The Cryosphere
Discussions

Deep learning applied to glacier evolution modelling

Jordi Bolibar[1,2], Antoine Rabatel[1], Isabelle Gouttevin[3], Clovis Galiez[4], Thomas Condom[1], Eric Sauquet[2]

[1]Univ. Grenoble Alpes, CNRS, IRD, G-INP, Institut des Géosciences de l'Environnement (IGE, UMR 5001), Grenoble, France
[2]Irstea, UR RiverLy, Lyon-Villeurbanne, France
[3]Univ. Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre d'Études de la Neige, Grenoble, France
[4]Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, Grenoble, France

Correspondence to: Jordi Bolibar (jordi.bolibar@univ-grenoble-alpes.fr)

Abstract. We present a parameterized glacier evolution model, with a surface mass balance (SMB) component based on a deep artificial neural network (*i.e.* deep learning). While most glacier models tend to incorporate more and more physical processes, here we take an alternative approach by creating a parameterized model based on data science. Annual glacier-wide SMBs can be simulated using either deep learning or Lasso (regularized multilinear regression), whereas the glacier geometry is updated using a glacier-specific parameterization. We compare and cross-validate our nonlinear deep learning SMB model against other standard linear statistical methods on a dataset of 32 French alpine glaciers. Deep learning is found to outperform linear methods, with improved explained variance (up to +64% in space and +108% in time) and accuracy (up to +47% in space and +58% in time), resulting in an estimated r^2 of 0.77 and RMSE of 0.51 m.w.e. Substantial nonlinear structures are captured by deep learning, with around 35% of nonlinear behaviour in the temporal dimension. For the glacier geometry evolution, the main uncertainties come from the ice thickness data used to initialize the model. These results should encourage the use of deep learning in glacier modelling as a powerful nonlinear tool, capable of capturing the nonlinearities of the climate and glacier systems, that can serve to reconstruct or simulate SMB time series for individual glaciers at regional scale for past and future climates.

1 Introduction

Glaciers are arguably one of the most important icons of climate change, being climate proxies which can clearly depict the evolution of climate for the global audience (IPCC, 2018). For the coming decades, mountain glaciers will be some of the most important contributors to sea level rise and will most likely drive important changes in the hydrological regime of glaciarized catchments (Beniston et al., 2018; Vuille et al., 2018; Hock et al., 2019). The reduction in ice volume may produce an array of consequences which requires to be properly predicted. These consequences will strongly depend on the future climatic scenarios, which will determine the timing and magnitude for the transition of hydrological regimes (Huss and Hock, 2018).

The Cryosphere
Discussions

Glacier and hydro-glaciological models can help answer these questions, giving several possible outcomes depending on multiple climate scenarios. For any glacier model, two main processes have to be simulated: the glacier surface mass balance (SMB) and the glacier dynamics; which can be done using approaches of different complexity. In order to model these processes at large scale (*i.e.* on several glaciers at a catchment scale), some compromises need to be made, which can be approached in different ways:

   (1) Parameterized models do not explicitly resolve any physical processes, but implicitly take them into account using parameterizations based on statistical or empirical relationships. This type of models range from very simple statistical models (Carlson et al., 2014) to more complex ones based on different approaches, such as a calibrated equilibrium-line altitude (ELA) model (Zemp et al., 2006) or volume/length-area scaling (Marzeion et al., 2012; Radić et al., 2014).

   (2) Process-based models, like GloGEM (Huss and Hock, 2015), GloGEMflow (Zekollari et al., 2019) and OGGM (Maussion et al., 2019), approximate a number of glacier physical processes and surface energy and mass balance in a simplified manner.

   (3) Physics-based models, like the finite elements Elmer/Ice model (Gagliardini et al., 2013), approach glacier dynamics by explicitly simulating physical processes and solving the full Stokes equations (Jouvet et al., 2009; Réveillet et al., 2015). They are generally not suitable for regional studies due to their high complexity.

At the same time, these different approaches strongly depend on available data, whose spatial and temporal resolutions have an important impact on the results' quality and uncertainties (*e.g.*, Réveillet et al., 2018). Parameterized models require a reference or training dataset to calibrate the relationships, which can then be used for projections with the hypothesis that relationships remain stationary. On the contrary, process-based and specially physics-based models have the advantage of representing physical processes, but they require larger datasets at higher spatial and temporal resolutions with a consequently higher computational cost (Réveillet et al., 2018). Meteorological reanalyses provide an attractive alternative to sparse point observations, although their spatial resolution and suitability to complex high-mountain topography are often not good enough for high-resolution physics-based glacio-hydrological applications. However, parameterized models are much more flexible, equally dealing with fewer and coarser meteorological data as well as the state of the art reanalyses, which allows to work at resolutions much closer to glaciers' scale and to reduce uncertainties. The current resolution of climate projections is still too low to adequately drive most glacier physical processes, but the ever-growing datasets of historical data are paving the way for the training of parameterized machine learning models.

In glaciology, statistical models have been applied for more than half a century, starting with simple multiple linear regressions on few meteorological variables (Hoinkes, 1968; Martin, 1974). Statistical modelling has made enormous progress in the last decades, specially thanks to the advent of machine learning. Compared to other fields in geosciences, the glaciological community has remained quite oblivious to these advances, mostly focusing on glacier physics and physical-based and process-

based models. Despite this fact, a number of studies have taken steps towards statistical approaches. Steiner et al., (2005) pioneered the very first study to use artificial neural networks (ANNs) in glaciology to simulate mass balances of the Grosse Aletschgletscher in Switzerland. They showed that a nonlinear model is capable of better simulating glacier mass balances compared to a conventional stepwise multiple linear regression. Furthermore, they found a significant nonlinear part within

5   the climate/glacier mass balance relationship. This work was continued in Steiner et al. (2008) and Nussbaumer et al., (2012) for the simulation of glacier length instead of mass balances. Later on, Maussion et al., (2015) developed an empirical statistical downscaling tool based on machine learning in order to retrieve glacier surface energy and mass balance (SEB/SMB) fluxes from large-scale atmospheric data. They used different machine learning algorithms, but all of them were linear, which are not necessarily the most suitable for modelling the nonlinear climate system (Houghton et al., 2001). Nonetheless, more recent

10 developments in the field of machine learning and optimization enabled the use of deeper network structures than the 3-layer ANN of Steiner et al. (2005). These deeper ANNs, which remain unexploited in glaciology, allow to capture more nonlinear structures in the data even for relatively small datasets (Ingrassia and Morlini, 2005; Olson et al., 2018).

Here, we present a parameterized open-source glacier model: the ALpine Parameterized Glacier Model (ALPGM, Bolibar,

15 2019). When most glacier models tend to incorporate more and more physical processes (Maussion et al., 2019; Zekollari et al., 2019), ALPGM takes an alternative approach based on data science. ALPGM simulates annual glacier-wide SMB and the evolution of glacier volume and surface area over time scales from a few years to a century. Glacier-wide SMBs are computed using a deep ANN, fed by several topographical and climatic variables, an approach which is compared to different linear methods in the present paper. In order to distribute these annual glacier-wide SMBs and to update the glacier geometry, a

20 refined version of the Δh methodology (*e.g.*, Huss et al., 2008) is used, for which we dynamically compute glacier-specific Δh functions. In order to validate this approach, we use a case study with 32 French alpine glaciers for which glacier-wide annual SMBs are available over the period 1984-2014 and 1959-2015 for certain glaciers. High resolution meteorological reanalyses for the same time period are used (SAFRAN: Durand et al., 2009) while the initial ice thickness distribution of glaciers are taken from Farinotti et al. (2019), for which we performed a sensitivity analysis based on field observations.

25

In the next section, we present an overview of the proposed glacier evolution model framework with a detailed description of the two components used to simulate the annual glacier-wide SMB and the glacier geometry update. Then, a case study using French alpine glaciers is presented, which enables to illustrate an example of application of the proposed framework including a rich dataset, the parameterized functions, as well as the results and their performance. In the end, several aspects regarding

30 machine and deep learning modelling in glaciology are discussed, from which we make some recommendations and draw the final conclusions.

2 Model overview and methods

In this section we present an overview of the ALPGM glacier model. Moreover, the two components of this model are presented in detail: the Glacier-wide SMB Simulation component and the Glacier Geometry Update component.

2.1 Model overview and workflow

5   ALPGM is an open-source glacier model coded in Python. The source code of the model is accessible in the project repository (see Code availability). It is structured in multiple files which execute specific separate tasks. The model can be divided into two main components: (1) the Glacier-wide SMB Simulation and (2) the Glacier Geometry Update. The Glacier-wide SMB Simulation component is based on machine learning, taking both meteorological and topographical variables as inputs. The Glacier Geometry Update component generates the glacier-specific parameterized functions and modifies annually the

10   geometry of the glacier (*e.g.* ice thickness distribution, glacier outline) based on the glacier-wide SMB values simulated by the Glacier-wide SMB simulation component.
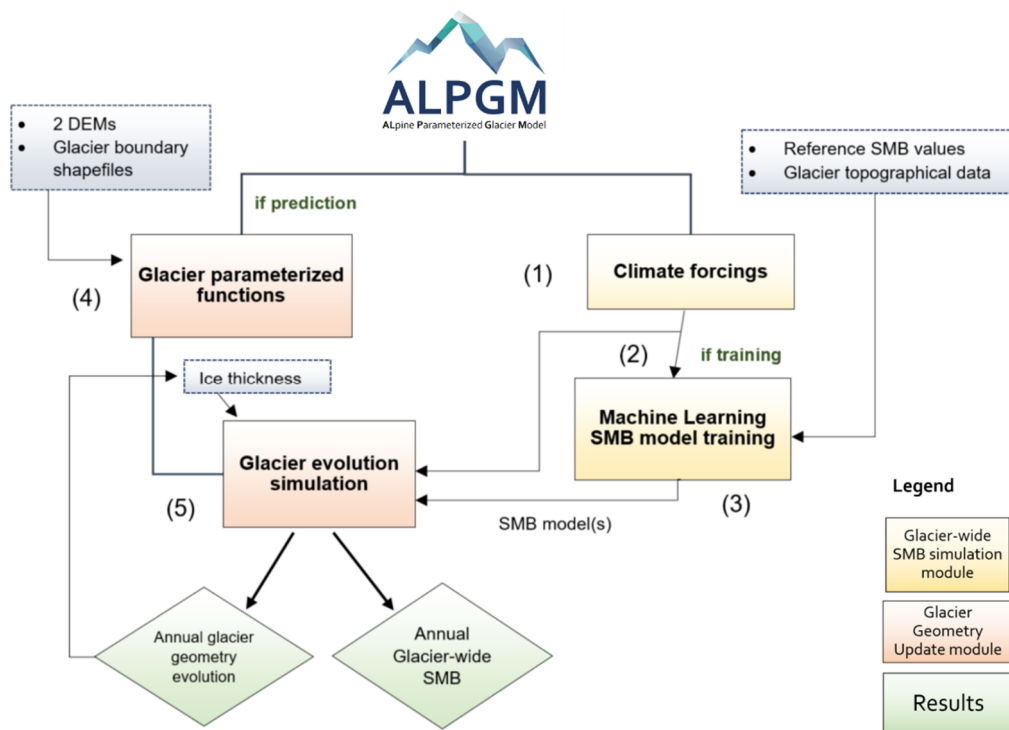

15   **Figure 1:** ALPGM structure and workflow

Fig. 1 presents ALPGM's basic workflow. The workflow execution can be configured *via* the model interface, allowing to run or skip any of the following steps:

(1) The meteorological forcings are preprocessed in order to extract the necessary data closest to each glacier's centroid. The meteorological features are stored in intermediate files in order to reduce computation times for future runs, automatically skipping this preprocessing step when the files have already been generated.

