Downscaling of surface wind forecasts using convolutional neural networks

Florian Dupuy[1,2], Pierre Durand[2], and Thierry Hedde[1]

[1]CEA, DES, IRESNE, DTN, Laboratory for Environmental Transfer Modeling, Cadarache
[2]Laboratoire d'Aérologie, Université de Toulouse, CNRS, UPS, Toulouse, France

Correspondence: Florian Dupuy (florian.dupuy@aero.obs-mip.fr)

Abstract. Near-surface winds over complex terrain generally feature a large variability at the local scale. Forecasting these winds requires high-resolution Numerical Weather Prediction (NWP) models, which drastically increases the duration of simulations and hinders to run them on a routine basis. Nevertheless, downscaling methods can help forecasting such wind flows at limited numerical cost. In this study, we present a statistical downscaling of WRF wind forecasts over south-eastern France
5 (including the south-western part of the Alps) from its original 9-km resolution onto a 1-km resolution grid (1-km NWP model outputs are used to fit our statistical models). Downscaling is performed using convolutional neural networks (CNNs), which are the most powerful machine learning tool for processing images or any kind of gridded data, as demonstrated by recent studies dealing with wind forecasts downscaling. The previous studies mostly focused on testing new model architectures. In this study, we aimed to extend these works by exploring different output variables and their associated loss function. We found
10 that there is no one approach that outperforms the others on both the direction and the speed at the same time. Finally, the best overall performance is obtained by combining two CNNs, one dedicated to the direction forecast, based on the calculation of the normalized wind components using a customized mean squared error (MSE) loss function, and the other dedicated to the speed forecast, based on the calculation of the wind components and using another customized MSE loss function. Local-scale, topography-related wind features, which were poorly forecast at 9 km, are now well reproduced, both for speed (e.g. acceler-
15 ation on the ridge, leeward deceleration, sheltering in valleys) and direction (deflection, valley channeling). There is a general improvement in the forecast, especially during the nighttime stable stratification period, which is the most difficult period to forecast. It results that, after downscaling, the wind speed bias and MAE are reduced from -0.55 m·s^{-1} and 1.02 m·s^{-1} initially to -0.01 m·s^{-1} and 0.69 m·s^{-1} respectively, while the wind direction MAE is reduced from 25.9^o to 15.5^o in comparison with the 9-km resolution forecast.

1 Introduction

20 Over complex terrain, topography and near surface processes affect low level winds: slope winds resulting from spatial thermal differences along sloping terrain; deviation around hills; channeling in valleys; speedup on mountain crests and acceleration across gaps and passes (Whiteman, 2000). It results that winds generally feature complex structures at the local scale. In consequence, forecasting these winds requires high-resolution (HR) Numerical Weather Prediction (NWP) models in order

25   to represent the complexity of the topography and its local impact on the flow. This can be achieved through dynamical downscaling, that is to say by using an HR NWP model on a limited domain forced by a lower-resolution forecast (Schmidli et al., 2018; de Bode et al., 2021). However, applying such methods over a relatively large domain and for long time periods drastically increases the duration of simulations and hinders the ability to run them on a routine basis.

Others downscaling methods can help forecasting such wind flows at limited numerical cost. TopoSCALE (Fiddes and
30   Gruber, 2014) and WindNinja (Wagenbrenner et al., 2016) are two physically-based downscaling schemes which formulate physical principles to account for the effect of a high-resolution topography on boundary-layer meteorology. Both provide wind forecast downscaling (albeit with some limitations) at a limited computational cost compared to NWP models (Fiddes and Gruber, 2014; Wagenbrenner et al., 2016; Kruyt et al., 2022). The other main downscaling approach is statistical downscaling. Contrary to physically-based models, such methods take advantage of past observations or HR forecasts, bringing local wind
35   information, which can help reproducing the local-scale flow structure. Several methods have been applied to weather forecast downscaling: generalized additive models for wind components (Salameh et al., 2009), random forests for wind speed (Zamo et al., 2016), artificial neural networks (ANNs) for wind components (Dupuy et al., 2021a) (see Vannitsem et al. (2021) for a recent overview).

Over the past decades, ANNs have become one of the most widely used machine learning methods and have transformed
40   many fields (e.g., image recognition, automatic translation, etc.), including science. Convolutional neural networks (CNNs) (LeCun et al., 2015) are a special kind of neural networks designed to extract hierarchical features from grid-like data, making them the state-of-the-art machine learning techniques for complex image processing like image super-resolution which consists in generating an HR image from a low-resolution (LR) image (see Yang et al. (2019) and Kulkarni et al. (2022) for an overview). Therefore, CNNs appear to be a suitable tool to work with geophysical data issued from numerical models in order to extract
45   spatial features, as well as to perform a downscaling. The atmospheric research community has already taken advantage of CNNs ability for diverse applications (see Reichstein et al. (2019) for an overview) including NWP output postprocessing (Vandal et al., 2018; Lagerquist et al., 2019; Dupuy et al., 2021b), including downscaling (see Leinonen et al. (2021) and Harris et al. (2022) for two examples on precipitation forecasts), CNNs outperforming other traditional methods in these studies.

50   However, studies dealing with statistical downscaling of both wind speed and direction forecast using CNNs are rare. Höhlein et al. (2020), Miralles et al. (2022) and Le Toumelin et al. (2023a, b) used a 2D-to-2D architecture producing a downscaled 2D field from LR 2D fields issued from a NWP model. Miralles et al. (2022) used a generative adversarial network (GAN), designed to produce realistic looking fields while Höhlein et al. (2020) and Le Toumelin et al. (2023a, b) both used a classic U-Net architecture, although they applied different training approaches: Höhlein et al. (2020) directly trained their model using
55   HR and LR forecasts while Le Toumelin et al. (2023a, b) first trained their model using HR and LR output from simulations of idealized conditions (controlled atmospheric conditions and idealized topographies), and then applied it to their real world LR forecasts. On the other hand, Dujardin and Lehning (2022) used a 2D-to-point architecture, meaning that their CNN uses 2D fields data to calculate the wind at a single point (the center of the input 2D data). This singular approach derives from the ground truth data they use, which come from weather stations observations, contrary to the 2D-to-2D approach where

60 target data are coming from NWP models. Nevertheless, Dujardin and Lehning (2022) also produced 2D wind fields (on a grid with an horizontal resolution as fine as 50 m) by providing their model with input data centered on different locations. It has to be noted that HR and LR do not refer to the same scales in these studies. Downscaling was performed from 31 km to 9 km (ratio close to 3), from 25 km to 1.1 km (ratio close to 20), from 1.1 km to 50 m (ratio close to 20) and from 1.3 km to 30 m (ratio close to 40) in Höhlein et al. (2020), Miralles et al. (2022) Dujardin and Lehning (2022) and Le Toumelin et al.

65 (2023a, b), respectively. Moreover, Le Toumelin et al. (2023b) and Miralles et al. (2022) only used topographical information as additional predictors to the LR wind forecast while Höhlein et al. (2020) and Dujardin and Lehning (2022) used others meteorological parameters. Thus, meteorological phenomena that are expected to be reproduced should differ. For instance, Miralles et al. (2022), Dujardin and Lehning (2022) and Le Toumelin et al. (2023b) report improvements in the representation of main orographic effects that are not resolved in the larger-scale data, like acceleration on the ridge and sheltering effect.

70 Moreover, Dujardin and Lehning (2022) noted a more realistic wind deflection while Le Toumelin et al. (2023b), as well as Miralles et al. (2022) over the Alps, got only a small impact on the direction from their downscaling.

This study aims to pursue the exploration of new strategies of wind forecast downscaling in line with the works introduced above. We present a 2D-to-2D statistical downscaling approach with the originality that wind variables are calculated by different ways, which generates different wind forecasts. We apply this strategy to WRF wind forecasts over south-eastern

75 France (including the south-western part of the Alps) from their original 9-km horizontal resolution to a 1-km resolution grid (section 2). We evaluate the performances of the different high-resolution forecasts and analyze their own advantages and disadvantages (section 3). Then we present the main conclusions of the study.

2 Methods

2.1 WRF forecasts

80 The WRF NWP model (Skamarock et al., 2019) was run in a grid-nested mode, with 3 nested domains (Fig. 1): D1 with a 9-km horizontal resolution (152×155 grid points, i.e. 1368×1395 km), covering France and its surroundings (especially to the south and to the east); D2 with a 3-km resolution (99×99 grid points, i.e. 297×297 km); and D3 with a 1-km resolution (99×99 grid points, i.e. 99×99 km), centered on southeastern France. WRF was routinely run in this operational forecast mode from the 24 December 2020 to the 5 May 2022, for lead times up to 72 hours. We only used simulations for lead times

85 from 12 to 72 hours (the first 12 hours are considered as the spin-up and are then discarded) resulting in a total of 29036 hourly outputs (after removing some dates for which D3 data are missing). More details on the WRF set up can be found in de Bode et al. (2023).
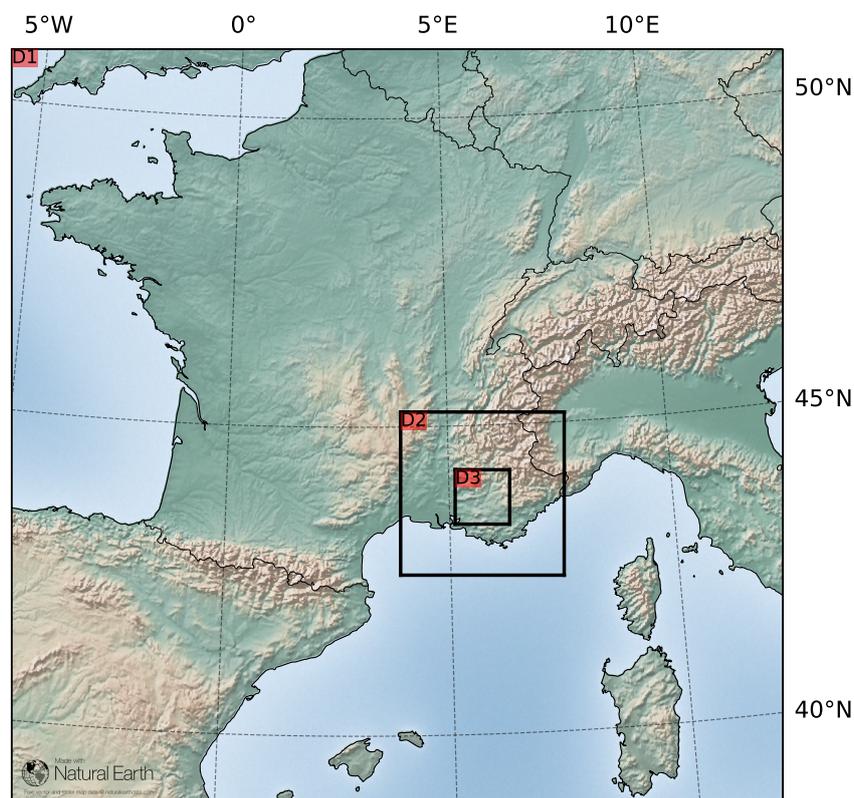
Figure 1. Representation of the 3 nested domains of the WRF model.

