The snow cover is more important than other climatic variables on the prediction of vegetation dynamics in the Pyrenees (1981–2014)

Esteban Alonso-González, Itsaso Ilzarbe-Senosiain, Juan Ignacio Lopez-Moreno, Manuel Esteban Lucas-Borja, Sergio M Vicente-Serrano, Santiago Beguería, Simon Gascoin

ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

The snow cover is more important than other climatic variables on the prediction of vegetation dynamics in the Pyrenees (1981–2014)

To cite this article: Esteban Alonso-González *et al* 2024 *Environ. Res. Lett.* **19** 064058

IOP Publishing Environ. Res. Lett. 19 (2024) 064058 https://doi.org/10.1088/1748-9326/ad4e4c

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
7 September 2023

REVISED
1 May 2024

ACCEPTED FOR PUBLICATION
21 May 2024

PUBLISHED
31 May 2024

LETTER

The snow cover is more important than other climatic variables on the prediction of vegetation dynamics in the Pyrenees (1981–2014)

Esteban Alonso-González[1,2,*], Itsaso Ilzarbe-Senosiain[3], Juan Ignacio Lopez-Moreno[1], Manuel Esteban Lucas-Borja[4], Sergio M Vicente-Serrano[1], Santiago Beguería[5] and Simon Gascoin[2]

[1] Instituto Pirenaico de Ecología (IPE), CSIC, Zaragoza, Spain
[2] Centre d'Etudes Spatiales de la Biosphère (CESBIO), UPS/CNRS/IRD/INRA/CNES, Toulouse, France
[3] Navarra de Infraestructuras Locales S.A. (NILSA), Tudela, Spain
[4] Departamento de Ciencia y Tecnología Agroforestal y Genética, Escuela Técnica Superior de Ingeniería Agronómica, de Montes Biotecnología, Universidad de Castilla-La Mancha, Albacete, Spain
[5] Estación Experimental de Aula Dei, CSIC, Zaragoza, Spain
[*] Author to whom any correspondence should be addressed.

E-mail: alonsoe@ipe.csic.es

Keywords: snow, vegetation dynamics, gradient boosting machines, NDVI, Pyrenees

Supplementary material for this article is available online

Abstract

The dynamics of the mountain vegetation is governed by multiple climatic drivers including temperature, precipitation, radiation and snow cover variability. However, in the Mediterranean environment, little is known about the relative importance of each variable. In this study we assess how different snowpack indices (the maximum annual accumulation, the length of the snow season, and the melt-out date) and key climate variables (precipitation, temperature and shortwave solar radiation) control the interannual variability of the maximum Normalized Difference Vegetation Index (peak NDVI) in the Pyrenees. We use a 33 year long remote sensing dataset (1981–2014) to build a statistical model relating the annual peak NDVI with snow and climate variables. In elevated areas characterized by a well developed seasonal snowpack the melt-out date was the most important climatic variable for predicting the annual peak NDVI. However, at lower elevations where snow presence is ephemeral, shortwave solar radiation was the most important variable. This change in the relative importance of climatic variables occurs around 1300 m a.s.l. The results do not show a significant contribution of maximum snow accumulation, suggesting that indicators of snow presence (i.e. melt-out date or snow season duration), which are significantly easier to obtain than snow mass indicators from remote sensing, could be used to model the influence of the snowpack on peak NDVI at regional scale.

1. Introduction

Mediterranean mountains are recognized as biodiversity hotspots and water towers for the surrounding lowlands (Viviroli *et al* 2007, Grobler and Cowling 2022). Despite the relatively mild winter temperatures of the Mediterranean climate, their mountains often exhibit deep and long-lasting snowpacks (approx. 3 m maximum depth and over 50% of days with snow presence above 2000 m in Iberia (Alonso-González *et al* 2020a)). Mediterranean snowpacks are also known for their high interannual variability (Fayad *et al* 2017). The snowpack in the Mediterranean region is generally close to isothermal for most of the season, making it particularly sensitive to climate warming (Bonsoms *et al* 2023).

Due to the air trapped between the ice crystals, the snowpack is a strong thermal insulator, which regulates the soil temperature (Edwards *et al* 2007). Additionally, fresh water released during the melt season replenishes soil moisture before the growing season (Harpold *et al* 2015). Although these characteristics tend to have a positive impact on plant activity, snow can also damage the vegetation through

avalanches (Camarero *et al* 2000) and occasional heavy snowfalls (Martín-Alcón *et al* 2010). As a result, snow-induced damages, in addition to wind-induced damages, are the primary factors responsible for the spatial patterns of the subalpine forest in the Pyrenees (Camarero *et al* 2000). All these effects are manifested in forest greening, which exhibits a response to the snowpack variations as found in the Sierra Nevada Mountains (California, USA) (Trujillo *et al* 2012).

Trees and shrubs also have great capacity to affect the snowpack distribution as trunks and canopy impact on the snow energy and mass balance (Mazzotti *et al* 2020). In the Pyrenees, forests reduce snow accumulation and snowmelt rates compared to open areas and gaps within the forests, leading to an overall earlier melt out date. Thus, forest management can be utilized to modify the snow storage capacity of forested areas. Managing forest density can regulate the dynamics of snowpack accumulation and melting rates, with a positive or negative correlation depending on the climatic conditions of the region (Gleason *et al* 2021).

In addition to the forest dynamics, the vegetation beneath the snowpack benefits from efficient thermal insulation and improved water resources. Thus, other biomes such as grasslands and shrubs have shown a strong relation to the snowpack dynamics. Revuelto *et al* (2022) proved that the snow melt-out dates explains the maximum greening in an experimental catchment in the Pyrenees. Xie *et al* (2020) proved that melt out dates affect the land surface phenology in the Alps, with a varying response among elevations. Francon *et al* (2020) found that the shrub radial growth is also controlled by the snow melt-out dates, specially at high elevations in the northern French Alps. In the Alps, the snow cover duration (or equivalently, the length of the snow free period) was identified as the primary driver of inter-annual variations in primary productivity of mountain grasslands over a 13 year period (Choler 2015). Thus, an improved understanding of the relationship between snow and NDVI may enhance the management capabilities of high mountain systems, especially under the current warming scenario where changes in the snowpack dynamics are expected. The recent availability of satellite data and snow reanalysis that spans longer time periods opens the opportunity to further investigate these effects. Despite this, the regional impact of snowpack on the dynamics of vegetation activity in the Pyrenees has not yet been investigated.

