Emulating the Adaptation of Wind Fields to Complex Terrain with Deep Learning

LOUIS LE TOUMELIN,[a] ISABELLE GOUTTEVIN,[a] NORA HELBIG,[b,c] CLOVIS GALIEZ,[d] MATHIS ROUX,[d] AND FATIMA KARBOU[a]

[a] *Université Grenoble Alpes, Université de Toulouse, Météo-France, CNRS, CNRM, Centre d'Études de la Neige, Grenoble, France*
[b] *WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland*
[c] *Eastern Switzerland University of Applied Sciences, Rapperswil, Switzerland*
[d] *Université Grenoble Alpes, CNRS, Grenoble Institute of Engineering, Jean Kuntzmann Laboratory, Grenoble, France*

(Manuscript received 19 May 2022, in final form 14 October 2022)

ABSTRACT: Estimating the impact of wind-driven snow transport requires modeling wind fields with a lower grid spacing than the spacing on the order of 1 or a few kilometers used in the current numerical weather prediction (NWP) systems. In this context, we introduce a new strategy to downscale wind fields from NWP systems to decametric scales, using high-resolution (30 m) topographic information. Our method (named "DEVINE") is leveraged on a convolutional neural network (CNN), trained to replicate the behavior of the complex atmospheric model ARPS, and was previously run on a large number (7279) of synthetic Gaussian topographies under controlled weather conditions. A 10-fold cross validation reveals that our CNN is able to accurately emulate the behavior of ARPS (mean absolute error for wind speed = 0.16 m s^{-1}). We then apply DEVINE to real cases in the Alps, that is, downscaling wind fields forecast by the AROME NWP system using information from real alpine topographies. DEVINE proved able to reproduce main features of wind fields in complex terrain (acceleration on ridges, leeward deceleration, and deviations around obstacles). Furthermore, an evaluation on quality-checked observations acquired at 61 sites in the French Alps reveals improved behavior of the downscaled winds (AROME wind speed mean bias is reduced by 27% with DEVINE), especially at the most elevated and exposed stations. Wind direction is, however, only slightly modified. Hence, despite some current limitations inherited from the ARPS simulations setup, DEVINE appears to be an efficient downscaling tool whose minimalist architecture, low input data requirements (NWP wind fields and high-resolution topography), and competitive computing times may be attractive for operational applications.

SIGNIFICANCE STATEMENT: Wind largely influences the spatial distribution of snow in mountains, with direct consequences on hydrology and avalanche hazard. Most operational models predicting wind in complex terrain use a grid spacing on the order of several kilometers, too coarse to represent the real patterns of mountain winds. We introduce a novel method based on deep learning to increase this spatial resolution while maintaining acceptable computational costs. Our method mimics the behavior of a complex model that is able to represent part of the complexity of mountain winds by using topographic information only. We compared our results with observations collected in complex terrain and showed that our model improves the representation of winds, notably at the most elevated and exposed observation stations.

KEYWORDS: Snow; Wind; Artificial intelligence; Data science; Deep learning; Machine learning

1. Introduction

The transport of snow particles by the wind, hereinafter referred to as drifting snow, is a key process for understanding the spatial distribution of mountain snowpacks (Mott et al. 2018). Drifting snow redistributes both falling hydrometeors before they reach the surface and snow originating from the surface through mechanisms of ablation and deposition. As the mountain snowpack acts as a major freshwater reservoir during winter and spring in continental areas, its spatial distribution prior to and during the melting periods is of high importance for human activities, with consequences in terms of flood hazard, hydropower management, and more generally water resource management (Lehning 2013; Jörg-Hess et al. 2015; Vionnet et al. 2020). At the scale of a mountain slope, drifting snow is also influencing the evolution of avalanche hazard (Schweizer et al. 2003; Lehning et al. 2000), thus impacting the safety of infrastructures and people.

In addition to its influence on snow preferential deposition, wind fields are the major driving factor of snow erosion over snow-covered areas (Xie et al. 2021). Topography has a strong influence on wind fields, first influencing the motion of large-scale air masses (Wanner and Furger 1990), and second introducing a strong spatial variability in wind fields at a very local scale (Lewis et al. 2008; Sharples et al. 2010; Butler et al. 2015). Dynamic modifications of incoming flows, often referred to as terrain-forced flows (Whiteman 2000), occur when air masses interact locally with topography. The most noticeable features of terrain-forced flows are speedup on mountain crests, accelerations across gaps and passes, or

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/AIES-D-22-0034.s1.

Corresponding author: Louis Le Toumelin, louis.letoumelin@gmail.com

DOI: 10.1175/AIES-D-22-0034.1 e220034

changes in wind direction with channeling in topographic depressions and around obstacles (Whiteman 2000).

Mountain wind fields commonly present a high variability at a local scale (<100 m) (Mott et al. 2018), intrinsically limiting the benefit of numerical weather prediction systems (NWP), which generally operate with a horizontal grid spacing above 1 km (e.g., Baldauf et al. 2011; Seity et al. 2011; Kehler et al. 2016; Pickering et al. 2020). Furthermore, basic interpolation methods (i.e., linear, polynomial), when applied to synoptic winds provided by NWP models, do not accurately represent the complex reality of mountain winds (Wagenbrenner et al. 2016). Consequently, dynamic downscaling methods are often used to infer the behavior of mountain winds at a local scale (Raderschall et al. 2008; Mott and Lehning 2010; Vionnet et al. 2014). These methods rely on complex atmospheric models to handle coarse-scale signal, and to generate high-resolution wind fields. They solve equations of state describing the flow, including its interaction with the terrain and more generally the representation of various physical processes that most directly determine the large spatial variability of wind fields in complex terrain (Mott and Lehning 2010). The counterpart of this complexity lies in large computing requirements, hence restricting the use of the method to small domains and/or limited time scales (Mott and Lehning 2010; Vionnet et al. 2014). Therefore, models with relatively low complexity have been developed and provide a good trade-off in terms of physical complexity versus numerical costs (Forthofer et al. 2014; Vionnet et al. 2021). Statistical parameterizations using topographic information have also been largely used to bridge the gap between coarse-scale-resolution wind fields provided by NWP models and the high-resolution forcings required by small-scale applications and, notably, drifting snow models (Liston and Elder 2006; Helbig et al. 2017; Winstral et al. 2017). Such downscaling methods identify parameters expected to capture the effect of topography on the wind fields and then apply statistical operations to transform the coarse-scale signal into a distributed signal at a higher target resolution. The choice of accurate parameters relies both on the identification of dominant physical processes at a local scale (e.g., sheltering, exposition, channeling) and their formulation through a mathematical expression (using, e.g., curvature, slope, and Laplacian). For example, the MicroMet model (Liston and Elder 2006) identified slope, terrain slope azimuth, and curvature as relevant parameters to account for the effect of local topography on wind fields, whereas Winstral et al. (2017) modeled sheltering/exposure of locations to wind using the terrain parameter Sx (Winstral et al. 2002; see section 3a) and the topographic position index (TPI; Weiss 2001). Complementarily, Helbig et al. (2017) identified the local Laplacian from terrain elevations and squared slope as valuable parameters to downscale wind speeds. Recently, new statistical approaches have emerged to downscale wind fields in complex terrain (Bonavita et al. 2021). Notably, Dujardin and Lehning (2022) proposed an architecture based on convolutional neural network (CNN) to process both topographic information and NWP data in order to perform pointwise predictions of wind fields in the Swiss Alps at high resolution (<100 m). Höhlein et al. (2020) similarly used CNN to downscale wind fields but at a larger spatial scale, capturing physical processes that influence the motion of synoptic air masses.

In this study, we leverage on a combination of simulations obtained with a complex atmospheric model and deep learning methods to tackle the issue of wind downscaling in complex terrain. We proceed as follows: we use a high number of existing atmospheric model simulations performed with the Advanced Regional Prediction System (ARPS) atmospheric model, all obtained under controlled atmospheric conditions over a set of synthetic Gaussian topographies (Helbig et al. 2017). Using those simulations, we derive a link between coarse-scale wind fields (such as winds provided by an NWP), topography, and high-resolution wind fields through a parameterization, in a manner similar to statistical downscaling. However, in our case, the statistical relationship is automatically determined using an artificial intelligence model and, more specifically, a CNN.

2. Data

a. *ARPS simulations*

1) ARPS CONFIGURATION

ARPS is an atmospheric model that solves the nonhydrostatic and compressible Navier–Stokes equations. A detailed description of the model implementation can be found in (Xue et al. 2000, 2001). Notably, the model is able to represent several features of terrain-forced flow such as speedup on crests, sheltering, separation, recirculation, topographic channeling (Raderschall et al. 2008), and thermally driven winds such as valley breezes (Anquetin et al. 1998).

Helbig et al. (2017) performed individual simulations with ARPS on synthetic topographies derived from isotropic and stationary Gaussian random fields (GRF). GRF are stochastic processes, that have been identified as a good proxy for real topographies, particularly in their ability to approximate real slope distributions (Helbig and Löwe 2012) and have already been successfully used to develop topographic parameterizations (Helbig and Löwe 2012, 2014; Helbig et al. 2017). In this study, we make use of ARPS simulations performed on individual Gaussian topographies (Helbig et al. 2017). Each simulation covers a rectangle of 79 × 69 pixels with a horizontal resolution of 30 m. Notably, a broad range of topographic characteristics was achieved by selecting nine combinations of the two characteristic length scales: the typical width ξ [200–1000 m; see Eq. (1)] and typical height σ_{DEM} (88–364 m) of topographic features, for 5 spatial mean square slopes μ (19°–36°) within a topography:

$$\xi = \frac{\sqrt{2} \times \sigma_{\mathrm{DEM}}}{\mu}. \tag{1}$$

Each combination of ξ, σ_{DEM} generated 200 realizations, resulting in a total of 9000 topographies [for more technical details see Table 2 in Helbig et al. (2017)]. About 80% of the topographies (7279) resulted in usable simulated wind fields and were used in this study.

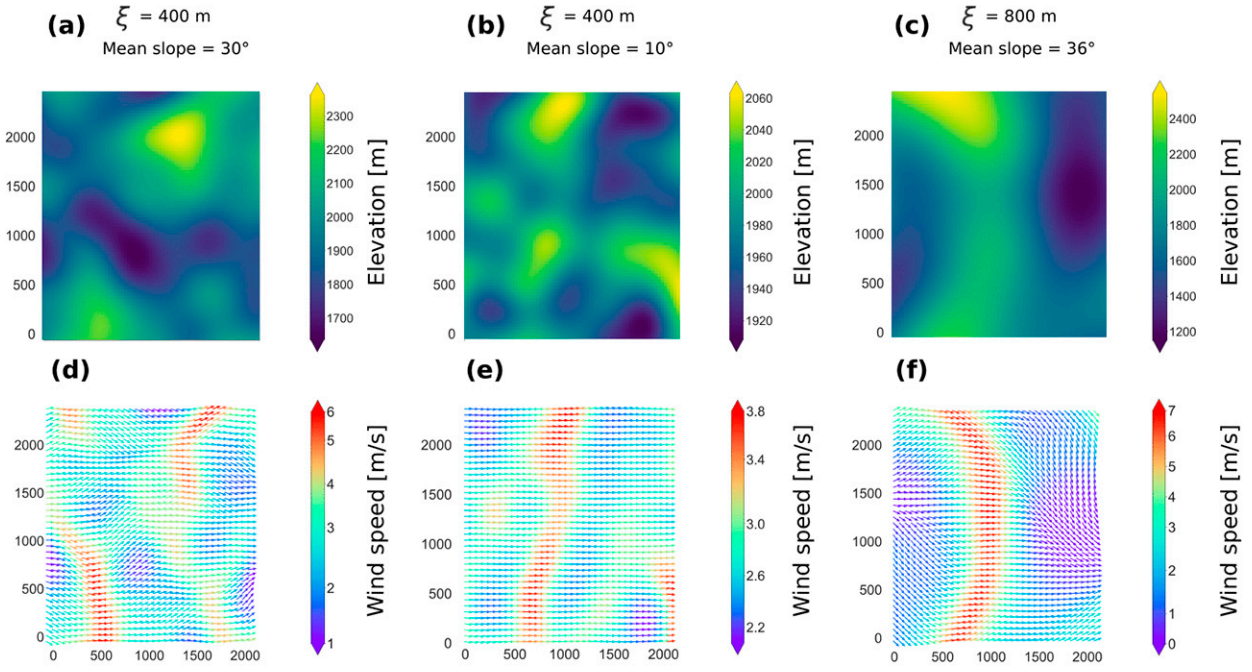

FIG. 1. (a)–(c) Maps featuring examples of Gaussian topographies, and (d)–(f) surface winds from the ARPS first layer on these topographies. The ARPS first layer has a mean elevation above ground of 2.95 m. These three simulations, as labeled, exemplify topography and model output couples that constitute our training database.

