



Low spatial autocorrelation in mountain biodiversity data and model residuals

MATHIEU CHEVALIER ^{1,†} HEIDI MOD,^{1,2} OLIVIER BROENNIMANN,^{1,3} VALERIA DI COLA,¹ SARAH SCHMID,¹
HÉLÈNE NICULITA-HIRZEL,⁴ JEAN-NICOLAS PRADERVAND,⁵ BENEDIKT R. SCHMIDT ^{6,7}
SYLVAIN URSENBACHER,^{6,8} LOÏC PELLISSIER,^{9,10} AND ANTOINE GUISAN^{1,3}

¹Department of Ecology and Evolution, University of Lausanne, Biophore, Lausanne CH-1015 Switzerland

²Department of Geosciences and Geography, University of Helsinki, Helsinki FI-00014 Finland

³Institute of Earth Surface Dynamics, University of Lausanne, Géopolis, Lausanne CH-1015 Switzerland

⁴Department Health, Work and Environment, Center for Primary Care and Public Health (Unisanté), University of Lausanne, Epalinges CH-1066 Switzerland

⁵Swiss ornithological institute, Valais Field station, Sion CH-1950 Switzerland

⁶Info Fauna - Karch, UniMail, Bâtiment G, Bellevaux 51, Neuchâtel 2000 Switzerland

⁷Institut für Evolutionsbiologie und Umweltwissenschaften, Universität Zürich, Winterthurerstrasse 190, Zürich 8057 Switzerland

⁸Department of Environmental Science, Section of Conservation Biology, University of Basel, Basel CH-4056 Switzerland

⁹Landscape Ecology, Institute of Terrestrial Ecosystems, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland

¹⁰Unit of Land Change Science, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

Citation: Chevalier, M., H. Mod, O. Broennimann, V. Di Cola, S. Schmid, H. Niculita-Hirzel, J.-N. Pradervand, B. R. Schmidt, S. Ursenbacher, L. Pellissier, and A. Guisan. 2021. Low spatial autocorrelation in mountain biodiversity data and model residuals. *Ecosphere* 12(3):e03403. 10.1002/ecs2.3403

Abstract. Spatial autocorrelation (SAC) is a common feature of ecological data where observations tend to be more similar at some geographic distance(s) than expected by chance. Despite the implications of SAC for data dependencies, its impact on the performance of species distribution models (SDMs) remains controversial, with reports of both strong and negligible impacts on inference. Yet, no study has comprehensively assessed the prevalence and the strength of SAC in the residuals of SDMs over entire geographic areas. Here, we used a large-scale spatial inventory in the western Swiss Alps to provide a thorough assessment of the importance of SAC for (1) 850 species belonging to nine taxonomic groups, (2) six predictors commonly used for modeling species distributions, and (3) residuals obtained from SDMs fitted with two algorithms with the six predictors included as covariates. We used various statistical tools to evaluate (1) the global level of SAC, (2) the spatial pattern and spatial extent of SAC, and (3) whether local clusters of SAC can be detected. We further investigated the effect of the sampling design on SAC levels. Overall, while environmental predictors expectedly displayed high SAC levels, SAC in biodiversity data was rather low overall and vanished rapidly at a distance of ~5–10 km. We found low evidence for the existence of local clusters of SAC. Most importantly, model residuals were not spatially autocorrelated, suggesting that inferences derived from SDMs are unlikely to be affected by SAC. Further, our results suggest that the influence of SAC can be reduced by a careful sampling design. Overall, our results suggest that SAC is not a major concern for rugged mountain landscapes.

Key words: correlograms; Mantel; Moran; mountains; spatial autocorrelation; species distribution models; western Swiss Alps.

Received 10 January 2020; accepted 22 October 2020. Corresponding Editor: Cory Merow.

Copyright: © 2021 The Authors. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

† **E-mail:** mathieu.chevalier38@gmail.com

INTRODUCTION

Assessing how species occurrence or abundance varies across space and along environmental gradients is a long-standing goal of ecology, recently revived by an increasing interest about the consequences of global change on biodiversity (Bellard et al. 2012, Guisan et al. 2017). Species distribution models (SDMs) have been widely used for this purpose (Guisan et al. 2017) and for a number of related applications including conservation planning (Rodríguez et al. 2007, Guisan et al. 2013) or estimating the potential distribution of species under various scenarios of climate change or of habitat fragmentation (Peterson et al. 2011, Guisan et al. 2017). SDMs are empirical models relating the occurrence or the abundance of species at known locations to a set of environmental variables characterizing those locations (Guisan and Zimmermann 2000). When constructing SDMs, several sources of uncertainty associated with both the data and the methods can be introduced into the modeling framework (Elith et al. 2002, Thuiller et al. 2019). Among them, spatial autocorrelation (SAC) is a recurrently reported concern because it violates the assumption that model residuals are independent, which could limit the predictive capacity of models (Cablak et al. 2002, Segurado et al. 2006, Dormann 2007).

SAC is frequently found in ecological or environmental data and characterizes the fact that locations tend to be more similar (e.g., in terms of species occurrences and/or physical characteristics) at some geographic distance(s) than expected by chance (Legendre and Fortin 1989, Legendre 1993). In the literature, two main types of SAC are commonly considered: the one caused by endogenous processes and the one induced by exogenous processes (Legendre and Legendre 2012). In the endogenous case, the spatial pattern is a result of contagious population processes such as growth, reproduction, dispersal and mortality (Kendall et al. 2000, Liebhold et al. 2004). It is known as true SAC (Legendre et al. 2002) or inherent SAC (Fortin and Dale 2005). SAC generated by exogenous processes results from independent responses to autocorrelated environmental gradients and is often referred as induced spatial dependence (Fortin and Dale 2005) or spatial dependency (Legendre

et al. 2002). In this case, spatially structured environmental factors such as geomorphology or climate are responsible for the spatial structure in species distribution data (Legendre et al. 2002, Diniz-Filho et al. 2003). Both types of SAC can be found in natural settings, and while their relative contribution may vary, both have been shown to generate complex spatial patterns, for example, spatial synchrony in population dynamics (Liebhold et al. 2004). Interestingly, while dispersal is putatively a major driver of SAC, most previous studies usually investigated spatial patterns in SAC using a measure of Euclidean distance between sampling plots (Fortin and Dale 2005, Borcard and Legendre 2012). While this measure might be appropriate for some taxa (e.g., birds), it is likely a poor descriptor of dispersal pathways for other taxa dispersing on the ground (e.g., reptiles). Least-cost modeling can be used to identify routes with the lowest cumulative resistance between target locations on a landscape (Adriaensen et al. 2003) and may reveal new insights about SAC.

SAC can be seen both as a source of information about spatial patterns and processes and as a methodological issue (Diniz-Filho et al. 2003). The latter arises because autocorrelated observations are less informative than independent observations implying that a new sample point is not carrying one complete degree of freedom (Fortin and Dale 2005). The lack of independence among observations can be problematic for statistical analyses, because if the pattern remains present in the residuals, then one of the key assumptions that residuals are independent and identically distributed is violated (Dormann et al. 2007*a, b*). In regression analyses, SAC causes the variance to be biased downward, which can affect model parameters and increase type I error rates (Dormann 2007). The impact of SAC on model parameters has been highlighted in multiple studies (Bini et al. 2009, Hodges and Reich 2010) and can have strong implications for model predictions (Kühn 2007). Nonetheless, other studies have shown that SAC does not necessarily induce bias (Diniz-Filho et al. 2003) and that it has a relatively small effect on model parameters and predictions compared with other factors (Thibaud et al. 2014).

In this study, we aim to provide the first regional assessment of the importance of SAC across

multiple taxonomic groups in a study area where intensive species distribution modeling has been carried out in the past years (Dubuis et al. 2011, D’Amen et al. 2015). For this purpose, we use presence–absence data collected for 850 species or genus belonging to nine major taxonomic groups (plants, bumblebees, lepidopterans, orthopterans, fungi, bacteria, protists, amphibians, and reptiles; see Mod et al. 2020). These data were collected within an interdisciplinary research program conducted in the western Swiss Alps (<http://rechalp.unil.ch>). Since the taxonomic resolution varies across taxonomic groups (owing to the poorly known taxonomy at the species level for microorganisms), we hereafter use the more general term “taxa” for the sake of clarity. Because SAC can be assessed in different ways (Dormann et al. 2007a, b), we use multiple statistical tools from different R packages to evaluate (1) the global level of SAC across the study area, (2) the spatial pattern (i.e., how SAC varies with distance between sampling plots) and the spatial extent of SAC (i.e., the distance at which SAC levels off), and (3) whether local clusters of SAC can be detected. These evaluations are performed on four types of variables: presence–absence data collected for 850 taxa, quantitative data for six predictors characterizing environmental conditions at the sampling plots, and residuals obtained from SDMs fitted with two different algorithms to the 850 taxa’s presence–absence data with the six predictors included as covariates. Multi-species presence–absence data are also frequently used to assess how the composition and the structure of communities are changing along environmental gradients (i.e., beta diversity; Legendre and De Cáceres 2013) using multivariate analyses (e.g., MANOVA, RDA, variance partitioning) that are also sensitive to SAC (Legendre et al. 2008, Cao et al. 2019). We thus also evaluate whether the dissimilarity in community composition displays evidence for spatial structuring for each taxonomic group. Finally, we investigate whether the sampling design can influence SAC by comparing our results (which are based on a particular sampling design; see below) with SAC levels computed on a dataset not following any sampling strategy.

