow threshold temperature (°C).

The rain-snow threshold temperature, here denoted by *rstt*, cannot be reliably determined *a priori* because it varies in time and space on large scales (Jennings et al., 2018). As such, its value was optimised, alongside several others, through our calibration approach. Incorporating atmospheric humidity data could have resulted in more reliable phase determination. The magnitude of precipitation underestimation is likewise highly uncertain, although solid precipitation measurements are usually more affected than liquid precipitation ones. For this reason, both *snob* and *snoa* were also calibrated, whilst *liqa* and *liqb* were

assigned fixed values (0.01 and 1.02, respectively). As neither the error characteristics of solid nor liquid precipitation were known at individual stations and/or for individual events, more targeted corrections were not possible.

The second implication is that spatial interpolation algorithms must be carefully selected. Generally speaking, given the pronounced and complex topography, both spatial and elevation dependencies in the various meteorological variables should ideally be accounted for. A 25 m resolution digital terrain model (DTM; swisstopo, 2018) defined the model grid. Then, an appropriate algorithm was applied to interpolate all available measurements of each variable at every time-step. To account for their strong elevational dependence, air temperature, wind speed, relative humidity, and vapour pressure measurements were interpolated using Elevation Dependent Regression (EDR). For air temperature, the possibility of the linear relationships varying across elevation bands (including full temperature inversions) per time-step was permitted. For precipitation, meanwhile, a linear combination of fields generated independently by Inverse Distance Weighting (IDW) and EDR was applied. The ratio of the former to the latter, *idwedr*, was also calibrated. In this way, a certain balance between spatial patterns and (spatially constant) elevational dependence in the station measurements was achieved. Since incoming shortwave radiation and sunshine duration demonstrate more limited elevation dependence, they were simply interpolated using IDW. Whenever IDW was applied, the maximum search radius was set such that no stations were excluded.

This approach differs from that taken in certain other studies, which applied either predefined or else calibrated constant linear temperature-elevation gradients or elevation-precipitation gradients. For example, Brauchli et al. (2017) distributed corrected single station precipitation measurements across their study catchment by applying a constant lapse rate of 2%/100 m. This and other studies (e.g. Naseer et al., 2019) suggest that in complex terrain, such constant lapse rates may be unrealistic. Avoiding the use of such constant gradients can therefore be considered a strength of the present approach, as more of the spatial and temporal structure of the local meteorological measurements should be retained.

That said, with such an approach, the temporal coverage or "crossover" between the underlying station data (shown in Figures S3 to S9) becomes important. This is because, for a given meteorological variable and time-step, only stations returning observations contribute to the resultant spatial field. In other words, no temporal gap filling or interpolation is undertaken, and each time-step's field is independent from the last. Consequently, uncertainty in the interpolated spatial fields is not constant in time, but rather varies with both the number and locations of stations providing measurements.

Finally, the interpolated hourly temperature and radiation grids were corrected to account for topographic shading effects using the scheme of Oke (1987). An empirical temperature factor involved, *radc*, was also calibrated. The aforementioned steps were all undertaken using the distributed model WaSiM (Schulla, 2017).

3.2. Satellite and in situ snow observations

Two complementary types of observed snow data were prepared to constrain the model; i) binary observed snow extent maps derived from Landsat 8 imagery, and ii) SWE time-series at two station locations. The former provides complete spatial coverage, but only for temporal "snapshots", and moreover provides no direct information on snowpack water storage. Conversely, the latter provide high-frequency, temporally continuous information on SWE, but only at discrete locations.

17 Landsat 8 scenes that fall within the period of meteorological data availability (i.e. the hydrological years 2015–2018) and were cloud-free over the study area were considired. For each, the Normalised Difference Snow Index (NDSI; Dozier, 1989) of every pixel was first calculated according to Equation (2):

$$NDSI_{L8} = \frac{B_3 - B_6}{B_3 + B_6} \qquad (2)$$

where B_3 and B_6 are Bands 3 (0.525–0.600 μm) and 6 (1.560–1.660 μm) of a given Landsat 8 (*L8*) image, respectively. Lakes were masked out because their reflectance signatures gave rise to NDSI value "peaks" that obscured the boundary between snow and snow-free pixels in some image histograms.

Snow extents were then delineated by applying a threshold to the NSDI maps. An iterative process of threshold adjustment and visual comparison of the binary snow maps and the corresponding True Colour Composite (TCC) images was followed until a satisfactory final classification was reached. Fig. 2 illustrates the main steps involved. The final thresholds varied somewhat between images – see Härer et al. (2018) for a dedicated exploration of NDSI threshold choice. Whilst the manual approach to threshold identification employed here was feasible for this fairly small image catalogue, generating accurate snow maps for a larger catalogue would require a more automated procedure.

To facilitate their integration in the automated calibration process, the maps were reprojected to the CH1903 system (EPSG: 21781) and clipped to the study catchment. Using the nearest neighbour approach, they were then downscaled from their native 30 m resolution to align with the 25 m resolution model grid. The full catalogue of observed snow extents, which encompasses practically the full range of possible snow cover conditions, is shown in Fig. 3. Fig. S11, meanwhile, provides comparisons of all TCC images and delineated observed snow extents (alongside the corresponding final simulated outputs).

Besides the more general issues of cloud cover and temporal gaps between satellite overpasses, dark forest canopies and areas of shadow represent the main challenges associated with using mapping snow extents using NDSI in steep mountainous terrain (Wang et al., 2015). In this case, whilst a small number of snow-covered pixels under dark forest and in heavily shaded snow-covered terrain may have been misclassified as snow-free, the maps generally compare very well the TCC images. Moreover, a sensitivity assessment of the classified snow extents to various plausible thresholds (not shown) found it to be small. As such, the maps can be taken to represent snow extents with a reasonably high degree of accuracy, especially in the more open, upper regions of the catchment, where snow patterns are of most interest.

To provide more direct information on snow water storage to the model, SWE time-series were reconstructed at two contrasting station locations. (SWE reconstruction was also required to enable direct comparisons with the model outputs, since the snow model configuration employed only provides total water storage, rather than full information on the interrelation between snowpack density, depth and SWE). Different methods were used at each station according to data availability.

The two stations providing regular snow measurements over the period in question are located just outside the study catchment, with each lying towards one extreme of its elevational range (see Fig. 1). The data were again obtained from IDAWEB. The higher station, Grand Cor (COR; Elevation: 2602 m), belongs to the SLF's Intercantonal Measurement and Information System (IMIS) (SLF, 2019b). These stations do not measure solid precipitation directly, but instead record snow depth and several other variables that can be used to drive the 1D physically-based, multi-layer model SNOWPACK (Lehning et al., 2002). An hourly time-series SWE evolution at COR over the entire four-year simulation period was constructed in this fashion.

The second, lower elevation station is located in Gryon (GRY; Elevation: 1146 m). Unlike at COR, only (manual) daily snow height measurements are made here, which prevented an application of SNOWPACK. The empirical model of Jonas et al. (2009), which was constructed using a large sample of snow observations from the Swiss Alps, was therefore applied. This approach facilitates the estimation of snow density as a function of geographic region, month of the year, and site elevation. Parameters corresponding to the elevation band < 1400 m and the "Region 1" offset were taken (see Tables 1 and 2 of Jonas et al., 2009); these being applicable to GRY. The resultant densities were then multiplied by measured snow heights to give daily SWE estimates.

Due to this reconstruction work, neither of the "observed" SWE time-series are actually direct measurements. The simulation domain was extended slightly to allow these data to be used in the model calibration.

4. Numerical modelling and calibration

4.1. Simulating snow accumulation, redistribution, and melt

As for the previous steps, WaSiM (Schulla, 2017) formed the foundation of the snow modelling approach. This decision was made following a review and testing of possible alternatives, including
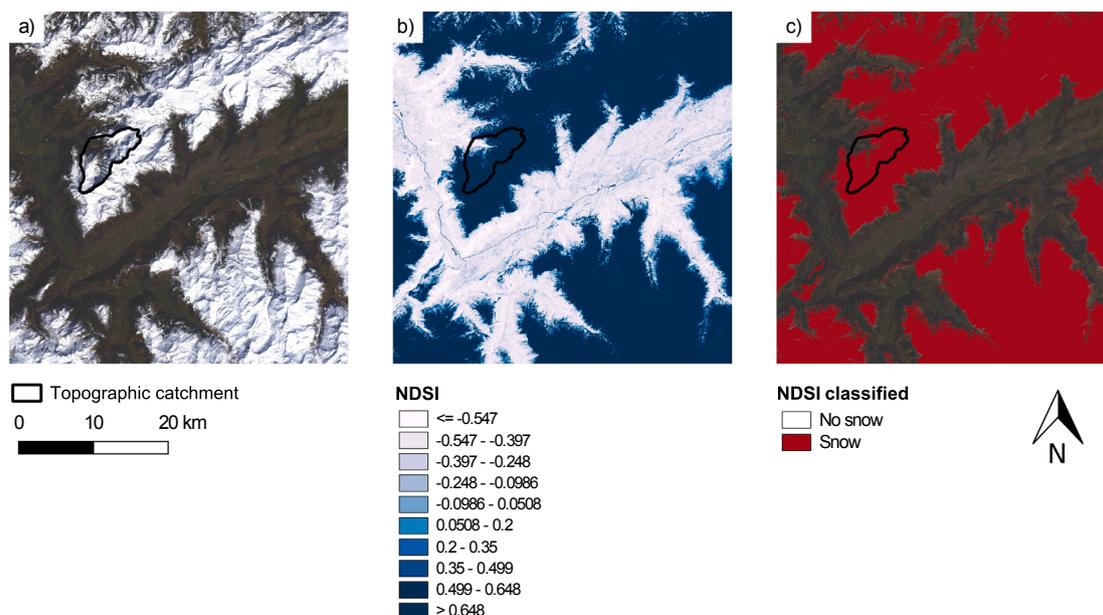
Fig. 2. Illustration of the binary snow extent map generation process, taking the 8 April 2015 as an example: a) True Colour Composite, b) calculated NDSI raster, and c) final binary observed extent developed by identifying and applying a threshold to (b). Each image in the catalogue was inspected and classified individually.
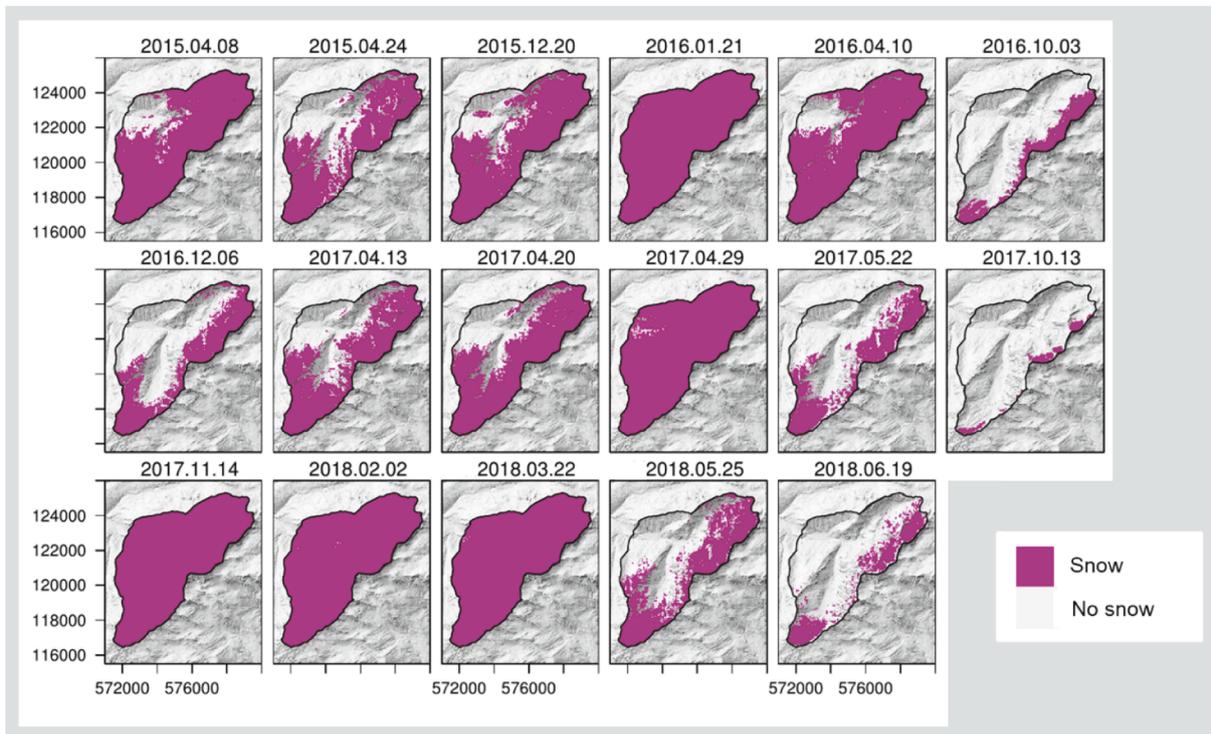
Fig. 3. The complete catalogue of 30-metre resolution observed snow extent maps compiled. The date of each image is indicated. Coordinates are in the CH1903 system.

Alpine3D which – although a strong contender – did not enable gravitational snow redistribution to be accounted for. Although slightly more sophisticated snow models are available, few fully-distributed codes offer such a broad range of functionality as WaSiM.

Snow accumulation, gravitational redistribution, and melt were calculated on the 25 m model grid on an hourly time-step. First, the (corrected) precipitation phase per pixel and time-step was estimated according to the interpolated air temperature and a transitional range within which both solid and precipitation can occur (Equation (3)):

$$S_{frac} = \frac{rstt + T_{trans} - TA}{2 \times T_{trans}} \; for \; (rstt - T_{trans}) \; < \; TA \; < \; (rstt + T_{trans}) \quad (3)$$

where S_{frac} is the fraction of the totalprecipitation that is snow (0–1), TA is the air temperature (°C), $rstt$ (°C) is the same rain-snow threshold temperature applied in Equation (1), and T_{trans} is half the rain-snow transition temperature range (°C).

