

Comparison of AI-based approaches for statistical downscaling of surface wind fields in the North Atlantic

Vadim Rezvov^{1,2}, Mikhail Krinitskiy¹, Alexander Gavrikov¹, and Sergey Gulev¹

rezvov.vyu@phystech.edu

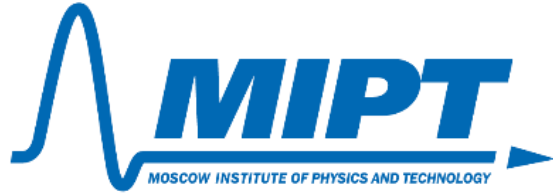
¹*Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia*

²*Moscow Institute of Physics and Technology (MIPT), Dolgoprudny, Russia*



Virtual Conference | 19-30 April 2021

Session NP4.1: Big Data and AI in the Earth Sciences



Moscow Institute of Physics and Technology,
Dolgoprudny, Russia



Shirshov Institute of Oceanology,
Moscow, Russia



Sea-Atmosphere Interaction Laboratory,
Shirshov Institute of Oceanology

Abstract

Surface winds — both wind speed and vector wind components — are fields of fundamental climatic importance. The character of surface winds greatly influences (and is influenced by) surface exchanges of momentum, energy, and matter. These wind fields are of interest in their own right, particularly concerning the characterization of wind power density and wind extremes. Surface winds are influenced by small-scale features such as local topography and thermal contrasts. That is why **accurate high-resolution prediction of near-surface wind fields** is a topic of central interest in various fields of science and industry. **Statistical downscaling** is the way for inferring information on physical quantities at a local scale from available low-resolution data. It is one of the ways **to avoid costly high-resolution simulations**. Statistical downscaling connects variability of various scales using **statistical prediction models**. This approach is fundamentally data-driven and can only be applied in locations where observations have been taken for a sufficiently long time **to establish the statistical relationship**. Our study considered **statistical downscaling of surface winds** (both wind speed and vector wind components) **in the North Atlantic**. **Deep learning methods** are among the most outstanding examples of state-of-the-art machine learning techniques that allow approximating sophisticated nonlinear functions. In our study, we applied various approaches involving **artificial neural networks for statistical downscaling of near-surface wind vector fields**. We used ERA-Interim reanalysis as low-resolution data and RAS-NAAD dynamical downscaling product (14km grid resolution) as a high-resolution target. We compared statistical downscaling results to those obtained with **bilinear/bicubic interpolation** with respect to downscaling quality. We investigated how network complexity affects downscaling performance. We will demonstrate the preliminary results of the comparison and propose the outlook for further development of our methods.

Problem review

- Accurate prediction of near-surface wind fields is a field of fundamental importance in science. However, memory and performance costs of numerical fields simulations limit the availability of high-resolution predictions, especially for large spatial domains
- Downscaling allows to circumvent costly high-resolution numerical simulations over large spatial scales. This approach uses suitable refinement processes to infer high-resolution information on physical quantities at local scale from available low-resolution simulation data.
- Meteorological and climatic research show a large variety of methods of physical parameters downscaling. Two broad types of such methods are dynamical and empirical-statistical approaches.
- **Dynamical downscaling:**
high-resolution numerical models are used over sub-domains of the large domain of interest, and their outputs provide boundary conditions for the numerical simulations on a finer scale. The restricted size of the model sub-domain leads to a significant reduction of computational costs compared to global domain simulations, but dynamical downscaling remains computationally demanding and time-consuming.

Statistical downscaling

- Statistical downscaling avoids simulation at the finer scales. It uses a low-resolution simulation (predictor data) to infer high-resolution information about the same physical quantities (predictand data). Correlations between the quantities at low and high resolutions are learned by **training statistical models** on a large set of known predictor–predictand data pairs.
- The empirical-statistical downscaling approaches apply **statistical regression methods** for downscaling purposes. These can be:
 - multilinear regression methods;
 - quantile mapping approaches;
 - nonlinear regression methods;
 - non-classical machine learning techniques.
- Additional value and downscaling quality of nonlinear machine learning approaches over classical methods is a question that is of current scientific interest.

