UNIL | Université de Lausanne

Faculté de biologie et de médecine

Ecole de biologie

IMPROVING ALPINE PLANT SPECIES DISTRIBUTION MODELS WITH HIGH RESOLUTION REMOTE SENSED SNOW COVER

Travail de Maîtrise universitaire ès Sciences en comportement, évolution et conservation
Master Thesis of Science in Behaviour, Evolution and Conservation

par

Julie BOSERUP

Directeur : Pr. Antoine Guisan
Superviseur (s) : Dr. Carmen Cianfrani
Expert (s) : Anonymous
Département d'Ecologie et d'Evolution (DEE)

January 2018

Improving alpine plant species distribution models with high resolution remote sensed snow cover

Abstract

The pivotal role of snow has since long been recognised as an important parameter shaping species distribution and plant communities. Despite this, it has only rarely been incorporated in plant species distribution models. Here, we developed three snow cover indexes using remote sensing of snow cover in the Western Swiss Alps from Landsat8, Sentinel and Worldview2 satellite images at different resolutions. We examined the changes of model predictive performance when adding each snow index in plant species distribution models based on other environmental variables (topography, climate, soil). When models contained only a few environmental variables the predictive power was significantly increased by the addition of our snow indexes at very high resolution and the magnitude was related to the resolution of the satellite images. However, this was not the case when models included more environmental variables. Snow indexes increased particularly the predictive performance of snow beds as well as anthropogenic meadows and pastures species. As this studied area only shows a reduced range of environmental conditions for the studied species, we therefore suggest that more studies are needed using snow indexes from remote sensing at high resolution over larger areas encompassing larger environmental gradients.

Résumé

La couverture neigeuse est depuis longtemps reconnue comme un paramètre clé contrôlant la distribution des espèces et des communautés végétales. Malgré cela, son incorporation dans les modèles de distribution d'espèces de plantes est rare. Ici, nous avons développé trois indices de couverture neigeuse au moyen d'images satellites de différentes résolutions spatiales – Landsat8, Sentinel and Worldview2 – dans les Alpes vaudoises en Suisse. Nous avons examiné les changements de pouvoir prédictif des modèles dû aux ajouts de chaque indice de neige dans les modèles de distribution basés sur un jeu de base de variables environnementales (topographie, climat, sol). Les résultats montrent que le pouvoir prédictif est significativement amélioré par l'addition des indices de neige à très haute résolution quand les modèles ne contiennent que quelques variables environnementales, et l'amélioration augmente avec la résolution des images satellites. Cependant, ce n'était pas le cas dans les modèles contenant davantage de variables environnementales. Les indices de neige augmentent plus particulièrement le pouvoir prédictif des modèles d'espèces de combes à neige et de prairies et pâturages anthropiques. La zone étudiée ne comprenant pas l'ensemble des variations environnementales possibles pour les espèces étudiées, nous encourageons le développement d'études incluant des indices de neige provenant d'images satellites à très haute résolution sur des régions plus grandes incluant des gradients plus larges.

Keywords
Species distribution models (SDMs); alpine flora; remote sensing of snow cover; VHR; Western Swiss Alps

Abbreviations
AUC = area under the ROC curve; DEM = digital elevation model; GIS = geographic information system; GAM = generalised additive models; GBM = generalised boosted method; GLM = generalised linear models; KAPPA = Cohen's Kappa; MMI = multimodel inference analysis; RF = random forest; SDMs = species distribution models; TSS = true skill statistic

1. Introduction

Plants in order to live, reproduce and produce viable offspring require some specific environmental conditions. The combination of those environmental conditions form a N-dimensional hypervolume in which the species can live, called the fundamental niche. However, in the field, some factors such as competition or dispersal limitations constrain the species to live in only part of their environmental niche, those parts form the realised niche (Hutchinson 1957). The realised niche can be modelled by using recordings of species presences and a set of environmental variables. Species distribution models (SDMs) are models that link species occurrences with environmental variables to model the realised niche of the species (Guisan *et al.* 2017). This realised niche can then be projected over a geographic area for which environmental variables are available as continuous maps to obtain the distribution of habitats where the species could potentially be found. Several examples exist that used species distribution in the alpine zone (Dirnböck & Dullinger 2004; Albert *et al.* 2008; Gallien *et al.* 2012; Pottier *et al.* 2014; Pradervand *et al.* 2014; Buri *et al.* 2017). Modelling alpine species distribution is important since alpine species are particularly threatened due to climate change (Thuiller *et al.* 2005; Schöb *et al.* 2009; Randin *et al.* 2010; Engler *et al.* 2011; Gottfried *et al.* 2012).

1.1 Effect of snow on alpine plant communities

Snow has an important effect on alpine plant communities by affecting the phenology, growth and distribution of their constituent plant species (Wipf *et al.* 2009). Microrelief and dominating wind cause the snow to be distributed unevenly (Walker *et al.* 1993; Palacios & Sánchez-Colomer 1997). Therefore, the melt out date varies, determining wind exposed areas, the length of the snow-free period and the location of the meltwater drainage (Walker *et al.* 1993). This then determines plant species distribution. For instance, while cushion plants are found on windy ridges, forbs tend to dominate in snow beds (Billings & Bliss 1959). The snow cover acts as an isolator. Soils under snow cover are usually warmer than uncovered soils, allowing the microbial

activity within the soil to be increased. The thicker the snow layer, the better the isolation effect (Walker *et al.* 1993; Schimel *et al.* 2004; Wahren *et al.* 2005; Wipf *et al.* 2009). While a longer lasting snow cover diminishes the length of the growing season, constraining the species to have a short vegetative cycle (Billings & Bliss 1959), it permits the plant to start its growth in a warmer period (Björk & Molau 2007; Jonas *et al.* 2008; Wipf *et al.* 2009). Furthermore, snow protects the plants against very low temperatures, winter desiccation, solar radiation (Körner 2003). Therefore, timing of the snow cover determines the air temperature and precipitation experienced by the plants just before snowfall and after snowmelt. Those have a high impact on plant height and plant growth rates (Jonas *et al.* 2008).

1.2 Snow modelling and its importance for species distribution models

On the one hand, mechanistic models can be used to model snow (Randin *et al.* 2009b, 2015; Dedieu *et al.* 2012). For instance, Randin *et al.* (2009b) modelled the snow cover by applying a snow transport model with wind data and a digital elevation model (DEM) to a one-metre-deep snow cover. However, due to the coarse resolution used and the low number of parameters included in the models to keep computing times short, such dynamic snow models have difficulties to reflect processes such as avalanches, fine-scale snow redistribution by wind or long lasting snow patches. On the other hand, remote sensing can be used to capture the observed patterns of snow cover, an approach that has been performed in a wide range of locations (Beck *et al.* 2005; Hüsler *et al.* 2014; Macander *et al.* 2015; Dedieu *et al.* 2016; Park *et al.* 2016; Satir 2016). Remote sensing has thus the potential to capture the exact time of snow arrival and melting, and so plants' growing season, but it can't be used to predict future snow distributions.

Even though, as seen previously, snow cover has long been recognised as an important parameter shaping plant communities, it has only rarely been incorporated into species distribution models due to the lack of satellites images with good-enough resolution and the difficulty to distinguish snow from clouds (Niittynen & Luoto 2017). The few recent attempts concluded that snow cover improves species distribution models and the predictive power is increased (Carlson *et al.* 2015; Niittynen & Luoto 2017). In the study of Niittynen & Luoto (2017) which modelled vascular plants, mosses and lichens species distribution in the Arctic, snow was the most important variable for 36% of the studied species and overall the second most important variable across all species, and it was also a very important variable for species living in very diverse snow conditions. Moreover, Carlson *et al.* (2015) showed that including climate variables without snow data in functional and taxonomic diversity modelling can give inaccurate predictions in the French Alps. The older study of Randin *et al.* (2009b) performed in the same area as the present study already showed the possible benefit to add snow. Overall, the model fit of alpine plants distribution models was increased but not their predictive power, and the predicted distribution maps were also ecologically more realistic. However, the

predictive power was only significantly raised by snow for 19 out of 91 species, and snow cover was only modelled with limited accuracy, as no remote sensing data were used. On the other hand, Pottier *et al.* (2014) showed the potential power of remote sensing data to model some species distributions in the same area, but without specific focus on snow and using summer hyperspectral satellite images at a single time step.

The aims of this study were consequently to:
1) develop snow indexes from different satellites images with different spatial and temporal resolutions for the study area
2) test the predictive power of the snow indexes in plant species distribution models (SDMs).
3) relate the change in predictive power of models including snow indexes to species traits or ecology.

Although one can expect the predictive power of the species distribution models to be generally improved by the addition of the three snow indexes, all species' models might not be improved in the same way by the snow indexes as species have different relationships with snow. We expect that the snow indexes will have a distinct variable importance depending on the species being modelled. We also hypothesise that model improvement as well as the importance of snow in the models could also depend on the resolutions of the satellite images.

2. Methods

2.1 Study area

The study area is located in the western Alps of Switzerland in the Vaud canton. It ranges from about 571700 to 581190 and between about 116900 and 128000. It is between the Diablerets massif to the North and the Muveran massif to the South-East. The elevation range from 1070 m in Les Plans-sur-Bex to 2850 m near la Tête Ronde (Figure 1). The vegetation in this area has been greatly impacted by pasturing activities (Randin *et al.* 2009a).

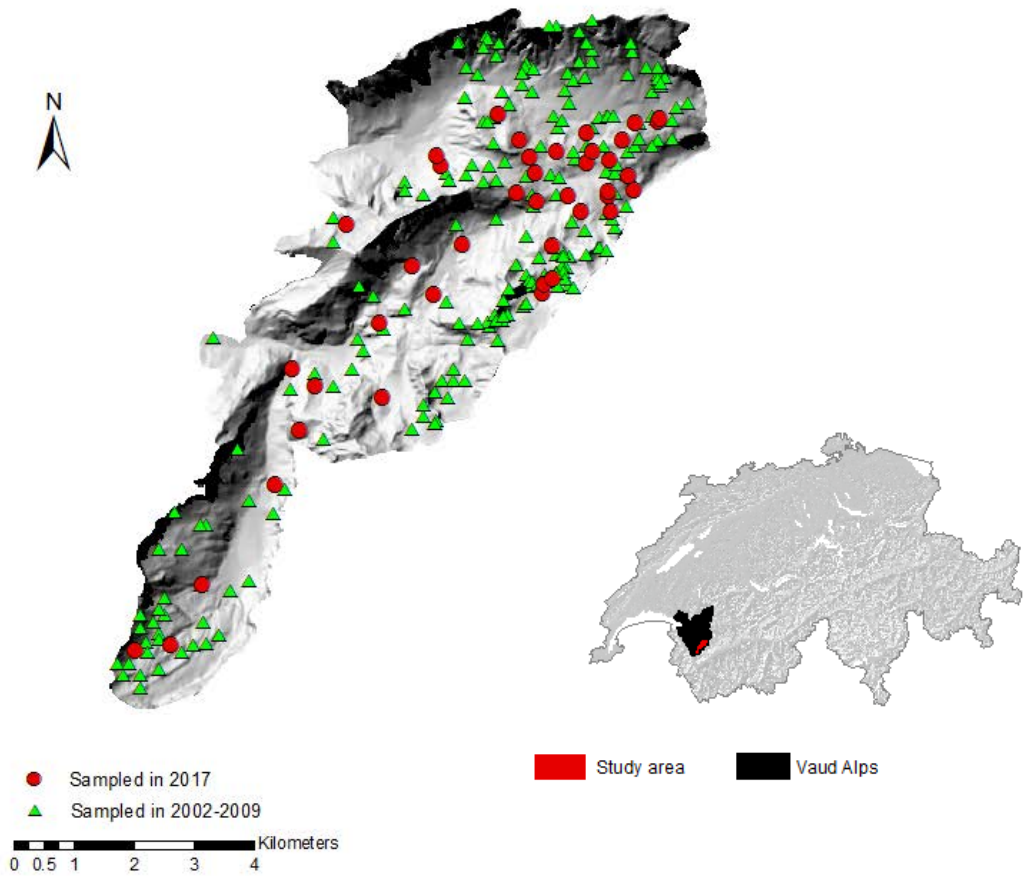


Figure 1. Location of the study area in Switzerland (right) and location of the sampled plots within the study area (left).

2.2 General framework

The study was conducted in two steps. The first step consisted in creating snow cover indexes. The second step consisted in the modelling of plant species distributions using the previously developed snow indexes and evaluating the models and predictions with and without snow (Figure 2).

2.3 Step 1 – Snow cover indexes

Snow cover indexes were obtained from satellite images for the period 2015-2017 using Landsat8 (30 m resolution), Sentinel (10 m resolution) and WorldView2 (1.3 m resolution), satellite images. Pictures where clouds were present were discarded for the analysis. A total of 34 Landsat images, 23 Sentinel images and 10 Worldview images were kept for the analyses.

A snow index was calculated for the images from each satellite: SI_L, SI_S and SI_W for Landsat8, Sentinel and Worldview2 respectively. For Landsat and Sentinel images, the normalised difference snow index (NDSI) was used. The NDSI is based on the difference

of reflectance of snow between the green part of the solar spectrum (GREEN) at 0.5 µm and the short wave infrared (SWIR) at 1.6 µm (Dozier 1989):

$$NDSI = \frac{\rho GREEN - \rho SWIR}{\rho GREEN + \rho SWIR}$$

where ρ is the reflectance. NDSI > 0.4 corresponds usually to snow.

As the SWIR band is not available from Worldview images, the Worldview snow index (SI_w) was calculated following a new formula proposed by Prof. Grégoire Mariéthoz at the University of Lausanne (Explanations in Annexe 1):

$$SI_W = \frac{\rho RED - \max(\rho NIR1,75)}{\rho RED + \max(\rho NIR1,75) + 0.01}$$

The RED band is band 5 at (0.630-0.690µm) and the NIR1 band corresponds to band 7 (0.770-0.895µm) of Worldview.

Then, the values of the different snow indexes were extracted from the satellites images at the locations of the vegetation sampling sites.
For Landsat and Sentinel, values of the snow indexes (SI_L and SI_S) lower than 0.4 were set to 0.4 to avoid the high impact of negative values in further calculations. Negative values are not meaningful as everything below 0.4 corresponds to the absence of snow. Thus, SI_L and SI_S values range typically from 0.4 to 1.
The Worldview snow index (SI_W) was transformed into a binary index by replacing all values higher than -0.2 by 1, indicating snow, and all values lower or equal to -0.2 by 0, indicating no snow (Annexe 1).

At each vegetation sampling location, the mean for each year and then across years was calculated for the three snow indexes in order to account for the fact that the number of images available for the analysis varied between years and satellites. We also tried to take into account only pictures from the three months of the growing season (June, July and August) but the results were better when SI were computed from images throughout the year (thus results for the growing season not reported).

2.4 Step 2 – Plant SDMs using snow cover indexes
2.4.1 – Vegetation sampling
84 plots were selected using a random equal stratified sampling strategy in the study area (Hirzel & Guisan 2002). The points were chosen among 254 plots previously sampled between 2002 and 2009. The stratification was made according to five categories of NDSI values from Landsat8 images over the 2016 growing season (March till September) on the 254 plots. Those five categories were: NDSI≤4, NDSI=5, NDSI=6,

NDSI=7, NDSI≥8. The important number of selected random stratified plots gave the possibility to increase or decrease the number of sampled plots during the field season according to field conditions and to replace plots impossible to reach. Attention was paid to try to keep as much as possible equal sample size between the different NDSI categories. In the end, 41 plots of 2 m by 2 m were sampled during summer 2017. The exact location of the middle of the sampled plots was recorded using a Trimble GeoExplorer GPS, with a precision of less than one metre after post-correction (66.9% of the points had a precision of less than 50 cm and 94.6% of the points had a precision of less than one metre).

The vegetation was sampled using the same strategy as done between 2002 and 2009, i.e. by recording presence-absence of all vascular plant species found in 4 m^2 quadrat. Nomenclature followed Lauber *et al.* (2012).
An assessment of the similarity, using a Sorensen Index, between the period 2002-2009 and 2017 for the 41 plots sampled in 2017 was carried out using the R package ade4 (Dray & Dufour 2007). Sorensen Index was computed as follows:

$$S=\frac{2a}{2a+b+c}$$

where a is the number of species in common between samples 1 and 2, b is the number of species in sample 1 and c is the number of species in sample 2. The Sorenson Index is used for presence-absence data and is useful when the difference between the samples is not due to sample size (Chao *et al.* 2005). Plots for which data were gathered over the 2017 field season were added to the bigger dataset of 254 plots and replaced the data collected in 2002-2009.

2.4.2 – Plant species distribution models (SDMs)
The predictors that we used are important to explain plant species distributions (Randin *et al.* 2009a; b; Dubuis *et al.* 2013b), and all had pairwise correlations lower than 0.7, as recommended by Dormann *et al.* (2013). There were two climatic predictors both at 25 m resolution: growing degree days (gdd) over the growing season and potential evapotranspiration (etp) over the growing season. There were two topographic predictors: topographic position index (tpi) at 5 m resolution and slope at 1 m resolution. Finally, there was one soil predictor, soil pH, at 25 m resolution. Growing degree days (gdd) was calculated as the average number of growing degree days above 0°C per month. We computed gdd and etp from monthly data from 2002 to 2015 (data not available in the database after 2015) by summing the three months of the growing season (June, July and August) for each year and then averaging across the 14 years. The topographic position index (tpi) describes the position of the pixel relative to its surroundings. Positive values indicate ridges (or bumps), negative values indicate

valleys (or holes) and values around zero corresponds to flat areas or areas of constant slopes.
Values of the different predictors were extracted at the locations of the vegetation sampling sites. Since the coordinate of the plots are taken in its centre, there are no issues due to the different resolution of the predictors.

All statistics were performed in the R 3.0.1 statistical software (R Core Team 2017).