The main objective of this work is to explore how various snow and climate indices impact the annual fluctuations of the maximum normalized difference vegetation index (peak NDVI) in the Pyrenees, to understand the underlying dynamics driving this interannual variability and how this relation evolves with elevation. This primary objective will be accomplished by developing statistical models capable of predicting the year-to-year variations in peak NDVI based on climatological and snow cover indicators. Despite its Mediterranean character, there is an evident climatological gradient in the Pyrenees, which makes it an interesting study area. We have used new sources of meteorological and phenological information, based on numerical simulations and remote sensing. This allowed us to study the response of the peak NDVI to the snowpack variability and other climatic factors for the period 1981–2014. Hence, we have used machine learning techniques to disentangle the complex relations between the snow and other climatic variables and the vegetation dynamics through NDVI retrievals, exploiting the recent availability of long term data.

2. Study area

The Pyrenees are the northern boundary of the Iberian Peninsula and constitute a natural border between Spain and France (figure 1). It runs from west to east over 400 km from the Atlantic Ocean to the Mediterranean Sea with 150 km at their widest point (along a north-south axis) in the central area. It is a cross-border mountainous area that includes the territory of three European countries: France, Spain and Andorra. The Pyrenean mountain range is a key source of runoff providing hydroelectric power and irrigation resources, as the headwaters include the main tributaries of important rivers as the Ebro (Spain) and Garonne (France). The Pyrenees have several peaks above 3,000 m a.s.l. being the Aneto the highest (3,404 m a.s.l.).

The climate of the Pyrenees is highly variable in an east-west direction, with a transition from the Atlantic influence to the Mediterranean climate. The northern and southern slopes exhibit a marked difference along the main axis of the range, the former often exhibiting higher precipitation rates. Annual precipitation ranges from 600 mm in the drier areas of the eastern part of the Spanish Pyrenees to 3000 mm in some areas of the northwest facing slopes (Serrano-Notivoli *et al* 2017). Above 2000 ma.s.l. the duration of the snowpack usually exceeds six months (Gascoin *et al* 2015).

The regional dynamics of the snowpack, which exhibits a decreasing trend towards the east of the range, induces changes in the vegetation, which tends to exhibit more xerophytic and continental behavior (Ninot *et al* 2007). The forest stands of the Pyrenees are highly exposed to the seasonal snowpack, which conditions its dynamics through its accumulation patterns (Sanmiguel-Vallelado *et al* 2019) and through the damages caused by avalanches and wind (Camarero *et al* 2000) being the main drivers of tree cover distribution at high elevations.

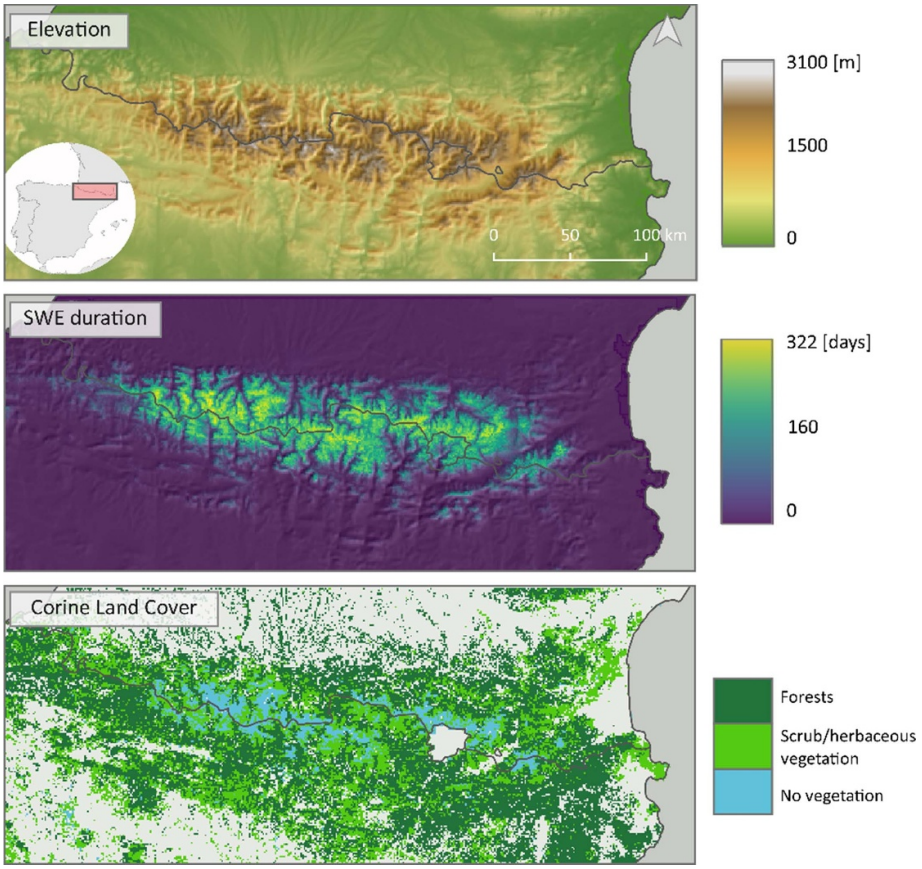

Figure 1. Boundary of the study area, with its topography, average snow water equivalent (SWE) duration and land use. The white areas in the lower panel represent the grid cells not included in the analysis due to no available information or non forest and semi-natural areas category derived from CORINE (see 3.4 Land cover map).

Pyrenean grasslands are also affected by the snow dynamics, as the melt out date controls the intensity of the greening (Revuelto *et al* 2022).

3. Data

3.1. Remotely sensed normalized difference vegetation index

The NDVI is a spectral index that can be estimated from information obtained from space-based sensors. This index has a strong correlation with various vegetation parameters such as the leaf area index (Shabanov *et al* 2005), green biomass (Wylie *et al* 2002) or vegetation cover (Gouveia *et al* 2016).

We have used a NDVI database (Vicente-Serrano *et al* 2020) generated from the daily observations of the Advanced Very-High-Resolution Radiometer onboard the National Oceanic and Atmospheric Administration (AVHRR-NOAA) orbiting platforms. The images correspond to the midday passes in ascending mode, which means more than 10 100 images covering the period 1981–2015 with a 1.1 km spatial resolution. Despite being a relatively old sensor, its performance in retrieving NDIV has proved to be close to more recent sensors such as MODIS or Spot (Fontana *et al* 2008). The product is a grid composite of the best (and also highest) observation available during two weeks without gaps as a consequence of the cloud cover, covering the period 1980–2015. Vicente-Serrano *et al* (2020), provides a full description of the algorithm to develop the product and more details about its validation. Since the database is a composite of the best and highest observed NDVI, it is particularly suitable for estimating peak NDVI values. In addition, the maximum NDVI ensures minimal contamination due to the sun-sensor-topography geometry and the atmosphere.