(2) The SMB machine learning component retrieves the preprocessed meteorological features and assembles the spatiotemporal training dataset, comprised by both climatic and topographical data. An algorithm is chosen for the SMB model, which can be loaded from a previous run or it can be trained again with a new dataset. These model(s) are stored in intermediate files, allowing to skip this step for future runs.

(3) Performances of the SMB models can be evaluated with a leave-one-glacier-out (LOGO) or a leave-one-year-out (LOYO) cross-validation. This step can be skipped when using already established models. Basic statistical performance metrics are given for each glacier and model, as well as plots with the simulated cumulative glacier-wide SMBs compared to their reference values with uncertainties for each of the glaciers from the training dataset.

(4) The Glacier Geometry Update component starts with the generation of the glacier specific parameterized functions, using a raster containing the difference of the two pre-selected digital elevation models (DEMs) covering the study area for two separate dates, as well as the glacier contours. These parameterized functions are then stored in individual files to be used in the final simulations.

(5) Once all previous steps have been run, the final simulations are launched. For each glacier, the initial ice thickness raster and the parameterized function are retrieved. The meteorological data at the glaciers' centroid is re-computed with an annual time step based on each glacier's evolving topographical characteristics. These forcings are used to simulate the annual glacier-wide SMB using the machine learning model. Once an annual glacier-wide SMB value is obtained, the changes in geometry are computed using the parameterized function, thus updating the glacier's DEM and ice thickness rasters. If all the ice thickness raster pixels of a glacier become zero, the glacier is considered as disappeared and is removed from the simulation pipeline. For each year, multiple results are stored in data files as well as the raster DEM and ice thickness values for each glacier.

2.2 Glacier-wide surface mass balance simulation

Annual glacier-wide SMBs are simulated using machine learning. Although the features used as input for the model are classical descriptors of the topographical and meteorological conditions of the glaciers, it is worth mentioning that applying the model in different areas or with different data sources would likely require a re-training of the model due to possible biases: different regions on the globe may have other descriptors of importance but also different measuring techniques will likely have different biases.

2.2.1 Selection of explanatory topographical and climatic variables

In order to narrow down which topographical and climatic variables best explain glacier-wide SMB in a given study area, a literature review as well as a statistical sensitivity analysis are performed. Typically used topographical predictors are longitude, latitude, glacier slope and mean altitude. As for meteorological predictors, cumulative positive degree days (CPDD),

5 but also mean monthly temperature, snowfall and possibly other variables that influence the surface energy budget are often used in the literature. Examples of both topographic and meteorological predictors can be found in the case study in Sect. 3. A way to prevent biases when making predictions with different climate data is to work with anomalies, calculated as differences of the variable with respect to its average value over a chosen reference period.

10 For the machine learning training, the relevant predictors must be selected, performing a sensitivity study of the annual glacier-wide SMB to topographical and climatic variables over the study training period. This can be performed with individual linear regressions between each variable and glacier-wide SMB data. After identification of the topographical and climatic variables that can potentially explain annual glacier-wide SMB variability for the region of interest, a training dataset is built. An effective way of expanding the training dataset in order to dig deeper into the available data is to combine the climatic and

15 topographical input variables. Such combinations can be expressed following Eq. (1):

$$SMB_{g,y} = f\left(\widehat{\Omega},\ \widehat{C}\right) + \varepsilon_{g,y} \qquad (1)$$

Where $\widehat{\Omega}$ is a vector of the selected topographical predictors, $\widehat{C}$ is a vector with the selected climatic features and $\varepsilon_{g,y}$ is the

20 residual error for each annual glacier-wide SMB value, $SMB_{g,y}$.

Once the training dataset is created, different algorithms f (two linear and one nonlinear, for the case of this study) can be chosen to create the SMB model: (1) OLS (Ordinary Least Squares) all-possible multiple linear regressions; (2) Lasso (Least absolute shrinkage and selection operator) (Tibshirani, 1996); and (3) a deep Artificial Neural Network (ANN). ALPGM uses

25 some of the most popular machine learning Python libraries: StatsModel (Seabold and Perktold, 2010), Scikit-learn (Pedregosa et al., 2012) and Keras (Chollet, 2015) with a TensorFlow backend. The overall workflow of the machine learning glacier-wide SMB model production in ALPGM is summarized in Fig. 2.
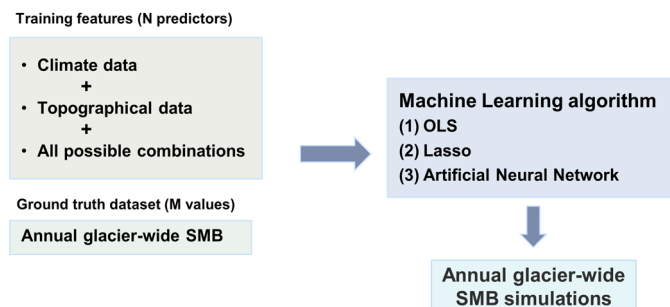
Figure 2: Glacier-wide SMB simulation component workflow. Machine learning models are dynamically created based on training data.

2.2.2 All-possible multiple linear regressions

With the ordinary least squares (OLS) all-possible multiple linear regressions, we attempt to find the best subset of predictors in Eq. 1 based on the resulting r^2 adjusted, while at the same time avoiding overfitting (Hawkins, 2004) and collinearity, and limiting the complexity of the model. As its name indicates, the goal is to minimize the residual sum of squares for each subset of predictors (Hastie et al., 2009). n models are produced by selecting all possible subsets of k predictors. It is advisable to narrow down the number of predictors for each subset in the search to reduce the computational cost. Models with low performance are filtered out, keeping only models with highest r^2 adjusted possible, a variance inflation factor (VIF) < 1.2 and a p-value < 0.01/n (in order to ensure the Bonferroni correction). Retained models are combined by averaging their predictions, thereby avoiding the pitfalls related to stepwise single model selection (Whittingham et al., 2006). These criteria ensure that the models explain as much variability as possible, avoid collinearity and are statistically significant.

2.2.3 Lasso

The Lasso (Least absolute shrinkage and selection operator) (Tibshirani, 1996) is a shrinkage method which attempts to overcome the shortcomings of the simpler step-wise and all-possible regressions. In these two classical approaches, predictors are discarded in a discrete way, giving subsets of variables which have the lowest prediction error. However, due to its discrete selection, these different subsets can exhibit high variance, which does not reduce the prediction error of the full model. The Lasso performs a more continuous regularization by shrinking some coefficients and setting others to zero, thus producing more interpretable models (Hastie et al., 2009). Because of its properties, it strikes a balance between subset selection (like all-possible regressions) and Ridge regression (Hoerl and Kennard, 1970). In order to determine the degree of regularisation applied to the coefficients used in the linear OLS regression, an alpha parameter needs to be chosen using cross-validation. ALPGM performs different types of cross-validations to choose from: the Akaike Information Criterion (AIC), the Bayes Information Criterion (BIC) and a classical cross-validation with iterative fitting along a regularization path. Alternatively, a Lasso model with Least Angle Regression, also known as Lasso Lars (Tibshirani et al., 2004), can also be chosen with a

classical cross-validation. All the input data is scaled and centered to zero before training the model. The generated coefficients from the model serve to determine the significant predictors to be kept for the artificial neural network training.

2.2.4 Deep artificial neural network

Artificial neural networks (ANNs) are nonlinear statistical models inspired by biological neural networks (Fausett, 1994; Hastie et al., 2009). A neural network is characterized by: (1) the architecture or pattern of connections between units and the number of layers; (2) the optimizer: the method for determining the weights of the connections between units; and (3) its (possibly nonlinear) activation functions (Fausett, 1994). The description of neural networks is beyond the scope of this study, so for more details and a full explanation please refer to Fausett (1994), Hastie et al. (2009), as well as Steiner et al. (2005, 2008) where the reader can find a thorough introduction to the use of ANNs in glaciology. ANNs gained recent interest thanks to improvements of optimization algorithms allowing to train deep neural networks, that lead to better representation of complex data patterns. As their learnt parameters are difficult to interpret, ANNs are best suited when the quality of predictions prevails over the interpretability of the model. Their ability to model complex functions of the input parameters makes them particularly suitable for modelling complex nonlinear systems such as the climate system (Houghton et al., 2001) and glacier systems (Steiner et al., 2005).

ALPGM uses a feedforward fully-connected ANN (Fig. 3). In such an architecture, the processing units - or neurons - are grouped into layers where all the units of a given layer are fully connected to all units of the next layer. The flow of information is directional, from the input layer (*i.e.* in which each neuron corresponds to one of the *N* explanatory variables) to the output neuron (*i.e.* corresponding to the target variable of the model, the SMB). For each connection of the ANN, weights are initialized in a random fashion following a specific distribution (generally centred around 0). In each unit of each hidden layer, the weighted values are summed before going through a nonlinear activation function, responsible for introducing the nonlinearities in the model. Using a series of iterations known as epochs, the ANN will try to minimize a specific loss function (the mean squared error (MSE) in our case) comparing the processed values of the output layer with the ground truth (y). In order to avoid falling into local minima of the loss function, some regularisation is needed to prevent the ANN from overfitting (Hastie et al., 2009). To prevent overfitting during the training process (*i.e.* to increase the ability of the model to generalize to new data), we used a classical regularization method called dropout, consisting in training iteratively smaller subparts of the ANN by randomly disconnecting a certain amount of connections between units. The introduction of Gaussian noise at the input of the ANN also helped to generalize, as it performs a similar effect to data augmentation. The main consequence of regularisation is generalization, for which the produced model is capable of better adapting to different configurations of the input data.
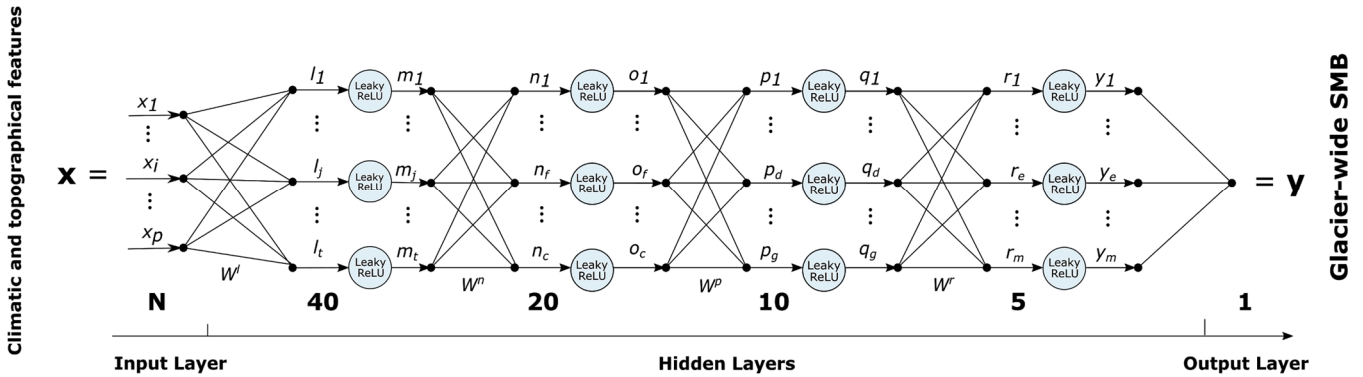
Figure 3: Deep Artificial Neural Network architecture used in ALPGM. The numbers indicate the number of neurons in each layer