2.2 Training data

The objective is to downscale the 9-km resolution WRF forecasts (called WRF LR, for low resolution, in the following) towards
a 1-km resolution over an area corresponding to the D3 domain. The 1-km WRF forecast (called WRF HR, for high resolution,
in the following) is considered as the target used to train the statistical models.

Wind variables as well as many other variables from the WRF LR forecasts are used as predictors for the CNNs: wind
components (u for the eastward component, v for the northward component), wind direction and wind speed; basic meteoro-
logical parameters (2-m temperature and potential temperature, surface pressure and precipitation); short-wave and long-wave
radiation fluxes; stability-related variables (boundary-layer height and friction velocity).

Moreover, in order to incorporate the geographical context, some HR parameters, called "static" since they do not vary in
the time, such as topography and a land/sea mask are added to the list of predictors. Moreover, following the recommendations
of Dujardin and Lehning (2022), slope and aspect (orientation of the slope) HR fields calculated from the HR topography are
added (see Fig. 2).
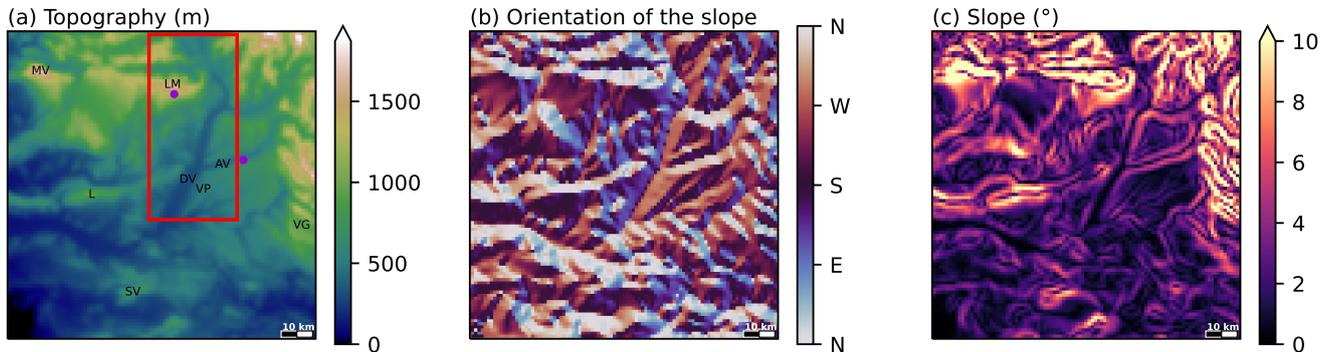
Figure 2. Illustration of static predictors on the D3 domain: (a) topography (in m a.s.l.), (b) orientation of the slope and (c) local slope. The red rectangle on (a) shows the area of the plots on Figs. 6, 7 and 8. Letters refer to some topographical sites: MV for Mont Ventoux, LM for Lure mountain, L for Luberon, DV for Durance valley, VP for Valensole plateau, AV for Asse valley, VG for Verdon Gorge and SV for Sainte-Victoire mountain. The two purple dots indicate the locations of the valley site and crest site depicted on Figs. 9 and 10 respectively.

100   Finally, we added the new predictors introduced by Dujardin and Lehning (2022), which combine wind and topography information, giving insight on wind–topography interactions. More specifically, they calculate a theoretical correction of the wind components in order to represent the speed modification caused by the exposure/sheltering to the wind, as well as the deflection caused by the relief. The reader can refer to Dujardin and Lehning (2022) for more details.

2.3   Convolutional Neural Network architecture

105   The objective of a neural network (NN) is to find a mathematical function linking a list of predictors to a list of predictants (considered as the truth). The function is composed of neurons (a linear combination of input variables transformed by a so-called activation function) interconnected between each other and arranged in layers. Training the NN consists in fitting its function to produce results as close as possible to the truth (the reader can refer to Goodfellow et al. (2016) for more explanations).

110   Convolutional layers can be introduced in NN when dealing with grid-like data in order to take advantage of information contained in spatial structures. In these layers, neurons actually correspond to a convolution function which is applied to a limited part of the grid. NNs using such layers are called convolutional neural networks (CNNs).

In this study, we used a U-Net architecture (Ronneberger et al., 2015), which is a fully convolutional network that generates images from images. Its name comes from its U-shaped architecture in which convolutional layers are separated first with
115   pooling layers and then with transposed convolutional layers. The first phase, with pooling layers, reduces the size of images, which is known to capture the context of input images. The second phase, with transposed convolutional layers, increases the size of the contracted images, enabling a precise localization.

The architecture of the CNN used is described in Fig. 3. Before entering the U-Net, the data follow a two steps process: LR input data (32×32 data with a resolution of 9 km centered over the D3 domain) are interpolated on the HR grid (288×288 data with a resolution of 1 km, which is larger than the D3 domain in order to incorporate information on a larger spatial scale) using a bicubic interpolation (the red layer in Fig. 3) in order to reduce grid patterns artefacts in the outputs, then all the predictors are standardized. We used a padding of 1 in order to produce outputs with the same size as inputs although we cropped the 288×288 outputs since we only focus on the $99 \, \text{km} \times 99 \, \text{km}$ central area corresponding to the D3 domain. In order to avoid overfitting, we added a batch normalization (Ioffe and Szegedy, 2015) and a drop out (Srivastava et al., 2014) after convolutional layers and we introduced an early stopping that stopped the learning when the loss function calculated on an independent validation dataset did not improve on 10 successive epochs. The ReLU activation function is applied after each convolutional layer, except for the final 1×1 convolutional layer in order to produce a not bounded regression. The mean squared error (MSE) is used as loss function. Additional modifications tested are described in the next sections. We used the PyTorch library of Python for the machine learning developments.
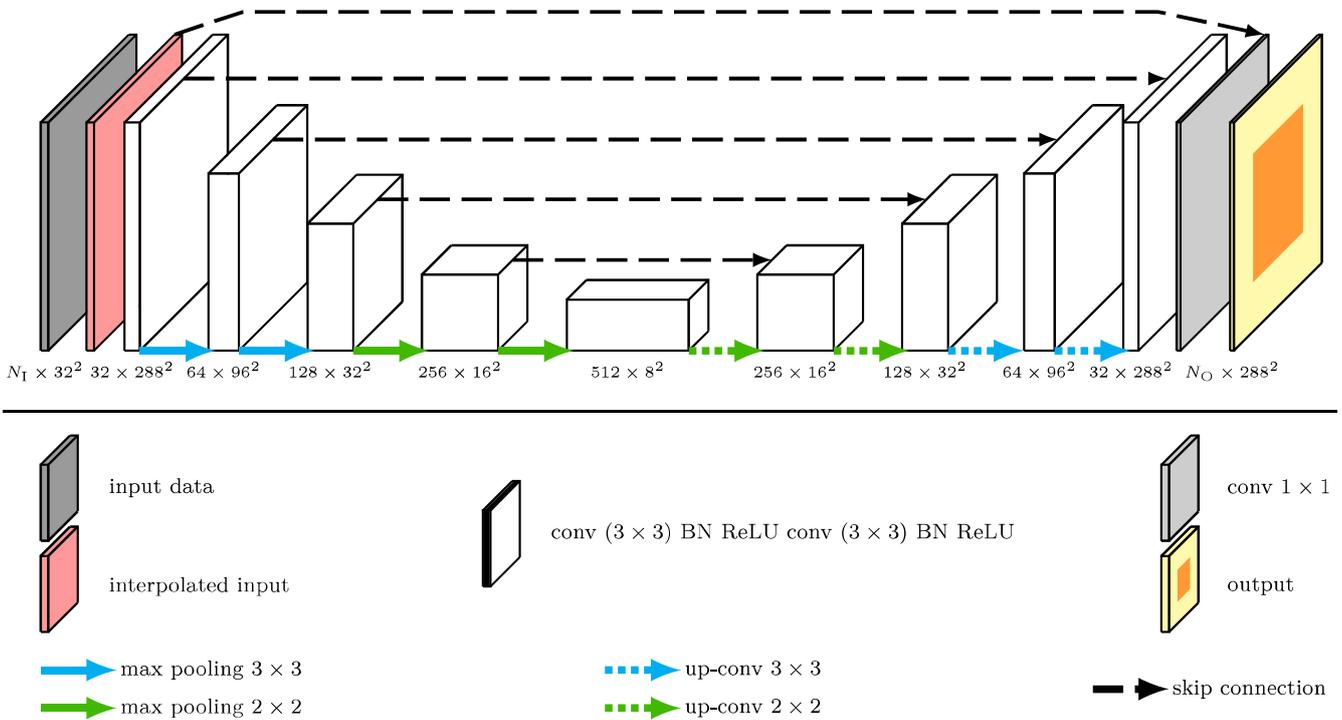
Figure 3. Schematic illustrating the architecture of the CNN used in the study. BN stands for batch normalization. The numbers under the different computation blocs indicate the dimension of the data at different stages of the network at the output of the corresponding bloc. N_I and N_O represent the number of input variables and the number of target variables, respectively. On the output, the orange area represents the crop (99×99) from the yellow part (288×288).

2.3.1 Choice of target variables

We are interested in calculating both wind speed and direction. However, the direction may be difficult to calculate directly because of its cyclic nature. Circular regression tools are generally based on the von Mises distribution (also called circular normal distribution), but they are challenging to optimize, even if Lang et al. (2020) developed a circular regression tool based on random forests (Breiman, 2001) which simplifies the optimization process. On the other hand, more classic regression approaches, for instance based on the estimation of the conditional mean (via the minimization of the mean squared error), seem inappropriate since there is no definition of the mean direction calculated directly on a set of direction values. Despite this, Le Toumelin et al. (2023a) performed a regression calculating directly the direction, which improved their direction forecast, using a cosine distance as loss function (see section 2.3.2). We tested their approach (model called CNN_{dir} hereafter).

Otherwise, the mean direction can be calculated based on its sine and cosine (or wind components in the case of wind data) values (Jammalamadaka and SenGupta, 2001). That is why it is more common to calculate the wind components when using regression tools (Dupuy et al., 2019, 2021a; Höhlein et al., 2020; Miralles et al., 2022; Dujardin and Lehning, 2022) since they carry information on both speed and direction. We apply this approach in this study. Therefore, the CNN outputs 2 variables aiming at representing the u and v wind components (this model is called $CNN_{u,v}$ hereafter).