3.2. Snowpack derived indices from numerical simulations

We have used a daily gridded snowpack database to estimate different indicators of the snowpack (Alonso-González *et al* 2018). This snowpack database has daily resolution, providing information about the snow water equivalent (SWE) for the period 1980–2014.

It was created by coupling numerical simulations of the atmosphere with a physically based snowpack model. More specifically an atmospheric simulation of 0.088° × 0.088° (∼9 km) resolution, generated using the Weather Research and Forecast (WRF) meso-atmospheric model was used to generate meteorological forcing over Iberia. A downscaling

algorithm was developed to improve the resolution of the WRF meteorological fields to 1 km before simulating the snowpack using the Factorial Snow Model (FSM, Essery 2015). The authors reported a coefficient of determination (R^2) of 0.72 compared with MODIS snow cover observations, and kappa values generally above 0.6 for most of the accumulation percentiles compared with in situ snow depth observations, demonstrating the ability of the database to reproduce the interannual variability of snow cover.

We have used this database to calculate different snowpack indicators. The indices calculated were annual maximum SWE (peak SWE), snow season length (duration SWE) and the melt-out day. These indices were only calculated on snow seasons (hydrological years, i.e. from October to September) that showed snow for more than 30 d with more than 30 mm SWE. These thresholds where selected to filter out ephemeral snowpack, as well as sporadic off-season snowfalls, following the parameters proposed by Sturm *et al* (1995), but adapted to the characteristics of the Pyrenean snowpack (Alonso-González *et al* 2020a).

3.3. Climatological indicators derived from reanalysis

The ERA5-Land reanalysis (Muñoz-Sabater *et al* 2021) has been used to obtain seasonal climatic data on temperature, total precipitation (TP) and shortwave radiation (SW). The use of ERA5-Land instead of WRF enhances the independence of the snowpack numerical simulations with respect to the climatic indicators, which would otherwise show a (non-linear) relationship through FSM. Using ERA5-Land, we have calculated the yearly values of mean temperature (Temp winter, Temp spring, Temp summer), shortwave radiation (SW winter, SW spring SW summer) and total precipitation (TP winter, TP spring, TP summer), for the winter (December, January, February), spring (March, April, May), and summer (June, July, August) seasons, respectively.

3.4. Land cover map

In order to mask areas out of the interest of the present study, and take into account the changes in the land cover along the time period (1980–2014), we have included information of the land cover extracted from the CORINE Land Cover inventory. The CORINE inventory was initiated in 1985, and is composed by different updates for the 2000, 2006, 2016 and 2018 years. We have included in our analyses the classes included in the forest and semi-natural areas category of CORINE, which involves Broad-leaved forest, Coniferous forest and Mixed forest, as well as Natural grassland, Transitional woodland/shrub and Sclerophyllous vegetation (These classes have been grouped into forest and shrub/herbaceous vegetation to facilitate map visualization in figure 1). We removed several land cover from the analyses (non-natural areas, areas with little or no vegetation, as non-vegetated expanses of sand or bare rock/ground, burnt areas and water bodies), as they are not relevant for the present study. Andorra, which is not included in CORINE, was also removed from the analyses.

4. Methods

We modeled the annual peak NDVI using climatological, land cover and snow-derived indices as predictors. However, before training the models, we performed various pre-processing steps. First we aligned the geometries of all the gridded products to make them consistent across all data sources using a bilinear interpolation for the continuous variables. In the case of the categorical variables, the labeled category was the dominant, covering in most cases more than 50% of the cell area (supplementary figure 1). To evaluate the effect of resampling, the experiment was repeated using only cells where a single class accounted for more than 80% of the cell area.

We removed any long-term temporal trends from the variables in the analysis, as otherwise its presence would force a correlation between some of the indices, influencing the modeling performance. Time series of each variable were extracted cell by cell, and then the slope of the linear trend was calculated with the Theil-Sen slope. This index is a robust way of obtaining the slope of a time series, which, unlike trend removal methods based on linear regression, is less sensitive to outliers. To eliminate the influence of land cover changes we used the CORINE products to mask those cells where changes were detected. We calculated the correlation matrix between the predictors in order to evaluate potential collinearities that could affect the modeling phase, particularly the interpretation of the relative importances of the predictors.

We used gradient boosting machines (GBMs) as the algorithm for modeling the peak NDVI value. We used the open-source implementation XGBoost (Chen and Guestrin 2016). GBMs are based on ensembles of weak models that are used to solve regression and prediction problems. We trained a different GMB for each elevation band in the range of 900–1800 m (50 m intervals) to completely isolate the effect of the elevation. The models were trained using 30% of the available data, using the remaining 70% for validation ensuring that the models were not overfitted. The selection of the training and validation samples was performed by sampling randomly among all the available data. As validation statistics we used the Pearson's R correlation, mean squared error (RMSE) and mean absolute error (MAE) between predicted and observed peak NDVI values.

The GBM model hyperparameters, namely the maximum depth of the trees, the percentage of subsampled data at each iteration, the learning rate and the minimum number of instances in each node

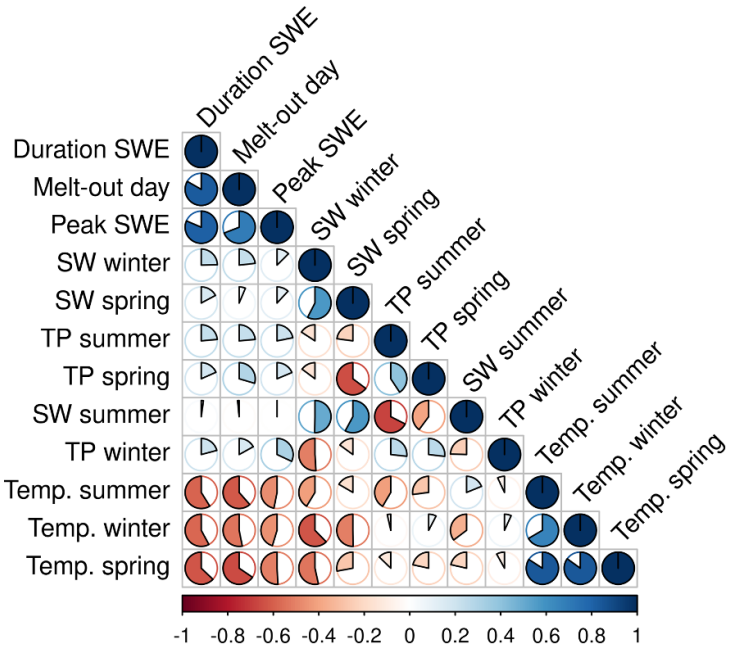

Figure 2. Correlation matrix between all non-categorical predictors.

was automatically selected using a grid approximation through the Machine Learning in R (mlr) package (Bischl *et al* 2016). In addition, we monitored the model improvement in each iteration, and used an early-stopping criteria when more than ten rounds without improvements occurred through cross-validation. The objective function to minimize was the squared error. Finally, from each model we extracted the relative feature importances and calculated the SHAP contributions (SHapley Additive exPlanation) (Lundberg and Lee 2017) of the most relevant snow index to estimate their marginal contribution to the peak NDVI prediction. The SHAP method is based on the concept of Shapley values from cooperative game theory, which are used to determine how much each player in a collaborative game contributes to its success.