Deep learning for statistical downscaling

- **Deep learning methods** are examples of state-of-the-art machine learning techniques.
- In particular, convolutional neural networks (CNNs) such as **ResNet** and **U-Net** have found manifold applications in complex image processing. One of these applications is **single-image super-resolution**, that is, **the generation of high-resolution images from low-resolution images**. This task can be formally thought of as a very similar task to climate variables downscaling.
- CNNs operate on a spatial domain as a **set of localized linear models (localized filter kernels)**, which are applied repeatedly at varying spatial positions across the domain through convolution operations. The restriction of the model parametrization to local filter kernels limits the number of trainable parameters, and thus **reduces the tendency of the model to overfit patterns in the data**.
- CNNs work most effectively with regular-gridded data in multi-dimensional array representations, while more complex mappings can be learned by stacking multiple layers of convolution operations (increasing the depth of the models) and applying these successively to generate more abstract feature representations. Nonlinear activation functions between convolution layers can **enable the model to learn nonlinear mappings**.
- There are sequential feature processing (**fully-convolutional NNs**), and more difficult methods such as skip connections between pairs of convolution layers (**U-Net**) or residual learning (**ResNet**) that **can improve model performance**.

Deep learning for statistical downscaling

- In [1], there was presented **SRGAN**, a generative adversarial network (GAN) for image super-resolution (SR). It is the first framework capable of inferring photo-realistic natural images for 4× upscaling factors.
- GANs provide a powerful framework for generating plausible-looking natural images with high perceptual quality. SRGAN is the first very deep ResNet architecture using the concept of GANs to form a photo-realistic single-image super-resolution.
- The general idea behind this GAN concept is that it allows one **to train a generative model G** with the goal of **fooling a differentiable discriminator D** that is trained to **distinguish super-resolved images from real high-resolution images**. With this approach **generator G can learn to create solutions that are highly similar to real high-resolution images** and thus difficult to classify by D . This encourages perceptually superior solutions residing in the subspace, the manifold, of natural images.

[1] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

Contribution

- Our study considered **statistical downscaling of near-surface winds** (both wind speed and vector wind components) **in the North Atlantic**.
- We perform a study of **different neural network architectures** for this task. The results are compared to those obtained by a bicubic interpolation, with respect to downscaling quality. We train models to predict the most likely outcome of a high-resolution simulation of near-surface winds 10 m above ground, based on low-resolution wind fields as primary predictors.
- We **compare the capabilities and downscaling quality** of different existing models, which reflect varying degrees of model complexity and elaboration. Starting with a simple bicubic interpolation, we continue the comparison with nonlinear convolutional models of increasing complexity (fully-convolutional NN, U-Net, Resnet-based SRGAN).
- The reconstruction quality of all downscaling models is compared to the high-resolution simulations of real-world weather situations for the North Atlantic.

