

Downscaling of Historical Wind Fields over Switzerland Using Generative Adversarial Networks

OPHÉLIA MIRALLES,^a DANIEL STEINFELD,^b OLIVIA MARTIUS,^b AND ANTHONY C. DAVISON^a

^a *Institute of Mathematics, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*

^b *Oeschger Centre for Climate Change Research and Institute of Geography, University of Bern, Bern, Switzerland*

(Manuscript received 7 March 2022, in final form 31 August 2022)

ABSTRACT: Near-surface wind is difficult to estimate using global numerical weather and climate models, because airflow is strongly modified by underlying topography, especially that of a country such as Switzerland. In this article, we use a statistical approach based on deep learning and a high-resolution digital elevation model to spatially downscale hourly near-surface wind fields at coarse resolution from ERA5 reanalysis from their original 25-km grid to a 1.1-km grid. A 1.1-km-resolution wind dataset for 2016–20 from the operational numerical weather prediction model COSMO-1 of the national weather service MeteoSwiss is used to train and validate our model, a generative adversarial network (GAN) with gradient penalized Wasserstein loss aided by transfer learning. The results are realistic-looking high-resolution historical maps of gridded hourly wind fields over Switzerland and very good and robust predictions of the aggregated wind speed distribution. Regionally averaged image-specific metrics show a clear improvement in prediction relative to ERA5, with skill measures generally better for locations over the flatter Swiss Plateau than for Alpine regions. The downscaled wind fields demonstrate higher-resolution, physically plausible orographic effects, such as ridge acceleration and sheltering, that are not resolved in the original ERA5 fields.

SIGNIFICANCE STATEMENT: Statistical downscaling, which increases the resolution of atmospheric fields, is widely used to refine the outputs of global reanalysis and climate models, most commonly for temperature and precipitation. Near-surface winds are strongly modified by the underlying topography, generating local flow conditions that can be very difficult to estimate. This study develops a deep learning model that uses local topographic information to spatially downscale hourly near-surface winds from their original 25-km resolution to a 1.1-km grid over Switzerland. Our model produces realistic high-resolution gridded wind fields with expected orographic effects but performs better in flatter regions than in mountains. These downscaled fields are useful for impact assessment and decision-making in regions where global reanalysis data at coarse resolution may be the only products available.

KEYWORDS: Topographic effects; Wind; Downscaling; Numerical weather prediction/forecasting; Deep learning; Neural networks

1. Introduction

Near-surface wind fields are of interest in applications such as wind energy projects (Emeis 2014; Staffell and Pfenninger 2016; Dujardin et al. 2021), risk and damage assessment for intense windstorms (Schwierz et al. 2010; Stucki et al. 2014; Welker et al. 2016; Stucki et al. 2016), snow distribution and avalanche forecasting (Lehning et al. 2008), and modeling the spread of wildfires (e.g., Sharples et al. 2012). Detailed wind information at high spatial and temporal resolution and for long time periods is needed to study wind-related impacts, space–time variability, and long-term trends, but the accurate representation of surface winds in complex terrain is challenging because winds fluctuate over a wide range of time and spatial scales. Surface weather stations provide accurate and long-term local wind measurements but are sparsely distributed and the spatial interpolation of wind between them is difficult (Kruyt et al. 2017; Harris et al. 2020). Climate and weather prediction models provide spatially

and temporally continuous gridded wind data that are physically consistent, but the observed wind field varies at much smaller spatial scales than those in global versions of such models (Koller and Humar 2016; Molina et al. 2021), whose grid resolutions range from tens to hundreds of kilometers and at best resolve only major topographical features. Models at these resolutions do not capture local flow effects such as wind speedup over ridges, flow channeling in valleys, flow deflection around and over mountain ranges, and thermally induced winds that alter the local flow field.

Reanalysis datasets produced by global weather prediction models, such as the state-of-the-art ERA5 reanalysis from the European Centre for Medium-Range Weather Forecasts (Hersbach et al. 2020), provide long-term gridded wind fields on a global scale, but their coarse spatial resolution (~25 km) limits their use for impact assessment in complex terrain. Although large-scale atmospheric flow conditions associated with surface winds are broadly well represented in reanalysis datasets (Molina et al. 2021), especially over flat regions (Ramon et al. 2019), such data are too coarse to accurately represent local surface wind conditions in regions with complex terrain, such as the Swiss mountains (Graf et al. 2019; Dörenkämper et al. 2020). The horizontal grid resolution in

Corresponding authors: Ophélie Miralles, ophelia.miralles@epfl.ch; Daniel Steinfeld, daniel.steinfeld@giub.unibe.ch

DOI: 10.1175/AIES-D-22-0018.1 e220018

© 2022 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](#)).

global reanalyses is relatively coarse, in part due to their high computational demands. On the other hand, high-resolution numerical model data are available, but typically only for short time periods. The Consortium for Small-Scale Modeling (COSMO) regional operational weather prediction model COSMO-1 (MeteoSwiss 2016) of the Swiss weather service has been successfully run at a grid resolution of 1.1 km over Switzerland since 2016, producing realistic representations of local wind conditions, but no long-term (>5 years) gridded climatology for wind exists (MeteoSwiss 2018). Hence there is a trade-off between geographic coverage and time span on the one hand and spatial detail on the other. This data gap can be filled by applying downscaling methods to long-term historical reanalysis and climate model outputs (Gutowski et al. 2016). This motivates the development of a downscaling technique to produce a gridded near-surface wind climatology at higher spatial and temporal resolution.

Statistical downscaling must address the question of what is considered to be the *ground truth*. Most statisticians would agree that field observations are a noisy version of the truth, whereas physicists tend to attribute value to reanalyzing such data, correcting for measurement errors, and smoothing it to fit physical theory. A consequence of these considerations for downscaling is that researchers favor either point-by-point modeling and forecasting based on a limited number of observation stations (Winstral et al. 2017; Nerini 2020) or mapping of low-resolution grids directly to high-resolution ones (Höhlein et al. 2020; Leinonen et al. 2020; Ramon et al. 2021).

Spatiotemporal regression models have been proposed for statistical downscaling (Winstral et al. 2017; Ramon et al. 2021), though they generally assume linear dependence and Gaussianity and often do not account for unobserved spatial phenomena. More complex statistical models have been avoided in the past because of the computational burden of dealing with very large datasets, which precludes applying the simulation-based methods widely used in other contexts. Latent variable models attempt to account for hidden or unobserved effects in high-dimensional data, and Gaussian processes can flexibly capture local correlations and uncertainties. Latent Gaussian models combine these concepts (Lawrence 2003; Rue et al. 2009) and provide a large class of statistical tools. The R integrated nested Laplace approximations (R-INLA) package (Rue et al. 2017) can estimate posterior distributions for latent Gaussian models, but the size of the latent field affects the complexity of precision matrix computation. Rue et al. (2017) argue that assuming Markov properties for the target process can greatly reduce the computational burden, and efficient solutions now exist for fitting multilayer statistical models to large numbers of data points and have been used for environmental applications. For example, Castro-Camilo et al. (2019) use R-INLA to fit a hierarchical Bayesian model involving a biphasic distribution for extreme and nonextreme wind speeds at 260 stations across the United States. However, there is little to no literature on downscaling climate time series using such models. In our study, we use grid-to-grid downscaling to produce entire maps of wind fields. Although we considered using a spatiotemporal Bayesian hierarchical model, the very large

number of data points (~10 billion in total) was impossible to handle using R-INLA.

Instead, we implement deep learning methods that deal with very large amounts of data by introducing a network hierarchy that allows a computer to build complicated structures from simple ones (Goodfellow et al. 2016). This hierarchy is commonly described as a series of layers; the deeper the network, the more layers there are and the more specific the role of each layer. Downscaling atmospheric fields using neural networks is a very recent development (Vandal et al. 2018; Reichstein et al. 2019; Baño Medina et al. 2020; Sha et al. 2020). Machine learning methods for downscaling environmental variables can provide good results, avoid information loss, and require reasonable computational effort if the structure has enough hidden layers (Höhlein et al. 2020). However, neural networks are mainly used to produce deterministic outcomes, which is an issue if one wants to know the distribution of the target process. This can be overcome with a recurrent generative adversarial network (GAN) that adds noise to the original input to make predictions more robust, as proposed by Leinonen et al. (2020) for rainfall data. A more probabilistic approach is to use neural networks to estimate the parameters of a given statistical model, for instance by estimating the parameters of gamma distribution for wind speed data (Nerini 2020). As far as we know, no existing neural network can efficiently downscale wind fields on complex terrain from different low- and high-resolution sources. In this paper, we propose a stochastic deep learning approach using a GAN to downscale historical maps of hourly near-surface wind fields over Switzerland from open-source ERA5 data and local topography. The target high-resolution maps are wind fields from the COSMO-1 model, provided by MeteoSwiss, which represent the local surface winds well. The resulting time series of downscaled wind fields can be used for detailed case studies of past weather events or climatological analyses.

This study is structured as follows. The data used for the downscaling and associated challenges are described in section 2, and the specific deep learning model and its training are explained in detail in sections 3 and 4. Quantitative analysis of the obtained predictions is performed in section 5, and the main findings are given in section 6.

2. Data

a. Geographical setting and typical wind systems in Switzerland

Switzerland has a complex and diverse topography with three main subregions (cf. Fig. 1): the Alps in the central and southern part of the country with high mountain ranges and deep valleys, covering ~60% of Switzerland; the Jura in the northwestern part with lower and narrow mountain ranges, covering ~10%; and, between them, the hilly and densely populated Swiss Plateau, covering ~30%. The elevation ranges from below 300 m to above 4500 m. Figure 1 shows how this topography is represented in the ERA5 reanalysis and in the COSMO-1 model, with respective horizontal grid resolutions of 25 and 1.1 km. The ERA5 grid cannot resolve the complex

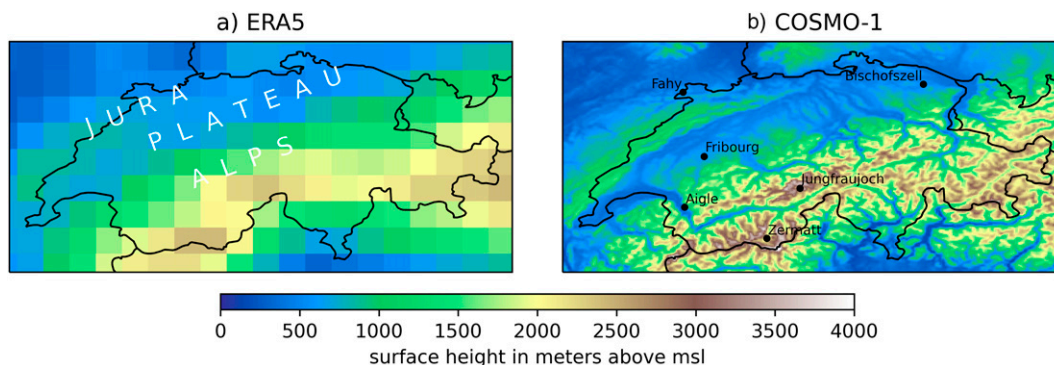


FIG. 1. Maps of Switzerland showing topography in meters above sea level for (a) ERA5 with 25-km resolution and (b) COSMO-1 with 1.1-km resolution. The three subregions of Switzerland are indicated in (a), and the locations of validation sites (see Fig. B1 in appendix B) are indicated in (b).

mountain terrain, but the high mountain ranges and deep inner alpine valleys are well resolved in COSMO-1. This terrain interacts with and modifies the synoptic-scale flow at different scales, generating region-specific surface winds (Barry 2008). At the larger (alpine) spatial scale, the frequent westerly winds are modified by the high mountains, for example, by horizontal and vertical deflection creating mountain waves, and by channeling of the flow (Jackson et al. 2013). A well-known example in Switzerland is the north–south foehn flow, which crosses the main Alpine ridge and leads to a warm and dry downslope windstorm in the lee, affecting many Alpine valleys (Richner and Hächler 2013; Sprenger et al. 2016). Another example is the Bise, an easterly wind that is enhanced in the Swiss Plateau region by channeling between the Jura and the Alps (MeteoSwiss 2015). At the more local scale, thermally driven diurnal mountain-valley winds are generated by temperature contrasts that form within the mountains and valleys due to radiative heating during the day and cooling at night (Weissmann et al. 2005; Zardi and Whiteman 2013).

b. Low-resolution input fields: ERA5 reanalysis

We use the ERA5 reanalysis, the fifth generation of global reanalysis datasets from the European Centre for Medium-Range Weather Forecasts (ECMWF), which has a spatial resolution of 0.25° (≈ 25 km) and is available hourly from 1979 onward. Long-term climate datasets such as this are built by assimilating observations from multiple data sources and solving the main atmospheric evolution equations, with the aim of representing past or current climates on a regular grid (Hersbach et al. 2020). Cycle 41r2 of the Integrated Forecast System (IFS), the global numerical forecast model of the ECMWF, and 4D variational data assimilation of past observations were used to produce the ERA5 reanalysis, which is freely available through the European Union (EU)-funded Copernicus Climate Change Service (C3S). It will eventually be extended back to 1950.