Nevertheless, for a given error on one or the other of the 2 components, the forecast error on the underlying direction varies in function of the speed (the lower the speed, the higher the error on the direction) and direction. It results that direction errors on lighter winds are artificially less penalized than direction errors on larger winds. Knowing that light winds directions are difficult to forecast (with a deterministic model) because of their higher spatial and temporal heterogeneity, this strategy reinforces the difficulty. The normalization of the components by the wind speed, which gives the cosine and sine values of the direction (noted $\tilde{u}$ and $\tilde{v}$), is a way to equally penalize all wind speed conditions. We thus tried to forecast these variables (this model is called $CNN_{\tilde{u},\tilde{v}}$) although they do not incorporate any information on the speed which has to be computed in some other way.

2.3.2 Loss function

In our case, the MSE loss produces a negative biased speed forecast. Dujardin and Lehning (2022) proposed a loss function, inspired by the Pinball function which is used to make quantile regressions, in order to produce unbiased wind speed predictions when they derive from the components (Eq. 1).

$$\mathcal{L}_{spd} = \frac{1}{N} \sum_{i=1}^{N} \tau_i \left[(\hat{u}_i - \beta_i u_i)^2 + (\hat{v}_i - \beta_i v_i)^2 \right], \tag{1}$$

$$\beta_i = \frac{\epsilon + \|U\|_i}{\epsilon + \|\hat{U}\|_i}, \quad \tau_i = \begin{cases} \tau & \text{if } \|\hat{U}\|_i \geq \|U\|_i, \\ 1-\tau & \text{if } \|\hat{U}\|_i < \|U\|_i. \end{cases}$$

with $\hat{u}$ and $\hat{v}$ the CNN outputs wind components, $\|U\|$ and $\|\hat{U}\|$ the target and output speed and τ and ϵ two constants. Using values of $\tau = 0.3$ and $\epsilon = 4.3$ m·s^{-1} (these values were chosen after numerous tests), we obtained an unbiased forecast (model called CNN$_{u,v,\mathcal{L}_{\mathrm{spd}}}$).

160   In order to get consistent couples of cosine and sine values when calculating the normalized components, that is to say $\tilde{u}^2 + \tilde{v}^2 = 1$, we tested a loss function combining the classic MSE and the absolute distance between unity and the sum of squared normalized components (Eq. 2):

$$\mathcal{L}_2 = \frac{1}{N} \sum_{i=1}^{N} \left[\left(\hat{\tilde{u}}_i - \hat{u}_i \right)^2 + \left(\hat{\tilde{v}}_i - \hat{v}_i \right)^2 + \alpha \left[1 - \left(\hat{\tilde{u}}_i^2 + \hat{\tilde{v}}_i^2 \right) \right] \right], \tag{2}$$

with α a weight to balance the two penalty terms. In our case, $\alpha = 0.2$ is the optimal value. For larger values the cosine/sine
165   values couples were globally more consistent, but at the expense of the direction forecast.

Finally, as described in the previous section, the CNN$_{\mathrm{dir}}$ is trained using its own loss function (Eq. 3), introduced in the study of Le Toumelin et al. (2023a):

$$\mathcal{L}_{\mathrm{dir}} = 1 - \frac{1}{N} \sum_{i=1}^{N} \left[\cos \left(\hat{\mathrm{dir}}_i - \mathrm{dir}_i \right) \right], \tag{3}$$

with dir and $\hat{\mathrm{dir}}$ the target and CNN output directions respectively. A summary of the tests performed in this study is given in
170   Table 1.

Table 1. Summary of models used for downscaling tests.

Name	Output variable(s)	Loss function
CNN$_{u,v}$	u and v	MSE
CNN$_{u,v,\mathcal{L}_{\mathrm{spd}}}$	u and v	$\mathcal{L}_{\mathrm{spd}}$
CNN$_{\tilde{u},\tilde{v}}$	$\tilde{u}$ and $\tilde{v}$	MSE
CNN$_{\tilde{u},\tilde{v},\mathcal{L}_2}$	$\tilde{u}$ and $\tilde{v}$	$\mathcal{L}_2$
CNN$_{\mathrm{dir}}$	direction	$\mathcal{L}_{\mathrm{dir}}$

2.4 Wind forecast evaluation

The performance was evaluated by comparing the difference between the target (deterministic 1-km WRF simulation) and the output of the various CNNs. The improvement brought by the CNNs translates as a reduction of the wind field error with respect to the initial 9-km WRF wind field error. The latter is computed on the difference between the deterministic 1-km

175 WRF simulation and the 9-km WRF wind field projected onto the 1-km grid (on the D3 domain) with a bicubic interpolation. Therefore, in the following, "WRF LR" stands for the interpolated field of the 9-km WRF forecast.

To evaluate the significance of the results, and considering the relatively small size of our dataset, we performed a k-fold cross validation in order to use as much data as possible during the training while evaluating the models on a large period. Then, we bootstrapped the test subset, yielding a distribution for each metric, in order to evaluate their dispersion (Wilks, 2011).

180 We mostly focused on evaluating the wind speed and direction using classic metrics such as the mean absolute error (MAE) and the mean bias error (MBE). Moreover, we compared the accuracy of the wind speed distribution using the earth mover's distance, also known as the Wasserstein distance (WD). For the direction distribution, we used a modified version called circular earth mover's distance suited to circular variables (Rabin et al., 2008). Finally, we evaluate the wind field spatial heterogeneity by calculating the standard deviation of the speed and direction fields for each of the 29036 map samples (one

185 value per sample). For the direction, which is a circular variable, we used the Yamartino (1984) method (Eq. 4):

$$\sigma_{\mathrm{dir}} = \arcsin{(\epsilon)} \left[1 + \left(\frac{2}{\sqrt{3}} - 1 \right) \epsilon^3 \right], \tag{4}$$

with

$$\epsilon = \sqrt{1 - (s_a^2 + c_a^2)}; \quad s_a = \frac{1}{N} \sum_{i=1}^{N} \sin(\mathrm{dir})_i; \quad c_a = \frac{1}{N} \sum_{i=1}^{N} \cos(\mathrm{dir})_i$$

2.5 Computational considerations

It takes only a few seconds to process a lead time of one simulation using one of the CNNs listed in Table 1 on a graphics processing unit (GPU), which is valuable considering that operational LR forecasts are available several hours ahead whereas

190 a dynamical downscaling performed with WRF would require hours of computation.

3 Results and discussion

3.1 Overall performance of the models

A summary of performance for all the downscaling models is given in Fig. 4. All the CNNs reduce the MAE of the direction and speed, as well as the wind speed bias compared to WRF LR forecasts. However, the bias on the wind components, which is

195 close to zero in WRF LR, is slightly degraded after the downscaling but remains low for all the CNNs (in the range $[-0.05; 0.05]$ m·s^{-1}).

Concerning the wind speed, the $\mathrm{CNN}_{u,v,\mathcal{L}_{\mathrm{spd}}}$ is better than the other CNNs, especially for the correction of the negative bias which is reduced to -0.01 m·s^{-1}, demonstrating the ability of the $\mathcal{L}_{\mathrm{spd}}$ loss function to reduce the bias. However, it has one of the worst performance among all the CNNs on the direction and the components (largest MAE values), meaning that the speed

200 improvement occurs at the expense of a degradation of the forecast of the couple of components.
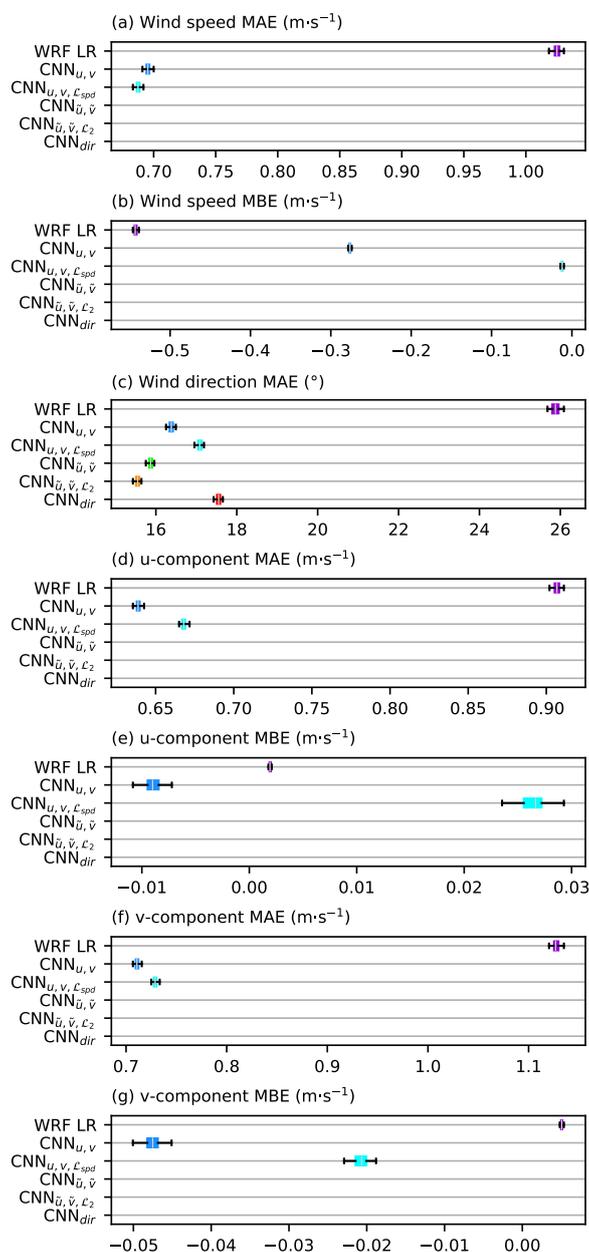
Figure 4. Summary of performance for the different models – (a) wind speed MAE, (b) wind speed MBE, (c) wind direction MAE, (d) u MAE, (e) u MBE, (f) v MAE and (g) v MBE.

Concerning the direction, all the CNNs greatly improve the performance with respect to the LR forecast, with a reduction in MAE of the order of 10 degrees. The $CNN_{\tilde{u},\tilde{v},\mathcal{L}_2}$ achieves the best performance. We detail the performance of the different direction forecasts in Fig. 5. On Fig. 5a, we compare the MAEs on the direction according to the wind direction (as computed

by WRF HR). The point of this figure is not to focus on the evolution of the MAE in function of the direction for a given model

205   since many factors should be considered (for instance, the MAE is lower for the northwesterly and southeasterly winds because of the higher occurrence of high speeds for which the direction is easier to forecast). Instead, we analyze the differences, for a given direction, between the different models. The $CNN_{\tilde{u},\tilde{v},\mathcal{L}_2}$ and $CNN_{\tilde{u},\tilde{v}}$ are really close, while the $CNN_{u,v}$ reaches slightly higher values for southerly winds for unidentified reasons. But above all, the CNN_{dir} behaves singularly, with larger MAE values for winds coming from the NW and NE quadrants while they are close to the other CNNs for the SW and SE quadrants.