For these ARPS simulations, constant initial atmospheric conditions were chosen. Notably, all simulations were initialized with a constant wind profile, initially oriented from left to right (wind coming from the west) with a speed of 3 m s^{-1}. The atmospheric stability was fixed as neutral (as frequently observed during drifting snow episodes) and radiation effects were neglected (Helbig et al. 2017). Thermally driven flows were neglected to solely represent the interaction between large-scale flow and topography. The total integration time was limited to 30 s (with an integration time step of 0.1 s), prohibiting the dominance of turbulence in the outputs, and restricting the simulated flow to the resultant of the adaptation of a mean flow to local topography (Raderschall et al. 2008; Mott and Lehning 2010). In all simulations, the surface was representative of uniform snow-covered areas, with an aerodynamic roughness length of 0.01 m. We give an example of three ARPS simulations, encompassing different mean slopes (30°, 10°, and 36°) and ξ (400 and 800 m), and their associated topography in Fig. 1 (see also the same figure with normalized axis in section S.6 and Fig. S4 in the online supplemental material). Notably, we observe accelerations on peaks (red arrows) and deceleration windward and leeward (blue arrows). The intensity of the modifications of the high-resolution wind differs with mean slope and ξ, the largest modulations occurring on the steepest topographies.

2) CHARACTERISTICS OF THE SIMULATIONS

We describe in this section the characteristics of ARPS wind outputs (Fig. 2), which constitute our training database. The speed of the wind outputs (three-dimensional outputs, speed computed using $\sqrt{u^2 + v^2 + w^2}$) are distributed following Fig. 2a. As noted in Helbig et al. (2017), the mean wind speed simulated by ARPS is always slightly less than 3 m s^{-1}, the speed that served as initialization. Notably, the steeper the mean slope, the lower the mean wind speed. This behavior exemplifies the mean drag exerted by the topography on the flow and the associated loss of momentum, which is intensified on rougher terrain. Oppositely, the distribution tails highlight more frequent intense wind speeds on steep topographies. We note that ARPS simulates accelerations up to 4 times the initial wind speed and reductions to almost null wind speeds.

ARPS simulated wind fields deviated from the direction of the input wind (west) in both directions according to Fig. 2b. Counterclockwise and clockwise deviations are equally represented in our dataset and range from 0° to 82°. The distribution of angular deviations is centered on zero for each category of mean slope and deviations introduced by ARPS are generally low. Such deviations can be representative of flow deflections around obstacles, alignment of the flow on ridges and more generally encompass an adaptation of the flow to local topography. The formation of turbulent structures was deliberately prevented in the ARPS simulations used here as training dataset: the ARPS wind fields thus do not describe more complex behaviors of mountain winds, such as turbulent recirculation or extremely strong deviations (e.g., barrier jets), which are generally epitomized by higher angular deviations (Raderschall et al. 2008; Sharples et al. 2010; Whiteman 2000). Similar to the situation with

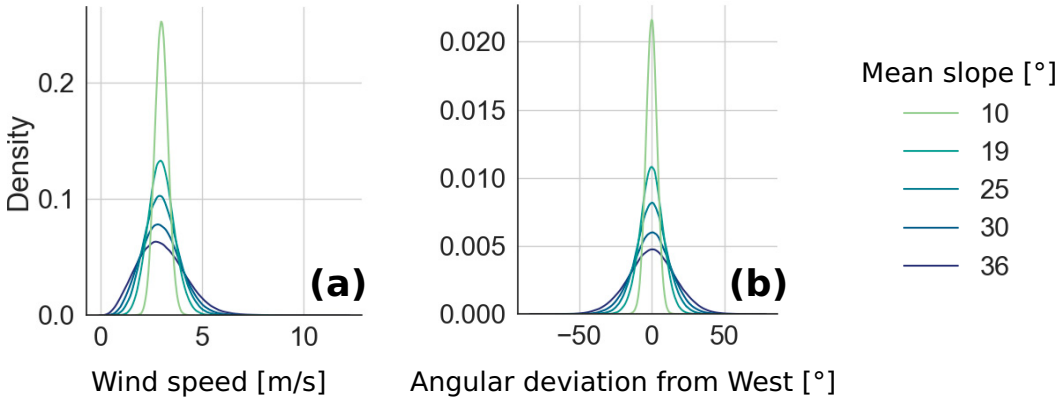

FIG. 2. (a) Wind speed and (b) angular deviation distributions as simulated by ARPS on the 7279 Gaussian topographies. Such winds constitute the training dataset used to fit the CNN of the DEVINE model.

wind speeds, we observe that the most intense wind direction modifications from the input wind occur on the steepest mean slopes.

b. *AROME simulations*

AROME is a limited-area NWP system used by Meteo-France (Seity et al. 2011). It provides short-term forecasts of atmospheric fields since 2008, over a domain encompassing the French Alps. Benefiting from its high horizontal resolution (1300 m) and complex physics and dynamics, the model has gained interest for mountain meteorology and snow sciences, progressively bridging the gap with coarser atmospheric products currently used to force snow models over the French mountain ranges (Quéno et al. 2016; Vernay et al. 2022; Gouttevin et al. 2023). AROME solves the nonhydrostatic fully compressible Euler equations system using hybrid pressure terrain-following coordinates. Notably, AROME uses a subgrid parameterization to describe the influence of unresolved orography on wind fields, via an effective roughness length described in Georgelin et al. (1994). We used here 10-m AROME wind fields initialized from the 0000 UTC analysis, from which we extracted daily forecasts between 0000 UTC + 7 h and 0000 UTC + 30 h at an hourly resolution. This way, we reconstructed continuous time series of gridded wind fields over the French Alps for a period of interest extending from 1 August 2017 to 31 May 2020.

c. *Observations*

Hourly observations of wind speed and direction have been collected and quality-checked in order to evaluate the downscaling scheme over real alpine topographies. A total of 61 automatic weather stations (AWS) acquiring wind measurements have been selected in the French Alps (Fig. 3). Most of them are part of Meteo-France operational observational network. Three stations: Vallot observatory (latitude = 45.83°, longitude = 6.85°; elevation = 4360 m), Argentieres glacier (latitude = 45.96°, longitude = 6.97°; elevation = 2434 m), and Saint-Sorlin glacier (latitude = 45.17°, longitude = 6.17°; elevation = 2720 m) are part of the glacier and climate observations {Glacier, an Observatory of the Climate [les Glacier, un Observatoire du Climat (GLACIOCLIM)]} network. Three other ones are located at Col du Lac Blanc (latitude = 45.12°, longitude = 6.11°; elevation = 2720 m) in the Grandes Rousses massif and belong to a high-mountain meteorological observatory dedicated to drifting snow and snow–atmosphere interactions (Vionnet et al. 2017; Guyomarc'h et al. 2019). The 61 sites cover the whole French Alps and a large variety of terrain, with some stations being located on flat surfaces, other on slopes, and some on exposed terrain (e.g., Aiguille du Midi: latitude = 45.87°, longitude = 6.88°; elevation = 3845 m). The observation stations are mainly located in nonforested areas and are mostly snow covered during the winter seasons.

Wind observations are commonly subject to measurement errors (DeGaetano 1997), particularly when collected in a challenging mountainous environment. These measurement errors can be of diverse nature and occur at different steps during the data collection process (Lucio-Eceiza et al. 2018a). A striking example of wind sensor dysfunction in mountain terrain is null and constant wind speed observations for several consecutive hours due to the accretion of ice on the sensor. Because our data come from different networks, their quality is unequal. Thus, we homogenized the quality standard of our dataset by applying a quality check, deeply inspired by Lucio-Eceiza et al. (2018a,b). These authors proposed a series of sequential tests designed to detect suspicious wind observations. We adapted the quality process of Lucio-Eceiza et al. (2018a,b) to fit the specificities of our dataset by selecting the most relevant tests and eventually introducing some modifications, as listed in section S.1 in the online supplemental material. We refer to Lucio-Eceiza et al. (2018a,b) for the evaluation of the quality process.

3. Method

This paper's method is organized as follows: we first build a statistical model by notably fitting a CNN to ARPS simulations. Then we use this statistical model to downscale wind fields from the AROME NWP system in the French Alps.

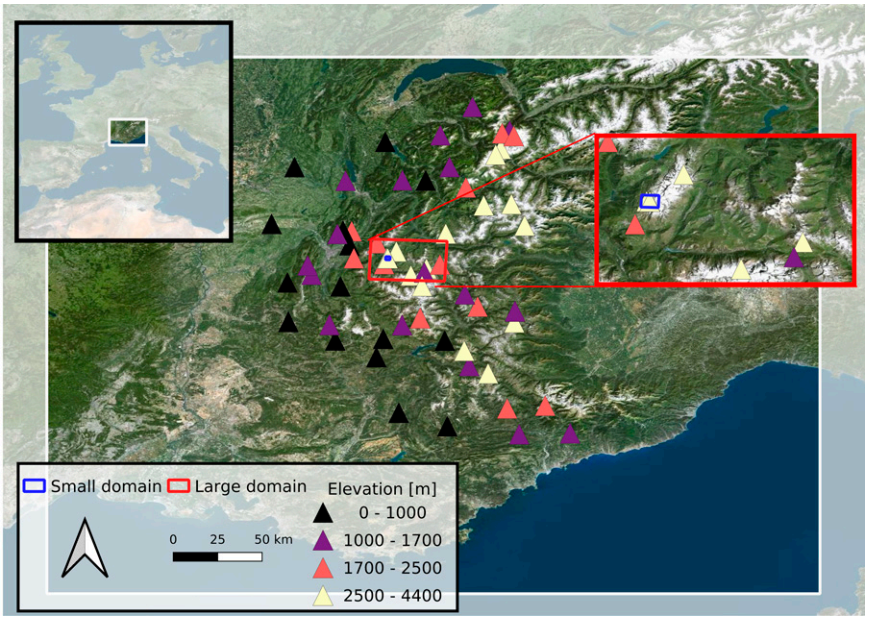

FIG. 3. Locations of observation stations (colored triangles) used for model evaluation. The colors of the observation sites indicate their elevation ranges. The small application domain used later in Fig. 6 is outlined in blue, and the larger domain that is used later in Fig. 7 is outlined in red and magnified in the zoom.

a. *Topographic descriptors*

In this study we make use of several parameters describing the topography and derived from digital elevation models (DEMs), all with a 30-m horizontal resolution dx. Here, we describe these parameters shortly with references: the TPI (Weiss 2001) compares the elevation of a DEM pixel with the mean elevation of the neighboring pixels given a fixed radius. The radius is equal to 500 m in our study, and consequently the TPI parameter is oriented toward the detection of topographic peaks/bowls on the slope scale. The Sx parameter (Winstral et al. 2002), is a direction-dependent parameter and quantifies how sheltered or exposed a pixel is within a given radius [here 300 m, as in Winstral et al. (2017)]. In detail, $\mathrm{Sx} = (z_j - z_i)/d_{ij}$, where $|\tan[(z_j - z_i)/d_{ij}]| = \max_k|\tan[(z_k - z_i)/d_{ik}]|$, with x_i being the cell of interest and k being the index of any pixel located in a zone starting from x_i and extending toward a direction defined by the incoming wind direction, within a 30° window and a 300-m maximum distance from x_i. Last, d_{ij} indicates the distance between x_i and x_j. In summary, positive values for Sx indicate sheltering for x_i, that is, how much x_i is protected from incoming wind within a 300-m radius, whereas negative Sx values quantify exposure. TPI and Sx are thus computed using information from neighboring pixels within a given radius and thus integrate information from areas located within a few hundreds of meters to characterize each DEM pixel. In contrast, the discrete Laplacian Δf:

$$\Delta(fx, y) = \frac{f(x + dx, y) + f(x - dx, y) + f(x, y + dx) + f(x, y - dx) - 4f(x, y)}{dx^2}, \tag{2}$$

which aims at detecting local peaks and bowls in topographic maps, and the squared slope, referred to as slope and computed following Helbig et al. (2017), only consider nearest-neighbor pixels in addition to the cell of interest and hence provide very local topographic information. We use these four parameters to characterize alpine and Gaussian topographies in sections 4 and 5.

b. *Fully convolutional neural networks*

CNN are a specific kind of neural network that benefits the use of convolution operations on tensors and are well suited for pattern recognition among spatialized data. Fully convolutional neural networks (FCN) are a specific type of CNN, proposing an end-to-end solution relying on convolutional and pooling layers without any use of dense networks, making them an efficient solution for gridded predictions. Convolutional layers consist of convolving a filter (i.e., a matrix with a predetermined size) to input data so as to detect spatial patterns. The product of convolutions goes through pooling layers that reduce their spatial resolution. Repeating both operations hence permits us to encode spatial features with a high level of abstraction. In FCN, encoding operations can be