Low-resolution surface (10 m) wind fields covering Switzerland are retrieved from the ERA5 reanalysis as predictors. These consist of gridded u (east–west) and v (south–north) hourly wind speed components on a horizontal grid of 0.25° (≈ 25 km)

from 2016 to 2020. We tested additional predictors from ERA5, also used by Höhlein et al. (2020), in the hope of obtaining information about the local wind systems and driving processes described in section 2a: boundary layer height, surface pressure, forecast surface roughness and geopotential height at 500 hPa. However, they did not improve the performance of the GAN and were not included in the final model.

c. Topographic descriptors

The terrain of Switzerland is complex: local topographic features strongly modify surface wind speeds, and to allow the GAN to learn this relationship we use the topography from the freely available 90-m-resolution SRTM3 digital elevation model (DEM) constructed by NASA and the National Geospatial-Intelligence Agency (NGA; Jarvis et al. 2008). We also tested a comprehensive set of DEM-derived descriptors, calculated using the Python package topo-descriptors (Nerini and Zanetta 2021) provided by MeteoSwiss: directional (south–north and east–west) derivatives, slope and aspect, the ridge/valley norm and direction, and the topographic position index (TPI), which evaluates a grid point's elevation relative to its surroundings (Winstral et al. 2017). However, best performance was reached with the raw DEM.

d. High-resolution target fields: COSMO-1

High-resolution 10-m target fields are from the COSMO-1 model. COSMO-1 is a nonhydrostatic deterministic limited-area numeric weather prediction model that is based on primitive, thermo-hydrodynamical equations describing compressible flow in a moist atmosphere (Consortium for Small-scale Modeling 2017). MeteoSwiss has run COSMO-1 at a grid resolution of 1.1 km with the domain centered over Switzerland operationally since March 2016 (MeteoSwiss 2016), which provides a little more than 4 yr of hourly (reanalysis) data. Boundary conditions are provided by the ECMWF Integrated Forecasting System, which is also the global weather model underlying the ERA5 reanalysis. The performance of COSMO-1 was assessed against weather stations in Krutyt et al. (2018) and found to give good overall wind speed results. We use surface wind estimates from the COSMO-1 analysis provided by MeteoSwiss

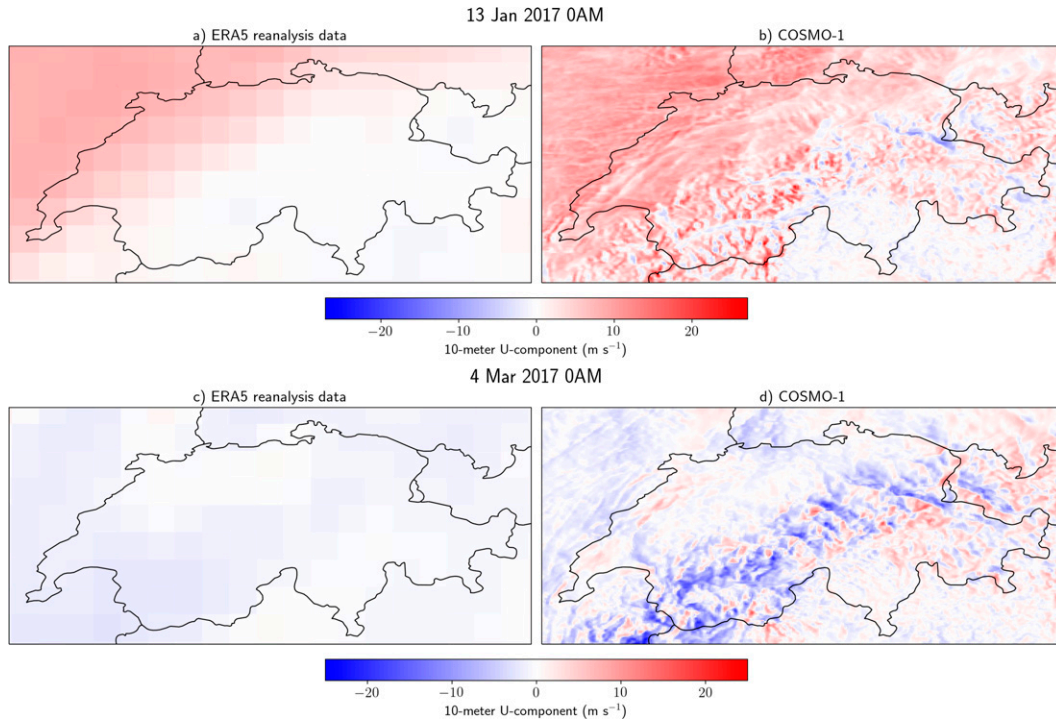


FIG. 2. Examples of (a),(c) ERA5 reanalysis input 10-m u wind component with resolution 25 km and (b),(d) target 10-m u wind component from COSMO-1 with resolution 1.1 km.

at hourly resolution from March 2016 to October 2020. The ERA5 and COMSO-1 10-m wind components u at 0000 UTC 13 January 2017 and at 0000 UTC 4 March 2017 are compared in Fig. 2.

3. GAN

Below we use the term “tensor” to refer to data provided as input to a neural network, data resulting from the transformation associated with a hidden layer of the network, or predictions made by the network. The network is fed with square frames, or “patches,” that are randomly selected from the input map. To ensure stability and speed of training, we do not update the parameters for each observation, but process a “batch” of observations at a time. In this study, all tensors are of dimension five: the first dimension is the batch size, the second is the time coordinate, the third and fourth are the spatial coordinates, and the last, the “channels,” refers to individual scalar variables.

a. General architecture

The generative adversarial network we use has a standard Wasserstein GAN with gradient penalty (WGAN-GP) architecture, comparable to that used for precipitation data by Leinonen et al. (2020). Such a network comprises two different deep neural networks with specific roles. The generator network, or “artist,” takes low-resolution sequences of wind and other covariates as input, convolves and upsamples them through sequential layers, and produces two output images

that are fitted to the high-resolution wind fields during training. The discriminator network, or “critic,” attributes a score that measures the match between the low-resolution input data and the high-resolution wind field prediction. Thus, the purpose of the discriminator is not so much that predicted winds look exactly like COSMO-1 winds, but to attribute a score assessing the consistency of a pair of low/high-resolution winds. The score function is obtained after compressing the information in the low- and high-resolution wind fields into a scalar value through convolutional layers. The critic is optimized throughout the training to make its output score as discriminating as possible. The goal is to clearly distinguish between fake wind fields created by the generator and their real counterparts. Its optimal parameters are found by minimizing a gradient-penalized version of the Wasserstein loss (Gulrajani et al. 2017),

$$\text{Loss}_D(x, y, z) = D(x, y) - D[x, G(x, z)] + \gamma[\|\nabla_{\tilde{y}} D(x, \tilde{y})\|_2 - 1]^2, \quad (1)$$

where x is the low-resolution input tensor, y is the true high-resolution wind field, z is a noise field, D is the score given by the discriminator to a pair of low and high-resolution fields, and $G(x, z)$ is the prediction of the generator (the fake high-resolution wind field). The score is obtained by minimizing the loss function. The final term of Eq. (1) is the gradient penalty, whose influence is determined by the positive scalar γ and that attracts the norm of the gradient toward unity. This term contains a random combination $\tilde{y} = \epsilon y + (1 - \epsilon)G(x, z)$

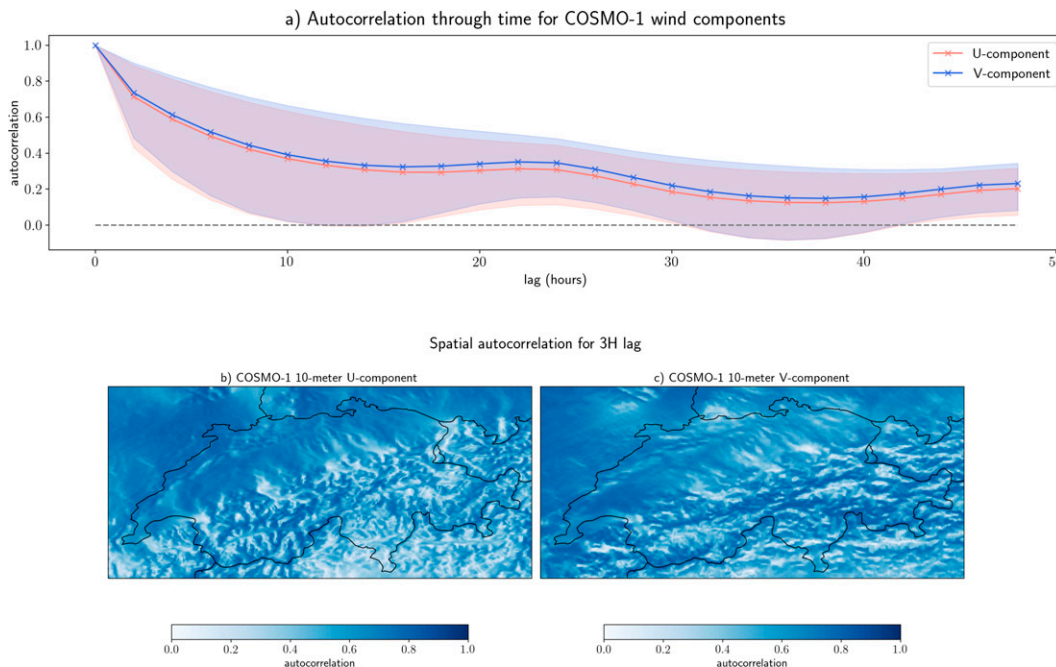


FIG. 3. (a) Mean autocorrelation as a function of lag (h) for COSMO-1 wind components u and v . The shaded area corresponds to 5% and 95% quantiles of the spatiotemporal distribution for u and v . Also shown is the spatial distribution of (b) u and (c) v autocorrelation for a 3-h lag.

of true and predicted wind fields, with ϵ a standard uniform random variable.

Scores attributed to both the true high-resolution and predicted winds should be robust in order that the discriminative ability of the network is reliable. The gradient term in Eq. (1) was introduced by Gulrajani et al. (2017) to enforce the 1-Lipschitz constraint on the discriminator's score relative to its inputs, but it also prevents gradient explosion at the start of the training, which is otherwise common when using deep structures (Huang et al. 2016). The artist's loss is simply the score given by the discriminator to the fake high-resolution output, that is,

$$\text{Loss}_G(x) = D[x, G(x, z)].$$

On the one hand, the critic should score unrealistic predictions as highly as possible so that the artist can improve, while reducing the score attributed to COSMO-1 high-resolution wind fields as much as possible. On the other hand, the optimum of the artist is reached when its loss is minimal, which means that the networks act on $D[x, G(x, z)]$ in opposite ways.

b. Modeling the wind time series

The wind time series of the two 10-m wind components u and v from COSMO-1 present strong short-term autocorrelation (Fig. 3a), which reduces to about 0.2 only after about 30 h. To allow the artist to accurately reproduce this, we augment the generator network with a long short-term memory (LSTM) layer (Hochreiter and Schmidhuber 1997) that uses a

hidden state to recall information about the past. The critic is also given such a layer so that scores are computed and optimized based on a wind sequence rather than on individual wind fields. As Fig. 3 shows, spatial autocorrelation depends on local topography: for both u (Fig. 3b) and v (Fig. 3c) components, autocorrelation is stronger in the plains of the Swiss Plateau and on top of the high mountain ridges than on steep slopes and in the valleys. Hence it is crucial that the topography is fed to the network before the activation of the LSTM layer in order to account for its effect on autocorrelation. The complete architecture of the network is displayed in appendix A (Fig. A1).

c. Generator network

The entry layer of the generator is a concatenation of the input low-resolution wind fields (of size $N_B \times N_T \times S \times S \times N_P$, where N_B is the batch size, N_T is the number of consecutive time steps used for building the sequences, S is the patch size, and N_P is the number of predictors) and random Gaussian noise (of size $N_B \times N_T \times S \times S \times N_N$, where N_N is the number of noise channels), which is used to robustify learning by making it less dependent on the precise data used. Introducing noise also allows for stochasticity in the model by sampling from the latent field distribution. The noise standard deviation of 0.1 m s^{-1} is chosen to represent small deviations from the input wind field.

After concatenating the input data and noise, we progressively increase the resolution of the input random vector to attain the desired resolution in the output. We decompose this step into two simultaneous substeps. The number of channels,

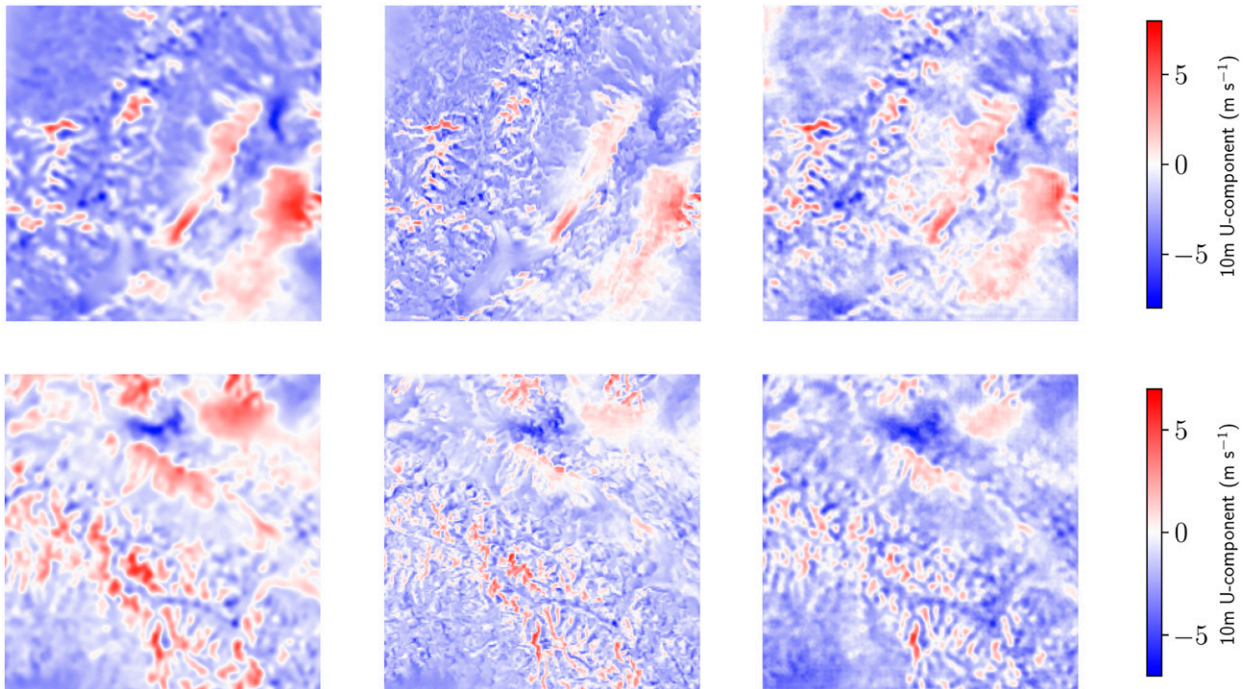


FIG. 4. Prediction of the u component of 10-m wind field by the generator model presented in section 3. The rows denote different 80-km patches at different times for inputs from (left) the COSMO-1 model at resolution 1.1 km blurred with a Gaussian filter with standard deviation 2, (center) the original raw high-resolution wind fields, and (right) the model prediction using RMSE loss.