Because mountains usually display strong environmental heterogeneity (Garcillán and

Ezcurra 2003), we expect the level of SAC to be low over the study area, although local clusters of SAC (Anselin 1995) may be present. If endogenous processes are causing SAC, one can expect taxonomic groups with high dispersal abilities (e.g., insects) to display SAC over a larger spatial extent than taxonomic groups with lower dispersal abilities (e.g., fungi, bacteria). This should be particularly the case when considering a measure of distance that accounts for dispersal costs. Based on previous studies, we expect environmental predictors to show high levels of SAC (Koenig 1999). Nonetheless, we expect taxa’s presence–absence data to display low to moderate levels of SAC owing to the rugged topography of the landscape, although some species may still display substantial levels of SAC. Finally, we expect residuals from SDMs to display low levels of SAC for most taxa, owing to the inclusion within the modeling framework of relevant environmental predictors, which have a direct impact on the physiology of the taxa.

MATERIALS AND METHODS

Study area

The study area is a priority area for interdisciplinary research at the University of Lausanne. It is located in the western Swiss Alps (Vaud, Switzerland, 46°10′–46°30′ N; 6°60′–7°10′ E) and is covering approximately 700 km². The region is formed by alpine meadows, forests, and agricultural lands including crop fields and pastures. It is characterized by a large elevation gradient, ranging from 375 to 3210 m. The climate is temperate with annual temperature and precipitation varying from 8°C and 1200 mm at an altitude of 600 m to –5°C and 2600 mm at 3000 m (Bouët 1985). A detailed description of the study area is provided at <http://rechalp.unil.ch> and in Randin et al. (2006). The average distance between sampling plots is 15 km (SD = 8.2 km; Appendix S1: Table S1).

Datasets

Reptiles and amphibians.—Observation points for reptiles and amphibians were obtained from the database maintained by infofauna—karch (<http://www.karch.ch/karch/de/home.html>). Following the target-group approach (Phillips et al. 2009), we transformed this presence-only dataset

into a presence–absence dataset by assuming that a taxon was absent if it did not occur in plots where other taxa were detected. To increase the reliability of absence observations, the reptile dataset was complemented with field observations collected during summer 2016 (Pittet 2017). This field campaign aimed to better examine zones where absence observations were reported and at possible distribution boundaries. For reptiles, the dataset contains 12 taxa observed over 1144 locations, while for amphibians, the dataset contains information for 14 taxa observed in 133 locations.

Plants.—The vegetation dataset contains presence–absence data for 795 vascular plants surveyed over 909 plots of 4 m². The plots were chosen following a balanced random-stratified sampling design (Hirzel and Guisan 2002), implemented here by subdividing the study area into environmental strata based on elevation, slope, and aspect, considering only regions outside of forested areas, and ensuring a minimum distance between plots of 150 m. This design ensures a more even sampling along the main environmental gradients, leading to better estimates of species' environmental response curves (Hirzel and Guisan 2002).

Insects.—Insects were sampled in 50 × 50 m plots centered on the coordinates of the above-mentioned vegetation sampling plots to increase the chance of capture for highly mobile species that can fly away from the sampled plots. For orthopterans and bumblebees, 202 plots were sampled (Pradervand et al. 2013, 2014a), while 208 plots were available for lepidopterans (Pelissier et al. 2013). The sampling was performed during the active hours of insects (10:00–17:00) under good weather conditions (i.e., little wind, sunny, and temperatures above 15°C). In total, the dataset contains presence–absence data for 41 orthopteran, 29 bumblebee, and 140 lepidopteran taxa.

Fungi.—Soil from 103 plots centered on the coordinates of the above-mentioned vegetation survey plots was sampled at the four cardinal points and in the center of a square of 2 × 2 m encompassing the 1–5 cm topsoil. Composition of fungal assemblage was determined by metabarcoding internal transcribed spacer 1 (ITS1), with Illumina HiSeq sequencing, followed by closed-reference clustering against the

ITS1_HiSeq dataset (97% threshold). The dataset contains presence–absence information for 190 taxa. For a detailed description of the methods and dataset, see Pinto-Figueroa et al. (2019).

Bacteria.—In addition to the 103 plots sampled for fungi, soil from additional 155 plots was sampled following the same strategy as the one defined for fungi. Composition of bacterial assemblage was determined by metabarcoding of the V5 hypervariable region of the 16S rRNA gene with Illumina HiSeq sequencing, followed by clustering and taxonomical assignment at the 97 % similarity threshold using the gg_13_8 database from Greengenes as a reference. Within the 258 sampled plots, 758 taxa were detected. For a detailed description of the methods and dataset, see Yashiro et al. (2016, 2018).

Protists.—Out of the soil samples collected from the 258 plots, 220 samples were investigated for protist composition. The metabarcoding of the V4 region of the ribosomal RNA small subunit gene was done by MiSeq Illumina, followed by clustering and taxonomical assignment using the trimmed PR2 database. Within the 220 samples, 496 taxa were detected. For the detailed description of the methods and dataset, see Seppey et al. (2019).

Environment.—For all considered plots, we extracted six environmental variables commonly used to model species distribution: the annual mean temperature (Tmean), the annual temperature range (Trange), the annual precipitation sum (Psum), the topographical position index (Topo; describing the position of plots in a gradient from ridge to hollow), the slope (Slp), and the potential amount of annual solar radiation (Srad). The three climatic variables were extracted from MeteoSwiss (meteoswiss.admin.ch) at a 1-km resolution covering the period 1981–2010, and subsequently downscaled to 25-m resolution by incorporating elevation information (see CHclim25 in <https://www.unil.ch/ecospat/en/home/menuguid/ecospat-resources/data.html>). Specifically, we used local regressions of temperature as a function of elevation to estimate the temperature–elevation lapse rate using 1-km resolution data in a 5 km radius moving window. Then, we applied this lapse rate to downscale temperature at 25 m using a 25-m digital elevation model. Given the rugged topography of the study area (i.e., adjacent pixels of 25 m can

have a difference in elevation up to 243 m), temperature can vary substantially even at a scale of 25 m (i.e., up to 1.13°C), suggesting that this resolution is appropriate to map climate for species distribution modeling in this mountain region (see also Pradervand et al. 2014b). While this downscaling procedure likely reduces SAC in climatic variables relative to their original resolutions, it was necessary to accurately represent the climatic variability within the study area (Pradervand et al. 2014b). The three topography-related variables were derived at 25-m resolution from digital elevation models (Federal Office of Topography; swisstopo.ch). For further information on the calculation of environmental data, see Dubuis et al. (2011), Descombes et al. (2016), and <https://www.unil.ch/ecospat/home/menuguid/ecospat-resources/data.html>. Note that insect data were sampled in 50 × 50 m plots, whereas predictors have a 25-m resolution, thus potentially causing a scale-mismatch problem (Guisan and Thuiller 2005). To check whether this could be a problem, we used focal windows to calculate average environmental values for the eight cells surrounding focal pixels (i.e., where sampling took place). We found that environmental values in surrounding pixels were highly correlated with the values of the focal pixel (Pearson's correlation coefficients ranged from 0.92 to 0.99), indicating that climatic conditions are rather homogeneous around sampled plots. Given that the sampled plots are located in topographically homogeneous sites, our results are unlikely to be strongly affected by this scale-mismatch problem.

Statistical analyses

To assess SAC within communities, we included all taxa with more than 10 occurrences, meaning 12 reptiles, 5 amphibians, 344 plants, 26 orthopterans, 20 bumblebees, 90 lepidopterans, 126 fungi, 409 bacteria, and 315 protists. For SAC analyses on presence–absence data and SDM residuals, we only considered taxa with more than 30 occurrences throughout all sampled plots, as recommended (Legendre and Fortin 1989) and classically done in SDM analyses (Randin et al. 2006, Dubuis et al. 2013). This selection resulted in 11 reptiles, 3 amphibians, 201 plants, 10 orthopterans, 14 bumblebees, 50 lepidopterans, 57 fungi, 288 bacteria, and 216 protists. All

analyses were performed in R version 3.6.3 (R Core Team 2019).

Species distribution models (SDMs).—For each taxa, we used presence–absence information together with environmental values extracted over the sampled plots to fit SDMs with two algorithms using the biomod2 package (Thuiller et al. 2019). Specifically, we fitted a generalized additive model (GAM) and a generalized linear model (GLM) using the default settings of biomod2. Both models were fitted with a binomial distribution and only included additive effects. Quadratic terms were considered for GLMs, while the number of smoothing terms was directly estimated from the GAM algorithm. A stepwise selection procedure based on AIC was used to remove unimportant predictors from GLMs. After running the models, we extracted the residuals for SAC analyses.