Materials and methods

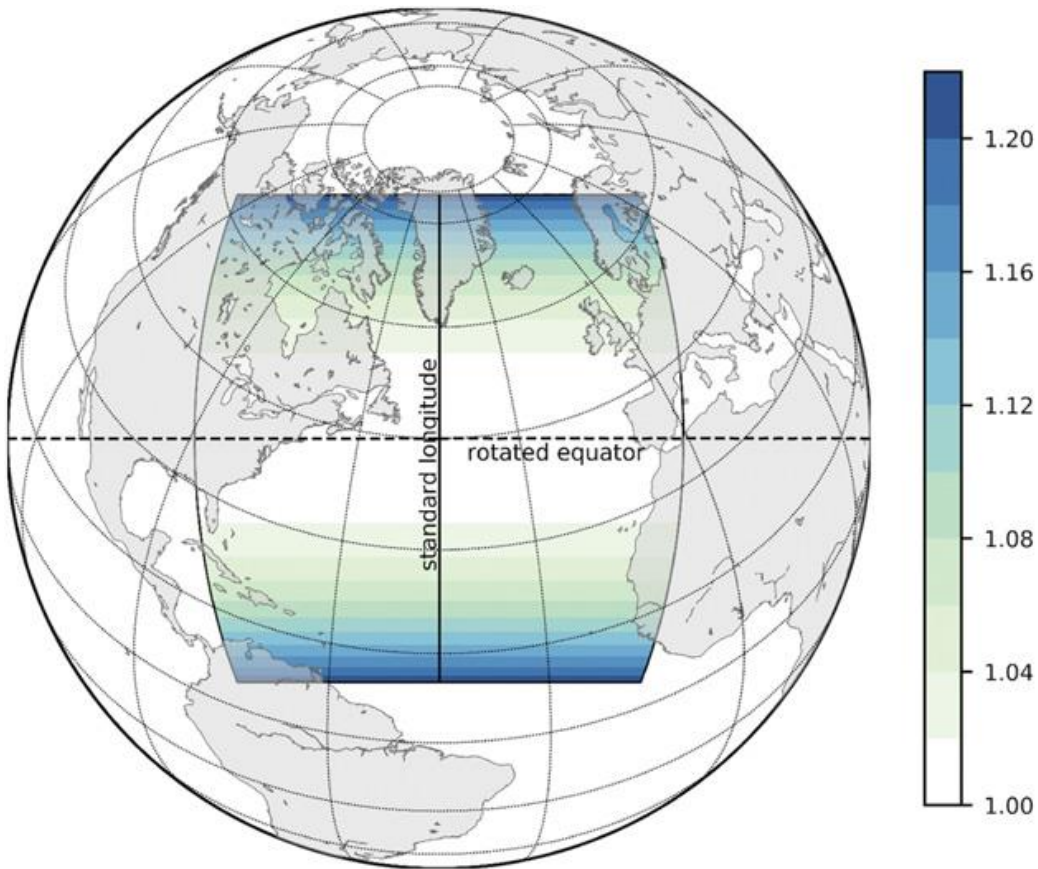


Fig.1. NAAD computational domain and map-scale factor for the HiRes simulation [2]

- Initial data were taken from [2]. This paper presents the results of the Russian Academy of Sciences North Atlantic Atmospheric Downscaling (RAS-NAAD) project, which provides a 40-yr hindcast of the North Atlantic atmosphere, performed for the period 1979–2018 and forced by ERA-Interim as a lateral boundary condition.
- The domain covers the North Atlantic from 10° to 80°N and from 90°W to 5°E (Fig. 1), with the center at 45°N, 45°W. The initial and lateral boundary conditions, including sea surface temperature (SST), were provided by the ERA-Interim reanalysis at 0.7° x 0.7° spatial resolution and 60 levels in the vertical direction.

[2] Gavrikov, A., Gulev, S. K., Markina, M., Tilinina, N., Verezhenskaya, P., Barnier, B., ... & Sokov, A. (2020). RAS-NAAD: 40-yr High-Resolution North Atlantic Atmospheric Hindcast for Multipurpose Applications (New Dataset for the Regional Mesoscale Studies in the Atmosphere and the Ocean). *Journal of Applied Meteorology and Climatology*, 59(5), 793-817.

Materials and methods

- The spatial resolution in the basic NAAD high-resolution experiment (**HiRes**) was 14 km (**550 x 550 points**) and 50 terrain-following, dry hydrostatic pressure levels, starting from around 10–12 m above the ocean surface to 50 hPa with.
- Besides the HiRes experiment, [2] also presents a moderately low resolution experiment (**LoRes**) with the hydrostatic setting at 77-km resolution (**110 x 110 points**) with 50 vertical levels (as in HiRes).
- All experiments in [2] were run for the 40-yr period from January 1979 to December 2018. Model output at **3-hourly time resolution** is freely available to the users.

The present study uses 38-yr period from January 1979 to December 2016:

- **Low-resolution data:** LoRes RAS-NAAD output – wind components (10 m above the ocean surface) and sea level pressure (SLP)
- **High-resolution data:** HiRes RAS-NAAD output – wind components (10 m above the ocean surface) and sea level pressure (SLP)

Entire dataset is divided into 2 parts. Training dataset is from 1979 to 2008 and validation dataset is from 2009 to 2016

We consider statistical downscaling of near-surface winds (both wind speed and vector wind components) generating high-resolution data from low-resolution data (LoRes). The high-resolution result of downscaling is compared with real high-resolution data (HiRes). Statistical downscaling is conducted by bicubic interpolation (baseline method), SRGAN and different types of convolutional neural networks (fully-convolutional NN, U-Net).