$N_p + N_N$, is first increased using padded convolutions to leave room for the information contained in the spatial dimension of the tensor, and convolutional layers with strides are simultaneously applied to the tensor to decrease the spatial dimension, triggering the transfer of information to the channels. This substep is shown in Fig. A1a of appendix A: it starts after the concatenation of input and noise channels and is terminated by an LSTM layer and a first split connection. At the end of the substep, the image size has been reduced by a factor of four. Layers in which the same operation is applied to different time steps are referred to as “TimeDistributed” in Fig. A1a. This substep can be seen as an organized and efficient destructuring by the generator of the information contained in the input layer in order to recreate a higher-resolution version of the image. The second substep increases the resolution by transferring information from the channels back to the spatial dimensions. The last upsampling layers of the generator use spatial bilinear interpolation rather than transposed convolutions, as this produces smoother outputs. All convolution layers from the upsampling step are activated with the leaky rectified linear unit (ReLU) function $x \mapsto x^+ - 0.2x^-$, where x^+ and x^- are the positive and negative parts of x .

Last, wind fields (of size $N_B \times N_T \times S \times S \times 2$) are predicted using padded convolution with linear activation (see the last convolution layer in Fig. A1a). Using bounded activation functions is known to increase the stability of training, especially on visual feature recognition problems

(Liew et al. 2016). The idea of constraining the generated wind fields using a normalization constant and a tanh activation function for the last layer was considered but not applied, primarily to avoid underestimating extreme winds.

To assess the functioning of the generator network, we blur COSMO-1 high-resolution fields using a Gaussian filter with a standard deviation of 2 and try to predict the unblurred high-resolution fields by minimizing the root mean squared error between generated and realized fields. No other predictor is added to this optimization problem, in order to check whether the generator alone can perform well on a very simple task. Figure 4 shows that the network produces good results when trained on a small number of steps, or “epochs”: The blurring pattern seems to be rapidly understood by the generator. The training validation metrics detailed in section 5 confirm that it performs well.

d. Discriminator network

The discriminator network, or critic, is used to determine whether a pair of low- and high-resolution wind fields (both of size $N_B \times N_T \times S \times S \times 2$) are a good match and to decide if the high-resolution wind field is from COSMO-1 or predicted by the generator. Accordingly, the first layer of the discriminator inputs concatenated low and high-resolution images in order to evaluate how well they match. In Fig. A1b of appendix A, this step is represented by the temporary creation of two branches, one in which the match between low and high-

resolution images is processed as a time-varying tensor, and one in which only the high-resolution image goes through this process. The two branches allow the generator to learn time series specificities for both the pair of low/high-resolution winds and the high-resolution field. The tensors containing information about the match and the high-resolution wind field alone are then concatenated and undergo a progressive information transfer from the spatial dimensions to the channels, as described for the first step of the generator network (the successive application of TimeDistributed convolution layers is shown in Fig. A1b). Last, a dense layer with linear activation is averaged on the time dimension to produce the final score with size $N_B \times 1$ for the two wind fields.

To check whether the critic can attribute different scores to realized and generated inputs, we train it alone by minimizing the loss introduced in Eq. (1). Inputs generated by the artist are not available when we train the critic alone, so we replace them with more obvious fake images, Gaussian random fields with a standard variation of 10. This task is similar to binary classification, although the scores here can take any value in \mathbb{R} . Figure 5 shows that the scores attributed by the trained critic to fake and real wind fields are clearly separated, so the critic performs correctly. The scores vary more for random wind fields, which could be interpreted as the network introducing uncertainty around the reliability of the classification in the presence of potentially fake winds.

e. Input data

The ERA5 inputs are grids of 12×24 pixels (25-km resolution), COSMO-1 targets are on 294×429 grids (1.1-km resolution), and topographic descriptors are constant on grids of size 3312×6912 (90-m resolution). To better control information reduction and expansion throughout the convolutional layers of the GAN, we created sequences of square patches as input for the model using the same geographical areas to create the inputs and the targets, while keeping the patch size constant. To do so we project all the inputs and outputs onto the COSMO-1 target grid. In particular, ERA5 inputs, topographic descriptors, and outputs are processed to reach 1.1-km grid resolution. The resolution of the ERA5 fields is artificially increased by filling the gaps with the nearest available value from the reanalysis dataset. Topographic predictors that have a higher spatial resolution than COSMO-1 fields are slightly blurred to meet the resolution standards.

Sequences of square patches of size S are created around random points in space and time for low-resolution inputs and corresponding high-resolution fields and are then randomly flipped or rotated before being input to the model. Data augmentation such as this has been found to enhance network efficiency and out-of-sample performance in other applications (Perez and Wang 2017).

4. Training

a. Adversarial training

Fitting a GAN can be difficult because the generator and discriminator networks may train at different speeds. For this

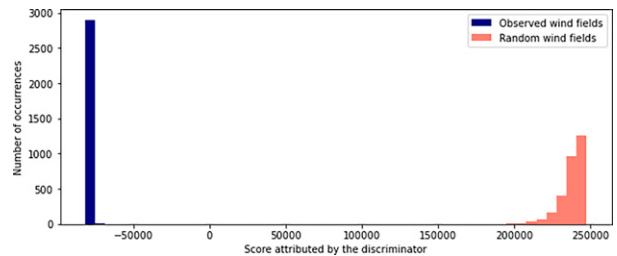


FIG. 5. Scores for 10-m wind fields predicted by the discriminator network presented in section 7. The score is a unit-free relative quality measure internal to the GAN and thus has no meaning in absolute terms.

study, the training was stabilized using spectral normalization that enforces a Lipschitz constraint on parameters in the convolution layers of both networks (Miyato et al. 2018), and different learning rates were chosen (Heusel et al. 2017). Indeed, a generator training against a poor discriminator does not generate better images because the discriminator cannot score them well. Typically, the learning rate of the discriminator is set 4 to 5 times above that of the generator to allow the scoring function to improve enough between successive updates of the generator. We further aided discriminator training by updating its network 3 times for each update of the generator to give the discriminator more time to process the change in the generator network (Gulrajani et al. 2017; see algorithm 1 given in appendix E). To avoid vanishing gradients the generator network includes not only split connections (Ronneberger et al. 2015; Srivastava et al. 2015), that is, shortcuts between deep layers, but also batch normalization layers (Santurkar et al. 2018) to normalize data across samples of a batch. For the same reason, the discriminator includes one split connection, and layer normalization of data aggregated on channels (Ba et al. 2016) was applied to the discriminator's convolutional layers. Normalization here should be understood as standardization, that is, transforming the data to have zero mean and unit standard deviation. The Adam optimizer, a stochastic gradient descent method based on adaptive estimation of first and second-order moments (Kingma and Ba 2014), was used with learning rates of 1×10^{-4} for the generator and 4×10^{-4} for the discriminator. The values for the first and second moment estimates $\beta_1 = 0.0$ and $\beta_2 = 0.9$ were derived from the calibration of the Adam optimizer for WGAN-GP by Gulrajani et al. (2017). Reconstruction loss was considered in this study to improve the training stability, by using an autoencoder to extract features from wind maps, but the results were more satisfactory with other techniques, such as layer normalization, split connections, and adjusting the optimizer hyperparameters. The small effect of inserting a reconstruction loss could be due to the very basic structure of the autoencoder, which we built ourselves for this study: efficiently extracting relevant features from the wind fields is a research project on its own. Moreover, the implementation of the GAN with a reconstruction loss had the undesired impact of keeping the prediction close to the original ERA5 pixelated style. Lowering the weight of the reconstruction

loss in the overall loss turned out to be equivalent to doing without reconstruction loss entirely.

b. Transfer learning

The GAN is trained using transfer learning (Bozinovski and Fulgosi 1976): after it is trained for one task, the learning curve for a similar task should be less steep and the training more efficient. Our downscaling problem is difficult for two main reasons. First, the difference in resolution between inputs and targets is large, as wind fields from ERA5 reanalysis data are available on 25-km grids, while COSMO-1 is on 1.1-km grids. Second, input and target winds come from two different sources, and discrepancies in modeling techniques make it more difficult for a network to understand how a high-resolution COSMO-1 field is linked to an ERA5 field than to an artificially blurred version of itself. In our case, no known transform of the high-resolution output data links it to the low-resolution predictors, so the network is first trained to downscale winds from artificially blurred COSMO-1 data to the high-resolution target wind fields, and then the training continues with low-resolution winds from the ERA5 reanalysis data.

c. Technical challenges

The MeteoSwiss data cover a period from April 2016 to October 2020, yielding 1673 days of hourly observations on grids of resolution 429×324 pixels. Our interest is in surface wind vectors with components u and v , so the number of individual data points to be predicted is about 10 billion. To capture daily patterns, we chose to train the GAN on 24-h sequences ($N_T = 24$) of square patches using two years of data (about 4.8 billion individual points). In the end, the only predictors inputted to the GAN are the low-resolution wind fields from COSMO-1 blurred data for the first training phase and the wind fields from ERA5 for the second training phase, mainly because the ratio of performance improvement to additional computational burden was too low when other predictors from ERA5 (see section 4b) were added. Using raw DEM as the only topographic predictor gave the best results. Small batches ($N_B = 8$) were chosen because we found that such microbatches stabilize the training. The generator contains 1.7 million parameters and the discriminator contains 3.3 million parameters, so the total number of parameters to be estimated is about 5 million. The training was done over about 200 epochs (training steps) on an Nvidia GPU with Volta microarchitecture provided by the EPFL Scientific IT and Application Support (SCITAS) system.

5. Metrics

The discriminator network scores the wind fields predicted by the generator according to its discrimination criteria by generating a model internal score function that is continuously improved during training. This function provides relative comparisons of the model at different training stages and cannot be interpreted in absolute terms. To gain a detailed understanding of the network's performance, we also use other metrics to monitor the training and the final results. The Fréchet inception distance is the most commonly used

metric to assess the performance of GANs (Heusel et al. 2017), but its implementation relies on the use of another neural network trained to recognize features that are friendly to human perception on static red–green–blue (RGB) images. This is irrelevant in our case because the two-dimensional field we aim to predict has only one channel per variable. Hence, we use standard metrics to assess the performance of image prediction, such as modified versions of the root-mean-squared error (RMSE), log spectral distance (LSD), angular cosine distance (ACD), and a spatially convolved version of the Kolmogorov–Smirnov statistic, which are detailed below.

a. RMSE

We use two versions of the original RMSE (Hyndman and Koehler 2006). Although weighting the metrics based on realized (COSMO-1) values rather than on predictions could bias model selection and validation (Lerch et al. 2017), these metrics were chosen to best meet the goal of the study, which is to provide accurate historical covariates for analysis of past weather and climate in which extreme winds might play a role as an aggravating factor. Although special attention was paid to extreme winds, other metrics and visual methods were also used to assess prediction reliability and thereby counterbalance any bias (Lerch et al. 2017).

The wind speed weighted RMSE (WSRMSE; Dujardin 2021) is defined as

$$\text{WSRMSE} = \sqrt{\frac{1}{N_T P} \sum_{t \leq N_T, i \leq P} \tau[(u_{it} - \beta \hat{u}_{it})^2 + (v_{it} - \beta \hat{v}_{it})^2]},$$

where N_T is the number of time steps, P the number of pixels in a single image, u and v are the 10-m high-resolution eastern and northern components of wind, \hat{u} and \hat{v} are their respective estimates, and

$$\beta = \frac{\epsilon + w}{\epsilon + \hat{w}}, \quad \tau = \begin{cases} t, & \hat{w} \geq w, \\ 1 - t, & \text{otherwise,} \end{cases}$$

where w is the 10-m high-resolution wind speed and \hat{w} is its estimate. The calibration of hyperparameters by Dujardin (2021) sets $\epsilon = 4$ and $t = 0.425$.

Another RMSE variant developed for this specific problem,

ExtremeRMSE

$$= \sqrt{\frac{1}{N_T P} \sum_{t \leq N_T, i \leq P} \frac{u_{it}^2}{\sum_{j,k} u_{jk}^2} (u_{it} - \hat{u}_{it})^2 + \frac{v_{it}^2}{\sum_{j,k} v_{jk}^2} (v_{it} - \hat{v}_{it})^2},$$

tries to condemn bad predictions of extreme components.