Plant SDMs were fitted using the R package biomod2 (Thuiller *et al.* 2016). Four modelling techniques were used to model the distribution of each species: generalised linear models (GLM) (McCullagh & Nelder 1989), generalised additive models (GAM) (Hastie & Tibshirani 1986), generalised boosted method (GBM) (Ridgeway 1999; Friedman *et al.* 2000) and random forest (RF) (Breiman 2001). For testing the predictive performance of the models, the dataset was randomly split into a calibration dataset (80%) and an evaluation dataset (20%). This was repeated 10 times per species. For each run, the area under the receiver-operating characteristic ROC curve (AUC) (Swets 1988), the true skill statistic (maxTSS) (Allouche *et al.* 2006) and the Cohen's Kappa (maxKAPPA) (Cohen 1960) were recorded, and a mean across the 10 runs was made for each agreement metric. The variables importance was computed by the estimation of decrease in explained variance across 3 permutations of each predictor variable (i.e in 3 models) and the mean was recorded. The higher the decrease in explained variance, the higher the variable importance (Thuiller et al. 2016).

AUC ranges from 0 to 1 and maxTSS and maxKAPPA both range from -1 to 1. AUC values between 0.5 and 0.7 indicate weak predictions, values between 0.7 and 0.9 indicate useful predictions and values above 0.9 are good predictions. This corresponds for both maxTSS and maxKAPPA to between 0 and 0.4 for weak predictions, between 0.4 and 0.8 for useful predictions and above 0.8 for good predictions (Thuiller *et al.* 2010). AUC values lower than 0.5 are not better than random (Swets 1988), which corresponds to values of 0 or below for maxKAPPA and maxTSS (see interpretation scales for KAPPA and TSS respectively; Cohen 1960; Allouche *et al.* 2006; Guisan *et al.* 2017).

We ran models considering environmental predictors, including one soil predictor and snow indexes. We included only one soil predictor (i.e. soil pH) since it improved the models of plant species distributions and was the most important soil predictor in several studies (Dubuis *et al.* 2013a; Buri *et al.* 2017). In order to compare the benefit of adding each of the three original snow indexes compared to adding the same variable randomised (i.e. not explanatory anymore), a permuted snow index ($SI_{permuted}$) was created by calculating the mean of the three snow indexes for each of the 254 plots and by shuffling their means (across sites) without replacement. This allowed comparing models with the same number of parameters, but with the randomised variable not

being expected to increase the model performance. To have a first idea of the power of each snow cover index in increasing individual species' models, univariate generalised additive models (GAM) were used to explore the relationship of each of the 47 species, with at least 30 occurrences across the sampled plots, with the three snow indexes each at a time, using the R package mgcv (Wood 2017). Family was set to binomial and the link function to logit. The response curves of the 47 species can be found in Annexe 2. The relationships between the species and the three snow indexes were classified as positive, negative, unimodal or showing no clear trend by visually looking at the response curves.

Due to the low number of observations, we chose to perform three sets of models with a different number of environmental variables and number of species considered. Each time, at least 10 species observations were used per predictor, as recommended by Harrell (2001) (Figure 2).

For the first set of models (SET 1), 149 species with at least 10 occurrences within the 254 plots were considered. Four groups of models were produced. Alternatively, the three snow indexes SI_L, SI_S and SI_W and the permuted snow index ($SI_{permuted}$) were used as predictors. For each model, three measures of agreement AUC, maxTSS and maxKAPPA were recorded. For this set, only GLM, GAM and RF modelling technique were used. For the second set (SET 2), 47 species with at least 30 occurrences within the 254 plots were considered. Four groups of models were produced. Gdd and etp were the predictors common to the four groups. Alternatively, the three snow indexes SI_L, SI_S and SI_W and the permuted snow index ($SI_{permuted}$) were added. For each model, three measures of agreement AUC, maxTSS and maxKAPPA were recorded. The variables importance was also recorded. For the third set (SET 3), 14 species with at least 60 occurrences within the 254 plots were considered. Four groups of models were produced. The variables gdd, etp, tpi, pH and slope were common to the four groups. Alternatively, the three snow indexes – SI_L, SI_S and SI_W - and the permuted snow index ($SI_{permuted}$) were added. The same elements as in the SET 2 were recorded.

2.4.3 – Plant SDMs comparison

We compared the model performance between the models within each set of models by comparing the AUC, maxTSS and maxKAPPA. That is to say that we compared the models computed with the three snow indexes – SI_L, SI_S and SI_W – with models computed with the permuted show index $SI_{permuted}$ using Wilcoxon signed rank tests. We also compared the models computed with the three snow indexes SI_L, SI_S and SI_W between each other using the same test. We calculated the percentage of plant species experiencing an increase in AUC, maxTSS and maxKAPPA due to the addition of snow indexes as predictors compared to the permuted snow index. Finally, we looked at the variables importance of the predictors in SET 2 and 3.

The choice was made to report median values for the model predictive performance and the variables importance as some values have a high impact on the mean, especially when the sample size is low.

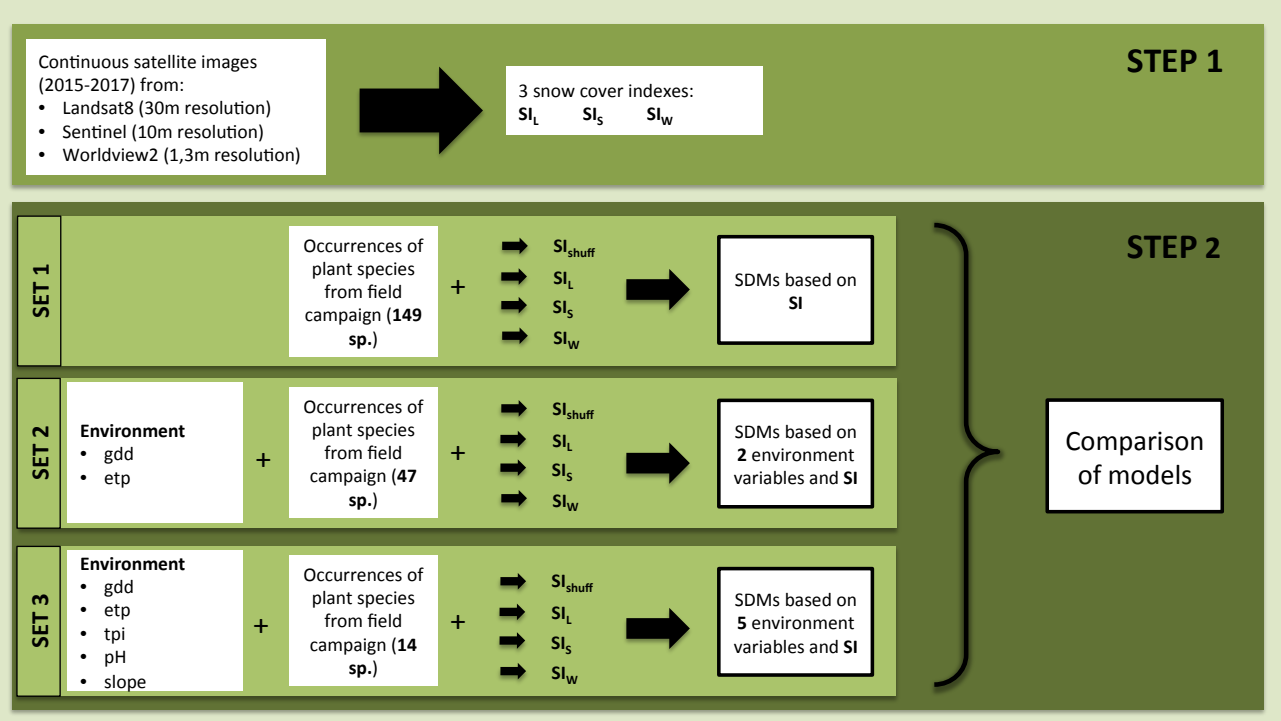


Figure 2. Analytical workflow of the study

2.5 Linking changes in model performance with species traits or ecology

We chose eight species characteristics: vegetation height, specific leaf area (SLA), soil moisture index, humus value, leaf duration, CRS life strategies (i.e. competitive, ruderal and stress-tolerant), growth form and habitat types. Vegetation height is related to competitive ability, stress avoidance and disturbance (Lavorel & Garnier 2002). SLA is the ratio between the surface of one side of a leaf and its dry biomass (Cornelissen *et al.* 2003) and is linked to the photosynthetic rate and carbon fixation (Lavorel & Garnier 2002). Soil moisture index, humus value, leaf duration and CRS life strategies come from *Flora Indicativa* (Landolt *et al.* 2010). Growth form comes from Pignatti (2005) and habitat types from Ellenberg *et al.* (1991). CRS life strategies and habitat types were simplified to a smaller number of categories (Tables 1-2).

To examine if some species were more improved than others regarding those species' characteristics, we first performed GLMs linking the difference of agreement measures (AUC, maxTSS, maxKAPPA) between models including the true snow index and models including $SI_{permuted}$ and all combinations of 3 by 3 traits. Distributions of the residuals were Gaussian. Multimodel inference analysis (MMI) (Burnham *et al.* 2011) using the

MuMin R package (Bartoń 2017) was used to measure the importance of species characteristics. These importance values were based on the weights given to the models in which each trait appeared.

3. Results

3.1 Step 2 – Plant SDMs using snow cover indexes

3.1.1 – Vegetation sampling

In the 41 complementary plots sampled during the 2017 field season, we found between 0 and 47 species per plots. The mean number of species was 24.66, sd= 12.06. In the 254 plots sampled between 2002 and 2017 taken for this analysis, the number of species ranged from 0 to 63, with a mean of 19.41, sd=14.92. The most frequent species were *Campanula scheuzeri, Carex sempervirens, Galium anisophyllon and Polygonum viviparum.* The difference in the mean species richness of the plots was due to the fact that there was a higher proportion of plots with none or only a few species within the 254 plots than within the 41 plots sampled in 2017. The plots discarded because they were impossible to reach during the 2017 field season were often plots with no or only a few species from the previous species lists of 2002-2009. The mean similarity index (Sorenson Index) for the 41 plots sampled in 2017 between the species list obtained in 2002-2009 and 2017 was 0.53, sd=0.19. The limited similarity might be due to the greater imprecision of the GPS coordinates of the plots sampled between 2002 and 2009 and stochastic species dynamics. Species on the lists of 2002-2009, that we could not find in the 2017 plots were often observed in the surroundings.

3.1.2 – Plant species distribution models (SDMs)

The response curves from GAMs of the 47 species with at least 30 occurrences showed the great potential of SI_W to improve individual species' models since only the response curve of 10.6% of the species showed no clear trend in relation to SI_W. This proportion was 53.2 and 48.9% with SI_L and SI_S respectively (Table 3). The relationship between the species and SI_W ranged for most of the species between a unimodal response curve with a distribution centred on low to middle snow values and a negative response curve (Annexe 2).

3.2.3- Plant SDMs comparison

Models in SET 1 make weak predictions as AUC is usually lower than 0.7 and maxKAPPA lower than 0.4. Models including SI_S and SI_W correspond to useful predictions according to maxTSS. The general weak performance was probably due to the fact that only snow was used as a predictor here. Models in SET 2 correspond to useful or close to being useful predictions. Models in SET 3 correspond to useful predictions (Figures 3-5, Tables 4-6).

The overall predictive performance of the three sets of models always increased gradually as the resolution of the satellite images taken for computing the snow index increased. In SET 1 and 2, models with $SI_{permuted}$ had the lowest predictive power, which indicates that snow cover improved the models. The differences between models including SI_S and SI_W and models with $SI_{permuted}$ were highly significant (Wilcoxon signed rank test; p-value<0.01) (Figures 3-4, Tables 4-5). The differences between models including SI_w and SI_L or SI_S were both significant in SET 1 (Wilcoxon signed rank test; p-value<0.05) (Table 7). In SET 2 models, all differences between models including the three snow indexes (i.e. not permuted) were highly significant (Wilcoxon signed rank test; p-value<0.01) (Table 8). In SET 3, models with $SI_{permuted}$ had usually not the lowest predictive power. However, there were no significant differences between models including any of the four snow indexes – $SI_{permuted}$, $SI_{L,}$ SI_S, SI_W – in this set of models (Figure 5, Tables 6 and 9).

For the overall variables importance in SET 2 and SET 3, SI had the lowest variable importance, apart from when SI_W was added. SI_W was the most important variable in SET 2 but the one before last in SET 3. There was a small tendency for the importance of the snow indexes to increase with the resolution of the images taken for the computation of the snow indexes. However, the importance of SI_L and SI_S remained similar. There was a great improvement in importance with SI_W. The importance of the permuted snow was very low in both sets of models, which gives an indication that snow cover did reasonably improve the models. Etp was the most important variable in SET 2, apart when SI_W was in the model. And pH was always the most important variable in the SET 3, independent of the inclusion of SI in the model (Figures 6-7, Tables 10-11).

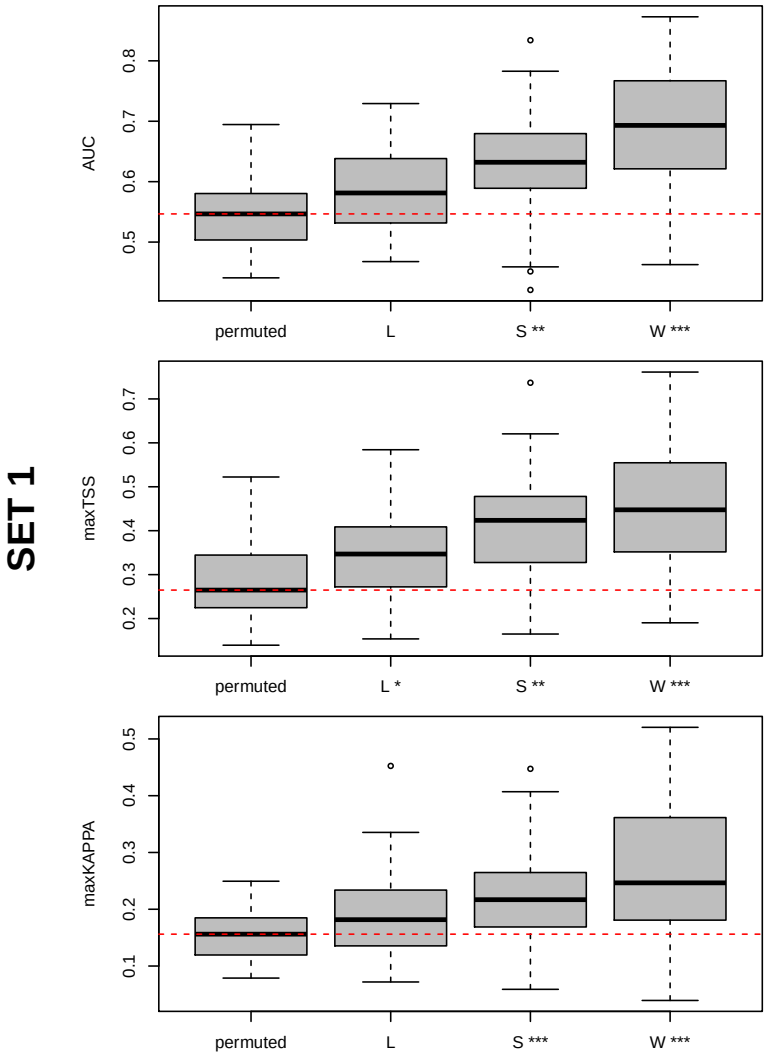


Figure 3. Agreement measures AUC, maxTSS and maxKAPPA for SET 1 models. The x-axis indicates the snow index used in the models ($SI_{permuted}$, SI_L, SI_S, and SI_W respectively). *: p-value<0.05; **: p-value<0.01; ***: p-value<0.001

Stars indicate the significance of the Wilcoxon signed rank test performed on the agreement metrics between the models including the related snow index and the models including $SI_{permuted}$. The dashed red line represents the median of the models including $SI_{permuted}$ to allow an easier comparison with the models containing the others snow indexes.