SRGAN

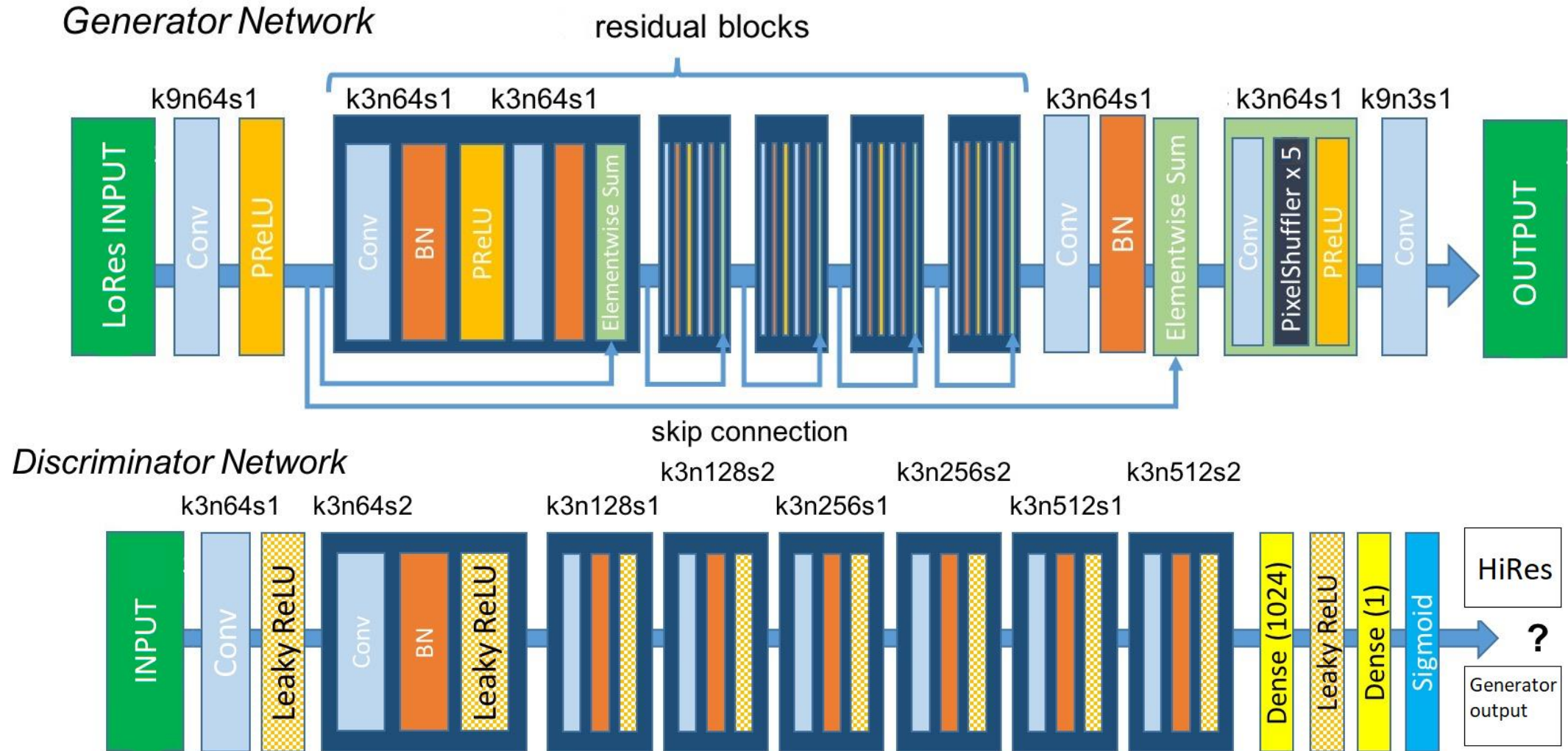


Fig.2. Architecture of Generator and Discriminator Network (based on [1]) with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