210   This is related to an under-prediction of northerly winds, as illustrated in Fig. 5b, which could result from an artifact around the 0–360° numerical discontinuity in the wind direction. It seems that Le Toumelin et al. (2023a, their Fig. 4e) experienced the same issue (important under-estimation of the occurrence of northerly winds), possibly with the same consequence on the MAE on wind direction, which confirms that calculating the direction as a direct output is not appropriate, as already explained in section 2.3.1. Besides, calculating $\tilde{u}$ and $\tilde{v}$ without any constraint on the couple they form leads to important inconsistencies

215   ($\hat{\tilde{u}}^2 + \hat{\tilde{v}}^2$ should be equal to 1 but is underestimated most of the time, see Fig. 5c), which are partly corrected when using the $\mathcal{L}_2$ loss (with more values close to 1), together with a slight improvement of the direction forecast. Finally, the geographical difference of MAE between the $CNN_{\tilde{u},\tilde{v},\mathcal{L}_2}$ and the $CNN_{u,v}$ (Fig. 5d) indicates that the improvement is not homogeneously spread over the domain, but mainly occurs over regions featuring lighter winds in average, generally corresponding to valleys. This behavior was expected since the calculation of the normalized components artificially increases the weight of the direction

220   errors on the lighter winds with respect to the calculation of u and v, as explained in section 2.3.1.
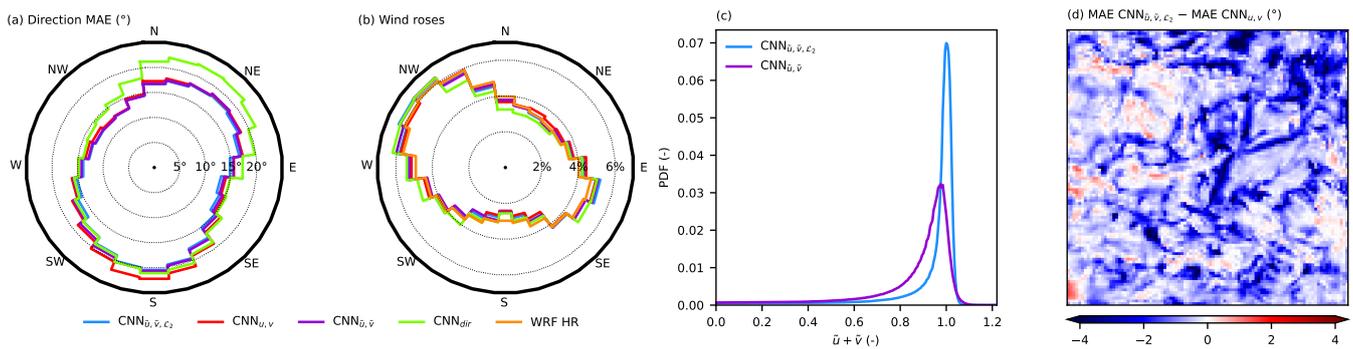
Figure 5. (a) MAE on wind direction according to the wind direction from the HR dataset. (b) Polar distributions of the directions computed by the different models. On (a) and (b), the results of $CNN_{u,v,\mathcal{L}_{spd}}$ are not represented in order to make the figures more readable, and because its performance on the direction is lower. (c) Distribution of $\hat{\tilde{u}}^2 + \hat{\tilde{v}}^2$ values for $CNN_{\tilde{u},\tilde{v}}$ and $CNN_{\tilde{u},\tilde{v},\mathcal{L}_2}$. (d) Difference of MAE on the wind direction between $CNN_{\tilde{u},\tilde{v},\mathcal{L}_2}$ and $CNN_{u,v}$ (negative values means that $CNN_{\tilde{u},\tilde{v},\mathcal{L}_2}$ performs better while positive values means that $CNN_{u,v}$ performs better).

In conclusion, there is no one CNN that outperforms the others. Compared to the classic approach ($CNN_{u,v}$), the modifications improved either the direction or the speed, but not at the same time. Finally, the best overall performance can be obtained by the combination of $CNN_{u,v,\mathcal{L}_{spd}}$ for the calculation of the speed and $CNN_{\tilde{u},\tilde{v},\mathcal{L}_2}$ for the direction. In the following, "CNN" results refer to such a combination of post-processing results.

3.2 Wind field characteristics

Before evaluating the downscaling overall performance, we illustrate the forecast improvement on two selected cases over a limited area of the D3 domain which is interesting since it encompasses important ridges (e.g. Lure mountain) and steep local valleys on its northern part and a smoother topography on its southern part, with the wide Durance valley crossing the area from the north towards the south (see the location of the area in Fig. 2a and its topographical characteristics in Fig. 6). The results for the two cases are presented in Figs. 7 and 8, respectively. These figures are composed of 6 panels split into two rows. There is a classical representation of wind forecasts on the top row (WRF LR (a), WRF HR (b) and CNN (c)) in which the wind is represented with arrows. The bottom row shows the forecast errors of WRF LR (d) and the CNN (f), as well as the correction of the CNN with respect to WRF LR (e): the longer the arrows, the larger the error/difference, while the color of arrows represents the speed error/modification. Therefore, on panels d and f (resp. e), small white arrows indicate a good forecast (resp. a little modified forecast), whereas long white arrows indicate important errors (resp. important modifications) on direction associated to low-speed errors (resp. small modifications). Long red/blue arrows on panels d and f indicate over/under-estimated speeds associated to, either good direction forecast if the orientation of arrows in the WRF LR (panel a) or CNN (panel c) forecast is the same (modulo 180°) as in the error of forecast (panels d or f), or bad direction forecast otherwise. The same logic applies to panel (e) for the correction by the CNN.
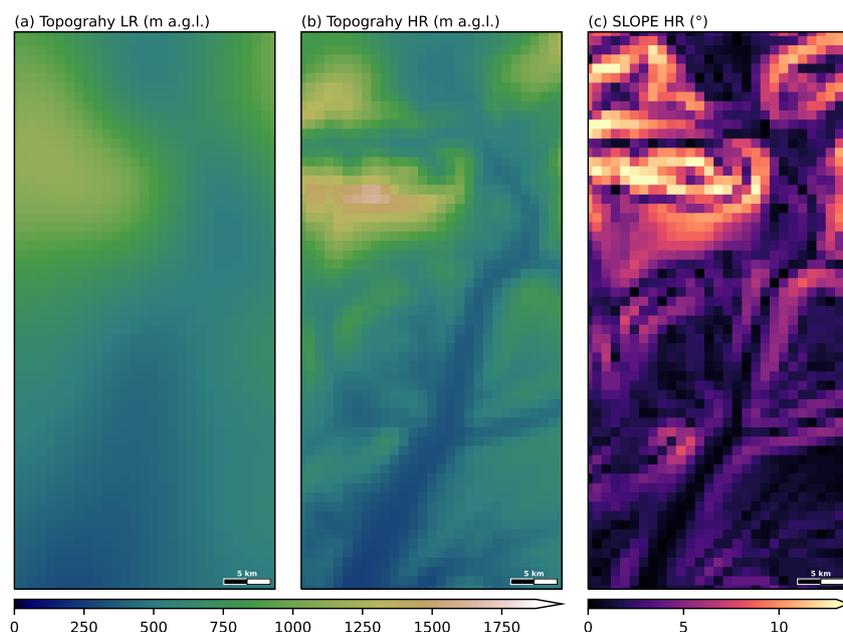
Figure 6. (a) LR topography after a bicubic interpolation towards the 1-km grid, (b) HR topography and (c) HR slope for the area of Figs. 7 and 8 (see the location of the area in Fig 2a).
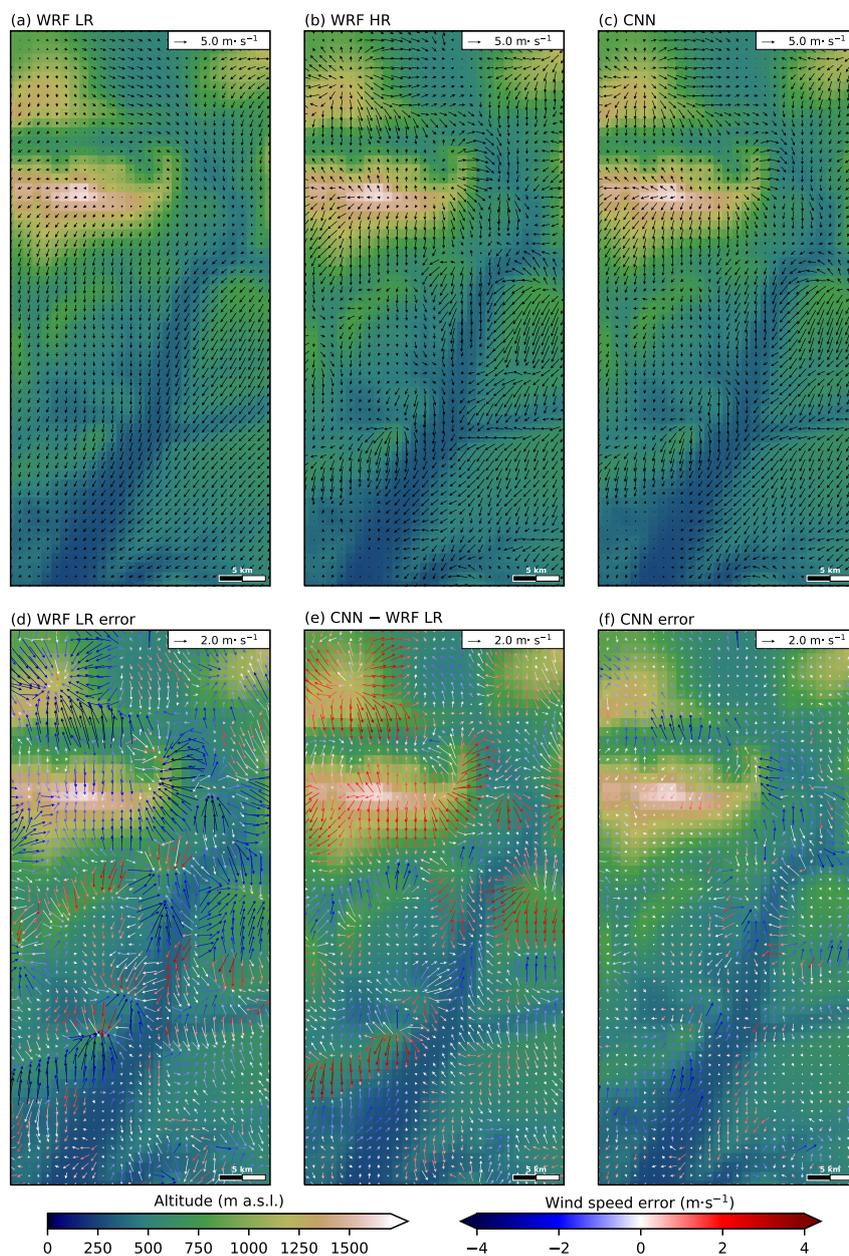
Figure 7. Situation of the 6 January 2021 at 7h UTC (simulation starting at 0h UTC on 5 January 2021, lead time 31h). The area corresponds to the red rectangle in Fig. 2a. Horizontal wind computed by (a) WRF LR, (b) WRF HR and (c) the CNN. Panels (d) and (f) represent the errors corresponding to the wind fields represented in (a) and (c), respectively. Panel (e) represents the vector difference between panels (c) and (a). On panels (d), (e) and (f), the color of the arrows indicates the error/difference in wind speed.