For each taxa, we evaluated model performance using the area under the ROC curve (AUC; Swets 1988). Specifically, we performed a cross-validation procedure based on repeated split sampling (80% for calibration and 20% for evaluation) with 10 runs to evaluate both the predictive power (average AUC over the 10 evaluation runs) and the goodness of fit (average AUC over the 10 calibration runs) of the models. Note that while evaluation metrics suggest a lack of fit for some taxa (e.g., AUC close to 0.5 on the calibration dataset), which can be problematic if the purpose is to predict taxa distributions; we considered all taxa for our assessment of SAC. To check whether this lack of fit can influence SAC, we performed linear regressions between AUC values (independent variable) and SAC estimated on presence–absence data or on SDMs' residuals (dependent variables).

SAC in presence–absence data, environmental predictors, and SDMs' residuals.—Presence–absence data, environmental predictors, and SDMs' residuals are all univariate vectors. Overall, 2604 of such vectors were tested for SAC: the 850 presence–absence variables, the six environmental predictors extracted at the sampled plots for each of the nine taxonomic groups (totaling 54 variables), and the 850 residual variables extracted from the two SDMs (totaling 1700 variables). Because various statistical approaches and R packages can be used to assess SAC and because there are no rules of thumb regarding which

approach/package to use, all analyses were repeated using different packages to evaluate the robustness of our results to this choice. Significance was assessed with 500 permutations, and p-values were adjusted for multiple testing using Holm's correction.

To evaluate the global level of SAC, we used five R packages: *ade4* (Dray and Dufour 2007), *ape* (Paradis et al. 2004), *ecodist* (Goslee and Urban 2007), *spdep* (Bivand and Piras 2015), and *vegan* (Oksanen et al. 2015). While these packages differ in how spatial weights are accounted for and the method used to assess the degree of spatial dependence among sampling plots (e.g., Moran's *I* or Mantel tests; see Table 1), they all provide a statistic varying from -1 to $+1$ with zero indicating an absence of SAC.

To assess how SAC varies depending on the distance between plots, we computed spatial correlograms using four packages: *ecodist*, *ncf* (Bjornstad 2019), *pgirmess* (Giraudoux 2016), and *vegan*. While the four packages provide measures of correlations in discrete distance classes, the package *ncf* also estimates the spatial dependence as a continuous function of distance, thus providing a finer assessment of the spatial dependence pattern. This package also returns a statistic indicating the distance at which SAC is equal to zero. We used this statistic as a measure of the spatial extent of SAC.

We computed local indicators of spatial association (LISA; Anselin 1995) to study how SAC varies locally over the sampling area and whether we can detect clusters in the spatial arrangement of sampling plots. For this purpose, we used the packages *ncf* and *spdep* and extracted the proportion of plots with significant LISA to evaluate the importance of this phenomenon over the study area.

The packages *vegan*, *ecodist*, and *ape* can include non-Euclidean distance matrices, which are interesting to consider in mountain landscapes where Euclidean distance is a poor descriptor of dispersal routes. We thus repeated our analyses where these packages were used (i.e., global level of SAC and spatial correlograms) using a measure of topographic distance between plots. Topographic distance was computed from a digital elevation model at a 25-m resolution using the package *topoDistance* (Wang 2019).

SAC in community composition.—To investigate the global level of SAC over the study area, we performed Mantel tests between a matrix of ecological distance and a matrix of geographical distance for each taxonomic group. For the latter, we used both Euclidean and topographic distances, while for the former, we considered two dissimilarity measures classically used in the literature for presence–absence data: Sørensen and Jaccard. To investigate the spatial pattern of SAC, we computed spatial correlograms. All analyses were performed using the packages *vegan*, *ncf*, and *ecodist*.

Influence of the sampling design on SAC.—To investigate the extent to which SAC is affected by the sampling design, we re-evaluated the global level of SAC on presence–absence data, environmental predictors, and SDMs' residuals using a gridded dataset containing a subset of the species considered above. This dataset does not follow any sampling strategy. It is a grid with a 100-m² resolution where observations mostly come from volunteers. As for reptiles and amphibians, we used the target-group approach (Phillips et al. 2009) to generate absences from this presence-only dataset. We found 202 species in common between the two datasets including six reptiles, three amphibians, 39 lepidopterans, 10 orthopterans, and 144 plants. Predictors were upscaled to a resolution of 100 m to match with the resolution of this dataset. SDM runs and residual extractions were performed as described above. SAC was measured using the package *spdep* with differences between the two datasets evaluated using the Mann–Whitney *U*-tests.

RESULTS

Because similar results were obtained using topographic distances, we here only present results based on Euclidean distances, as this measure is the one commonly used in most studies (Borcard and Legendre 2012). Topographic distance-based results are available in Appendix S1 (Figs. S1–S4).

Global SAC

Environmental predictors.—The global level of SAC varied from -0.03 to 1 (Fig. 1). Regardless of the package or the taxonomic group considered, the highest levels of SAC were detected for

Table 1. Characteristics of the R packages used to assess SAC.

	ape	spdep	pgirmess	vegan	ncf	ecodist	ade4
Global measure of autocorrelation	Yes	Yes	No	Yes	No	Yes	Yes
Statistic used	Moran	Moran	Moran	Mantel	Mantel or Moran	Mantel	Moran
Spatial correlogram (distance classes)	No	Yes	Yes	Yes	Yes	Yes	No
Spatial correlogram (spline)	No	No	No	No	Yes	No	No
Spatial extent of autocorrelation (X intercept)	No	No	No	No	Yes	No	No
Local indicator of spatial association (LISA)	No	Yes	No	No	Yes	No	No
Spatial weights	Euclidean or other distances	Neighborhood contiguity	Neighborhood contiguity	Euclidean or other distances	Euclidean distances	Euclidean or other distances	Neighborhood contiguity
Multivariate data	No	No	No	Yes	Yes	Yes	No
Randomization test	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Related functions	Moran.I	moran.test, moran.mc, localmoran, sp.correlogram	correlog	mantel.correlog, mantel	correlog, correlog.nc, mantel.correlog, spline.correlog	mantel	gearymoran, mantel.randtest, mantel.rtest

Note: The package ecodist heavily relies on functions from the package vegan and thus provides identical results in most cases.

climatic variables ($\overline{SAC} = [0.52, 0.39, 0.54]$; $SD_{SAC} = [0.22, 0.24, 0.28]$ for Tmean, Trange, and Psum, respectively; Appendix S1: Table S2) with all values being significantly different from zero (Appendix S1: Table S3), while the lowest levels of SAC were detected for topographic variables ($\overline{SAC} = [0.10, 0.11, 0.10]$; $SD_{SAC} = [0.10, 0.10, 0.11]$ for Slp, Srad, and Topo; Appendix S1: Table S2). For the latter, SAC levels were not systematically different from zero (Appendix S1: Table S3). Estimates of SAC varied depending on the package considered (Appendix S1: Table S2). Low levels of SAC were obtained using the packages vegan and ecodist, and to a lower extent using the package ape (Fig. 1, Appendix S1: Table S2). Higher levels of SAC were obtained using the packages ade4 and spdep.

Presence-absence data.—The global level of SAC varied from -0.1 to 0.6 (Fig. 1). Orthopterans and reptiles were the groups displaying the highest level of SAC ($\overline{SAC} = [0.16, 0.12]$; $SD_{SAC} = [0.12, 0.14]$, respectively). The remaining groups presented average levels of SAC below 0.10 with low standard deviations, suggesting low variation across taxa, regardless of the package

considered (Appendix S1: Table S2). Nonetheless, despite low SAC levels at the group scale, a substantial proportion of taxa displayed significant levels of SAC in most groups, particularly when using some packages (Appendix S1: Table S3). For instance, using the packages spdep and ape, the average level of SAC for plants was rather low (0.12), but more than 96% of SAC measures were significant. Using the three other packages, this proportion was zero.

Community data.—The global level of SAC was low, regardless of the dissimilarity measure used (Sørensen or Jaccard) or the package considered (Appendix S1: Fig. S5). It varied from -0.01 for amphibians to 0.19 for lepidopterans (Appendix S1: Table S4).

Residuals from SDMs.—Species distribution models presented an overall good performance with 70% (GLM) and 98% (GAM) of species presenting an AUC above 0.7 on the calibration dataset (Appendix S1: Fig. S6). We found very low evidence for SAC in model residuals, regardless of the taxonomic group considered, the algorithm used to fit the models, or the package used to compute SAC (Fig. 1, Appendix S1: Table S2).