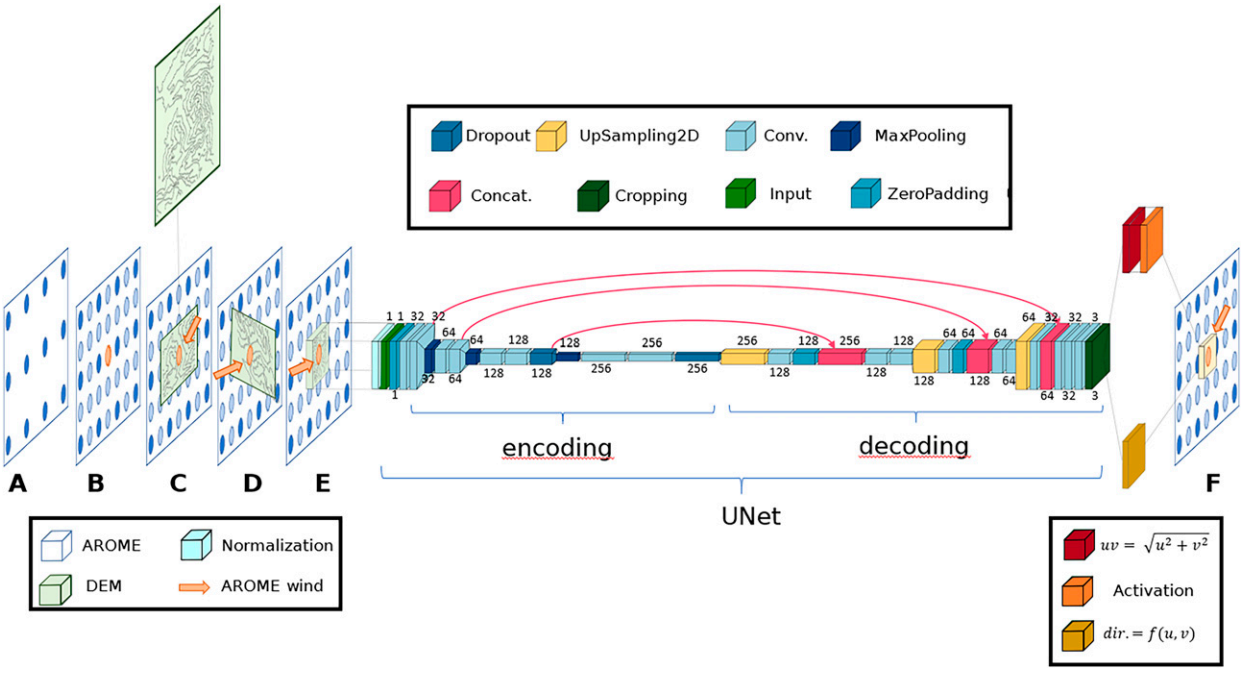

FIG. 4. Workflow of the downscaling model DEVINE. Both preprocessing operations (labeled A–E) and postprocessing operations (labeled F) are required before and after calling the CNN for predictions. In detail, label A corresponds to the selection of wind fields in the form of gridded outputs provided by an NWP system. This grid is interpolated (label B), and the following operations are done pixelwise: DEM data around each pixel are first selected (label C), then rotated with respect to the initial wind direction provided by the NWP system, and finally cropped (label E) to match the CNN input size. The CNN is then called and outputs high-resolution maps of wind fields. Within the CNN, the following standard operations are used: normalization, padding maps with zeros ("ZeroPadding"), convolutions ("Conv."), dropout connection ("Dropout," only during training), maximum pooling operation ("MaxPooling"), concatenations ("Concat."), cropping map borders ("Cropping"), and increasing the size of a matrix by repeating its rows and columns ("UpSampling2D"). Small numbers next to each layer represent the number of features maps. The scaled outputs of the CNN go through an activation layer to ensure that plausible values are produced. Wind patches are ultimately rotated back and placed on the high-resolution grid to constitute a continuous map of wind fields (label F).

followed by a decoding stage, where convolutions are mixed with spatial interpolations of the encoded signal, to sequentially increase the spatial resolution. The Unet architecture has been introduced in 2015 (Ronneberger et al. 2015) and constitutes a specific type of FCN frequently used for meteorological applications (Trebing et al. 2021; Fernández and Mehrkanoon 2021). In Unet (Fig. 4), encoding and decoding stages are connected through concatenation operations, which makes it possible to transfer high-resolution information to lower-resolution information within the model architecture. Indeed, data in the first layers of the Unet have not been through almost any pooling operation. Hence, this "raw" (or moderately encoded) information from the encoding stage is used to complement encoded and processed information from the decoding stage. Such an operation is also frequently referred to as "skip connection" (Lagerquist et al. 2021).

Here, two-dimensional topographic maps are fed into a Unet architecture. The model then outputs three features maps, each one of them representing a component of the wind vector. To determine the appropriate filters used in the convolutional layers, the Unet is fitted during a training step using Gaussian topographies as inputs and ARPS simulations as labels (see section 1). Model architecture and performance are investigated using 10-fold cross validation. Cross validation consists of randomly partitioning our database into "training" data (90% of the data) and "test" data (the remaining 10%), which permits us to fit the CNN on the first group and evaluate its performance on the second group. For a more robust evaluation, the process is repeated 10 times by rolling over 10 random training/test splits. Furthermore, we extracted validation data from the training data (i.e., 10% of the remaining 90% among the 10 folds) to follow a validation loss (mean absolute error on validation data) during training. We sequentially reduced the learning rate when the loss reached a plateau ("reduce on plateau"), and eventually stopped the learning process ("early stopping") whenever the validation loss stopped decreasing for 15 epochs ("patience"). This approach, coupled with the fact that after validation the CNN outputs are evaluated on an independent test set, aims at limiting the risk of overfitting the training set. In our specific case, hyperparameter tuning did not prove crucial to converge toward an efficient CNN architecture, as the training statistics highlighted a low sensibility to the different hyperparameters. We adopted a shallower version of the initially published Unet with only two additional layers corresponding to dropout connections, added to limit overfitting during the training phase. The selected

hyperparameters are summarized in section S.4 and Table S.1 in the online supplemental material.

c. DEVINE downscaling model

Several preprocessing and postprocessing steps are required before and after calling the CNN in order to generate high-resolution wind predictions on real topographies (Fig. 4). Their combination with the CNN forms a full downscaling scheme that we named Downscaling Wind Fields in Complex Terrain with Deep Learning, Applications for Nivology and Associated Challenges [Descente d'Echelle de Vent par Méthodes d'Apprentissage Profond, Application pour la Nivologie et ses Enjeux (DEVINE)].

First, the initial NWP gridded outputs (Fig. 4: label A) are upsampled using bilinear interpolation (using a factor of 2) (Fig. 4: label B). This step is necessary to ensure that the initial NWP signal is (i) dense enough to produce continuous wind fields at 30-m spatial resolution, and (ii) smoothly densified before calling the Unet, so that the final high-resolution maps produced are not affected by chesslike patterns. First, the initial NWP gridded outputs (Fig. 4: label A) are interpolated using bilinear interpolation and an interpolation rate of two (the dimensions of the interpolated grid are twice the original dimensions) (Fig. 4: label B). Two reasons justify the use of interpolation: firstly since the outputs of our CNN are of predetermined size (79 × 69 pixels), we have to ensure that predicted maps (around each NWP pixel) are sufficiently large to be juxtaposed one next to each other without introducing any hole in the final downscaled map. In our case, the grid spacing configuration (DEM = 30 m; NWP = 1300 m) ensures having no holes in the final maps. But for applications with, for example, another NWP system of coarser grid spacing, an initial interpolation of the NWP grid may be used to ensure spatial continuity. Second the use of interpolation is motivated by the necessity to reduce chessboard-like patterns. Such patterns appear at the NWP gridcell boundaries when downscaling gridded data (Winstral et al. 2017; Dujardin and Lehning 2022); see section 5c. They can be largely diminished by downscaling an interpolated map, that is, downscaling a higher-resolution grid where the transition between wind values of two neighboring grid points is less abrupt than in the original grid. Then, each interpolated NWP pixel is downscaled sequentially as described hereinafter. For each interpolated NWP pixel, high-resolution topography patches surrounding the pixel are first extracted (patches of size 140 × 140 pixels, Fig. 4: label C) and then rotated with respect to the NWP wind direction so that the rotated direction is from the west (Fig. 4: label D), similar to the ARPS simulations. The rotated patches of topography are then cropped (to 79 × 69 pixel size), normalized using their mean value and the standard deviation of the Gaussian topographies used during the training phase, and then go through the CNN (Fig. 4: label E), which produces three maps corresponding to the three components of wind speed. Horizontal wind speed and direction are then computed from wind components. This step eventually involves a linear scaling [$UV_{scaling}$; Eq. (3)] applied to the speeds outputted by the CNN [UV_{cnn}; Eq. (3)] using AROME wind speed [UV_{AROME}; Eq. (3)], as CNN outputs are only valid for an input speed of 3 m s^{-1} [$UV_{initARPS}$; Eq. (3) and section 2a]:

$$UV_{scaling} = UV_{AROME} \times \frac{UV_{cnn}}{UV_{initARPS}}. \quad (3)$$

We discuss the implication of this assumption in section 5. Given that we extrapolate results obtained on a set of Gaussian topographies under controlled atmospheric conditions to the very vast diversity of real topographies and atmospheric conditions (section 5a), we introduced a custom activation layer using a modified arctan function g that eventually curbs predicted unrealistic large wind speeds at extreme locations to plausible wind speeds:

$$g(x) = \alpha \arctan(x/\alpha). \quad (4)$$

This function does not act as a bias corrective term, but rather as a safety guard: we defined a maximum acceptable mean speed in our model (60 m s^{-1}), which is above the maximum observed speed in our historical observational dataset (36 m s^{-1}) and scaled the parameter α (=38.2) so that the g limit in positive infinity is equal to 60. The choice of the function arctan is motivated by the shape of the function g (monotonically increasing and progressively departing from the 1–1 line for speeds above 20 m s^{-1}; see section S.5 and Fig. S3 in the online supplemental material). Following these properties, g has an almost null impact on downscaled wind speeds lower than 20 m s^{-1} (i.e., most of the time) and starts lowering speeds for outputs larger than >20 m s^{-1}. Hence, g limits the simulation of unrealistic wind speeds that could potentially be obtained when particularly high NWP wind speeds are associated with the highest accelerations, generally obtained on the most exposed locations in real topographies.

In the final step of the DEVINE model, the downscaled pixels generated by the CNN are rotated back to their original position (Fig. 4: label F), in the opposite direction of the first rotation (Fig. 4: label D), and assembled to reconstruct a continuous regularly spaced grid of wind predictions, with a horizontal grid spacing of 30 m. CNN outputs can overlap one on each other and are thus cropped (patch size = 23 pixels) to avoid overlapping areas in the final predictions.