Both of these metrics put more weight on extreme winds, which explains their similar evolution during training (see Fig. 6). However, WSRMSE penalizes extremes in a relative sense (β becomes large when the wind speed is underestimated), whereas ExtremeRMSE directly puts component-wise weights that increase with the realized extremeness of each direction, whether or not the estimated component is also extreme.

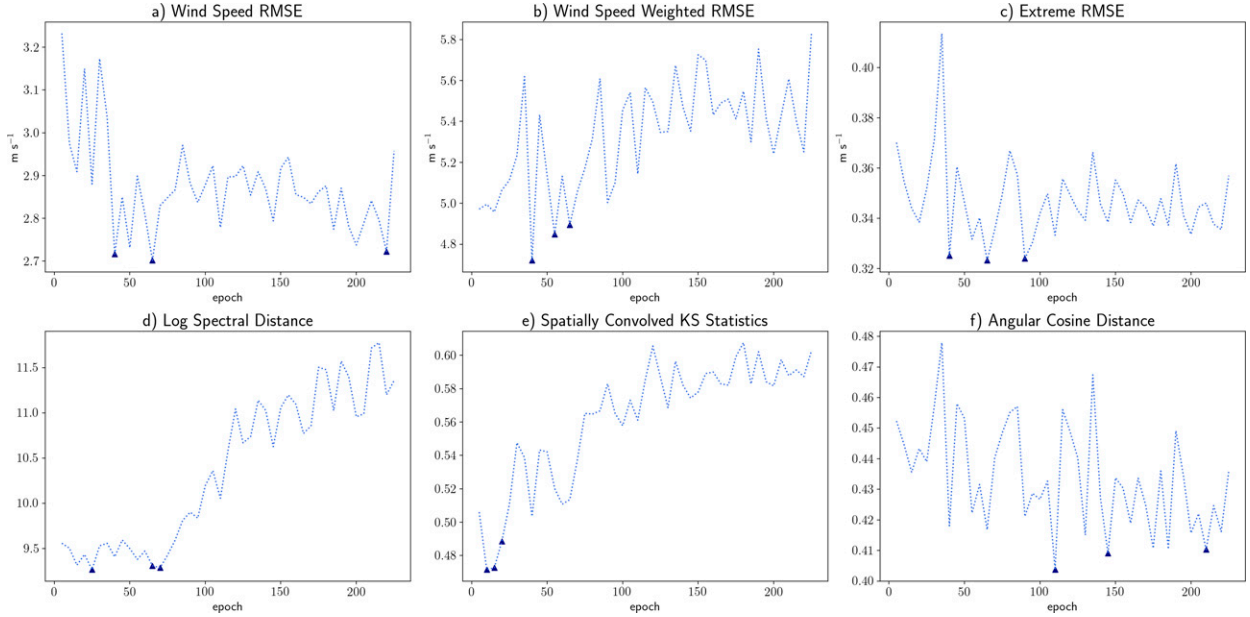


FIG. 6. Behavior of the different verification metrics throughout the training of the GAN. The x axis represents the epochs, with one epoch sampling patches of wind fields for all available dates. The three smallest values attained by each metric during training are denoted by triangles. Metrics are evaluated every day over random square patches in Switzerland on a 2-month validation set going from September to November 2019. Results are averaged over space and time.

b. ACD

The angular cosine distance (Foreman 2013) computes the average angle between the target and generated vectors,

$$\text{ACD} = \frac{1}{N_T P} \sum_{t \leq N_T, i \leq P} \arccos \left(\frac{u_u \hat{u}_u + v_u \hat{v}_u}{\sqrt{u_u^2 + v_u^2} \sqrt{\hat{u}_u^2 + \hat{v}_u^2}} \right),$$

and thus quantifies agreement between predicted and observed wind field directions. The ACD and the RMSE metrics complement each other, as ACD measures the performance of the network in terms of wind direction, whereas RMSE evaluates the distance between realized and predicted wind speed, which is the wind vector's norm. Both are needed for an accurate performance assessment.

c. Spatially convolved SKSS

This new metric, developed in the scope of this research, represents the disagreement between the distributions of the generated and observed wind fields. It is computed as the maximum absolute difference of empirical cumulative distribution functions for the generated and realized fields, summed over 10×10 patches of the image of interest. The aim is to obtain a metric with properties close to those of the Fréchet inception distance (Heusel et al. 2017) for RGB images by assessing the match between input wind fields and images produced by the generator, as a human eye would. Indeed, a GAN's performance can be hard to assess and visual checks of the generator's output may be preferred by users. The Kolmogorov–Smirnov statistic (SKSS) assesses whether the output is visually pleasing by checking whether the u and v

fields on small spatial patches look similar to those in the original image. First, M spatial patches of constant size are extracted from the target and predicted images. We then set

$$\text{SKSS} = \sum_{t \leq N_T} \sum_{j \leq M} \sum_{c \in \{u, v\}} \max_{x \in \mathbb{R}} |F_{jt}^c(x) - \hat{F}_{jt}^c(x)|,$$

where F_{jt}^c represents the empirical cumulative distribution function for a single spatial patch j , point in time t and channel c (u or v component of wind) and \hat{F}_{jt}^c is its analog for generated data. This metric is intended to evaluate the agreement between two local distributions rather than focusing on individual pixels.

d. LSD

The LSD metric (Rabiner and Juang 1993) is expressed as the log-difference of power spectra between the generated and realized samples,

$$\text{LSD} = \sqrt{\frac{1}{2N_T P} \sum_{t \leq N_T, i \leq P} \sum_{c \in \{u, v\}} \left[10 \log_{10} \left(\frac{|f(c_{it})|^2}{|f(\hat{c}_{it})|^2} \right) \right]^2},$$

where f is the Fourier transform, $|f(\cdot)|^2$ the power spectrum, c is the wind component and \hat{c} is its estimate. The LSD evaluates whether the generated images reproduce the spatial structures noticeable in the target images.

6. Results

The GAN described above is a stochastic model: predictions may vary with different samples of the input noise. In

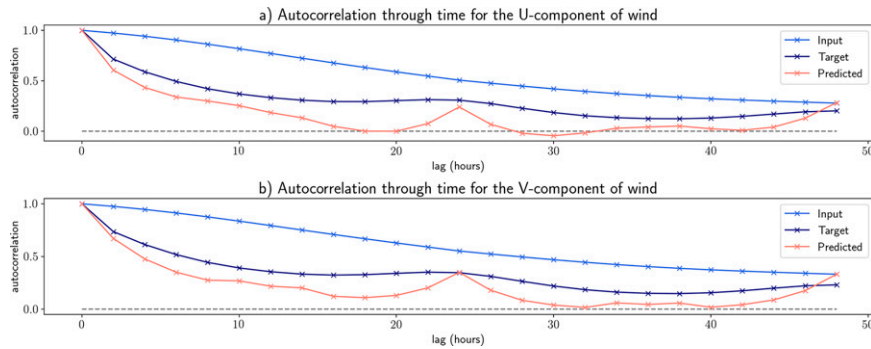


FIG. 7. Spatially and temporally averaged estimated autocorrelation for input, predicted, and target (a) u and (b) v wind components.

the following analysis of the results, the average prediction for the test set over 200 different noise samples is used to construct graphs and maps. As explained in section 6b, the network is trained in two phases that are evaluated separately below. Unless specified otherwise, the years 2016–18 were used for training, 2019 was used as the validation set, and 2020 was used as the test set for all results and plots.

a. Training phase I: Downscaling COSMO-1 blurred wind fields

1) MODEL SELECTION

The first step of the quantitative analysis entails model selection, as the network produces a different set of parameters for every training epoch, that is, every complete training round of the GAN on all wind fields. The best model must be selected to perform predictive analyses and make diagnostic plots. Metrics are expected to be nonmonotonic throughout training, as the generator and discriminator improve in an adversarial way. As the generator improves, it is more difficult for the discriminator to determine whether a given image is observed or predicted data, and is, therefore, more likely to attribute an incorrect score. As the discriminator's loss decreases, the classification becomes more accurate and the images produced by the generator are more likely to receive very positive scores, increasing the generator's loss. Figure 6 shows that the six metrics considered partially agree on the best-performing epoch, that is, the training step with the minimum value for a given metric. All three RMSE metrics (Figs. 6a–c) and the LSD (Fig. 6d) indicate a local minimum at epoch 55, while the local minima for the KS-statistic (Fig. 6e) occur at the very beginning of the training. The angular cosine distance (Fig. 6f) shows no clear minima. Epoch 55 is used to produce the diagnostic plots and computations that follow.

2) QUANTITATIVE ANALYSIS OF WIND TIME SERIES

The network is built to capture time series features in the target data. To evaluate the performance of the model, input, target and predicted autocorrelation are compared for lags from 2 h to 2 days in Fig. 7. The artificial blurring of the COSMO-1 wind fields (input) introduces additional autocorrelation that the network can successfully remove, as seen in

Fig. 7. However, the predicted wind components show lower autocorrelation for lags below 48 h than do the COSMO-1 data (target), with a marked increase for multiples of 24-h lags, perhaps because we process the wind field time series in 24-h sequences. Wind exhibits very specific subdaily patterns that vary with the topography. To evaluate whether these are well captured by the network, Fig. B1 of appendix B shows the subdaily wind variability averaged over the validation set for locations in valleys, plains, and on mountaintops. The GAN can accurately reproduce average daily patterns for both u and v components in relatively flat zones, for example, Fahy lies in the Jura and Aigle in a valley in the Alps (Figs. B1a,b), and Bischofszell and Fribourg are in the Swiss Plateau (Figs. B1c,d). In extremely complex terrain (Zermatt and Jungfrauoch are located on mountain passes), the model does not capture the subdaily pattern of the u component (Figs. B1e,f) well.

3) VISUALIZATION OF HISTORICAL WIND MAPS

The GAN is trained on square patches of size $S = 96$ pixels, while COSMO-1 maps of Switzerland have size 429×324 (section 4c). We combine the patches to predict entire wind fields. One possibility would be to crop the initial COSMO-1 map into a 384×288 map focusing on Switzerland only, predict a grid of 4×3 patches of size S and accept that there will be a discontinuity at their borders, but as our goal is to create realistic-looking historical wind fields we prefer to predict a grid of overlapping 5×4 patches and average them to give smoother borders. Maps of target and predicted u and v components of wind averaged over the 1-yr test period are displayed in Fig. 8. Specific patterns, such as local acceleration at exposed sites (ridges) and sheltering in valleys, are very well reproduced by the network for both u and v components in all three subregions of Switzerland. Both COSMO-1 (Fig. 8c) and the predictions (Fig. 8d) show strong regional patterns of the mean v wind direction depending on the topography, with southerly winds on north-facing slopes and northerly winds on south-facing slopes. This is probably the fingerprint of the foehn, an intense, warm, and dry downslope windstorm that occurs frequently on both the northern and southern sides of the Alps (Richner and Hächler 2013). Appendixes C and D contain examples of hourly maps produced from blurred

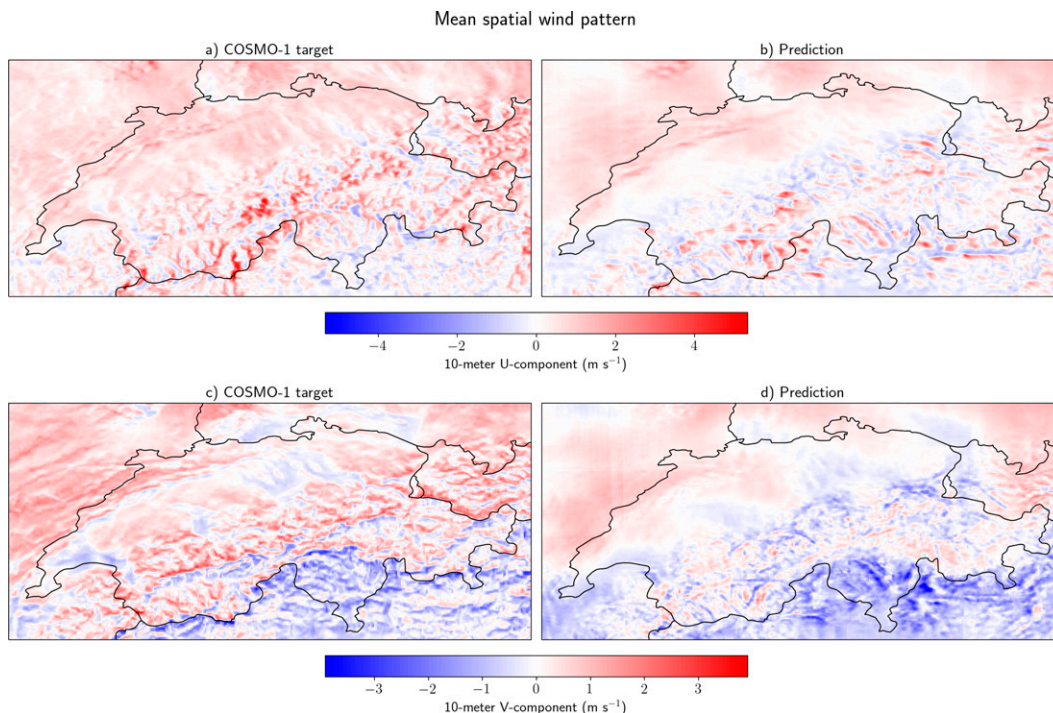


FIG. 8. (left) Target and (right) predicted wind components (a),(b) u and (c),(d) v averaged over the test period of 1 year in 2020.