T_{trans} was fixed to 1°C (i.e. the total transition range was 2°C), whilst $rstt$ took the same (calibrated) value as that applied to distinguish precipitation phase (Equation (1)). Snowmelt was then calculated by solving the surface energy balance for the energy available for melt following the approach of Warscher et al. (2013). In this scheme, the snowpack is treated as a single homogeneous layer beneath the surface, for which the energy balance is computed (Equation (4)):

$$Q + H + E + A + G + M_{ae} = 0 \quad (4)$$

where Q is the shortwave and longwave radiation balance, H is the sensible heat flux, E is the latent heat flux, A is the advective energy supplied by solid or liquid precipitation, G is the soil heat flux (which is small compared to other fluxes and here was set equal to 2, and M_{ae} is the energy potentially available for melting during a given time-step (the units of all terms are $W \cdot m^{-2}$).

Melting and non-melting conditions were distinguished according to $rstt$. When the energy balance is positive (i.e. $M_{ae} > 0$) and air temperature favourable, melt (M) can occur (see Warscher et al. (2013) for an explanation of the use of air temperature as a proxy to differentiate

melting from non-melting conditions). Finally, M per time-step, dt, is expressed in mm of water by introducing the latent heat of fusion, c_i (Equation (5)).

$$M = \frac{M_{ae} \times dt}{c_i} \quad (5)$$

Sublimation, which can be an important component of Alpine water balances (Strasser et al., 2008), is explicitly accounted for in this approach.

Two additional scaling parameters, $lwin$ and $lwout$, were available to fine-tune the incoming and outgoing longwave components of the energy balance, respectively. Raleigh et al. (2016) showed that behind temperature and precipitation, longwave radiation estimates most strongly affect energy balance snow simulations results, and moreover noted that longwave measurements are uncommon in high elevation terrain; such considerations justify including parameters related to longwave radiation in the calibration. Here, both parameters were subjected to calibration, albeit within relatively narrow bounds (see Table 1). In this way, potential errors in both surface albedo and cloudiness (as determined from the interpolated sunshine duration fields, used in the calculation of incoming longwave radiation) could be accounted for.

In addition, gravitational redistribution was simulated using a mass-conservative algorithm that is underpinned by a topographic analysis. This algorithm was implemented in WaSiM by Warscher et al. (2013). Several steps are involved; as with the previously summarised algorithms, they are comprehensively described in Warscher et al. (2013) and Schulla (2017). Here, only the main parameters are discussed. Two represent critical local slope limits; $mids$ is the lower inclination limit for gravitational slides to occur, and $mads$ is the upper inclination angle (above which all incoming snow is immediately transported downslope). Because these slope angles are dependent on the model grid scale, they cannot easily be transferred from previous studies and were therefore calibratated. Following the advice of Schulla (2017), two further parameters related to gravitational redistribution were also

Table 1

WaSiM model parameters that were subject to calibration. The final estimated parameter values are also reported here to prevent the later duplication of a very similar table. *These parameters had a lower bound of zero, but this had to be raised marginally for practical implementation in PEST.

Parameter	Description	Lower Bound	Upper Bound	Estimated value
rstt	Snow-rain temperature threshold (°C)	0.0*	3.5	**0.0266**
snoa	Snow precipitation correction (-)	0.0*	0.15	**0.0283**
snob	Snow precipitation correction (-)	1.0	1.45	**1.4500**
radc	Factor for temperature correction *radc* · (-1.6 .... + 1.6) in the radiation correction module	0.1	8.0	**0.1731**
mads	Maximum slope for snow deposition (°)	45.0	75.0	**73.5269**
scmd	Upper deposition limit for gravitational redistribution (mm)	0.0*	10.0	**1.1497**
mids	Minimum slope for gravitational slides (°)	20.0	48.0	**43.3811**
frss	Fraction of snowpack that forms the slide (0–1)	0.001	0.05	**0.0076**
idwedr	Relative weight of IDW to EDR in the interpolation of precipitation (0–1)	0.05	0.85	**0.05**
lwin	Correction factor for incoming long wave radiation for energy balance fine tuning(accounting for errors in cloudiness and albedo)	0.7	1.3	**1.2167**
lwou	Correction factor for outgoing long wave radiation for energy balance fine tuning(accounting for errors in cloudiness and albedo)	0.7	1.3	**1.1920**

calibrated. The first is the fraction of the current snow storage in a cell that can form a slide in any given time-step, *frss*. It is recommended that its value be set to some small fraction, typically ~1%, although this depends somewhat on the time-step at which the model is run. The second, *scmd,* is an upper depositional mass limit (mm) for snow flows, i. e. the maximum permitted transfer from one cell to another, again per time-step. Whilst such an approach can estimate plausible snow redistribution patterns, it should be noted that the specific timing of avalanches cannot be predicted (Warscher et al., 2013).

Warscher et al. (2013) also proposed a simple algorithm designed to account for snow redistribution by wind. More recently still, methods seeking to improve WaSiM's representation of snow and coniferous forest canopy interactions have been published (Förster et al., 2018). However, for several reasons, neither of these algorithm sets was included in our final model here. Firstly, given the study area's steepness, we decided to focus on gravitational redistribution. Secondly, including wind algorithm actually resulted in poorer model-observation fits (see Section 5.7). Finally, the WaSiM release containing the extensions of Förster et al. (2018) came too late in the course of our work to be evaluated thoroughly.

4.2. Multi-objective calibration

The 11 empirical parameters that required estimation are listed, alongside the upper and lower bounds that were assigned to each based on prior knowledge, in Table 1. For numerical reasons, all parameters were log transformed. This step is recommended when using PEST so that the linearity assumption (of model outputs to varied parameters) holds better, and to normalize parameters with respect to their inherent variability. To this end, the lower bounds of three parameters which would ordinarily have been zero were marginally raised. The final estimated values are also listed in Table 1 to prevent later duplication being necessary.

A novel, multi-objective calibration approach that incorporated both the spatial snow extents and the reconstructed SWE time-series was then developed. For each of the 17 days with an observed extent map and

model iteration, the spatial component of goodness-of-fit was quantified as follows:

1. Simulated SWE maps at the end of the days for which observed maps were available were extracted and clipped to the study catchment.
2. Pixels in the simulated SWE maps were reclassified to either snow or no-snow using a 5 mm exceedance threshold (i.e. pixels with SWE > 5 mm were classified as snow covered).
3. All pixels were binned into one of the quadrants of the contingency matrix shown in Table 2 according to whether snow presence/absence had been correctly simulated with respect to the observed maps.
4. Three related performance metrics were calculated after Aronica et al. (2002) using Equations (6) to (8).

$$F_1 = \frac{\sum_{i=1}^n a + \sum_{i=1}^n d}{n} \tag{6}$$

$$F_2 = \frac{\sum_{i=1}^n a}{\sum_{i=1}^n a + \sum_{i=1}^n b + \sum_{i=1}^n c} \tag{7}$$

$$F_3 = \frac{\sum_{i=1}^n a - \sum_{i=1}^n b}{\sum_{i=1}^n a + \sum_{i=1}^n b + \sum_{i=1}^n c} \tag{8}$$

where *a, b,* and *c* are the quadrants of the contingency matrix (Table 2), and *n* is the total number of pixels.

F_1 corresponds to the overall proportion of correctly simulated pixels. F_2 and F_3 expressly discount pixels that are snow free in both simulations and observations, and so typically result in lower scores – on

Table 2

Contingency matrix used for the per-pixel classification of distributed snow model outputs.

	Observed snow	Observed no snow
Simulated snow	*a*	*b*
Simulated no snow	*c*	*d*

many days this quadrant can be heavily populated as snow-free pixels at lower elevations are generally relatively easy to reproduce (Warscher et al., 2013). In each case, a perfect fit between simulations and observations would return a value of one. Therefore, for each model iteration, the squared residuals between the three F-statistics obtained and one was calculated. Model performance with respect to the "observed" SWE time-series was quantified using the squared residuals between simulated and observed values per time-step.

To construct a single, multi-objective function that could be minimised and thus produce the best overall fit according to both types of observations, weights had to be assigned to each squared residual. This (subjective) process aimed to ensure that each observation maintained a certain "visibility" in the calibration process, whilst an appropriate balance between the two data types was achieved. The hierarchical weighting scheme illustrated in Fig. S2 was ultimately implemented.

In summary, a slightly higher weighting was applied to the spatial data than the time-series (60:40). To reflect their more stringent nature, F_2 and F_3 values were assigned double the weight of F_1 values. Finally, to account for the disparity in measurement frequency at the snow stations – hourly at COR but only daily at GRY – observations at the latter were assigned weights 24 times higher than those at the former. The objective function (OF) is expressed in Equation (9):

$$OF = \sum_{i=1}^{17}\left[wF_1(1-F_1)^2\right] + \sum_{i=1}^{17}\left[wF_2(1-F_2)^2\right] + \sum_{i=1}^{17}\left[wF_3(1-F_3)^2\right]$$
$$+ \sum_{i=1}^{35063}\left[wCOR(COR.SWE_{sim} - COR.SWE_{obs})^2\right]$$
$$+ \sum_{i=1}^{1461}\left[wGRY(GRY.SWE_{sim} - GRY.SWE_{obs})^2\right]$$

(9)

where $wF_1, wF_2, wF_3, wCOR,$ and $wGRY$ are the relative weights that were assigned to each observation belonging to the different observation groups, as illustrated in Fig. S2, i.e. 0.706, 1.412, 1.412, 0.0057 and 0.01368%, respectively, F_1, F_2, and F_3 are the fit statistics calculated for each of the 17 pairs of images according to Equations (6) to (8), and $COR.SWE_{sim}$, and $COR.SWE_{obs}$, $GRY.SWE_{sim}$, and $GRY.SWE_{obs}$ are the observed and simulated SWE values at each time time-step at the COR and GRY stations, respectively.

The WaSiM model was then linked with PEST (Doherty, 2019) – a model-independent, gradient-based parameter estimation tool which uses the Levenberg-Marquardt algorithm. PEST repeatedly runs the model, altering the calibration parameter values each iteration in an attempt to minimise the objective function (in a least squares sense). PEST was selected primarily due to its efficiency, which is considerably higher than that of commonly applied Monte Carlo-oriented approaches. Indeed, parameter search efficiency was crucial given the relatively high computational demands of the energy balance snow model. The coupling was achieved by implementing routines to extract the spatial and temporal model outputs corresponding to the observations, and then calculate the required statistics (see the Appendix A). The final parameters values obtained are presented in Table 1. Note that another valid approach to assign weights (and one that may have accounted better for their contrasting magnitudes) would have been to run the model once and then automatically adjust the weights using PEST's PWTADJ1 utility such that each observation group gave an approximately equal contribution.

4.3. Predictive uncertainty and data worth analyses

A linear analysis was then conducted to quantify the pre- and post-calibration uncertainty associated with selected individual "predictions" of interest (the term predictions is not used here to allude to the future); namely, the SWE at COR on 1 April in each of the simulated hydrological years (2015–2018), and the spatial F_1 metric for 22 May

2017. This strategy enables any reduction in uncertainty achieved through the calibration process to be evaluated.

1 April SWE at station locations is an indicator commonly employed by environmental managers in snowmelt-dependent regions to predict water availability throughout the subsequent summer. The spatial prediction was included because few (if any) previous studies have specifically considered uncertainty in snow pattern predictions. Analyses quantifying the contribution of individual parameters to pre- and post-calibration uncertainty variance, as well as the information provided by the five observation groups the calibration process (i.e. "data worth"), were also undertaken. To achieve these tasks, tools from PEST's GENLINPRED suite were applied. For a through description see Doherty (2010, 2019).

In contrast to the model calibration phase, for these analyses, identical weights were assigned to all non-zero weighted observations (zero-weighed observations being the predictions of interest). Following the advice of Doherty (2010), this weight was estimated by taking the number of non-zero weighted observations (here 36,570), calculating its square root, and dividing the result by the calibrated model objective function (91,150) – giving a value of 0.002098.

4.4. Estimating glacial melt, liquid precipitation, and potential evapotranspiration

To generate comprehensive forcing data for subsequent distributed, integrated hydrological simulations, four additional datasets were also generated using the model; liquid precipitation, firn melt, ice melt, and ET_p. Firstly, to account for liquid precipitation in addition to snowmelt, "snowcover outflow" grids were written at each time-step. These represent the snowmelt (as calculated by the snow model) from snow-covered pixels plus any liquid precipitation falling on snow-free pixels. As Section 3.1 explained, modest fixed corrections were applied to the raw liquid precipitation measurements to account for undercatch. Accordingly, for snow-free pixels, the "snowcover outflow" values correspond to any (corrected, interpolated) rainfall falling.

As the coverage of glaciers is small (<3%), glacial melt makes a much more modest contribution to annual, catchment-averaged meltwater input than snowmelt. Nevertheless, glacial meltwater generation can be locally considerable in summer. As such, a "dynamic" glacier model (employing a simple volume-area scaling relationship, with default parameter values; see Schulla, 2017) that accounts for accumulation, dynamics, and ablation (with radiation correction) was applied in WaSiM. The parameters of this model were not calibrated due to the overall dominance of snowmelt and a lack of glacial data; the Glacier des Martinets, for instance, has not been actively monitored since 1975 (SCNAT, 2018). Rather, the intention was simply to ensure that ice melt was not entirely neglected.