240   On Fig. 7, the meteorological situation corresponds to a calm night with a weak synoptic forcing. The HR forecast wind field (Fig. 7b) has a high spatial heterogeneity, for both speed and direction, resulting from typical topography-dependent features under stable stratification such as channeled valley winds and down-slope winds, e.g. in the northern part of the domain, as well as around less steep hills, e.g. in the southern part of the domain. The LR forecast (Fig. 7a) also features valley and down-slope winds but in relation to its LR topography (Fig. 6a), that is to say a single wide and flared valley, bending around the only

245   main ridge resolved on the northwest part of the area. This results in important errors of direction and speed around all the unresolved topographical features (Fig. 7d). The CNN output is very different from the LR wind field over the whole area (long arrows on Fig. 7e), with important speed and/or direction modification. The downscaled wind field (Fig. 7c) appears to be very close to the HR field, with a good representation of all the topography-dependent features, which is confirmed by the errors which are generally very low over the whole area (Fig. 7f).

250   The meteorological situation presented in Fig. 8 corresponds to a Mistral (regional wind) event, characterized by northwesterly winds at the synoptic scale over the area considered. Over the southern part of the domain, the HR wind field (Fig. 8b) shows a moderate variability with slight deflections and variations in speed around small scale reliefs (acceleration over ridges and leeward deceleration). The LR forecast (Fig. 8a) does not represent these topographical features since local topography vanishes at the 9-km resolution. The wind field is therefore very homogeneous, with small direction changes hence an over-

255   estimation of leeward speeds and an under-estimation windward (Fig. 8d). The CNN only marginally corrects the LR forecast over this area (small arrows in Fig. 8e). The deflection around reliefs (Fig. 8c) as well as the leeward deceleration and ridge acceleration (red arrows windward and blue arrows leeward on Fig. 8e) are only partially represented, resulting in significant errors remaining after downscaling (Fig. 8f).

Over the northern part of the domain, the HR wind field (Fig. 8b) is much more impacted by the topography (which features

260   deeper valleys and hills). It results a flow channeling in the narrow valleys, deflections around main hills and a large acceleration over the main ridges. The LR wind field (Fig. 8a) presents features constrained by the highest ridge (Lure mountain), which is still represented even at the 9-km resolution, with a notable deflection around it as well as a deceleration on the leeward side. However, deflection and acceleration are lower than those in the HR field, which translates as important errors on the speed over the relief (Fig. 8d). Moreover, topographic effects by small hills and valleys are not represented, which generates large

265   direction and speed errors (Fig. 8d). The correction brought by the CNN is more important over this area (long arrows on Fig. 8e) than over the southern area, allowing to correctly represent the topographical features present in the HR wind field.

To summarize, the CNN learned from the HR WRF dataset the wind-topography interaction features occurring under weak synoptic forcing as well as under strong wind conditions even if, for the latter, the corrections over local and thin topographical elements are moderate. In the following section, we analyze whether these improvements can be generalized to the whole

270   period.

3.3   Wind climatology at specific sites

We begin this wind climatology with the description of results on two sites featuring very different characteristics, a valley site (Fig. 9) and a crest site (Fig. 10). The location of these sites is indicated in Fig. 2a.
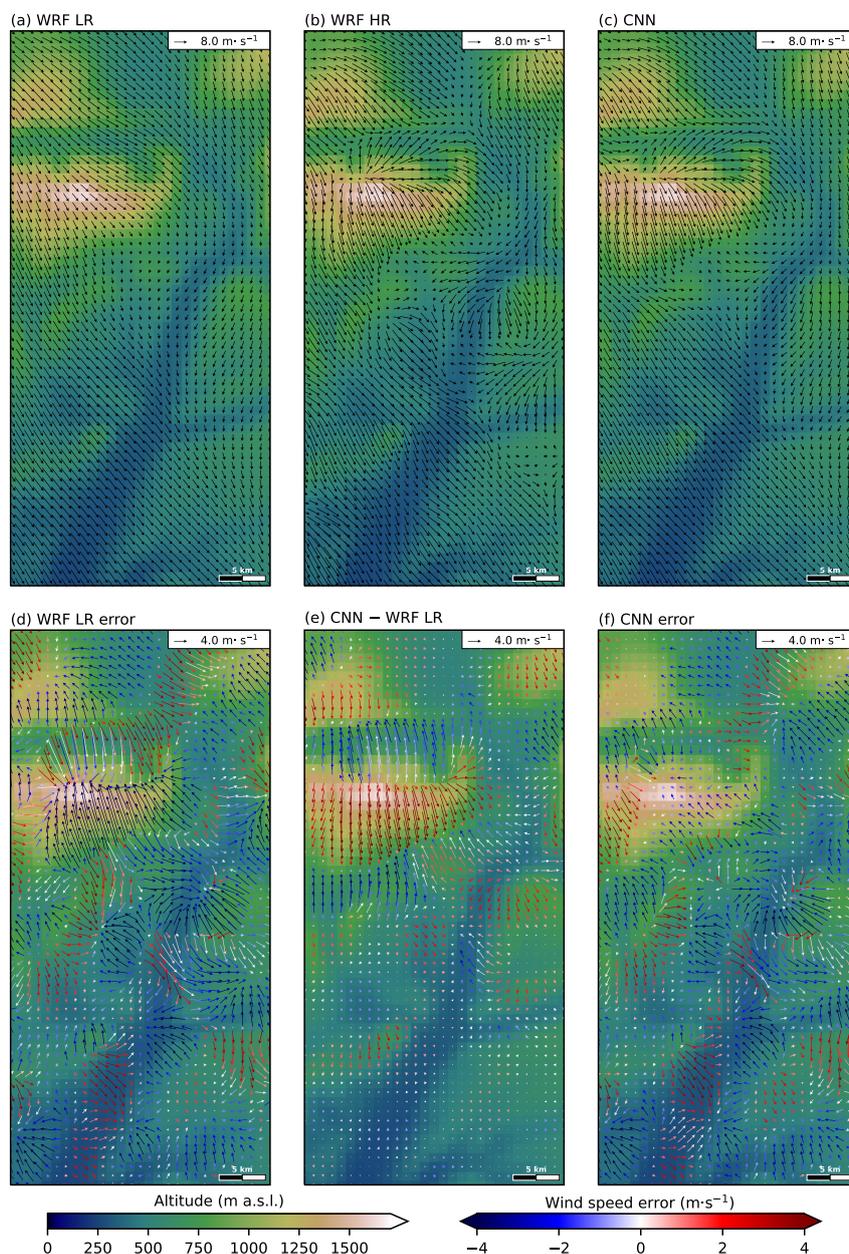
Figure 8. Same as Fig. 7 (only the scales of the arrows differ) for the situation of the 13 January 2021 at 13h UTC (simulation starting at 0h UTC on the 11 January 2021, lead time 61h).

The valley site features two different kinds of winds (Fig. 9e). Firstly, there are winds greater than $6\,\mathrm{m\cdot s^{-1}}$ which are mainly oriented northwesterly and southeasterly, corresponding to Mistral events and cloudy/rainy weather, respectively, the directions of which are little dependent on the local topography (important large scale forcing). These winds are well reproduced in the
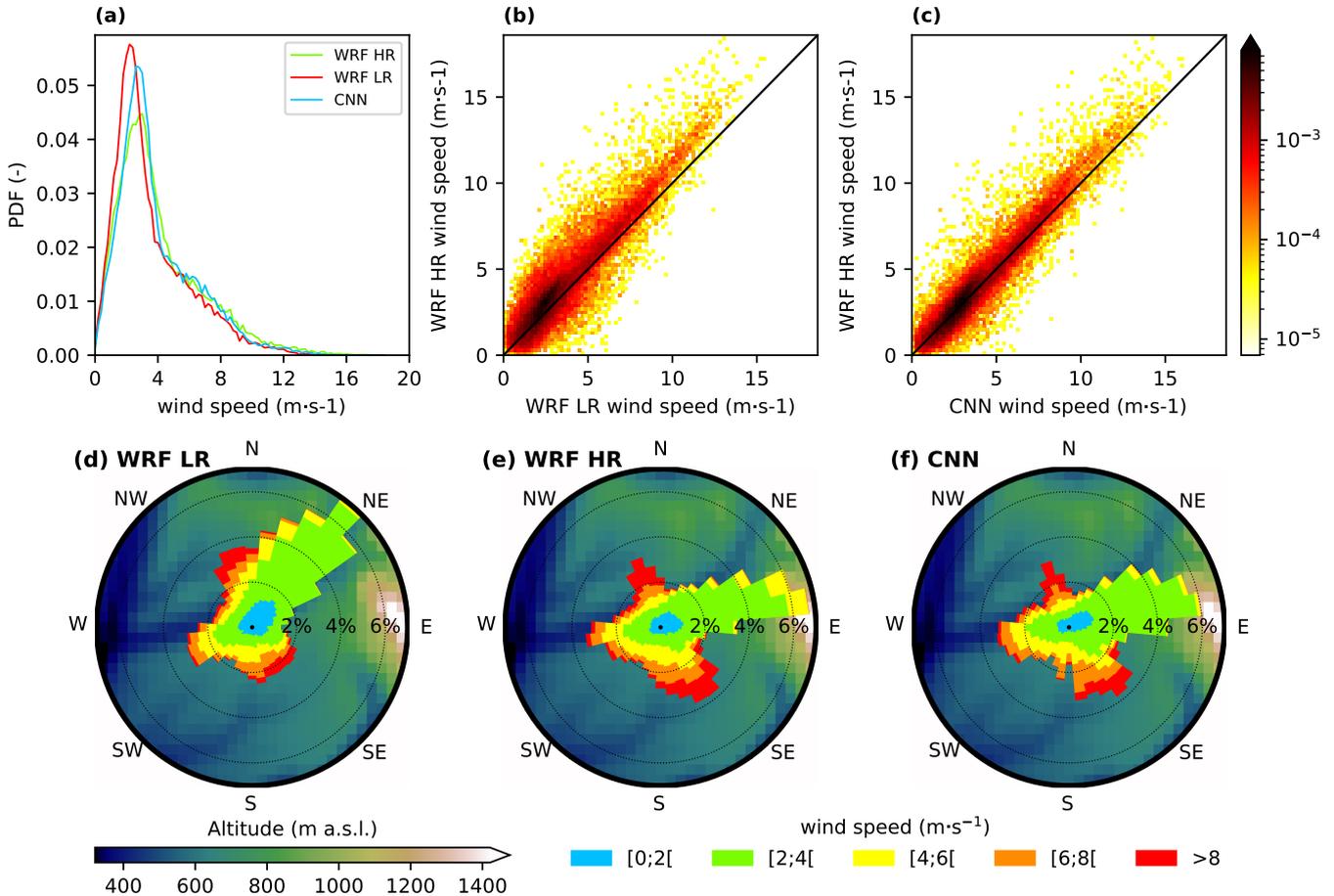
Figure 9. Comparison of wind climatology between WRF HR, WRF LR and the CNN on a single grid cell in a valley (cf. purple dot on Fig. 2a for the exact location). (a) Comparison of wind speed probability density functions. (b) Density scatter plot comparing wind speeds from the HR and LR forecasts. (c) Same as (b) for the comparison between the HR forecast and the CNN. (d), (e) and (f) Wind roses of the WRF LR, the WRF HR and the CNN, respectively. The colors on the wind roses represent the wind speed, and the background in the disks represents the topography within a radius of 15 km around the valley site.