COSMO-1 after the first training phase (Fig. C1 in appendix C) and from ERA5 low-resolution inputs after the second training phase (Fig. D1 in appendix D).

The spatial quality of the predicted wind fields is evaluated by plotting the median values for WSRMSE and ACD, computed for the test year 2020 (Fig. 9). The former was made unitless by applying a hyperbolic tangent transform to facilitate interpretation. Predictions for wind speed (Fig. 9a) and direction (Fig. 9b) are good in the Swiss Plateau and Jura. Differences in wind direction occur in valleys and at the feet of mountains, while wind direction is well predicted on upper slopes and ridges (Fig. 9b). In the Swiss Alps, where the terrain is more complex, the wind speed is predicted less well at

the high-wind exposed mountain sites (ridges) than at the sheltered valley sites (Fig. 9a).

b. Training phase 2: Downscaling ERA5 wind fields

In the second phase, the training continues and the parameters of the best model of the first training phase are used as initial parameter values. When the second training phase is completed, the steps of the first training phase are repeated to find the epoch with the best performance. Maps of wind fields from target and predicted from ERA5 averaged over the 1-yr test period are shown in Fig. 10 and median values of the metrics WSRMSE and ACD are shown in Fig. 11. The prediction

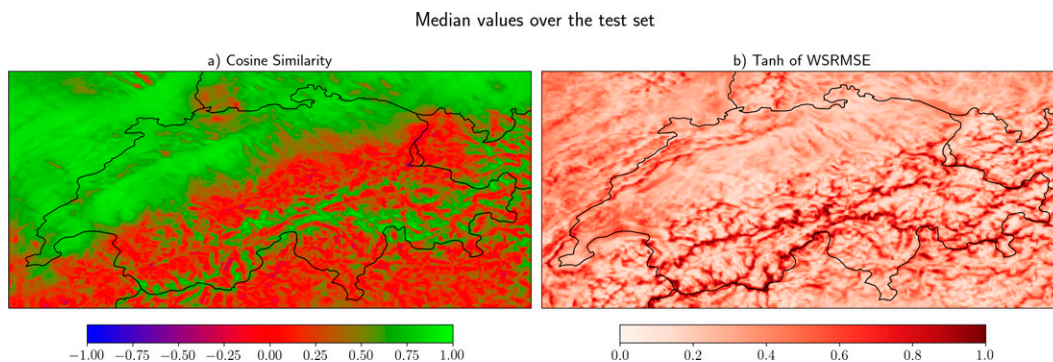


FIG. 9. Visualization of GAN performance after the first phase of training: (a) median cosine similarity (1 is perfect and -1 is bad) and (b) median wind speed weighted RMSE (WSRMSE) after a bounded transform (0 is perfect and 1 is bad).

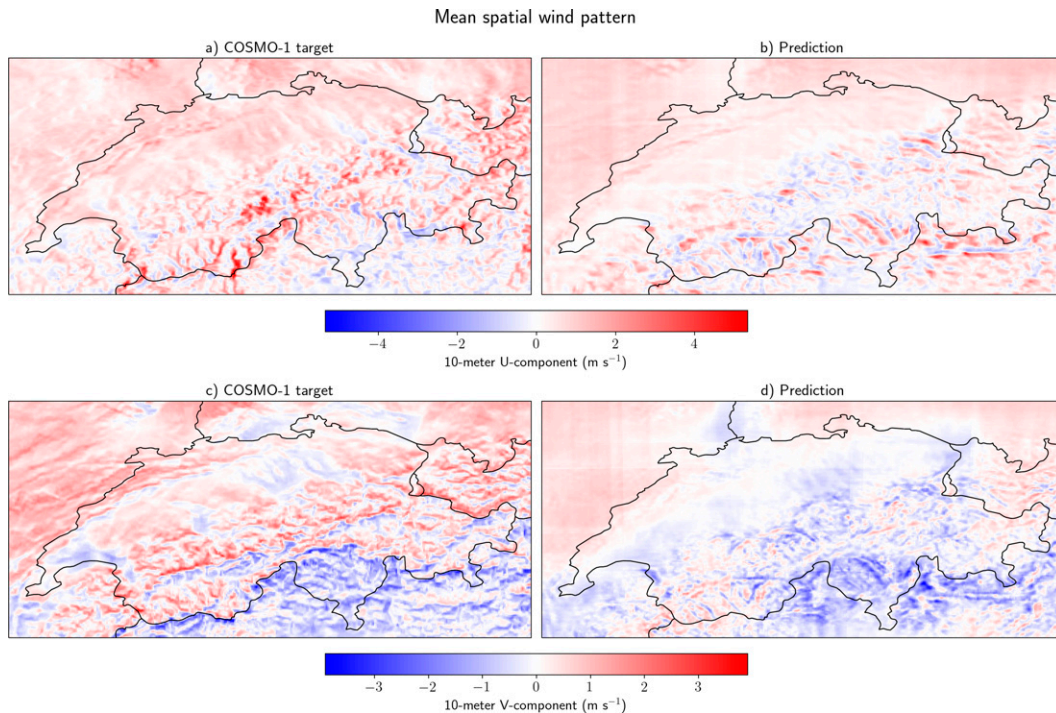


FIG. 10. (left) Target and (right) predicted (from ERA5) (a),(c) u and (b),(d) v components of wind fields averaged over the test period (1 year).

(Figs. 10b and 10d) reproduces the high-resolution mean wind pattern of the COSMO-1 target (Figs. 10a and 10c) with stronger westerly winds at the exposed mountain sites in the Alps and Jura and weaker or easterly winds at the sheltered valley sites. Looking at the mean wind direction, we see that more regions have northerly and easterly winds in comparison with COSMO-1. These differences in wind direction between prediction and COSMO-1 are also seen in the median cosine similarity in Fig. 11b and occur over the entire Alps. The wind speed is predicted less well at the high-wind exposed mountain sites (ridges) than at the sheltered valley sites (Fig. 11a). Results from the first (Fig. 9) and the second (Fig. 11) training phase using ACD and WSRMSE show better predictive performance when downscaling from COSMO-1 blurred inputs. Indeed, the Gaussian filter used to create low-resolution winds

from the COSMO-1 high-resolution target produces smoother maps than those created with ERA5 winds, which may facilitate pattern detection.

The ultimate goal of this project is to build accurate historical wind fields for the analysis of specific extreme events, such as windstorms and the role of foehn in forest fires, so the main question addressed in the diagnostics for the second training phase is whether the network accurately represents extreme wind speeds. Strong winds during storms are known to cause damage in populated regions at lower altitudes (Swiss Plateau and valleys) (Schwierz et al. 2010; Welker et al. 2016) and are important in spreading forest fires (Sun et al. 2009; Sharples et al. 2012; Cruz et al. 2020), so we desire them to be accurately downscaled from the low-resolution fields of ERA5.

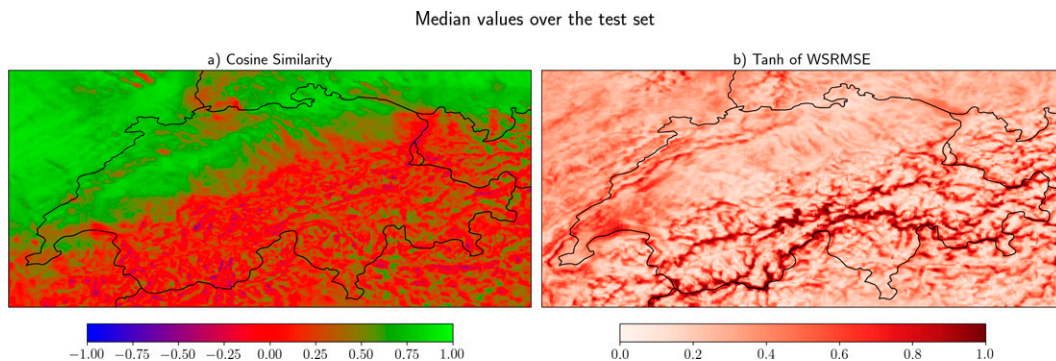


FIG. 11. As in Fig. 9, but after the second phase of training.

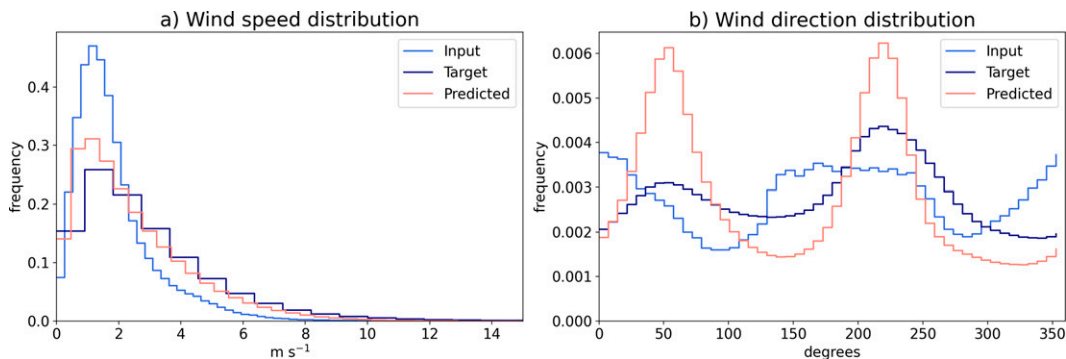


FIG. 12. Estimated distribution of input (ERA5), target (COSMO-1), and predicted (a) wind speed and (b) wind direction.

Speed and direction distributions for predicted (from ERA5) and target winds are displayed and compared in Fig. 12. We consider the wind speed computed from the mean predicted maps of u and v components rather than the mean wind speed over the noise samples. The predictions of the average wind speed distribution are very close to the target distribution, although the predicted wind speed is slightly less long tailed than that for COSMO-1 (Fig. 12a). The network underestimates very high wind speeds and predicts more southwestern (0° – 90°) and northeastern (180° – 270°) winds than are observed in COSMO-1 (Fig. 12b). The wind speed distribution of ERA5 winds shows a much thinner tail than the target and predicted wind speed (Fig. 12a), whereas the ERA5 wind direction seems completely out of sync in comparison with the target and predicted distributions (Fig. 12b).

Further analysis of wind speed predictions within 10 km around the largest Swiss cities is encouraging. Figure 13 shows that the extremes of wind speed are well captured by the model, especially for cities in the Swiss Plateau: Zürich (Fig. 13a, ~400 000 inhabitants), Lausanne (Fig. 13d, ~135 000 inhabitants), Winterthur (Fig. 13f, ~108 000 inhabitants), and Luzern (Fig. 13g, ~81 000 inhabitants).

How much do the downscaled wind fields improve on the original ERA5 fields? We compare the metrics WSRMSE, LSD, and SKSS, averaged over the test set, between predicted winds and ERA5 winds in Figs. 14b–d. Each point represents the average at a specific grid cell and is colored according to the subregion shown in Fig. 14a to highlight regional differences. The diagonal line ($x = y$) represents identical values of the metric before and after prediction by the network. Average values of the metrics for each geographical subregion, given in Table 1, show that for all regions the LSD is much smaller when comparing predicted winds with the target COSMO-1 than for the ERA5 inputs. The SKSS is smaller for predicted winds than for ERA5 inputs in the Alps but slightly higher at some points on the Swiss Plateau, maybe because ERA5 predicts the winds sufficiently well on homogeneous and flat zones, while the GAN could add artifacts, that is, undesired signals with nonphysical origins, at these locations. There is no reduction in WSRMSE, which is essentially preserved by the GAN. This is expected because pointwise comparisons do not need to detect the visual improvements highlighted by LSD

and SKSS. Table 1 shows clear improvements using the GAN prediction instead of ERA5 winds in Alpine regions (Alpes Valaisannes, Alpes Vaudoises, Alpi Lepontine, Alpi Retiche, Berner Alpen, Glarner Alpen, and Urner Alpen), especially for LSD and SKSS. On the Swiss Plateau (Mittelland) and in the Jura, a sharp decrease of the LSD can be noted, while the SKSS is preserved.

7. Conclusions

In this paper, we developed and applied a deep learning model for downscaling hourly near-surface gridded wind fields in complex terrain using low-resolution (~ 25 km) inputs from ERA5 reanalysis and high-resolution (1.1 km) targets from the numerical weather prediction model COSMO-1. Topographic information from high-resolution (90 m) digital elevation data from the SRTM3 was used as a static input to incorporate local orographic effects that modify airflow. A Wasserstein recurrent generative adversarial network (GAN) with a gradient penalty architecture was chosen with an autoencoder-like structure for the upsampling part of the generator. Adapted normalization of layer outputs was introduced in both networks, and weight normalization was used to speed up and smooth the training. Careful attention was paid to coordinating the networks so that neither became too strong for the other to train, and a good balance was achieved by using different learning rates and updating the discriminator more frequently than the generator. Due to the complexity of the problem, an approach based on transfer learning was chosen to train the GAN. Segmentation of the learning curve greatly improved the performance of the network relative to a more direct approach. This appears to be the first deep learning model trained using transfer learning that can efficiently perform such an extreme ($25\times$, from 25 to 1.1 km) downscaling of wind fields from two different data sources. Its performance was tested over the complex terrain of Switzerland for the period 2016–20, but if the local topography is available, our model could be applied to wind fields elsewhere in order to generate long-term, high-resolution wind climatologies.