[1] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

U-Net

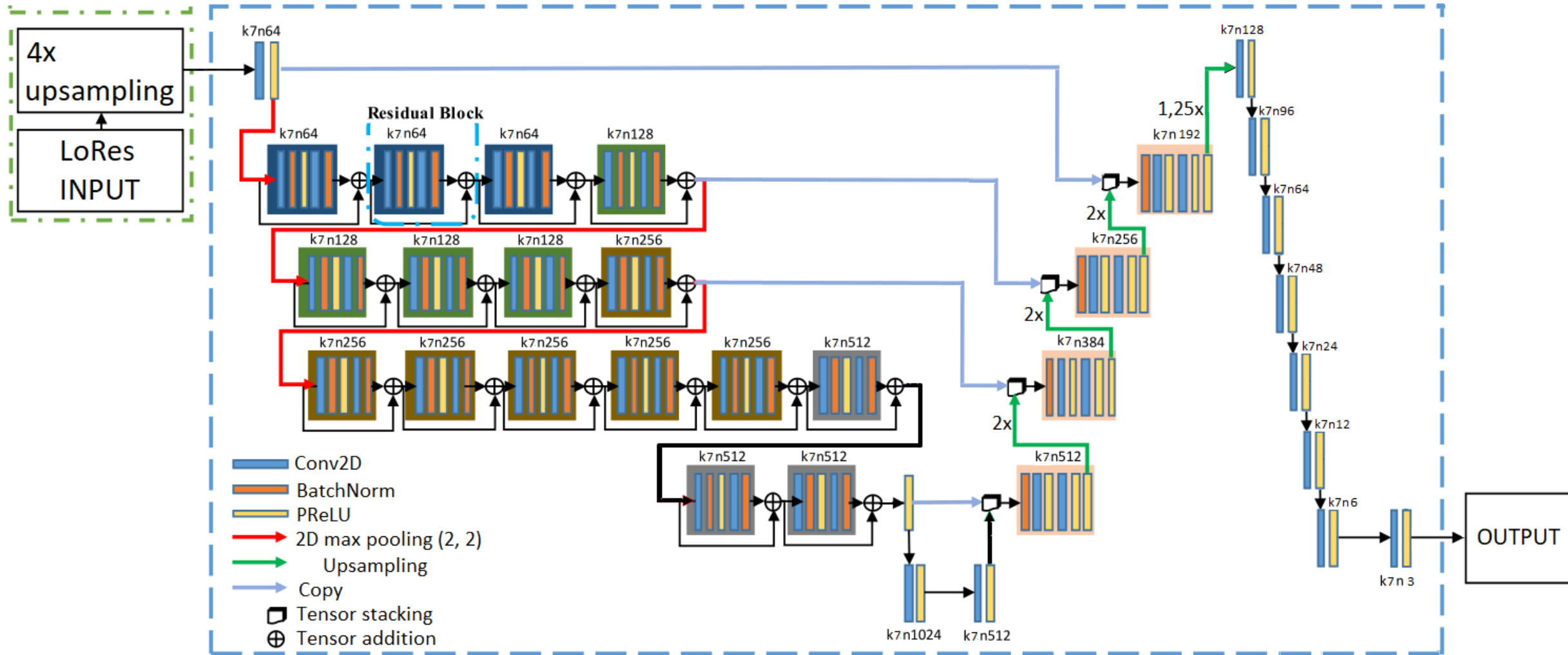


Fig.3. An overview of the U-Net based on [3] where kz_1nz_2 represents a convolutional layer with kernel size of $z_1 \times z_1$ and with z_2 feature maps

[3] Hu, X., Naiel, M. A., Wong, A., Lamm, M., & Fieguth, P. (2019). RUNet: A robust UNet architecture for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.