For snow on the glaciers, an identical approach to the main snow model was taken. In this way, distinct (hourly, 25 m resolution) grids representing snowmelt, firn melt, and ice melt from glacierised areas were produced. Besides containing all the optimised snow-related parameters, the WaSiM control file in Appendix A indicates the values of the (fixed) parameters that were applied to generate these additional datasets. With only slight modification, possibly related to the distinction between snow and bare glacier ice in the NDSI images, the approach could be easily transferred to more glacierised catchments.

The glacier meltwater estimates had to be normalized according to the glacier-covered fraction of each cell per time-step, which ranged from 0 to 1. Having done this, the (off glacier) "snowcover outflow" rasters were summed together with the corresponding normalised "snowmelt on glacier", "firn melt", and "ice melt" grids per time-step to produce a single set of hourly "all liquid water" grids. These calculations were carried out by executing batch GDAL scripts (GDAL, 2019; OSGeo4W, 2019). Units were also converted for subsequent hydrological modelling.

Finally, the Penman-Monteith method was used to estimate 25 m

resolution, hourly ET_p. To achieve this, classes in a land cover map developed from existing swisstopo datasets (swisstopo, 2019; Fig. S10) were attributed with appropriate physical parameters (e.g. Leaf Area Index; LIA). Fuhrer and Jasper (2012) describe a dedicated application of WaSiM to this end. Apart from unit conversion for subsequent hydrological modelling, no additional processing of the ET_p grids was required

In possessing identical spatio-temporal resolution and having been generated using predominantly physically-based approaches, all resultant datasets can be considered broadly commensurate with one another.

5. Results and discussion

5.1. Estimated parameter values

The parameter values estimated via inversion (Table 1) constitute an initial set of "results". Two in particular are interesting to consider briefly. Firstly, the high value of 1.45 taken by the wind speed-independent snow correction constant, *snob*, which actually reached the permitted upper bound, attests to the considerable underestimation bias generally contained within the winter station precipitation measurements. Recall here that this constant factor was combined with the wind speed-dependent factor, *snoa*, which was estimated to be 0.0283 (i.e. an increase of 2.83% per additional m·s^{-1} of wind speed) and the interpolated wind speed to determine effective precipitation. Secondly, *idwedr* took its lowest permitted value, i.e. improved fits were obtained when any elevational gradients present in the station data were enforced upon the interpolated precipitation fields. This likely reflects substantial differences in cumulative precipitation between the lowest elevation stations and those at elevations that correspond more closely to the study catchment.

The estimated solid precipitation correction factor magnitude is broadly consistent with existing literature. For instance, Sevruk (1985) suggested that precipitation in Switzerland is underestimated by between 4% (in summer at low elevations) and approximately 40% (at high altitudes in winter). In a simulation of a 200 km^2 region of Switzerland, Bavera et al. (2014) applied a fixed 30% factor to solid precipitation. Pan et al. (2016), meanwhile, found the sheltering effect of surrounding vegetation to be an important influence on the magnitude of any underestimation in northern Canada, with well-sheltered sites requiring much less correction. At open sites, the bias corrections they applied increased annual precipitation by between 15 and 34%. At windy sites with high snowfall proportions, even larger corrections – sometimes exceeding 50% – were sometimes be necessary. Finally, Jimeno-Sáez et al. (2020) found that very large undercatch uplift factors

(78% for solid precipitation) were required in a basin in the Sierra Nevada, southern Spain. Therefore, the upper bounds of our solid precipitation correction factors could perhaps have been set more generously.

5.2. Correspondence with observations

The spatial goodness-of-fit statistics (i.e. F_1, F_2, and F_3) obtained following calibration are shown in Table 3. The corresponding observed and simulated snow maps from which these statistics were calculated are shown in Fig. S11.

The F-statistics returned are generally high; across the 17 days, the average percentage of correctly simulated pixels is 85%. As expected, the scores decline progressively from F_1 to F_3. Moreover, the variability in the statistics between images was observed, with the highest scores naturally being achieved on completely snow covered, mid-winter days. The lowest F_1 value was for the 29 April 2017, when the model missed a late season snowstorm that briefly blanketed the catchment.

Fig. 4 shows the comparison between simulated and observed SWE time-series at the two measurement stations. The dynamics of snowpack evolution are replicated adequately, including the contrast between seasonal snowpack at the higher elevation station (COR) and the more intermittent pattern at the lower one (GRY). The colder prevailing conditions at COR allow the results to be discussed explicitly in terms of the accumulation and ablation phases. This simulated onset of accumulation closely matches the observations here, as do changes in accumulation rate. The timing and rate of ablation is likewise broadly consistent with the observations. However, there does seem to be a general tendency towards underestimation of accumulation totals. A similar underestimation is evident in the first three years at GRY. Across both stations, observations from winter 2017/2018 are reproduced best, perhaps due to increased local data availability.

In evaluating these results, it must be highlighted that the key meteorological variables of precipitation at COR, and both precipitation and temperature at GRY, were not actually measured at these locations, but rather had to be spatially interpolated (onto the corresponding model cells) from station measurements elsewhere. As such, much of the remaining post-calibration difference is likely to be associated with uncertainties in these interpolated fields. In addition, as explained earlier, the "observations" themselves were reconstructed. Finally, the observations were made at discrete locations, whilst the simulated values correspond to the 25 m pixel within which each station was located. Hence, should the terrain at the station locations not be entirely representative of its immediate surroundings, this would represent another potential source of mismatch. A shaded region corresponding to $\pm$ 20% around the observations – that being roughly the maximum SWE mismatch one could expect purely for this reason (T. Jonas, *pers. comm.*) – is shown to reflect this.

5.3. Comparison of simulated spatial statistics with previous studies

In Fig. 5, the F-statistics obtained are compared to those reported in previous studies that also employed these metrics to quantify the fit between distributed snow models and maps derived from Landsat imagery. Three such studies are known, each of which reported statistics for only a small number of days (between one and three); Schöber et al. (2010) presented F_1 values for two days but numerous catchments, while Bernhardt et al. (2012) and Warscher et al. (2013) provided all three F-statistics for their respective study catchments and selected days. In contrast to the present study, which also included mid-winter and late summer days in the calibration catalogue, the previously published statistics correspond to spring and early summer periods exclusively (i.e. partially snow-covered conditions). As such, to ensure the fairest possible comparisons, only our statistics corresponding to this spring and early summer are included. The underlying data are compiled in Table S2.

Table 3
Post-calibration F-statistics quantifying the spatial goodness-of-fit for each of the 17 days.

Date	F_1	F_2	F_3
08/04/2015	0.852	0.847	0.696
24/04/2015	0.786	0.698	0.511
20/12/2015	0.838	0.798	0.699
21/01/2016	0.999	0.999	0.999
10/04/2016	0.871	0.849	0.819
03/10/2016	0.910	0.621	0.399
06/12/2016	0.750	0.596	0.199
13/04/2017	0.744	0.569	0.504
20/04/2017	0.790	0.752	0.508
29/04/2017	0.567	0.561	0.561
22/05/2017	0.846	0.626	0.434
13/10/2017	0.958	0.131	0.073
14/11/2017	1.000	1.000	1.000
02/02/2018	0.997	0.997	0.996
22/03/2018	0.963	0.963	0.963
25/05/2018	0.773	0.584	0.348
19/06/2018	0.844	0.328	0.089
mean	**0.852**	**0.701**	**0.576**

Fully Grd. Cor [COR] (2602 m)

Gryon [GRY] (1147 m)

Fig. 4. Observed and simulated Snow Water Equivalent (SWE) time-series at the two measurement stations, Fully Grand Cor (a: COR, hourly) and Gryon (b: GRY, daily). The observations are not direct measurements of SWE, but rather are reconstructions based on snow depth and other measurements at COR, and purely snow depth at GRY. The grey bar of ± 20% is added around the reconstructed "observed" series to reflect the pixel-point nature of these comparisons.

The F-statistic distributions obtained seem to have a tendency to be slightly higher than those of previous studies, even if the mean and median of F_1 are marginally lower than those calculated from the published values. This tendency appears to be most pronounced for F_2 and F_3; in both cases, the mean values obtained here are higher than their previously published counterparts. Importantly, with the possible exception of Schöber et al. (2010) who may have used the spatial snow observations in an informal fashion to adjust certain model parameters (ambiguity remains as this step was not fully explained), the previous studies used the spatial data purely for model evaluation (in contrast to the present study, in which they formed calibration targets).

As such, while perhaps slightly disappointing that the explicit calibration did not yield higher F_1 scores, a real benefit of calibration can arguably be seen in the noticeably higher F_2 and F_3 scores. These metrics were assigned enhanced weight in the calibration processes. That said, the small sample sizes must be borne in mind when making such interpretations. An additional consideration is that our calibration was not informed solely by the observed snow extent maps, but rather sought to achieve an acceptable balance between fits according to both the maps and SWE time-series. Improved spatial fits could probably have been achieved if the maps alone comprised the objective function, but this may have come at the expense of reduced accuracy in simulated
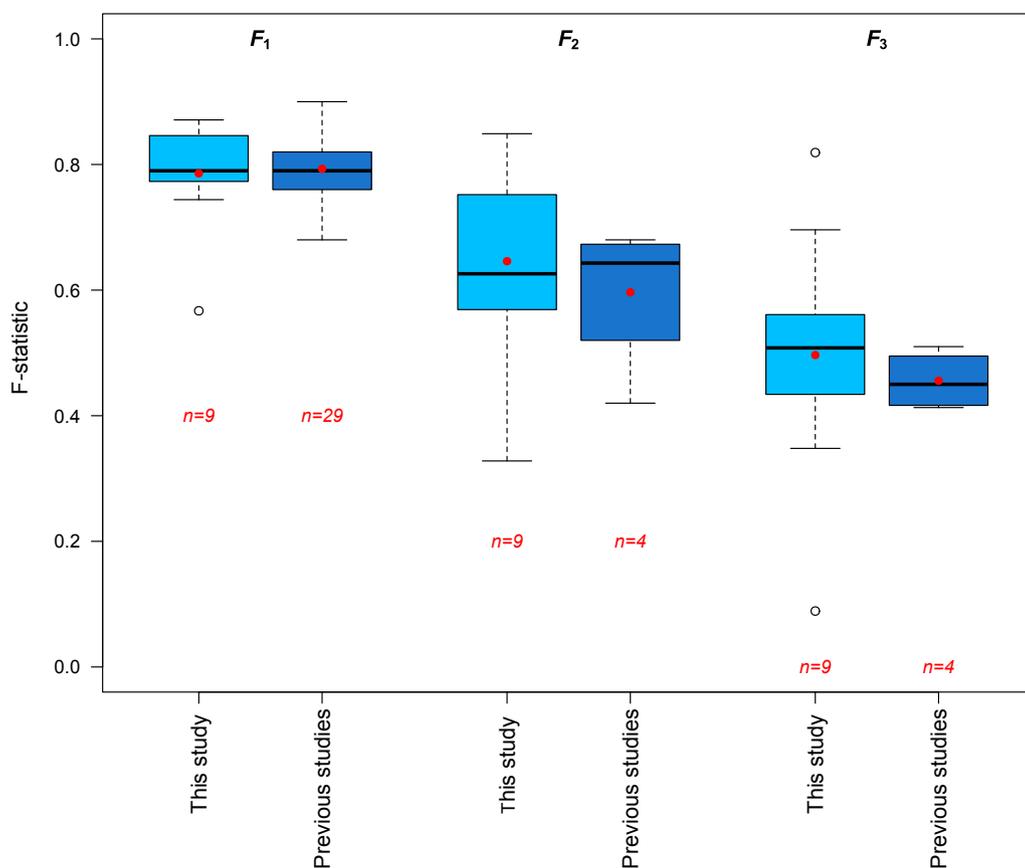
Fig. 5. Comparison between the spatial model fits obtained in the present study and previously reported equivalents (i.e. the same metrics generated with respect to Landsat imagery). In the interests of fairness, only the F-statistics from this study that corresponded to spring and early summer (i.e. 08/04/2015, 24/04/2015, 10/04/2016, 13/04/2017, 20/04/2017, 29/04/2017, 22/05/2017, 25/05/2018, 19/06/2018) are included. The medians are represented using a thick black line in the traditional fashion, whilst the red dots indicate the means. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

catchment-wide snow water storage.

A much larger catalogue of published F-statistics would be required to assess the statistical significance of these results, i.e. whether the calibration approach proposed does lead to real improvements in the capacity of distributed snow models to reproduce high-resolution observed extents. Nevertheless, the comparisons presented confirm the overall appropriateness of the approach taken, and perhaps even alludes to some added value. Finally, given the variability in the F-statistics obtained between days, the larger number of days for which performance statistics were presented here (which elucidate this) is a strength of this study; because previous studies considered fewer days, it is unclear how consistent in time their (generally good) model performance might be.

5.4. Snowpack evolution and hydrological plausibility

A key benefit of the model (and indeed any distributed, transient simulator) is that it "fills in the gaps" in space and time between the available observations. To visualise this, Animation S1 presents the simulated evolution of SWE at a daily time-step during Winter 2017/2018. The redistribution of snow from steep slopes is particularly apparent, with patterns being broadly consistent with both our local field experience during the period (Fig. S1) and the avalanche activity that occurred in the region more generally (Bühler et al., 2019).

Fig. 6 illustrates the spatio-temporal distribution of a) "all liquid water" (comprised of liquid precipitation, snowmelt, firn melt and ice melt) and b) ET_p over the last hydrological year of the simulation period, aggregated on a monthly basis. As expected, very little simulated meltwater input occurs during the winter months, when temperatures are generally below freezing. Conversely, the highest meltwater volumes are generated during the spring melt, especially the months of April, May, and June. The elevation at which the majority of melt water is produced increases as the season progresses, and the localised

contribution of the glaciers during the summer months is also evident. Indeed, extremely high values are generated locally from the glaciers, especially in Summer 2018, when heatwave conditions were experienced. In our simulation, the glaciers can exhibit strongly negative mass balance.

