Response to reviewer #1

We thank the reviewer for his/her comments, which are reproduced in black hereafter. Our responses are in blue. In the revised version of the manuscript, all the modifications are in red.

Summary

The authors build on downscaling methods using convolutional neural networks for wind forecasts over south-eastern France. Here, they focus on exploring different output variables and loss functions. They find that there is not one that is better than the others in terms of both direction and speed simultaneously. But combining two different CNNs one for direction and one for speed results in a better performance. This is then analyzed using error metrics in a quantitative and qualitative way.

The manuscript is well written and gives new insights trough the extensive evaluation. Parts of the method section need to be improved to better understand which variables are used and which data is used for training and testing. Finally, the results that the authors already have for unseen areas should be included in the study.

Comments

In the abstract, could you use % to explain how the MAE is reduced instead of just numbers?

We added the relative improvement on the MAE in the abstract. It amounts to 32% for the speed and 40% for the direction after downscaling. (lines 17-19)

In section „2 Methods" could you state your objective and which variables are used to predict the speed and direction in a small paragraph?

We think it would be redundant to recall the objective at this place in the manuscript, since it is already stated near this section at:

- the end of the section "Introduction", lines 72-77
- the beginning of section "2.2 Training data", lines 89-90

The variables used are detailed in section 2.2. We added the list of the 35 variables used in the revised manuscript (see section 2.2 and Table 1). They are all used to train the different CNNs.

In section „2.2 Training data“ could you state which data you used for training and which for testing?

As the other reviewer made a similar comment, a joint response is provided below.

For each CNN model tested, we performed a k-fold cross validation (k=4, so 4 different trainings for each CNN model tested are performed) so that we have the largest dataset to test the models. For each of the 4 trainings, 25% of the dataset is used for testing while the remaining 75% are used for training. By combining the results of the 4 trainings applied to the 4 test sets, we get a testing dataset for the whole period (that is to say the largest dataset possible in our case).

As this is highly related to the evaluation process, we think it is more appropriate to describe this method in section 2.4. We added more details in that section.

L 119: 32x32 data and 288x288 data, by data do you mean grid points?

Yes. We changed “data” by “grid points” in the new version of the text (lines 124 and 127).

L 119: which domain corresponds to the HR grid of 288x288 and why do you need this larger domain which is than cropped?

The 288x288 grid is centered on the D3 domain and roughly corresponds to the D2 domain of the WRF simulations (297x297 km, cf. Fig. 1 in the revised manuscript). Using data on an enlarged area helps give more information on the regional atmospheric conditions.

We added this justification to the new version of the text (lines 125-126).

Figure 3: where are N_I and N_O (the number of input variables and the number of target variable) specified?

We have used the same set of input variables for all the CNNs. Therefore, N_I is the same for all the CNNs (N_I = 35).

N_O depends on the number of output variables, which is 2 for all the CNNs (either u and v, or ũ and ṽ) except for the CNNdir where N_O=1 since only the direction is calculated.

We added this information in the caption of Fig. 3 as well as in Table 2.

L 131: Please explain what you mean by „speed“ and „direction“

We mean that we are interested not only by the wind speed, as in most studies about wind forecast downscaling, but we also want to calculate the wind direction. We modified the sentence in the article (lines 137-138).

L 138: „We tested their approach“. Please rephrase such that it is clear that you trained another CNN using the described loss.

We modified the sentence:

"We tested their approach (model called CNNdir hereafter)."

by

"We tested their approach via a specific CNN training (this model is called CNNdir hereafter)." (line 145)

L 150: which other way is used to compute the wind speed?

In case the CNN delivers the direction only, we used the speed extracted from the $CNN_{u,v}$ or $CNN_{u,v,L_{spd}}$. We added this clarification in the new version of the text (line 158).

L 160 onward: Why do you incorporate the condition u^2+v^2=1 in the loss and not predict u and compute v or have it built in the neural network architecture?