CNN (Fig. 9f), and reasonably well in WRF LR (Fig. 9d), although the northwesterly winds are more dispersed. Moreover, there is a negative bias on speed associated to these winds in WRF LR (Fig. 9b) which is only marginally corrected by the CNN (Fig. 9c). Secondly, there are west-southwesterly and east-northeasterly winds corresponding to up-valley and down-valley winds, respectively, which are highly dependent on the local topography. Both WRF LR and the CNN correctly predict the up-valley winds. However, WRF LR fails to predict the down-valley winds, which are anti-clockwise rotated of approximately 45º due to the south-westerly slope of the LR topography at this place (consistent with explanations of the Fig. 7a), contrary to the CNN which correctly represents them. Therefore, those winds correspond to down-slope winds in both models. Finally, there is an over-representation of winds lower than $3 \, \mathrm{m \cdot s^{-1}}$ in WRF LR which is only marginally corrected by the CNN (Fig. 9a).

280

285   The wind at the crest site features two major orientations, slightly dispersed around the northerly-to-north-easterly and southerly directions (Fig. 10e), mainly resulting from large scale flows forcing. The WRF LR wind rose also presents a bi-modal distribution, but slightly anti-clockwise rotated, and with a higher scatter for the southeasterly winds (Fig. 10d). The comparison of the probability density functions of WRF LR and WRF HR (Fig. 10a) highlights a large disagreement, with an over-estimation (resp. under-estimation) of the occurrence of lighter (resp. stronger) winds in WRF LR, which is generalized

290   to all the directions according to Figs. 10d and 10e. This is consistent with the results depicted in Fig. 8 (under-estimation of the ridge acceleration effect). The CNN corrects both the speed and direction (Figs. 10a, c and f). The probability density functions of the CNN and WRF HR are very close, even for strong speeds (whereas WRF LR is unable to predict any wind speed over 15 m·s^{-1} at this site), and the corresponding wind roses look very similar for all directions.
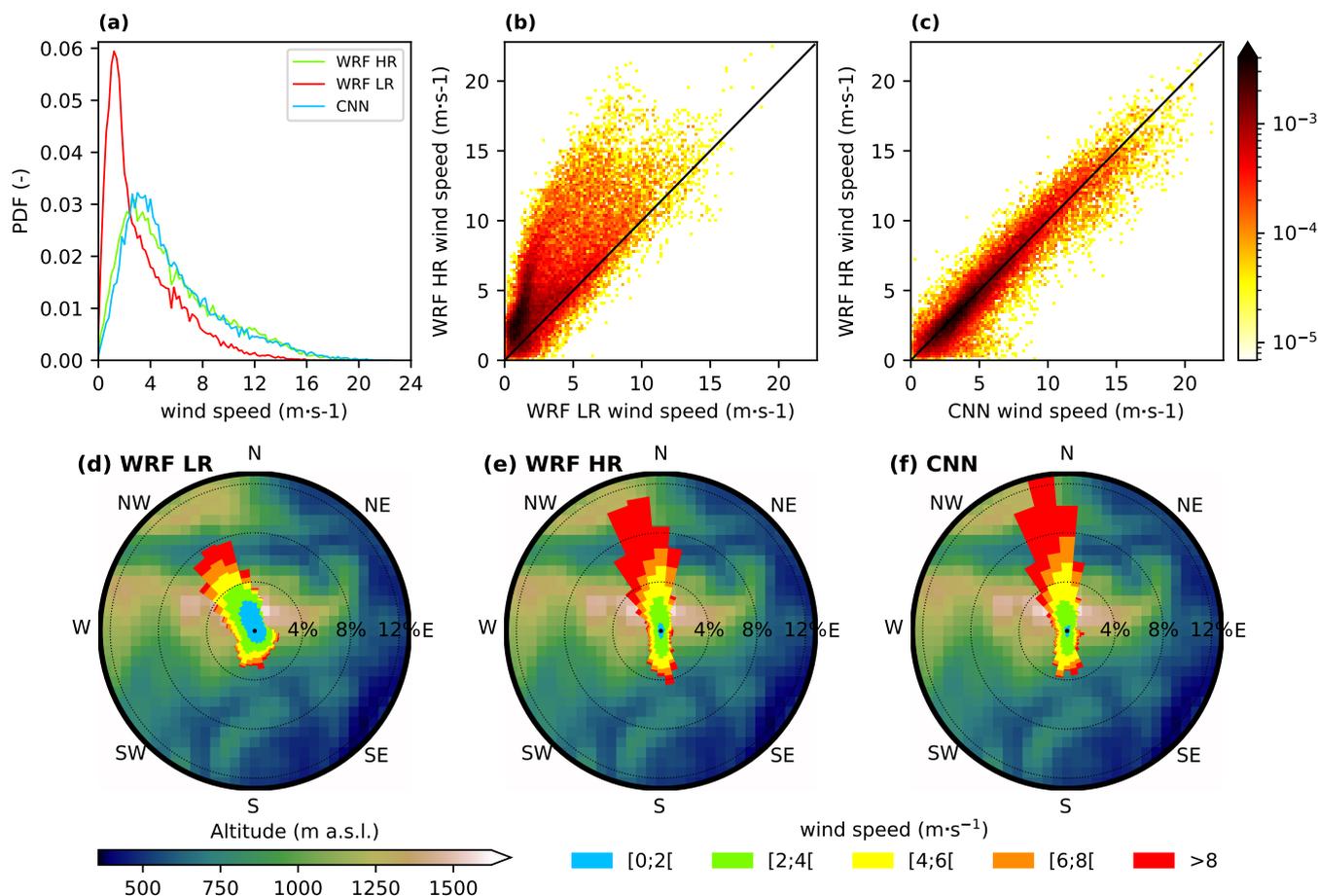
Figure 10. Same as figure 9 for a crest location.