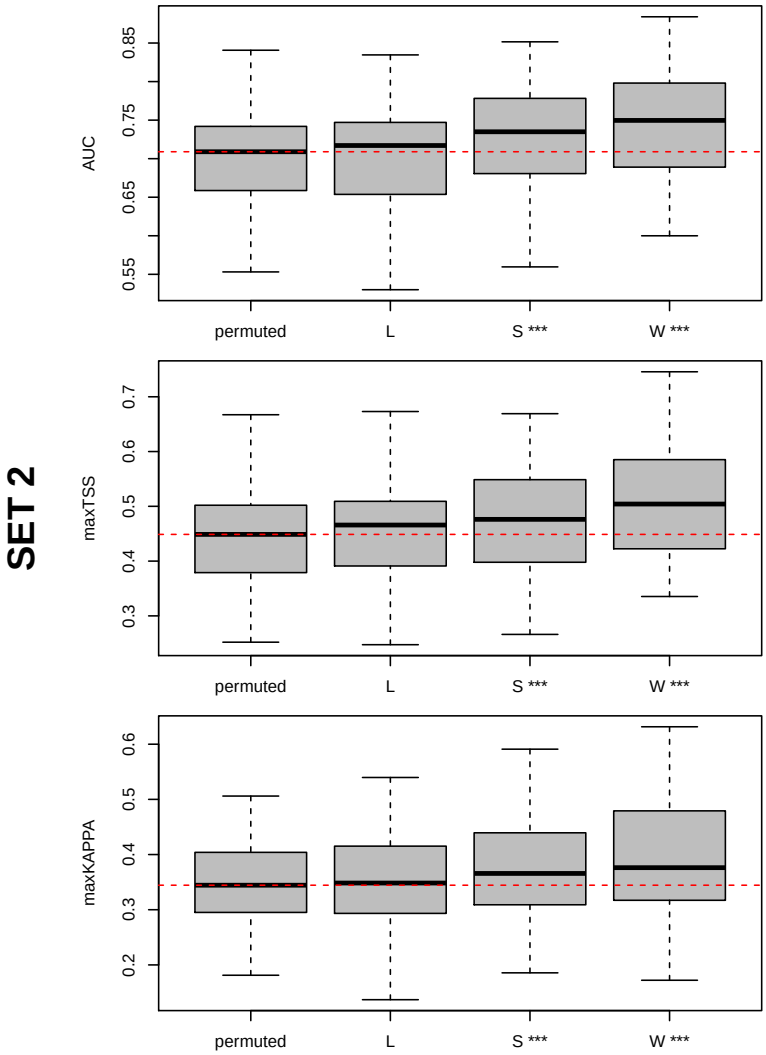


Figure 4. Like Figure 3 but with SET 2. *: p-value<0.05; **: p-value<0.01; ***: p-value<0.001

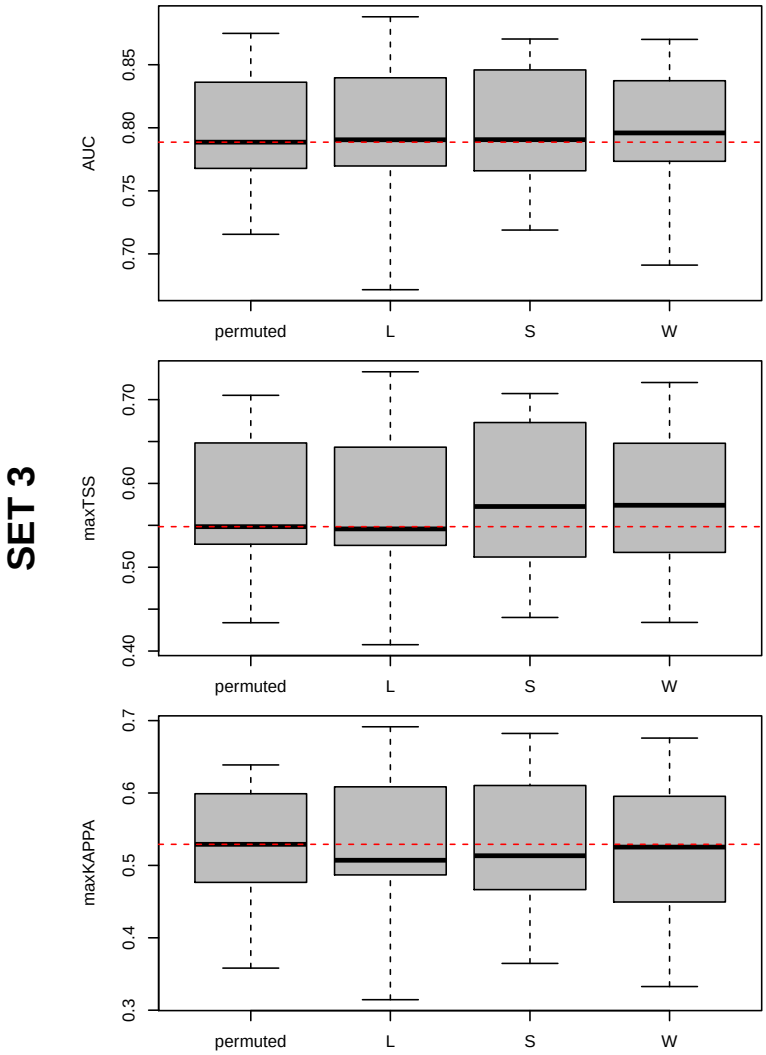


Figure 5. Like Figure 3 but with SET 3. *: p-value<0.05; **: p-value<0.01; ***: p-value<0.001

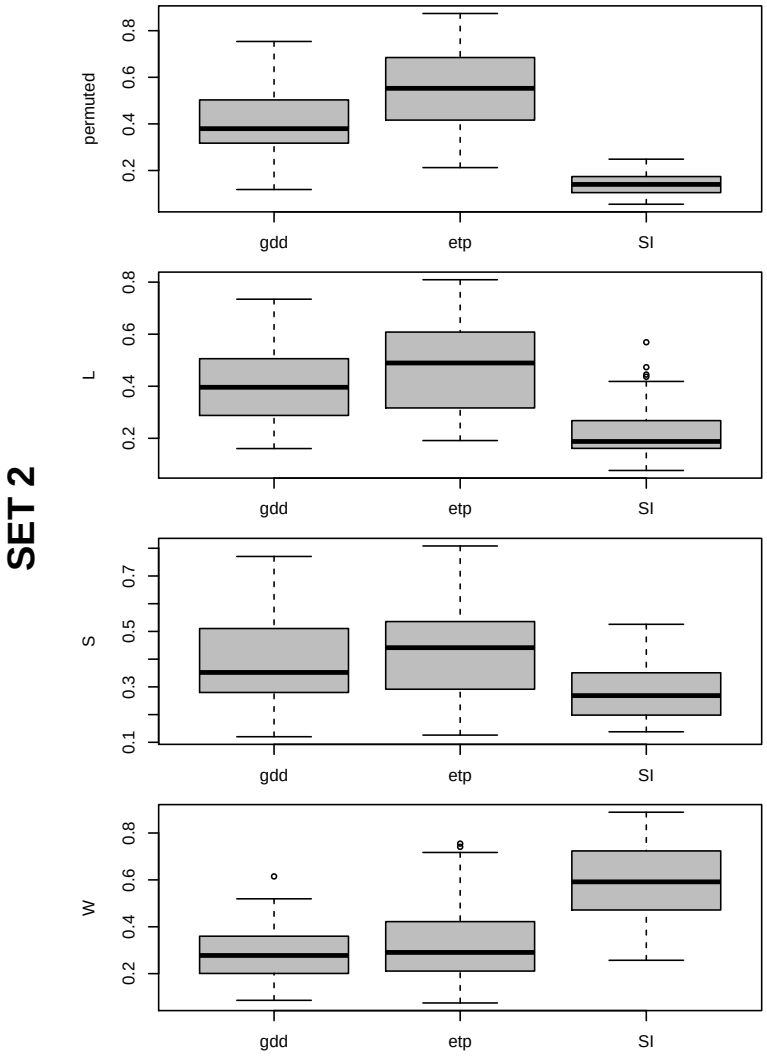


Figure 6. Variables importance in models including $SI_{permuted}$, SI_L, SI_S, SI_W respectively (y-axis) for SET 2

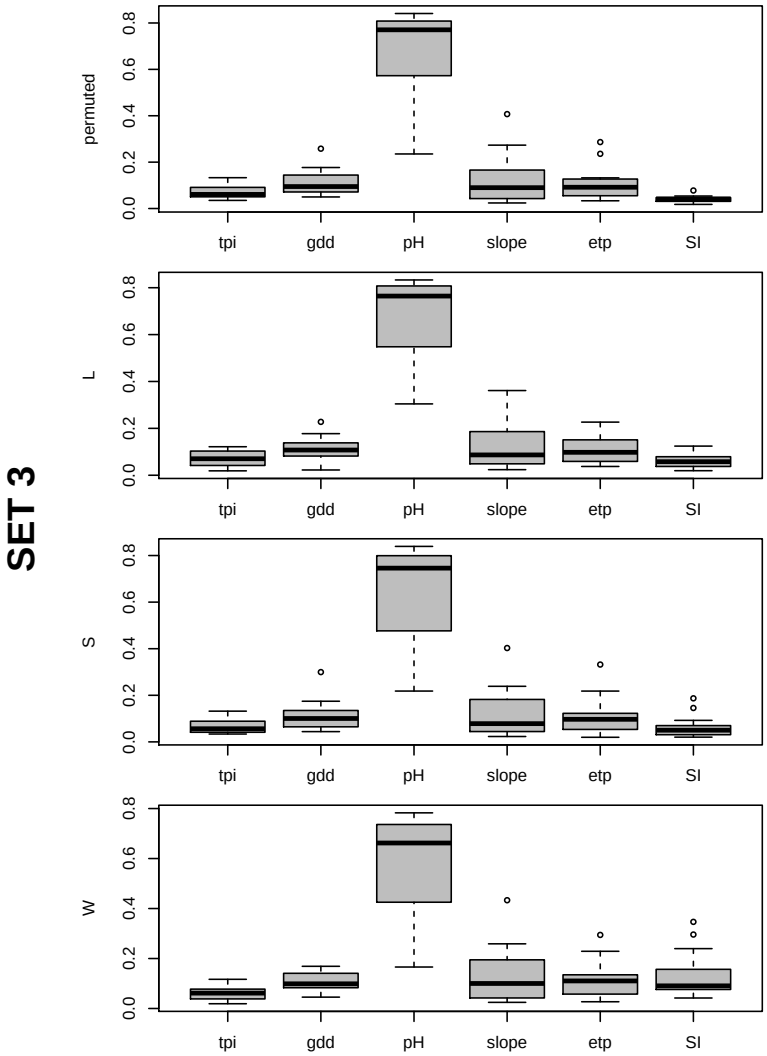


Figure 7. Like Figure 6 but with SET 3

If we look closer at individual species, we see that the proportion of species experiencing an increase of any of the three agreement metrics when the true snow indexes were used instead of $SI_{permuted}$ was quite high and was increasing as the resolution of the satellite images increases in SET 1 and 2 (Tables 12-13). In SET 3 models, the same ratio was the highest when SI_S was used (Table 14).

As the number of predictors in the models increased, the proportion of species experiencing an increase of predictive power due to the addition of the real snow indexes decreased (Tables 12-14).

The importance of the snow indexes varied between species. For the species modelled in SET 2, snow was the most important variable for 4.9 % and 9.3% when SI_L and SI_S were used respectively. This number increased to 72.3% when SI_w was used (Table 15). These species had various relationships with SI_w when we looked at the GAMs (Annexe 2), which may suggest that snow is an important predictor for species living in different

snow regimes. In SET 3 models, snow was the most important variable for none of the species when SI_L and SI_S were used, and for 2 out of 14 species when SI_W was used (Table 16).

Throughout this study, the three agreement metrics showed congruent trends (Figures 3-5, Tables 4-9 and 12-14).

3.2 Linking changes in model performance with species traits or ecology

Results from the multimodel inference analysis (MMI) showed that SLA, soil moisture index, growth form and habitat types were all important characteristics (Table 17) regarding the change in predictive performance between $SI_{permuted}$ and the other SI. We present results from SET 1 and 2 since the impact of snow was higher (Figures 3-5, Tables 4-6) as well as the number of species modelled. As the three agreement metrics gave congruent results regarding important variables (Table 17), we only explored more closely the relationship between the important variables and maxKAPPA. We also only present the results of models in which SI_W was the snow index, as these were the ones with the best predictive performance (Tables 4-5). Species living in anthropogenic meadows and pastures as well as snow beds were species for which the models were improved by the addition of the snow indexes (Figure 8). There were no clear trends between the increase in model predictive performance and SLA, the moisture index and the growth form (Figures 9-11).

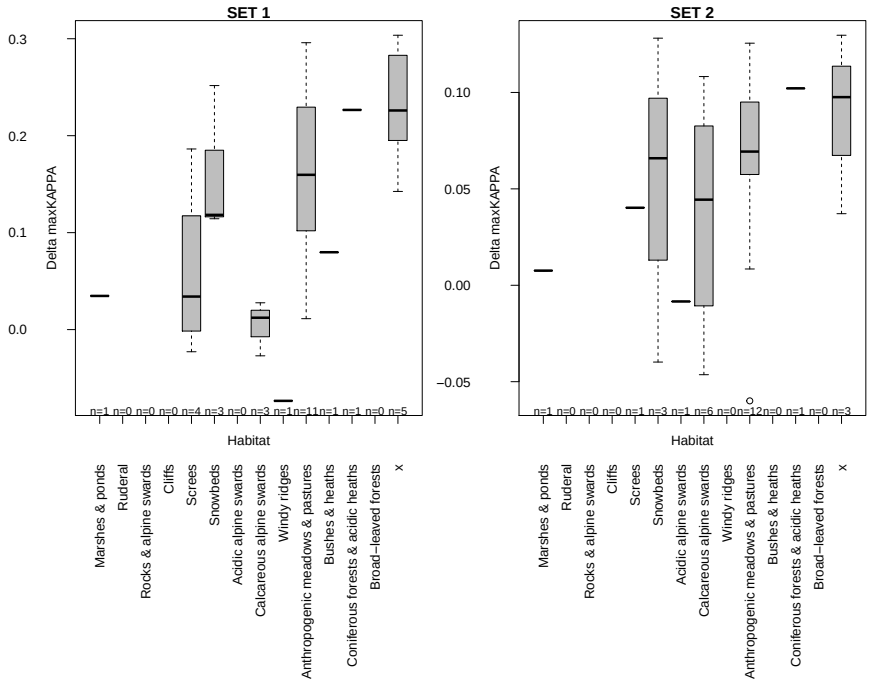


Figure 8. Relationship between the difference in predictive performance (Delta maxKAPPA) between models with SI_W and $SI_{permuted}$ in SET 1 and 2 and the habitat types

4. Discussion

In this study, we added three snow cover indexes computed from satellite images at high resolution in alpine plant species distribution models for the first time. The predictive power of the models with few environmental variables significantly increased and this was related to the resolution of the satellites images. However, this was not the case with models including more environmental variables. The predictive power was increased by snow indexes for species living in snow beds and anthropogenic meadows and pastures.

Some earlier studies have already used snow cover as a predictor to model species distribution. Gottfried *et al.* (1998) highlighted the importance of snow cover since curvature, which they used as a representation of snow cover, was a very important predictor. Guisan *et al.* (1998) as well as Guisan & Theurillat (2000) used a few aerial photographs to capture snow cover. The latter reported that the snow index performing the best out of two indexes calculated was ranked as the 5th most important predictors in both presence-absence models and abundance models out of 15 predictors. Fischer (1990), Dirnböck *et al.* (2003), Dirnböck & Dullinger (2004) and Beck *et al.* (2005) all used satellite images. In Dirnböck *et al.* (2003), snow was the 9th variable with the highest predictive power, which was low, and the 8th most important variable out of 11 variables. However, snow remained a significant predictor to model species distribution. Beck *et al.* (2005) showed that snow cover was influencing the distribution of *Dryas octopetala* in the Arctic and emphasised the possible advantage to add snow in species distribution models. Randin *et al.* (2006) and Randin *et al.* (2009b) used snow cover predictions from dynamic snow models. While the predictive power of alpine plant distribution models was improved significantly for only 19 out of 91 species, the overall model fit was raised by snow in Randin *et al.* (2009b). Finally, Carlson *et al.* (2015) and Niittynen & Luoto (2017) both used a high number of satellites images. The predictive power was increased by snow in both studies. Furthermore, in the study of Niittynen & Luoto (2017) performed in the Arctic, snow also increased the model fit. Snow was the second most important variable across all species and the most important for 36% of the species.

In the present study, the overall predictive performance was generally slightly increased in SET 3, which had the same number of predictors as the study of Niittynen & Luoto (2017), with the resolution of the satellites images. SI_W, which was the SI that improved the models the most, was the 5th overall most important predictor out of 6 and was the most important variable for 2 out of 14 species in SET 3. The lower improvement of SI_W in this study compared to Niittynen & Luoto (2017) may be explained by the low number of sampled plots (254 plots) and consequently only 14 vascular plant species could be modelled in SET 3. In contrast, Niittynen & Luoto (2017) used 1200 vegetation

inventories to model the distribution of 273 species comprising lichens, mosses and vascular plants. Besides, those three groups have different relationships with snow, which might make SDMs more likely to display the importance of the incorporation of snow cover in the models. Moreover, while Niittynen & Luoto (2017) used a total of 124 satellite images from 1984 to 2016 to produce one snow cover metric, we used 34 Landsat images, 23 Sentinel images and 10 Worldview images over a period of 3 years to compute our three snow indexes. Finally, we did not relate the snow cover to the date of the year for each satellite images to model the length of the snow cover.

The species for which SI_W was the most important variable in SET 2 had various GAM response curves with SI_w, which might imply that snow is an important predictor for species living in different snow regimes. This is in accordance with results found by Carlson *et al.* (2015) and Niittynen & Luoto (2017).

The importance of the three snow indexes (i.e. not permuted) was not as high as expected compared to the models with the permuted index. However, the importance increased with the resolution of the satellites images. Besides, the predictive performance was increasing with the resolution of the satellites images in all sets of models and it was significantly greater when SI_w was used than with SI_S or SI_L in SET 1 and 2 models. Finally, the proportion of species experiencing an increase of the agreement metrics was also increasing with the resolution of the satellite images in SET 1 and 2. It can therefore be recommended to use high-resolution satellite images for mapping snow cover in further studies trying to incorporate snow cover in SDMs in mountain areas, where environmental conditions can change rapidly over short distances due to the complex topography. The fact that snow was not as important as expected compared to the five environmental variables in SET 3 models might be due to the fact that the resolution was already high for those environmental variables. Since snow distribution is greatly influenced by the microtopography, the high resolution of the environmental variables might allow them to capture the microtopographic climatic variations and thus snow cover.

A great limitation was the relatively small sample size in all sets of models, particularly when we wanted to add more predictors in our models. Even though Worldview images could not have been acquired for a larger area in this study, Landsat8 and Sentinel images are currently available for the whole Vaud Alps area, where more vegetation plots are also available. In further studies it could be useful to have access to Worldview images for the whole Vaud Alps, although this is still uneasy to do in practice.

Climate is meant to play the major role in SDMs (Shao & Halpin 1995). For instance, temperature was shown to be the most important variable in Buri *et al.* (2014), Dubuis *et al.* (2013a) and Pradervand *et al.* (2014) and soil pH was the second most important

variable in the two studies where it was included. In contrast, in the present study, soil pH was by far the most important variable in SET 3. Climatic factors (gdd and etp) were still more important than topography (apart for gdd when SI_W is used), which indicates that climate had still an important effect in our models. This difference of importance of pH might be explained by the fact that the three previously mentioned studies took place in the Vaud Alps where elevation ranges from 375 to 3210 m asl, whereas our study took place in only a reduced part of this area, where elevation ranges from 1070 to 2850 m asl. Furthermore, the climate in the Vallon de Nant, the lowest part of our study area, is greatly influenced by the Martinets Glacier with cold air flowing down. Consequently, our results concerning the variables importance might reflect the important variables impacting species distribution models when working towards high elevations.

5. Perspectives

In the present study, we did not model the daily presence of snow over the growing season since it would have been too tedious for a master thesis. In further studies this could be done by linking presence and absence of snow cover from satellite images to topographic variables and the day of the year in a generalised additive model (GAM), as explored in Carlson *et al.* (2015), or use general linear model (GLM) to link snow cover with the day of the year as done in Niittynen & Luoto (2017). Dedieu *et al.* (2016) used satellites images at high temporal and spatial resolution to obtain the date of snowmelt. From any of those techniques, the number of days with snow in the plant growing season for each pixel could be calculated.