Liquid precipitation during the summer and autumn months, which can be highly concentrated in space and time, is naturally "smoothed out" in these plots, appearing as a fairly low and constant daily mean value in non-glacierised areas. That said, Fig. S12 presents the same underlying data in an alternative fashion – as hourly catchment-averaged series – indicating the dynamism of the system. Strong seasonality and elevational influences are apparent in the spatial patterns of ET_p, with estimated values being widely low in winter but restricted to higher elevations in summer.

To further verify the hydrological plausibility of these results (without recourse to a full hydrological model), simulated "snowcover outflow" and observed discharge were compared for the Vallon de Nant sub-catchment. A concrete weir gauging station at the outlet of this sub-catchment (Avançon Weir; Fig. 1) provides high-frequency streamflow estimates from spring 2016 onwards, from which hourly mean flows were calculated. The empirical stage-discharge relationship (i.e. rating curve) was developed by salt dilution gauging (Ceperley et al., 2018). Despite the regular cross-section, these data are somewhat uncertain, especially at flow extremes (at low flows, this is due to shifting channel configurations immediately upstream of the weir).

In Fig. 7(a), hourly catchment-averaged "snowcover outflow" (i.e. snowmelt from non-glaciated areas plus any liquid precipitation) is plotted against hourly observed, catchment area-normalised discharge measured at the Avançon Weir over spring 2018 (April to June inclusive).

Simply comparing the two cumulative totals gives an estimated runoff ratio of 0.61 over the three-month period, although this value is only tentative given uncertainties associated with precipitation, the
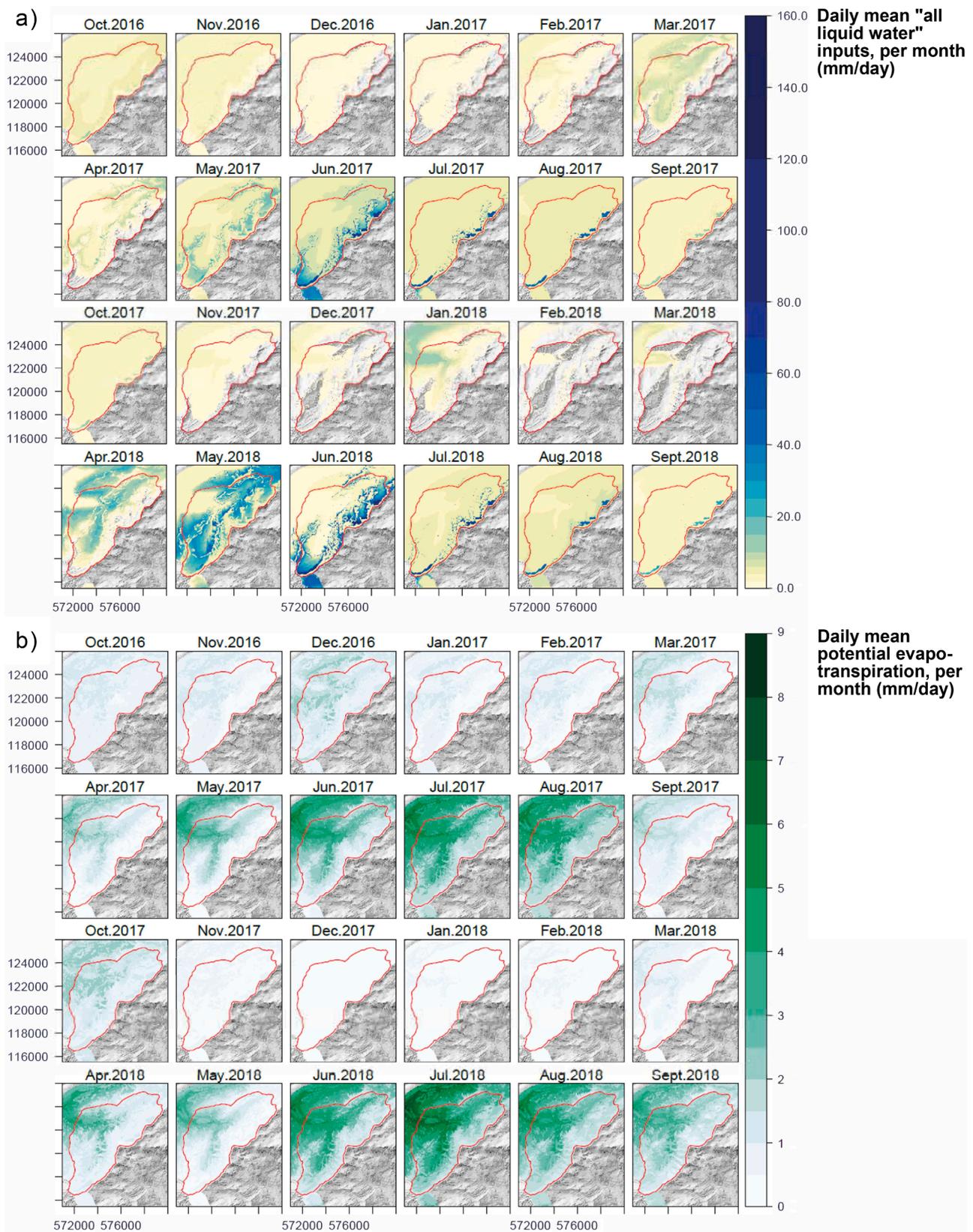
Fig. 6. Simulated spatio-temporal patterns of a) "all liquid water" arriving at the surface (i.e. liquid precipitation + snowmelt + firn melt + ice melt), and b) potential evapotranspiration (ET_p) generated using the optimised model configuration over the two hydrological years 2017–2018. The underlying hourly data are expressed here as daily mean values in mm, averaged across calendar months.
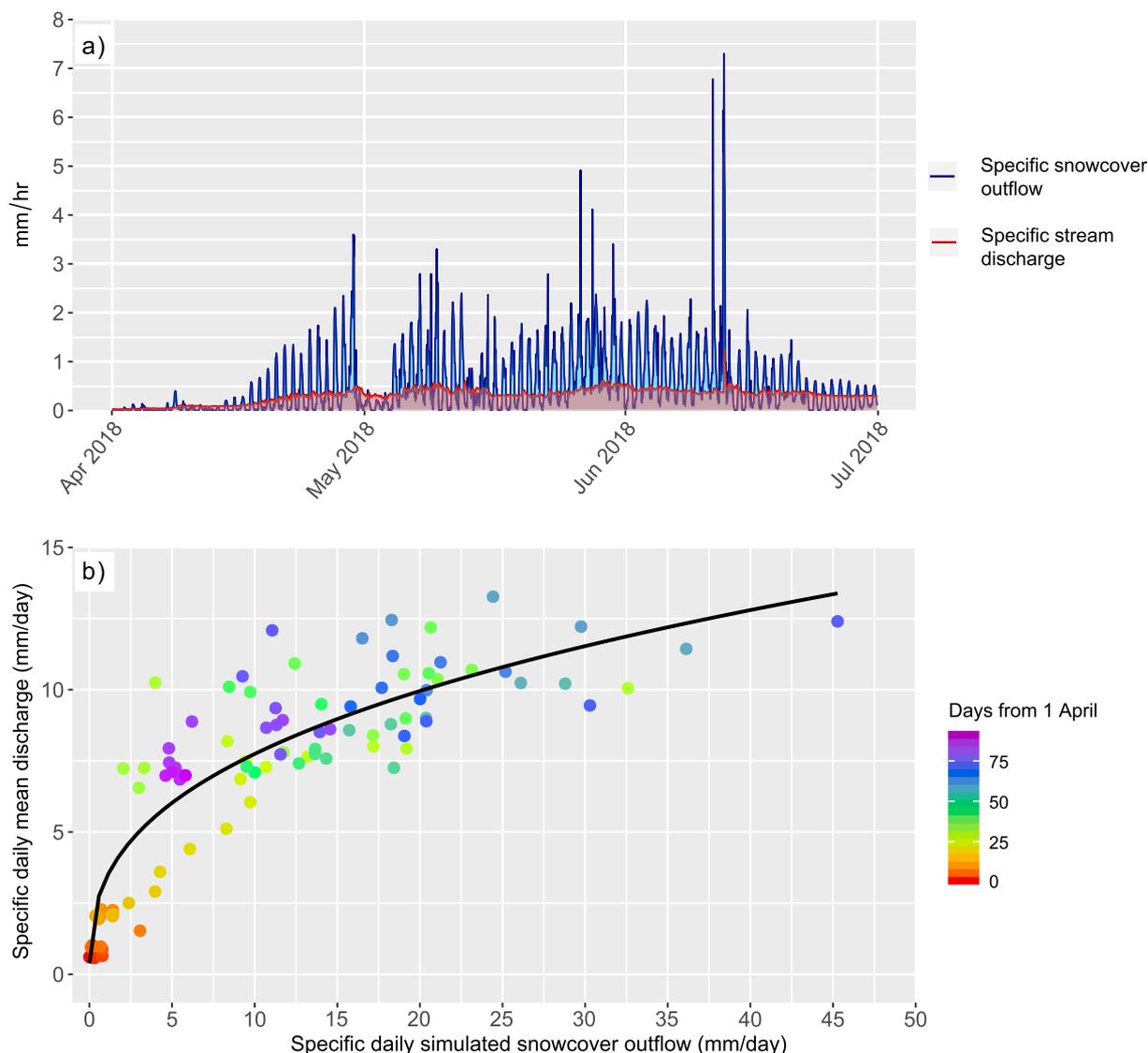
Fig. 7. a) Simulated hourly specific (i.e. catchment-averaged) "snowcover outflow" (i.e. snowmelt from non glaciated areas, plus any liquid precipitation) vs. hourly observed stream specific discharge for the Vallon de Nant sub-catchment during spring 2018 (discharge gauged at Avancon Weir and normalised according to catchment area), and (b) simulated daily sum specific "snowcover outflow" vs. observed daily mean specific stream discharge, again at the Avancon Weir station, over the same 3-month period.

model, and the observed discharge data. At the diurnal timescale, simulated snowmelt is associated with increasing observed streamflow, which seems reasonable. Dependence remains present on slightly longer frequencies also. For example, the decrease in measured streamflow just before the start of May coincides with a marked reduction in simulated water inputs. Extending this plot to later summer (not shown) revealed that a certain proportion of the "excess" spring melt inputs arrive in the stream later, reflecting the "buffering" capacity of relatively shallow groundwater storage. In reality, a proportion of meltwater will of course also be lost to actual evapotranspiration, and perhaps also to deeper groundwater storage and/or groundwater exportation across topographic divides.

Fig. 7(b), meanwhile, shows the relationship between these data aggregated to a daily time-step (as the lagged and strongly dampened observed streamflow response relative to the simulated melt inputs complicates hourly comparisons). A clear relationship between the variables is apparent which can be approximated by a power-law function (illustrated by the estimated non-linear least squares regression line). A certain hysteresis is also present, meaning that less meltwater is required to produce a given magnitude of flow response as the

Table 4
Pre- and post-calibration uncertainty standard deviation of selected snow predictions. The SWE predictions are at COR.

Prediction	Pre-calibration uncertainty standard deviation	Post-calibration uncertainty standard deviation
SWE 01/04/ 2015	219.09 (mm)	57.87 (mm)
SWE 01/04/ 2016	200.81 (mm)	57.82 (mm)
SWE 01/04/ 2017	628.54 (mm)	151.63 (mm)
SWE 01/04/ 2018	320.29 (mm)	82.13 (mm)
F_1 22/05/ 2017	0.0537 (-)	0.0365 (-)

season progresses. This is consistent with increasing groundwater storage/saturation throughout the period.