4. Results

a. Performance at emulating ARPS

We first assess the performance of the CNN at emulating ARPS on synthetic topographies on the test datasets, using 10-fold cross validation (Fig. 5 and Table 1). We computed the error between all predicted maps and corresponding labels pixelwise and categorized the CNN errors (i.e., the difference between modeled and observed value, defined at each time step) in terms of topographic characteristics (TPI, Sx, Laplacian, and slope) to interpret the results encapsulated in integrated metrics. Each TPI, Sx, Laplacian, and slope values are sorted according to their position in their respective distribution using the 0.25 quantile q_{25}, 0.5 quantile q_{50}, and 0.75 quantile q_{75}. We

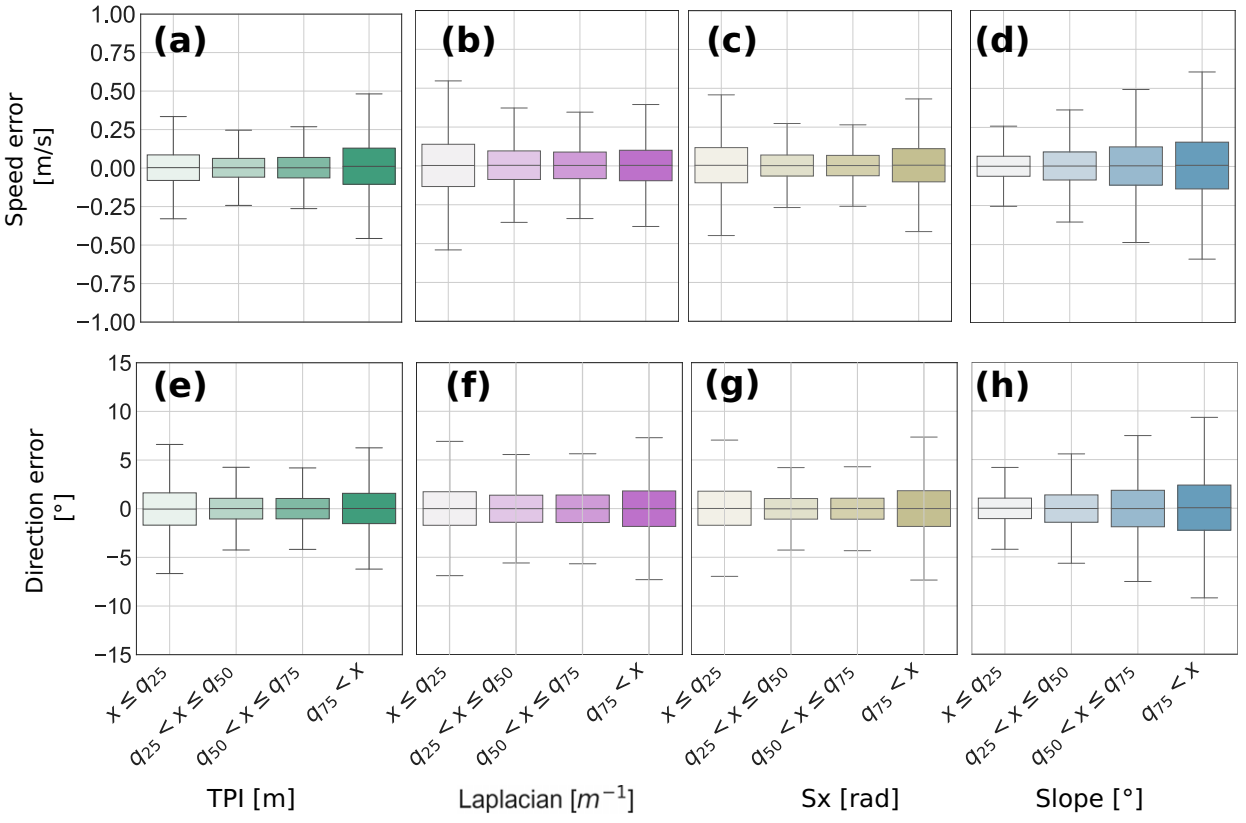

FIG. 5. Performance (errors) of the CNN at reproducing ARPS behavior on Gaussian topographies, using a 10-fold cross validation. All of the statistics presented have been acquired using test data (i.e., data not used during the training phase) and explored using parameters describing the topography (TPI, Sx, Laplacian, and slope). All parameters are categorized in four classes according to the quantiles of the parameter distribution (q_{25}, q_{50}, and q_{75}). Each boxplot has been computed using approximately 10 million samples.

observe that all errors on the test groups are centered on 0, suggesting that the CNN is able to produce unbiased estimations of ARPS wind fields. Errors are relatively low: boxplots suggest that most of the speed errors are lower than 0.25 m s^{-1}, with a mean absolute error of 0.16 m s^{-1} (5% of the initial wind speed; Table 1). Similarly, most of wind direction errors range between −5 and 5°. More specifically, we observe that the pixels corresponding to the most complex terrain (i.e., steepest slopes or tails of the parameter distributions) lead to larger errors, suggesting the difficulty to capture wind behavior in the most extreme terrains (Fig. 5). However, this could be explained by the fact that the largest wind speeds and angular deviations are also observed on the most complex terrain: when relative errors (i.e., errors at each time steps divided by ARPS wind speed) are considered, the performances of the CNN are more balanced among all topographic classes (see section S.2 and Fig. S1 in the online supplemental material). We note that the meridional component of the wind presents lower errors than the zonal component (Table 1), also in line with stronger zonal wind speeds due to ARPS initial conditions (wind coming from the west). Last, we analyzed the spatial location of the errors and observed that most errors are located

TABLE 1. Performance of the Unet model at emulating ARPS on Gaussian topographies, using a 10-fold cross validation. All the statistics presented have been acquired using test data (i.e., data not used during the training phase); U, V, and W respectively refer to the zonal, meridional, and vertical component of wind speed; UV designates the horizontal wind speed ($\sqrt{U^2 + V^2}$); and UVW is the three-dimensional wind speed ($\sqrt{U^2 + V^2 + W^2}$). Also, AE refers to absolute error and ρ is the Pearson correlation coefficient.

	U (m s^{-1})	V (m s^{-1})	W (m s^{-1})	UV (m s^{-1})	UVW (m s^{-1})	Wind direction (°)
Mean bias	<0.01	<0.01	<0.01	<0.01	<0.01	—
Mean AE	0.15	0.13	0.06	0.16	0.16	3
q25 AE	0.04	0.04	0.01	0.04	0.04	1
q50 AE	0.10	0.08	0.03	0.10	0.10	2
q75 AE	0.19	0.17	0.08	0.20	0.21	3
ρ	0.96	0.94	0.99	0.96	0.95	—

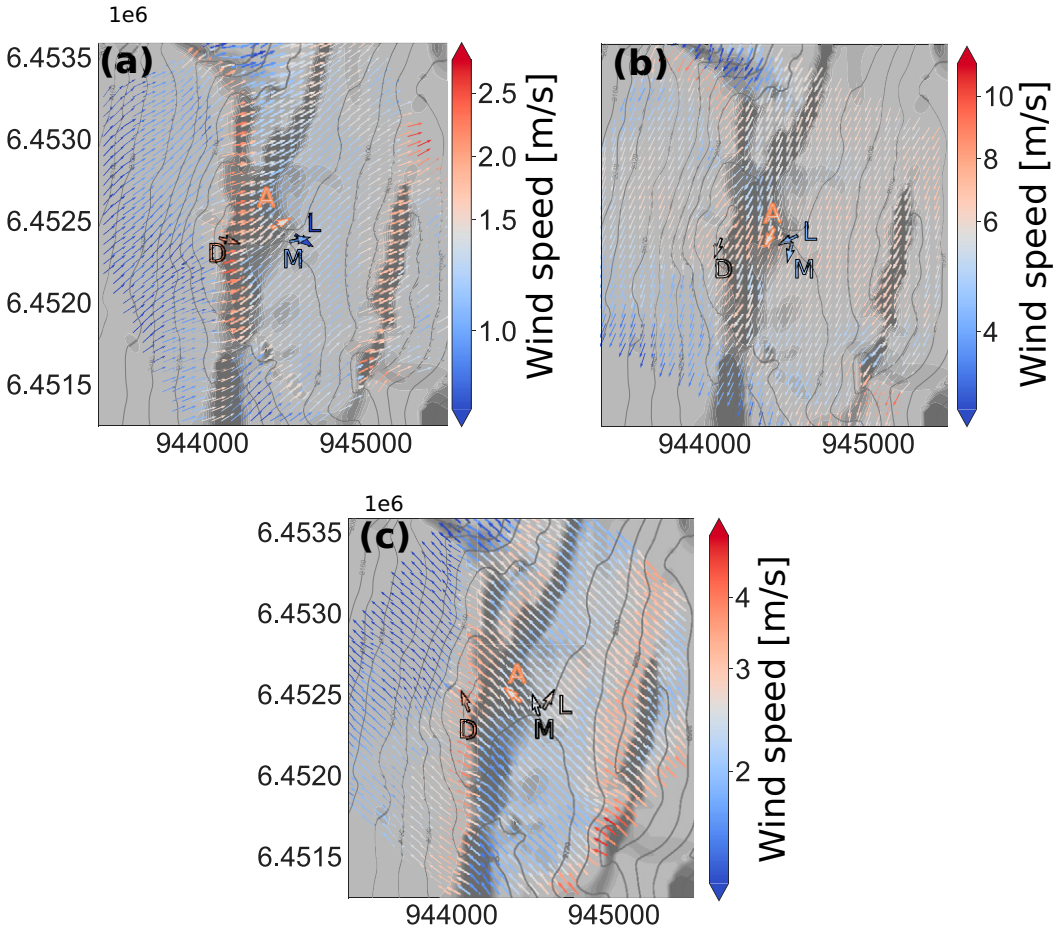

FIG. 6. Downscaling AROME wind fields at Col du Lac Blanc using DEVINE for three specific dates: (a) 1100 UTC 7 Mar 2018, (b) 0900 UTC 6 Apr 2021, and (c) 0000 UTC 9 Apr 2021; M, L, and D correspond to three in situ wind observations (Muzelle, Lac, and Dome stations), and A corresponds to the AROME forecast wind field. The x and y coordinates are expressed in Lambert 93 projection, i.e. EPSG 2154. Note that the left y axes are divided by 1×10^6.

on the margins of topography maps (see section S.3 and Fig. S2 in the online supplemental material).

b. *Case study: Application to a small domain*

We explore the behavior of DEVINE on real topographies by first downscaling a single AROME grid cell surrounding the Col du Lac Blanc experimental site in the French Alps (Fig. 6). This observation site (see section 2c) is composed of three distinct AWS located between tens to a few hundreds of meters from each other. The Lac station (L in Fig. 6; elevation = 2720 m) and Muzelle (M in Fig. 6; elevation 2722 m) are both located along a north–south pass. The relative proximity of the two stations makes them subject to very similar local wind conditions. On the contrary the Dome station (D in Fig. 6; elevation = 2808 m) is located on a small hill, dominating the pass by 85 m in elevation. This station is more subject to the influence of the hill on the local flow and is more exposed (TPI at Dome = 164 m, TPI at Muzelle = −24 m, and TPI at Lac = −27 m). In our case study, information from the nearest interpolated AROME grid point (A in Fig. 6; distance to the Lac station = 148 m, and pixel elevation = 2681 m) is first extracted for three dates (Fig. 6a: 1100 UTC 7 March 2018; Fig. 6b: 0900 UTC 6 April 2021; Fig. 6c: 0000 UTC 9 April 2021) and then downscaled using DEVINE. We selected the dates using the following criteria: (i) three distinct initial wind direction (respectively, west, north, and east as simulated by AROME), (ii) different AROME speeds ($\lesssim$3, $\approx$3, and $\gtrsim$3 m s^{-1}), (iii) AROME roughly in phase with local observations (speed error less than 2 m s^{-1} and direction error less than 90°), and (iv) including one "close to training" condition, that is, a neutrally stratified boundary layer ($-0.1 \leq$ observed Richardson number ≤ 0.1) with an AROME speed close to 3 m s^{-1} (Fig. 6c).

At 1100 UTC 7 March 2018, a west flux blows above the massif, with AROME locally simulating a low wind speed (1.4 m s^{-1}) with a west-southwest direction, (243°). DEVINE downscales the AROME signal and increases the speed on D (2 m s^{-1}). The flow is decelerated on both the windward and the leeward areas, but it is almost unchanged when compared with AROME at M and L (1.3 m s^{-1} at L and 1.4 m s^{-1} at M;

model). DEVINE acceleration patterns at D are consistent with the observations (observed wind speed at $D = 2$ m s^{-1}), but the model fails to fully capture deceleration at M and L (respectively, 1 and 0.4 m s^{-1} for the observations). In terms of direction, the observed wind aligns perpendicularly to the topographic barrier at D (285°, observation), a feature partially captured by DEVINE (253°; model). The model then suggests a small deviation of the flow toward the north of the pass (227°; model), which is slightly confirmed by observation at M (257°; observation). However, L suggests an orientation of the flow toward the south of the pass (303°; observation), but the direction observation is acquired for an almost null wind speed (0.4 m s^{-1}; observation).

On the second example (Fig. 6b; 0900 UTC 6 April 2021), the synoptic conditions indicate a flux from the north above the western French Alps, and AROME locally forecasts a 27° wind direction (north-northeast) characterized by larger speeds (5.5 m s^{-1}). DEVINE accelerates AROME wind at D (6.8 m s^{-1}; model), and across the pass at L and M (6.9 and 7 m s^{-1}). The observed speed is larger at D (5.8 m s^{-1}; observation) than at L and M (respectively, 4.1 and 4.6 m s^{-1}; observation). Interestingly, the acceleration at the pass, captured by DEVINE but not retrieved in the observation at this specific date, is a well-known behavior of wind fields at M and L, where drifting snow measuring devices have been specifically installed to monitor wind driven processes. In terms of direction, DEVINE suggests almost no deviation at D, L and M (33°, 19°, and 18°; model), a pattern not confirmed at L (64°; observation). The explanation of this divergence can also be retrieved in the observation acquisition process, because the quality control flagged as suspicious the direction observation at L for this specific date.