5. Results

First, we computed a correlation matrix among the non-categorical (all but the ones derived from CORINE) predictor variables. As expected, several variables exhibited high values of correlation (figure 2), with many of them exhibiting Pearson's *R* values greater than 0.5. This is particularly relevant for the snowpack variables (namely the melt rate, season duration and melt-out day), as they exhibit an obvious relationship.

Cross-validation of the GMB models consistently show statistical significance of the relationship between predicted and observed values (p-value <0.05) and generally high Pearson's *R* values (ranging between a maximum of 0.92 and a minimum of 0.64) and very low MAE and RMSE metrics across the entire range of elevations, as depicted in figure 3.

Additionally, the variance of the residuals remained relatively constant across the entire range of peak NDVI values (figure 3), not showing obvious heteroscedasticity. The GBM models accurately captured the interannual variability of the peak NDVI, exhibiting high correlation with the validation dataset (Pearson's *R* between 0.64 and 0.92 in all the elevation bands, figure 4). The model performances declined with increasing elevation, both in terms of cell by cell correlation and in temporal dynamics. Despite the high collinearity between the predictor variables, GBM has proved to be a powerful tool for explaining a significant portion of the variance of the peak NDVI based on the predictor variables.

Figure 5 illustrates the ranking of the top ten most important variables, with 1 being the most important and 10 the least important, at each elevatiol band. Certain snow indices were found to have the most significant impact on predicting the peak NDVI in the elevations where they are present, specially the melt-out date and to a lesser degree the snow season duration.

The maximum annual snow accumulation (Peak SWE) has not been found to be a significant variable in any of the elevation bands. Thus, considering the results, the snow mass is not a determining factor in influencing peak NDVI, despite its obvious relationship with the duration of the snow season (Duration SWE). As both predictors are highly correlated, we can not conclude that snow mass is not important, as it can exert its influence throughout its relationship

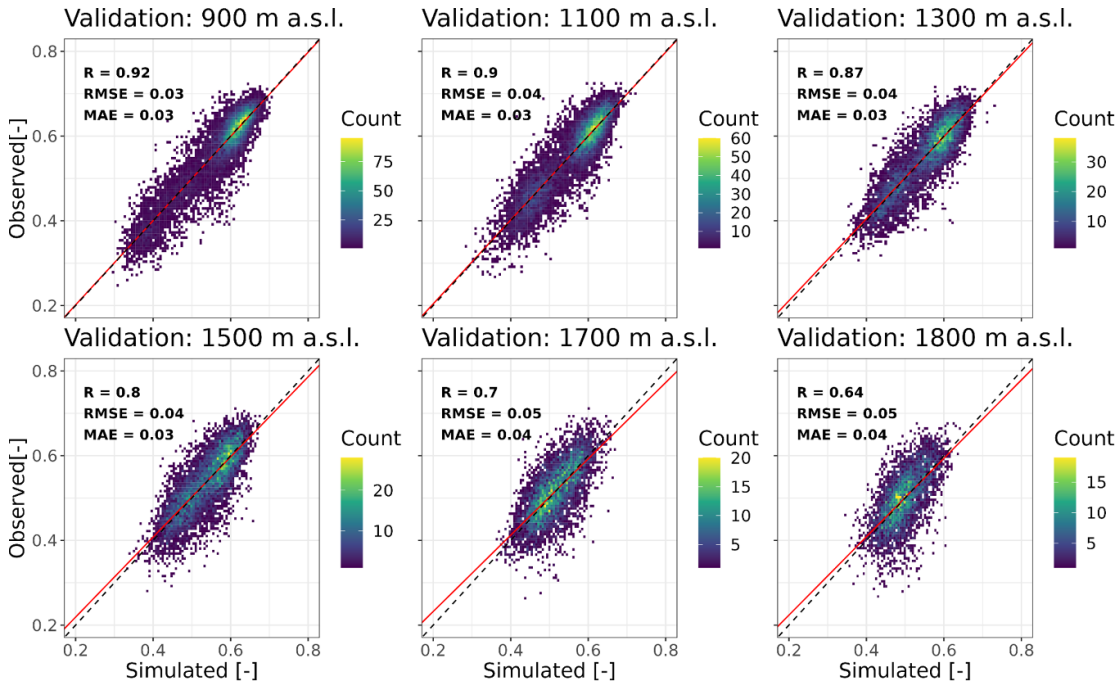

Figure 3. Simulated and observed peak NDVI values in the validation data set.

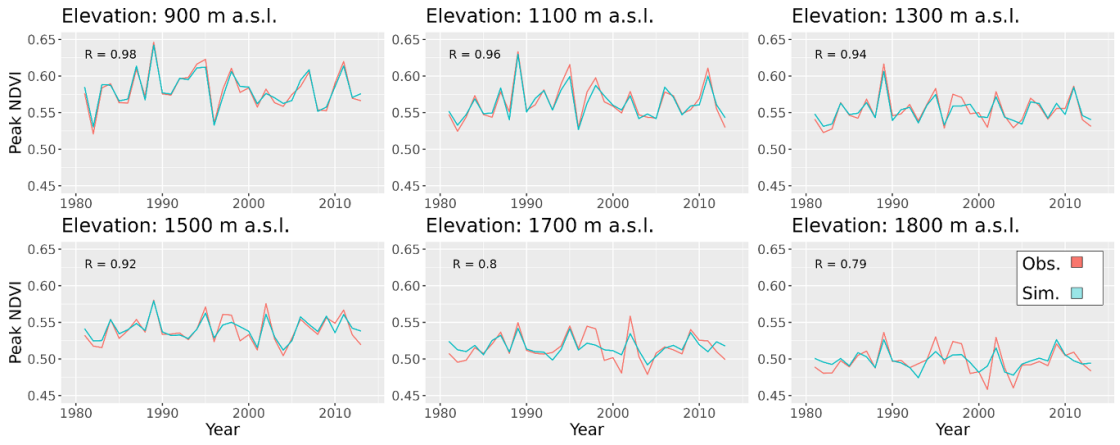

Figure 4. Simulated and observed peak NDVI time series averaged over all the grid cells in the validation data set.