The hyperparameters used to configure the ANN are determined using cross-validation, in order to find the best performing combination of number of units, hidden layers, activation function, learning rate and regularisation method. Due to the relatively small size of our dataset, we encountered the best performances with a quite small deep ANN, with a total of 6 layers (4 hidden layers) with a (N, 40, 20, 10, 5, 1) architecture (Fig. 3), where N is the number of selected features. Since the ANN already performs all the possible combinations between features (predictors), we use a reduced version of the training matrix from Eq. 1, with no combination of climatic and topographical features. Due to the relatively small size of the architecture, the best dropout rates are small (Srivastava et al., 2014), and range between 0.3 and 0.01 depending on the number of units of each hidden layer. Leaky ReLUs have been chosen as the activation function, because of their widespread reliability and the fact they help prevent the "dead ReLU" problem, where certain neurons can stop "learning" (Xu et al., 2015). The He uniform initialization (He et al., 2015) has been used as it is shown to work well with Leaky ReLUs, and all unit bias were initialized to zero. In order to optimize the weights of the gradient descent, we used the RMSprop optimizer, for which we fine-tuned the learning rate, obtaining the best results at 0.0005 in space and 0.02 in time. Each batch was normalized before applying the activation function in order to accelerate the training (Ioffe and Szegedy, 2015).

Like for many other geophysical processes found in nature, extreme annual glacier-wide SMB values occur much less often than average values, approximately following an unbounded Gumbel-type distribution (Thibert et al., 2018). From a statistical point of view, this means that ANN will "see" few extreme values and will accord less importance to them. For future projections in a warmer climate, extreme positive glacier-wide SMB balances should not be the main concern of glacier models. However, extreme negative annual glacier-wide SMB values should likely increase in frequency, so it is in the modeller's interest to reproduce them as well as possible. Setting the sample weights as the inverse of the probability density function during the ANN training can partly compensate for the imbalance of a dataset. This boosts the performance of the model for the extreme values, at the cost of sacrificing some performance on more average values, which can be seen as a

r^2/RMSE trade-off (see Fig. 6 and 9 from the case study). The correct setting of the sample weights allows the modeller to adapt the ANN to each dataset and application.

2.3 Glacier geometry update

Since the first component of ALPGM simulates annual glacier-wide SMBs, these changes in mass need to be redistributed over the glacier surface-area in order to reproduce glacier dynamics. This redistribution is applied using the Δh parameterization. The idea was first developed by Jóhannesson et al. (1989) and then adapted and implemented by Huss et al. (2008). The main idea behind it is to use two or more DEMs covering the study area. These DEMs should have dates covering a period long enough (which will be later discussed in detail). By subtracting them, the changes in glacier surface elevation over time can be computed, which corresponds to a change in thickness (considering no basal erosion). Then, these thickness changes are normalized and considered as a function of the normalized glacier altitude. This Δh function is specific for each glacier and represents the normalized glacier thickness evolution over its altitudinal range. One advantage of such a parametrized approach is that it implicitly considers the ice flow which redistributes the mass from the accumulation to the ablation area. In order to make the glacier volume evolve in a mass-conserving fashion, we apply this function to the annual glacier-wide SMB values in order to scale and distribute its change in volume.

As discussed in Vincent et al. (2014), the time period between the two DEMs used to calibrate the method needs to be long enough to show important ice thickness differences. The criteria will of course depend on each glacier and each period, but it will always be related to the achievable signal-to-noise ratio. Vincent et al. (2014) concluded that for their study on the Mer de Glace glacier (28.8 km², mean altitude = 2868 m.a.s.l.) in the French Alps, the 2003-2008 period was too short, due to the delayed response of glacier geometry to a change in surface mass balance. Indeed, the results for that 5-year period diverged from the results from longer periods. Moreover, the period should be long enough to be representative of the glacier evolution, which will often encompass periods with strong ablation and others with no retreat or even with positive SMBs.

Therefore, by subtracting the two DEMs, the ice thickness difference is computed for each specific glacier. These values can then be classified by altitude, thus obtaining an average glacier thickness difference for each pixel altitude. As a change to previous studies (Vincent et al., 2014; Huss and Hock, 2015; Hanzer et al., 2018; Zekollari et al., 2019), we no longer work with altitudinal transects, but with individual pixels. In order to filter noise and artefacts coming from the DEM raster files, different filters are applied to remove outliers and pixels with unrealistic values, namely at the border of glaciers or where the surface slopes are high (refer to Supplements for detailed information). Our methodology thus allows to better exploit the available spatial information based on its quality, and not on arbitrary location within transects.

3 Case study: French alpine glaciers

3.1 Data

All data used in this case study is based on the French Alps, located in the westernmost part of the European Alps, between 5.08° and 7.67°E, and 44° and 46°13'N. This region is particularly suited for the validation of a glacier evolution model

5 because of the wealth of available data. Moreover, ALPGM has been developed as part of a hydro-glaciological study to understand the impact of the retreat of French alpine glaciers in the Rhône river catchment (97,800 km^2).

3.1.1 Glacier-wide surface mass balance

An annual glacier-wide SMB dataset reconstructed using remote sensing based on changes in glacier volume and the snow line altitude is used (Rabatel et al., 2016). This dataset is constituted by annual glacier-wide SMB values for 30 glaciers in the

10 French Alps (Fig. 4) for 31 years, between 1984-2014. The great variety in topographical characteristics of the glaciers included in the dataset, with a good coverage of the three main clusters or groups of glaciers in the French Alps (Fig. 4), makes them an ideal training dataset for the model. Each of the clusters represents a different setup of glaciers with different contrasting latitudes (Écrins and Mont-Blanc), longitudes (Écrins and Vanoise), glacier size (smaller glaciers in Écrins and Vanoise *vs.* larger ones in Mont-Blanc) and climatic characteristics with a Mediterranean influence towards the south of the study region.

15 For more details regarding this dataset refer to Rabatel et al. (2016). Data from the Mer de Glace, Saint-Sorlin, Sarennes and Argentière glaciers is also used, coming from field observations from the GLACIOCLIM observatory. For some of these glaciers, glacier-wide SMB values are available since 1949, although only values from 1959 onwards were used to match the meteorological reanalysis. This makes a total of 32 glaciers (Argentière and Saint-Sorlin glaciers belonging to the two datasets), representing 1048 annual glacier-wide SMB values (taking into account some gaps in the dataset).
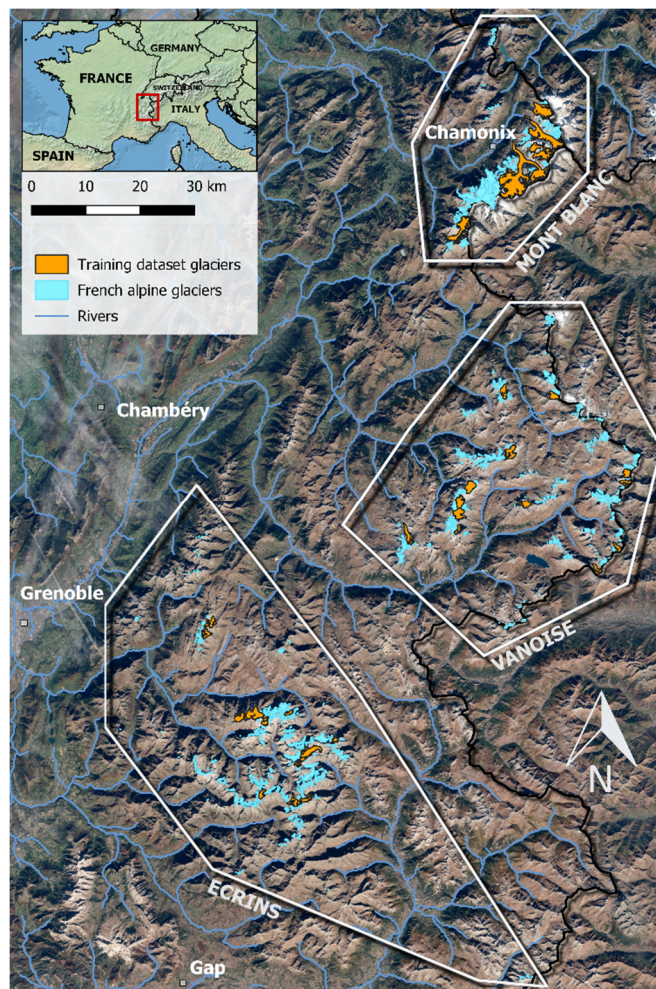
20

Figure 4: French alpine glaciers used for model training and validation and their

classification into 3 clusters/regions (Ecrins, Vanoise, Mont Blanc)

5   3.1.2 Topographical glacier data and altimetry

All topographical data for this study has been taken from the multitemporal inventory of the French Alps glaciers (*e.g.*, Gardent et al., 2014) partly available through the GLIMS Glacier Database (GLIMS and NSIDC, 2005). We worked with the 1967, 1985, 2003 and 2015 inventories (Gardent et al., 2014; with 2015 update). Two DEMs covering the whole French Alps have been used: (1) one from 2011 generated from SPOT5 stereo-pair images, acquired on 15 October 2011; and (2) a 1979 aerial

10   photogrammetric DEM from the French National Geographic Institute (Institut Géographique National, IGN), processed from aerial photographs taken around 1979. Both DEMs have an accuracy between 1 and 4 meters (Rabatel et al., 2016), and their uncertainties are negligible compared to many other parameters in this study.

3.1.3 Glacier ice thickness

Glacier ice thickness data come from Farinotti et al. (2019), hereafter F19, based on the Randolph Glacier Inventory (RGIv6.0). The ice thickness values represent the latest consensus estimate, averaging an ensemble of different methods based on the principles of ice flow dynamics to invert the ice thickness from surface characteristics.

5

We also have ice thickness data acquired by diverse field methods (seismic, ground penetrating radar or hot water drilling, Rabatel et al., 2018) for four glaciers of the GLACIOCLIM observatory. We compared these *in situ* thickness data, with the simulated ice thicknesses from F19 (refer to Supplements for detailed information). Although differences can be found (locally up to 100% in the worst cases), no systematic biases were found with respect to glacier local slope nor glacier altitude;
10 therefore, no systematic correction was applied to the dataset. The simulated ice thicknesses for Saint-Sorlin (2 km^2, mean altitude = 2920 m.a.s.l., Écrins cluster) and Mer de Glace (28 km^2, mean altitude = 2890 m.a.s.l., Mont-Blanc cluster) glaciers are quite satisfactorily modelled by F19, with a good overall agreement except for the local distribution in certain altitudinal ranges. The ice thicknesses for Argentière Glacier (12.8 km^2, mean altitude = 2808 m.a.s.l., Mont-Blanc cluster) and Glacier Blanc (4.7 km^2, mean altitude = 3196 m.a.s.l., Écrins cluster) are underestimated by F19 with an almost constant bias with
15 respect to altitude, as seen in Rabatel et al. (2018). Therefore, a manual correction was applied to the F19 datasets for these two glaciers based on the field observations from the GLACIOCLIM observatory. A detailed plot (Fig. S2) presenting these results can be found in the supplementary material.