Historical maps, created for all of Switzerland using overlapping patches of predicted wind fields, were visually appealing for both training phases, with maps predicted from either blurred COSMO-1 or ERA5 inputs consistently resembling

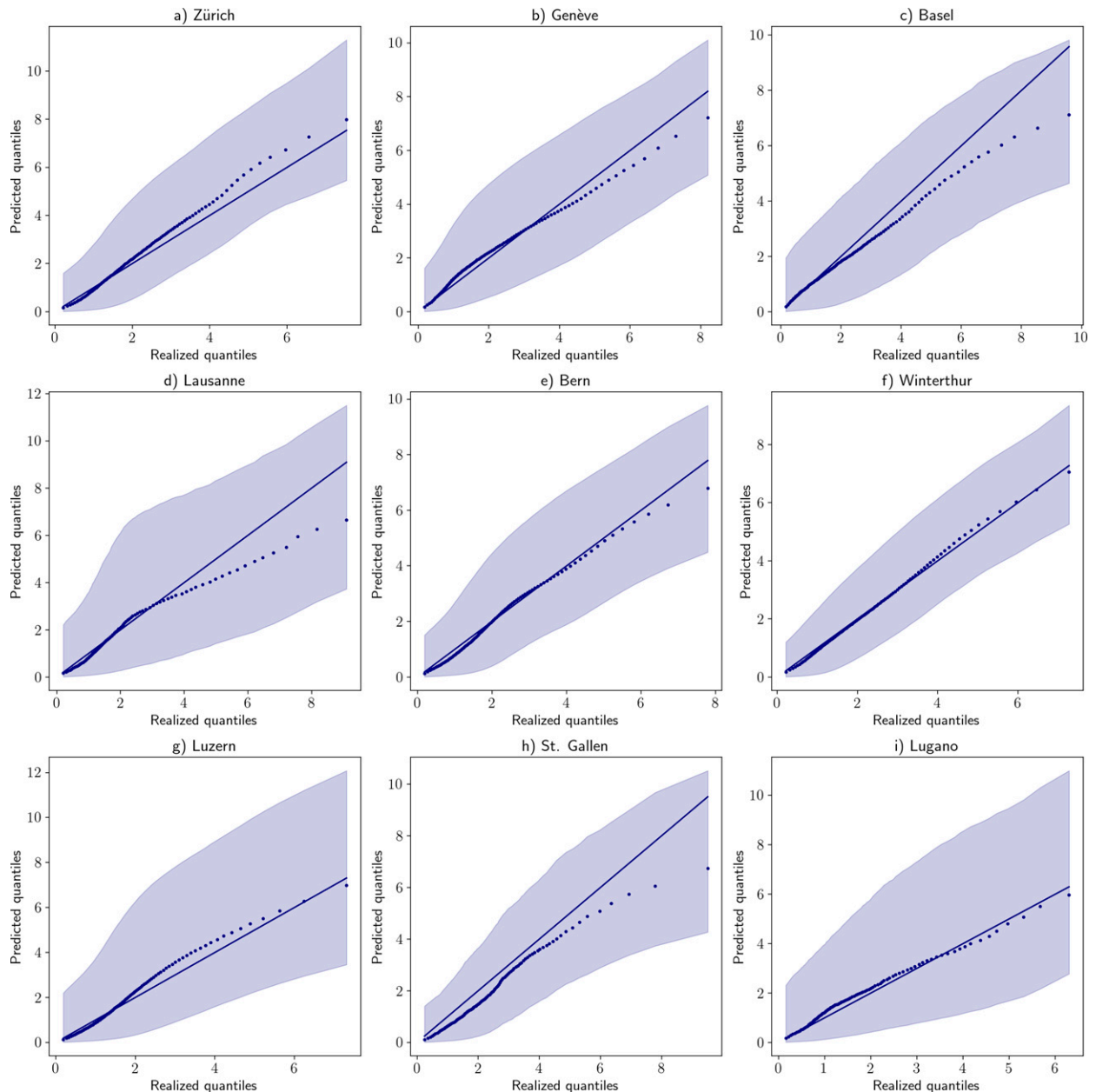


FIG. 13. Q–Q plots of predicted (using ERA5 as input) vs target (COSMO-1) quantiles of the wind speed distribution within 10 km of the largest Swiss cities: (a) Zürich, (b) Genève, (c) Basel, (d) Lausanne, (e) Bern, (f) Winterthur, (g) Luzern, (h) St. Gallen, and (i) Lugano. Shaded areas are plotted between quantiles computed from the distribution of the minimum and maximum wind speed across 200 different noise samples given as inputs to the GAN. The diagonal line represents $x = y$.

the COSMO-1 target. As our goal was to produce wind field predictions for the analysis of historical extreme weather events and long-term climatology, the second phase of training diagnostics focused on the wind speed distribution to verify how well extreme values were captured. The findings indicate an excellent prediction of the aggregated wind speed distribution around densely populated areas. Quantitative analysis of time series and spatial averages after the first training phase showed that the network missed some

autocorrelations and that there were differences in the predictive performance between flat and mountainous regions. Wind speed and mean daily patterns were less well predicted in high altitudes of the mountainous terrain than in the hilly plains and valleys. While wind direction was well predicted on mountaintops, the network struggled when predicting wind direction in valleys.

Because most issues stem from differences in topography, the global architecture could be improved by building different

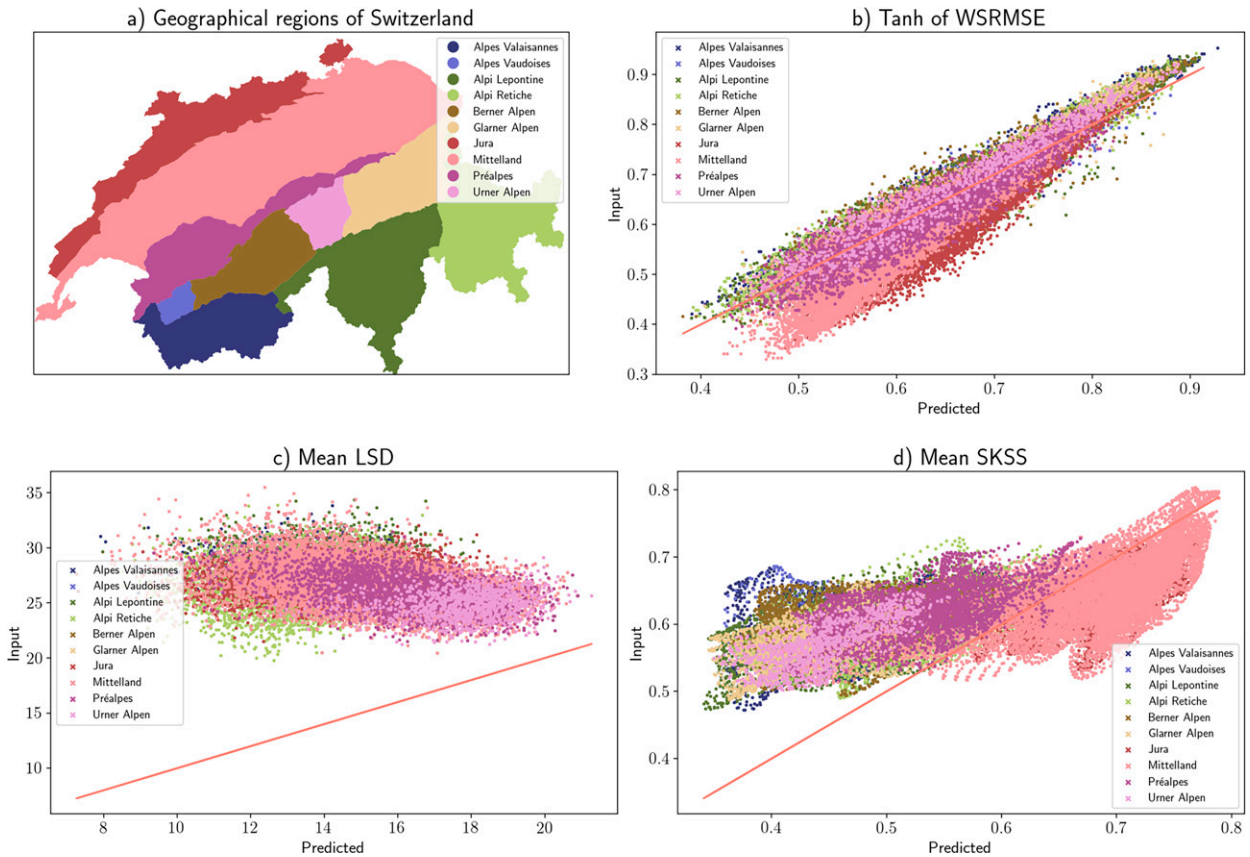


FIG. 14. Before-to-after regional comparison between ERA5 inputs (y axis) and predicted high-resolution winds (x axis) averaged over the test set. (a) The geographical regions, and results for the (b) WSRMSE, (c) LSD, and (d) SKSS metrics.

models for predicting patches that are mainly over mountainous areas or over the Swiss Plateau, rather than using a single network for the entire country. Another deep structure trained for image classification based on topography could use an input sensor to select which of those two networks should be applied. Different parameters would thus be used to predict winds in plains and in complex terrain, perhaps leading to less topographic variation in model performance. The training of such a

structure would require more time and resources but might overcome most remaining issues with our model.

Acknowledgments. We thank the Federal Office of Meteorology and Climatology (MeteoSwiss) for providing the COSMO-1 analysis data. We are grateful to Daniele Nerini (MeteoSwiss) and Jérôme Dujardin (EPFL) for fruitful discussions. Author Ophélie Miralles acknowledges funding from the

TABLE 1. Before-to-after comparison of the regional log spectral distance, the hyperbolic tangent of WSRMSE, and SKSS averaged over the time dimension of the test set.

Geographical region	LSD			Tanh WSRMSE			SKSS		
	Input	Predicted	Percent variation	Input	Predicted	Percent variation	Input	Predicted	Percent variation
Alpes Valaisannes	28.04	13.98	−50.2	0.69	0.68	−2.3	0.77	0.69	−10.5
Alpes Vaudoises	27.98	13.96	−50.1	0.69	0.68	−1.9	0.78	0.68	−13.1
Alpi Lepontine	26.86	14.97	−44.3	0.66	0.66	−0.9	0.77	0.69	−9.9
Alpi Retiche	26.60	13.12	−50.7	0.67	0.66	−1.5	0.77	0.71	−7.8
Berner Alpen	25.48	16.82	−34.0	0.69	0.67	−2.7	0.78	0.70	−10.6
Glarner Alpen	25.74	16.35	−36.5	0.66	0.65	−0.8	0.76	0.68	−11.0
Jura	27.06	14.21	−47.5	0.59	0.64	10.2	0.80	0.82	3.3
Mittelland	26.41	15.56	−41.1	0.56	0.60	6.8	0.80	0.83	3.0
Préalpes	25.78	16.42	−36.3	0.62	0.63	0.2	0.79	0.73	−7.4
Urner Alpen	24.77	17.76	−28.3	0.68	0.66	−2.6	0.76	0.68	−10.6

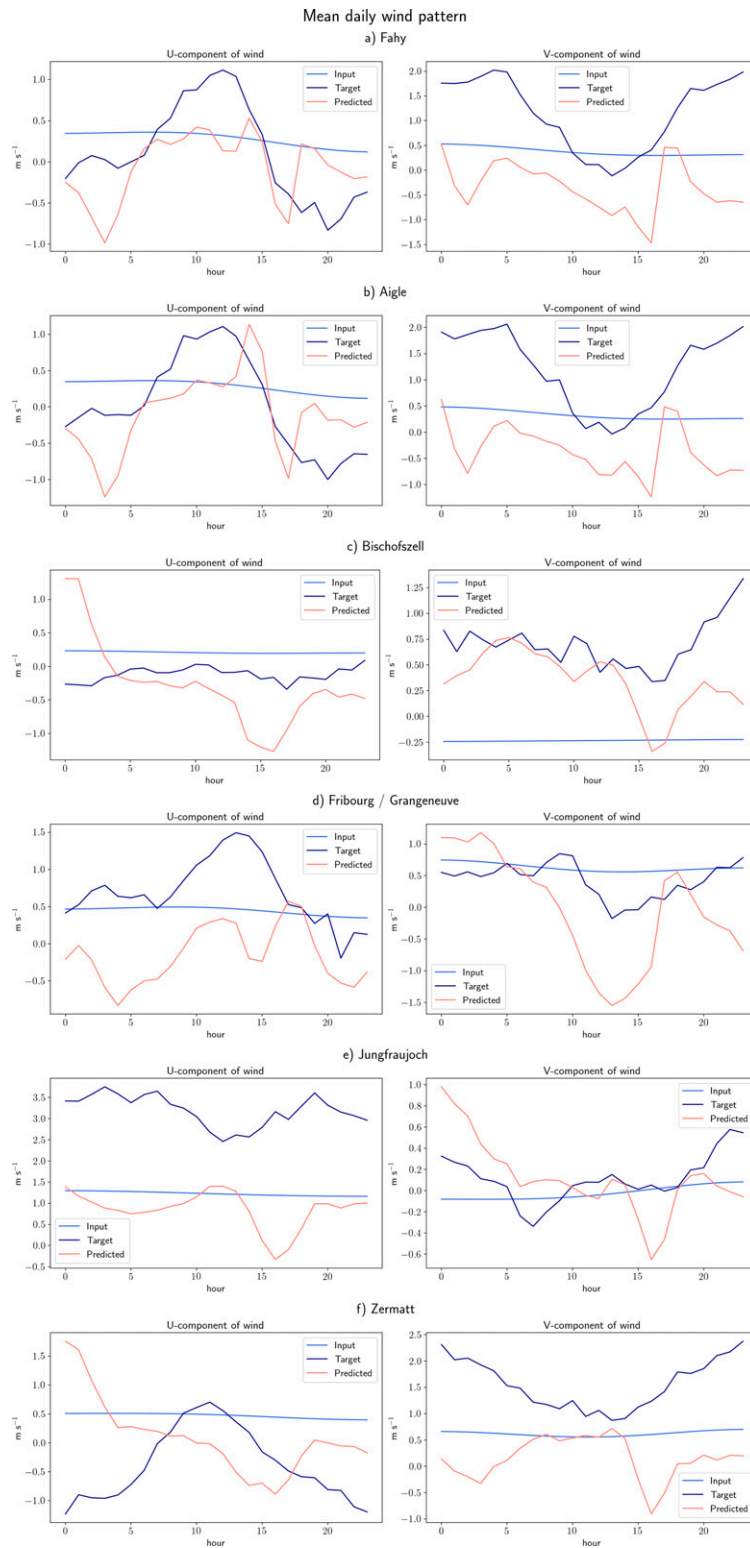


FIG. B1. Mean daily pattern for (left) u and (right) v wind components (a),(b) in the Jura, (c),(d) on the Swiss Plateau, and (e),(f) in mountainous areas averaged over time. The locations of the validation sites are shown in Fig. 1.