On the third and final case (0000 UTC 9 April 2021), synoptic conditions feature a flux from the south with AROME simulating a wind speed = 2.4 m s^{-1} and a wind direction = 130° (southeast). As in Figs. 6a and 6b, observations suggest stronger winds at D (3.4 m s^{-1}; observation) than L and M and this variability is captured by DEVINE (3.7 m s^{-1} at D; model). This could be interpreted by the incidence of the incoming synoptic wind forecast by AROME, which is rather perpendicular to the hill around D. AROME speeds are almost unmodified by DEVINE at M and L (respectively, 2.4 and 2.4 m s^{-1}; model), a pattern confirmed by the observation at M and L (2.4 m s^{-1} at M and 2.6 m s^{-1} at L; observations). DEVINE agrees with AROME direction without introducing any important directional shift at D, M, and L (respectively, 126°, 143°, and 144°). Oppositely, the observations suggest a flow from the south and a tendency to align along the pass at M and L (respectively, 163°, 208°, and 161° at D). In this example, a small difference in both speed and direction occurs between M and L, which might highlight very local phenomena and more generally the spatial variability of wind speed at scales below 30 m in complex terrain.

All the above examples illustrate how DEVINE produces accelerations and decelerations with respect to the underlying topographic features, and how zones of accelerations/deceleration and their relative intensity are a function of the AROME wind direction that served as initialization. This is of high interest for the modeling of drifting snow, as, for example, the zones of deceleration simulated in Fig. 6 could potentially favor flux convergence and hence shape drifting snow deposition (Vionnet et al. 2021). Further work might use more dense observation networks such as in Taylor and Teunissen (1987), Butler et al. (2015), and Wagenbrenner et al. (2016) to better characterize the spatial variability of DEVINE wind fields.

c. *Case study: Application to a large domain*

The DEVINE architecture can also be deployed on larger domains than Fig. 6 to downscale wind fields provided by gridded outputs of an NWP (Fig. 7). We selected a 40 km by 30 km domain in the French Alps (Fig. 7a), and downscaled AROME wind forecast at 1500 UTC 11 July 2019 (Figs. 7b–d). For this specific date, the dominant wind direction was from the west with maximum AROME speeds reaching 7.7 m s^{-1}. Figure 7 shows the initial AROME wind with a 1300-m horizontal resolution (green/yellow arrows) and the downscaled speeds at 30-m horizontal resolution (violet/orange color). We observe strong modifications of speeds on this domain (maximum downscaled speed = 16.2 m s^{-1}). Highest downscaled wind speeds appear on clear lines in Fig. 7b, which mostly correspond to mountains ridges and summits as identified in Fig. 7a. Oppositely, dark violet areas correspond to low wind speeds as simulated by DEVINE and are often in phase with low AROME speeds (small arrows), suggesting that even if DEVINE modifies the AROME signal, it still respects important features provided by the NWP system (areas of high wind speeds vs areas of low wind speeds). Thus, even though DEVINE has been fitted to simulations performed assuming certain weather conditions and is not able to reproduce some processes of mountain winds at a local scale (recirculation areas, thermally driven flows, etc.), the larger-scale wind fields simulated by AROME, which can be obtained under all types of weather conditions, can be used to drive the downscaling model. Accelerations and decelerations lie in the range of accelerations/decelerations as simulated by ARPS on synthetic topographies (section 2). We also observe strong accelerations with DEVINE on the Grandes Rousses massif (Fig. 7a; white ellipse) and on the Aiguille d'Arves massif (Fig. 7a; red ellipse), which are both oriented toward a north–south axis (thus perpendicular to the dominant synoptic wind on 11 July 2019). Oppositely, the north of the Ecrin massif (Fig. 7a; black ellipse) is oriented east–west, along the main wind direction, and DEVINE does not suggest any strong acceleration on this more complex and higher massif. Thus, we draw the conclusion that on this example DEVINE is able not only to detect ridges and complex terrain but also to interpret the incidence angle between AROME wind fields and topographic features. We also observe angular deviations with respect to the AROME initial direction (Fig. 7c; red and blue colors). These deviations mainly lie in the range of ARPS deviations on Gaussian topographies (deviations from AROME direction up to 81° on this specific case study). The presence of red features (counterclockwise deviations) in Fig. 7 next to blue features (clockwise deviations) suggests local circumventions of the flow

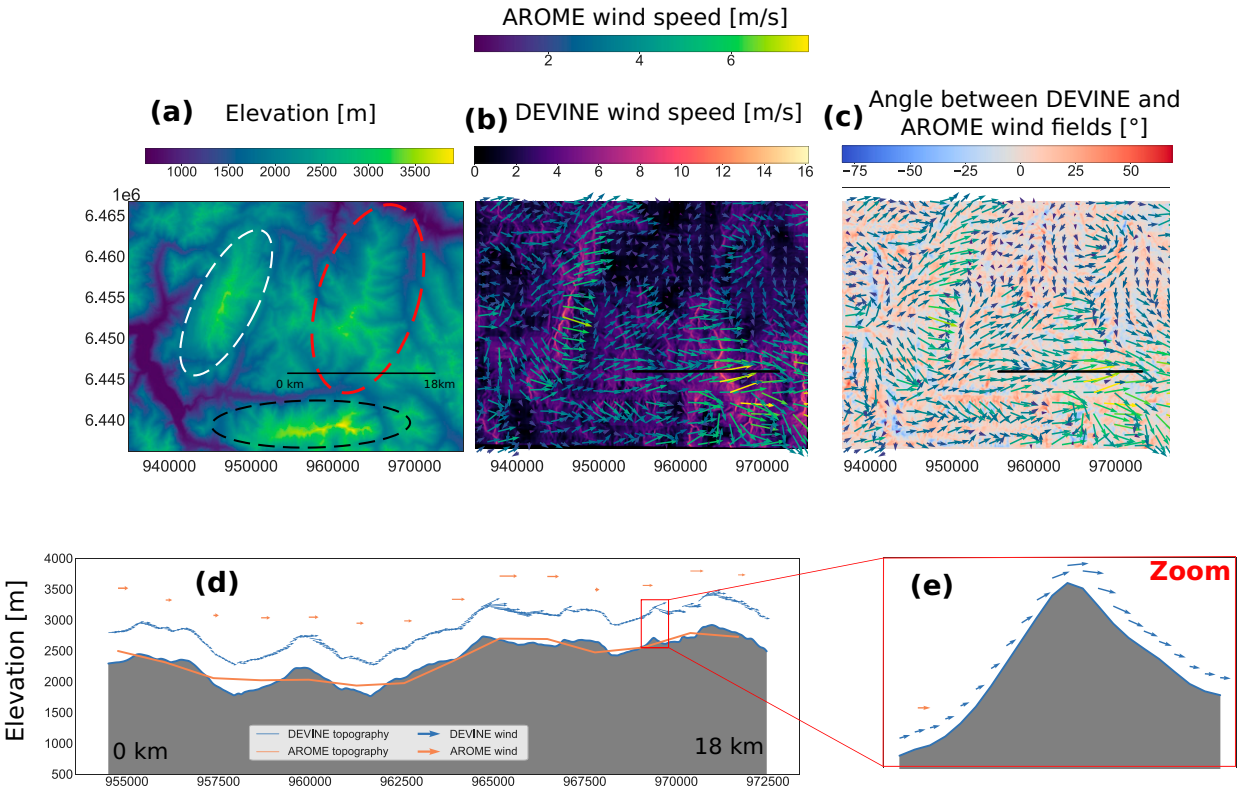

FIG. 7. (a) Topography of an alpine domain and wind field simulations at 1500 UTC 11 Jul 2019, with (b) AROME (colored arrows) and DEVINE wind speed (color shading) on the domain, (c) AROME (colored arrows) and DEVINE angular deviations from AROME (color shading) on the domain, and (d) topography and DEVINE wind speeds projected on a vertical west–east transect, with (e) an enlargement of the area within the red-outlined box in (d). On this transect, the bases of the arrows locate the associated wind field. AROME wind fields are two-dimensional 10-m wind fields, whereas DEVINE wind fields also incorporate a vertical dimension.

around topographic obstacles and channeling within valleys. Last, we projected in Fig. 7d DEVINE 3D wind fields on an 18 km west–east transect (see the black horizontal bar in Figs. 7a–c). We observe on this transect accelerations (larger arrows) on peaks and changes in the vertical component of the wind field, which generally tend to follow the terrain and to be oriented with respect to topographic features. From a modeling perspective, topography imposes surface conditions that force the wind speed to orient following the terrain.

d. *Performance on real topographies*

When compared with in situ observations collected at the 61 observation sites in the French Alps, AROME wind fields are characterized by a negative speed bias for high wind speeds (>10 m s^{-1}) (Fig. 8a). This behavior has already been documented (Vionnet et al. 2016; Gouttevin et al. 2023) and might be to some extent explained by the 1.3-km horizontal grid spacing of the model that fails to capture wind accelerations on ridges and speed variability due to subkilometric variations of elevation. For lower wind speeds (<10 m s^{-1}), AROME is closer to the observations. The departure of the points from the 1–1 line for high wind speeds in Fig. 8a is partly corrected by DEVINE downscaling (Fig. 8b), supporting our above hypothesis. Conversely, some low observed wind speeds are overestimated by DEVINE. Further analyses show that the overestimation of low wind speeds by DEVINE is, among other factors, imputable to the behavior of the model at a few exposed observation stations, such as the Aiguille du Midi station (TPI = 288 m). These behaviors are discussed in section 5a.

A quantile-to-quantile analysis complements the comparison of model wind speeds with in situ observations and indicates that the observed wind distribution is better captured by DEVINE than by AROME (Fig. 8c). The underestimation of high wind speeds in AROME is reflected by a departure from the 1–1 line in Fig. 8c for high quantiles. As partially observed in Fig. 8b, high wind speeds are better captured by DEVINE, which is reflected by a better representation of the highest wind speed quantiles. However, as previously pointed out in Fig. 8b the overestimation of low wind speeds contributes to simulating larger speeds and hence to obtain a better match on the 1–1 plot with DEVINE than with AROME in Fig. 8c.

Further analysis reveals that DEVINE reduces wind speed mean biases at most of the observation stations, whereas direction is only slightly affected. In Fig. 9, wind speed errors have first been normalized by the observed wind speed and then categorized by TPI. Only speeds above 1 m s^{-1} are

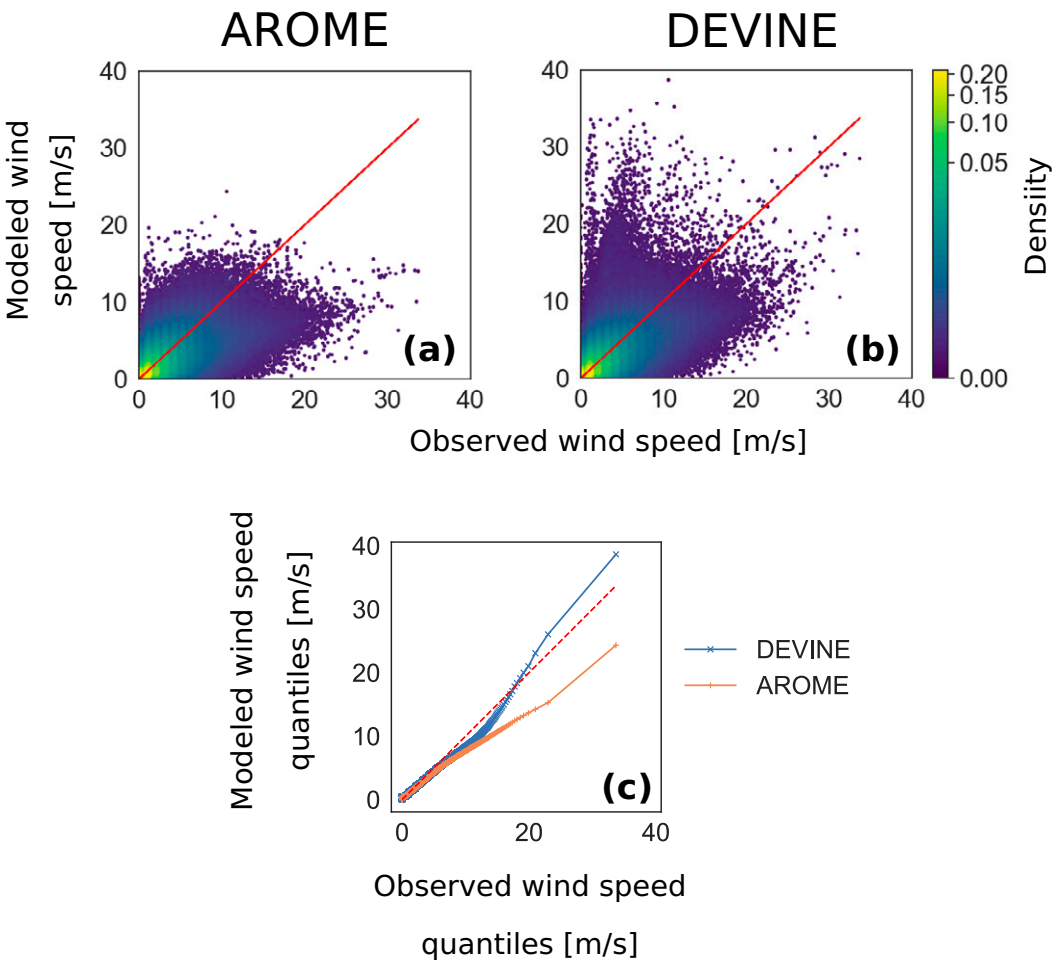

FIG. 8. (a) AROME 1–1 plot (modeled vs observed hourly values), (b) DEVINE 1–1 plot, and (c) quantile–quantile plot (each point being a quantile in the respective distributions, both discretized in 10 000 quantiles). In (a)–(c), the red line represents the 1–1 line. The color bar in (b) represents data density for (a) and (b).