3.1.5 Climate data

In our French-Alps case study, ALPGM is forced with daily mean near-surface (2 m) temperatures, daily cumulative snowfall
20 and rain. The SAFRAN dataset is used to provide this data close to the glaciers' centroids. SAFRAN meteorological data (Durand et al., 2009) is a reanalysis of weather data including observations from different networks, and specific to the French mountain regions (Alps, Pyrenees and Corsica). Instead of being structured as a grid, data is provided at the scale of massifs, which are in turn divided into altitude bands of 300 meters and into 5 different aspects (north, south, east, west and flat).

3.2 Glacier-wide surface mass balance simulations: validation and results
25

In this section, we go through the selection of SMB predictors, the building of the machine learning SMB models, we assess their performance in space and time and we show some results of simulations using the French alpine glaciers dataset.

3.2.1 Selection of predictors

Statistical relationships between meteorological and topographical variables with respect to glacier-wide SMB are frequent in the literature for the European Alps (Hoinkes, 1968). Martin (1974) performed a sensitivity study on the SMB of the Saint-Sorlin and Sarennes glaciers (French Alps) with respect to multi-annual meteorological observations for the 1957-1972 period.

5   Martin (1974) obtained a multiple linear regression function based on annual precipitation and summer temperatures, and he concluded that it could be further improved by differentiating winter and summer precipitations. Six and Vincent (2014) studied the sensitivity of the SMB to climate change in the French Alps from 1998 until 2014. They found that the variance of summer SMB is responsible for over 90% of the variance of the annual glacier-wide SMB. Rabatel et al. (2016, 2013) performed an extensive sensitivity analysis of different topographical variables (slope of the lowermost 20% of the glacier

10   area, mean elevation, surface area, length, minimum elevation, maximum elevation, surface area change and length change) with respect to glacier ELA and annual glacier-wide SMBs of French alpine glaciers. Together with Huss (2012), who performed a similar study with SMB, the most significant statistical relationships were found for the lowermost 20% area slope, the mean elevation, glacier surface area, aspect and easting and northing. Rabatel et al. (2013) also determined that the climatic interannual variability is mainly responsible for driving the glacier equilibrium-line altitude temporal variability,

15   whereas the topographical characteristics are responsible for the spatial variations in the mean ELA.

Summer ablation is often accounted for by means of cumulative positive degree days (CPDD). However, in the vast majority of studies, accumulation and ablation periods are defined between fixed dates (*e.g.*, 1st October - 30th April for the accumulation period in the northern mid-latitudes) based on optimizations. As discussed in Zekollari and Huybrechts (2018), these fixed

20   periods may not be the best to describe SMB variability through statistical correlation. Moreover, the ablation season will likely evolve in the coming century, due to climate warming. In order to overcome these limitations, we dynamically calculate each year the transition between accumulation and ablation seasons (and *vice-versa*) based on a chosen quantile in the CPDD (Fig. S3). We found higher correlations between annual SMB and ablation-period CPDD calculated using this dynamical ablation season. On the other hand, it was not the case for the separation between summer and winter snowfall. Therefore, we

25   decided to keep constant periods to account for winter (1st October-1st May) and summer (1st May-1st October) snowfalls, and to keep them dynamical for the CPDD calculation.

Following this literature review, vectors $\widehat{\Omega}$ and $\widehat{C}$ from (Eq. 1) read as:

$$\widehat{\Omega} = [\ \overline{Z}\quad Z_{max}\quad \alpha_{20\%}\quad Area\quad Lat\quad Lon\quad \underline{\Phi}\ ] \qquad (2)$$

$$\widehat{C} = [\ \Delta CPDD\quad \Delta WS\quad \Delta SS\quad \Delta \overline{T}_{mon}\quad \Delta \overline{S}_{mon}\ ] \qquad (3)$$

5

Where:

$\overline{Z}$: Mean glacier altitude

Z_{max}: Maximum glacier altitude

$\alpha_{20\%}$: Slope of the lowermost 20% glacier altitudinal range

10 $Area$: Glacier surface area

Lat: Glacier latitude

Lon: Glacier longitude

$\underline{\Phi}$: Cosine of the glacier's aspect (North = 0º)

$\Delta CPDD$: CPDD (Cumulated Positive Degree Days) anomaly

15 ΔWS: Winter snow anomaly

ΔSS: Summer snow anomaly

$\Delta \overline{T}_{mon}$: Average temperature anomaly for each month for the hydrological year

$\Delta \overline{S}_{mon}$: Average snowfall anomaly for each month for the hydrological year

20 For the linear machine learning model training, we chose a function f that combines $\widehat{\Omega}$ and $\widehat{C}$, generating new combined predictors (Eq. 4):

$$
\begin{aligned}
SMB_{g,y} = &(a_1\overline{Z} + a_2 Z_{max} + a_3\alpha_{20\%} + a_4 Area + a_5 Lat + a_6 Lon + a_7\underline{\Phi} + a_8)\Delta CPDD\ + \\
&(b_1\overline{Z} + b_2\ Z_{max} + b_3\alpha_{20\%} + b_4 Area + b_5 Lat + b_6 Lon + b_7\underline{\Phi} + b_8)\Delta SS\ + \\
&(c_1\overline{Z} + c_2 Z_{max} + c_3\alpha_{20\%} + c_4 Area + c_5 Lat + c_6 Lon + c_7\underline{\Phi} + c_8)\Delta WS\ + \\
&d_1\overline{Z} + d_2 Z_{max} + d_3\alpha_{20\%} + d_4 Area + d_5 Lat + d_6 Lon + d_7\underline{\Phi} + d_n\Delta\overline{T}_{mon} + d_m\Delta\overline{S}_{mon} + \varepsilon \Big)_{g,y}
\end{aligned}
$$

25

$$(4)$$

This combination produces a total of 50 predictors for our 32 glaciers over variable periods between 31 and 57 years, resulting in a 1048x50 matrix used for OLS and Lasso. For the training of the ANN, no combination of topo-climatic features is done as previously mentioned.

3.2.2 Causal analysis

By running the Lasso algorithm on the dataset based on Eq. 4, we obtain the contribution of each predictor in order to explain the annual glacier-wide SMB variance. Interestingly, not all topographical variables found in the literature were kept by Lasso, which only found significant linear relationships with the average slope of the lowermost 20% of the glacier and with the latitude and longitude. Regarding the climatic variables, the annual CPDD as well as the winter and summer snowfall appear as significant predictors as well as several monthly mean temperatures and snowfall values (Fig. 5). Latitude and longitude seem to play an important role when combined with snowfall. Indeed, glaciers located in the eastern part of the Vanoise cluster (Fig. 4) are likely more affected by eastern precipitation fluxes, whereas the western glaciers in the Écrins and Mont-Blanc clusters are mostly affected by western disturbances. Latitude plays an important role as well, with Mont-Blanc glaciers receiving higher amounts of snowfall compared to the Mediterranean-influenced ones in Écrins. These findings are in agreement with Rabatel et al. (2013) and the general snow climatology in the French Alps (Durand et al., 2009). On the other hand, the meteorological conditions are also crucial for the annual glacier-wide SMB in the French Alps, especially in the transition months: 1) October temperature is determinant for the transition between the ablation and the accumulation season, favouring a lengthening of melting when temperature remains positive, or conversely allowing snowfalls that protect the ice and contribute to the accumulation when temperatures are negative; 2) March snowfall has a similar effect: positive anomalies contribute to the total accumulation at the glacier surface, and a thicker snow pack will delay the snow/ice transition during the ablation season leading to a less negative ablation rate (*e.g.*, Fig. 6b in Réveillet et al., 2018). Therefore, meteorological conditions of these transition months seem to strongly impact the annual glacier-wide SMB variability, since their variability oscillates between positive and negative values, unlike the months in the heart of summer or winter.
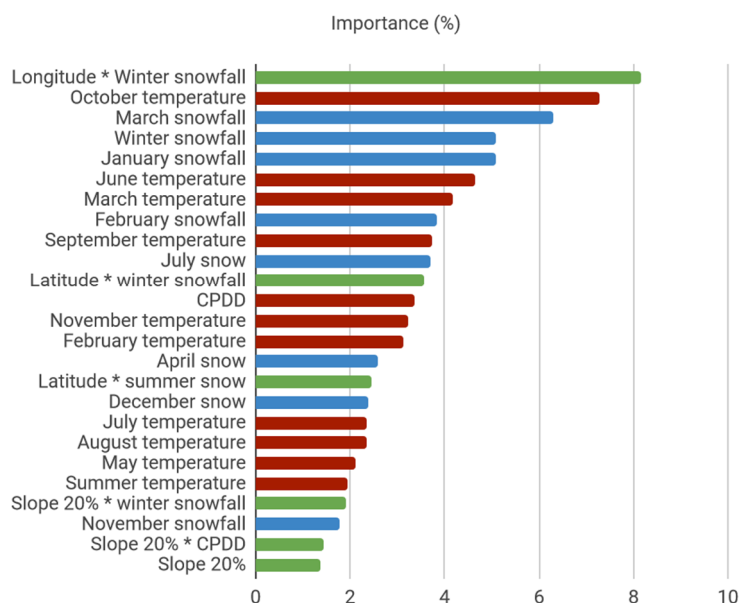
Figure 5: Contribution to the total variance of the 25 top topo-climatic predictors out of 50 predictors using Lasso. Green bars indicate predictors including topographical features, blue ones including accumulation-related features, and red ones including ablation-related features

3.2.3 Spatial predictive analysis

In order to evaluate the performance of the machine learning SMB models in space, we perform a leave-one-glacier-out (LOGO) cross-validation. For relatively small datasets like the one used in this study, cross-validation ensures that the model is validated on the full dataset. Such validation aims at understanding the model's performance for predictions on other glaciers for the same time period as during the training.

An important aspect is the comparison between linear and nonlinear machine learning algorithms used in this study. Steiner et al. (2005) already proved that a nonlinear ANN improved the results with respect a classic stepwise multiple linear regression. Here, we draw a similar comparison using more advanced methods for a larger dataset: OLS and Lasso as linear machine learning algorithms and a deep ANN as a nonlinear one. We observed significant differences between OLS, Lasso and deep learning, both in terms of explained variance (r^2) and accuracy (RMSE) of predicted glacier-wide SMBs. On average, we found improvements between +55% and +61% in the explained variance (from 0.49 to 0.76-0.79) using the nonlinear deep ANN compared to Lasso, whereas the accuracy was improved up to 45% (from 0.74 to 0.51-0.62). This means that 27% more variance is explained with a nonlinear model in the spatial dimension for glacier-wide SMB in this region. See Fig. 6 for a full summary of the results.