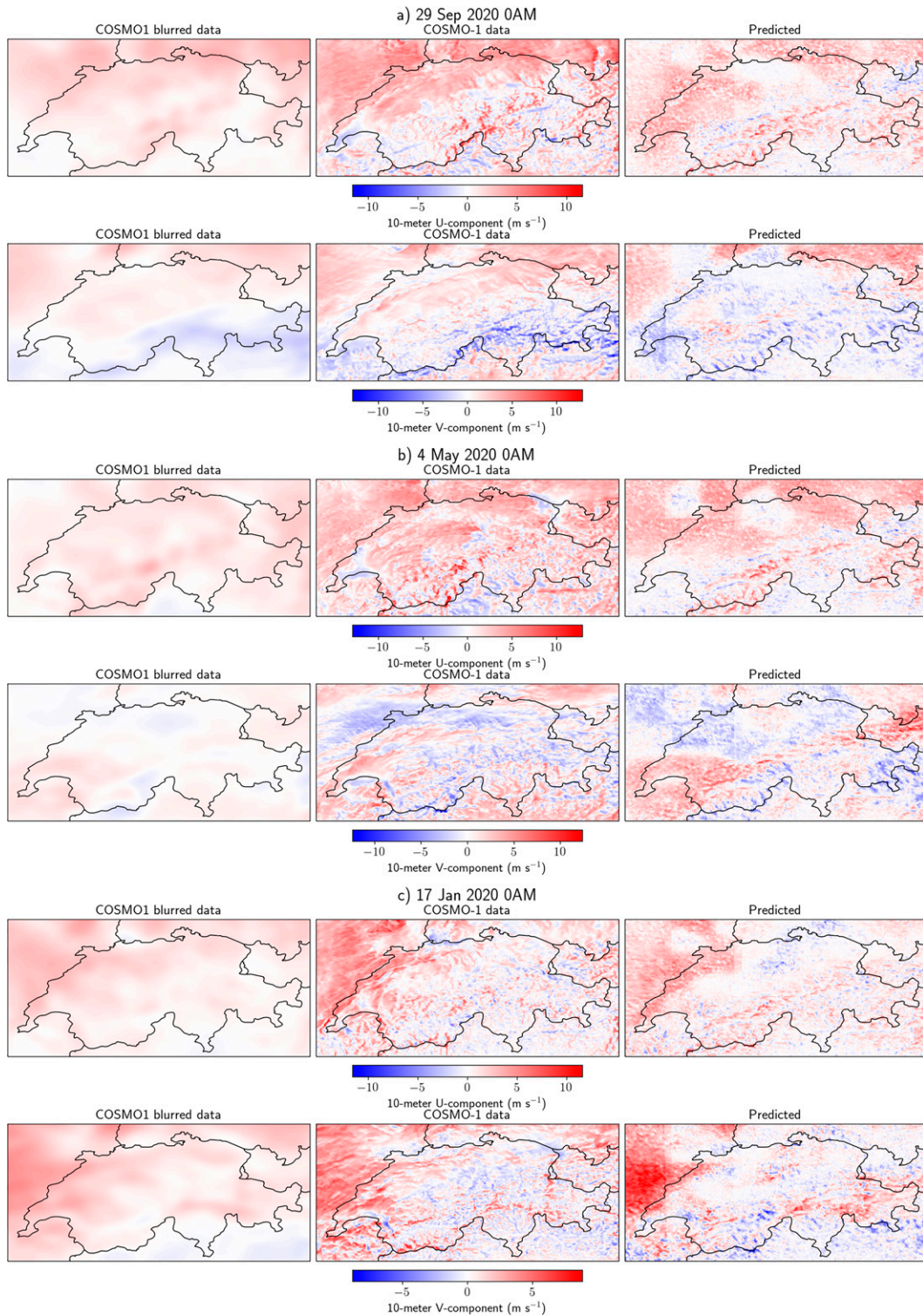


FIG. C1. Prediction of the u and v components of 10-m wind field by the GAN presented in section 3 for three different dates, with (left) inputs from COSMO-1 blurred, (center) the outputs from the COSMO-1 model at 1.1-km resolution, and (right) the model prediction.

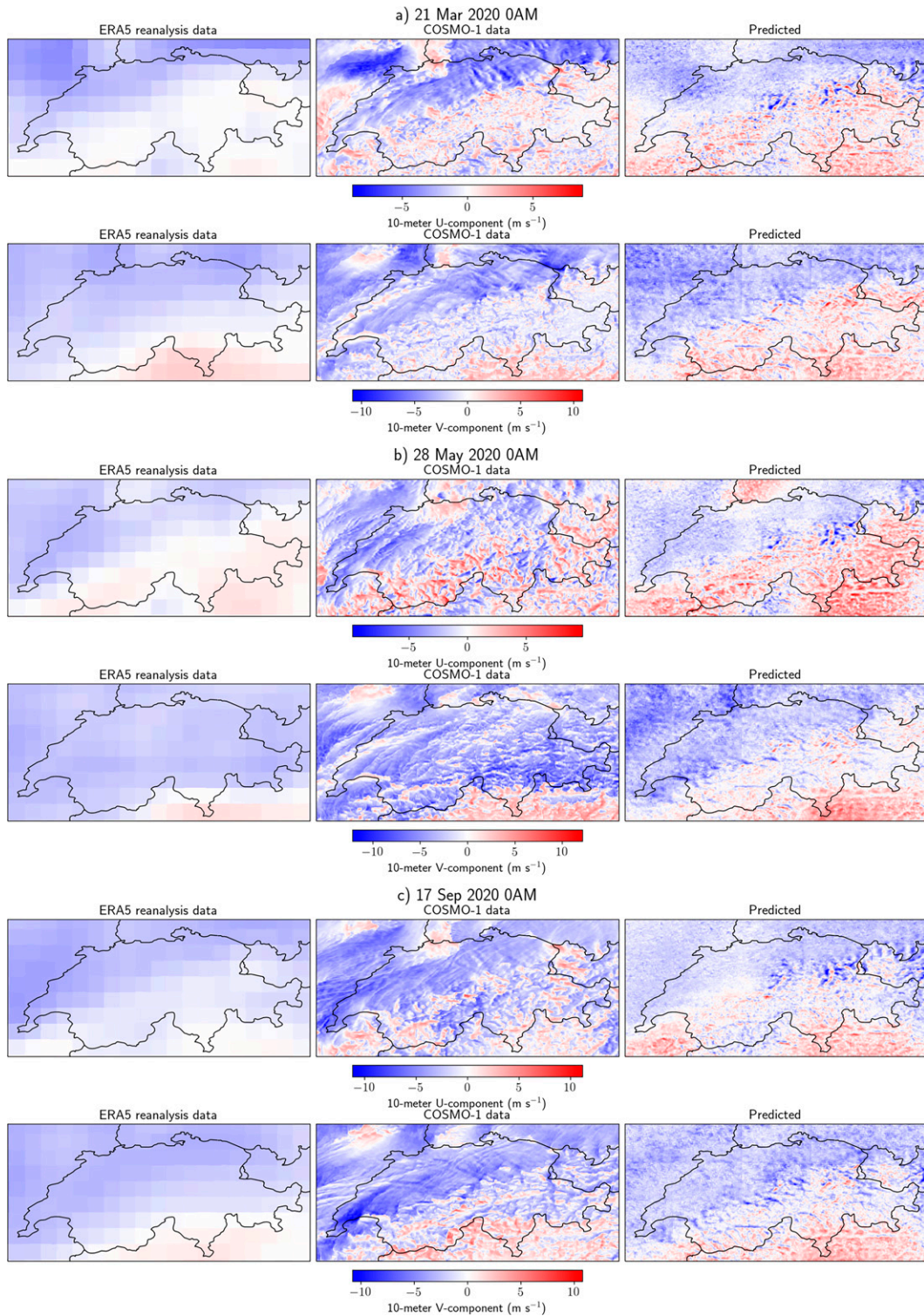


FIG. D1. Prediction of the u and v components of 10-m wind field by the GAN presented in section 3 for three different dates, with (left) inputs from ERA5 25-km-resolution grids, (center) the outputs from the COSMO-1 model at 1.1-km resolution, and (right) the model prediction.

APPENDIX B

Examples of Wind Mean Daily Patterns from COSMO-1 Blurred Test Sample

Figure B1 shows the mean daily pattern for u and v wind components for validation sites in the Jura, on the Swiss Plateau, and in mountainous areas, averaged over time.

APPENDIX C

Examples of GAN Prediction from COSMO-1 Blurred Test Sample

Figure C1 shows an example of hourly maps produced from blurred COSMO-1 after the first training phase, showing the prediction of the u and v components of the 10-m wind field by the GAN presented in section 3.

APPENDIX D

Examples of GAN Prediction from ERA5 Test Sample

Figure D1 shows an example of hourly maps produced from ERA5 low-resolution inputs after the second training phase, showing the prediction of the u and v components of the 10-m wind field by the GAN presented in section 3.

APPENDIX E

Algorithm 1

Algorithm 1 implements WGAN-GP with different update rates for the generator and discriminator networks, as proposed by Gulrajani et al. (2017). Here, θ and w represent respectively the generator and discriminator's parameters throughout training. The steps are as follow.

Require: For data processing: batch size N_B , time steps N_T , patch size S , and number of predictors N_P . The number of noise channels N_N is also required.

Require: Learning rates α_G and α_D for the generator and discriminator networks, optimizer hyperparameters $\beta_1 = 0$ and $\beta_2 = 0.9$, and the number of consecutive discriminator updates $n_{\text{critic}} = 3$ for one generator update.

Require: Initial discriminator parameters w_0 and initial generator parameters θ_0 .

while θ has not converged **do**

for $t = 1, \dots, n_{\text{critic}}$ **do**

Create inputs batches x (of size $N_B \times N_T \times S \times S \times N_P$) and corresponding target batches y ($N_B \times N_T \times S \times S \times 2$)

Sample noise $z \sim \mathcal{N}(0, 10^{-2})$ of size $N_B \times N_T \times S \times S \times N_N$

Sample a random number $\epsilon \sim \mathcal{U}(0, 1)$

$\hat{y} \leftarrow G_\theta(x, z)$

$\tilde{y} \leftarrow \epsilon y + (1 - \epsilon)\hat{y}$

$\text{Loss}_D \leftarrow D(x, y) - D(x, \hat{y}) + \gamma[\|\nabla_y D\{x, \tilde{y}\}\|_2 - 1]^2$

$w \leftarrow \text{Adam}(\nabla_w \text{Loss}_D, \alpha_D, \beta_1, \beta_2)$

end for

Sample noise $z \sim \mathcal{N}(0, 10^{-2})$ of size $N_B \times N_T \times S \times S \times N_N$
 $\theta \leftarrow \text{Adam}[\nabla_\theta D\{x, G_\theta(x, z)\}, \alpha_G, \beta_1, \beta_2]$
end while

REFERENCES

- Ba, J. L., J. R. Kiros, and G. E. Hinton, 2016: Layer normalization. arXiv, 1607.06450v1, <https://doi.org/10.48550/arXiv.1607.06450>.
- Baño-Medina, J., R. Manzananas, and J. M. Gutiérrez, 2020: Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geosci. Model Dev.*, **13**, 2109–2124, <https://doi.org/10.5194/gmd-13-2109-2020>.
- Barry, R. G., 2008: *Mountain Weather and Climate*. 3rd ed. Cambridge University Press, 532 pp., <https://doi.org/10.1017/CBO9780511754753>.
- Bozinovski, S., and A. Fulgosi, 1976: The influence of pattern similarity and transfer learning upon training of a base perceptron B2. arXiv, 2110.02879v1, <https://arxiv.org/pdf/2110.02879.pdf>.
- Castro-Camilo, D., R. Huser, and H. Rue, 2019: A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting. *J. Agric. Biol. Environ. Stat.*, **24**, 517–534, <https://doi.org/10.1007/s13253-019-00369-z>.
- Consortium for Small-scale Modeling, 2017: COSMO model. Accessed 15 December 2021, <http://www.cosmo-model.org/>.
- Cruz, M. G., M. E. Alexander, P. M. Fernandes, M. Kilinc, and Á. Sil, 2020: Evaluating the 10% wind speed rule of thumb for estimating a wildfire's forward rate of spread against an extensive independent set of observations. *Environ. Modell. Software*, **133**, 104818, <https://doi.org/10.1016/j.envsoft.2020.104818>.
- Dörenkämper, M., and Coauthors, 2020: The making of the New European Wind Atlas—Part 2: Production and evaluation. *Geosci. Model Dev.*, **13**, 5079–5102, <https://doi.org/10.5194/gmd-13-5079-2020>.
- Dujardin, J. F. S., 2021: The complex winds of the Alps: An unseen asset for the energy transition. EPFL Thesis, 224 pp., <https://doi.org/10.5075/epfl-thesis-9253>.
- , A. Kahl, and M. Lehning, 2021: Synergistic optimization of renewable energy installations through evolution strategy. *Environ. Res. Lett.*, **16**, 064016, <https://doi.org/10.1088/1748-9326/abfc75>.
- Emeis, S., 2014: Current issues in wind energy meteorology. *Meteor. Appl.*, **21**, 803–819, <https://doi.org/10.1002/met.1472>.
- Foreman, J. W., 2013: *Data Smart: Using Data Science to Transform Information into Insight*. John Wiley and Sons, 432 pp.
- Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep Learning*. MIT Press, 787 pp.
- Graf, M., S. C. Scherrer, C. Schwierz, M. Begert, O. Martius, C. C. Raible, and S. Brönnimann, 2019: Near-surface mean wind in Switzerland: Climatology, climate model evaluation and future scenarios. *Int. J. Climatol.*, **39**, 4798–4810, <https://doi.org/10.1002/joc.6108>.
- Gulrajani, I., F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, 2017: Improved training of Wasserstein GANs. arXiv, 1704.00028v3, <https://doi.org/10.48550/arXiv.1704.00028>.
- Gutowski, W. J., Jr., and Coauthors, 2016: WCRP coordinated regional downscaling experiment (CORDEX): A diagnostic MIP for CMIP6. *Geosci. Model Dev.*, **9**, 4087–4095, <https://doi.org/10.5194/gmd-9-4087-2016>.