In this preliminary study, we calculated the environmental variables only at the locations of the sampled vegetation plots. However, we would suggest to first compute those environmental variables for the whole Vaud Alps in the raster format and then extract the data at the location of the vegetation plots for the modelling procedure. This could allow to make projection maps for the species that had their model improved by snow and we could also model species distributions in the future if some snow simulations for the future becomes available.

Projections would be particularly interesting regarding climate change. Indeed, the increase in mean temperature in Switzerland has been twice that of the Northern hemisphere during the 20th century and the pace of the increase has been quicker between 1975 and 2004 (Rebetez & Reinhard 2008). Moreover, it is forecasted that the mean global temperature will increase by 0.3 to 0.7°C for the period 2016-2035 compared to 1986-2005 (PCC Report 2013).

As a result, warming will have a high impact on snow cover: snow depth is forecasted to decrease by 2085, despite the higher amount of precipitation. This might provoke a decrease of the length of the snow season, and thus a longer growing season for plants, even at high elevations (Appenzeller *et al.* 2014). However, not all plant species are able to flower earlier since some plants require a minimal day length to start the growing season and this day length differs between species (Keller & Körner 2003). As a consequence, the species composition may change due to the earlier snowmelt. The species that can flower earlier will become more frequent (Keller & Körner 2003; Keller *et al.* 2005).

7. Acknowledgements

I deeply thank all the members of the Guisan group for their precious help. My special thanks to Heidi Mod for her help with coding.
Thanks to Grégoire Mariethoz for his help with the calculations of the snow indexes and Pascal Vittoz for his advises on the choice of species traits.
I am thankful to François Bonnet to have provided me with accommodation during part of the field work. Thanks to him and Patrice Descombes for their help with plant identification. I also would like to thank all my internship students for their help during the field work.
I'm thankful to the IntegrAlp project as my work was part of this big project, which provided the satellite images used here.
Finally, I'm particularly grateful to my director, Antoine Guisan for his support and advice and to my supervisor Carmen Cianfrani for her great help and support.

Bibliography

Albert CH, Thuiller W, Lavorel S, Davies ID, Garbolino E (2008) Land-use change and subalpine tree dynamics: colonization of Larix decidua in French subalpine grasslands. *Journal of Applied Ecology*, **45**, 659–669.

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models : prevalence , kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

Appenzeller C, Fischer EM, Fuhrer J *et al.* (2014) *CH2014 – Impacts Toward quantitative scenarios of climate change impacts in Switzerland*. Bern.

Bartoń K (2017) MuMIn: Multi-Model Inference. R package version 1.40.0.

Beck PSA, Kalmbach E, Joly D, Stien A, Nilsen L (2005) Modelling local distribution of an Arctic dwarf shrub indicates an important role for remote sensing of snow cover. *Remote Sensing of Environment*, **98**, 110–121.

Billings WD, Bliss LC (1959) An Alpine Snowbank Environment and Its Effects on Vegetation, Plant Development, and Productivity. *Ecology*, **40**, 388–397.

Björk RG, Molau U (2007) Ecology of Alpine Snowbeds and the Impact of Global Change. *Arctic, Antartic and Alpine Research*, **39**, 34–43.

Breiman L (2001) Random Forests. *Machine Learning*, **45**, 5–32.

Buri A, Cianfrani C, Pinto-Figueroa E *et al.* (2017) Soil factors improve predictions of plant species distribution in a mountain environment. *Progress in Physical Geography*, **41**, 703–722.

Buri A, Pradervand J-N, Pinto-Figueroa E *et al.* (2014) Predicting plant distribution: does edaphic factor matter. Master thesis.

Burnham KP, Anderson DR, Huyvaert KP (2011) AIC model selection and multimodel inference in behavioral ecology : some background , observations , and comparisons. *Behavioral Ecology and Sociobi- ology*, **65**, 23–35.

Carlson BZ, Choler P, Renaud J, Dedieu J, Thuiller W (2015) Modelling snow cover duration improves predictions of functional and taxonomic diversity for alpine plant communities. *Annals of Botany*, **116**, 1023–1034.

Chao A, Chazdon RL, Colwell RK, Shen T-J (2005) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters*, **8**, 148–159.

Cohen J (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, **20**, 37–46.

Cornelissen JHC, Lavorel S, Garnier E *et al.* (2003) A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian Journal of Botany*, **51**, 335–380.

Dedieu J, Carlson BZ, Bigot S *et al.* (2016) On the Importance of High-Resolution Time Series of Optical Imagery for Quantifying the Effects of Snow Cover Duration on Alpine Plant Habitat. *Remote Sensing*, **8**.

Dedieu J, Randin C, Zappa M (2012) Validation par télédétection spatiale de l'enneigement dans les Alpes autrichiennes pour l'approvisionnement en eau de la ville de Vienne. In: *25ème Colloque de l'Association Internationale de Climatologie, Grenoble, France*, pp. 207–212.

Dirnböck T, Dullinger S (2004) Habitat distribution models, spatial autocorrelation, functional traits and dispersal capacity of alpine plant species. *Journal of Vegetation Science*, **15**, 77–84.

Dirnböck T, Dullinger S, Grabherr G (2003) A regional impact assessment of climate and

land-use change on alpine vegetation. *Journal of Biogeography*, **30**, 401–417.
Dormann CF, Elith J, Bacher S *et al.* (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.
Dozier J (1989) Spectral signature of alpine snow cover from the Landsat Thematic Mapper. *Remote Sensing of Environment*, **28**, 9–22.
Dray S, Dufour AB (2007) The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, **22**, 1–20.
Dubuis A, Giovanettina S, Pellissier L *et al.* (2013a) 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, Rossier L, Pottier J *et al.* (2013b) Predicting current and future spatial community patterns of plant functional traits. *Ecography*, **36**, 1158–1168.
Ellenberg H, Weber HE, Düll R *et al.* (1991) Zeigerwerte von Pflanzen in Mitteleuropa. *Scripta Geobotanica*, **18**, 1–248.
Engler R, Randin CF, Thuiller W *et al.* (2011) 21st century climate change threatens mountain flora unequally across Europe. *Global Change Biology*, **17**, 2330–2341.
Farooq S Spectral reflectance of land covers.
Fischer HS (1990) Simulating the distribution of plant communities in an Alpine landscape. *Coenoses*, **5**, 37–43.
Friedman JH, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting. *Annuals of Statistics*, **28**, 337–374.
Gallien L, Douzet R, Pratte S, Zimmermann NE, Thuiller W (2012) Invasive species distribution models - how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, **21**, 1126–1136.
Gottfried M, Pauli H, Futschik A *et al.* (2012) Continent-wide response of mountain vegetation to climate change. *Nature Climate Change*, **2**, 111–115.
Gottfried M, Pauli H, Grabherr G (1998) Prediction of vegetation patterns at the limits of plant life: a new view of the alpine-nival ecotone. *Arctic and Alpine Research*, **30**, 207–221.
Guisan A, Theurillat J-P (2000) Equilibrium modeling of alpine plant distribution: how far can we go? *Phytocoenologia*, **30**, 353–384.
Guisan A, Theurillat J-P, Kienast F (1998) Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science*, **9**, 65–74.
Guisan A, Thuiller W, Zimmermann NE (2017) *Habitat suitability and distribution models: with applications in R* (CU Press, Ed,). Cambridge.
Harrell FE (2001) *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Springer-Verlag, New York.
Hastie TJ, Tibshirani RJ (1986) Generalized additive models (C and Hall, Ed,). *Statistical Sciences*, **1**, 297–318.
Hirzel A, Guisan A (2002) Which is the optimal sampling strategy for habitat suitability modelling. *Ecological Modelling*, **157**, 331–341.
Hüsler F, Jonas T, Riffler M, Musial JP, Wunderle S (2014) A satellite-based snow cover climatology (1985–2011) for the European Alps derived from AVHRR data. *The Cryosphere*, **8**, 73–90.
Hutchinson GE (1957) Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.
Jonas T, Rixen C, Sturm M, Stoeckli V (2008) How alpine plant growth is linked to snow cover and climate variability. *Journal of Geophysical Research*, **113**, 1–10.
Keller F, Goyette S, Beniston M (2005) Sensitivity analysis of snow cover to climate

change scenarios and their impact on plant habitats in alpine terrain. *Climatic Change*, **71**, 299–319.
Keller F, Körner C (2003) The Role of Photoperiodism in Alpine Plant Development. *Arctic, Antartic, and Alpine Research*, **35**, 361–368.
Körner C (2003) *Alpine plant life: functional plant ecology of high mountain ecosystems*. Springer, Berlin [etc.].
Landolt E, Bäumler B, Erhardt A *et al.* (2010) *Flora indicativa: Ecological indicator values and biological attributes of the Flora of Switzerland and the Alps* (Haupt, Ed,). Bern, CH.
Lauber K, Wagner G, Gygax A (2012) *Flora Helvetica: flore illustrée de Suisse*. Haupt, Berne [etc.].
Lavorel S, Garnier E (2002) Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Functional Ecology*, **16**, 545–556.
Macander MJ, Swingley CS, Joly K, Raynolds MK (2015) Landsat-based snow persistence map for northwest Alaska. *Remote Sensing of Environment*, **163**, 23–31.
McCullagh P, Nelder JA (1989) *Generalized Linear Models, Second Edition* (Chapman & Hall, Ed,). London.
Niittynen P, Luoto M (2017) The importance of snow in species distribution models of arctic vegetation. *Ecography*, **40**, 1–13.
Palacios D, Sánchez-Colomer MG (1997) The distribution of high mountain vegetation in relation to snow cover : Peñalara , Spain. *Catena*, **30**, 1–40.
Park S-H, Lee M-J, Jung H-S (2016) Journal of Atmospheric and Solar-Terrestrial Physics Spatiotemporal analysis of snow cover variations at Mt . Kilimanjaro using multi-temporal Landsat images during 27 years. *Journal of Atmospheric and Solar-Terrestrial Physics*, **143–144**, 37–46.
PCC Report (2013) *Working Group 1 Contribution to the IPCC Fifth Assessment Report, Climate Change 2013: The Physical Science Basis*. Cambridge, UK.
Pignatti S (2005) Valori di bioindicazione delle piante vascolari della flora d'Italia. *Braun-Blanquetia*, **39**, 1–97.
Pottier J, Malenovsky Z, Psomas A *et al.* (2014) Modelling plant species distribution in alpine grasslands using airborne imaging spectroscopy. *Biology Letters*, **10**, 1–4.
Pradervand J-N, Dubuis A, Pellissier L, Guisan A, Randin C (2014) Very high resolution environmental predictors in species distribution models: Moving beyond topography? *Progress in Physical Geography*, **38**, 79–96.
R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Randin CF, Dedieu J, Zappa M, Long L, Dullinger S (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**, 1181–1193.
Randin CF, Dirnböck T, Dullinger S *et al.* (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.
Randin CF, Engler R, Pearman PB, Vittoz P, Guisan A (2010) Using Georeferenced Databases to Assess the Effect of Climate Change on Alpine Plant Species and Diversity. In: *Data mining for global trends in mountain biodiversity* (ed Spehn E.M. KC), pp. 149–163.
Randin CF, Jaccard H, Vittoz P, Yoccoz NG, Guisan A (2009a) Land use improves spatial predictions of mountain plant abundance but not presence-absence. *Journal of*

Vegetation Science, **20**, 996–1008.
Randin CF, Liston GE, Vittoz P, Guisan A (2009b) Introduction of Snow and Geomorphic Disturbance Variables into Predictive Models of Alpine Plant Distribution in the Western Swiss Alps. *Arctic, Antarctic, and Alpine Research*, **41**, 347–361.
Rebetez M, Reinhard M (2008) Monthly air temperature trends in Switzerland 1901-2000 and 1975-2004. *Theoretical and Applied Climatology*, **91**, 27–34.
Ridgeway G (1999) The state of boosting. *Computing Science and Statistics*, **31**, 172–181.
Satir O (2016) Comparing the satellite image transformation techniques for detecting and monitoring the continuous snow cover and glacier in Cilo mountain chain Turkey. *Ecological Indicators*, **69**, 261–268.
Schimel JP, Bilbrough C, Welker JM (2004) Increased snow depth affects microbial activity and nitrogen mineralization in two Arctic tundra communities. *Soil Biology & Biochemistry*, **36**, 217–227.
Schöb C, Kammer PM, Choler P, Veit H (2009) Small-scale plant species distribution in snowbeds and its sensitivity to climate change. *Plant Ecology*, **200**, 91–104.
Shao G, Halpin PN (1995) Climatic controls of eastern North American coastal tree and shrub distributions. *Journal of Biogeography*, **22**, 1083–1089.
Swets JA (1988) Measuring the Accuracy of Diagnostic Systems. *Science*, **240**, 1285–1293.
Thuiller W, Georges D, Engler R, Breiner F (2016) biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7.
Thuiller W, Lafourcade B, Araujo M (2010) Presentation Manual for BIOMOD. , 35p.
Thuiller W, Lavorel S, Araujo MB, Sykes MT, C. PI (2005) Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 8245–8250.
Wahren C-HA, Walker MD, Bret-Harte M. (2005) Vegetation responses in Alaskan arctic tundra after 8 years of a summer warming and winter snow manipulation experiment. *Global Change Biology*, **11**, 537–552.
Walker DA, Halfpenny JC, Walker MD, Wessman CA (1993) Long-term Studies of Snow-Vegetation Interactions. *BioScience*, **43**, 287–301.
Wipf S, Stoeckli V, Bebi P (2009) Winter climate change in alpine tundra : plant responses to changes in snow depth and snowmelt timing. *Climatic Change*, **94**, 105–121.
Wood SN (2017) *Generalized Additive Models: An Introduction with R (2nd edition)*. Chapman & Hall/CRC, Boca Raton [etc.].

Annexe 1

Quelques tests pour identifier la neige sur des images WorldView: (par Pr. Grégoire Mariéthoz)

Plusieurs problèmes ont été identifiés:

1) Sur les images WorldView, il n'y a pas de bande SWIR à 1.6 micromètre qui est habituellement utilisée pour calculer le NDSI.
2) Je pensais regarder uniquement le visible, mais il y a beaucoup d'ombres qui font qu'il est impossible d'utiliser une seule bande et de mettre un seuil dessus. Il faut utiliser des rapports de bandes qui sont robustes aux ombres.
3) Apparemment il y a un problème de capteur sur certaines bandes WorldView où on a des sortes de lignes/zones obscures. En particulier la bande 7 qui a parfois des valeurs proches de zéro à des endroits pas logiques.
4) Pas de chance, c'est justement la bande 7 qui arrive bien à distinguer la neige.

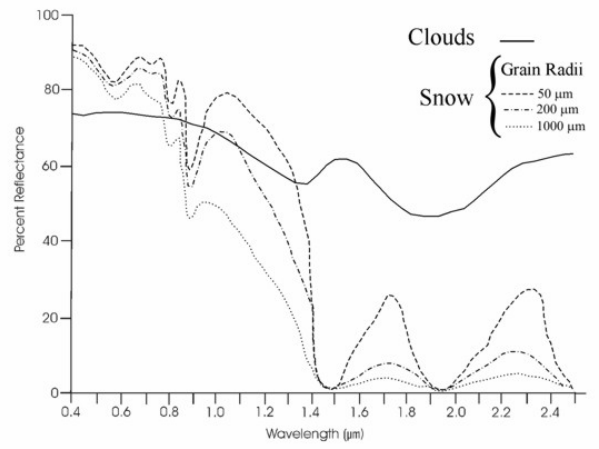
A la fin, j'ai trouvé cette combinaison de bandes qui marche plutôt bien:

$$SI = \frac{B5 - \max(B7, 75)}{B5 + \max(B7, 75) + 0.01}$$

Ensuite, j'attribue la neige à tout ce qui vaut

$SI > -0.2$

La logique de la formule est de trouver quelque chose qui ressemble à la bande 1.6 micromètre utilisée pour le NDSI. Quand on regarde la courbe de réflectance de la neige (image ci-dessous) on voit qu'à environ 0.8-0.9 micromètres (bande 7) il y a un petit creux qui ressemble (en moins prononcé) à celui de 1.6 micromètre.



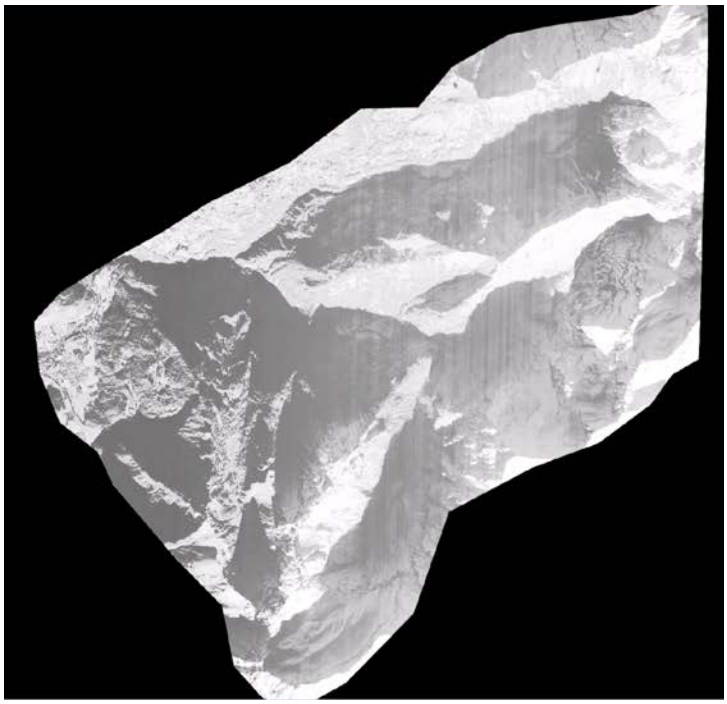
Snow and clouds reflectance curves (Farooq)

Dans la formule, le 75 corrige les endroits où la bande 7 a des valeurs trop faibles. Le 0.01 est un tout petit nombre pour éviter des éventuelles divisions par zéro. J'ai mis le seuil à -0.2, et ça a l'air de marcher pas mal.