An interesting consequence of the nonlinearity of the ANN is the fact that it better captures extreme SMB values compared to a linear model. A linear model can correctly approximate the main cluster of values around the median, but the linear approximation performs poorly for extreme annual glacier-wide SMB values. The ANN solves this problem, with an increased explained variance which translates into a better accuracy for extreme SMB values, even without the use of sample weights
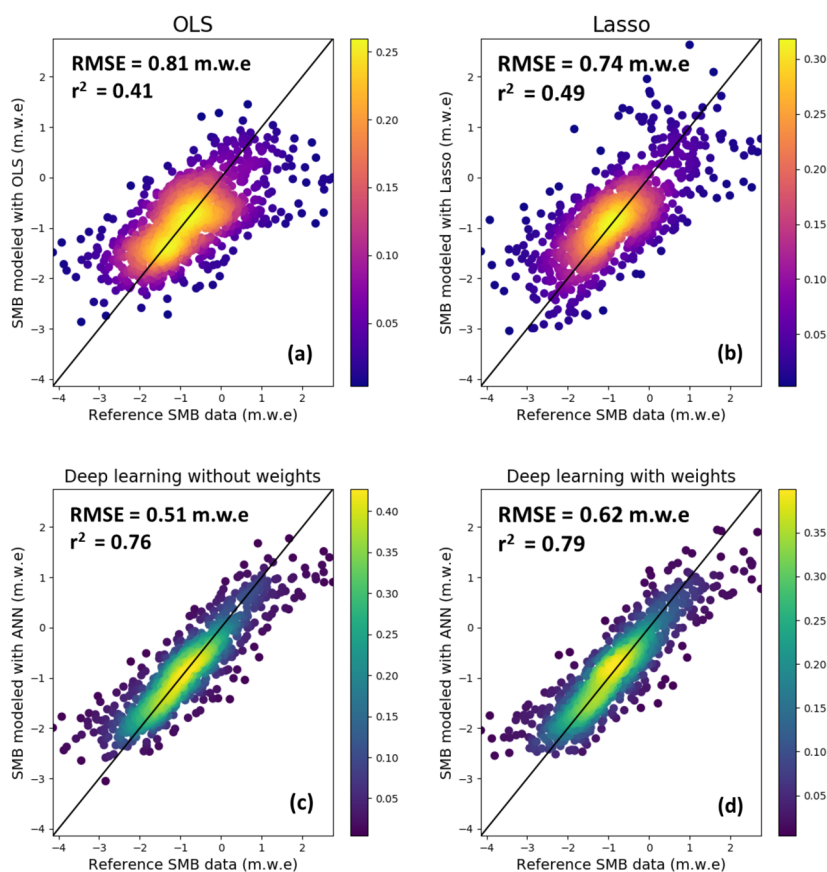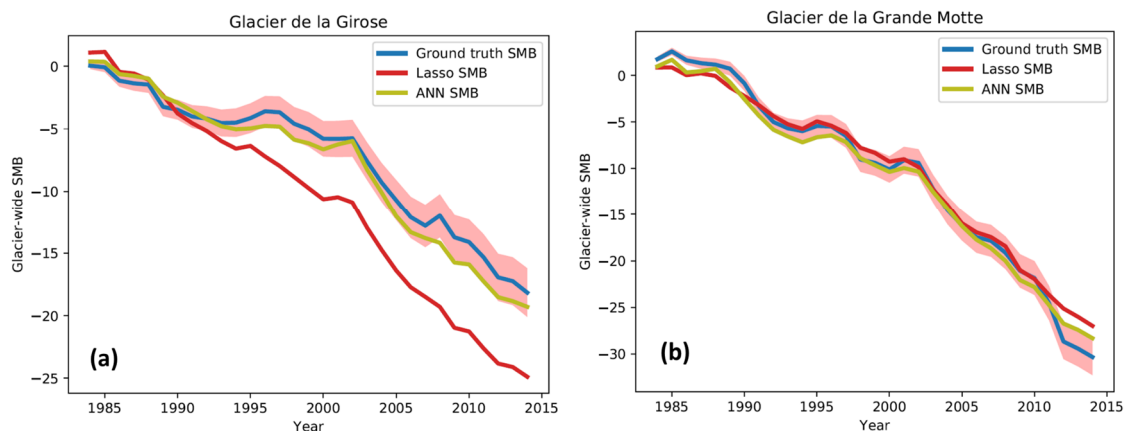
5 (Fig. 6).

Figure 6: Evaluation of modelled annual glacier-wide SMB against the ground truth SMB data (both in m.w.e. yr[-1]) using Leave-One-Glacier-Out cross-validation. The colour (purple-orange for linear; blue-green for nonlinear) indicates frequency based on the probability density function. The black line indicates the reference one-to-one line. **a)** Scatter plot of the OLS model results; **b)** Scatter plot of the Lasso

10 linear model results; Scatter plots of the deep artificial neural network nonlinear models without **(c)** and with sample weights **(d)**.

As a consequence, the added value of deep learning is especially relevant on glaciers with steeper annual changes in glacier-wide SMB (Fig. 7a). The use of sample weights can scale up or down this factor, thus playing with a performance trade-off depending on how much one wants to improve the model's behaviour for extreme SMB values.

Figure 7: Examples of cumulative glacier-wide SMB (m.w.e) simulations against the ground truth SMB data. The pink envelope indicates the accumulated uncertainties from the ground truth data. The deep learning SMB model has not been trained with sample weights in these illustrations.

Overall, deep learning results in a lower error throughout all the glaciers in the dataset when evaluated using LOGO cross-validation (Fig. 8). Moreover, the bias is also systematically reduced, but it is strongly correlated to the one from Lasso.
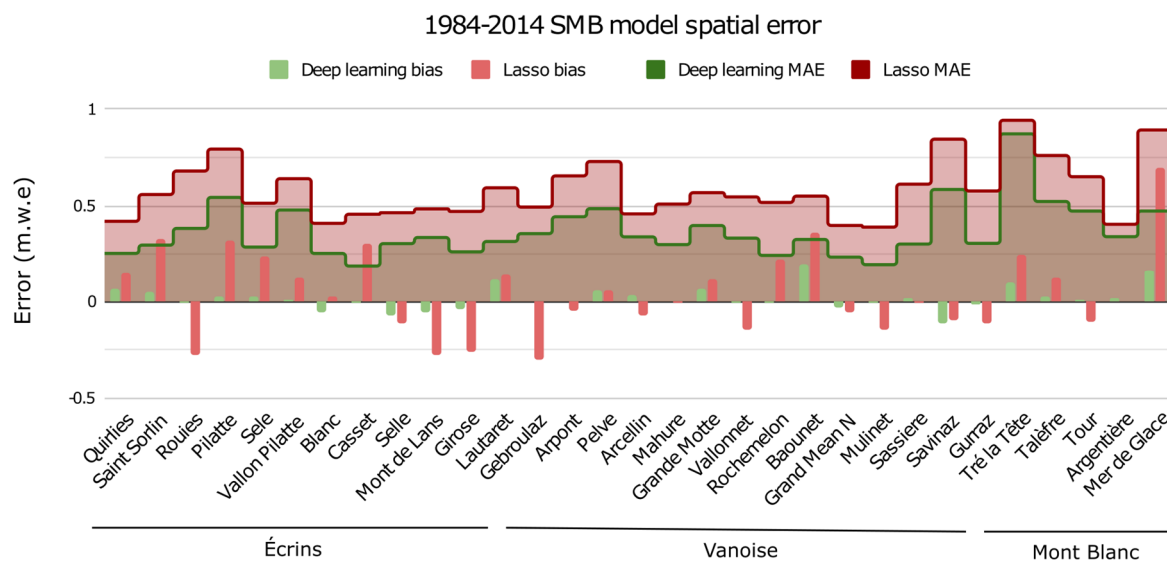
Figure 8: Mean average error (MAE) and bias (vertical bars) for each glacier of the training dataset structured by clusters for the 1984-2014 LOGO glacier-wide SMB simulation

3.2.4 Temporal predictive analysis

In order to evaluate the performance of the machine learning SMB models in time, we perform a leave-one-year-out (LOYO) cross-validation. This validation serves to understand the model's performance for past or future periods outside the training time period. The best results achieved for Lasso make no use of any monthly average temperature or snowfall, suggesting that these features are not relevant for temporal predictions unlike the spatial case.

As in Sect. 3.2.3, the results between the linear and nonlinear machine learning algorithms were compared. Interestingly, using LOYO, the differences between the different models were even greater than for spatial validation, revealing the more complex nature of the information in the temporal dimension. As illustrated by Fig. 9, we found remarkable improvements between the linear Lasso and the nonlinear deep learning in both the explained variance (between +94% and +108%) and accuracy (between +32% and +58%). This implies, that 35% more variance is explained using a nonlinear model in the temporal dimension for glacier-wide SMB balance in this region. Deep learning manages to keep very similar performances between the spatial and temporal dimensions, whereas the linear methods see their performance affected most likely due to the increased nonlinearity of the SMB reaction to meteorological conditions.
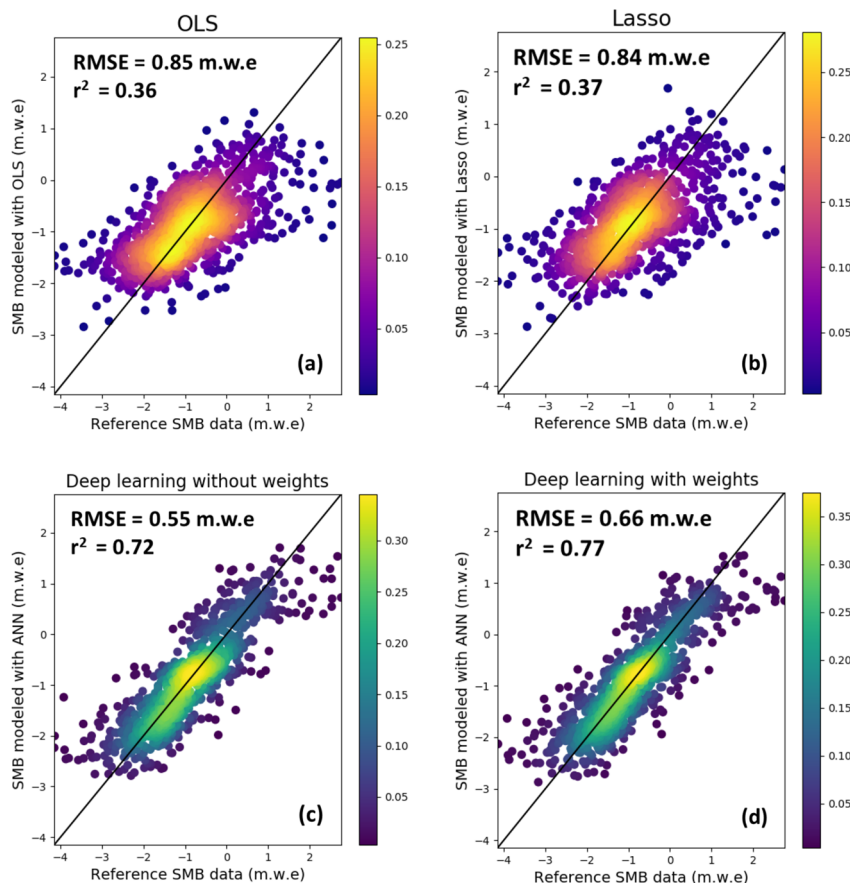
Figure 9: Evaluation of modelled annual glacier-wide SMB against the ground truth SMB data (both in m.w.e yr⁻¹) using Leave-One-Year-Out cross-validation. The colour (purple-orange for linear; blue-green for nonlinear) indicates frequency based on the probability density function. The black line indicates the reference one-to-one line. a) Scatter plot of the OLS model results; b) Scatter plot of the Lasso linear model results; Scatter plots of the deep artificial neural network nonlinear models without (c) and with sample weights (d).