- Harris, I., T. J. Osborn, P. Jones, and D. Lister, 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data*, **7**, 109, <https://doi.org/10.1038/s41597-020-0453-3>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Heusel, M., H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, 2017: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. arXiv, 1706.08500v6, <https://doi.org/10.48550/arXiv.1706.08500>.
- Hochreiter, S., and J. Schmidhuber, 1997: Long short-term memory. *Neural Comput.*, **9**, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.
- 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>.
- Huang, G., Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, 2016: Deep networks with stochastic depth. *Computer Vision—ECCV 2016*, B. Leibe et al., Eds., Lecture Notes in Computer Science, Vol. 9908, Springer International Publishing, 646–661.
- Hyndman, R. J., and A. B. Koehler, 2006: Another look at measures of forecast accuracy. *Int. J. Forecasting*, **22**, 679–688, <https://doi.org/10.1016/j.ijforecast.2006.03.001>.
- Jackson, P. L., G. Mayr, and S. Vosper, 2013: Dynamically-driven winds. *Mountain Weather Research and Forecasting*, F. Chow, S. De Wekker, and B. Snyder, Eds., Springer Atmospheric Sciences, Springer, 121–218, https://doi.org/10.1007/978-94-007-4098-3_3.
- Jarvis, A., H. Reuter, A. Nelson, and E. Guevara, 2008: Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture, accessed September 2021, <http://srtm.csi.cgiar.org>.
- Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. arXiv, 1412.6980v9, <https://doi.org/10.48550/arXiv.1412.6980>.
- Koller, S., and T. Humar, 2016: Windpotentialanalyse für windatlas.ch: Jahresmittelwerte der modellierten windgeschwindigkeit und windrichtung (Wind potential analysis for windatlas.ch: Annual mean values of the modeled wind speed and wind direction). Bundesamt für Energie Final Rep., 29 pp., <https://pubdb.bfe.admin.ch/de/publication/download/8302>.
- Krutz, B., M. Lehning, and A. Kahl, 2017: Potential contributions of wind power to a stable and highly renewable Swiss power supply. *Appl. Energy*, **192**, 1–11, <https://doi.org/10.1016/j.apenergy.2017.01.085>.
- , J. Dujardin, and M. Lehning, 2018: Improvement of wind power assessment in complex terrain: The case of COSMO-1 in the Swiss Alps. *Front. Energy Res.*, **6**, 102, <https://doi.org/10.3389/fenrg.2018.00102>.
- Lawrence, N. D., 2003: Gaussian process latent variable models for visualisation of high dimensional data. *NIPS'03: Proceedings of the 16th International Conference on Neural Information Processing Systems*, S. Thrun, L. K. Saul, and B. Schölkopf, Eds., MIT Press, 329–336.
- Lehning, M., H. Löwe, M. Ryser, and N. Raderschall, 2008: Inhomogeneous precipitation distribution and snow transport in steep terrain. *Water Resour. Res.*, **44**, W07404, <https://doi.org/10.1029/2007WR006545>.
- Leinonen, J., D. Nerini, and A. Berne, 2020: Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Trans. Geosci. Remote Sens.*, **59**, 7211–7223, <https://doi.org/10.1109/TGRS.2020.3032790>.
- Lerch, S., T. L. Thorarinsdottir, F. Ravazzolo, and T. Gneiting, 2017: Forecaster's dilemma: Extreme events and forecast evaluation. *Stat. Sci.*, **32**, 106–127, <https://doi.org/10.1214/16-STS588>.
- Liew, S. S., M. Khalil-Hani, and R. Bakhteri, 2016: Bounded activation functions for enhanced training stability of deep neural networks on visual pattern recognition problems. *Neurocomputing*, **216**, 718–734, <https://doi.org/10.1016/j.neucom.2016.08.037>.
- MeteoSwiss, 2015: Typische wetterlagen im Alpenraum (Typical weather conditions in the Alpen region). MeteoSwiss, 28 pp., https://www.meteoswiss.admin.ch/dam/jcr:7fc86ff0-23d6-41df-85c8-c66c92eee634/Typische_Wetterlagen_DE_low.pdf.
- , 2016: The new weather forecasting model for the Alpine region. MeteoSwiss, <https://www.cscs.ch/publications/press-releases/2016/the-new-weather-forecasting-model-for-the-alpine-region/>.
- , 2018: Documentation of MeteoSwiss grid-data products. Hourly precipitation estimation through rain-gauge and radar: CombiPrecip. MeteoSwiss, 5 pp., https://www.meteoswiss.admin.ch/dam/jcr:2691db4e-7253-41c6-a413-2c75c9de11e3/ProdDoc_CPC.pdf.
- Miyato, T., T. Kataoka, M. Koyama, and Y. Yoshida, 2018: Spectral normalization for generative adversarial networks. arXiv, 1802.05957v1, <https://doi.org/10.48550/arXiv.1802.05957>.
- Molina, M. O., C. Gutiérrez, and E. Sánchez, 2021: Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. *Int. J. Climatol.*, **41**, 4864–4878, <https://doi.org/10.1002/joc.7103>.
- Nerini, D., 2020: Probabilistic deep learning for postprocessing wind forecasts in complex terrain. ECMWF video, <https://vimeo.com/465719202>.
- , and F. Zanetta, 2021: Topo-descriptors. MeteoSwiss, <https://github.com/MeteoSwiss/topo-descriptors>.
- Perez, L., and J. Wang, 2017: The effectiveness of data augmentation in image classification using deep learning. arXiv, 1712.04621v1, <https://doi.org/10.48550/arXiv.1712.04621>.
- Rabiner, L., and B.-H. Juang, 1993: *Fundamentals of Speech Recognition*. 1st ed. Prentice-Hall, 544 pp.
- Ramon, J., L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, 2019: What global reanalysis best represents near-surface winds? *Quart. J. Roy. Meteor. Soc.*, **145**, 3236–3251, <https://doi.org/10.1002/qj.3616>.
- , —, P.-A. Bretonnière, M. Samsó, and F. J. Doblas-Reyes, 2021: A perfect prognosis downscaling methodology for seasonal prediction of local-scale wind speeds. *Environ. Res. Lett.*, **16**, 054010, <https://doi.org/10.1088/1748-9326/abe491>.
- Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195–204, <https://doi.org/10.1038/s41586-019-0912-1>.
- Richner, H., and P. Hächler, 2013: Understanding and forecasting alpine foehn. *Mountain Weather Research and Forecasting*, F. Chow, S. De Wekker, and B. Snyder, Eds., Springer Atmospheric Sciences, Springer, 219–260, https://doi.org/10.1007/978-94-007-4098-3_4.
- Ronneberger, O., P. Fischer, and T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, N. Navab et al., Eds., Lecture Notes in Computer Science, Vol. 9351, Springer, 234–241.

- Rue, H., S. Martino, and N. Chopin, 2009: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. Roy. Stat. Soc.*, **71**, 319–392, <https://doi.org/10.1111/j.1467-9868.2008.00700.x>.
- , A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren, 2017: Bayesian computing with INLA: A review. *Annu. Rev. Stat. Appl.*, **4**, 395–421, <https://doi.org/10.1146/annurev-statistics-060116-054045>.
- Santurkar, S., D. Tsipras, A. Ilyas, and A. Mądry, 2018: How does batch normalization help optimization? arXiv, 1805.11604v5, <https://doi.org/10.48550/arXiv.1805.11604>.
- Schwierz, C., P. Köllner-Heck, E. Zenklusen Mutter, D. N. Bresch, P.-L. Vidale, M. Wild, and C. Schär, 2010: Modelling European winter wind storm losses in current and future climate. *Climatic Change*, **101**, 485–514, <https://doi.org/10.1007/s10584-009-9712-1>.
- Sha, Y., D. J. G. Ii, G. West, and R. Stull, 2020: Deep-learning-based gridded downscaling of surface meteorological variables in complex terrain. Part I: Daily maximum and minimum 2-m temperature. *J. Appl. Meteor. Climatol.*, **59**, 2057–2073, <https://doi.org/10.1175/JAMC-D-20-0057.1>.
- Sharples, J., R. McRae, and S. Wilkes, 2012: Wind-terrain effects on the propagation of wildfires in rugged terrain: Fire channelling. *Int. J. Wildland Fire*, **21**, 282–296, <https://doi.org/10.1071/WF10055>.
- Sprenger, M., B. Dürr, and H. Richner, 2016: Foehn studies in Switzerland. *From Weather Observations to Atmospheric and Climate Sciences in Switzerland—Celebrating 100 years of the Swiss Society for Meteorology*, S. Willemse and M. Furger, Eds., vdf Hochschulverlag AG an der ETH Zürich, 215–247.
- Srivastava, R. K., K. Greff, and J. Schmidhuber, 2015: Highway networks. arXiv, 1505.00387v2, <https://doi.org/10.48550/arXiv.1505.00387>.
- Staffell, I., and S. Pfenninger, 2016: Using bias-corrected reanalysis to simulate current and future wind power output. *Energy*, **114**, 1224–1239, <https://doi.org/10.1016/j.energy.2016.08.068>.
- Stucki, P., S. Brönnimann, O. Martius, C. Welker, M. Imhof, N. von Wattenwyl, and N. Philipp, 2014: A catalog of high-impact windstorms in Switzerland since 1859. *Nat. Hazards Earth Syst. Sci.*, **14**, 2867–2882, <https://doi.org/10.5194/nhess-14-2867-2014>.
- , S. Dierer, C. Welker, J. J. Gómez-Navarro, C. C. Raible, O. Martius, and S. Brönnimann, 2016: Evaluation of down-scaled wind speeds and parameterised gusts for recent and historical windstorms in Switzerland. *Tellus*, **68A**, 31820, <https://doi.org/10.3402/tellusa.v68.31820>.
- Sun, R., S. K. Krueger, M. A. Jenkins, M. A. Zulauf, and J. J. Charney, 2009: The importance of fire–atmosphere coupling and boundary-layer turbulence to wildfire spread. *Int. J. Wildland Fire*, **18**, 50–60, <https://doi.org/10.1071/WF07072>.
- Vandal, T., E. Kodra, S. Ganguly, A. Michaelis, R. Nemani, and A. R. Ganguly, 2018: Generating high resolution climate change projections through single image super-resolution: An abridged version. *Proc. 27th Int. Joint Conf. on Artificial Intelligence*, Stockholm, Sweden, IJCAI, 5389–5393, <https://doi.org/10.24963/ijcai.2018/759>.
- Weissmann, M., F. J. Braun, L. Ganter, G. J. Mayr, S. Rahm, and O. Reitebuch, 2005: The Alpine mountain–plain circulation: Airborne Doppler lidar measurements and numerical simulations. *Mon. Wea. Rev.*, **133**, 3095–3109, <https://doi.org/10.1175/MWR3012.1>.
- Welker, C., O. Martius, P. Stucki, D. Bresch, S. Dierer, and S. Brönnimann, 2016: Modelling economic losses of historic and present-day high-impact winter windstorms in Switzerland. *Tellus*, **68A**, 29546, <https://doi.org/10.3402/tellusa.v68.29546>.
- Winstral, A., T. Jonas, and N. Helbig, 2017: Statistical downscaling of gridded wind speed data using local topography. *J. Hydrometeorol.*, **18**, 335–348, <https://doi.org/10.1175/JHM-D-16-0054.1>.
- Zardi, D., and C. D. Whiteman, 2013: Diurnal mountain wind systems. *Mountain Weather Research and Forecasting*, F. Chow, S. De Wekker, and B. Snyder, Eds., Springer, 35–119, https://doi.org/10.1007/978-94-007-4098-3_2.