Thank you for this suggestion. As the other reviewer made a similar comment, a joint response is provided below.

We tried to implement the suggested modification by computing the value of $\widehat{\tilde{v}}$ knowing only the value of $\widehat{\tilde{u}}$ using the formula $\widehat{\tilde{u}}^2 + \widehat{\tilde{v}}^2 = 1$ in order to get couples of $\widehat{\tilde{u}}$ and $\widehat{\tilde{v}}$ values that are consistent. However, it is not possible to derive the sign of $\widehat{\tilde{v}}$ from this formula. Therefore, using the results of $CNN_{\tilde{u},\tilde{v}}$, we used the sign from the output $\tilde{v}$ in addition to the $\tilde{u}$ output values to calculate $\widehat{\tilde{v}}$ as follows:

$$\widehat{\tilde{v}} = sign(\widehat{\tilde{v}}) \times \sqrt{1 - \widehat{\tilde{u}}^2}$$

Similarly, we computed $\widehat{\tilde{u}}$ as:

$$\widehat{\tilde{u}} = sign(\widehat{\tilde{u}}) \times \sqrt{1 - \widehat{\tilde{v}}^2}$$

These two new tests are called $CNN_{\tilde{u}\rightarrow\tilde{v}}$ and $CNN_{\tilde{v}\rightarrow\tilde{u}}$ in the revised version of the article (note that no additional CNNs were trained since we used the results of the $CNN_{\tilde{u},\tilde{v}}$). We got results that were worse than with $CNN_{\tilde{u},\tilde{v}}$ and $CNN_{\tilde{u},\tilde{v},L2}$ on the direction forecast.

This new approach is described in section 2.3.2 (lines 175-180). The results of $CNN_{\tilde{u}\rightarrow\tilde{v}}$ and $CNN_{\tilde{v}\rightarrow\tilde{u}}$ are presented in Fig. 4 and lines 245-248 in the revised manuscript.

In section „2.4 Wind forecast evaluation“ could you state all the formulas for the evaluation metrics used?

We added the formulas for the Wasserstein distance (Eqs. 5 and 6). We do not think it is relevant to add the formulas for the MAE, MBE and standard deviation as these are basic.

In section „2.5 Computational considerations“ could you state how long you trained the neural networks and on which GPU?

The training lasts approximately 4 hours on a NVidia GeForce GTX TITAN V GPU.

We added these details in the new version of the article (section 2.5).

In section „3.3 Wind climatology at specific sites“, over which period is the climatology computed? And why only at 2 different locations?

The climatology is made on the whole period of our dataset, that is to say from the 24 December 2020 to the 5 May 2022.

It is not possible to make such an analysis for numerous grid cells of the domain (the domain contains 99x99=9801 grid cells). That is why we picked a limited number of sites. We thought it was interesting to highlight the characteristics of a crest site and a valley site, which are very different, since they correspond to challenging locations for wind forecast.

Moreover, Figs. 11 and 12 (corresponding to Figs. 12 and 13 in the revised manuscript) confirm that the improvements found on the valley site (mainly on the direction) and crest site (mainly on the speed) are generalized to all the crests and valleys of the domain, justifying the choice of this kind of sites for climatological analysis.

We added this justification to the text (lines 305-306).

Figure 11: Please specify „whole period“.

“whole period” refers to the dataset we used, that is to say from the 24 December 2020 to the 5 May 2022. We added this information to the caption of Fig. 12.

The last sentence in the conclusion states that you have more results for evaluating the CNNs over unseen areas. If you already have these results why not include them?

As the other reviewer made a similar comment, a joint response is provided below.

As stated in the paragraph mentioned, this is only a preliminary work. Indeed, for now, the dataset on the other sites is really small (a simulation of 72 hours on a single site, cf. Figs. 1 and 2 below), preventing a significant analysis. Therefore, this work must be extended by generating a larger (in time and spatially) HR dataset in order to get significant insight from the results. For this reason, we think these results are not worth publishing at the present state and must remain as a mention to future work to make.