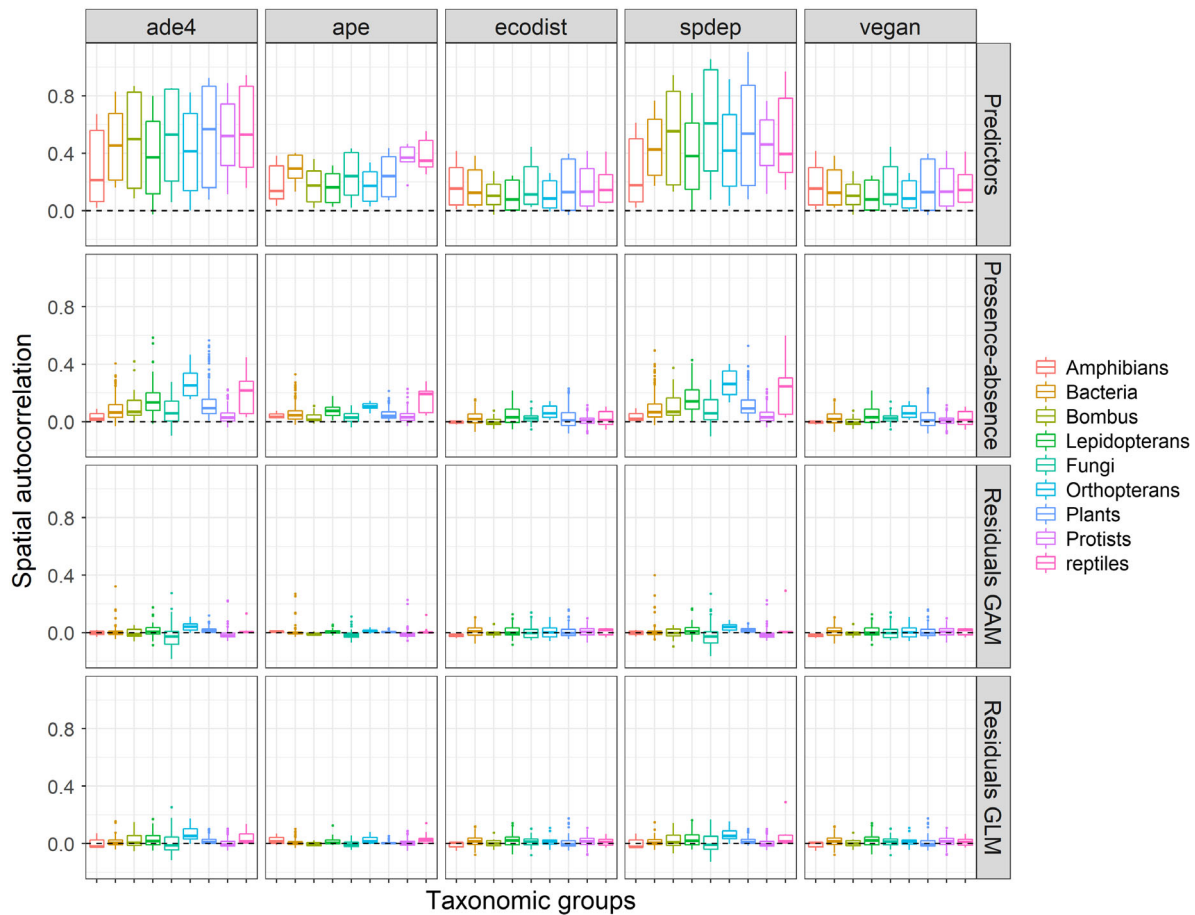


Fig. 1. Global level of SAC measured over the study area for nine taxonomic groups. This assessment was made using five different R packages for six environmental predictors extracted at the plots where each taxonomic group was sampled (first row), 850 taxa's presence–absence data (second row), residuals obtained from 850 SDMs fitted with a GAM (third row), and residuals obtained from 850 SDMs fitted with a GLM (last row).

Nonetheless, variation was evident across taxa and models. For GAMs, the level of SAC varied between -0.18 and 0.4 , while for GLM, it varied between -0.12 and 0.29 . For reptiles and orthopterans, a constantly non-null proportion of taxa displayed significant levels of SAC when considering GLM residuals, though in different proportions depending on the package considered (Appendix S1: Table S3). For all taxonomic groups, a lower proportion of taxa displayed significant levels of SAC when considering GAM residuals (Appendix S1: Table S3).

Link between model performance (AUC) and SAC.—We found that both the goodness of fit and the predictive power of the models dramatically increased when presence–absence data

displayed higher levels of SAC (Appendix S1: Fig. S7). This was particularly the case when SAC was estimated with the packages *ade4* and *spdep* with slope coefficients ranging from 0.57 to 0.67 , thus indicating that a 0.1 increase in AUC values would result in a 0.62 increase in SAC, on average. In contrast, we found weak relationships between SAC estimated on model residuals and AUC values with a maximum slope coefficient reaching 0.18 (Appendix S1: Figs. S8–S9).

Spatial pattern and spatial extent of SAC

Environmental predictors.—Spatial correlograms showed that the level of SAC was moderate in the first distance classes and rapidly vanished at a distance of ~ 10 – 12 km for climatic variables

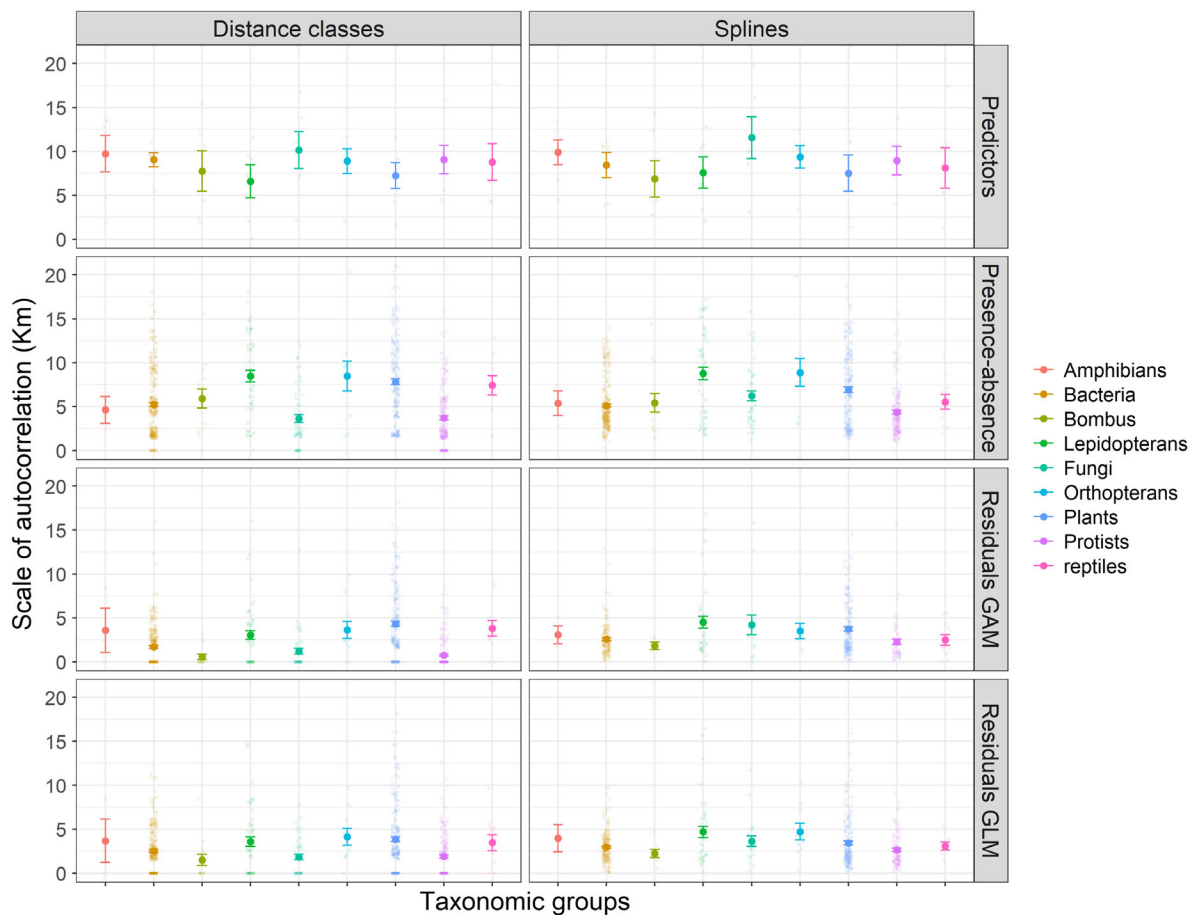


Fig. 2. Spatial extent of SAC over the study area for nine taxonomic groups. The spatial extent of SAC was measured as the distance at which the level of SAC is equal to zero for environmental predictors, taxa's presence-absence data, and residuals extracted from two SDMs. This assessment was made using the package *ncf* using either spatial correlograms with distance classes or spatial correlograms with splines. Points correspond to the average spatial extent of SAC, while vertical bars represent the associated standard error. Shaded dots point to the estimated spatial extent of SAC for each taxa (or predictors depending on the type of variable considered).

and a distance of four to seven kilometers for topographic variables (Fig. 2, Appendix S1: Table S5). This pattern was similar across packages, despite differences in the magnitude of SAC in the different distance classes (Fig. 3, Appendix S1: Fig. S10, Tables S6–S7).