considered in this process. Error normalization is preferred in our case as generally elevated stations are characterized by stronger winds and consequently larger errors, thus scrambling the interpretation when compared with stations located at lower elevations that are less prone to strong winds. DEVINE reduces speed median errors among half of the TPI classes, with a notable improvement for the most exposed stations, characterized by the largest TPI (q_{75} < TPI), where the negative speed mean bias is reduced by 66%. Among the 61 observation stations, DEVINE significantly reduces wind speed mean bias at 37 stations (dependent Student's t test at a 95% confidence interval, using Bonferroni method for multiple-comparison correction), a feature concerning 58% of the stations located above 1500 m. We also notice that correlation in speed signals is only slightly improved with DEVINE (0.56 for AROME vs 0.58 for DEVINE; Table 2). However, Fig. 9b indicates that errors in direction are almost unchanged between AROME and DEVINE (average change in direction error less than 1°). Furthermore, we observe that AROME can be affected by strong direction errors (>90°), for which only small modifications are brought by DEVINE: the maximum modification in direction introduced by DEVINE, at the 61 stations, is equal to 37°. We link this to the fact that (i) our CNN is fitted on ARPS model outputs where wind direction is generally only slightly modified by topography (Fig. 2a) and (ii) the largest wind direction modifications are expected on the most complex terrain (Fig. 2a), which corresponds to areas not necessarily equipped with AWS (section 5).

We finally investigate the errors of DEVINE when compared with AROME, in a seasonal and diurnal perspective (Fig. 10). This is done by considering the difference in the normalized absolute error (defined as the absolute value of the error at each time step divided by the observed speed) of both wind products, categorized into three distinct groups based on the observed wind speed. The evolution of the normalized absolute error confirms that DEVINE mostly improves forecast for higher wind speeds (Fig. 10). Lowest wind speeds are generally characterized by an increased absolute error with DEVINE (brown color in Fig. 10), and wind speeds above 3 m s^{-1} are on the contrary characterized by a

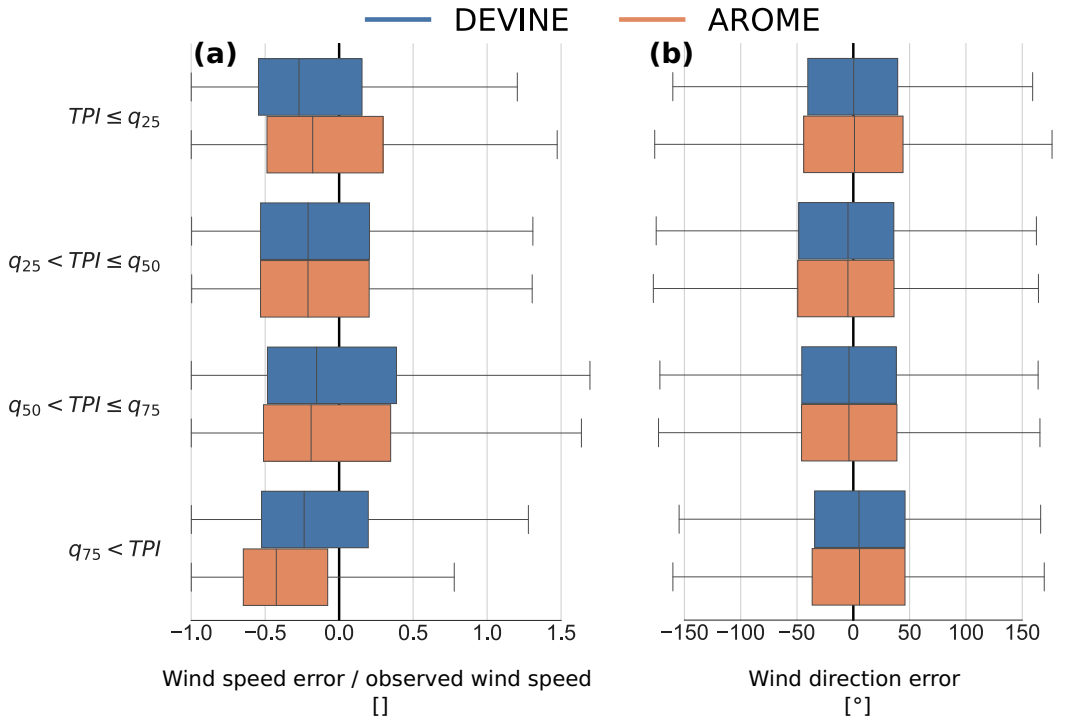

FIG. 9. Performances in wind (a) speed and (b) direction of AROME and DEVINE on real topographies with respect to in situ observations (for observed wind speed ≥ 1 m s^{-1}). Results are clustered by TPI according to the quantiles of the TPI distribution (q_{25} = −15 m, q_{50} = 3 m, and q_{75} = 32 m).

decreased error with DEVINE (green color in Fig. 10). It is interesting to note that for wind speeds below 3 m s^{-1}, the largest increases in normalized absolute error occur between April and September, and between 0900 and 1700 UTC. This tends to occur at periods of highest incoming shortwave radiations and temperatures, suggesting some influence of thermal processes (section 5c). Oppositely, visible improvements occur with DEVINE during nights of summer months where atmospheric stability could be closer to neutral conditions than during afternoons of the same periods. More generally, we observe that improvements/degradations with DEVINE tend to be more linked with validity hour rather than forecast lead time hour, also suggesting an influence of thermal processes.

5. Discussion

a. *Representativity of Gaussian topographies*

The characteristics of Gaussian topographies accurately approximate the mean characteristics of alpine topographies (Fig. 11). For every pixel of a digital elevation model covering the French Alps, thus including the observation sites that served for model evaluation in section 4d, we computed the TPI, Sx (using a fixed direction of 270°), Laplacian, and slope.

TABLE 2. Evaluation statistics of AROME and DEVINE when compared with 61 AWS located in the French Alps, in terms of wind speed and direction. DEVINE$_c$ designates DEVINE performances when DEVINE is initialized by realistic AROME forecasts. AROME forecasts are here considered to be realistic when the speed error is less than 3 m s^{-1} and the direction error is less than 30° (AROME$_c$).

Variable	Metric	AROME	DEVINE	AROME$_c$	DEVINE$_c$
Speed	Mean bias (m s^{-1})	−0.33	−0.24	−0.29	−0.17
	ρ (—)	0.56	0.58	0.73	0.72
	Mean AE (m s^{-1})	1.40	1.37	1.06	1.07
	q_{25} AE (m s^{-1})	0.44	0.44	0.41	0.40
	q_{50} AE (m s^{-1})	0.98	0.96	0.91	0.89
	q_{75} AE (m s^{-1})	1.89	1.84	0.91	0.89
Direction	Mean AE (°)	58	57	14	14
	q_{25} AE (°)	18	17	7	6
	q_{50} AE (°)	43	41	14	13
	q_{75} AE (°)	88	87	14	13

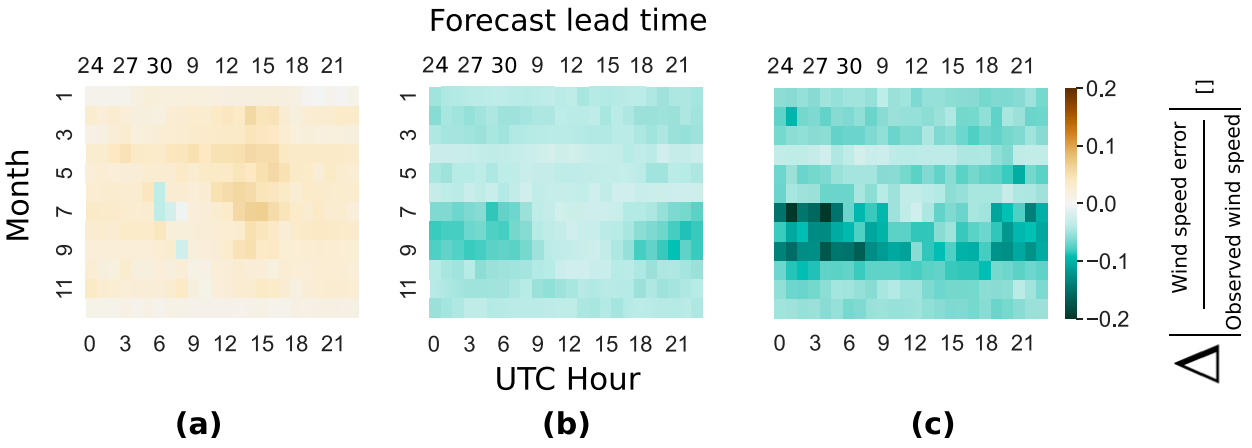

FIG. 10. Evolution of the mean of the normalized absolute error (nAE) for wind speed between AROME and DEVINE ($nAE_{DEVINE} - nAE_{AROME}$), categorized by hour of the day and month of the year (the mean of the nAE corresponds to the mean of the absolute value of the error at each time step divided by the observed wind speed) (a) only considering observed wind speed between 1 and 3 m s^{-1}, (b) for wind speeds between 3 and 7 m s^{-1}, and (c) for wind speeds above 7 m s^{-1}. Negative values correspond to improvements (green), and positive values correspond to degradations (brown).

We then compare the results with the same parameters obtained on the Gaussian topographies forming our training dataset (Fig. 11). We observe that for each parameter, the distribution obtained on Gaussian topographies (yellow distribution) overlaps most of the distribution obtained on real topographies (green distribution) suggesting that, with respect to the chosen parameters, most of the alpine topographic pixels are represented in our Gaussian topographies dataset. This strengthens the results of Helbig and Löwe (2012), who showed that Gaussian statistics outperformed other statistical models when representing slopes of real topographies in complex terrain.

However, for individual pixels, some terrain parameters derived on the real alpine topographies are not encompassed in the range of the same parameters derived on the Gaussian topographies. These correspond to pixels located in extremely complex terrain. Notably, the tail of the distribution of Laplacians [Δf; Eq. (2)] computed on real topographies

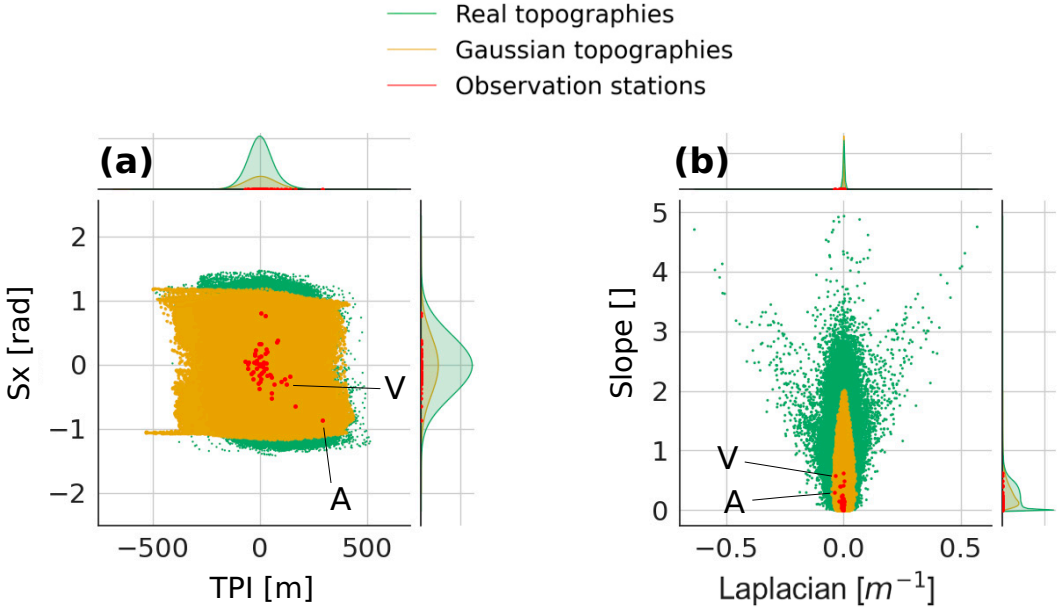

FIG. 11. Parameters computed for each point of a real digital elevation model ("real topographies"; green), only at sites with wind observations ("observation stations"; red), and on Gaussian topographies used in our training dataset ("Gaussian topographies"; orange): (a) TPI vs Sx and (b) Laplacian [Eq. (2)] vs slope. The Sx values are computed using a wind direction of 270°; A indicates the Aiguille du midi station, and V indicates the Vallot station.