A more detailed year by year analysis reveals interesting information about the glacier-wide SMB data structure. As seen in Fig. 10, the years with the worst deep learning precision are 1984, 1985 and 1990. All these three hydrological years present a high spatial variability in observed (or remotely-sensed) SMBs: very positive SMB values in general for 1984 and 1985 with few slightly negative values, and extremely negative SMB values in general for 1990 with few almost neutral values. These complex configurations are clearly outliers within the dataset, which push the limits of the nonlinear patterns found by the ANN. The situation becomes even more evident with Lasso, which struggles to resolve these complex patterns and often performs poorly where the ANN succeeds (*e.g.*, years 1996, 2012 or 2014). The important bias present only with Lasso is representative of its lack of complexity towards nonlinear structures: the average error is often good, but it is constantly biased unlike the ANN.
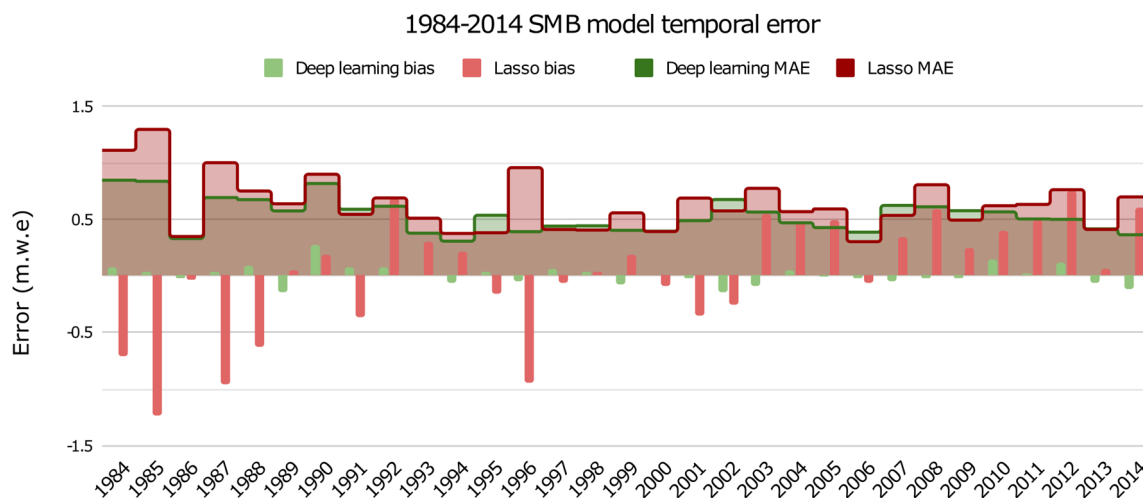
21

Figure 10: Mean average error (MAE) and bias (vertical bars) for each year of the training dataset for the 1984-2014 LOYO glacier-wide SMB simulation

3.2.5 Spatiotemporal predictive analysis

Once the specific performances in the spatial and temporal dimensions have been assessed, the performance in both dimensions at the same time is evaluated using Leave-Some-Glaciers-and-Years-Out (LSYGO) cross-validation. 64 folds were built, with test folds being comprised of data for 2 random glaciers on 2 random years, and train folds of all the data except the 2 years (for all glaciers) and the 2 glaciers (for all years) present in the test fold. These combinations are quite strict, implying that for every 4 tested values we need to drop between 123 and 126 values for training, depending on the glacier and year, to respect the spatiotemporal independence (Roberts et al., 2017).

The performance of LSYGO is similar to LOYO, with a RMSE of 0.51 m.w.e and a coefficient of determination of 0.77 (Fig. S5). This is reflected in the fact that very similar ANN hyperparameters were used for the training. This means that the deep learning SMB model is successful in generalizing and it does not overfit the training data.

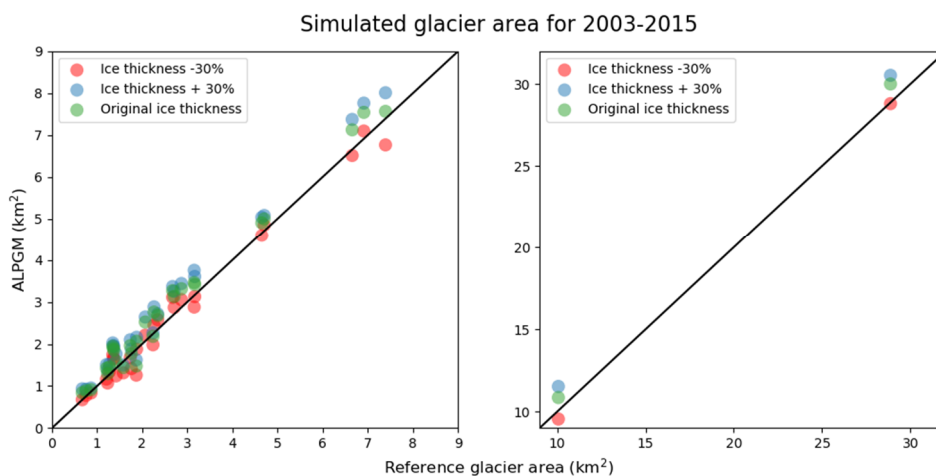### 3.3 Glacier geometry evolution: Validation and results

As mentioned in Sect. 2.3, the Δh parameterization has been widely used in many studies (*e.g.*, Huss et al., 2008, 2010; Vincent et al., 2014; Huss and Hock, 2015, 2018; Hanzer et al., 2018). It is not in the scope of this study to evaluate the performance of this method, but we present the approach developed in ALPGM to compute the Δh functions and show some examples for single glaciers to illustrate how these glacier-specific functions perform compared to observations. For the studied French alpine glaciers, the 1979-2011 period is used. This period was proved by Vincent et al. (2014) to be representative of Mer de

Glace's secular trend. Other sub-periods could have been used, but it was shown that they did not necessarily improve the performance. In addition, the 1979 and 2011 DEMs are the only ones available that cover all the French alpine glaciers. Within this period, some years with neutral to even positive surface mass balances in the late 1970s and early 1980s can be found, as well as a remarkable change from 2003 onward with strongly negative surface mass balances, following the heatwave that severely affected the western Alps in summer 2003.

The glacier-specific Δh functions are computed for glaciers ≥ 0.5 km², which represented about 80% of the whole glaciarized surface of the French Alps in 2015 (some examples are illustrated in the Supplement Fig. S4). For the rest of very small glaciers (< 0.5 km²), a standardized flat function is used in order to make them shrink equally at all altitudes. This is done to simulate the fact that generally, the equilibrium line of very small glaciers has surpassed the glacier's maximum altitude, thus shrinking from all directions and altitudes in summer. Moreover, due to their reduced size and altitudinal range, the ice flow no longer has the same importance as for larger or medium sized glaciers.

In order to evaluate the performance of the parameterized glacier dynamics of ALPGM, we compared the simulated glacier area of the 32 studied glaciers with the observed area in 2015 from the most up-to-date glacier inventory in the French Alps. Simulations were started in 2003, for which we used the F19 ice thickness dataset. In order to take into account the ice thickness uncertainties, we ran three simulations with different versions of the initial ice thickness: the original data, -30% and +30% of the original ice thickness in agreement with the uncertainty estimated by the authors (Fig. 11).

Figure 11: Simulated glacier areas for the 2003-2015 period for the 32 study glaciers using a deep learning SMB model without weights. Colours indicate the different F19 initial ice thicknesses used taking into account their uncertainties. For better visualisation, the figure is split in two with the two largest French glaciers on the right.

Overall, the results illustrated in Fig. 11 show a good agreement with the observations. Even for a 12-year period, the initial ice thickness remains the largest uncertainty, with almost all glaciers falling within the observed area when taking it into account. The mean error in simulated surface area was of 12% with the original F19 ice thickness dataset. Other studies using the Δh parameterization already proved that the initial ice thickness is the most important uncertainty in glacier evolution simulations, together with the choice of a GCM for future projections (Huss and Hock, 2015).

4 Discussion and perspectives

4.1 Linear methods still matter

Despite the fact that deep learning often outperforms linear machine learning and statistical methods, there is still a place for such methods in modelling. Indeed, unlike ANNs, simpler regularised linear models such as Lasso allow an easy interpretation of the coefficients associated to each input feature, which helps to understand the contribution of each of the chosen variables to the model. This means that linear machine learning methods can be used for both prediction and causal analysis. Training a linear model in parallel to an ANN has therefore the advantage to provide a simpler linear alternative which can be used to understand the dataset. Moreover, seeing the contribution of each coefficient, one can reduce the complexity of the dataset by keeping only the most significant predictors. Finally, a linear model serves as well as a reference to highlight and quantify the nonlinear gains obtained by deep learning.

4.2 Training deep learning models with spatiotemporal data

The creation and training of a deep ANN requires a certain knowledge and strategy with respect to the data and study focus. When working with spatiotemporal data, the separation between training and validation becomes tricky. The spatial and temporal dimensions in the dataset cannot be ignored, and strongly affect the independence between training and validation data (Roberts et al., 2017; Oliveira et al., 2019). Depending on how the cross-validation is performed, the obtained performance will be indicative of one of these two dimensions. As it is shown in Sect. 3.2.3, the ANNs and especially the linear modelling approaches had more success in predicting SMB values in space than in time. This is mostly due to the fact that the glacier-wide SMB signal has a greater variability and nonlinearities in time than in space, with climate being the main driver of the annual fluctuations in SMB, whereas geography, and in particular the local topography, modulates the signal between glaciers (Huss, 2012; Rabatel et al., 2016; Vincent et al., 2017). Consequently, linear models find it easier to make predictions on a given period of time for other glaciers elsewhere in space, than for time periods outside the training. Nonetheless, the deep learning SMB models were capable of equally capturing the complex nonlinear patterns in both the spatial and temporal dimensions.

In order to cope with the specific challenges related to each type of cross-validation, there are several parameters that can be modified to adapt the ANN's behaviour. The ANN architecture plays an important role: the number of neurons as well as the

number of hidden layers will determine the ANN's complexity and its capabilities to capture hidden patterns in the data. But the larger the architecture, the higher are the chances to overfit the data. This undesired effect can be counterbalanced using regularization. The amount of regularization (dropout and Gaussian noise in our case, see Sect. 2.2.4) used in the training of the ANN necessarily introduces some trade-offs. The greater the dropout, the more we will constrain the learning of the ANN

5   so the higher the generalization will be, until a certain point. On the other hand, the learning rate to compute the stochastic gradient descent, which tries to minimize the loss function, also plays an important role: smaller learning rates result in a slower convergence towards the absolute minima, thus producing models with better generalization. By balancing all these different effects, one can achieve the accuracy *versus* generalization ratio that best suits a certain dataset and model. Nonetheless, one key aspect in machine learning models is data: expanding the training dataset in the future will allow to

10 increase the complexity of the model and its performance. Consequently, machine learning models see their performance improved as time goes by, with new data becoming available for training.