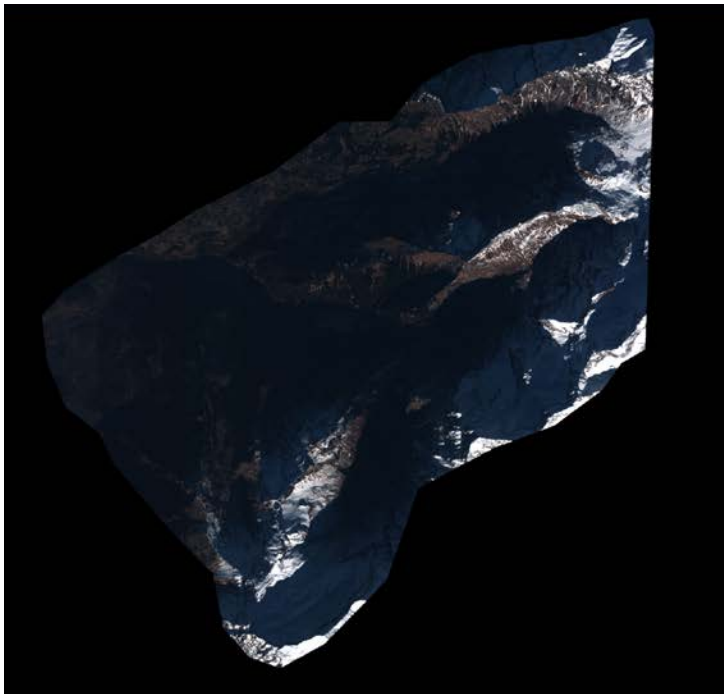
Julie : j'ai encore réfléchi et à mon avis pour tes accumulations de neige, ce serait mieux d'utiliser juste une variable binaire avec des 0 quand il n'y a pas de neige et 1 quand il y en a. C'est beaucoup plus propre.

Ci-dessous quelques captures d'écran. J'ai fait tous mes tests sur l'image du 4 décembre 2016.

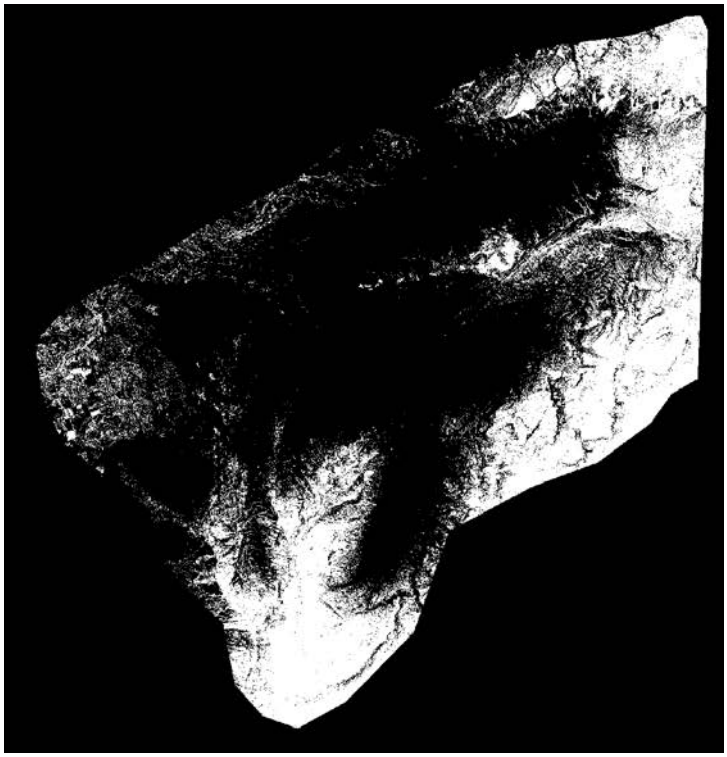
Artefacts sur la bande 7 (mis en évidence par une égalisation d'histogramme).



L'image en couleurs vraies (RGB). On voit les ombres qui nous dérangent.

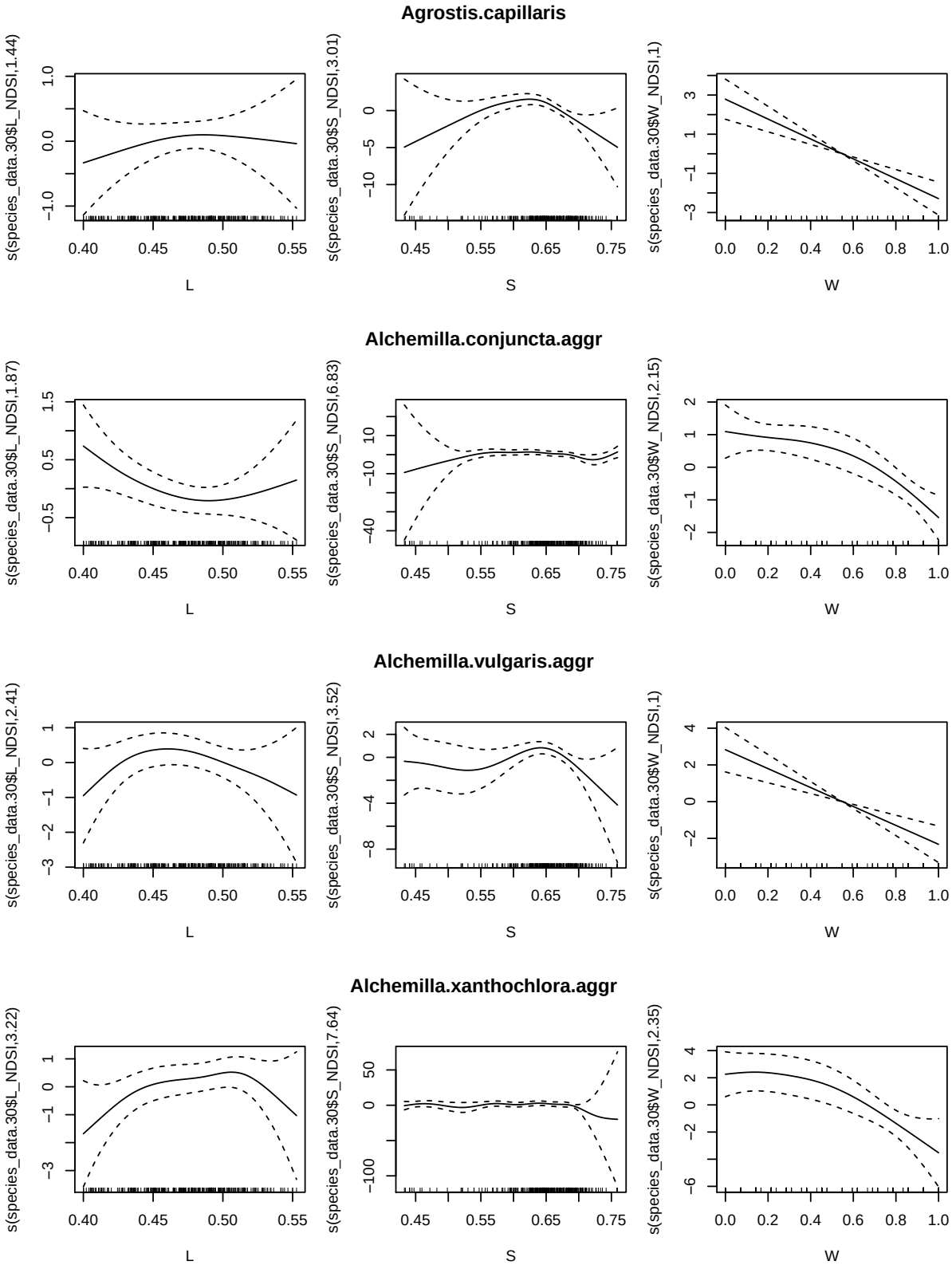


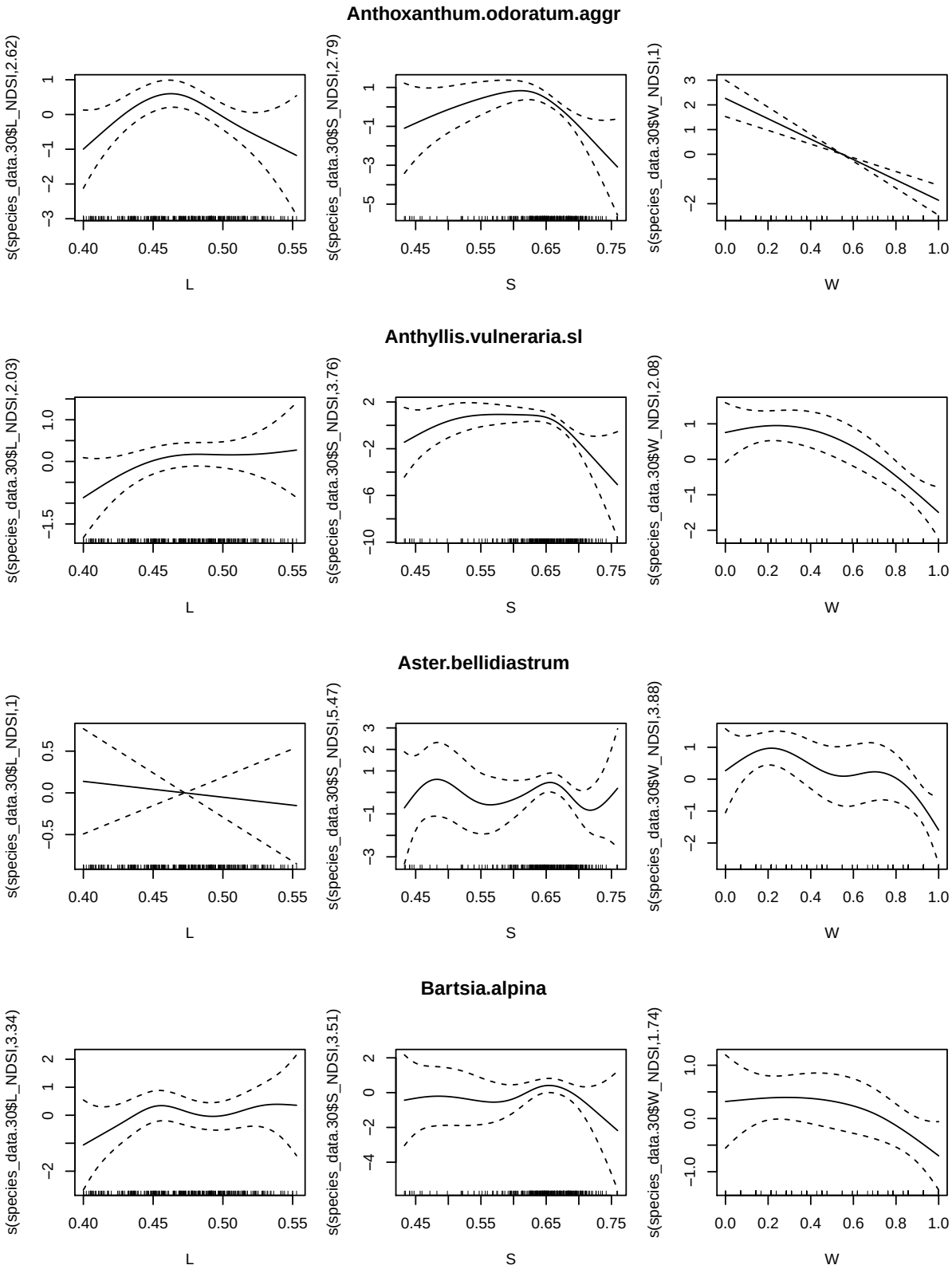
L’indice SI calculé comme ci-dessus, après application du seuil. Ça correspond grosso modo à la neige qu’on voit sur l’image RGB, sauf une zone urbaine à gauche de l’image qui réagit un peu différemment.