Presence-absence data.—We found a similar pattern for taxa, but the magnitude of SAC in the different distance classes was considerably lower than for predictors and close to zero even in the first distance classes (Fig. 3, Appendix S1: Fig. S10, Table S6). The proportion of taxa displaying significant levels of SAC was low in all distance classes, regardless of the taxonomic

group considered (Appendix S1: Table S7). Nonetheless, large variations were evident across taxa with levels of SAC varying between, for example, -0.15 and 0.69 in the first distance class. The spatial extent of SAC ranged from 4 km for protists to more than eight kilometers for lepidopterans and orthopterans (Fig. 2, Appendix S1: Table S5).

Community data.—The level of SAC decreased as the distance between plots increased for all packages, regardless of the dissimilarity measure used and the taxonomic group considered (Appendix S1: Fig. S11, Table S8). The level of SAC was low in all distance classes using the

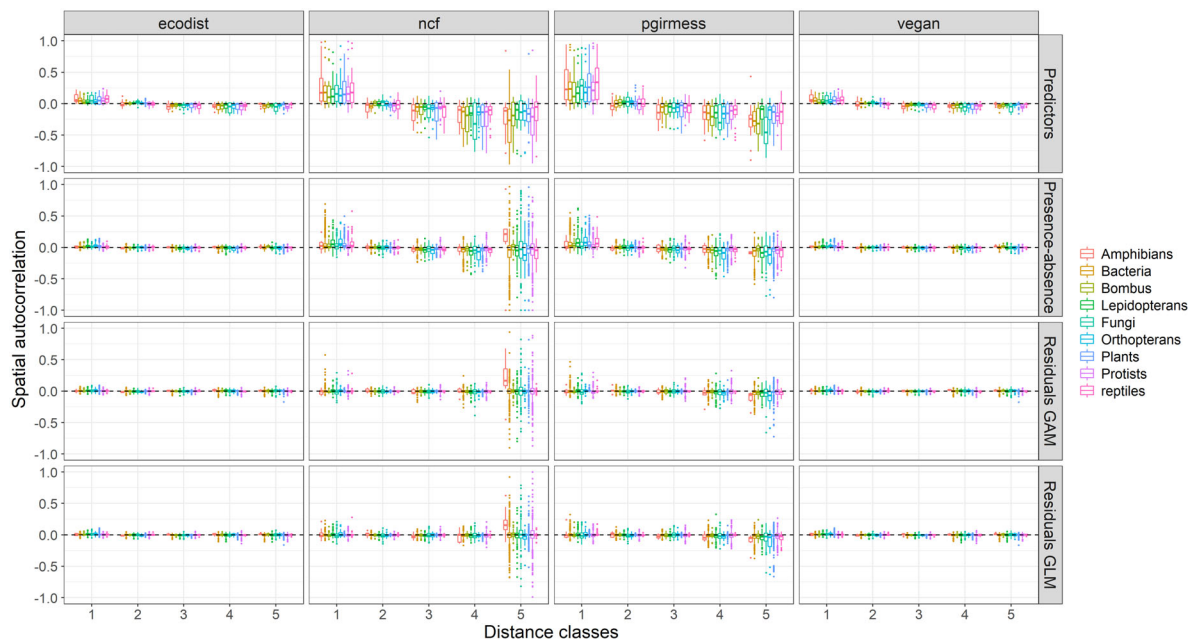


Fig. 3. Spatial correlograms derived from four different R packages presenting SAC levels in discrete distance classes for nine taxonomic groups. Spatial correlograms were computed for environmental predictors, taxa's presence–absence data, and residuals extracted from two SDMs from top to down, respectively. The correspondence with distance classes is as follows: 1 = from 0.968 to 9.06 km, 2 = from 9.06 to 17.1 km, 3 = from 17.1 to 25.2 km, 4 = from 25.2 to 33.2 km, and 5 = from 33.2 to 41.3 km. A similar assessment performed with spline correlograms using the package *ncf* is available in the supplementary material (Appendix S1: Fig. S10).

packages *ecodist* and *vegan* but higher using the package *ncf* (Appendix S1: Fig. S11).

Residuals from SDMs.—The spatial pattern of SAC was flat and close to zero in all distance classes, regardless of the taxonomic group considered, the package used, or the model fitted (Fig. 3, Appendix S1: Fig. S10, Tables S6–S7). Some variations were evident among taxa with SAC levels ranging between -0.13 and 0.27 for GLMs and between -0.17 and 0.57 for GAMs. The distance at which SAC vanishes was short with an estimated maximum of 4.4 km for orthopterans when considering GLM residuals (Fig. 2). The spatial extent of SAC was lower when considering GAM residuals.

Local SAC

Considering predictors and model residuals, the proportion of plots displaying significant level of SAC was low and below 20% for most of the taxonomic groups (Fig. 4). Nonetheless, considering presence–absence data, some taxa presented a high proportion of plots with significant

levels of SAC. Fig. 5 provides an overview of the spatial clustering of SAC over the study area for predictors extracted at insects' sampled plots and for presence–absence data collected for the species *Metrioptera saussuriana* and associated SDM residuals.

Influence of the sampling design on SAC

When considering the gridded dataset, SAC tended to be higher for predictors, and model residuals, but to be slightly lower for presence–absence data (Appendix S1: Fig. S12). Interestingly, GAMs performed much better than GLMs in removing SAC from the gridded dataset.

DISCUSSION

Complementing the long history of spatial autocorrelation (SAC) studies, in terms of both mathematical developments and their applications to data, this study provides a comprehensive assessment of the importance of SAC

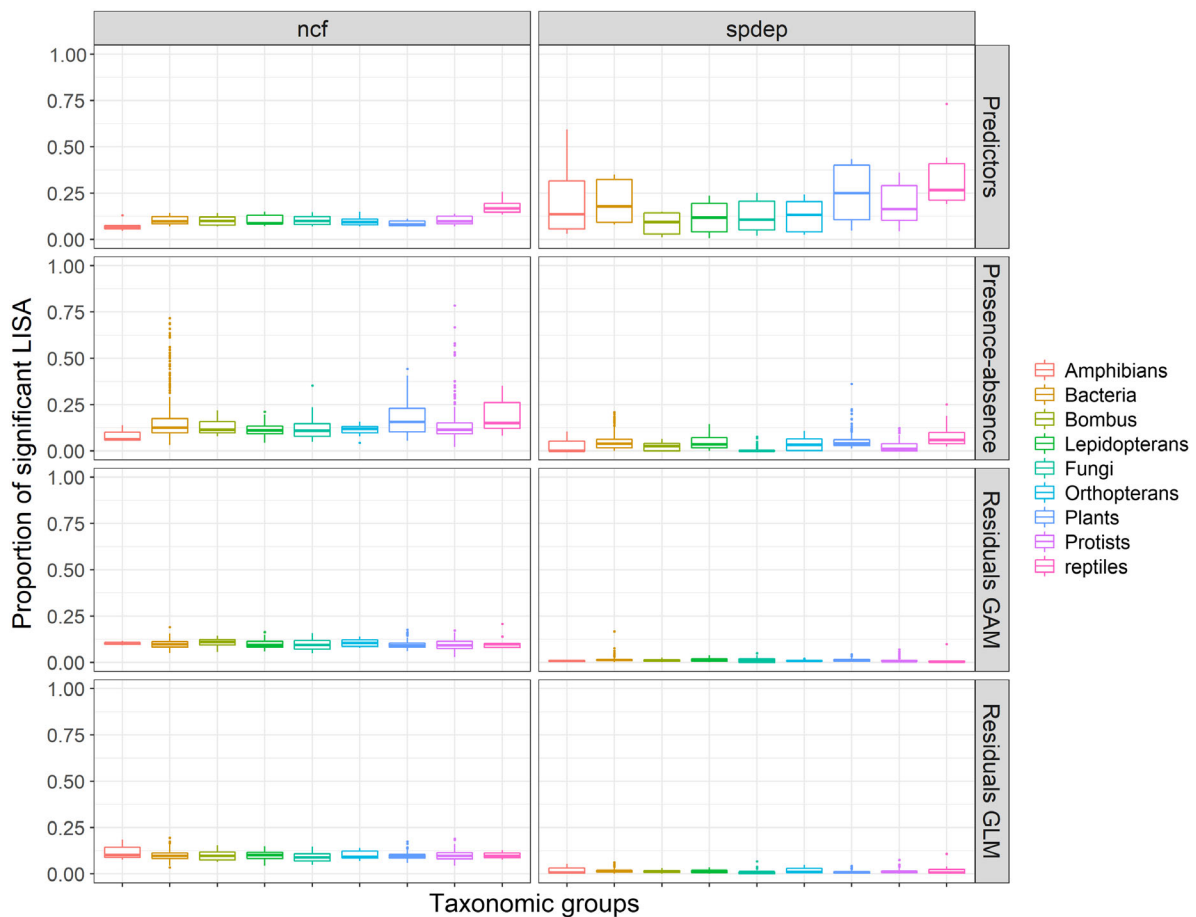


Fig. 4. Proportion of plots displaying significant levels of SAC for each taxonomic group. Local SAC was measured with local indicators of spatial association using two different R packages for environmental predictors, taxa's presence-absence data, and residuals extracted from two SDMs from top to down, respectively.