In subsequent work, the gridded estimates generated here have been combined with a specifically developed 3D model of bedrock geology
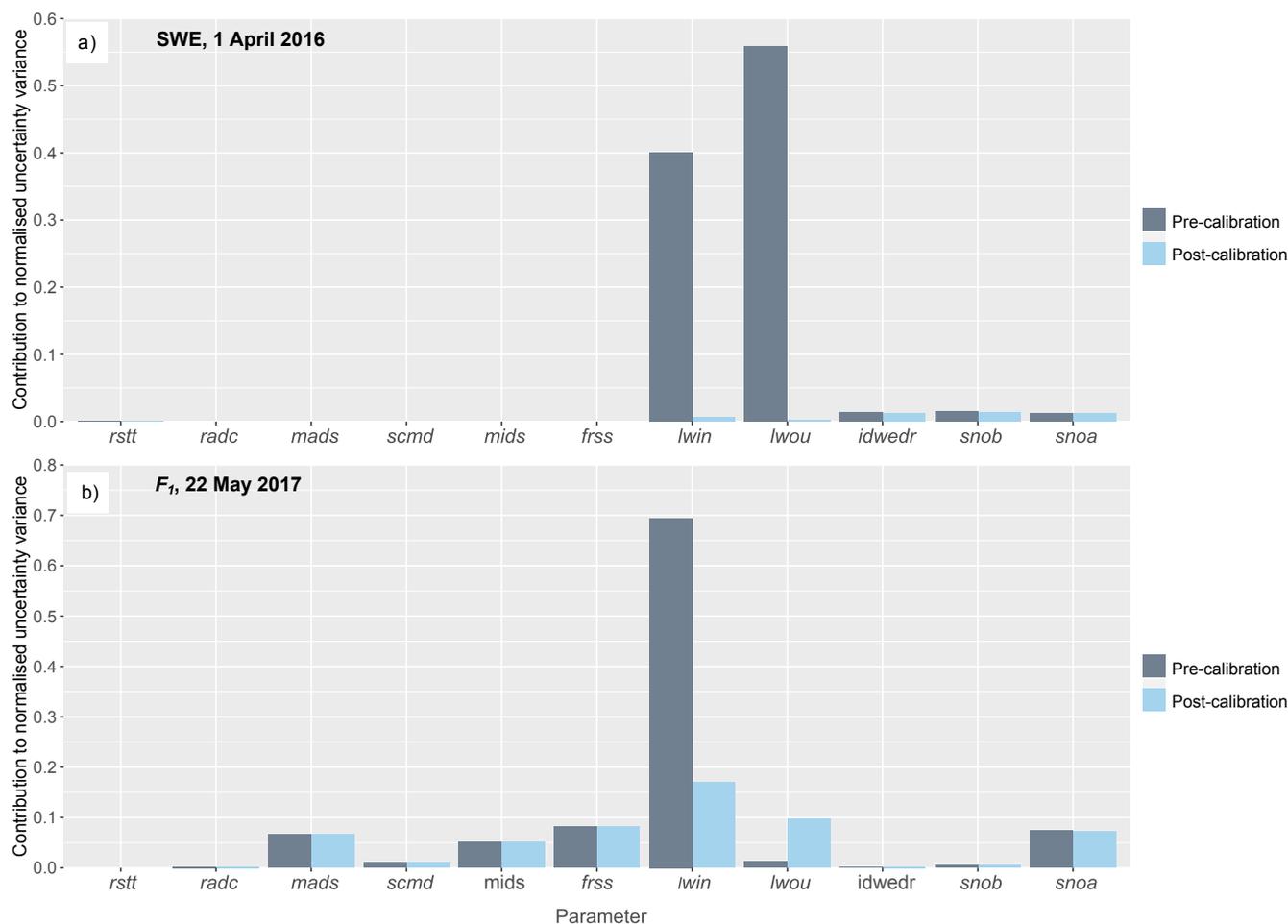
Fig. 8. Parameter contributions to predictive uncertainty variance pre- and post- calibration for the predictions of a) Snow Water Equivalent on 1 April 2016, and b) snow pattern (summarised by F_1) on 22 May 2017.

(Thornton et al., 2018) and other data to inform a sophisticated, fully-integrated surface–subsurface hydrological model (Thornton et al., under review). The generally good simulated streamflow and groundwater level results presented therein further reinforce the hydrological plausibly of our snow simulations.

5.5. Predictive uncertainty and data worth analyses

Table 4 shows the estimated pre- and post-calibration uncertainty standard deviation of the selected predictions. The calibration process substantially reduced the uncertainty associated with the predictions (by a factor of approximately four).

Fig. 8 provides an indication of contribution of the different model parameters to the uncertainty variance, both pre- and post-calibration, for two of the five predictions. In these plots, the uncertainty variance contributions have been normalised with respect to the pre-calibration uncertainty variance associated with the respective predictions.

Fig. 8(a), which concerns the prediction of SWE on 1 April 2016, reveals firstly that many parameters that either do not (or hardly) contribute to predictive uncertainty either before or after calibration, i.e. the prediction is insensitive to these parameters. A slight reduction in the contribution to uncertainty variance can however be observed for *idwedr, snob,* and *snoa.* The presence of many parameters that do not contribute to either the pre- or post-calibration uncertainty in the SWE prediction is to be expected. This is because most of these parameters concern the model's gravitational redistribution component, whereas the COR measurement station will have undoubtedly been sited strategically such that the measurements are generally unaffected by such processes. Another striking feature of this plot

is the large reduction in the predictive uncertainty associated with the longwave correction parameters, *lwin* and *lwout*, that calibration induced. The results for the three other 1 April SWE predictions were similar, and so are not presented in the interests of space.

For the prediction of spatial snow extent on the 22 May 2017 quantified according to the F_1 statistic (Fig. 8(b)), practically all parameters make some discernible contribution to uncertainty variance both pre-and post-calibration. Interestingly, as with the SWE prediction, a large reduction in the post-calibration uncertainty associated with *lwin* is observed, but the post-calibration uncertainty associated with *lwou* in relation to this prediction is actually higher than the pre-calibration value. This counterintuitive situation can occasionally arise when a parameter to which the prediction is insensitive can only be made in conjunction with another parameter to which the prediction is indeed sensitive; see Doherty (2010) for further explanation. In this instance, it may be because these two parameters are not entirely independent of one another. For all other parameters, the uncertainty contribution post-calibration is very similar to the pre-calibration level, suggesting a certain insensitivity of the simulated snow extents to varying these parameter values. The "robustness" that the similarity between pre- and post-calibration parameter contributions to uncertainty in the spatial prediction indicates could, in fact, be particularly beneficial in applications where only spatial snow patterns, as opposed to water volumes, are of primary importance (e.g. assessing the influence of snow patterns on vegetation).

Fig. 9 provides two alternative representations of the worth of the observations belonging to the five different groups in the calibration process. Fig. 9(a) shows the increase in post-calibration predictive
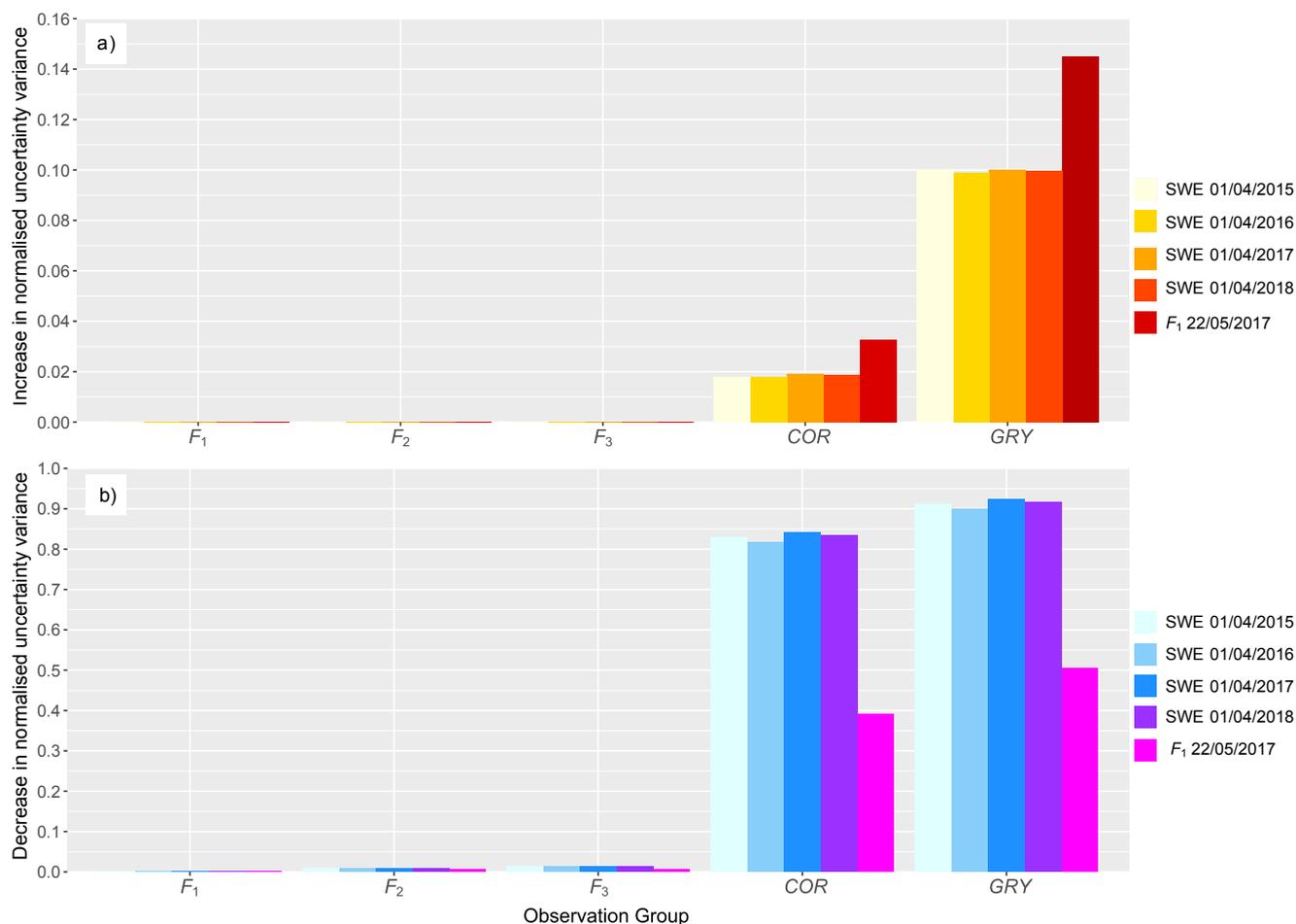
Fig. 9. a) Increase in relative (to the pre-calibration uncertainty variance) post-calibration predictive uncertainty variance associated with each of the five selected predictions incurred by omitting each observation group from the calibration dataset in turn, and b) decrease in relative (again to the pre-calibration uncertainty variance) uncertainty variance accrued when each observation group comprises the sole member of the calibration dataset. The "redundancy", or commonality of information, between the three spatial observation groups (i.e. F_1, F_2, and F_3) is clearly apparent.

uncertainty variance associated with each of the five selected predictions – again relative to pre-calibration uncertainty variance – that is incurred by omitting each observation group from the calibration dataset in turn. Removing either F_1, F_2, or F_3 has very little adverse effect on any of the predictions. This may be explained by the fact that when one of these groups is omitted, similar information is retained in other two groups. The comparatively small number of observations in these groups, coupled with the uniform weighing applied to all observations for the purposes of this analysis, could also partially explain these results. Combining these observations into a single group, and/or adjusting the relative weights assigned, could have suggested enhanced importance of these observations.

This plot furthermore reveals the notable contribution that both time-series, but most especially that at GRY, make to the prediction; the prediction's uncertainty variance increases markedly if either of these observation groups is removed. Four of the five predictions under consideration correspond to the SWE predicted at the high elevation COR station. In light of this, the analysis suggests, perhaps surprisingly, that removing the data at GRY from the calibration dataset has a more detrimental effect than removing the other observations at COR itself (the very location of the predictions).

Fig. 9(b) provides an indication of the decrease in uncertainty variance accrued relative to the pre-calibration uncertainty variance when each observation group comprises the sole member of the calibration dataset. Including any one of the observation groups F_1, F_2, or F_3 alone in the calibration datasets leads to only modest reductions in the pre-

calibration uncertainty variance associated with the predictions (although this is not to say that they do not have a more pronounced effect in combination). In contrast, including either of COR or GRY SWE time-series observations as the sole calibration dataset leads to similarly large reductions in the uncertainty associated with the prediction of SWE at COR. The uncertainty around the F_1 prediction on 22 May 2017 is also greatly reduced by including either of these groups as the sole calibration dataset, although only by about half as much as the reduction seen for the 1 April SWE predictions. This result, that even including only one of the time-series as the sole calibration dataset substantially reduces the uncertainty in the spatial prediction, indicates an important flow of information from the time-series to the predicted spatial snow patterns, and can probably be generalised to the other days on which simulated spatial patterns were compared with observations.

The apparent significance of the GRY data that both analyses indicate is especially notable given that the number of observations at this site is substantially lower than at GRY (due to lower measurement frequency). It could be that, being straddled more frequently by the 0°C isotherm, the SWE time-series at GRY contains more important information about temperature (and therefore temperature gradients) and snow limits than the COR data, which contains distinct accumulation and ablation phases are apparent.

More generally, the notable contributions that the time-series data seem to make demonstrates the importance of obtaining various complementary data types and employing them within multi-objective approaches. This result is consistent with the Tuo et al. (2018), for

instance, who also showed that SWE data can be included to good effect in the hydrological model calibration in alpine catchments.

Lastly, because a consideration in this study was to generate the best possible inputs for subsequent hydrological modelling (that also coincide in time with other measurements from the study region, such as groundwater levels; not shown), no snow observations were specifically withheld for evaluation. Future research should certainly explore the influence of the chosen calibration period and/or assess model's performance under different condition. Although the uncertainty analysis can be considered a partial replacement, traditional split-sample model evaluation should be undertaken whenever sufficiently long time-series are available.

5.6. Potential sources of residual mismatch

Uncertainty in snow observations aside, a large proportion of the residual spatio-temporal mismatch between the simulations and observations can probably be attributed to the meteorological forcing data. Due to the combination of relatively low station density and variable (but sometimes high) data gap frequency (Figures S3 to S9), the interpolated spatial fields of meteorological variables used to drive the model are undoubtedly uncertain.

More specifically, whilst the temporal coverage and therefore crossover of the meteorological data varies throughout the simulation period, the parameters in the model are global. This means that the rain-snow threshold temperature (*rstt*) and solid precipitation correction factors (*snoa* and *snob*), for example, remained constant in time and space. Hence, when estimated though the calibration processes, values producing "best overall" outcomes with respect to the observations were returned. In reality, however, the error distribution associated with precipitation measurements likely varies per station and per event. Smith et al. (2020) reported difficulties in finding undercatch correction factors that perform comparably well across multiple sites.

In other words, it may be that in not permitting "bespoke" per station/event corrections, the model structure is insufficiently flexible to compensate fully for deficiencies in meteorological measurements during certain periods and/or at certain locations. In this sense, improved fits could perhaps have been achieved by simply scaling relatively complete time-series measured at (an) individual location(s) using linear, elevation-dependent relationships, although such an approach would have been less satisfactorily during periods with high meteorological data availability since much of it would have essentially been discarded. In addition, the spatial dependencies in meteorological variables away from the station locations, which the interpolation processes attempted to characterise, probably also demonstrate some temporal non-stationarity in reality (e.g. spatial structures may differ depending on whether precipitation is frontal or convective).