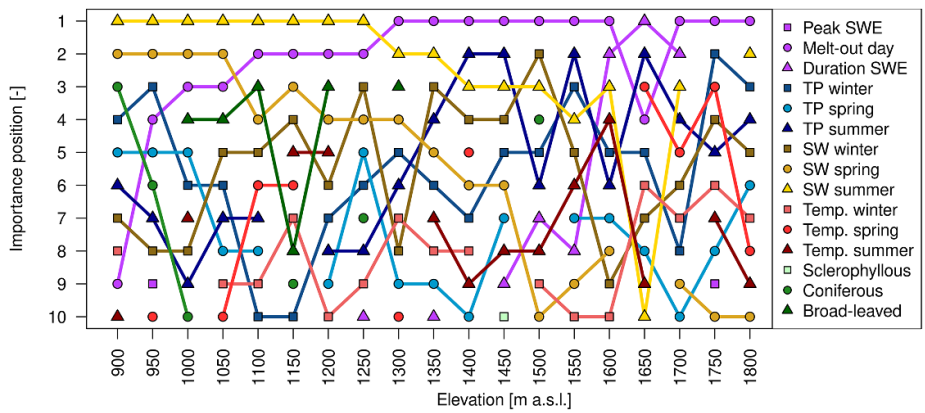

Figure 5. Relative importance of each feature to predict the annual maximum NDVI at each elevation band.

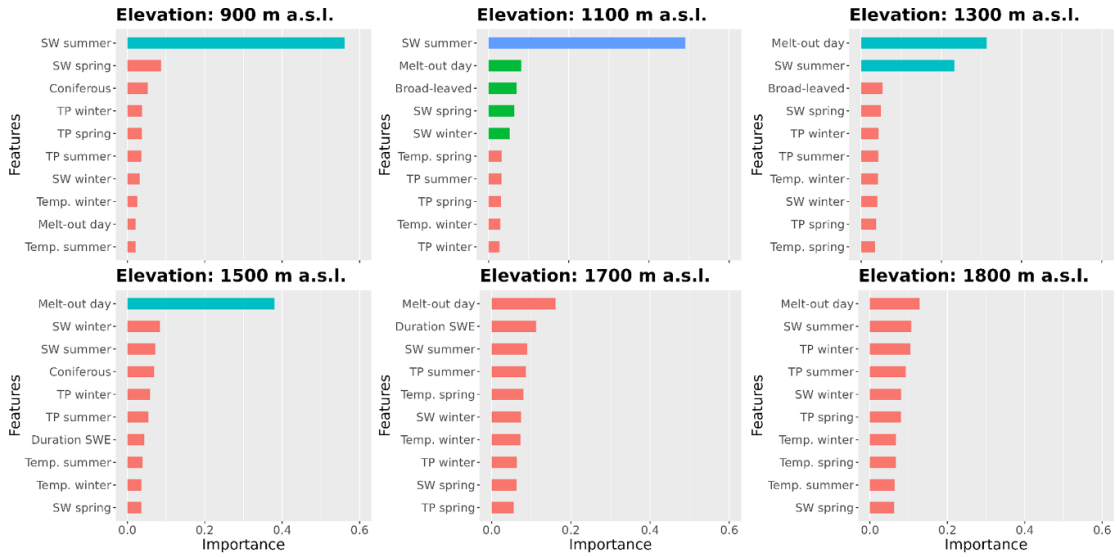

Figure 6. Relative contribution of each predictor to predict the annual peak NDVI at different elevation bands. The colors indicate groups of similar contributions within each elevation band estimated automatically from a one dimensional cluster.

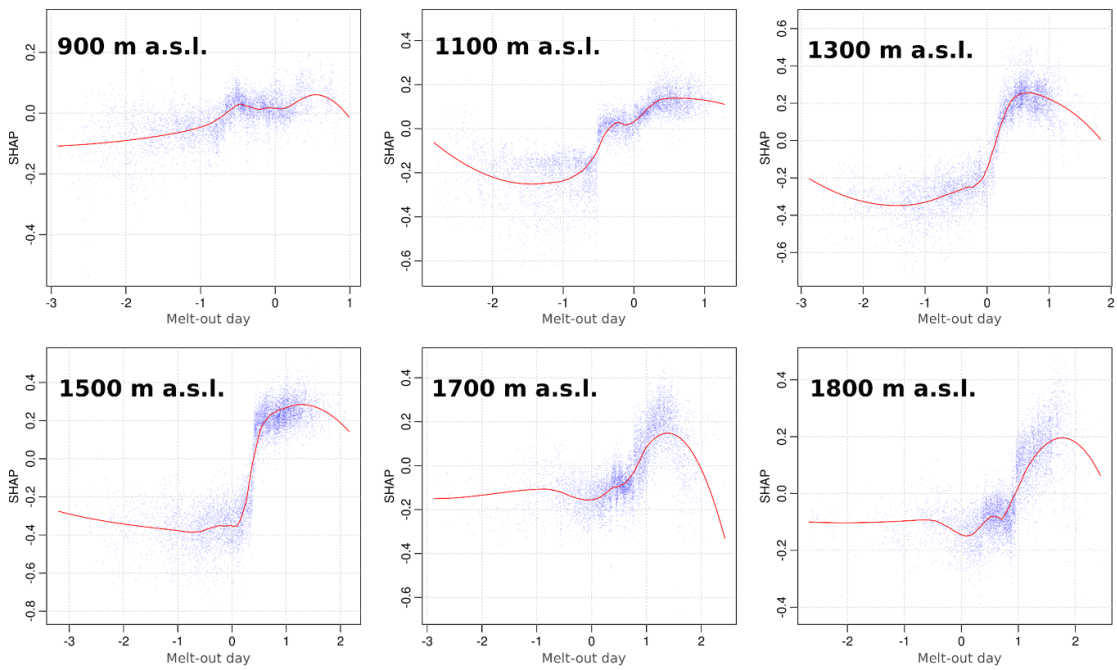

Figure 7. SHAP values showing the relative contribution of the melt-out date to predict the maximum NDVI at different elevations.

with other variables that show a stronger signal. Land cover variables were not found to be relevant even if only those cells with more than 80% of the same coverture were used (supplementary figure 2).

Although figure 5 depicts the top ten variables, it should be noted that there are no discernible patterns among the less important predictors. Figure 6 shows that the largest contributions of the snow variables occur at mid-elevation, while at higher elevation their relative contribution decreases.