simultaneously for a large number of taxa, environmental predictors, and residuals from spatial models relating presence-absence data to environmental predictors (i.e., species distribution models, SDMs) for a whole study area at the regional scale. Although in ecology, data are often thought to be strongly spatially autocorrelated (Legendre 1993, Koenig 1998, 1999, Dormann 2007), here using data for nine distinct taxonomic groups and 850 single taxa, we showed that this is not necessarily the case in rugged mountain landscapes. Despite the fact that environmental predictors expectedly showed a substantial degree of SAC, taxa observations were only weakly spatially autocorrelated at the scale of the study area ($\sim 700 \text{ km}^2$) and SDMs' residuals accordingly presented low

evidence for SAC, supporting a safe use of the models and of their predictions from that perspective. While the rugged topography of mountains can explain the low prevalence of SAC highlighted here, we also showed that an appropriate sampling design (here balanced random stratified, except for reptiles and amphibians) allows reducing the prevalence of SAC in ecological data and SDM residuals.

The level of SAC was low across the study area for most taxa. However, this does not necessarily mean that SAC is not present because the analyses we have used assume that SAC is homogeneous over the study area (Ord and Getis 2001). Local indicators of spatial association can be used to relax this assumption and to detect local clusters of sites displaying different levels of

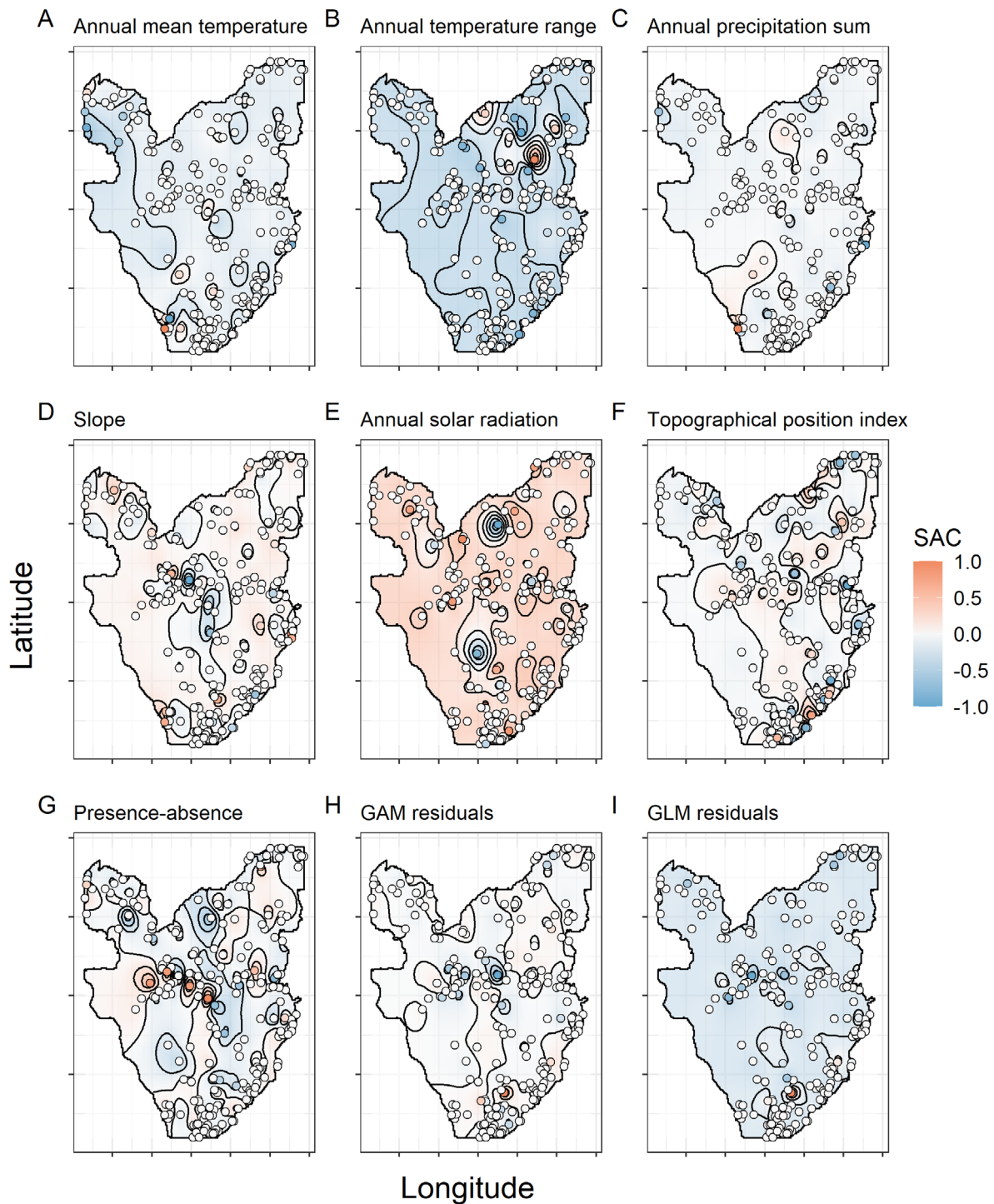


Fig. 5. Local indicators of spatial autocorrelation (LISA) over the study area as estimated with the package *ncf* for the six environmental predictors (abbreviations are Tmean, annual mean temperature; Trange, annual temperature range; Psum, annual precipitation sum; Topo, topographical position index; Slp, slope; and Srad, potential amount of annual solar radiation) characterizing environmental conditions at the plots where orthopteran

(Fig. 5. *Continued*)

were sampled (A–F), for presence–absence data collected for the species *Metrioptera saussuriana* (G) and for residuals obtained from two species distribution models fitted to species presence–absence data with the six predictors included as covariates (H–I). The background surface was approximated from LISA values using multi-level B-Splines. Red colors are indicative of positive SAC, while blue colors are indicative of negative SAC. White colors indicate no SAC. Points indicate the location of sampling plots with white colors indicating non-significant LISA values. Significant LISA values are indicated with colors corresponding to estimated LISA values.

SAC (Anselin 1995). Using this approach, we found that some taxa presented a substantial proportion of plots with high levels of SAC, thus suggesting that SAC can be important locally. The fact that some of these plots displayed negative levels of SAC, while others presented positive levels of SAC explains why the level of SAC can be low at the scale of the study area. Regarding model residuals, the proportion of plots presenting evidence for local SAC was close to zero, suggesting that SDMs are, nevertheless, able to deal with this local clustering of SAC.

Using spatial correlograms, we found that the level of SAC decreases as the distance between plots increases and tends to level-off at a distance of five to ten km. Except for environmental predictors, where the level of SAC was high in the first distance class, SAC was low in all distance classes for both presence–absence data and model residuals, regardless of the distance measure considered (i.e., topographic or Euclidean). A decrease in SAC as the distance between plots increases is a common ecological pattern (Koenig 1999, Bell 2001). This phenomenon can be induced by different mechanisms including decreased similarity of environmental properties (Nekola and White 1999), limited dispersal abilities of organisms (Hubbell 2001), and complex spatial arrangement of landscape characteristics (Garcillán and Ezcurra 2003). Although the mechanisms responsible for this pattern are difficult to disentangle, similarly low levels of SAC at short distances have previously been highlighted in heterogeneous landscapes presenting important dispersal barriers (Garcillán and Ezcurra 2003). Thus, the heterogeneous nature of the study area might explain the low SAC highlighted in the different distance classes.

Several methods have been proposed to account for SAC. These include a priori procedures such as sampling design ensuring sufficient space between plots (Guisan and Theurillat 2000) or a posteriori methods that consist in

correcting the number of degrees of freedom in statistical analyses (Legendre et al. 2002) or incorporating an auto-covariate term in regression models (Lichstein et al. 2002, Crase et al. 2012). Random-stratified, or other systematic, sampling designs—as performed for most of the taxonomic groups considered here—are advocated to prevent SAC across sampling plots (Legendre and Fortin 1989) and might explain why SAC is generally low over the study area, a desirable outcome when building and projecting species distribution models (Record et al. 2013, Guélat and Kéry 2018). The fact that we found higher SAC levels when considering a gridded dataset (though still low) confirms that appropriate sampling designs can help prevent SAC in ecological data.