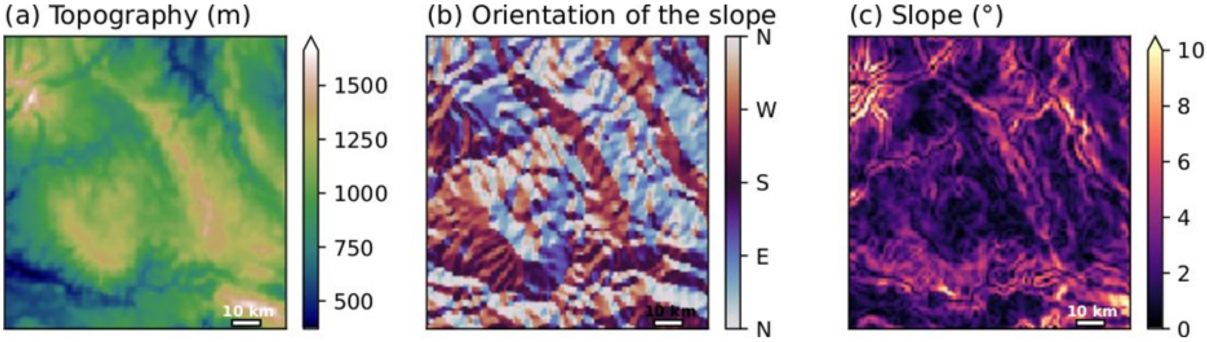

Figure 1: New domain tested (center of France) – (a) topography (in m a.s.l.), (b) orientation of the slope and (c) local slope.

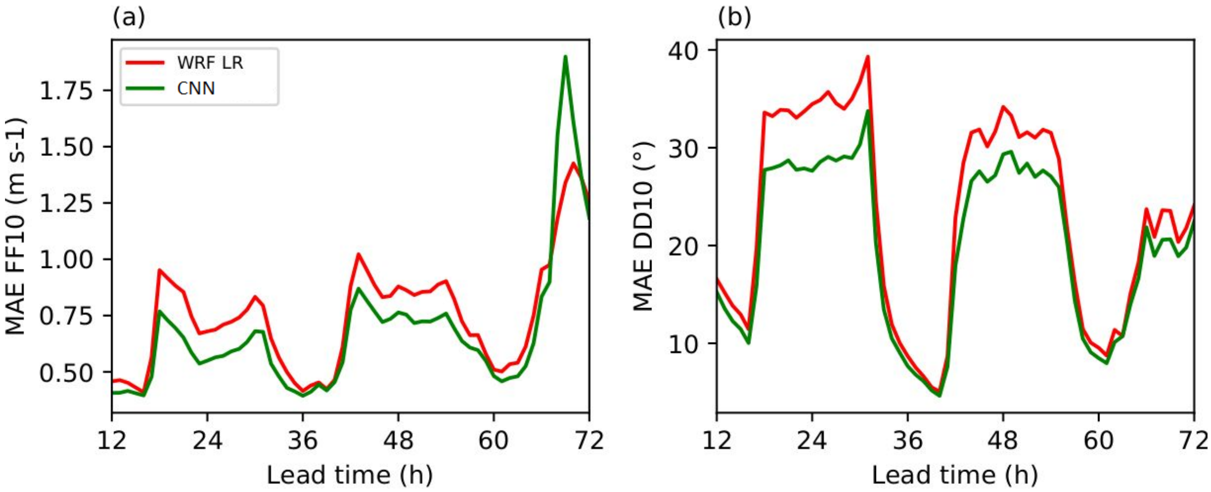

Figure 2: Evolution of the MAE on the wind speed (a) and direction (b) average over the whole domain for a 3-day simulation period.