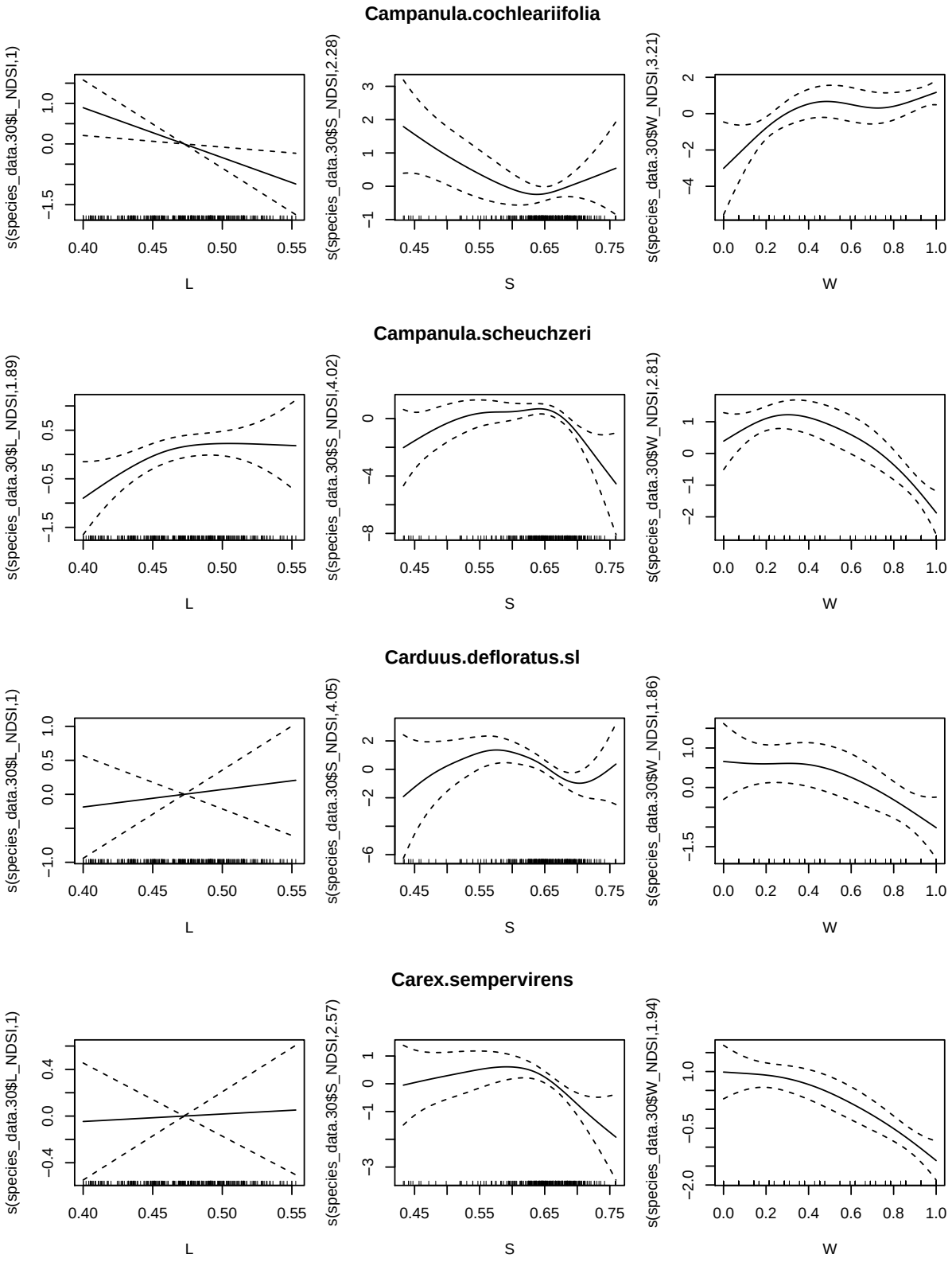


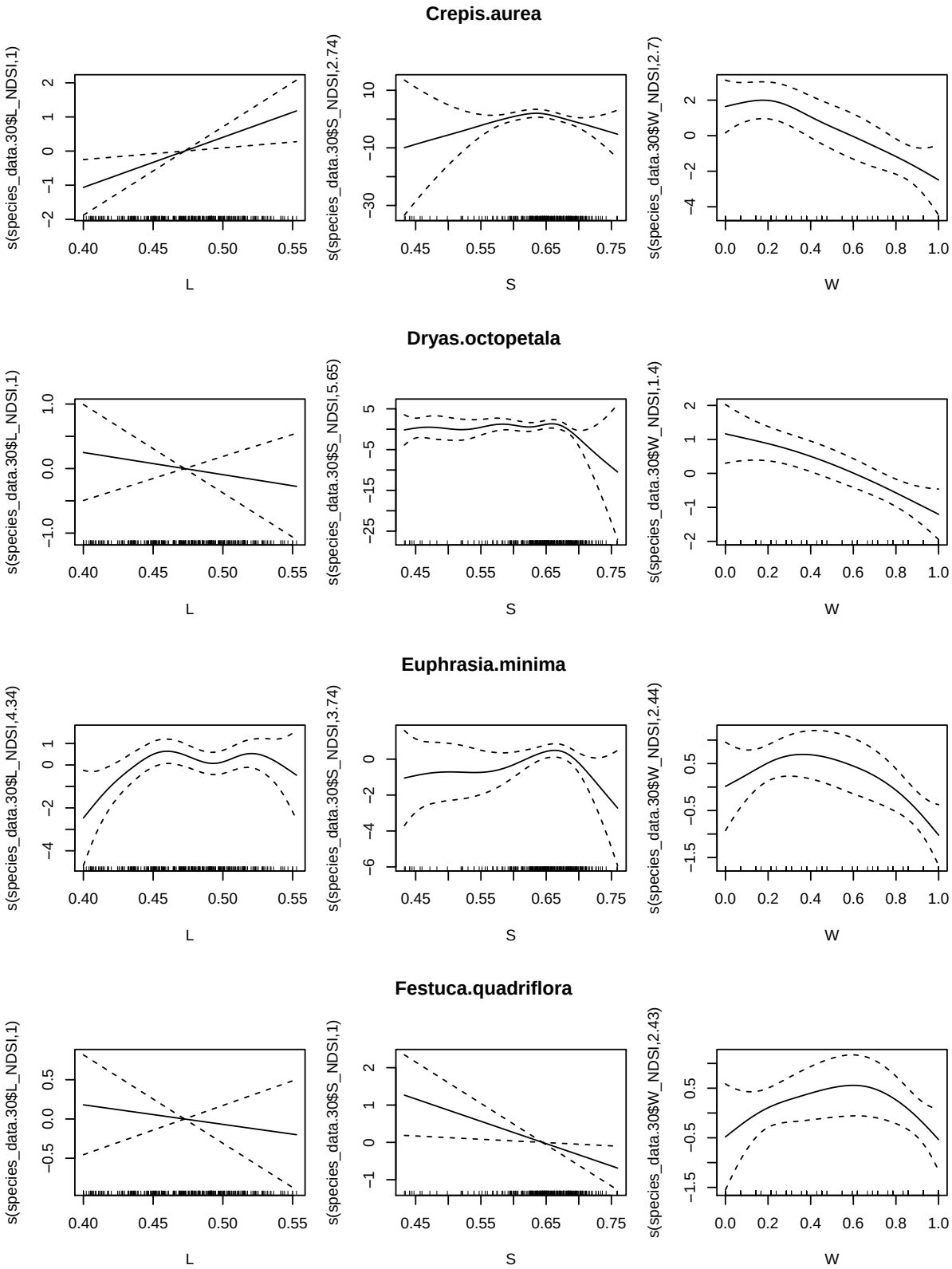
Annexe 2

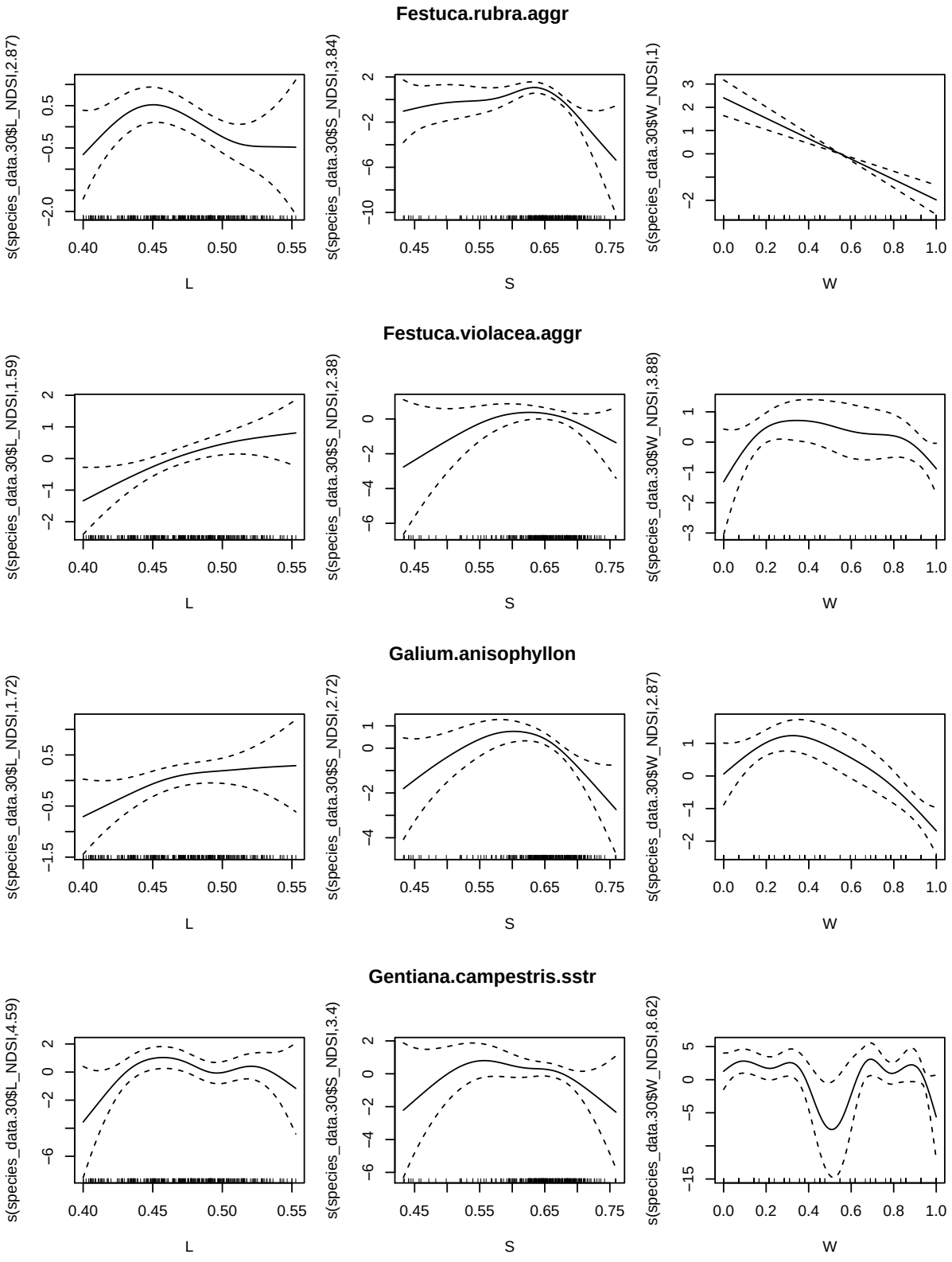
Univariate GAM showing the relationship between the species and the three snow indexes SI_L, SI_S, SI_W respectively (x-axis).

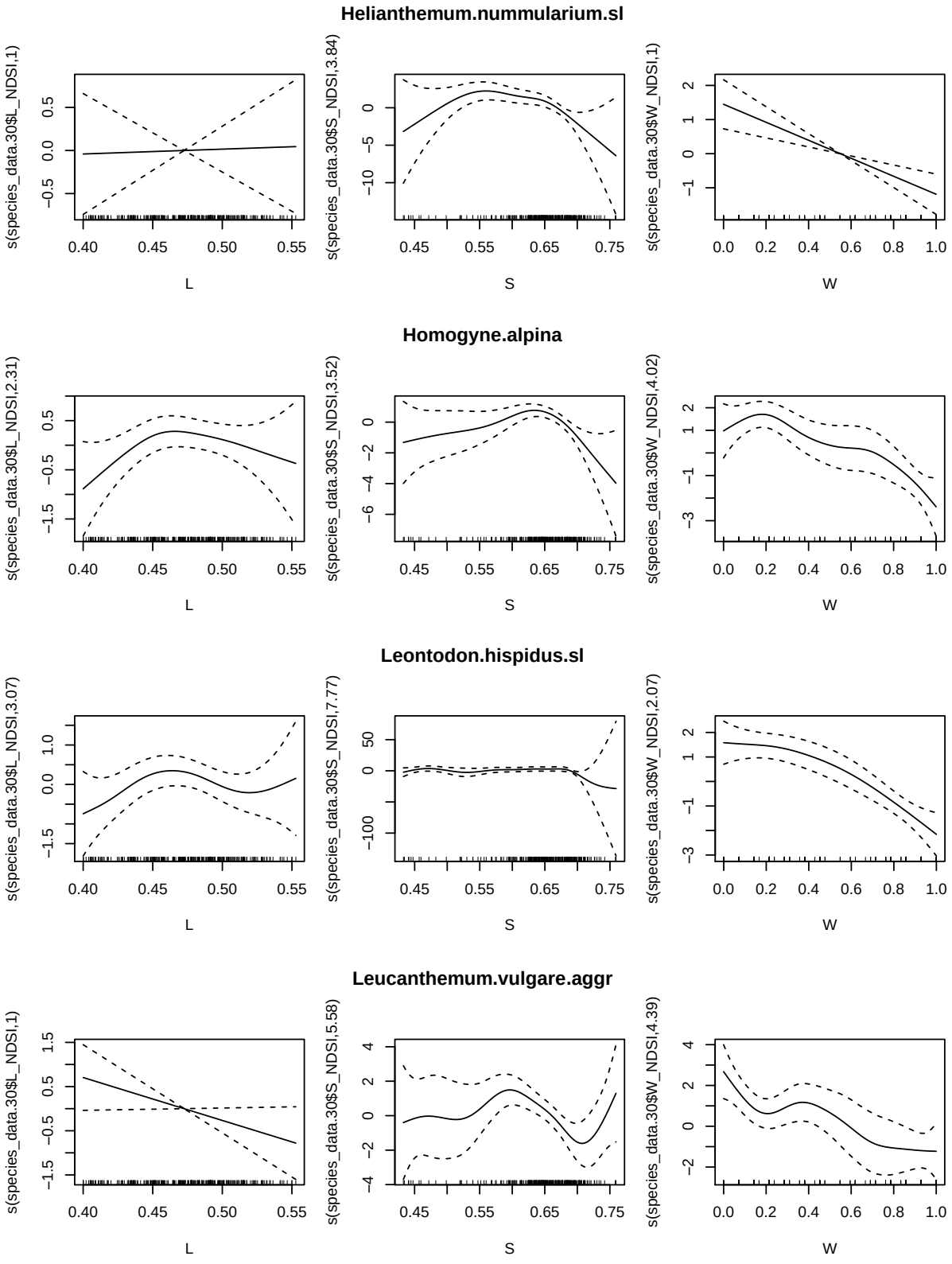


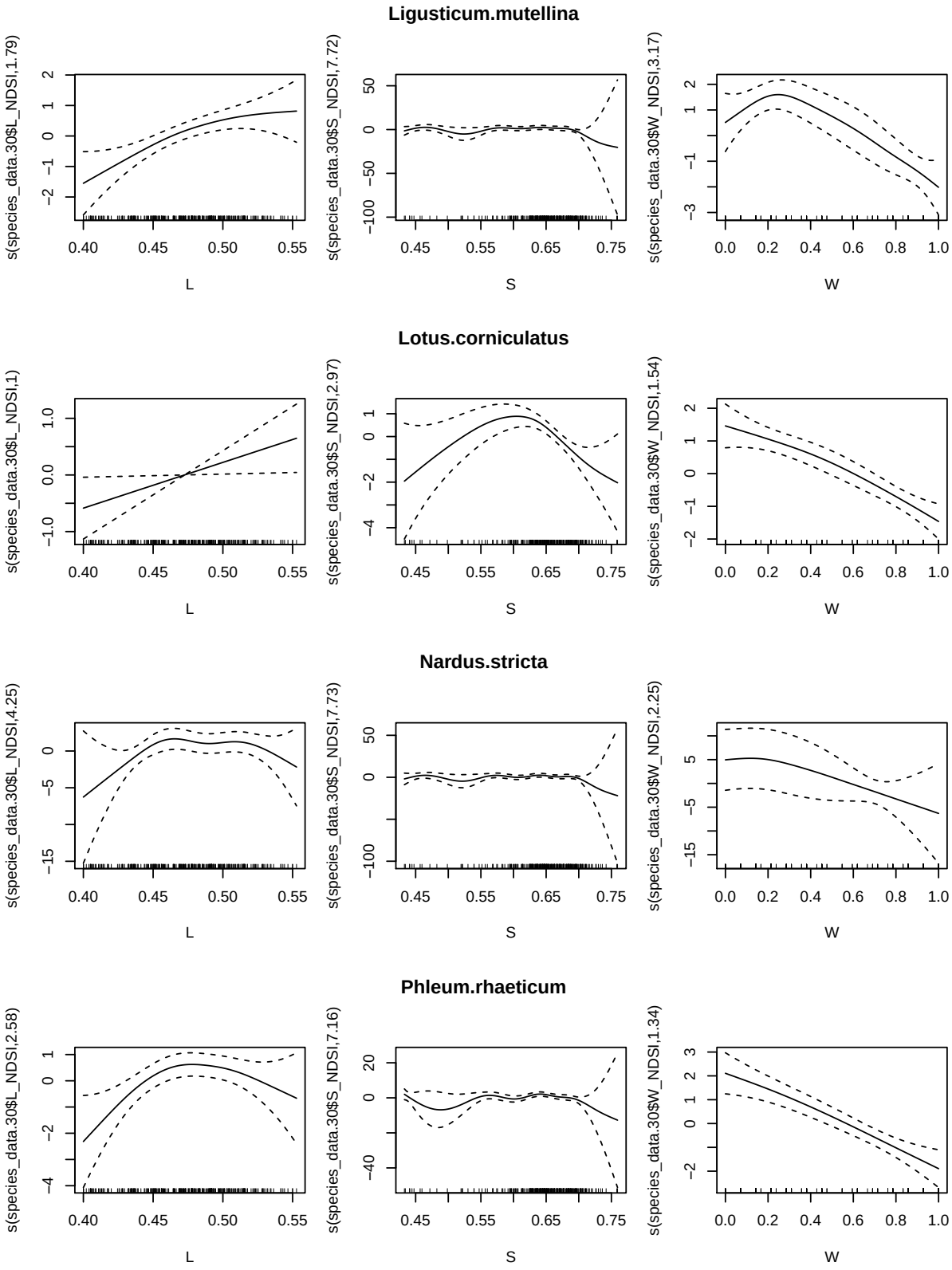

Anthoxanthum.odoratum.aggr
s(species_data.30$L_NDSI,2.62)
s(species_data.30$S_NDSI,2.79)
s(species_data.30$W_NDSI,1)
L
S
W
Anthyllis.vulneraria.sl
s(species_data.30$L_NDSI,2.03)
s(species_data.30$S_NDSI,3.76)
s(species_data.30$W_NDSI,2.08)
L
S
W
Aster.bellidiastrum
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,5.47)
s(species_data.30$W_NDSI,3.88)
L
S
W
Bartsia.alpina
s(species_data.30$L_NDSI,3.34)
s(species_data.30$S_NDSI,3.51)
s(species_data.30$W_NDSI,1.74)
L
S
W

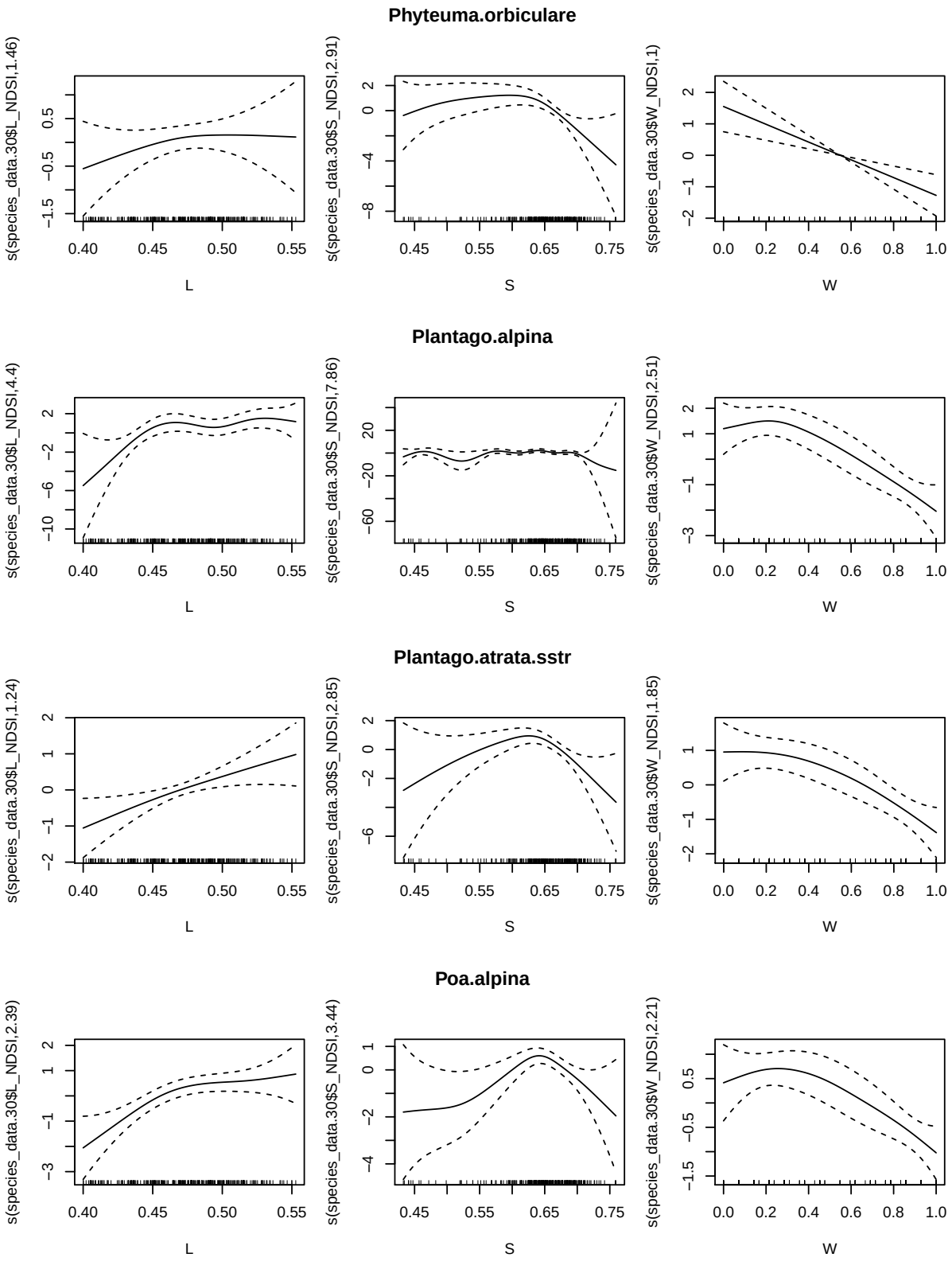

Campanula.cochleariifolia
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,2.28)
s(species_data.30$W_NDSI,3.21)
L
S
W
Campanula.scheuchzeri
s(species_data.30$L_NDSI,1.89)
s(species_data.30$S_NDSI,4.02)
s(species_data.30$W_NDSI,2.81)
L
S
W
Carduus.defloratus.sl
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,4.05)
s(species_data.30$W_NDSI,1.86)
L
S
W
Carex.sempervirens
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,2.57)
s(species_data.30$W_NDSI,1.94)
L
S
W