The low SAC levels can be explained by the spatial scale of the study area. SAC is expected to be more pronounced at small spatial scales (Collingham et al. 2000) where the influence of endogenous factors (e.g., dispersal) is more prevalent than at larger scales (Dormann et al. 2007a, b). However, SAC was here measured over a relatively coarse spatial scale relative to the environmental and topographic variability present in the study area, with an average distance between plots of 15 km. This large distance implies that the influence of endogenous processes is likely difficult to detect. By considering a measure of topographic distance, we attempted to highlight the influence of such processes but obtained similar results to Euclidean distance-based analyses, confirming that these processes are complex to detect empirically, unless additional data are considered (Mielke et al. 2020). Our results, however, suggest that SAC is likely of exogenous origin in this study area. Indeed, if exogenous factors are driving species occurrences to be more similar than expected by chance at neighboring plots, then including these factors in statistical models may considerably lower or even eliminate SAC in model residuals

(Higgins et al. 1999, Fisher et al. 2002, Dormann et al. 2007a, b, Record et al. 2013). Accordingly, we found that model performance strongly increased when presence–absence data displayed higher levels of SAC while no such relationship existed with model residuals, thus suggesting that predictors (which display high levels of SAC) are efficient in explaining SAC in presence–absence data. Nonetheless, the fact that some taxonomic groups presenting low dispersal abilities (e.g., fungi, protists) showed a tendency to present lower SAC levels than others (e.g., orthopterans, lepidopterans) suggests that endogenous processes might still be at play.

The need to investigate the prevalence of SAC in model residuals, including SDMs, has been stressed in several studies (Dormann et al. 2007a, b, Dormann 2007, Kühn 2007, Kissling and Carl 2008). The presence of SAC in model residuals could be the result of different, non-mutually exclusive, processes including the influence of an unmeasured and spatially structured environmental variable (i.e., hidden variable) or the influence of endogenous processes related to population dynamics (Dray et al. 2012). While spatial models incorporating an auto-covariate term can be used to remove the spatial dependence in model residuals (Lichstein et al. 2002, Dormann et al. 2007a, b), their use is controversial. A major concern is that models with an auto-covariate term are poorly transferable in time and space, because sites where predictions are to be made can be different than sites used to fit the models in terms of, for example, landscape structure or biotic interactions (Guisan and Thuiller 2005, Segurado et al. 2006; but see Record et al. 2013). Moreover, it remains unknown in which circumstances spatial models provide more accurate estimates of model parameters compared with non-spatial models (Kissling and Carl 2008). Here, we found a low prevalence of SAC in model residuals for the majority of the taxa, thus suggesting that in our study, SDMs are likely robust with respect to SAC (Diniz-Filho et al. 2003; see also Thibaud et al. 2014 for similar evidence in the same area). On average, residuals of GAMs tended to be less spatially autocorrelated than those of GLMs, suggesting that GAMs may be an alternative to GLMs in case SAC is present in model residuals (see also Crase et al. 2012). Both GAMs and

GLMs usually display good transferability and although predictive accuracy is usually higher with GLMs (Randin et al. 2006, Heikkinen et al. 2012, Norberg et al. 2019) we here found that these two approaches have a very similar predictive performance. Note, however, that an absence of SAC in model residuals does not imply that the model is unbiased (Dormann 2007, Dormann et al. 2007a, b) and additional tests have to be performed depending on the objective (e.g., normally distributed residuals, multicollinearity among predictors, predictive accuracy of the model).

Using different R packages, we found that our results were generally robust to the choice of the package used to assess SAC. The spatial pattern of SAC was similar across packages and all suggested low SAC in model residuals. There were nonetheless some discrepancies regarding the magnitude of SAC or the proportion of taxa presenting significant levels of SAC both at the scale of the study area and in the different distance classes. These discrepancies occurred for different reasons (Table 1). For instance, in the packages *vegan* and *ecodist* (the latter is actually based on the former and both provide very similar results), SAC is assessed using the Mantel's r statistic, whereas in the packages *ape*, *spdep*, and *ade4*, SAC is assessed through the Moran's I statistic. Both statistics can be used to assess SAC and have been shown to display similar statistical power (Borcard and Legendre 2012). They nevertheless differ with respect to the null hypothesis tested: The Moran statistic is testing for an absence of SAC in the variable of interest, whereas the Mantel statistic is testing for an absence of relationship between sample dissimilarities and spatial distance. Even when the same statistic is used, discrepancies can occur depending on how spatial weights are accounted for. For instance, while the packages *spdep* and *ape* both rely on the Moran statistic, *spdep* relies on a neighborhood contiguity matrix, whereas *ape* relies on a matrix of Euclidean distances between plots. We here used different packages to assess the robustness of our results but also because the different assessments performed cannot all be done with the same package (but see package *ncf*; Table 1). Depending on the purpose of the study, the use of one specific package may be enough, but whenever possible, we

advise to test SAC using multiple approaches and metrics.

CONCLUSIONS

This study provides an assessment of SAC at three different levels including presence–absence data, environmental predictors, and model residuals. In general, we found a low prevalence of SAC, which can be explained by the spatial heterogeneity of the mountain landscape, the use of a random-stratified sampling design, and the type of data considered (i.e., presence–absence instead of abundance data; see Guélat and Kéry 2018). This result combined with the fact that, in the same area, SAC has been shown to only have a small effect on SDM parameters compared with other commonly reported factors (Thibaud et al. 2014), suggest that SDMs can be safely used in heterogeneous landscapes, providing that the models—at least from that perspective—are fit for purpose (Guillera-Arroita et al. 2015). Whether these results can be generalized to other study areas and datasets is not warranted, but is likely for datasets collected with an environmentally random-stratified sampling design in landscapes with similar topographic and climatic settings. Additional studies conducted in landscapes displaying different levels of environmental heterogeneity would make it possible to determine whether some landscape configurations are more prone to display SAC than others. Considering different sampling strategies (e.g., uniform, random) and data types (e.g., presence–absence or abundance) would also be welcome.

ACKNOWLEDGMENTS

We are grateful to Quentin Blandenier, Enrique Lara, Edward Mitchell, Eric Pinto-Figueroa, Christophe Sèpey, David Singer, Jan van der Meer, and Erika Yashiro for their contribution to some microbial data. Data were sampled during previous SNF projects to AG: BIOASSEMBLE (grant nr 31003A_125145), SESAM'ALP (grant nr 31003A-1528661), and Soil Microbial Biogeography (grant nr PDFMP3-135129) and the FP6 project ECOCHANGE funded by the European Commission (Grant agreement ID 36866). AG obtained further support for this study from the Swiss National Science Foundation (SNF), INTEGRALP project (grant number CR23 I2_162754), and SOMETALP project (grant number 315230_184908).

LITERATURE CITED

- Adriaensen, F., J. P. Chardon, G. De Blust, E. Swinnen, S. Villalba, H. Gulinck, and E. Matthysen. 2003. The application of “least-cost” modelling as a functional landscape model. *Landscape and Urban Planning* 64:233–247.
- Anselin, L. 1995. Local indicators of spatial analysis — LISA. *Geographical Analysis* 27:93–115.
- Bell, G. 2001. Neutral macroecology. *Science* 293:2413–2418.
- Bellard, C., C. Bertelsmeier, P. Leadley, W. Thuiller, and F. Courchamp. 2012. Impacts of climate change on the future of biodiversity. *Ecology Letters* 15:365–377.
- Bini, L. M., et al. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography* 32:193–204.
- Bivand, R., and G. Piras. 2015. Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software* 63:1–36.
- Bjornstad, O. N. 2019. ncf: spatial Covariance Functions. <https://cran.r-project.org/package=ncf>
- Borcard, D., and P. Legendre. 2012. Is the Mantel correlogram powerful enough to be useful in ecological analysis? A simulation study. *Ecology* 93:1473–1481.
- Bouët, M. 1985. Pages 473–148 in *Climat et météorologie de la Suisse romande*, Nouv. éd., Payot, Switzerland.
- Cablk, M., D. White, and A. Kiester. 2002. Assessment of spatial autocorrelation in empirical models in ecology. Pages 429–440 in *Predicting species occurrences: issues of scale and accuracy*. Island Press, Washington, D.C., USA.
- Cao, K., X. Mi, L. Zhang, H. Ren, M. Yu, J. Chen, J. Zhang, and K. Ma. 2019. Examining residual spatial correlation in variation partitioning of beta diversity in a subtropical forest. *Journal of Plant Ecology* 12:636–644.
- Collingham, Y. C., R. A. Wadsworth, B. Huntley, and P. E. Hulme. 2000. Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology* 37:13–27.
- Crane, B., A. C. Liedloff, and B. A. Wintle. 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography* 35:879–888.
- D’Amen, M., J. N. Pradervand, and A. Guisan. 2015. Predicting richness and composition in mountain insect communities at high resolution: a new test of the SESAM framework. *Global Ecology and Biogeography* 24:1443–1453.