TABLE 3. DEVINE computing performances on a 1250-km^2 domain presented in Fig. 7 (40.9 km by 30.5 km; horizontal resolution of 30 m). The computing time does not account for data loading overhead (loading DEM map, NWP gridded outputs, etc.). Note that because the downscaling operation is not sequential through time it can be easily parallelized across time by using different processing units.

	Name	Value
Domain	Domain size	40.9 km by 30.5 km
	Horizontal resolution (input)	1300 m
	Initial interpolation rate	2
	Horizontal resolution (interpolated)	650 m
	No. of interpolated NWP grid points to downscale	3072
Performance on CPU	CPU model	Intel Core i7-10610U CPU at 1.80 GHz
	No. of CPU	1
	Prediction (downscaling 1 time step)	64 s
	Prediction (downscaling 24 time steps)	1461 s
Performance on a GPU + CPU	GPU model	Nvidia Tesla V100
	No. of GPU	1
	Prediction (downscaling 1 time step)	14 s
	Prediction (downscaling 24 time steps)	97 s

exhibits the largest discrepancies with respect to their Gaussian counterparts.

Most of the topographies surrounding the observation stations, however, have characteristics well represented in our Gaussian topographies (red dots in Fig. 11). We nevertheless note that some extreme locations such as the Aiguille du Midi (TPI = 288 m, Sx = −0.88 rad, Laplacian = −0.039 m s^{-1}, and slope = 0.29) or the Vallot station (TPI = 120 m, Sx = −0.3 rad, Laplacian = −0.033 m^{-1}, and slope = 0.58) are localized on the tails of the distributions obtained on the Gaussian topographies. We note that DEVINE frequently overestimates low wind speeds at Aiguille du midi, suggesting current limitations of our method on such extremely exposed and complex terrain, for which only few information can be derived from the training dataset. It would be questionable that adding more complex topographies to our training dataset would be beneficial for DEVINE since numerical limitations and errors can arise when computing wind fields on very steep slopes with mesoscale models (e.g., Lundquist et al. 2012).

b. *Efficient downscaling by using a CNN with reduced complexity*

Wind downscaling with DEVINE is particularly useful as it comes with a low computational cost, when compared with much more computationally expensive atmospheric models such as ARPS. Here, we optimized the model implementation using a strategy leveraged on the graphical processing unit (GPU) only for raw CNN predictions plus computationally expensive rotations and left other pre- and postprocessing operations (interpolation, normalization, final activation, etc.) for the CPU. As a result, it is possible to downscale AROME fields as presented in Fig. 7, in 14 s (Table 3). These performances pave the way for the use of our method in time constrained applications, for instance as a downscaling tool to reach decametric scales within operational forecasting systems in complex terrain.

Such a low computational cost is also attributable to our choice of limiting input channels of the CNN to topographic maps only. As a consequence, DEVINE only requires minimal inputs (i.e., topography and initial wind fields provided by an NWP system) to output downscaled wind predictions at a high resolution. Following the topographic nature of the selected inputs, we observed that DEVINE is able to detect main features of terrain-forced flow, including the representation of acceleration on ridges, deceleration on leeward areas, flow deflections and moderate deviations around obstacles (Figs. 6 and 7). It is worth noting that contrary to Dujardin and Lehning (2022), who reached state-of-the-art results using a downscaling model also based on CNN, we did not used precomputed topographic parameters (TPI, aspect, etc.) as inputs. Indeed, we converged to low errors on the test dataset by simply using raw topographic maps. Three factors emerge to explain our choice to use raw topographic data as inputs versus preprocessed topographic data as in Dujardin and Lehning (2022): our training dataset encompasses more topographies [7279 vs 261 for Dujardin and Lehning (2022)], our CNN learns topographic features related to Gaussian topographies versus real topographies in Dujardin and Lehning (2022) and as we only have one channel in input, feature detection is directly oriented toward topographic characteristics whereas Dujardin and Lehning (2022) join topographic maps with many other atmospheric variables when constituting the input channels, which could eventually make the detection of topographic characteristic less direct.

The impact of high-resolution topography on wind fields is reflected in the evaluation statistics, including a decrease in the mean bias, a more moderate reduction in MAE and a slightly increasing correlation. We emphasize the fact that our method is only based on an emulation of the atmospheric model ARPS and does not need to be calibrated with any observation, in contrast to Pohl et al. (2006), Liston and Elder (2006), Winstral et al. (2017), and Dujardin and Lehning (2022). It highlights that DEVINE upsamples NWP wind fields but does not explicitly involve a bias correction, as was

done, for example, in Winstral et al. (2017) and Dujardin and Lehning (2022). Thus, DEVINE is independent from the NWP system providing the initial information. However, as the initial errors of the NWP are not compensated with any calibration step, they can eventually be propagated (Wagenbrenner et al. 2016) and amplified through DEVINE. For instance, we observe that when the wind fields simulated by AROME are in phase with the observation, that is, a direction error less than 30° and a speed error less than 3 m s^{-1}, DEVINE mean bias is even more reduced (Table 2).

c. *Limits of our approach*

Apart from error propagation due to the initial errors of the NWP (section 5b), limitations on the most extreme alpine terrain (section 5a), and even though the CNN reproduces particularly well the ARPS simulations, some errors remain when predicting wind on real topographies (section 4). Indeed, for simplicity and computational efficiency, we pruned DEVINE to a minimalist architecture. In particular this was possible following the assumptions used in the setup of the ARPS model (Helbig et al. 2017). Inheriting the assumption of the ARPS configuration, DEVINE assumes (i) a neutral stratification of the atmosphere, (ii) an absence of thermal processes, (iii) mostly nonturbulent flow, and (iv) a linear behavior between the wind flow obtained for a 3 m s^{-1} initial speed and output obtained with any other speed.

For DEVINE, the assumption i on the neutral stratification of the atmosphere may explain the model's limited ability to drastically change wind direction. The thermal stability of the atmosphere influences the motion of air masses in complex terrain, and more particularly, is responsible for large deviations of stable and heavy air masses (Whiteman 2000) that tend to get around obstacles rather than above. Additionally, assumption iii is responsible for the absence of small-scale turbulent eddies in DEVINE simulations and thus also explain the model difficulty to simulate large modifications in wind direction.

We also attribute increasing normalized errors observed in Fig. 10, for the lowest wind speeds, to the absence of thermal processes in DEVINE. Butler et al. (2015) and Sharples et al. (2010) indicated the prominence of along-slope and valley flows during spring/summer months, underlining that the intensity of some thermal flows largely depends on seasonality. Interestingly, we observe that the largest degradations with DEVINE occur for observed wind speeds less than 3 m s^{-1} during days of spring and summer month.

Moreover, using additional ARPS simulations performed by Helbig et al. (2017) on a small group of the Gaussian topographies using a different initial wind speed (5 m s^{-1} instead of 3 m s^{-1}), we challenged our assumption iii on linearity. We observed that the acceleration rates obtained with an initial speed of 3 m s^{-1} are consistent with the acceleration rates obtained at 5 m s^{-1}, thus suggesting that the assumption on linearity holds (not shown). However, we can expect that a nonlinear relation would arise with higher wind speeds, which is still to be benchmarked.

As described in section 1, DEVINE exhibits features of a mostly nonturbulent flow, including the absence of recirculation zones (Raderschall et al. 2008; Sharples et al. 2010), whose impact on snow deposition is important (Vionnet et al. 2021). Avoiding the inclusion of turbulent features favors the development of simple yet generalizable outputs and is in phase with the choice of an initial speed of 3 m s^{-1} in ARPS simulations, where the generation of turbulent eddies due to wind–topography interaction is probably low (Whiteman 2000). Again, this assumption might not stand for higher speeds.

We underline that simultaneous occurrence of assumptions i, ii, iii, and iv in the atmosphere is probably rare. However, even though DEVINE is limited in terms of physical processes it can represent at the slope scale, it proves to add value as a downscaling tool of NWP wind fields in all weather conditions (Table 2, Fig. 10). Furthermore, the NWP system driving DEVINE is not bound to these assumptions so that it provides a representation of all atmospheric situations, to the extent enabled by its spatial resolution and inherent assumptions.

Note that AROME uses a parameterization of the subgrid topography following Georgelin et al. (1994). Consequently, high-resolution wind fields obtained using DEVINE have interacted twice with the topography: a first time through AROME subgrid parameterization and a second time through DEVINE. This redundancy could contribute to errors in wind fields estimations.

Also note that AROME wind fields (arrows in Fig. 7) are first interpolated to double the horizontal resolution and limit the establishment of chesslike patterns in the output signal (section 3c). Chessboard-like patterns correspond to colored squares in the downscaled signal: as (i) two neighbors in an NWP grid can forecast different wind conditions and (ii) as each (interpolated) NWP grid point is treated independently from its neighbors by DEVINE, discontinuity can appear at the border of each grid cell in the downscaled signal. Such patterns progressively disappear by increasing the interpolation rate, at the expense of more computing time. Ultimately, we note that the impact of chesslike patterns on the downscaled signal could impact drifting snow modeling, a task that requires spatially coherent wind forcing for the computation of snow flux divergences. This impact still has to be quantified through the evaluation of distributed snowpack simulations.

6. Summary and conclusions

DEVINE is a downscaling scheme based on deep learning, relying on a fully convolutional neural network (Unet-like), that downscales NWP gridded wind fields from a grid spacing on the order of several kilometers to tens of meters using topographical information only. This model has been fitted using simulations obtained with the model ARPS on a set of 7279 Gaussian topographies. We demonstrated that the Unet architecture is performant on a cross-validation dataset to emulate the behavior of ARPS on synthetic topographies. By evaluating our model using simulations performed on real topographies and by using quality-checked data from 61 observation stations in the French Alps, we showed that DEVINE partially improves AROME wind speed forecasts, and is able to reproduce

observed wind speed patterns, thus providing a numerically efficient alternative to complex atmospheric models for simulations of high-resolution wind fields. Most notably, DEVINE reduces AROME mean bias, slightly reduces the absolute error, and increases the correlation. Moreover, DEVINE outputs are consistent with the well-known influence of main topographic features (peaks, slopes, and ridges) on airflow at local scale. Our method is developed for snow-transport applications and therefore does not account for some processes that may be controlling the wind patterns at a local scale in other conditions like thermal stability and thermal winds. Using transfer learning and additional model simulations could be of interest to complement current capabilities of DEVINE. This would probably require thousands of new high-resolution simulations to be used as labels and may induce modifying DEVINE architecture. Additionally, we discussed the influence of several factors that could restrict DEVINE applicability and conclude that future work might focus on an indirect distributed evaluation of the downscaling model, through the use, for example, of remotely sensed data and drifting snow models. Moreover, we note that reducing the initial biases of the NWP could limit error propagation when increasing the spatial resolution of wind fields using DEVINE. A comprehensive intercomparison exercise of state-of-the-art downscaling models could help to benchmark solutions for drifting snow applications.

Acknowledgments. This research is supported by the French Meteorological Institute (Météo-France). The authors thank the national observation service GLACIOCLIM (CNRS-INSU, OSUG, IRD, INRAE, and IPEV) for the data provided. The authors thank Eric Bazile, Yann Seity, Hugo Merzisen, and Ange Haddjeri for the meaningful discussions that helped to build the study.

Data availability statement. AROME outputs and all AWS data that are not otherwise mentioned below can be requested online (https://donneespubliques.meteofrance.fr/). Training data are available from norahelbig@gmail.com upon request. DEVINE code and model weights are available at https://github.com/louisletoumelin/wind_downscaling_cnn. GLACIOCLIM data are available at https://glacioclim.osug.fr/. Col du Lac Blanc data can be requested online (https://doi.osug.fr/public/CRYOBSCLIM_CLB/).