According to the SHAP values (figure 7), the relation between the peak NDVI and the melt-out date exhibits a high nonlinear behavior. While low melt out dates (below the average) seems to slightly negatively affect the peak NDVI values, its influence becomes high with intermediate snowmelt dates. In particular, values close to the mean (or slightly above the mean) seem to maximize the peak NDVI as found in previous field experiments (Revuelto *et al* 2022). However the peak NDVI exhibits an abrupt decline as the melt-out day delays towards the latest values, in line with previous findings (Choler 2015, Sanmiguel-Vallelado *et al* 2019, Francon *et al* 2020).

6. Discussion

In this work we have studied the main factors controlling the interannual variability of the annual peak NDVI. We have used GBMs to model the peak NDVI from different climatological variables and land categories including vegetation types derived from both remote sensing and numerical modeling using physically based models. Despite the associated uncertainty to the predictors derived from numerical simulations and remotely sensed variables, GBMs have been able to predict annual peak NDVI with Pearson's R values ranging from a maximum of 0.9 to a minimum of 0.6, exhibiting decreasing performance with elevation. It is likely that predictive ability has overcome uncertainty, given that we use standardized seasonal values. Since we find inter-annual variability much easier to simulate than intra-annual variability, other experimental designs aiming at simulating higher frequency time series are likely to present greater challenges. The decreased model performance with increasing elevation may be attributed to the lack of interannual variability in some of the predictors. Particularly those related to the snowpack, exhibit an important decrease in interannual variability as elevation increases (Alonso-González *et al* 2020a). Moreover, in biomes associated with higher elevations and cold regions such as shrublands and grasslands, inter-annual climate variability has a stronger impact on the date of occurrence of the peak NDVI rather than the peak annual NDVI (Paruelo and Lauenroth 1998, Pirk *et al* 2023).

Snowpack has proven to be a crucial factor for controlling the dynamics of NDVI at the regional scale. These findings could assess the basis of a potential seasonal prediction of peak NDVI, as there is potential in the predictability of snow cover dynamics through atmospheric teleconnection indices. More specifically for the Iberian Peninsula, the North Atlantic Oscillation (NAO) has proved to be an important predictor of the snow season (López-Moreno and Vicente-Serrano 2007, Alonso-González *et al* 2020b). Actually, the relationship between the NAO index and the NDVI dynamics in Iberia has already been proven (Vicente-Serrano and Heredia-Laclaustra 2004). Some potential exists to predict the NAO in a year in advance (Dunstone *et al* 2016), and even up to a decade (Eden *et al* 2002).

The snowpack influence in predicting the maximum NDVI has been mostly manifested through the melt-out date, but also through the length of the snow season, in line with previous findings in the Alps (Jonas *et al* 2008). This control exhibits a non-linear behavior, where late melt-out dates seem to exert a higher influence on the peak NDVI (figure 7). This is in line with previous findings in grasslands in the Alps (Choler 2015) and the Pyrenees (Revuelto *et al* 2022), where the influence of the snow season was not linear. Similar conclusions were drawn from dendrochronological studies in shrubs in the Alps (Francon *et al* 2020), and in mountain forests in the Pyrenees (Sanmiguel-Vallelado *et al* 2019). All these biomes showed the same response to snow cover duration according to previous studies, which explains why the relative importance of the land cover was not high in our study compared to the snow and climatic variables (figure 5), suggesting the potential generalization of the conclusions to different biomes.

In addition to the fact that the length of the snow season limits the growing season in the low vegetation areas, previous studies have related NDVI patterns to soil temperature. For example, Piedallu *et al* (2019) conducted a study looking at the impact of some climatic and soil factors on NDVI patterns on the French side of the Eastern Pyrenees (Occitania, France). They found that the major constraint for high NDVI values was low soil temperatures in spring together with low water supply in summer. These two factors are strongly related to the melt out day, due to the insulating effect and water storage capabilities of the snowpack. Furthermore, local studies using dendrochronology and soil temperature monitoring in the Central Pyrenees (Aragon, Spain) demonstrated that snowpack influences radial growth of high mountain pine forests through the control of soil temperature (Sanmiguel-Vallelado *et al* 2021). Figure 4 shows how the melt-out day progressively increases importance as elevation increases. In its place, summer SW radiation and spring SW become progressively less important. In the absence of a snowpack, shortwave radiation influences ground temperature dynamics. The increasing importance of the melt-out day with elevation in predicting the peak NDVI (also observed in previous studies, but manifested through the snow depth (Trujillo *et al* 2012), together with the decreasing importance of shortwave solar radiation with elevation, could suggest that the most important mechanism of snow cover influence on peak NDVI would be the control of soil temperature dynamics. This effect has been proven at the local scale through field experiments (Sanmiguel-Vallelado *et al* 2021), but to our knowledge this is the first time it has been shown at regional scales. The results highlight the influence of the snowpack in the dynamics of mountain vegetation communities, and thus recommend to consider the expected evolution of the snowpack in land management climate change adaptation policies.

On the other hand, peak SWE has not been shown to be a determining variable in predicting peak NDVI. Although large accumulations of snow can trigger avalanches that may cause damage to forest stands, and exceptional snow falls may damage the vegetation, no correlation has been found between the maximum annual snow accumulation and the peak NDVI. Avalanches tend to occur in

topographically favorable locations, and the damage they cause is a crucial factor in determining the distribution of vegetation in high mountain areas (Camarero *et al* 2000). As these events occur consistently in the same areas, they maintain similar NDVI values over time, regardless of snow accumulation levels. Another effect that seems to not influence the peak NDVI is the fact that the snowpack is an effective water storage system. This surprising conclusion highlights the complex dynamics of the hydrological system, where even other directly related phenomena such as the annual water balance is not necessarily related with the peak accumulation values (López-Moreno *et al* 2020). However, this conclusion should be taken with caution. Previous studies have found a very high correlation between melt out day and maximum soil moisture, independently of the snow accumulation (Harpold *et al* 2015). Although GBM is a robust algorithm able to deal with collinearity in the predictors, it still has some influence on the results. Collinearity may not have a significant impact on the overall predictability of the models but the relative importance metrics may suffer in case of high correlations. This effect should be taken into consideration, as if one of the snowpack variables shows much higher relative importance than the others, it may be a consequence of this effect. Therefore, the model could be underestimating the relative importance of some variables favoring its correlated counterpart. Thus, it is possible that the models are underestimating the hydrological effect given the high correlation between the melt out day and the water availability (figure 2). Including soil moisture in the modeling pipeline could resolve this issue, but more research is needed as soil moisture is a complicated variable to estimate from both numerical simulations and observations. A potential line of future work will be the hybridization of mechanistic models with machine learning techniques instead of the use of purely empirical data driven models, which could help to solve this kind of issues (Razavi *et al* 2022). Also, the use of purely empirical models makes it difficult to draw conclusions in the medium term where conditions will differ significantly from the reference period as a result of climate change impacts. The use of mechanistic models may be the way to draw robust conclusions about future climate change impacts, as the reduction of assumptions allows for a better performance in extrapolation mode.