4.3 Perspectives on future applications of deep learning in glaciology

The currently used meteorological variables in the deep ANN of ALPGM's SMB component are based on the classic degree-day approach, which relies only on temperature and precipitation. However, the model could be trained with variables involved

15 in more complex models, such as SEB-type models, for which the longwave and shortwave radiation, as well as the turbulent fluxes and albedo intervene. The current model framework allows flexibility in the choice and number of input variables that can reflect different degrees of complexity for the resolved processes. Despite it has been shown that for glaciers in the European Alps there is almost no added value in transitioning from a simple degree-day to a SEB model for annual glacier-wide SMB simulations (*e.g.*, Réveillet et al., 2017), it could be an interesting way to expand the training dataset for glaciers in

20 tropical and subtropical regions, where shortwave radiation plays a much more important role (Benn and Evans, 2014). Maussion et al. (2015) followed a similar approach with linear machine learning in order to calibrate a regression-based downscaling model that linked local SEB/SMB fluxes to atmospheric reanalysis variables. For ALPGM, the ANN was trained with glacier-wide SMB data, but the same approach could be used for point SMB data from field observations.

25 In this work, we also evaluated the resilience of the deep learning approach: since many glaciarized regions in the world do not have the same amount of data used in this study, we trained an ANN only with monthly average temperature and snowfall, without any topographical predictors, to see until which point the algorithm is capable of learning from minimal data. The results were quite astonishing, with a coefficient of determination of 0.68 (against 0.76 from the full model) and a RMSE of 0.59 (against 0.51 from the full model). These results indicate that meteorological data is the primary source of information

30 regarding the annual glacier-wide SMB modelling, and the "bonus" of topographical data helps to better differentiate glaciers whose specificities do not stand out by just using climate data. Consequently, the potential of deep learning in glacier modelling should not be ignored. A nonlinear deep learning SMB component like the one used for ALPGM could provide an interesting alternative to classical SMB models used for regional modelling. The comparison with other SMB models is beyond the scope

of this study, but it would be worth investigating to quantify the specific gains that could be achieved by switching to a deep learning modelling approach. Nonetheless, the linear machine learning models trained with the CPDD and cumulative snowfall used in this study behave in a similar way to a calibrated temperature-index model.

5　Deep learning can be of special interest once applied in the reconstruction of SMB time series. More and more SMB data is becoming available thanks to the advances in remote sensing (*e.g.,* Brun et al., 2017; Rabatel et al., 2017; Zemp et al., 2019), but these datasets often cover limited areas and the most recent time period in the studied regions. An interesting way of expanding a dataset would be to use a deep learning approach to fill the data gaps. SMB past time series of vast glaciarized regions could thereby be obtained with unprecedented efficiency. Such an approach would be an excellent way of obtaining

10　more SMB data in remote glaciarized regions such as the Andes or the Himalayas. It could also be applied in data-rich regions benefiting from regionalized climate reanalyses (*e.g.*, Caillouet et al., 2016, covering the 1871-present period for France) Another possibility would be to completely bypass both the SMB and glacier dynamics of a classic glacier evolution model by training a deep ANN which would directly simulate changes in glacier thicknesses. If the ANN is trained with enough glacier thickness changes, climatic and topographical data, it could be able to simulate the 3D evolution of the glacier straight

15　from the raw data. It might still be too soon for such models to be implemented, but once enough data will be available in the future, this could be a promising new way of tackling glacier evolution modelling.

Finally, yet another research perspective for deep learning and ANNs in glaciology is the estimation of glacier ice thickness. Clarke et al. (2009) already took the first steps towards this direction, but as in Steiner et al. (2005), they used a rather simple

20　ANN, with sigmoid and hyperbolic tangent activation functions and just two hidden layers. Glacier ice thickness estimation is probably a much more complex problem than SMB simulation, so a more complex deep ANN fed by the Glacier Thickness Database (GlaThiDa Consortium, 2019) could provide an interesting alternative to physical approaches.

5 Conclusions

We presented ALPGM (Bolibar, 2019), a parameterized glacier evolution model, following an alternative approach to most

25　physical and process-based glacier models. Annual glacier-wide SMBs are simulated using nonlinear deep learning, and the glacier geometry is updated annually using a glacier-specific parameterization. Deep learning is shown to outperform linear methods for the simulation of glacier-wide SMB with a case study of French alpine glaciers. By means of cross-validation, we demonstrated how important nonlinear structures (up to 35%) coming from the glacier and climate systems in both the spatial and temporal dimensions are captured by the deep ANN. Taking into account this nonlinearity substantially improved the

30　explained variance and accuracy compared to linear statistical models, especially in the more complex temporal dimension. As we have shown in our case study, deep ANNs are capable of dealing with relatively small datasets, and they present a wide

range of configurations to generalize and prevent overfitting. Machine learning models benefit from the increasing number of available data, which makes their performance constantly improve as time goes by.

Deep learning should be seen as an opportunity by the glaciology community. Its good performance in both the spatial and

5   temporal dimensions shows how relevant it can be for a broad range of applications. Combined with *in situ* or remote sensing SMB estimations, it can serve to extend SMB time series for regions or glaciers with already available data for past and future periods; or it could help extend SMB datasets to other unmeasured glaciers for remote regions with only few monitored glaciers such as the Andes or the high mountains of Asia. Moreover, deep learning could be used as an alternative to classical SMB models; important nonlinearities from the glacier and climate systems are potentially ignored by these mostly linear models,

10   which could give an advantage to deep learning models in regional studies. It might still be too early for the development of such models in certain regions, but as new data becomes available the gap is slowly being closed towards real big data approaches in glaciology.

Code availability

The source code of ALPGM is available at https://github.com/JordiBolibar/ALPGM with its DOI (10.5281/zenodo.3269678)

15   for the v1.0 release. All scripts used to generate plots and results are included in the repository. The detailed information regarding glaciers used in the case study and the SMB model performance per glacier are included in the Supplementary material.

Author contribution

J. Bolibar developed ALPGM, analysed the results and wrote the paper. A. Rabatel did the glaciological analysis, contributed

20   with the remote sensing SMB data, and conceived the study together with I. Gouttevin, T. Condom and J. Bolibar. C. Galiez and J. Bolibar developed the deep and machine learning modelling approach, and E. Sauquet contributed in the statistical analysis. All authors discussed the results and helped to develop the paper.

Competing interests

The authors declare that they have no conflict of interest.

25

Acknowledgements

References

Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., Fantini, A., Giacona, F., Hauck, C., Huss, M., Huwald, H., Lehning, M., López-Moreno, J.-I., Magnusson, J., Marty, C., Morán-Tejéda, E., Morin, S., Naaim, M., Provenzale, A., Rabatel, A., Six, D., Stötter, J., Strasser, U., Terzago, S. and Vincent, C.: The European mountain cryosphere: a review of its current state, trends, and future challenges, The Cryosphere, 12(2), 759–794, doi:10.5194/tc-12-759-2018, 2018.

Benn, D. I. and Evans, D. J. A.: Glaciers & glaciation, 2nd ed., Routledge, New York, NY, USA. [online] Available from: http://www.imperial.eblib.com/EBLWeb/patron/?target=patron&extendedid=P_615876_0 (Accessed 4 April 2019), 2014.

Bolibar, J.: JordiBolibar/ALPGM: ALPGM v1.0, , doi:10.5281/zenodo.3269678, 2019.

Brun, F., Berthier, E., Wagnon, P., Kääb, A. and Treichler, D.: A spatially resolved estimate of High Mountain Asia glacier mass balances from 2000 to 2016, Nat. Geosci., 10(9), 668–673, doi:10.1038/ngeo2999, 2017.

Caillouet, L., Vidal, J.-P., Sauquet, E. and Graff, B.: Probabilistic precipitation and temperature downscaling of the Twentieth Century Reanalysis over France, Clim. Past, 12(3), 635–662, doi:10.5194/cp-12-635-2016, 2016.

Carlson, B. Z., Georges, D., Rabatel, A., Randin, C. F., Renaud, J., Delestrade, A., Zimmermann, N. E., Choler, P. and Thuiller, W.: Accounting for tree line shift, glacier retreat and primary succession in mountain plant distribution models, edited by M. Rouget, Divers. Distrib., 20(12), 1379–1391, doi:10.1111/ddi.12238, 2014.

Chollet, F.: Keras. [online] Available from: https://keras.io, 2015.

Clarke, G. K. C., Berthier, E., Schoof, C. G. and Jarosch, A. H.: Neural Networks Applied to Estimating Subglacial Topography and Glacier Volume, J. Clim., 22(8), 2146–2160, doi:10.1175/2008JCLI2572.1, 2009.

Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B. and Mérindol, L.: Reanalysis of 44 Yr of Climate in the French Alps (1958–2002): Methodology, Model Validation, Climatology, and Trends for Air Temperature and Precipitation, J. Appl. Meteorol. Climatol., 48(3), 429–449, doi:10.1175/2008JAMC1808.1, 2009.

Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F. and Pandit, A.: A consensus estimate for the ice thickness distribution of all glaciers on Earth, Nat. Geosci., 12(3), 168–173, doi:10.1038/s41561-019-0300-3, 2019.

Fausett, L. V.: Fundamentals of neural networks: architectures, algorithms, and applications, Prentice Hall, Englewood Cliffs, N.J., 1994.

Gagliardini, O., Zwinger, T., Gillet-Chaulet, F., Durand, G., Favier, L., de Fleurian, B., Greve, R., Malinen, M., Martín, C., Råback, P., Ruokolainen, J., Sacchettini, M., Schäfer, M., Seddik, H. and Thies, J.: Capabilities and performance of Elmer/Ice, a new-generation ice sheet model, Geosci. Model Dev., 6(4), 1299–1318, doi:10.5194/gmd-6-1299-2013, 2013.

Gardent, M., Rabatel, A., Dedieu, J.-P. and Deline, P.: Multitemporal glacier inventory of the French Alps from the late 1960s to the late 2000s, Glob. Planet. Change, 120, 24–37, doi:10.1016/j.gloplacha.2014.05.004, 2014.

GlaThiDa Consortium: Glacier Thickness Database 3.0.1, 2019.

GLIMS and NSIDC: Global Land Ice Measurements from Space glacier database. Compiled and made available by the international GLIMS community and the National Snow and Ice Data Center, 2005.

Hanzer, F., Förster, K., Nemec, J. and Strasser, U.: Projected cryospheric and hydrological impacts of 21st century climate change in the Ötztal Alps (Austria) simulated using a physically based approach, Hydrol. Earth Syst. Sci., 22(2), 1593–1614, doi:10.5194/hess-22-1593-2018, 2018.

Hastie, T., Tibshirani, R. and Friedman, J.: The Elements of Statistical Learning, Springer New York, New York, NY., 2009.

Hawkins, D. M.: The Problem of Overfitting, J. Chem. Inf. Comput. Sci., 44(1), 1–12, doi:10.1021/ci0342472, 2004.

He, K., Zhang, X., Ren, S. and Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, 2015 IEEE Int. Conf. Comput. Vis. ICCV, doi:10.1109/iccv.2015.123, 2015.