295   Based on those results, it seems that some of the conclusions made upon the analysis of Figs. 7 and 8 (better representation of the ridge acceleration and the channeling in valleys) can be generalized to the whole period. We will now investigate whether the results illustrated in Figs. 9 and 10 could be generalized to the other areas of the domain.
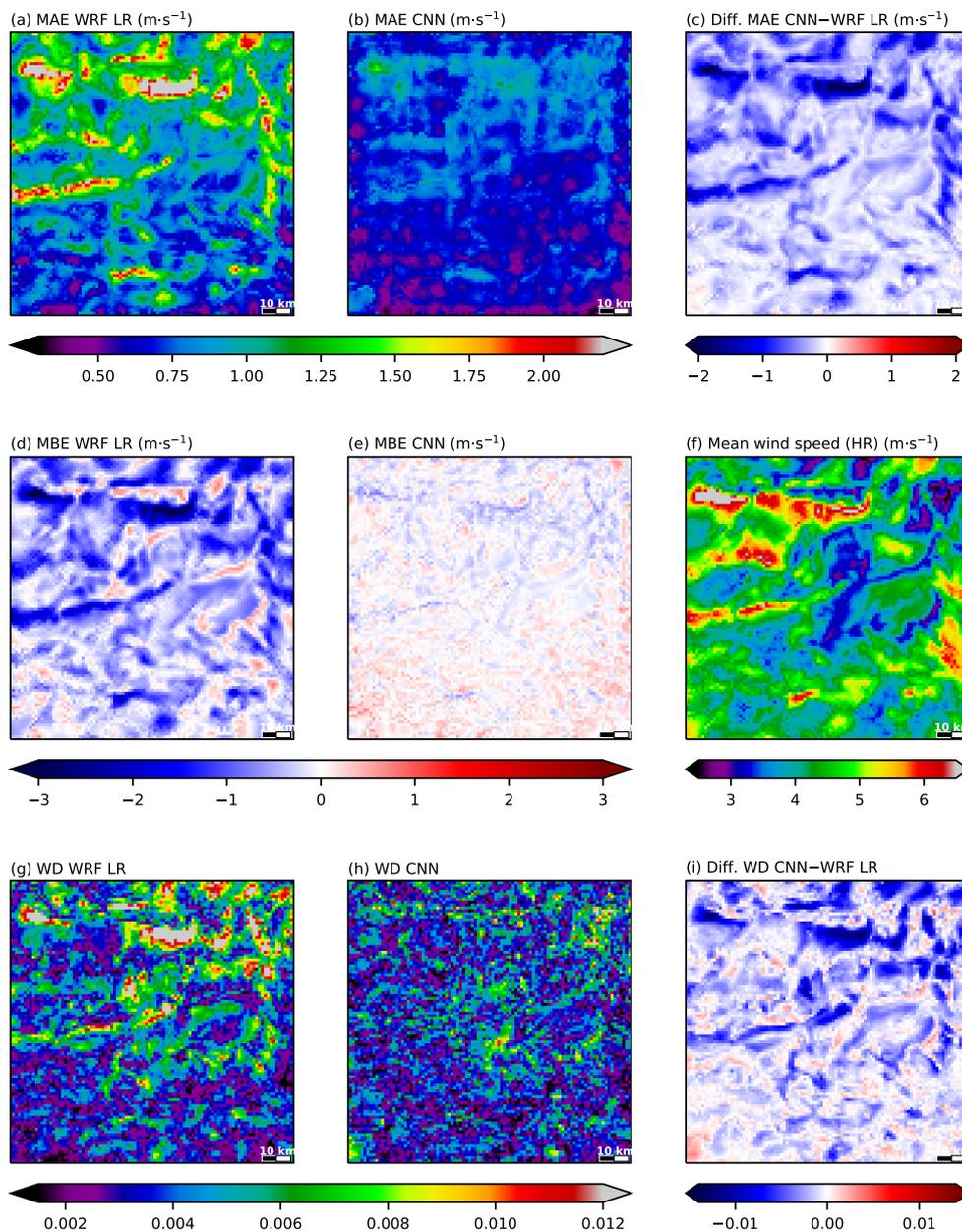
Figure 11. Summary of wind speed metrics averaged over the whole period. Mean absolute error (MAE, a and b), mean bias error (MBE, d and e) and Wasserstein distance (WD, g and h) for the WRF LR forecast and the CNN, respectively. Panels (c) and (i) show the modification brought by the CNN with respect to WRF LR (negative values for improvements) on the MAE and WD, respectively. Panel (f) shows the mean wind speed calculated from the HR forecast.
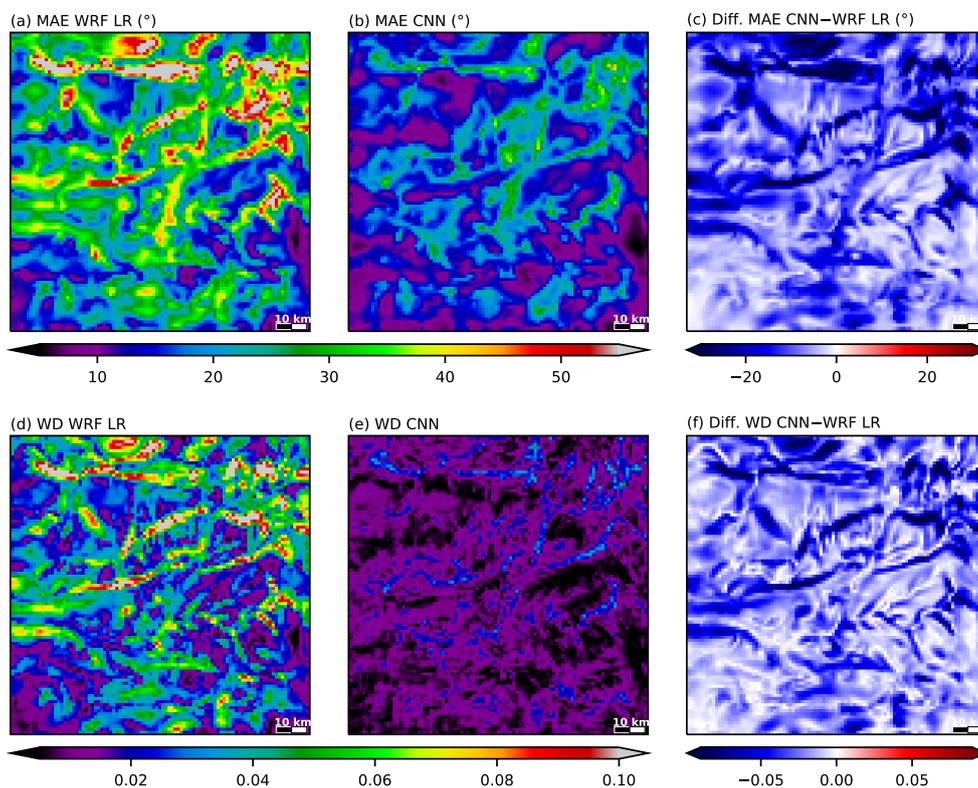
Figure 12. Same as Fig. 11 for wind direction metrics and restricted to the MAE (a, b and c) and the WD (d, e and f).

First of all, the mean speed (over the whole period of simulation) from the HR WRF forecast (Fig. 11f) features some expected patterns related to the topography, such as maxima over the highest topographical elements (ridge acceleration) and minima in valleys (sheltering). The WRF LR mean speed is generally under-estimated over a large part of the domain (Fig.

300   11d), resulting to a global MBE of -0.55 m·s^{-1} (cf. Fig. 4b), with largest negative biases corresponding to crests. Indeed, the ridge acceleration (cf. the mean wind speed maxima from WRF HR in Fig. 11f over the Mont Ventoux, the Lure mountain, the Luberon, . . .) is under-estimated in WRF LR, resulting to large negative bias values (Fig. 11d). This is consistent with the results over the crest site depicted in Fig. 10. The CNN is able to reproduce the features of the mean wind speed since the MBE is low over most of the domain (Fig. 11e), resulting in a global MBE close to zero (cf. Fig. 4b).

305   In WRF LR, largest MAE values are related to largest topographic features: high mountains for the speed, where the ridge acceleration is the most important (Fig. 11a), valleys for the direction, where the channeling cannot be represented without a fine description of the topography (Fig. 12a). The CNN either reduces or at least does not degrade the MAE over the main part of the domain (the only exception being on the southwestern corner which corresponds to a pond area) for both the speed and direction, with larger improvements (in dark blue on Figs. 11c and 12c) corresponding to regions originally featuring the

310   largest errors in WRF LR.

The WD indicates that in WRF LR the wind climatology is the worst either over the crests for the speed (Fig. 11g), resulting from the large under-estimation of ridge acceleration, or in the valleys for the direction (Fig. 12g), resulting from a lack of channeling, which is consistent with the MAE results. The improvement in wind direction brought about by the CNN is translated in a very similar appearance on the MAE (Fig. 12c) or WD (Fig. 12f) field. For the wind speed, there is no similar

315 consensus between the two metrics: the overall structure of the improvement field resembles each other for the MAE (Fig. 11c) and WD (Fig. 11i), but there are local differences, with some spots even reflecting a degradation of the performance evaluated by the WD, as revealed by a few red cells in Fig. 11i. Note that the corresponding MAEs (Fig. 11c) are not degraded at these places. Therefore, from a climatological point of view, WRF LR has mostly a direction issue in valleys, and mostly a speed issue over crests, which is consistent with the results of Figs. 9 and 10, and these shortcomings are well corrected by the CNN.

320 ## 3.4 Diurnal cycle

In this section, we examine how the different metrics evolve according to the time elapsed since the launch of the daily simulations. The results, shown in Fig. 13, represent the average of all the simulations used. Concerning WRF LR, the MAE exhibits a clear diurnal cycle for both the speed (Fig. 13a) and the direction (Fig. 13c), with higher values during the night. This can be explained by the fact that under stable nocturnal conditions, flows are highly dependent on the local topography,

325 resulting in a high spatial heterogeneity (green lines on Figs. 13b and 13d for the speed and direction, respectively). After downscaling, the MAE is reduced for all times. The diurnal cycle remains, although its magnitude is reduced due to a greater improvement during the nighttime when the higher errors were encountered. The nocturnal evolution of the MAE generally exhibits two maxima, at the evening and morning transitions (even if this effect is smoothed on the graphic due to seasonal effect), except for the MAE on the direction in WRF LR which slightly increases towards its maxima reached at the morning

330 transition. Morning and evening transitions are known to be highly difficult periods to forecast, especially for wind parameters which can experience a high temporal variability, which could explain these error maxima.

Note that the lower spatial heterogeneity values encountered in WRF LR in comparison with WRF HR, which is consistent with the results presented on Figs. 7 and 8, were expected since the local topography is not resolved in this simulation. The CNN increases the spatial heterogeneity in comparison with WRF LR (also consistent with Figs. 7 and 8), but not enough to

335 reach the values of the HR forecast, demonstrating that the downscaled forecast is still too smooth. A way to improve this issue would be to use a GAN, which is a specific kind of CNN designed to produce very realistic fields (see for example Miralles et al. (2022)).

4 Conclusions

CNNs are becoming the most popular deep learning tool and their specialization for extracting spatial information is well

340 suited for use in atmospheric sciences. Recent studies demonstrated their ability to downscale wind forecasts. In this study, we aimed to move forward in this question by exploring different strategies for downscaling low-level wind forecast using CNNs, especially regarding the output variables and their associated loss function.
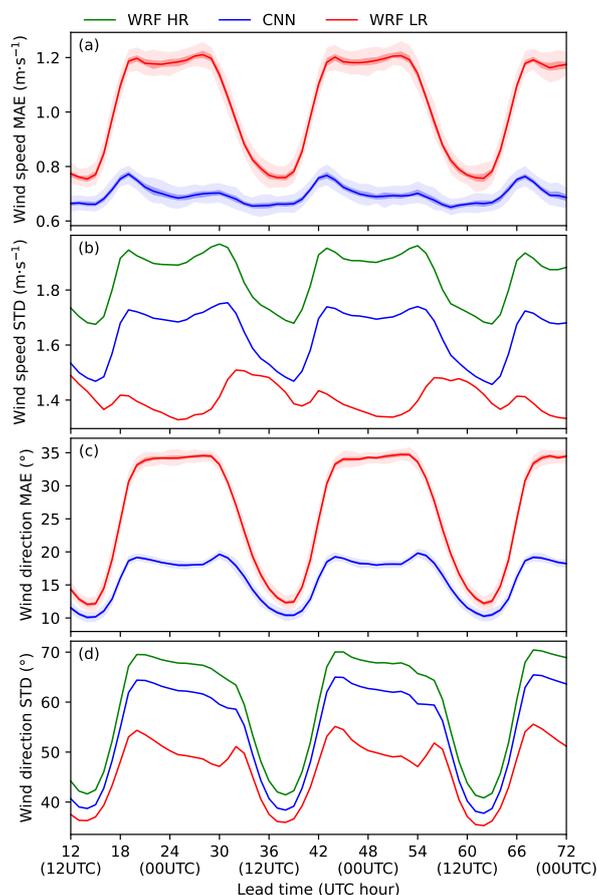
Figure 13. Evolution of the MAE for the speed (a) and direction (c) for lead times from 12 to 72 hours of the WRF LR forecast (in red) and the CNN (in blue). On these panels, the solid line represents the median value, the dark area is the 25 to 75 quantile interval and the light areas are the whiskers, according to the boxplot construction based on the bootstrap distribution, as explained in section 2.4. Panels (b) and (d) represent the time evolution of the heterogeneity (cf. section 2.4) for the speed and the direction, respectively.

The downscaling was applied on WRF wind forecasts over south-eastern France (including the south-western part of the Alps) from its original 9-km resolution onto a 1-km resolution grid. The 1-km resolution data used for training consists of a series of WRF simulations over a 99 km × 99 km domain, launched for each day of a 16-month period.

Among the approaches tested (i.e. computing the wind components, the normalized components or the wind direction, using the MSE loss function in its classical version or with specific adaptations), there is no one that outperforms the others for both the direction and the speed at the same time. Nevertheless, combining two different CNNs, dedicated to the direction or speed forecast respectively, yields the better overall performance. The best direction forecast is derived from the normalized components, which we found to be more accurate when the loss function is customized by adding a penalty term designed to produce a more physically consistent couple of components. The best speed forecast is derived from the wind components,

using a modified MSE loss function designed to remove the speed bias, even if the performance on the individual components is degraded.