Fully-convolutional NN

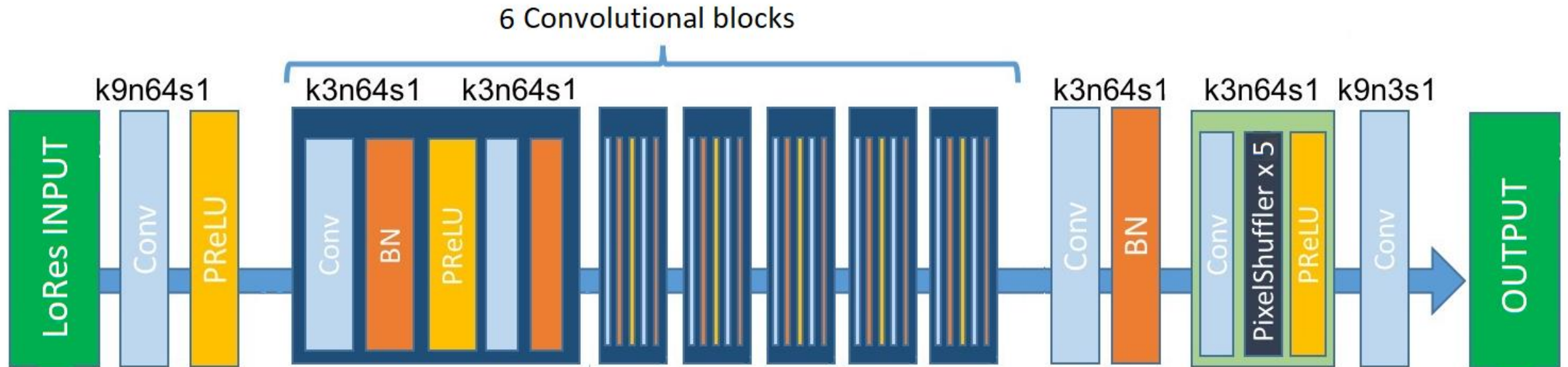


Fig.4. Architecture of fully-convolutional neural network (based on [1]) with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

[1] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

Metrics

We compared statistical downscaling results to real HiRes data with respect to downscaling quality. To estimate quality of statistical downscaling, we use following metrics:

- root-mean-square error (**RMSE**). The RMSE represents the square root of the second sample moment of the differences between predicted values (model output) and observed values (HiRes) or the quadratic mean of these differences. RMSE is **always non-negative**, and a value of 0 (almost never achieved in practice) would indicate a perfect fit to the data. In general, **a lower RMSE is better than a higher one**:

$$RMSE = \sqrt{\frac{\sum_{i,j} (\hat{u}_{i,j} - u_{i,j})^2}{N} + \frac{\sum_{i,j} (\hat{v}_{i,j} - v_{i,j})^2}{N}}$$

where u and v – wind components,

$\hat{u}$, $\hat{v}$ - model outputs, u, v – HiRes data,

$N = 550 \times 550 = 302\,500$

i, j – indexes in 550 x 550 grid

RMSE is defined for every single point of time and then averaged over dataset

Metrics

- **RMSE-95** is calculated using the definition of root-mean-square error. The difference is that for the calculation of the RMSE-95, only values that are above the 95th percentile for every single 550 x 550 grid node are taken into account. Consequently, **a lower RMSE-95 is better than a higher one.**
- Peak signal-to-noise ratio (**PSNR**) is an engineering term for the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. In general, **a higher PSNR is better than a lower one.** PSNR is usually expressed as a logarithmic quantity using the decibel scale:

$$PSNR = 20 \cdot \log_{10} \left(\frac{MAX}{RMSE} \right)$$

Here, *MAX* is the maximum value of the 3 x 550 x 550 model output, and *RMSE* is root-mean-square error for the output at the same point of time.

Results – baseline method

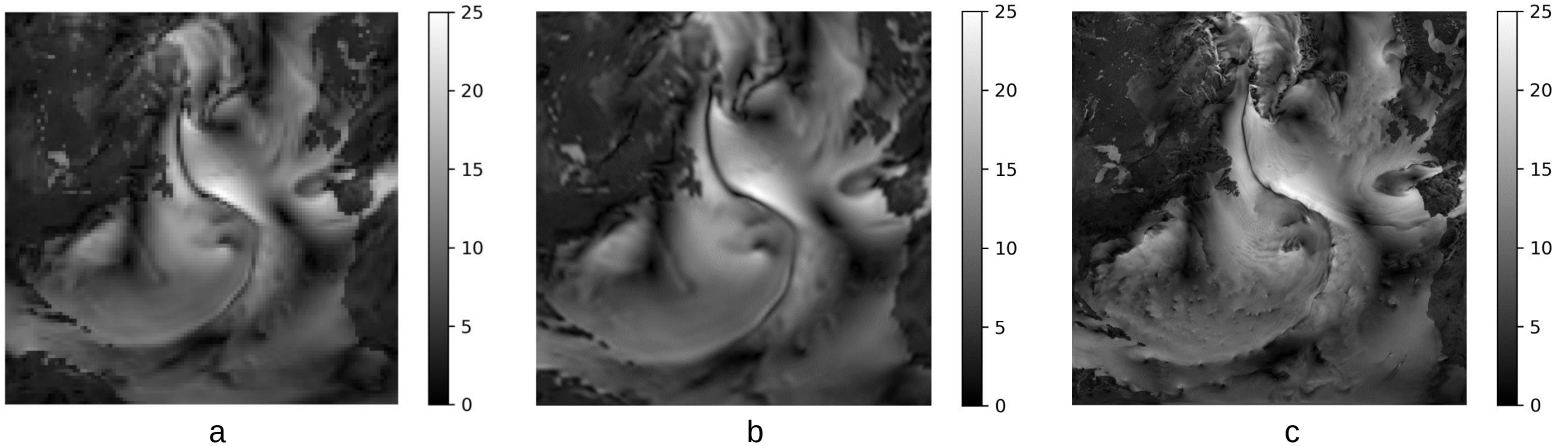


Fig.5. Wind magnitude (m/s, 10 m above surface), 00:00 1 January 2010
(a) LoRes; (b) Bicubic interpolation of LoRes; (c) HiRes