REFERENCES

Anquetin, S., C. Guilbaud, and J.-P. Chollet, 1998: The formation and destruction of inversion layers within a deep valley. *J. Appl. Meteor. Climatol.*, **37**, 1547–1560, https://doi.org/10.1175/1520-0450(1998)037<1547:TFADOI>2.0.CO;2.

Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1.

Bonavita, M., and Coauthors, 2021: Machine learning for Earth system observation and prediction. *Bull. Amer. Meteor. Soc.*, **102**, E710–E716, https://doi.org/10.1175/BAMS-D-20-0307.1.

Butler, B. W., and Coauthors, 2015: High-resolution observations of the near-surface wind field over an isolated mountain and in a steep river canyon. *Atmos. Chem. Phys.*, **15**, 3785–3801, https://doi.org/10.5194/acp-15-3785-2015.

DeGaetano, A. T., 1997: A quality-control routine for hourly wind observations. *J. Atmos. Oceanic Technol.*, **14**, 308–317, https://doi.org/10.1175/1520-0426(1997)014<0308:AQCRFH>2.0.CO;2.

Dujardin, J., and M. Lehning, 2022: Wind-topo: Downscaling near-surface wind fields to high-resolution topography in highly complex terrain with deep learning. *Quart. J. Roy. Meteor. Soc.*, **148**, 1368–1388, https://doi.org/10.1002/qj.4265.

Fernández, J. G., and S. Mehrkanoon, 2021: Broad-Unet: Multi-scale feature learning for nowcasting tasks. *Neural Network*, **144**, 419–427, https://doi.org/10.1016/j.neunet.2021.08.036.

Forthofer, J. M., B. W. Butler, and N. S. Wagenbrenner, 2014: A comparison of three approaches for simulating fine-scale surface winds in support of wildland fire management. Part I. Model formulation and comparison against measurements. *Int. J. Wildland Fire*, **23**, 969–981, https://doi.org/10.1071/WF12089.

Georgelin, M., E. Richard, M. Petitdidier, and A. Druilhet, 1994: Impact of subgrid-scale orography parameterization on the simulation of orographic flows. *Mon. Wea. Rev.*, **122**, 1509–1522, https://doi.org/10.1175/1520-0493(1994)122<1509:IOSSOP>2.0.CO;2.

Gouttevin, I., V. Vionnet, Y. Seity, A. Boone, M. Lafaysse, Y. Deliot, and H. Merzisen, 2023: To the origin of a wintertime screen-level temperature bias at high altitude in a kilometric NWP model. *J. Hydrometeor.*, **24**, 53–71, https://doi.org/10.1175/JHM-D-21-0200.1.

Guyomarc'h, G., and Coauthors, 2019: A meteorological and blowing snow data set (2000–2016) from a high-elevation alpine site (Col du Lac Blanc, France, 2720 m a.s.l.). *Earth Syst. Sci. Data*, **11**, 57–69, https://doi.org/10.5194/essd-11-57-2019.

Helbig, N., and H. Löwe, 2012: Shortwave radiation parameterization scheme for subgrid topography. *J. Geophys. Res.*, **117**, D03112, https://doi.org/10.1029/2011JD016465.

——, and ——, 2014: Parameterization of the spatially averaged sky view factor in complex topography. *J. Geophys. Res. Atmos.*, **119**, 4616–4625, https://doi.org/10.1002/2013JD020892.

——, R. Mott, A. Van Herwijnen, A. Winstral, and T. Jonas, 2017: Parameterizing surface wind speed over complex topography. *J. Geophys. Res. Atmos.*, **122**, 651–667, https://doi.org/10.1002/2016JD025593.

Höhlein, K., M. Kern, T. Hewson, and R. Westermann, 2020: A comparative study of convolutional neural network models for wind field downscaling. *Meteor. Appl.*, **27**, e1961, https://doi.org/10.1002/met.1961.

Jörg-Hess, S., N. Griessinger, and M. Zappa, 2015: Probabilistic forecasts of snow water equivalent and runoff in mountainous areas. *J. Hydrometeor.*, **16**, 2169–2186, https://doi.org/10.1175/JHM-D-14-0193.1.

Kehler, S., J. Hanesiak, M. Curry, D. Sills, and N. Taylor, 2016: High Resolution Deterministic Prediction System (HRDPS) simulations of Manitoba lake breezes. *Atmos.–Ocean*, **54**, 93–107, https://doi.org/10.1080/07055900.2015.1137857.

Lagerquist, R., J. Q. Stewart, I. Ebert-Uphoff, and C. Kumler, 2021: Using deep learning to nowcast the spatial coverage of convection from *Himawari-8* satellite data. *Mon. Wea. Rev.*, **149**, 3897–3921, https://doi.org/10.1175/MWR-D-21-0096.1.

Lehning, M., 2013: Snow-atmosphere interactions and hydrological consequences. *Adv. Water Resour.*, **55**, 1–3, https://doi.org/10.1016/j.advwatres.2013.02.001.

——, J. Doorschot, N. Raderschall, and P. Bartelt, 2000: Combining snow drift and SNOWPACK models to estimate snow loading in avalanche slopes. *Snow Engineering Recent Advances and Developments*, Routledge, 113–122.

Lewis, H. W., S. D. Mobbs, and M. Lehning, 2008: Observations of cross-ridge flows across steep terrain. *Quart. J. Roy. Meteor. Soc.*, **134**, 801–816, https://doi.org/10.1002/qj.259.

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

Lucio-Eceiza, E. E., J. F. González-Rouco, J. Navarro, and H. Beltrami, 2018a: Quality control of surface wind observations in northeastern North America. Part I: Data management issues. *J. Atmos. Oceanic Technol.*, **35**, 163–182, https://doi.org/10.1175/JTECH-D-16-0204.1.

——, ——, ——, ——, and J. Conte, 2018b: Quality control of surface wind observations in northeastern North America. Part II: Measurement errors. *J. Atmos. Oceanic Technol.*, **35**, 183–205, https://doi.org/10.1175/JTECH-D-16-0205.1.

Lundquist, K. A., F. K. Chow, and J. K. Lundquist, 2012: An immersed boundary method enabling large-eddy simulations of flow over complex terrain in the WRF model. *Mon. Wea. Rev.*, **140**, 3936–3955, https://doi.org/10.1175/MWR-D-11-00311.1.

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

——, V. Vionnet, and T. Grünewald, 2018: The seasonal snow cover dynamics: Review on wind-driven coupling processes. *Front. Earth Sci.*, **6**, 197, https://doi.org/10.3389/feart.2018.00197.

Pickering, B., C. M. Grams, and S. Pfenninger, 2020: Sub-national variability of wind power generation in complex terrain and its correlation with large-scale meteorology. *Environ. Res. Lett.*, **15**, 044025, http://doi.org/10.1088/1748-9326/ab70bd.

Pohl, S., P. Marsh, and G. E. Liston, 2006: Spatial-temporal variability in turbulent fluxes during spring snowmelt. *Arct. Antarct. Alp. Res.*, **38**, 136–146, https://doi.org/10.1657/1523-0430(2006)038[0136:SVITFD]2.0.CO;2.

Quéno, L., V. Vionnet, I. Dombrowski-Etchevers, M. Lafaysse, M. Dumont, and F. Karbou, 2016: Snowpack modelling in the Pyrenees driven by kilometric-resolution meteorological forecasts. *Cryosphere*, **10**, 1571–1589, https://doi.org/10.5194/tc-10-1571-2016.

Raderschall, N., M. Lehning, and C. Schär, 2008: Fine-scale modeling of the boundary layer wind field over steep topography. *Water Resour. Res.*, **44**, W09425, https://doi.org/10.1029/2007WR006544.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Springer, 234–241, https://doi.org/10.1007/978-3-319-24574-4_28.

Schweizer, J., J. B. Jamieson, and M. Schneebeli, 2003: Snow avalanche formation. *Rev. Geophys.*, **41**, 1016, https://doi.org/10.1029/2002RG000123.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, https://doi.org/10.1175/2010MWR3425.1.

Sharples, J. J., R. H. D. McRae, and R. O. Weber, 2010: Wind characteristics over complex terrain with implications for bushfire risk management. *Environ. Modell. Software*, **25**, 1099–1120, https://doi.org/10.1016/j.envsoft.2010.03.016.

Taylor, P. A., and H. W. Teunissen, 1987: The Askervein Hill project: Overview and background data. *Bound.-Layer Meteor.*, **39**, 15–39, https://doi.org/10.1007/BF00121863.

Trebing, K., T. Stanczyk, and S. Mehrkanoon, 2021: SmaAt-Unet: Precipitation nowcasting using a small attention-Unet architecture. *Pattern Recognit. Lett.*, **145**, 178–186, https://doi.org/10.1016/j.patrec.2021.01.036.

Vernay, M., M. Lafaysse, D. Monteiro, P. Hagenmuller, R. Nheili, R. Samacoïts, D. Verfaillie, and S. Morin, 2022: The S2M meteorological and snow cover reanalysis over the French mountainous areas: Description and evaluation (1958–2020). *Earth Syst. Sci. Data*, **14**, 1707–1733, https://doi.org/10.5194/essd-14-1707-2022.

Vionnet, V., E. Martin, V. Masson, G. Guyomarc'h, F. Naaim-Bouvet, A. Prokop, Y. Durand, and C. Lac, 2014: Simulation of wind-induced snow transport and sublimation in alpine terrain using a fully coupled snowpack/atmosphere model. *Cryosphere*, **8**, 395–415, https://doi.org/10.5194/tc-8-395-2014.

——, I. Dombrowski-Etchevers, M. Lafaysse, L. Quéno, Y. Seity, and E. Bazile, 2016: Numerical weather forecasts at kilometer scale in the French Alps: Evaluation and application for snowpack modeling. *J. Hydrometeor.*, **17**, 2591–2614, https://doi.org/10.1175/JHM-D-15-0241.1.

——, F. Naaim-Bouvet, Y. Deliot, H. Bellot, F. Karbou, M. Naaim, G. Guyomarc'h, and Y. Durand, 2017: Col du Lac Blanc: Un site pour l'observation et la modélisation de la neige en haute montagne. *La Meteor.*, **99**, 35–44, https://doi.org/10.4267/2042/63588.

——, V. Fortin, E. Gaborit, G. Roy, M. Abrahamowicz, N. Gasset, and J. W. Pomeroy, 2020: Assessing the factors governing the ability to predict late-spring flooding in cold-region mountain basins. *Hydrol. Earth Syst. Sci.*, **24**, 2141–2165, https://doi.org/10.5194/hess-24-2141-2020.

——, C. B. Marsh, B. Menounos, S. Gascoin, N. E. Wayand, J. Shea, K. Mukherjee, and J. W. Pomeroy, 2021: Multi-scale snowdrift-permitting modelling of mountain snowpack. *Cryosphere*, **15**, 743–769, https://doi.org/10.5194/tc-15-743-2021.

Wagenbrenner, N. S., J. M. Forthofer, B. K. Lamb, K. S. Shannon, and B. W. Butler, 2016: Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja. *Atmos. Chem. Phys.*, **16**, 5229–5241, https://doi.org/10.5194/acp-16-5229-2016.

Wanner, H., and M. Furger, 1990: The Bise—Climatology of a regional wind north of the Alps. *Meteor. Atmos. Phys.*, **43**, 105–115, https://doi.org/10.1007/BF01028113.

Weiss, A., 2001: Topographic position and landforms analysis. *Esri User Conf.*, San Diego, CA, Esri, http://www.jennessent.com/downloads/tpi-poster-tnc_18x22.pdf.

Whiteman, C. D., 2000: *Mountain Meteorology: Fundamentals and Applications*. Oxford University Press, 355 pp.

Winstral, A., K. Elder, and R. E. Davis, 2002: Spatial snow modeling of wind-redistributed snow using terrain-based parameters. *J. Hydrometeor.*, **3**, 524–538, https://doi.org/10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2.

——, T. Jonas, and N. Helbig, 2017: Statistical downscaling of gridded wind speed data using local topography. *J. Hydrometeor.*, **18**, 335–348, https://doi.org/10.1175/JHM-D-16-0054.1.

Xie, Z., and Coauthors, 2021: Decision tree-based detection of blowing snow events in the European Alps. *Hydrol. Earth Syst. Sci.*, **25**, 3783–3804, https://doi.org/10.5194/hess-25-3783-2021.

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part I: Model dynamics and verification. *Meteor. Atmos. Phys.*, **75**, 161–193, https://doi.org/10.1007/s007030070003.

——, and Coauthors, 2001: The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys.*, **76**, 143–165, https://doi.org/10.1007/s007030170027.