Hock, R., Bliss, A., Marzeion, B., Giesen, R. H., Hirabayashi, Y., Huss, M., Radić, V. and Slangen, A. B. A.: GlacierMIP – A model intercomparison of global-scale glacier mass-balance models and projections, J. Glaciol., 65(251), 453–467, doi:10.1017/jog.2019.22, 2019.

Hoerl, A. E. and Kennard, R. W.: Ridge Regression: Biased Estimation for Nonorthogonal Problems, Technometrics, 12(1), 55–67, doi:10.1080/00401706.1970.10488634, 1970.

Hoinkes, H. C.: Glacier Variation and Weather, J. Glaciol., 7(49), 3–18, doi:10.3189/S0022143000020384, 1968.

Houghton, J. T., Ding, Y., Griggs, D. J., Noguer, M., van der Linden, P. J., Dai, X., Maskell, K. and Johnson, C.: Climate change 2001: the scientific basis, The Press Syndicate of the University of Cambridge., 2001.

Huss, M.: Extrapolating glacier mass balance to the mountain-range scale: the European Alps 1900–2100, The Cryosphere, 6(4), 713–727, doi:10.5194/tc-6-713-2012, 2012.

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Front. Earth Sci., 3, doi:10.3389/feart.2015.00054, 2015.

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

Huss, M., Farinotti, D., Bauder, A. and Funk, M.: Modelling runoff from highly glacierized alpine drainage basins in a changing climate, Hydrol. Process., 22(19), 3888–3902, doi:10.1002/hyp.7055, 2008.

Ingrassia, S. and Morlini, I.: Neural Network Modeling for Small Datasets, Technometrics, 47(3), 297–311, doi:10.1198/004017005000000058, 2005.

Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift., 2015.

Jóhannesson, T., Raymond, C. and Waddington, E.: Time–Scale for Adjustment of Glaciers to Changes in Mass Balance, J. Glaciol., 35(121), 355–369, doi:10.3189/S002214300000928X, 1989.

Jouvet, G., Huss, M., Blatter, H., Picasso, M. and Rappaz, J.: Numerical simulation of Rhonegletscher from 1874 to 2100, J. Comput. Phys., 228(17), 6426–6439, doi:10.1016/j.jcp.2009.05.033, 2009.

Martin, S.: Correlation bilans de masse annuels-facteurs météorologiques dans les Grandes Rousses, Z. Für Gletscherkunde Glazialgeol., 1974.

Marzeion, B., Jarosch, A. H. and Hofer, M.: Past and future sea-level change from the surface mass balance of glaciers, The Cryosphere, 6(6), 1295–1322, doi:10.5194/tc-6-1295-2012, 2012.

Maussion, F., Gurgiser, W., Großhauser, M., Kaser, G. and Marzeion, B.: ENSO influence on surface energy and mass balance at Shallap Glacier, Cordillera Blanca, Peru, The Cryosphere, 9(4), 1663–1683, doi:10.5194/tc-9-1663-2015, 2015.

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., Gregor, P., Jarosch, A. H., Landmann, J., Oesterle, F., Recinos, B., Rothenpieler, T., Vlug, A., Wild, C. T. and Marzeion, B.: The Open Global Glacier Model (OGGM) v1.1, Geosci. Model Dev., 12(3), 909–931, doi:10.5194/gmd-12-909-2019, 2019.

Nussbaumer, S., Steiner, D. and Zumbühl, H.: Réseau neuronal et fluctuations des glaciers dans les Alpes occidentales., 2012.

5 Oliveira, M., Torgo, L. and Santos Costa, V.: Evaluation Procedures for Forecasting with Spatio-Temporal Data, in Machine Learning and Knowledge Discovery in Databases, vol. 11051, edited by M. Berlingerio, F. Bonchi, T. Gärtner, N. Hurley, and G. Ifrim, pp. 703–718, Springer International Publishing, Cham., 2019.

Olson, M., Wyner, A. J. and Berk, R.: Modern Neural Networks Generalize on Small Data Sets, in NeurIPS., 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., 10 Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. and Louppe, G.: Scikit-learn: Machine Learning in Python., 2012.

Rabatel, A., Letréguilly, A., Dedieu, J.-P. and Eckert, N.: Changes in glacier equilibrium-line altitude in the western Alps from 1984 to 2010: evaluation by remote sensing and modeling of the morpho-topographic and climate controls, The Cryosphere, 7(5), 1455–1471, doi:10.5194/tc-7-1455-2013, 2013.

15 Rabatel, A., Dedieu, J. P. and Vincent, C.: Spatio-temporal changes in glacier-wide mass balance quantified by optical remote sensing on 30 glaciers in the French Alps for the period 1983–2014, J. Glaciol., 62(236), 1153–1166, doi:10.1017/jog.2016.113, 2016.

Rabatel, A., Sirguey, P., Drolon, V., Maisongrande, P., Arnaud, Y., Berthier, E., Davaze, L., Dedieu, J.-P. and Dumont, M.: Annual and Seasonal Glacier-Wide Surface Mass Balance Quantified from Changes in Glacier Surface State: A Review on 20 Existing Methods Using Optical Satellite Imagery, Remote Sens., 9(5), 507, doi:10.3390/rs9050507, 2017.

Rabatel, A., Sanchez, O., Vincent, C. and Six, D.: Estimation of Glacier Thickness From Surface Mass Balance and Ice Flow Velocities: A Case Study on Argentière Glacier, France, Front. Earth Sci., 6, doi:10.3389/feart.2018.00112, 2018.

Radić, V., Bliss, A., Beedlow, A. C., Hock, R., Miles, E. and Cogley, J. G.: Regional and global projections of twenty-first century glacier mass changes in response to climate scenarios from global climate models, Clim. Dyn., 42(1–2), 37–58, 25 doi:10.1007/s00382-013-1719-7, 2014.

Réveillet, M., Rabatel, A., Gillet-Chaulet, F. and Soruco, A.: Simulations of changes to Glaciar Zongo, Bolivia (16° S), over the 21st century using a 3-D full-Stokes model and CMIP5 climate projections, Ann. Glaciol., 56(70), 89–97, doi:10.3189/2015AoG70A113, 2015.

Réveillet, M., Vincent, C., Six, D. and Rabatel, A.: Which empirical model is best suited to simulate glacier mass balances?, 30 J. Glaciol., 63(237), 39–54, doi:10.1017/jog.2016.110, 2017.

Réveillet, M., Six, D., Vincent, C., Rabatel, A., Dumont, M., Lafaysse, M., Morin, S., Vionnet, V. and Litt, M.: Relative performance of empirical and physical models in assessing the seasonal and annual glacier surface mass balance of Saint-Sorlin Glacier (French Alps), The Cryosphere, 12(4), 1367–1386, doi:10.5194/tc-12-1367-2018, 2018.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, 35 B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F. and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, 40(8), 913–929, doi:10.1111/ecog.02881, 2017.

Seabold, S. and Perktold, J.: Statsmodels: Econometric and Statistical Modelingwith Python, PROC 9th PYTHON Sci. CONF, 2010.

Six, D. and Vincent, C.: Sensitivity of mass balance and equilibrium-line altitude to climate change in the French Alps, J. 40 Glaciol., 60(223), 867–878, doi:10.3189/2014JoG14J014, 2014.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting, J Mach Learn Res, 15(1), 1929–1958, 2014.

Steiner, D., Walter, A. and Zumbühl, H. J.: The application of a non-linear back-propagation neural network to study the mass balance of Grosse Aletschgletscher, Switzerland, J. Glaciol., 51(173), 313–323, doi:10.3189/172756505781829421, 2005.

Steiner, D., Pauling, A., Nussbaumer, S. U., Nesje, A., Luterbacher, J., Wanner, H. and Zumbühl, H. J.: Sensitivity of European glaciers to precipitation and temperature – two case studies, Clim. Change, 90(4), 413–441, doi:10.1007/s10584-008-9393-1, 2008.

Thibert, E., Dkengne Sielenou, P., Vionnet, V., Eckert, N. and Vincent, C.: Causes of Glacier Melt Extremes in the Alps Since 1949, Geophys. Res. Lett., 45(2), 817–825, doi:10.1002/2017GL076333, 2018.

Tibshirani, R.: Regression Shrinkage and Selection via the Lasso, J. R. Stat. Soc. Ser. B Methodol., 58(1), 267–288, 1996.

Tibshirani, R., Johnstone, I., Hastie, T. and Efron, B.: Least angle regression, Ann. Stat., 32(2), 407–499, doi:10.1214/009053604000000067, 2004.

Vincent, C., Harter, M., Gilbert, A., Berthier, E. and Six, D.: Future fluctuations of Mer de Glace, French Alps, assessed using a parameterized model calibrated with past thickness changes, Ann. Glaciol., 55(66), 15–24, doi:10.3189/2014AoG66A050, 2014.

Vincent, C., Fischer, A., Mayer, C., Bauder, A., Galos, S. P., Funk, M., Thibert, E., Six, D., Braun, L. and Huss, M.: Common climatic signal from glaciers in the European Alps over the last 50 years: Common Climatic Signal in the Alps, Geophys. Res. Lett., 44(3), 1376–1383, doi:10.1002/2016GL072094, 2017.

Vuille, M., Carey, M., Huggel, C., Buytaert, W., Rabatel, A., Jacobsen, D., Soruco, A., Villacis, M., Yarleque, C., Elison Timm, O., Condom, T., Salzmann, N. and Sicart, J.-E.: Rapid decline of snow and ice in the tropical Andes – Impacts, uncertainties and challenges ahead, Earth-Sci. Rev., 176, 195–213, doi:10.1016/j.earscirev.2017.09.019, 2018.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B. and Freckleton, R. P.: Why do we still use stepwise modelling in ecology and behaviour?: Stepwise modelling in ecology and behaviour, J. Anim. Ecol., 75(5), 1182–1189, doi:10.1111/j.1365-2656.2006.01141.x, 2006.

Xu, B., Wang, N., Chen, T. and Li, M.: Empirical Evaluation of Rectified Activations in Convolutional Network, CoRR, abs/1505.00853 [online] Available from: http://arxiv.org/abs/1505.00853, 2015.

Zekollari, H. and Huybrechts, P.: Statistical modelling of the surface mass-balance variability of the Morteratsch glacier, Switzerland: strong control of early melting season meteorological conditions, J. Glaciol., 64(244), 275–288, doi:10.1017/jog.2018.18, 2018.

Zekollari, H., Huss, M. and Farinotti, D.: Modelling the future evolution of glaciers in the European Alps under the EURO-CORDEX RCM ensemble, The Cryosphere, 13(4), 1125–1146, doi:10.5194/tc-13-1125-2019, 2019.

Zemp, M., Haeberli, W., Hoelzle, M. and Paul, F.: Alpine glaciers to disappear within decades?, Geophys. Res. Lett., 33(13), doi:10.1029/2006GL026319, 2006.

Zemp, M., Huss, M., Thibert, E., Eckert, N., McNabb, R., Huber, J., Barandun, M., Machguth, H., Nussbaumer, S. U., Gärtner-Roer, I., Thomson, L., Paul, F., Maussion, F., Kutuzov, S. and Cogley, J. G.: Global glacier mass changes and their contributions to sea-level rise from 1961 to 2016, Nature, 568(7752), 382–386, doi:10.1038/s41586-019-1071-0, 2019.