Metrics	Bicubic interpolation
RMSE, m/s	1,435
RMSE-95, m/s	1,898
PSNR	35,159

Results – SRGAN

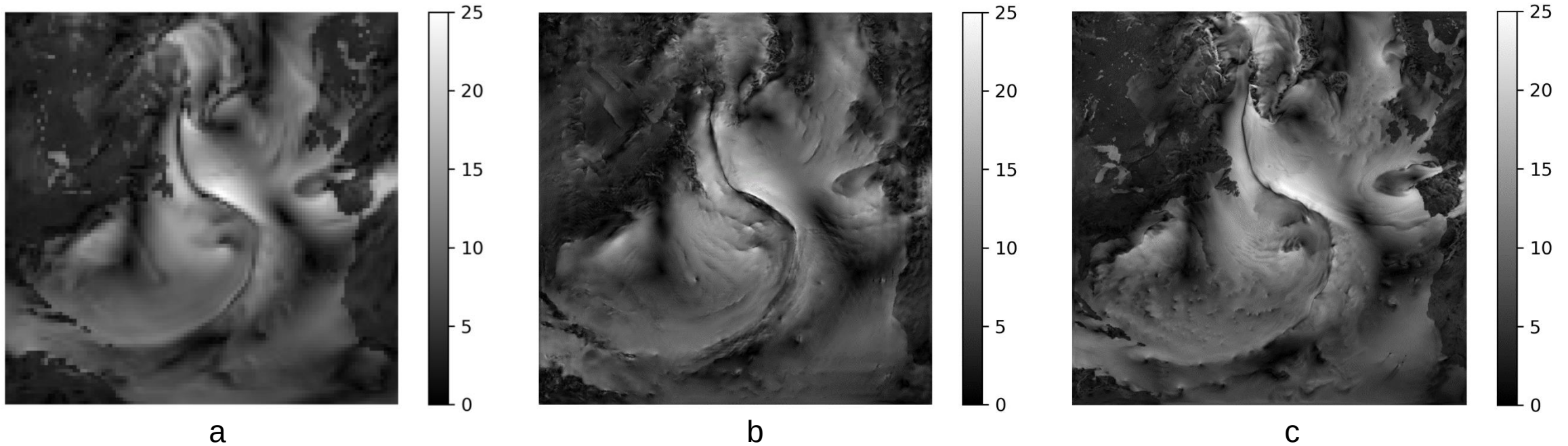


Fig.6. Wind magnitude (m/s, 10 m above surface), 00:00 1 January 2010
(a) LoRes; (b) SRGAN output; (c) HiRes

Metrics	Bicubic interpolation	SRGAN
RMSE, m/s	1,435	1,902
RMSE-95, m/s	1,898	3,031
PSNR	35,159	30,505