5.7. Some remarks on wind redistribution

Spatially distributed drift/solid precipitation correction factors have been applied to account for the influence of wind transport processes on snowpack heterogeneity (e.g. Hanzer et al., 2016; Marshall et al., 2019), including based on LiDAR-derived snow depth maps where available (Vögeli et al., 2016). However, as mentioned earlier, after testing, WaSiM's wind redistribution algorithm was not included in our final model because doing so led to poorer observation fits.

The algorithm in question computes a temporally invariant, spatially-distributed correction factor grid that acts as a multiplier to the interpolated precipitation fields, such that precipitation falling on predominantly sheltered slopes and on the leeward side of ridges is augmented (deposited), whilst that falling on exposed slopes is reduced (scoured). To generate such a grid, a single prevailing wind direction (in fact, a sector) must be prescribed. However, an analysis of the relationship between high elevation winter wind speeds (these conditions being those under which snow redistribution by wind is most relevant)

and directions revealed that no such single prevailing wind direction can be identified across the study area (Fig. S13); strong winter winds can apparently originate from contrasting directions, probably according to larger-scale synoptic meteorology. Some influence of the complex local topography on wind patterns is also evident. Similar patterns could be expected in other mountainous regions.

Furthermore, the calculated wind redistribution factor range seemed somewhat high, leading to both too much "deposition" and "scouring", depending upon pixel exposition, with respect to the (admittedly limited) SWE data. Another potential issue is that unlike the gravitational redistribution algorithm applied, the wind redistribution approach does not conserve mass within a given area (e.g. a catchment). Ultimately, it may be that in mountainous regions where large-scale meteorological phenomena interacting with extremely complex topography give rise to considerable spatio-temporal variability in near-surface wind fields, such comparatively simple algorithms are unable to match highly resolved, site-specific snow observations.

Thus, the development of extended empirical approaches that could include directly measured high-elevation wind directions and/or are calibrated explicitly to observed redistribution magnitudes near ridges could form an appropriate intermediate objective, until physics-based wind-induced snow transport can be simulated in a physically-based fashion at high resolution across entire catchments. In any case, in this particular study area, snow redistribution by wind is likely of secondary hydrological importance to that by gravity.

5.8. Ongoing debates regarding hydrological model calibration

Whilst it is increasingly clear that internal states should be verified at some stage in the calibration or evaluation of hydrological models, the most appropriate approach for including snow data remains under debate. On the one hand, it has been argued that since the volumetric information contained within discharge measurements complements the internal spatial pattern information embedded in distributed snow observations (Finger et al., 2011), these two observation types should be considered simultaneously (e.g. Finger et al., 2011; 2014; Shrestha et al., 2014). The argument runs that the discharge constraint can help ensure that, in aggregate, the total simulated system water volume is approximately correct – which is important given uncertain precipitation measurements and gridded products – whilst the snow pattern constraints help ensure that runoff is being generated in the right areas.

Others, however, posit that calibration is best tackled more sequentially, whereby snow simulations are initially optimised independently before one proceeds to simulate "intermediate" hydrological variables and, ultimately, discharge. The principal argument in favour of this approach is that simultaneous calibration may enable error compensation (Ragettli & Pellicciotti, 2012; Magnusson et al., 2015) that can be hidden or easily overlooked, at least unless extremely careful evaluative work with respect to observed spatial patterns in undertaken (which remains rare).

Provided a given set of snow observations are sufficiently informative (and it is hoped those employed here are, despite being somewhat limited by practical considerations, such as the risks associated with conducting regular winter snow surveys in such terrain), then the latter approach, which was taken here, should automatically ensure that the total water volumes are reasonably accurate. Indeed, both binary spatial and volumetric time-series at contrasting sites were considered in the snow model calibration for this precise reason. The hydrological plausibly assessment undertaken (Section 5.4) and subsequent work (Thornton et al., under review) provides reassurance that this is indeed the case. Splitting the calibration phases essentially reduced the potential for model parameters related to the surface or subsurface to compensate for poor snow simulations (or vice versa), and additionally enabled a more advanced, fully integrated code to be applied for the simulation of the remaining hydrological processes (Thornton et al., under review).

6. Conclusions

This paper has presented and exemplified a novel, computationally efficient approach for the calibration and uncertainty analysis of distributed snow models in steep, rugged alpine terrain. The physically-based core of the model explicitly captures the spatio-temporal variability in energy balance components, which is largely responsible for heterogeneous snow patterns and therefore melt rates in such terrain. Substantial uncertainties related to biased solid precipitation measurements and other observational deficiencies were addressed using empirical correction factors. Gravitational redistribution can also substantially influence meltwater patterns in steep regions, but cannot presently be represented on an entirely physical basis at catchment or larger scales. A pragmatic empirical algorithm was therefore also included to represent this process.

Reliably simulating the spatio-temporal evolution of SWE thus hinged upon the estimation of several parameters, which was not trivial given the reasonable computational expense associated with the forward model; a single parallelized simulation took several hours. To this end, a novel multi-objective calibration strategy that involved a gradient-based algorithm and incorporated two complementary types of high-resolution snow observations – snow extent maps and SWE time-series – was developed. The study represents one of the first such occasions on which a distributed snow model has actually been calibrated (rather than merely evaluated) according to an objective function that is partially spatially explicit, whereby mismatches are penalised at the pixel level. This was achieved by quantifying model performance with respect to Landsat-derived observed snow maps using so-called F-statistics. Substantial corrections to measured winter precipitation totals were required to minimise model-data mismatches.

Following calibration, spatio-temporal snow dynamics could be satisfactorily reproduced with respect to the available observations, and the spatial fit metrics obtained moreover compared favourably with the few equivalent statistics reported previously. Subsequent uncertainty and data worth analyses indicated that: i) the uncertainty variance of indicative predictions of snow states, both spatial and volumetric, were substantially reduced through calibration, ii) including two parameters that enable the longwave component of the surface energy balance to be adjusted, and thus potential errors in cloudiness and albedo compensated for, was especially beneficial, and iii) the SWE time-series at the lower elevation station (GRY) was particularly informative, despite the comparatively small number of observations at this site. Any snow observation uncertainties notwithstanding, much of the residual mismatch between simulations and observations is likely associated with the meteorological forcing data; both the raw measurements (especially for precipitation) and interpolated spatial fields. As such, efforts to better characterise and account for the uncertainties and biases in mountain meteorological measurements and interpolated/downscaled gridded products should be prioritised.

The generic model-data integration framework presented extends well beyond standard treatments of snow in hydrological modelling, and could be easily applied in different settings. In addition, in our view, the specific application of it presented here achieved an appropriate balance between model complexity and data availability. Nevertheless, the implications of some potential limitations associated with the "core" of the model employed here could be investigated. For instance, fairly basic spatial interpolation schemes were used, a single rather than multi-layer snow model was employed, and wind redistribution was not accounted for. Indeed, acknowledging that even sophisticated multi-physics snow models contain uncertain parameters that must be identified (see also Günther et al., 2020), future work should explore the extent to which using even more sophisticated snow simulation approaches might further reduce both pre- and post-calibration predictive uncertainties. Specifically, such work could involve testing more sophisticated interpolation algorithms (e.g. that of Liston and Elder (2006) for wind), deploying a more process-rich and advanced snow code that includes at least three snow layers, and/or accounts for wind redistribution. Provided increased data availability and quality accompanies increased model complexity, a further reduction in uncertainty should be realised, although this remains to be tested. However, the outcomes of such experiments must be interpreted with respect to the magnitudes of "irreducible" uncertainties associated with both forcing and observed calibration data, i.e. uncertainties that will persist under many circumstances.

In summary, distributed meltwater datasets generated via calibration-constrained simulations of snow dynamics demonstrate great potential to inform the next generation of comprehensive, physically-based, spatially explicit hydrological simulations in complex alpine terrain. Such efforts are urgently needed to provide a sound basis for decision making under hydrological system change. A range of other applications which also require spatio-temporally comprehensive snow information could also benefit from such an approach. Employing model-data integration frameworks such as that presented here more routinely, perhaps eventually in conjunction with even more advanced snow codes (where data availability makes it appropriate to do), will lead to improved snow simulations but also greater appreciation of their (currently somewhat overlooked) outstanding uncertainties.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://dx.doi.org/10.17632/84ph9rpnhr.1.

References

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., Seibert, J., 2014. Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss

catchments. Water Resour. Res. 50 (10), 7541–7562. https://doi.org/10.1002/2014WR015549.

Aronica, G., Bates, P.D., Horritt, M.S., 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrol. Process. 16 (10), 2001–2016. https://doi.org/10.1002/(ISSN)1099-108510.1002/hyp.v16:1010.1002/hyp.398.

Barnett, T.P., Adam, J.C., Lettenmaier, D.P., 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. Nature 438 (7066), 303–309. https://doi.org/10.1038/nature04141.

Bavera, D., Bavay, M., Jonas, T., Lehning, M., De Michele, C., 2014. A comparison between two statistical and a physically-based model in snow water equivalent mapping. Adv. Water Resour. 63, 167–178. https://doi.org/10.1016/j.advwatres.2013.11.011.

Behrangi, A., Bormann, K.J., Painter, T.H., 2018. Using the Airborne Snow Observatory to Assess Remotely Sensed Snowfall Products in the California Sierra Nevada. Water Resour. Res. 54 (10), 7331–7346. https://doi.org/10.1029/2018WR023108.

Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N.,… Vincent, C. (2018). The European mountain cryosphere: a review of its current state, trends, and future challenges. The Cryosphere, 12 (2), 759–794. doi: 10.5194/tc-12-759-2018.

Benoit, L., Allard, D., & Mariethoz, G. (2018, 6). Stochastic rainfall modeling at sub-kilometer scale. Water Resources Research, 54 (6), 4108–4130. doi: 10.1029/2018WR022817.

Berezowski, Tomasz, Chormański, Jarosław, Batelaan, Okke, 2015. Skill of remote sensing snow products for distributed runoff prediction. J. Hydrol. 524, 718–732. https://doi.org/10.1016/j.jhydrol.2015.03.025.

Bernhardt, M., Schulz, K., Liston, G.E., Zängl, G., 2012. The influence of lateral snow redistribution processes on snow melt and sublimation in alpine regions. J. Hydrol. 424–425, 196–206. https://doi.org/10.1016/j.jhydrol.2012.01.001.

Bini, A., Buonchristiani, J., Couterand, S., Ellwanger, D., Felber, M., Florineth, D.,… Schoeneich, P. (2009). Switzerland during the Last Glacial Maximum (LGM), 1: 500,000. Federal Office of Topography, swisstopo.

Bolch, T., Kulkarni, A., Kaab, A., Huggel, C., Paul, F., Cogley, J.G., Frey, H., Kargel, J.S., Fujita, K., Scheel, M., Bajracharya, S., Stoffel, M., 2012. The State and Fate of Himalayan Glaciers. Science 336 (6079), 310–314. https://doi.org/10.1126/science:1215342.

Bormann, K.J., Brown, R.D., Derksen, C., Painter, T.H., 2018. Estimating snow-cover trends from space. Nat. Clim. Change 8 (11), 924–928. https://doi.org/10.1038/s41558-018-0318-3.

Brauchli, T., Beria, H., Michelon, A., Larsen, J., & Schaefli, B. (2018). Estimating the precipitation in a high-alpine catchment combining local meteo stations and Swiss-wide meteo products. In Swiss Geoscience Meeting 2018. Bern.

Brauchli, T., Trujillo, E., Huwald, H., & Lehning, M. (2017, 12). Influence of slope-scale snowmelt on catchment response simulated with the Alpine3D model. Water Resources Research, 53 (12), 10723–10739. doi: 10.1002/2017WR021278.

Bründl, M., Hafner, E., Bebi, P., Bühler, Y., Margreth, S., Marty, C.,…Schweizer, J. (2019). Analyse des évènements de la situation avalancheuse de janvier 2018 (Tech. Rep.). WSL Institut pour l'étude de la neige et des avalanches SLF, Davos. https://www.wsl.ch/de/publikationen/analyse-des-evenements-de-la-situation-avalancheuse-de-janvier-2018.html.

Bühler, Y., Hafner, E.D., Zweifel, B., Zesiger, M., Heisig, H., 2019. Where are the avalanches? Rapid SPOT6 satellite data acquisition to map an extreme avalanche period over the Swiss Alps. The Cryosphere 13 (12), 3225–3238. https://doi.org/10.5194/tc-13-3225-2019.

Casassa, G., López, P., Pouyaud, B., & Escobar, F. (2009). Detection of changes in glacial run-off in alpine basins: examples from North America, the Alps, central Asia and the Andes. Hydrological Processes, 23 (1), 31–41. doi: 10.1002/hyp.7194.

Ceperley, N., Michelon, N., Escoffier, A. Mayoraz G., Boix Canadell, M., Horgby, A., Hammer, F.,…Boss, S. (2018). Salt gauging and stage-discharge curve, Avancon de Nant, outlet Vallon de Nant catchment. doi: 10.5281/zenodo.1154798.

Clark, Martyn P., Hendrikx, Jordy, Slater, Andrew G., Kavetski, Dmitri, Anderson, Brian, Cullen, Nicolas J., Kerr, Tim, Örn Hreinsson, Einar, Woods, Ross A., 2011. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. Water Resour. Res. 47 (7) https://doi.org/10.1029/2011WR010745.

Clark, M.P., Slater, A.G., Barrett, A.P., Hay, L.E., McCabe, G.J., Rajagopalan, B., Leavesley, G.H., 2006. Assimilation of snow covered area information into hydrologic and land-surface models. Adv. Water Resour. 29 (8), 1209–1221. https://doi.org/10.1016/j.advwatres.2005.10.001.