In this work we have focused on the regional scale. This has allowed us to benefit from long time series. The fact that snow presence indicators (such as the season length or the melt-out date) may be sufficient to predict the annual peak NDVI paves the way for a more systematic inclusion of snowpack variables in ecological studies, where the melt out date seems to be a key predictor even at hyper-resolution scales (Revuelto *et al* 2022). Obtaining information of snow cover from remote sensing is significantly easier and more accurate than other products that involve estimating snowpack mass, and it is even possible at high resolutions (Macander *et al* 2015). This is particularly relevant in the context of operational forest management, where resolution becomes a key factor to consider. In addition, given the limitations of resolution, our models are not spatially explicit, which makes it difficult to extrapolate the conclusions at the local scale, a limitation that could be addressed in the future with the use of higher resolution products. Using high-resolution satellites such as Landsat and/or Sentinel 2 could allow the present work to be extended to a wide number of ecosystems, even if snow mass data is not readily accessible. The obvious increase of data available with the increase of resolution and the more detailed representation of the vegetation spatial dynamics paves the way to develop more detailed studies in the future.

7. Conclusions

In this work, we have used remote sensing data sources and long-term numerical simulations to model the influence of different climatic variables on the control of peak NDVI at the regional scale. We have used GBMs to model the ability of different climate and snow cover indices to predict peak NDVI at different elevation bands, to study how the relative importance of the predictors evolves with the elevation.

The results showed that snowpack plays a crucial role in the dynamics of peak NDVI, becoming the most important predictor of all those chosen at the elevations where it is present. Among the snowpack indices used, the models suggest that melt-out day is the most informative predictor. The importance of melt-out day increases progressively with elevation, replacing indicators related to shortwave radiation which were found to be the most informative at lower elevations. This could indicate that the control mechanism exerted by snow on peak NDVI could be manifested through the control of soil temperature.

The results have not shown that SWE-dependent indicators of the snowpack are of significant benefit over indicators based on the presence of snowcover. The latter being significantly easier to obtain than the former by remote sensing especially at higher resolutions, this opens the door to a multitude of studies relating snowpack to vegetation dynamics. This is especially relevant in the context of the new generation of satellites, where high resolution information is readily available.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.20350/digitalCSIC/15466 and https://doi.org/10.5281/zenodo.854619.

Acknowledgments

Esteban Alonso-González has been funded by the French Space Agency (Centre National D'Etudes Spatiales (CNES)) postdoctoral fellowship and the Climate Change Innitiative European Space Agency postdoctoral Fellowship. This project has been partially funded by the project: 'Cartografía de alta resolución espacial del manto de nieve y su variabilidad reciente en los PPNN de montaña, y los impactos del cambio climático para el horizonte 2050'. from the Spanish Autonomous Organism of National Parks

ORCID iDs

Esteban Alonso-González https://orcid.org/0000-0002-1883-3823
Juan Ignacio Lopez-Moreno https://orcid.org/0000-0002-7270-9313
Manuel Esteban Lucas-Borja https://orcid.org/0000-0001-6270-8408
Sergio M Vicente-Serrano https://orcid.org/0000-0003-2892-518X
Santiago Beguería https://orcid.org/0000-0002-3974-2947
Simon Gascoin https://orcid.org/0000-0002-4996-6768

References

Alonso-González E, López-Moreno J I, Gascoin S, García-Valdecasas Ojeda M, Sanmiguel-Vallelado A, Navarro-Serrano F, Revuelto J, Ceballos A, Esteban-Parra M J and Essery R 2018 Daily gridded datasets of snow depth and snow water equivalent for the Iberian Peninsula from 1980 to 2014 *Earth Syst. Sci. Data* **10** 303–15

Alonso-González E, López-Moreno J I, Navarro-Serrano F M and Revuelto J 2020b Impact of North Atlantic oscillation on the snowpack in Iberian Peninsula mountains *Water* **12** 105

Alonso-González E, López-Moreno J I, Navarro-Serrano F, Sanmiguel-Vallelado A, Revuelto J, Domínguez-Castro F and Ceballos A 2020a Snow climatology for the mountains in the Iberian Peninsula using satellite imagery and simulations with dynamically downscaled reanalysis data *Int. J. Climatol.* **40** 477–91

Bischl B, Lang M, Kotthoff L, Schiffner J, Richter J, Studerus E, Casalicchio G and Jones Z M 2016 mlr: machine learning in R *J. Mach. Learn. Res.* **17** 1–5 (available at: https://mlr.mlr-org.com/authors.html)

Bonsoms J, López-Moreno J I and Alonso-González E 2023 Snow sensitivity to temperature and precipitation change during compound cold–hot and wet–dry seasons in the Pyrenees *Cryosphere* **17** 1307–26

Camarero J J, Gutiérrez E and Fortin M J 2000 Spatial pattern of subalpine forest-alpine grassland ecotones in the Spanish Central Pyrenees *For. Ecol. Manag.* **134** 1–16

Chen T and Guestrin C 2016 XGBoost: a scalable tree boosting system *Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* vol 13–17 (Association for Computing Machinery) pp 785–94 (available at: https://github.com/dmlc/xgboost)

Choler P 2015 Growth response of temperate mountain grasslands to inter-annual variations in snow cover duration *Biogeosciences* **12** 3885–97

Dunstone N, Smith D, Scaife A, Hermanson L, Eade R, Robinson N, Andrews M and Knight J 2016 Skilful predictions of the winter North Atlantic oscillation one year ahead *Nat. Geosci.* **9** 809–14

Eden C, Greatbatch R J and Lu J 2002 Prospects for decadal prediction of the North Atlantic oscillation (NAO) *Geophys. Res. Lett.* **29** 104-1–104–4

Edwards A C, Scalenghe R and Freppaz M 2007 Changes in the seasonal snow cover of alpine regions and its effect on soil processes: a review *Quat. Int.* **162–163** 172–81

Essery R 2015 A factorial snowpack model (FSM 1.0) *Geosci. Model Dev.* **8** 3867–76

Fayad A, Gascoin S, Faour G, López-Moreno J I, Drapeau L, Page M L and Escadafal R 2017 Snow hydrology in Mediterranean mountain regions: a review *J. Hydrol.* **551** 374–96

Fontana F, Rixen C, Jonas T, Aberegg G and Wunderle S 2008 Alpine grassland phenology as seen in AVHRR, VEGETATION, and MODIS NDVI time series—a comparison with in situ measurements *Sensors* **8** 2833–53