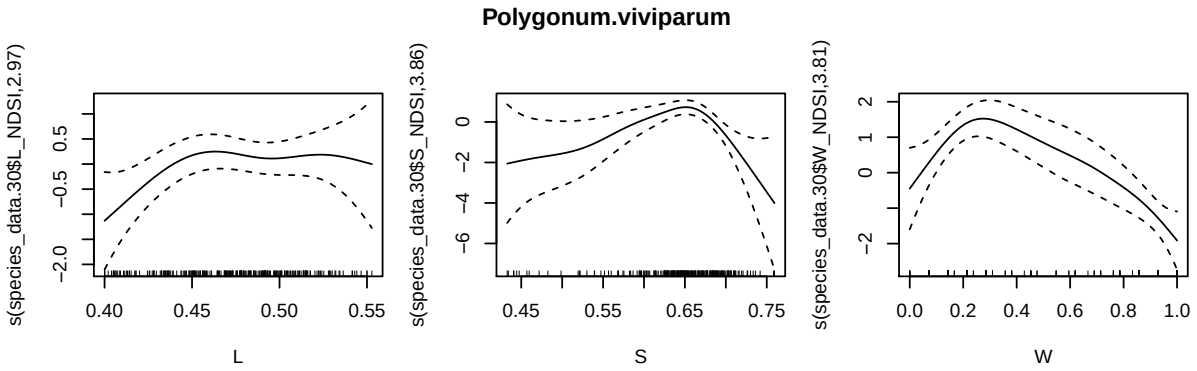

Crepis.aurea
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,2.74)
s(species_data.30$W_NDSI,2.7)
L
S
W
Dryas.octopetala
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,5.65)
s(species_data.30$W_NDSI,1.4)
L
S
W
Euphrasia.minima
s(species_data.30$L_NDSI,4.34)
s(species_data.30$S_NDSI,3.74)
s(species_data.30$W_NDSI,2.44)
L
S
W
Festuca.quadriflora
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,1)
s(species_data.30$W_NDSI,2.43)
L
S
W

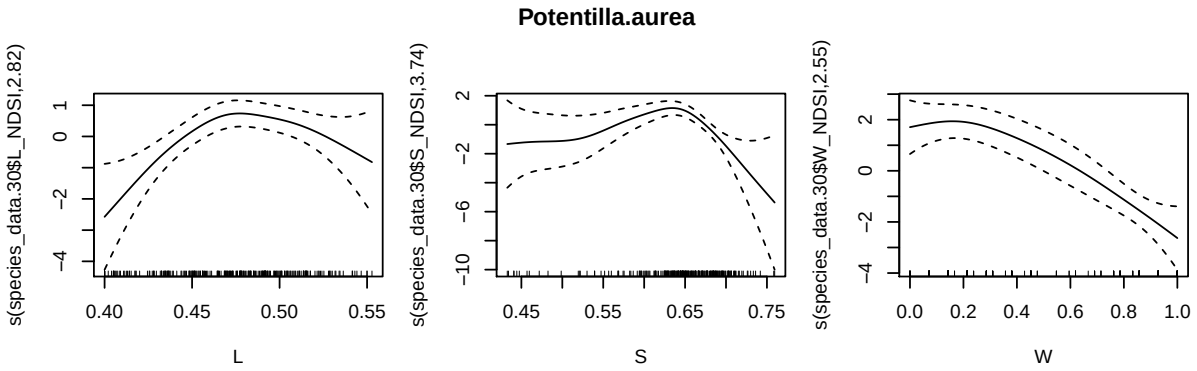

Festuca.rubra.aggr
s(species_data.30$L_NDSI,2.87)
L
s(species_data.30$S_NDSI,3.84)
S
s(species_data.30$W_NDSI,1)
W
Festuca.violacea.aggr
s(species_data.30$L_NDSI,1.59)
L
s(species_data.30$S_NDSI,2.38)
S
s(species_data.30$W_NDSI,3.88)
W
Galium.anisophyllon
s(species_data.30$L_NDSI,1.72)
L
s(species_data.30$S_NDSI,2.72)
S
s(species_data.30$W_NDSI,2.87)
W
Gentiana.campestris.sstr
s(species_data.30$L_NDSI,4.59)
L
s(species_data.30$S_NDSI,3.4)
S
s(species_data.30$W_NDSI,8.62)
W

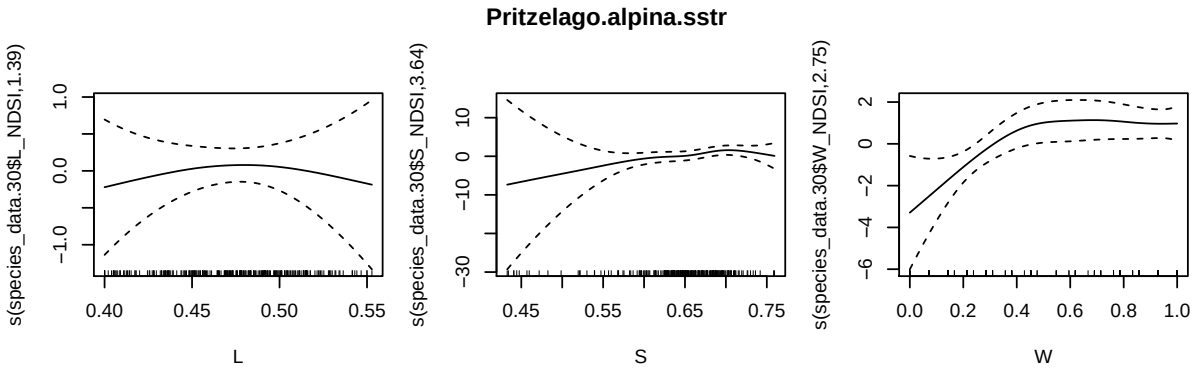

Helianthemum.nummularium.sl
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,3.84)
s(species_data.30$W_NDSI,1)
L
S
W
Homogyne.alpina
s(species_data.30$L_NDSI,2.31)
s(species_data.30$S_NDSI,3.52)
s(species_data.30$W_NDSI,4.02)
L
S
W
Leontodon.hispidus.sl
s(species_data.30$L_NDSI,3.07)
s(species_data.30$S_NDSI,7.77)
s(species_data.30$W_NDSI,2.07)
L
S
W
Leucanthemum.vulgare.aggr
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,5.58)
s(species_data.30$W_NDSI,4.39)
L
S
W

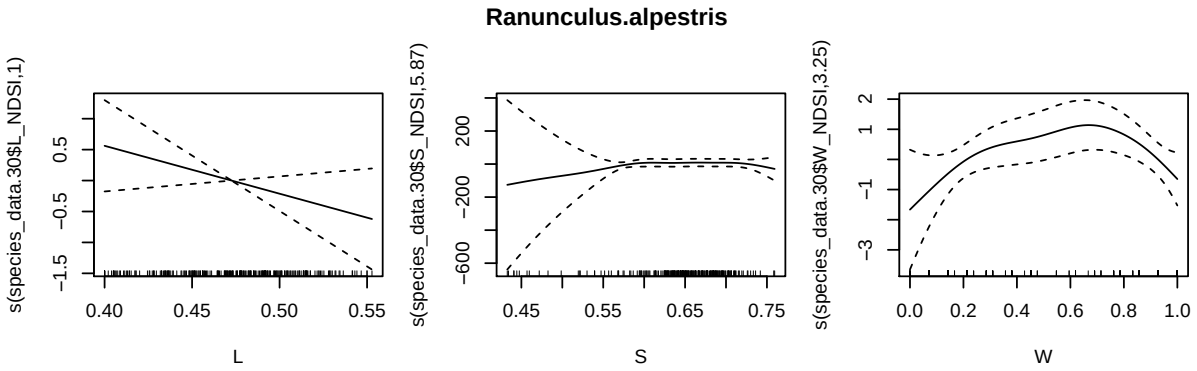

Ligusticum.mutellina
s(species_data.30$L_NDSI,1.79)
s(species_data.30$S_NDSI,7.72)
s(species_data.30$W_NDSI,3.17)
L
S
W
Lotus.corniculatus
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,2.97)
s(species_data.30$W_NDSI,1.54)
L
S
W
Nardus.stricta
s(species_data.30$L_NDSI,4.25)
s(species_data.30$S_NDSI,7.73)
s(species_data.30$W_NDSI,2.25)
L
S
W
Phleum.rhaeticum
s(species_data.30$L_NDSI,2.58)
s(species_data.30$S_NDSI,7.16)
s(species_data.30$W_NDSI,1.34)
L
S
W

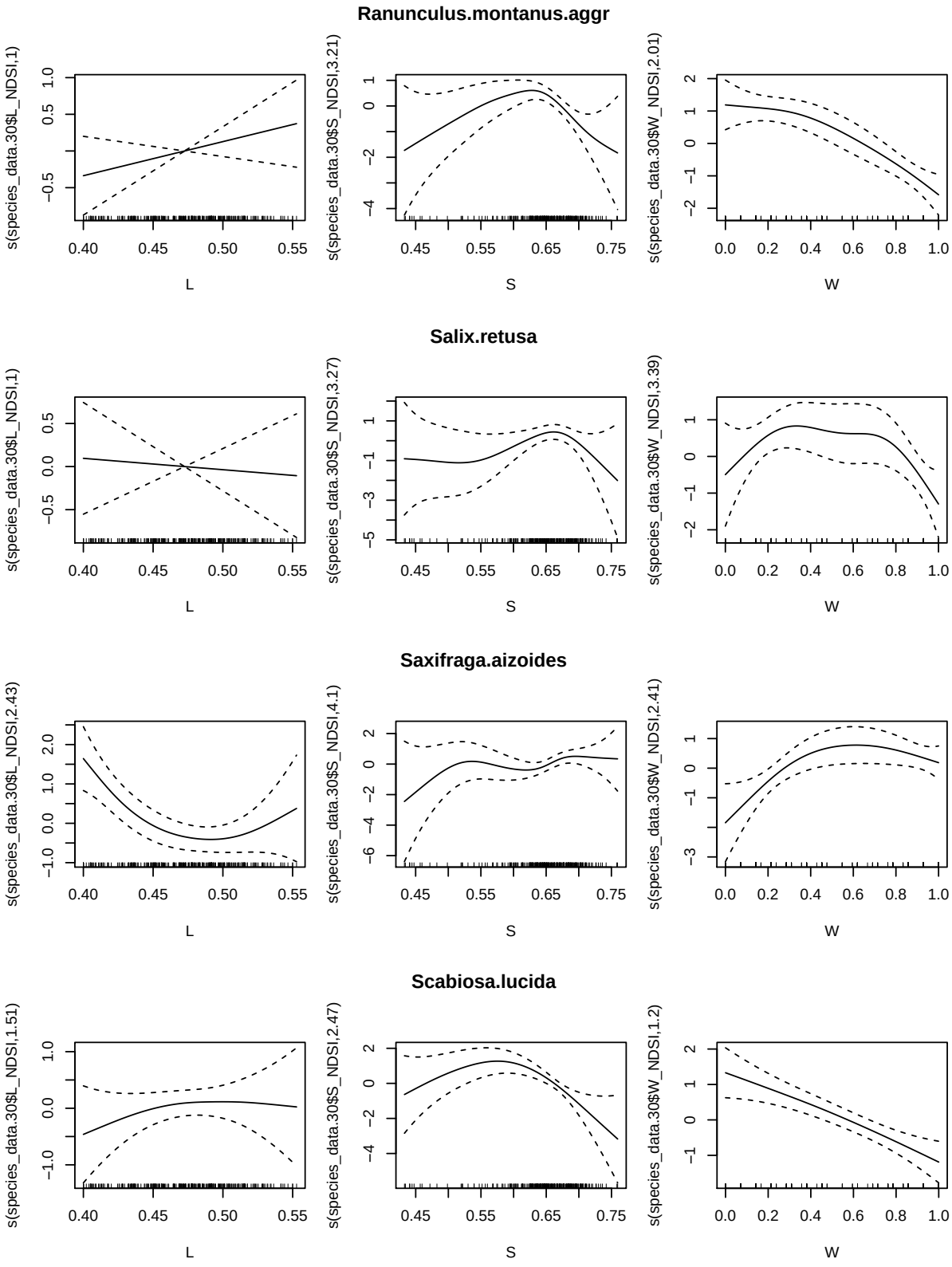

Phyteuma.orbiculare
s(species_data.30$L_NDSI,1.46)
s(species_data.30$S_NDSI,2.91)
s(species_data.30$W_NDSI,1)
L
S
W
Plantago.alpina
s(species_data.30$L_NDSI,4.4)
s(species_data.30$S_NDSI,7.86)
s(species_data.30$W_NDSI,2.51)
L
S
W
Plantago.atrata.sstr
s(species_data.30$L_NDSI,1.24)
s(species_data.30$S_NDSI,2.85)
s(species_data.30$W_NDSI,1.85)
L
S
W
Poa.alpina
s(species_data.30$L_NDSI,2.39)
s(species_data.30$S_NDSI,3.44)
s(species_data.30$W_NDSI,2.21)
L
S
W

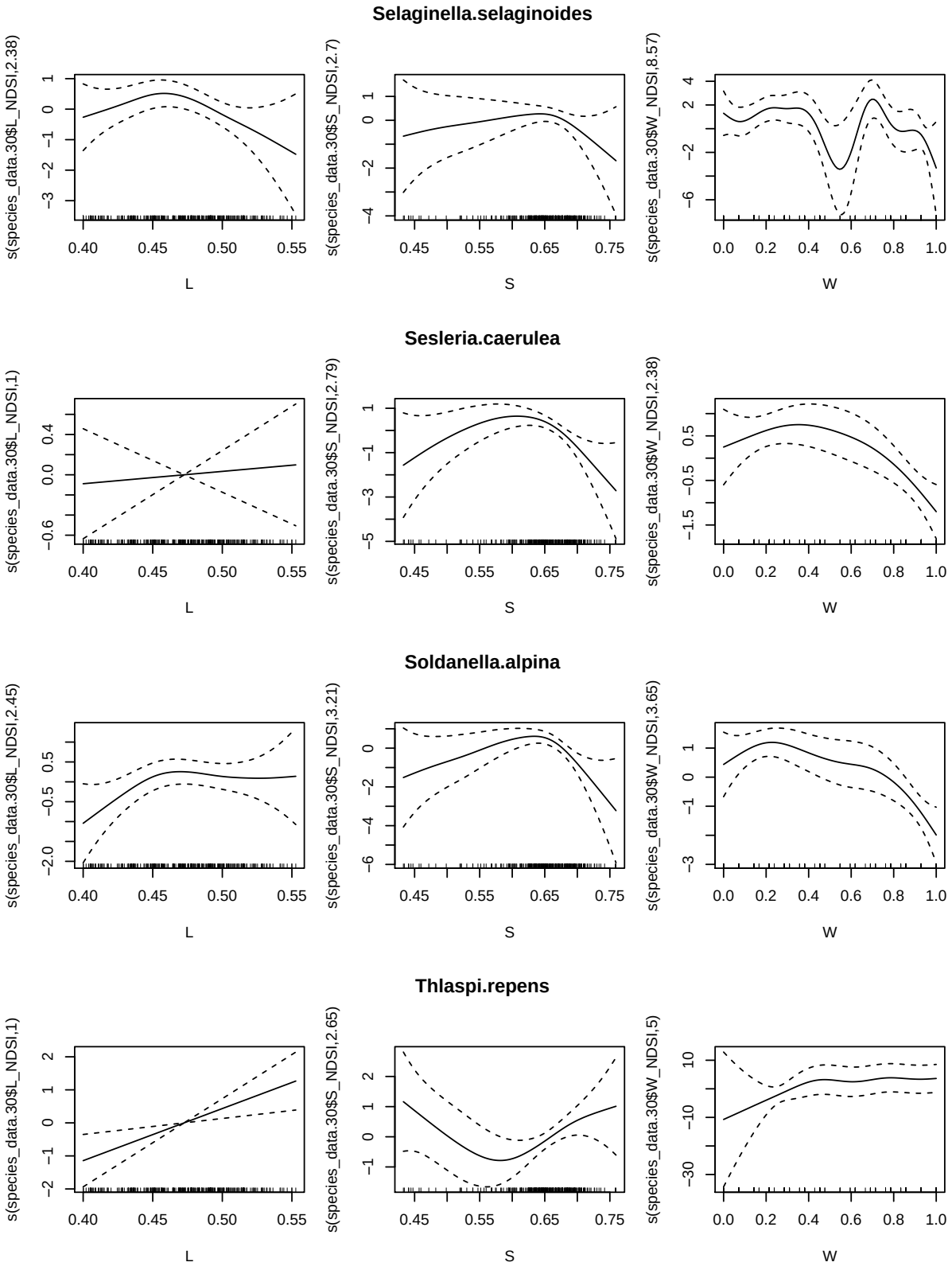

Polygonum.viviparum
s(species_data.30$L_NDSI,2.97)
s(species_data.30$S_NDSI,3.86)
s(species_data.30$W_NDSI,3.81)
L
S
W