- Descombes, P., J. N. Pradervand, J. Golay, A. Guisan, and L. Pellissier. 2016. Simulated shifts in trophic niche breadth modulate range loss of alpine butterflies under climate change. *Ecography* 39:796–804.
- Diniz-Filho, J. A. F., L. M. Bini, and B. A. Hawkins. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography* 12:53–64.
- Dormann, C. F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography* 16:129–138.
- Dormann, C., et al. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30:609–628.
- Dray, S., et al. 2012. Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs* 82:257–275.
- Dray, S., and A. B. Dufour. 2007. The ade4 Package: implementing the duality diagram for ecologists. *Journal of Statistical Software* 22:1–20.
- Dubuis, A., S. Giovanettina, L. Pellissier, J. Pottier, P. Vittoz, and A. Guisan. 2013. Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science* 24:593–606.
- Dubuis, A., J. Pottier, V. Rion, L. Pellissier, J. P. Theurillat, and A. Guisan. 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions* 17: 1122–1131.
- Elith, J., M. Burgman, and H. Regan. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling* 157:313–329.
- Fisher, R. N., A. V. Suarez, and T. J. Case. 2002. Spatial patterns in the abundance of the coastal horned lizard. *Conservation Biology* 16:205–215.
- Fortin, M.-J., and M. R. T. Dale. 2005. *Spatial analysis: a guide for ecologists*. Second edition. Cambridge University Press, Cambridge, UK.
- Garcillán, P. P., and E. Ezcurra. 2003. Biogeographic regions and β -diversity of woody dryland legumes in the Baja California peninsula. *Journal of Vegetation Science* 14:859–868.
- Giraoudoux, P. 2016. *pgirmess: data Analysis in Ecology*. <http://cran.r-project.org/package=pgirmess>
- Goslee, S. C., and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* 22:1–19.
- Guélat, J., and M. Kéry. 2018. Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution* 9:1614–1625.
- Guillera-Aroita, G., J. J. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P. E. Lentini, M. A. McCarthy, R. Tingley, and B. A. Wintle. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* 24:276–292.
- Guisan, A., et al. 2013. Predicting species distributions for conservation decisions. *Ecology Letters* 16: 1424–1435.
- Guisan, A., and J. P. Theurillat. 2000. Equilibrium modeling of alpine plant distribution: How far can we go? *Phytocoenologia* 30:353–384.
- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8:993–1009.
- Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. *Habitat suitability and distribution models: with applications in R*. First edition. Cambridge University Press, Cambridge, UK.
- Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186.
- Heikkinen, R. K., M. Marmion, and M. Luoto. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* 35:276–288.
- Higgins, S. I., D. M. Richardson, R. M. Cowling, and T. H. Trinder-Smith. 1999. Predicting the landscape-scale distribution of alien plants and their threat to plant diversity. *Conservation Biology* 13:303–313.
- Hirzel, A., and A. Guisan. 2002. Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling* 157:331–341.
- Hodges, J. S., and B. J. Reich. 2010. Adding spatially-correlated errors can mess up the fixed effect you love. *American Statistician* 64:325–334.
- Hubbell, S. P. 2001. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press, Princeton, New Jersey, USA.
- Kendall, B., O. Bjørnstad, J. Bascompte, T. Keitt, and F. William. 2000. Dispersal, environmental correlation, and spatial synchrony in population dynamics. *American Naturalist* 155:628–636.
- Kissling, W. D., and G. Carl. 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography* 17:59–71.
- Koenig, W. D. W. 1998. Spatial autocorrelation in California land birds. *Conservation Biology* 12:612–620.
- Koenig, W. W. D. 1999. Spatial autocorrelation of ecological phenomena. *Trends in Ecology and Evolution* 14:22–26.
- Kühn, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions* 13:66–69.

- Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? *Ecology* 74:1659–1673.
- Legendre, P., D. Borcard, and P. R. Peres-Neto. 2008. Analyzing or explaining beta diversity? *Comment. Ecology* 89:3238–3244.
- Legendre, P., M. R. T. Dale, M. J. Fortin, J. Gurevitch, M. Hohn, and D. Myers. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25:601–615.
- Legendre, P., and M. De Cáceres. 2013. Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. *Ecology Letters* 16:951–963.
- Legendre, P., and M.-J. Fortin. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80:107–138.
- Legendre, P., and L. F. J. Legendre. 2012. *Numerical ecology*. Third edition. Elsevier, Amsterdam, The Netherlands.
- Lichstein, J. W., T. R. Simons, S. A. Shriener, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72:445–463.
- Liebholt, A. M., W. Koenig, and O. Bjørnstad. 2004. Spatial synchrony in population dynamics. *Annual Review of Ecology, Evolution and Systematics* 35:467–490.
- Mielke, K. P., T. Claassen, M. Busana, T. Heskes, M. A. J. Huijbregts, K. Koffijberg, and A. M. Schipper. 2020. Disentangling drivers of spatial autocorrelation in species distribution models. *Ecography* 43:1–11.
- Mod, H. K., et al. 2020. Greater topoclimatic control of above- versus below-ground communities. *Global Change Biology* 26:6715–6728.
- Nekola, J. C., and P. S. White. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* 26:867–878.
- Norberg, A., et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs* 89:1–24.
- Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2015. *vegan: community Ecology Package*. <http://cran.r-project.org/package=vegan>
- Ord, K. J., and A. Getis. 2001. Testing for local spatial autocorrelation in the presence of global autocorrelation. *Journal of Regional Science* 41:411–432.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Pellissier, L., N. Alvarez, A. Espíndola, J. Pottier, A. Dubuis, J. N. Pradervand, and A. Guisan. 2013. Phylogenetic alpha and beta diversities of butterfly communities correlate with climate in the western Swiss Alps. *Ecography* 36:541–550.
- Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo. 2011. *Ecological niches and geographic distributions*. Princeton University Press, Princeton, New Jersey, USA.
- Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* 19:181–197.
- Pinto-Figueroa, E. A., E. Seddon, E. Yashiro, A. Buri, H. Niculita-Hirzel, J. R. Van Der Meer, and A. Guisan. 2019. Archaeorhizomycetes spatial distribution in soils along wide elevational and environmental gradients reveal co-abundance patterns with other fungal saprobes and potential weathering capacities. *Frontiers in Microbiology* 10:1–13.
- Pittet, M. 2017. Impact of global warming on the distribution and dispersal of reptiles in the Western Swiss Alps. Lausanne, Switzerland.
- Pradervand, J.-N., L. Pellissier, C. F. Randin, and A. Guisan. 2014a. Functional homogenization of bumblebee communities in alpine landscapes under projected climate change. *Climate Change Responses* 1:1–10.
- Pradervand, J. N., A. Dubuis, L. Pellissier, A. Guisan, and C. Randin. 2014b. Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Progress in Physical Geography* 38:79–96.
- Pradervand, J. N., A. Dubuis, A. Reymond, V. Sonnay, A. Gelin, and A. Guisan. 2013. Quels facteurs influencent la richesse en orthoptères des Préalpes vaudoises? *Bulletin De La Societe Vaudoise Des Sciences Naturelles* 93:155–173.
- R Core Team. 2019. *R: a language and environment for statistical computing*. Austria, Vienna. <https://www.r-project.org/>
- Randin, C. F., T. Dirnböck, S. Dullinger, N. E. Zimmermann, M. Zappa, and A. Guisan. 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography* 33:1689–1703.
- Record, S., M. C. Fitzpatrick, A. O. Finley, S. Veloz, and A. M. Ellison. 2013. Should species distribution models account for spatial autocorrelation? A test of model projections across eight millennia of climate change. *Global Ecology and Biogeography* 22:760–771.
- Rodríguez, J. P., L. Brotons, J. Bustamante, and J. Seoane. 2007. The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions* 13:243–251.

- Segurado, P., M. B. Araújo, and W. E. Kunin. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology* 43:433–444.
- Seppey, C. V. W., et al. 2019. Soil protist diversity in the Swiss western Alps is better predicted by topo-climatic than by edaphic variables. *bioRxiv*:571760.
- Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293.
- Thibaud, E., B. Petitpierre, O. Broennimann, A. C. Davison, and A. Guisan. 2014. Measuring the relative effect of factors affecting species distribution model predictions. *Methods in Ecology and Evolution* 5:947–955.
- Thuiller, W., D. Georges, R. Engler, and F. Breiner. 2019. *biomod2: ensemble Platform for Species Distribution Modeling*. <https://cran.r-project.org/package=biomod2>
- Wang, I. 2019. *topoDistance: calculating Topographic Paths and Distances*. <https://cran.r-project.org/package=topoDistance>
- Yashiro, E., E. Pinto-Figueroa, A. Buri, J. E. Spangenberg, T. Adatte, H. Niculita-Hirzel, A. Guisan, and J. R. van der Meer. 2016. Local environmental factors drive divergent grassland soil bacterial communities in the western Swiss Alps. *Applied and Environmental Microbiology* 82:6303–6316.
- Yashiro, E., E. Pinto-Figueroa, A. Buri, J. E. Spangenberg, T. Adatte, H. Niculita-Hirzel, A. Guisan, and J. R. Van Der Meer. 2018. Meta-scale mountain grassland observatories uncover commonalities as well as specific interactions among plant and non-rhizosphere soil bacterial communities. *Scientific Reports* 8:1–12.

SUPPORTING INFORMATION

Additional Supporting Information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/ecs2.3403/full>