Francon L *et al* 2020 Assessing the effects of earlier snow melt-out on alpine shrub growth: the sooner the better? *Ecol. Indic.* **115** 106455

Gascoin S, Hagolle O, Huc M, Jarlan L, Dejoux J F, Szczypta C, Marti R and Sánchez R 2015 A snow cover climatology for the Pyrenees from MODIS snow products *Hydrol. Earth Syst. Sci.* **19** 2337–51

Gleason K E, Bradford J B, D'Amato A W, Fraver S, Palik B J and Battaglia M A 2021 Forest density intensifies the importance of snowpack to growth in water-limited pine forests *Ecol. Appl.* **31** e02211

Gouveia C M, Páscoa P, Russo A and Trigo R M 2016 Land degradation trend assessment over iberia during 1982–2012 *Cuad. Investig. Geogr.* **42** 89–112

Grobler B A and Cowling R M 2022 Which is the richest of them all? Comparing area-adjusted plant diversities of Mediterranean- and tropical-climate regions *Front. Biogeogr.* **14** e56241

Harpold A A, Molotch N P, Musselman K N, Bales R C, Kirchner P B, Litvak M and Brooks P D 2015 Soil moisture response to snowmelt timing in mixed-conifer subalpine forests *Hydrol. Process.* **29** 2782–98

Jonas T, Rixen C, Sturm M and Stoeckli V 2008 How alpine plant growth is linked to snow cover and climate variability *J. Geophys. Res. Biogeosci.* **113** G03013

López-Moreno J I, Pomeroy J W, Alonso-González E, Morán-Tejeda E and Revuelto J 2020 Decoupling of warming mountain snowpacks from hydrological regimes *Environ. Res. Lett.* **15** 114006

López-Moreno J I and Vicente-Serrano S M 2007 Atmospheric circulation influence on the interannual variability of snow pack in the Spanish Pyrenees during the second half of the 20th century *Hydrol. Res.* **38** 33–44

Lundberg S M and Lee S I 2017 A unified approach to interpreting model predictions *Advances in neural information processing systems* (*December 2017*) pp 4766–75

Macander M J, Swingley C S, Joly K and Raynolds M K 2015 Landsat-based snow persistence map for northwest Alaska *Remote Sens. Environ.* **163** 23–31

Martín-Alcón S, González-Olabarria J R and Coll L 2010 Wind and snow damage in the Pyrenees pine forests: effect of stand attributes and location *Silva Fenn.* **44** 399–410

Mazzotti G, Essery R, Moeser C D and Jonas T 2020 Resolving small-scale forest snow patterns using an energy balance snow model with a one-layer canopy *Water Resour. Res.* **56** e2019WR026129

Muñoz-Sabater J *et al* 2021 ERA5-Land: a state-of-the-art global reanalysis dataset for land applications *Earth Syst. Sci. Data* **13** 4349–83

Ninot J M, Carrillo E, Font X, Carreras J, Ferré A, Masalles R M, Soriano I and Vigo J 2007 Altitude zonation in the Pyrenees. A geobotanic interpretation *Phytocoenologia* **37** 371–98

Paruelo J M and Lauenroth W K 1998 Interannual variability of NDVI and its relationship to climate for North American shrublands and grasslands *J. Biogeogr.* **25** 721–33

Piedallu C, Chéret V, Denux J P, Perez V, Azcona J S, Seynave I and Gégout J C 2019 Soil and climate differently impact NDVI patterns according to the season and the stand type *Sci. Total Environ.* **651** 2874–85

Pirk N *et al* 2023 Snow–vegetation–atmosphere interactions in alpine tundra *Biogeosciences* **20** 2031–47

Razavi S, Hannah D M, Elshorbagy A, Kumar S, Marshall L, Solomatine D P, Dezfuli A, Sadegh M and Famiglietti J 2022 Coevolution of machine learning and process-based modelling to revolutionize Earth and environmental sciences: a perspective *Hydrol. Process.* **36** e14596

Revuelto J *et al* 2022 Intermediate snowpack melt-out dates guarantee the highest seasonal grasslands greening in the Pyrenees *Sci. Rep.* **12** 18328

Sanmiguel-Vallelado A, Camarero J J, Gazol A, Morán-Tejeda E, Sangüesa-Barreda G, Alonso-González E, Gutiérrez E, Alla A Q, Galván J D and López-Moreno J I 2019 Detecting snow-related signals in radial growth of Pinus uncinata mountain forests *Dendrochronologia* **57** 125622

Sanmiguel-Vallelado A, Camarero J J, Morán-Tejeda E, Gazol A, Colangelo M, Alonso-González E and López-Moreno J I 2021 Snow dynamics influence tree growth by controlling soil temperature in mountain pine forests *Agric. For. Meteorol.* **296** 108205

Serrano-Notivoli R, Beguería S, Saz M Á, Longares L A and De Luis M 2017 SPREAD: a high-resolution daily gridded precipitation dataset for Spain—an extreme events frequency and intensity overview *Earth Syst. Sci. Data* **9** 721–38

Shabanov N V *et al* 2005 Analysis and optimization of the MODIS leaf area index algorithm retrievals over broadleaf forests *IEEE Trans. Geosci. Remote Sens.* **43** 1855–65

Sturm M, Holmgren J and Liston G E 1995 A seasonal snow cover classification system for local to global applications *J. Climatol.* **8** 1261–83

Trujillo E, Molotch N P, Goulden M L, Kelly A E and Bales R C 2012 Elevation-dependent influence of snow accumulation on forest greening *Nat. Geosci.* **5** 705–9

Vicente-Serrano S M *et al* 2020 Vegetation greening in Spain detected from long term data (1981–2015) *Int. J. Remote Sens.* **41** 1709–40

Vicente-Serrano S M and Heredia-Laclaustra A 2004 NAO influence on NDVI trends in the Iberian peninsula (1982–2000) *Int. J. Remote Sens.* **25** 2871–9

Viviroli D, Dürr H H, Messerli B, Meybeck M and Weingartner R 2007 Mountains of the world, water towers for humanity: typology, mapping, and global significance *Water Resour. Res.* **43** W07447

Wylie B K, Meyer D J, Tieszen L L and Mannel S 2002 Satellite mapping of surface biophysical parameters at the biome scale over the North American grasslands a case study *Remote Sens. Environ.* **79** 266–78

Xie J, Jonas T, Rixen C, de Jong R, Garonna I, Notarnicola C, Asam S, Schaepman M E and Kneubühler M 2020 Land surface phenology and greenness in Alpine grasslands driven by seasonal snow and meteorological factors *Sci. Total Environ.* **725** 138380