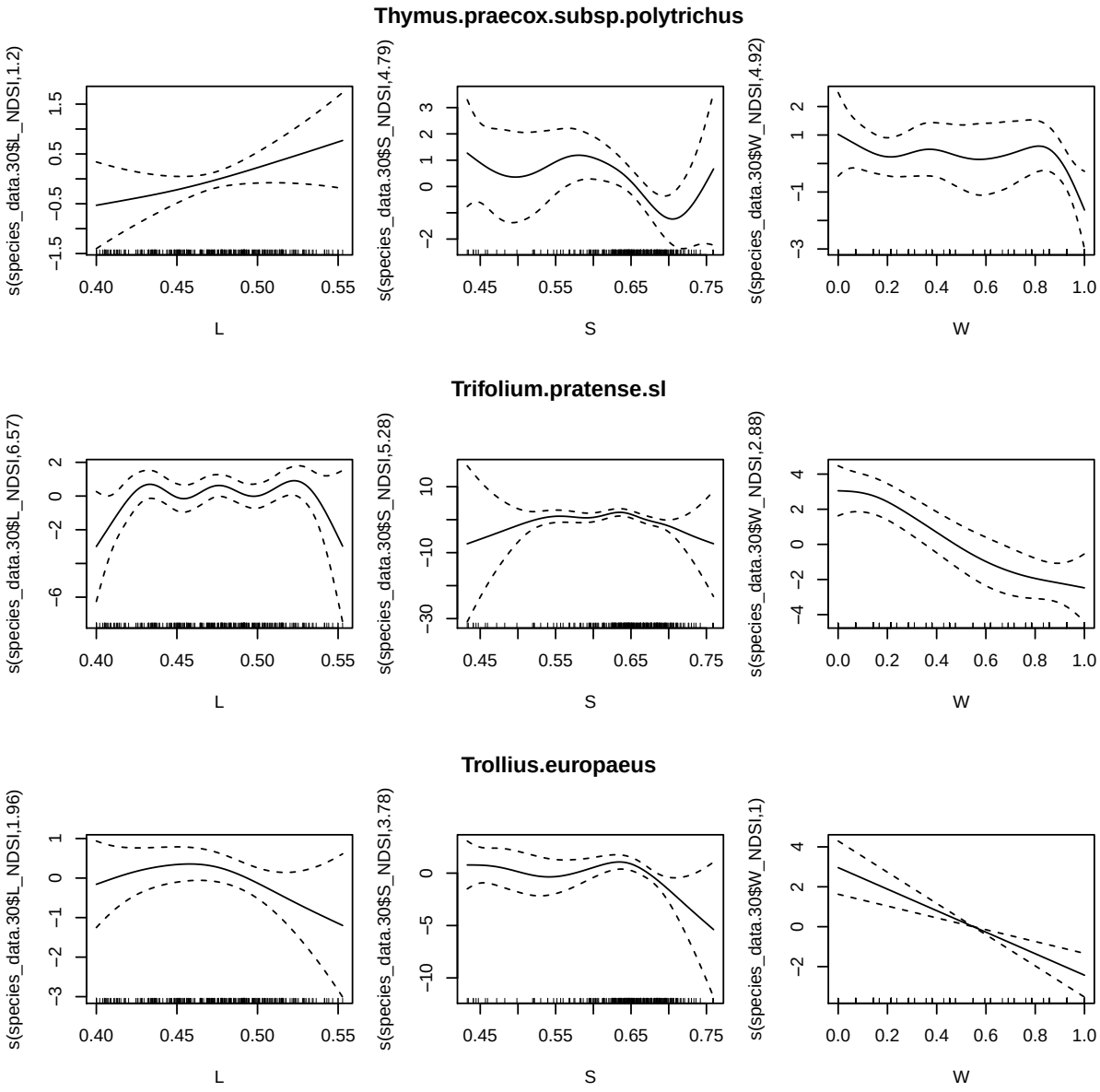

Potentilla.aurea
s(species_data.30$L_NDSI,2.82)
s(species_data.30$S_NDSI,3.74)
s(species_data.30$W_NDSI,2.55)
L
S
W


Pritzelago.alpina.sstr
s(species_data.30$L_NDSI,1.39)
s(species_data.30$S_NDSI,3.64)
s(species_data.30$W_NDSI,2.75)
L
S
W


Ranunculus.alpestris
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,5.87)
s(species_data.30$W_NDSI,3.25)
L
S
W


Ranunculus.montanus.aggr
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,3.21)
s(species_data.30$W_NDSI,2.01)
L
S
W
Salix.retusa
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,3.27)
s(species_data.30$W_NDSI,3.39)
L
S
W
Saxifraga.aizoides
s(species_data.30$L_NDSI,2.43)
s(species_data.30$S_NDSI,4.1)
s(species_data.30$W_NDSI,2.41)
L
S
W
Scabiosa.lucida
s(species_data.30$L_NDSI,1.51)
s(species_data.30$S_NDSI,2.47)
s(species_data.30$W_NDSI,1.2)
L
S
W


Selaginella.selaginoides
s(species_data.30$L_NDSI,2.38)
s(species_data.30$S_NDSI,2.7)
s(species_data.30$W_NDSI,8.57)
L
S
W
Sesleria.caerulea
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,2.79)
s(species_data.30$W_NDSI,2.38)
L
S
W
Soldanella.alpina
s(species_data.30$L_NDSI,2.45)
s(species_data.30$S_NDSI,3.21)
s(species_data.30$W_NDSI,3.65)
L
S
W
Thlaspi.repens
s(species_data.30$L_NDSI,1)
s(species_data.30$S_NDSI,2.65)
s(species_data.30$W_NDSI,5)
L
S
W


Thymus.praecox.subsp.polytrichus
s(species_data.30$L_NDSI,1.2)
s(species_data.30$S_NDSI,4.79)
s(species_data.30$W_NDSI,4.92)
L
S
W
Trifolium.pratense.sl
s(species_data.30$L_NDSI,6.57)
s(species_data.30$S_NDSI,5.28)
s(species_data.30$W_NDSI,2.88)
L
S
W
Trollius.europaeus
s(species_data.30$L_NDSI,1.96)
s(species_data.30$S_NDSI,3.78)
s(species_data.30$W_NDSI,1)
L
S
W

Annexe 3

Table 1. Reclassification for the CRS life strategies from Landolt *et al.* (2010)

CRS life strategies	
Initial categories	Simplification
ccc	competitive
rrr	ruderal
sss	stress-tolerant
ccr, crr	competitive ruderals
ccs, css	stress-tolerant competitors
rrs, rss	stres-tolerant ruderals
crs	CRS

Table 2. Reclassification for the habitat types from Ellenberg *et al.* (1991). We simplified the number of categories by keeping only the broad categories for categories 1,2,3,5,6,7,8 but we kept the subdivisions for category 4. There were not any species that we modelled in habitats 2, 4.1, 4.3.

Habitat types	
Category	Meaning
1	Marshes & ponds
2	Salty habitat, sea shore
3	Ruderal
4	Rocks & alpine swards
4.1	Walls
4.2	Cliffs
4.3	Mine terrils
4.4	Screes
4.5	Snowbeds
4.6	Acidic alpine swards
4.7	Calcareous alpine swards
4.8	Windy ridges
5	Anthropogenic meadows & pastures
6	Bushes & heaths
7	Coniferous forests & acidic heaths
8	Broad-leaved forests
x	Habitat not indicated

Annexe 4

Table 3. Percentage of species showing the different relationship with the respective SI. Graphics are in Annexe 2.

Trends	SI_L	SI_S	SI_W
negative	6.4	2.1	46.8
positive	17.0	0.0	4.3
unimodal	23.4	48.9	38.3
no clear trend	53.2	48.9	10.6

Table 4. Median values of agreement metrics for the models in SET 1 computed with $SI_{permuted}$, SI_L, SI_S, SI_W and the p-values of Wilcoxon signed rank test performed on the agreement metrics between the models including the related snow index and the models including $SI_{permuted}$. P-values in bold are significant.

SET 1							
	$SI_{permuted}$	SI_L	p-value (SI_L)	SI_S	p-value (SI_S)	SI_W	p-value (SI_W)
AUC	0.5466	0.5813	0.0501	0.6321	**0.0076**	0.6931	**2.46E-09**
maxTSS	0.2648	0.3469	**0.0359**	0.4236	**0.0052**	0.4476	**4.02E-08**
maxKAPPA	0.1561	0.1816	0.0794	0.2168	**0.0008**	0.2465	**5.79E-07**

Table 5. Like Table 4 but with SET 2

SET 2							
	$SI_{permuted}$	SI_L	p-value (SI_L)	SI_S	p-value (SI_S)	SI_W	p-value (SI_W)
AUC	0.7090	0.7171	0.1991	0.7349	**4.40E-05**	0.7496	**5.03E-09**
maxTSS	0.4488	0.4659	0.1542	0.4762	**2.45E-05**	0.5042	**2.70E-09**
maxKAPPA	0.3445	0.3484	0.3498	0.3658	**0.0005**	0.3762	**9.58E-07**

Table 6. Like Table 4 but with SET 3

SET 3							
	$SI_{permuted}$	SI_L	p-value (SI_L)	SI_S	p-value (SI_S)	SI_W	p-value (SI_W)
AUC	0.7886	0.7905	1	0.7906	0.6698	0.7959	0.7148
maxTSS	0.5484	0.5457	0.9032	0.5725	0.2412	0.5739	0.8077
maxKAPPA	0.5291	0.5071	0.8552	0.5133	0.2676	0.5253	1

Table 7. P-values of Wilcoxon signed rank test performed on the agreement metrics between the models including the corresponding snow indexes in SET 1. P-values in bold are significant.

SET 1			
	SI_L-SI_S	SI_L-SI_W	SI_S-SI_W
AUC	0.1073	**5.26E-08**	**1.41E-04**
maxTSS	0.1485	**5.50E-06**	**0.02135**
maxKAPPA	0.0667	**1.13E-04**	**0.0275**

Table 8. Like Table 7 but with SET 2

	SET 2		
	SI_L-SI_S	SI_L-SI_W	SI_S-SI_W
AUC	**0.0025**	**1.35E-09**	**0.0008**
maxTSS	**0.0025**	**1.35E-08**	**0.0004**
maxKAPPA	**0.0040**	**1.27E-07**	**0.0023**

Table 9. Like Table 7 but with SET 3

	SET 3		
	SI_L-SI_S	SI_L-SI_W	SI_S-SI_W
AUC	0.583	0.6698	0.4263
maxTSS	0.3258	0.2958	0.3258
maxKAPPA	0.7609	1	0.3575

Table 10. Medians values of variables importance in models including the different snow indexes for SET 2

	SET 2			
	$SI_{permuted}$	SI_L	SI_S	SI_W
gdd	0.380	0.396	0.352	0.278
etp	0.553	0.489	0.441	0.291
snow	0.140	0.188	0.268	0.592

Table 11. Like Table 10 but with SET 3

	SET 3			
	$SI_{permuted}$	SI_L	SI_S	SI_W
tpi	0.061	0.070	0.056	0.062
gdd	0.095	0.108	0.100	0.099
ph	0.771	0.764	0.746	0.662
slope	0.090	0.087	0.078	0.100
etp	0.092	0.098	0.097	0.110
SI	0.041	0.058	0.050	0.091

Table 12. Percentage of species experiencing an increase of predictive performance (AUC, maxTSS, maxKAPPA) for models including each of the snow indexes compared to the models with $SI_{permuted}$ in SET 1

	SET 1		
	SI_L	SI_S	SI_W
AUC	68.2	81.3	91.9
maxTSS	63.6	81.3	91.9
maxKAPPA	63.6	81.3	83.8

Table 13. Like Table 12 but with SET 2

	SET 2		
	SI_L	SI_S	SI_W
AUC	69.2	75.0	87.5
maxTSS	59.0	80.0	90.0
maxKAPPA	56.4	72.5	82.5

Table 14. Like Table 12 but with SET 3

	SET 3		
	SI_L	SI_S	SI_W
AUC	50.0	64.3	42.9
maxTSS	42.9	64.3	50.0
maxKAPPA	50.0	64.3	50.0

Table 15. Percentage of species for which SI is the most important variable in models including each SI in SET 2

SET 2			
$SI_{permuted}$	SI_L	SI_S	SI_W
0.0	4.9	9.3	72.3

Table 16. Like Table 15 but with SET 3

SET 3			
$SI_{permuted}$	SI_L	SI_S	SI_W
0.0	0.0	0.0	14.3

Table 17. Characteristics importance from multimodel inference analysis (MMI)

SET 1 AUC	Moisture index	Habitat	Growth form	Leaf duration	Life strategies	SLA	Vegetation height	Humus
	1	1	0.68	0.3	0.03	<0.01	<0.01	<0.01
SET 1 maxTSS	Moisture index	Habitat	Growth form	Leaf duration	Life strategies	SLA	Vegetation height	Humus
	1	1	0.94	0.03	0.02	<0.01	<0.01	<0.01
SET 1 maxKAPPA	Moisture index	Habitat	Leaf duration	Life strategies	Growth form	SLA	Vegetation height	Humus
	1	1	0.98	0.02	<0.01	<0.01	<0.01	<0.01
SET 2 AUC	SLA	Vegetation height	Growth form	Humus	Life strategies	Moisture index	Leaf duration	Habitat
	1	1	1	<0.01	<0.01	<0.01	<0.01	<0.01
SET 2 TSS	Growth form	SLA	Vegetation height	Life strategies	Humus	Moisture index	Leaf duration	Habitat
	1	1	1	<0.01	<0.01	<0.01	<0.01	<0.01
SET 2 maxKAPPA	Growth form	SLA	Vegetation height	Life strategies	Humus	Moisture index	Leaf duration	Habitat
	1	1	1	<0.01	<0.01	<0.01	<0.01	<0.01
SET 3 AUC	Moisture index	Habitat	Leaf duration	SLA	Vegetation height	Life strategies	Humus	Growth form
	1	1	1	<0.01	<0.01	<0.01	<0.01	<0.01

SET 3 maxTSS	Moisture index	Habitat	SLA	Vegetation height	Life strategies	Leaf duration	Humus	Growth form
	1	1	0.99	<0.01	<0.01	<0.01	<0.01	<0.01
SET 3 maxKAPPA	Moisture index	Habitat	SLA	Leaf duration	Vegetation height	Life strategies	Humus	Growth form
	1	1	1	<0.01	<0.01	<0.01	<0.01	<0.01

Annexe 5

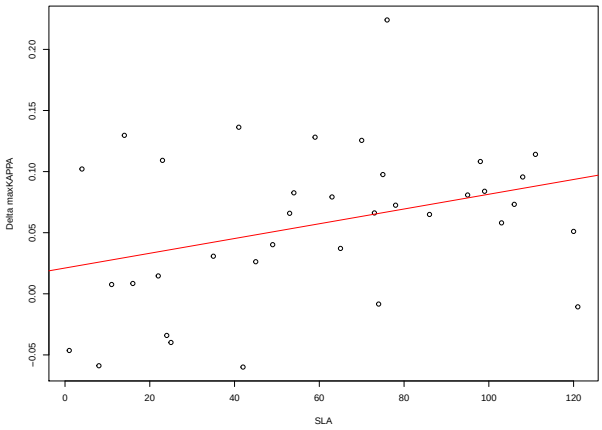


Figure 9. Relationship between the difference in model predictive performance (Delta maxKAPPA) between models with SI_W and $SI_{permuted}$ in SET 2 and the specific leaf area (SLA). The red line is the linear regression line. The slope of the regression line was positive no matter which snow index was considered for SET 2. In SET 1, the slope was usually negative.

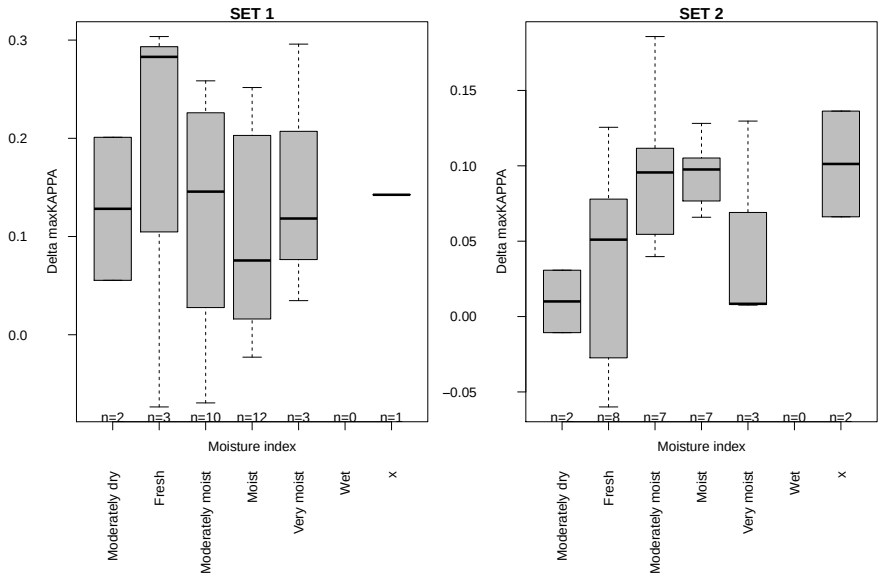


Figure 10. Relationship between the difference in maxKAPPA between models with SI_W and $SI_{permuted}$ in SET 1 and 2 and the moisture index

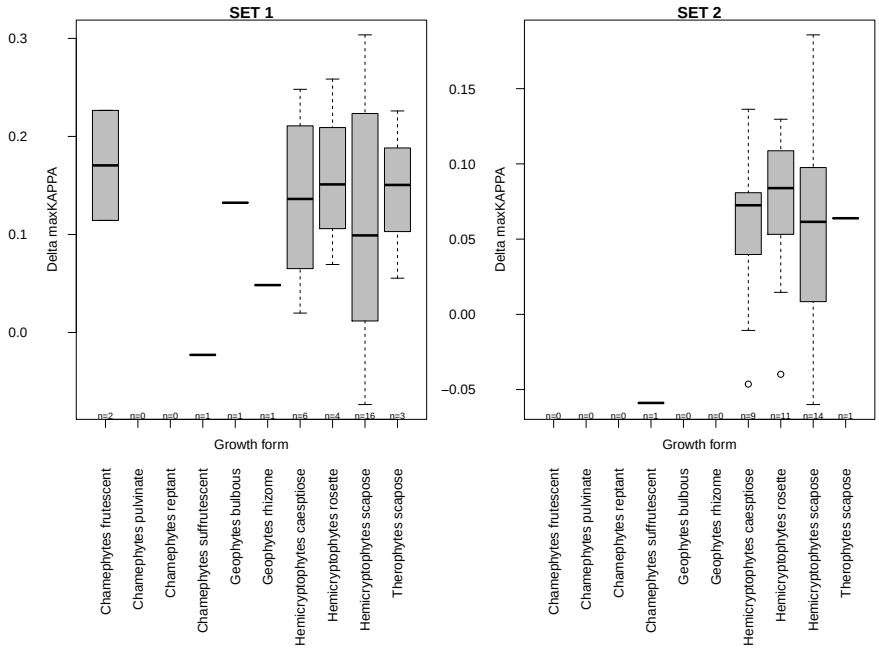


Figure 11. Relationship between the difference in maxKAPPA between models with SI_W and $SI_{permuted}$ in SET 1 and 2 and the growth form