Cochand, M., Christe, P., Ornstein, P., & Hunkeler, D. (2019, 4). Groundwater storage in high Alpine catchments and its contribution to streamflow. Water Resources Research, 55 (4), 2613–2630. doi: 10.1029/2018WR022989.

Costa, Anna, Molnar, Peter, Stutenbecker, Laura, Bakker, Maarten, Silva, Tiago A., Schlunegger, Fritz, Lane, Stuart N., Loizeau, Jean-Luc, Girardclos, Stéphanie, 2018. Temperature signal in suspended sediment export from an Alpine catchment. Hydrol. Earth Syst. Sci. 22 (1), 509–528. https://doi.org/10.5194/hess-22-509-2018.

Dierauer, J.R., Whitfield, P.H., Allen, D.M., 2018. Climate controls on runoff and low flows in mountain catchments of western North America. Water Resour. Res. 54 (10), 7495–7510. https://doi.org/10.1029/2018WR023087.

Doherty, J. (2010). Methodologies and Software for PEST-Based Model Predictive Uncertainty Analysis (Tech. Rep.). Watermark Numerical Computing.

Doherty, J., 2019. PEST: Model-Independent Parameter Estimation, 7th Edition. Watermark Numerical Computing, Brisbane, Australia.

Dozier, J., 1989. Spectral signature of alpine snow cover from the landsat thematic mapper. Remote Sens. Environ. 28 (C), 9–22. https://doi.org/10.1016/0034-4257(89)90101-6.

Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., Güntner, A., 2014. The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. Water Resour. Res. 50 (3), 2002–2021. https://doi.org/10.1002/2013WR014382.

Engel, Michael, Notarnicola, Claudia, Endrizzi, Stefano, Bertoldi, Giacomo, 2017. Snow model sensitivity analysis to understand spatial and temporal snow dynamics in a high-elevation catchment. Hydrol. Process. 31 (23), 4151–4168. https://doi.org/10.1002/hyp.v31.2310.1002/hyp.11314.

Etter, S., Addor, N., Huss, M., Finger, D., 2017. Climate change impacts on future snow, ice and rain runoff in a Swiss mountain catchment using multi-dataset calibration. J. Hydrol.: Reg. Stud. 13, 222–239. https://doi.org/10.1016/j.ejrh.2017.08.005.

Fatichi, S., Rimkus, S., Burlando, P., Bordoy, R., Molnar, P., 2015. High- resolution distributed analysis of climate and anthropogenic changes on the hydrology of an Alpine catchment. J. Hydrol. 525, 362–382. https://doi.org/10.1016/j.jhydrol.2015.03.036.

Finger, D., Pellicciotti, F., Konz, M., Rimkus, S., Burlando, P., 2011. The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model. Water Resour. Res. 47 (7), 1–14. https://doi.org/10.1029/2010WR009824.

Förster, K., Garvelmann, J., Meißl, G., & Strasser, U. (2018). Modelling forest snow processes with a new version of WaSiM. Hydrological Sciences Journal, 63 (10), 1540–1557. Retrieved from https://doi.org/10.1080/.

Franz, K.J., Butcher, P., Ajami, N.K., 2010. Addressing snow model uncertainty for hydrologic prediction. Adv. Water Resour. 33 (8), 820–832. https://doi.org/10.1016/j.advwatres.2010.05.004.

Freudiger, D., Kohn, I., Seibert, J., Stahl, K., Weiler, M., 2017. Snow redistribution for the hydrological modeling of alpine catchments. Wiley Interdisciplinary Reviews: Water 4 (5), 1232. https://doi.org/10.1002/wat2.1232.

Fuhrer, J., Jasper, K., 2012. Demand and supply of water for agriculture: Influence of topography and climate in pre-Alpine, mesoscale catchments. Natural Resources 03 (3), 145–155. https://doi.org/10.4236/nr.2012.33019.

Gallice, A., Bavay, M., Brauchli, T., Comola, F., Lehning, M., Huwald, H., 2016. StreamFlow 1.0: An extension to the spatially distributed snow model Alpine3D for hydrological modelling and deterministic stream temperature prediction. Geosci. Model Dev. 9 (12), 4491–4519. https://doi.org/10.5194/gmd-9-4491-2016.

GDAL (2019). GDAL. https://gdal.org/. Last accessed April 2019.

Giaccone, Elisa, Luoto, Miska, Vittoz, Pascal, Guisan, Antoine, Mariéthoz, Grégoire, Lambiel, Christophe, 2019. Influence of microclimate and geomorphological factors on alpine vegetation in the Western Swiss Alps. Earth Surf. Proc. Land. 44 (15), 3093–3107. https://doi.org/10.1002/esp.v44.1510.1002/esp.4715.

Grand, Stephanie, Rubin, Aurélie, Verrecchia, Eric P., Vittoz, Pascal, Araujo, Wagner L., 2016. Variation in soil respiration across soil and vegetation types in an Alpine valley. PLoS ONE 11 (9), e0163968. https://doi.org/10.1371/journal.pone.0163968.

Grünewald, T., Schirmer, M., Mott, R., Lehning, M., 2010. Spatial and temporal variability of snow depth and ablation rates in a small mountain catchment. The Cryosphere 4 (2), 215–225. https://doi.org/10.5194/tc-4-215-2010.

Günther, D., Hanzer, F., Warscher, M., Essery, R., Strasser, U., 2020. Including Parameter Uncertainty in an Intercomparison of Physically-Based Snow Models. Front. Earth Sci 8, 542599. https://doi.org/10.3389/feart.2020.542599.

Hanzer, F., Helfricht, K., Marke, T., Strasser, U., 2016. Multilevel spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments. The Cryosphere 10 (4), 1859–1881. https://doi.org/10.5194/tc-10-1859-2016.

Härer, S., Bernhardt, M., Siebers, M., & Schulz, K. (2018). On the need for a time- and location-dependent estimation of the NDSI threshold value for reducing existing uncertainties in snow cover maps at different scales. The Cryosphere, 12 (5), 1629–1642. doi: 10.5194/tc-12-1629-2018.

He, M., Hogue, T.S., Franz, K.J., Margulis, S.A., Vrugt, J.A., 2011. Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes. Adv. Water Resour. 34 (1), 114–127. https://doi.org/10.1016/j.advwatres.2010.10.002.

Hedrick, Andrew R., Marks, Danny, Havens, Scott, Robertson, Mark, Johnson, Micah, Sandusky, Micah, Marshall, Hans-Peter, Kormos, Patrick R., Bormann, Kat J., Painter, Thomas H., 2018. Direct Insertion of NASA Airborne Snow Observatory-Derived Snow Depth Time Series Into the iSnobal Energy Balance Snow Model. Water Resour. Res. 54 (10), 8045–8063. https://doi.org/10.1029/2018WR023190.

Herrnegger, M., Nachtnebel, H.-P., Haiden, T., 2012. Evapotranspiration in high alpine catchments – an important part of the water balance! Hydrol. Res. 43 (4), 460. https://doi.org/10.2166/nh.2012.132.

Hock, R., 2003. Temperature index melt modelling in mountain areas. J. Hydrol. 282 (1–4), 104–115. https://doi.org/10.1016/S0022-1694(03)00257-9.

Huss, M., Bookhagen, B., Huggel, C., Jacobsen, D., Bradley, R., Clague, J.,…Winder, M. (2017). Toward mountains without permanent snow and ice. Earth's Future, 5 (May), 418–435. doi: 10.1002/eft2.207.

Huss, M., & Hock, R. (2018, 2). Global-scale hydrological response to future glacier mass loss. Nature Climate Change, 8 (2), 135–140. doi: 10.1038/s41558-017-0049-x.

Jenicek, M., Seibert, J., Zappa, M., Staudinger, M., Jonas, T., 2016. Importance of maximum snow accumulation for summer low flows in humid catchments. Hydrol. Earth Syst. Sci. 20 (2), 859–874. https://doi.org/10.5194/hess-20-859-2016.

Jennings, K.S., Winchell, T.S., Livneh, B., Molotch, N.P., 2018. Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere. Nat. Commun. 9 (1), 1–9. https://doi.org/10.1038/s41467-018-03629-7.

Jimeno-Sáez, P., Pulido-Velazquez, D., Collados-Lara, A.J., Pardo-Igúzquiza, E., Senent-Aparicio, J., Baena-Ruiz, L., 2020. A Preliminary Assessment of the "Undercatching" and the Precipitation Pattern in an Alpine Basin. Water 12 (4), 1061. https://doi.org/10.3390/w12041061.

Jonas, T., Marty, C., Magnusson, J., 2009. Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. J. Hydrol. 378 (1–2), 161–167. https://doi.org/10.1016/j.jhydrol.2009.09.021.

Kerr, T., Clark, M., Hendrikx, J., Anderson, B., 2013. Snow distribution in a steep mid-latitude alpine catchment. Adv. Water Resour. 55, 17–24. https://doi.org/10.1016/j.advwatres.2012.12.010.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resour. Res. 42 (3) https://doi.org/10.1029/2005WR004362.

Klein, G., Vitasse, Y., Rixen, C., Marty, C., Rebetez, M., 2016. Shorter snow cover duration since 1970 in the Swiss Alps due to earlier snowmelt more than to later snow onset. Clim. Change 139 (3–4), 637–649. https://doi.org/10.1007/s10584-016-1806-y.

Kochendorfer, John, Nitu, Rodica, Wolff, Mareile, Mekis, Eva, Rasmussen, Roy, Baker, Bruce, Earle, Michael E., Reverdin, Audrey, Wong, Kai, Smith, Craig D., Yang, Daqing, Roulet, Yves-Alain, Buisan, Samuel, Laine, Timo, Lee, Gyuwon, Aceituno, Jose Luis C., Alastrué, Javier, Isaksen, Ketil, Meyers, Tilden, Brækkan, Ragnar, Landolt, Scott, Jachcik, Al, Poikonen, Antti, 2017. Analysis of single-Alter-shielded and unshielded measurements of mixed and solid precipitation from WMO-SPICE. Hydrol. Earth Syst. Sci. 21 (7), 3525–3542. https://doi.org/10.5194/hess-21-3525-2017.

Lane, S.N., Borgeaud, L., Vittoz, P., 2016. Emergent geomorphic-vegetation interactions on a subalpine alluvial fan. Earth Surf. Proc. Land. 41 (1), 72–86. https://doi.org/10.1002/esp.3833.

Lane, S.N., Nienow, P.W., 2019. Decadal-scale climate forcing of Alpine glacial hydrological Systems. Water Resour. Res. 55 (3), 2478–2492. https://doi.org/10.1029/2018WR024206.

Lehning, M., Bartelt, P., Brown, B., Fierz, C., 2002. A physical SNOWPACK model for the Swiss avalanche warning Part III: Meteorological forcing, thin layer formation and evaluation. Cold Reg. Sci. Technol. 35 (3), 169–184. https://doi.org/10.1016/S0165-232X(02)00072-1.

Lehning, Michael, Völksch, Ingo, Gustafsson, David, Nguyen, Tuan Anh, Stähli, Manfred, Zappa, Massimiliano, 2006. ALPINE3D: A detailed model of mountain surface processes and its application to snow hydrology. Hydrol. Process. 20 (10), 2111–2128. https://doi.org/10.1002/hyp.6204.

Leonarduzzi, E., Molnar, P., McArdell, B.W., 2017. Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. Water Resour. Res. 53 (8), 6612–6625. https://doi.org/10.1002/2017WR021044.

Liston, G.E., Elder, K., 2006. A meteorological distribution system for high-resolution terrestrial modeling (MicroMet). J. Hydrometeorol. 7 (2), 217–234. https://doi.org/10.1175/JHM486.1.

Magnusson, Jan, Wever, Nander, Essery, Richard, Helbig, Nora, Winstral, Adam, Jonas, Tobias, 2015. Evaluating snow models with varying process representations for hydrological applications. Water Resour. Res. 51 (4), 2707–2723. https://doi.org/10.1002/2014WR016498.

Marshall, A.M., Link, T.E., Abatzoglou, J.T., Flerchinger, G.N., Marks, D.G., Tedrow, L., 2019. Warming Alters Hydrologic Heterogeneity: Simulated Climate Sensitivity of Hydrology-Based Microrefugia in the Snow-to-Rain Transition Zone. Water Resour. Res. 55 (3), 2122–2141. https://doi.org/10.1029/2018WR023063.

Mas, A., Baraer, M., Arsenault, R., Poulin, A., Préfontaine, J., 2018. Targeting high robustness in snowpack modeling for Nordic hydrological applications in limited data conditions. J. Hydrol. 564, 1008–1021. https://doi.org/10.1016/j.jhydrol.2018.07.071.

McGrath, Daniel, Webb, Ryan, Shean, David, Bonnell, Randall, Marshall, Hans-Peter, Painter, Thomas H., Molotch, Noah P., Elder, Kelly, Hiemstra, Christopher, Brucker, Ludovic, 2019. Spatially Extensive Ground-Penetrating Radar Snow Depth Observations During NASA's 2017 SnowEx Campaign: Comparison With In Situ, Airborne, and Satellite Observations. Water Resour. Res. 55 (11), 10026–10036. https://doi.org/10.1029/2019WR024907.

Meeks, J., Moeck, C., Brunner, P., Hunkeler, D., 2017. Infiltration under snow cover: Modeling approaches and predictive uncertainty. J. Hydrol. 546, 16–27. https://doi.org/10.1016/j.jhydrol.2016.12.042.

MeteoSwiss. (2019). IDAWEB. https://gate.meteoswiss.ch/idaweb/. Last accessed June 2019.

Micheletti, N., Lane, S.N., 2016. Water yield and sediment export in small, partially glaciated Alpine watersheds in a warming climate. Water Resour. Res. 52 (6), 4924–4943. https://doi.org/10.1002/2016WR018774.