In comparison with the initial 9-km resolution forecast, the CNN reduced the wind speed bias from -0.55 m·s^{-1} to -0.01

355   m·s^{-1}, the wind speed MAE from 1.02 m·s^{-1} to 0.69 m·s^{-1}, and the wind direction MAE from 25.9^o to 15.5^o. Moreover, some typical topographical features, poorly represented in the LR forecast, are well reproduced in the downscaled wind fields, both for speed (ridge acceleration, leeward deceleration, sheltering in valleys) and direction (deflection, valley channeling). Regarding the diurnal cycle, there is a general improvement in the forecast, especially during the nighttime, stable stratification period, which is the most difficult to simulate. Finally, the downscaling creates a spatial heterogeneity in the wind fields, but

360   not as the same level as in the HR forecast. In the future, this issue could be solved by using generative networks, which are specifically designed to produce realistic fields, but with the risk of impacting the overall performance.

We will extend this study by evaluating the performance of the method presented in this paper when applying over areas other than those where the CNNs have been trained. The first tests already performed have shown an improvement over the initial low-resolution wind fields, although this improvement is much less than that obtained over the area of the present study.

365   *Author contributions.* FD developed and implemented the CNNs-based models, performed the analyses and wrote the initial version of the manuscript. PD and TH contributed to the analyses and the writing. TH performed the WRF simulations.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We acknowledge the CEA MRISQ project for the funding of the research. We also thank the Laboratoire d'Aérologie for its support, particularly with regard to GPU computing resources.

References

Breiman, L.: Random forests, Machine learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

de Bode, M., Hedde, T., Roubin, P., and Durand, P.: Fine-Resolution WRF Simulation of Stably Stratified Flows in Shallow Pre-Alpine Valleys: A Case Study of the KASCADE-2017 Campaign, Atmosphere, 12, https://doi.org/10.3390/atmos12081063, 2021.

de Bode, M., Hedde, T., Roubin, P., and Durand, P.: A Method to Improve Land Use Representation for Weather Simulations Based on High-Resolution Data Sets—Application to Corine Land Cover Data in the WRF Model, Earth and Space Science, 10, e2021EA002 123, https://doi.org/10.1029/2021EA002123, 2023.

Dujardin, J. and Lehning, M.: Wind-Topo: Downscaling near-surface wind fields to high-resolution topography in highly complex terrain with deep learning, Quarterly Journal of the Royal Meteorological Society, 148, 1368–1388, https://doi.org/10.1002/qj.4265, 2022.

Dupuy, F., Duine, G.-J., Durand, P., Hedde, T., Roubin, P., and Pardyjak, E.: Local-Scale Valley Wind Retrieval Using an Artificial Neural Network Applied to Routine Weather Observations, J. Appl. Meteorol. Clim., 58, 1007–1022, https://doi.org/10.1175/JAMC-D-18-0175.1, 2019.

Dupuy, F., Duine, G.-J., Durand, P., Hedde, T., Pardyjak, E., and Roubin, P.: Valley Winds at the Local Scale: Correcting Routine Weather Forecast Using Artificial Neural Networks, Atmosphere, 12, https://doi.org/10.3390/atmos12020128, 2021a.

Dupuy, F., Mestre, O., Serrurier, M., Kivachuk Burdá, V., Zamo, M., Cabrera-Gutiérrez, N. C., Bakkay, M. C., Jouhaud, J.-C., Mader, M.-A., and Oller, G.: ARPEGE Cloud Cover Forecast Postprocessing with Convolutional Neural Network, Weather and Forecasting, 36, 567 – 586, https://doi.org/10.1175/WAF-D-20-0093.1, 2021b.

Fiddes, J. and Gruber, S.: TopoSCALE v.1.0: downscaling gridded climate data in complex terrain, Geoscientific Model Development, 7, 387–405, https://doi.org/10.5194/gmd-7-387-2014, 2014.

Goodfellow, I., Bengio, Y., and Courville, A.: Deep Learning, MIT Press, http://www.deeplearningbook.org, 2016.

Harris, L., McRae, A. T. T., Chantry, M., Dueben, P. D., and Palmer, T. N.: A Generative Deep Learning Approach to Stochastic Downscaling of Precipitation Forecasts, Journal of Advances in Modeling Earth Systems, n/a, e2022MS003 120, https://doi.org/10.1029/2022MS003120, 2022.

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

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

Jammalamadaka, S. R. and SenGupta, A.: Topics in Circular Statistics, WORLD SCIENTIFIC, https://doi.org/10.1142/4031, 2001.

Kruyt, B., Mott, R., Fiddes, J., Gerber, F., Sharma, V., and Reynolds, D.: A Downscaling Intercomparison Study: The Representation of Slope- and Ridge-Scale Processes in Models of Different Complexity, Frontiers in Earth Science, 10, https://doi.org/10.3389/feart.2022.789332, 2022.

Kulkarni, A., Shivananda, A., and Sharma, N. R.: Image Super-Resolution, pp. 261–295, Apress, Berkeley, CA, https://doi.org/10.1007/978-1-4842-8273-1_8, 2022.

Lagerquist, R., McGovern, A., and Gagne II, D. J.: Deep Learning for Spatially Explicit Prediction of Synoptic-Scale Fronts, Weather and Forecasting, 34, 1137–1160, https://doi.org/10.1175/WAF-D-18-0183.1, 2019.

Lang, M. N., Schlosser, L., Hothorn, T., Mayr, G. J., Stauffer, R., and Zeileis, A.: Circular Regression Trees and Forests with an Application to Probabilistic Wind Direction Forecasting, Journal of the Royal Statistical Society Series C: Applied Statistics, 69, 1357–1374, https://doi.org/10.1111/rssc.12437, 2020.

Le Toumelin, L., Gouttevin, I., Galiez, C., and Helbig, N.: A two-folds deep learning strategy to correct and downscale winds over mountains, Nonlinear Processes in Geophysics Discussions, 2023, 1–32, https://doi.org/10.5194/npg-2023-10, 2023a.

Le Toumelin, L., Gouttevin, I., Helbig, N., Galiez, C., Roux, M., and Karbou, F.: Emulating the Adaptation of Wind Fields to Complex
410   Terrain with Deep Learning, Artificial Intelligence for the Earth Systems, 2, e220 034, https://doi.org/10.1175/AIES-D-22-0034.1, 2023b.

LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, Nature, 521, 436, https://doi.org/10.1038/nature14539, 2015.

Leinonen, J., Nerini, D., and Berne, A.: Stochastic Super-Resolution for Downscaling Time-Evolving Atmospheric Fields With a Generative Adversarial Network, IEEE Transactions on Geoscience and Remote Sensing, 59, 7211–7223, https://doi.org/10.1109/TGRS.2020.3032790, 2021.

415 Miralles, O., Steinfeld, D., Martius, O., and Davison, A. C.: Downscaling of Historical Wind Fields over Switzerland Using Generative Adversarial Networks, Artificial Intelligence for the Earth Systems, 1, e220 018, https://doi.org/10.1175/AIES-D-22-0018.1, 2022.

Rabin, J., Delon, J., and Gousseau, Y.: Circular Earth Mover's Distance for the comparison of local features, in: 2008 19th International Conference on Pattern Recognition, pp. 1–4, https://doi.org/10.1109/ICPR.2008.4761372, 2008.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat: Deep learning and process understanding
420 for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, CoRR, abs/1505.04597, http://arxiv.org/abs/1505.04597, 2015.

Salameh, T., Drobinski, P., Vrac, M., and Naveau, P.: Statistical downscaling of near-surface wind over complex terrain in southern France, Meteorology and Atmospheric Physics, 103, 253–265, https://doi.org/10.1007/s00703-008-0330-7, 2009.

425 Schmidli, J., Böing, S., and Fuhrer, O.: Accuracy of Simulated Diurnal Valley Winds in the Swiss Alps: Influence of Grid Resolution, Topography Filtering, and Land Surface Datasets, Atmosphere, 9, https://doi.org/10.3390/atmos9050196, 2018.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M., et al.: A description of the advanced research WRF model version 4, National Center for Atmospheric Research: Boulder, CO, USA, 145, 550, https://doi.org/10.5065/1dfh-6p97, 2019.

430 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, 1929–1958, 2014.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R.: Generating High Resolution Climate Change Projections through Single Image Super-Resolution: An Abridged Version, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18, pp. 5389–5393, International Joint Conferences on Artificial Intelligence Organization,
435 https://doi.org/10.24963/ijcai.2018/759, 2018.

Vannitsem, S., Bremnes, J. B., Demaeyer, J., Evans, G. R., Flowerdew, J., Hemri, S., Lerch, S., Roberts, N., Theis, S., Atencia, A., Bouallègue, Z. B., Bhend, J., Dabernig, M., Cruz, L. D., Hieta, L., Mestre, O., Moret, L., Plenković, I. O., Schmeits, M., Taillardat, M., den Bergh, J. V., Schaeybroeck, B. V., Whan, K., and Ylhaisi, J.: Statistical Postprocessing for Weather Forecasts: Review, Challenges, and Avenues in a Big Data World, Bulletin of the American Meteorological Society, 102, E681 – E699, https://doi.org/10.1175/BAMS-D-19-0308.1,
440 2021.

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

Whiteman, C. D.: Mountain Meteorology: Fundamentals and Applications, Oxford University Press, https://doi.org/10.1093/oso/9780195132717.001.0001, 2000.

Wilks, D.: Chapter 8 - Forecast Verification, in: Statistical Methods in the Atmospheric Sciences, edited by Wilks, D. S., vol. 100 of *International Geophysics*, pp. 301 – 394, Academic Press, https://doi.org/10.1016/B978-0-12-385022-5.00008-7, 2011.

Yamartino, R. J.: A Comparison of Several "Single-Pass" Estimators of the Standard Deviation of Wind Direction, Journal of Applied Meteorology and Climatology, 23, 1362 – 1366, https://doi.org/10.1175/1520-0450(1984)023<1362:ACOSPE>2.0.CO;2, 1984.

Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.-H., and Liao, Q.: Deep Learning for Single Image Super-Resolution: A Brief Review, IEEE Transactions on Multimedia, 21, 3106–3121, https://doi.org/10.1109/TMM.2019.2919431, 2019.

Zamo, M., Bel, L., Mestre, O., and Stein, J.: Improved Gridded Wind Speed Forecasts by Statistical Postprocessing of Numerical Models with Block Regression, Weather and Forecasting, 31, 1929 – 1945, https://doi.org/10.1175/WAF-D-16-0052.1, 2016.