Michelon, A., Schaefli, B., Ceperley, N. C., & Beria, H. (2017). Weather dataset from Vallon de Nant, Switzerland, until July 2017. doi: 10.5281/ zenodo.1042472.

Mott, R., Lehning, M., 2010. Meteorological modeling of very high-resolution wind fields and snow deposition for mountains. J. Hydrometeorol. 11 (4), 934–949. https://doi.org/10.1175/2010JHM1216.1.

Mott, R., Scipíon, D., Schneebeli, M., Dawes, N., Berne, A., & Lehning, M. (2014). Orographic effects on snow deposition patterns. Journal of Geophysical Research: Atmospheres, 119, 1419–1439. doi: 10.1002/2013JD019880.

Mott, R., Wolf, A., Kehl, M., Kunstmann, H., Warscher, M., Grünewald, T., 2019. Avalanches and micrometeorology driving mass and energy balance of the lowest perennial ice field of the Alps: a case study. The Cryosphere 13, 1247–1265. https://doi.org/10.5194/tc-13-1247-2019.

Musselman, K.N., Pomeroy, J.W., Essery, R.L.H., Leroux, N., 2015. Impact of windflow calculations on simulations of alpine snow accumulation, redistribution and ablation. Hydrol. Process. 29 (18), 3983–3999. https://doi.org/10.1002/hyp.10595.

Mutzner, R., Weijs, S.V., Tarolli, P., Calaf, M., Oldroyd, H.J., Parlange, M.B., 2015. Controls on the diurnal streamflow cycles in two subbasins of an alpine headwater catchment. Water Resour. Res. 51 (5), 3403–3418. https://doi.org/10.1002/2014WR016581.

Naseer, A., Koike, T., Mohamad, R., Ushiyama, T., & Shrestha, M. (2019). Distributed hydrological modelling framework for quantitative and spatial bias correction for rainfall, snowfall, and mixed-phase precipitation using Vertical Profile of Temperature. Journal of Geophysical Research: Atmospheres, 2018JD029811. doi: 10.1029/2018JD029811.

Oke, T., 1987. Boundary Layer Climates, 2nd Ed. Routledge, London and New York.

OSGeo4W. (2019). OSGeo4W. https://trac.osgeo.org/osgeo4w/. Last accessed April 2019.

Pan, Xicai, Yang, Daqing, Li, Yanping, Barr, Alan, Helgason, Warren, Hayashi, Masaki, Marsh, Philip, Pomeroy, John, Janowicz, Richard J., 2016. Bias corrections of precipitation measurements across experimental sites in different ecoclimatic regions of western Canada. The Cryosphere 10 (5), 2347–2360. https://doi.org/10.5194/tc-10-2347-2016.

Papathoma-Köhle, M., Kappes, M., Keiler, M., Glade, T., 2011. Physical vulnerability assessment for alpine hazards: State of the art and future needs. Nat. Hazards 58 (2), 645–680. https://doi.org/10.1007/s11069-010-9632-4.

Parajka, J., Blöschl, G., 2008. The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models. J. Hydrol. 358 (3–4), 240–258. https://doi.org/10.1016/j.jhydrol.2008.06.006.

Pepin, N., Bradley, R. S., Diaz, H. F., Baraer, M., Caceres, E. B., Forsythe, N.,…Yang, D. Q. (2015). Elevation dependent warming in mountain regions of the world. Nature Climate Change, 5 (5), 424–430. doi: 10.1038/nclimate2563.

Pritchard, H.D., 2019. Asia's shrinking glaciers protect large populations from drought stress. Nature 569 (7758), 649–654. https://doi.org/10.1038/s41586-019-1240-1.

Ragettli, S., Cortés, G., Mcphee, J., Pellicciotti, F., 2014. An evaluation of approaches for modelling hydrological processes in high-elevation, glacierized Andean watersheds. Hydrol. Process. 28 (23), 5674–5695. https://doi.org/10.1002/hyp.10055.

Ragettli, S., Pellicciotti, F., 2012. Calibration of a physically based, spatially distributed hydrological model in a glacierized basin: On the use of knowledge from glaciometeorological processes to constrain model parameters. Water Resour. Res. 48 (3), 1–20. https://doi.org/10.1029/2011WR010559.

Raleigh, M.S., Livneh, B., Lapo, K., Lundquist, J.D., 2016. How does availability of meteorological forcing data impact physically based snowpack simulations? J. Hydrometeorol. 17 (1), 99–120. https://doi.org/10.1175/JHM-D-14-0235.1.

Randin, Christophe F., Dedieu, Jean-Pierre, Zappa, Massimiliano, Long, Li, Dullinger, Stefan, 2015. Validation of and comparison between a semidistributed rainfall-runoff hydrological model (PREVAH) and a spatially distributed snow-evolution model (SnowModel) for snow cover prediction in mountain ecosystems. Ecohydrology 8 (7), 1181–1193. https://doi.org/10.1002/eco.1570.

Rittger, K., Painter, T.H., Dozier, J., 2013. Assessment of methods for mapping snow cover from MODIS. Adv. Water Resour. 51, 367–380. https://doi.org/10.1016/j.advwatres.2012.03.002.

Rössler, O., Froidevaux, P., Börst, U., Rickli, R., Martius, O., Weingartner, R., 2014. Retrospective analysis of a nonforecasted rain-on-snow flood in the Alps-A matter of model limitations or unpredictable nature. Hydrol. Earth Syst. Sci. 18 (6), 2265–2285. https://doi.org/10.5194/hess-18-2265-2014.

Ruelland, D., 2020. Should altitudinal gradients of temperature and precipitation inputs be inferred from key parameters in snow-hydrological models? Hydrol. Earth Syst. Sci. 24 (5), 2609–2632. https://doi.org/10.5194/hess-24-2609-2020.

Schattan, P., Schwaizer, G., Schöber, J., & Achleitner, S. (2020). The complementary value of cosmic-ray neutron sensing and snow covered area products for snow hydrological modelling. Remote Sensing of Environment, 239 (December 2019), 111603. doi: 10.1016/j.rse.2019.111603.

Schöber, J., Achleitner, S., Kirnbauer, R., Schöberl, F., & Schönlaub, H. (2010). Hydrological modelling of glacierized catchments focussing on the validation of simulated snow patterns – applications within the flood forecasting system of the Tyrolean river Inn. Advances in Geosciences, 27 (September 2010), doi: 10.5194/adgeo-27-99-2010.

Schulla, J. (2017). WaSiM (Water balance Simulation Model) Model Description (Tech. Rep.). Zürich: Hydrology Software Consulting J. Schulla. Retrieved from http://www.wasim.ch/downloads/doku/wasim/wasim 2015 en.pdf.

SCNAT. (2018). GLAMOS 1881-2018, The Swiss Glaciers 1880-2016/17. https://doi.glamos.ch/.

Sevruk, B. (1985). Systematischer Niederschlagsmessfehler in der Schweiz. In B. Sevruk (Ed.), Der niederschlag in der schweiz. Beitr. Geol. Schweiz Hydrol., 31, 65–75.

Shrestha, M., Wang, L., Koike, T., Tsutsui, H., Xue, Y., Hirabayashi, Y., 2014. Correcting basin-scale snowfall in a mountainous basin using a distributed snowmelt model and remote-sensing data. Hydrol. Earth Syst. Sci. 18 (2), 747–761. https://doi.org/10.5194/hess-18-747-2014.

Slater, A.G., Barrett, A.P., Clark, M.P., Lundquist, J.D., Raleigh, M.S., 2013. Uncertainty in seasonal snow reconstruction: Relative impacts of model forcing and image availability. Adv. Water Resour. 55, 165–177. https://doi.org/10.1016/j.advwatres.2012.07.006.

SLF (2019a). niVis. https://models.slf.ch/p/niviz/. Last accessed May 2019.

SLF. (2019b). Description of automated stations. https://www.slf.ch/en/avalanche-bulletin-and-snow-situation/measured-values/description-of-automated-stations.html. Last accessed June 2019.

Smith, Craig D., Ross, Amber, Kochendorfer, John, Earle, Michael E., Wolff, Mareile, Buisán, Samuel, Roulet, Yves-Alain, Laine, Timo, 2020. Evaluation of the WMO Solid Precipitation Intercomparison Experiment (SPICE) transfer functions for adjusting the wind bias in solid precipitation measurements. Hydrol. Earth Syst. Sci. 24 (8), 4025–4043. https://doi.org/10.5194/hess-24-4025-2020.

Sommer, C.G., Lehning, M., Mott, R., 2015. Snow in a Very Steep Rock Face: Accumulation and Redistribution During and After a Snowfall Event. Front. Earth Sci. 3 (December), 1–13. https://doi.org/10.3389/feart.2015.00073.

Staudinger, Maria, Stoelzle, Michael, Seeger, Stefan, Seibert, Jan, Weiler, Markus, Stahl, Kerstin, 2017. Catchment water storage variation with elevation. Hydrol. Process. 31 (11), 2000–2015. https://doi.org/10.1002/hyp.v31.1110.1002/hyp.11158.

Strasser, U., Bernhardt, M., Weber, M., Liston, G.E., Mauser, W., 2008. Is snow sublimation important in the alpine water balance? The Cryosphere 2, 53–66. https://doi.org/10.5194/tc-2-53-2008.

Sturm, M., Goldstein, M.A., Parr, C., 2017. Water and life from snow: A trillion dollar science question. Water Resour. Res. 53 (5), 3534–3544. https://doi.org/10.1002/2017WR020840.

swisstopo (2018). DHM25. https://shop.swisstopo.admin.ch/en/products/height models/dhm25. Last accessed October 2018.

swisstopo (2019). Land cover. https://www.geo.admin.ch/en/geo-information-switzerland/geodata-index-inspire/surface-representation/land-cover.html. Last accessed March 2019.

Thornton, J.M., Mariethoz, G., Brunner, P., 2018. A 3D geological model of a structurally complex alpine region as a basis for interdisciplinary research. Sci. Data 5, 1–20. https://doi.org/10.1038/sdata.2018.238.

Thornton, J.M., Therrien, R., Mariethoz, G., Linde, N. and Brunner, P., under review. Simulating fully-integrated hydrological dynamics in complex Alpine headwaters. Under review, Water Resources Research. Pre-print available at: https://doi.org/10.31223/X5RG7Q.

Tobin, C., Nicotina, L., Parlange, M.B., Berne, A., Rinaldo, A., 2011. Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region. J. Hydrol. 401 (1–2), 77–89. https://doi.org/10.1016/j.jhydrol.2011.02.010.

Tuo, Y., Marcolini, G., Disse, M., Chiogna, G., 2018. A multi-objective approach to improve SWAT model calibration in alpine catchments. J. Hydrol. 559, 347–360. https://doi.org/10.1016/j.jhydrol.2018.02.055.

Vittoz, P., Randin, C., Dutoit, A., Bonnet, F., Hegg, O., 2009. Low impact of climate change on subalpine grasslands in the Swiss Northern Alps. Glob. Change Biol. 15 (1), 209–220. https://doi.org/10.1111/j.1365-2486.2008.01707.x.

Viviroli, D., Archer, D. R., Buytaert, W., Fowler, H. J., Greenwood, G. B., Hamlet, A. F., . . . Woods, R. (2011). Climate change and mountain water resources: overview and recommendations for research, management and policy. Hydrology and Earth System Sciences, 15 (2), 471–504. doi: 10.5194/hess-15-471-2011.

Vögeli, C., Lehning, M., Wever, N., Bavay, M., 2016. Scaling precipitation input to spatially distributed hydrological models by measured snow distribution. Front. Earth Sci. 4 (December), 1–15. https://doi.org/10.3389/feart.2016.00108.

Vuille, M., Carey, M., Huggel, C., Buytaert, W., Rabatel, A., Jacobsen, D.,…Sicart, J.-E. (2018). Rapid decline of snow and ice in the tropical Andes – Impacts, uncertainties and challenges ahead. Earth-Science Reviews, 176 (October 2017), 195–213. doi: 10.1016/j.earscirev.2017.09.019.

Wagner, T., Themeßl, M., Schüppel, A., Gobiet, A., Stigler, H., Birk, S., 2017. Impacts of climate change on stream flow and hydro power generation in the Alpine region. Environmental Earth Sciences 76 (1), 4. https://doi.org/10.1007/s12665-016-6318-6.

Wang, X.-Y., Wang, J., Jiang, Z.-Y., Li, H.-Y., Hao, X.-H., 2015. An Effective Method for Snow-Cover Mapping of Dense Coniferous Forests in the Upper Heihe River Basin Using Landsat Operational Land Imager Data. Remote Sensing 7 (12), 17246–17257. https://doi.org/10.3390/rs71215882.

Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H., Kunstmann, H., 2013. Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. Water Resour. Res. 49 (5), 2619–2637. https://doi.org/10.1002/wrcr.20219.

Wayand, N.E., Marsh, C.B., Shea, J.M., Pomeroy, J.W., 2018. Globally scalable alpine snow metrics. Remote Sens. Environ. 213 (May), 61–72. https://doi.org/10.1016/j.rse.2018.05.012.

Wulf, H., Bookhagen, B., Scherler, D., 2016. Differentiating between rain, snow, and glacier contributions to river discharge in the western Himalaya using remote-sensing data and distributed hydrological modeling. Adv. Water Resour. 88, 152–169. https://doi.org/10.1016/j.advwatres.2015.12.004.

Würzer, S., Wever, N., Juras, R., Lehning, M., Jonas, T., 2017. Modelling liquid water transport in snow under rain-on-snow conditions–considering preferential flow. Hydrol. Earth Syst. Sci. 21, 1741–1756. https://doi.org/10.5194/hess-21-1741-2017.