Results – U-Net

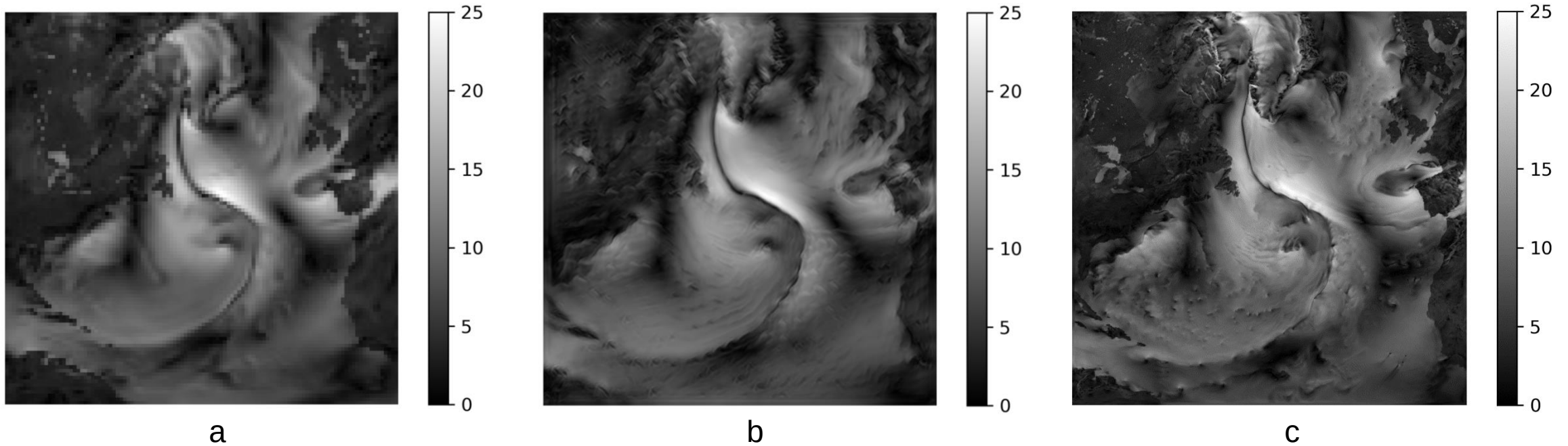


Fig.7. Wind magnitude (m/s, 10 m above surface), 00:00 1 January 2010
(a) LoRes; (b) U-Net output; (c) HiRes

Metrics	Bicubic interpolation	U-Net
RMSE, m/s	1,435	1,769
RMSE-95, m/s	1,898	2,793
PSNR	35,159	31,555

Results – fully-convolutional NN

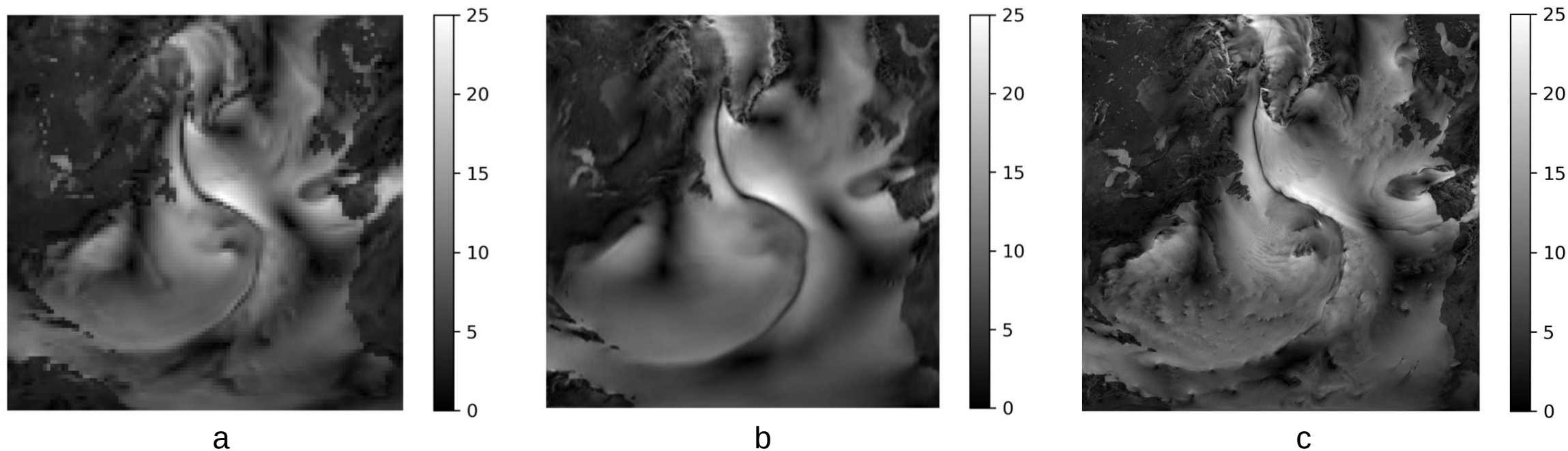


Fig.8. Wind magnitude (m/s, 10 m above surface), 00:00 1 January 2010
(a) LoRes; (b) CNN output; (c) HiRes

Metrics	Bicubic interpolation	Fully-convolutional NN
RMSE, m/s	1,435	1,290
RMSE-95, m/s	1,898	1,972
PSNR	35,159	33,929

Results – Summary

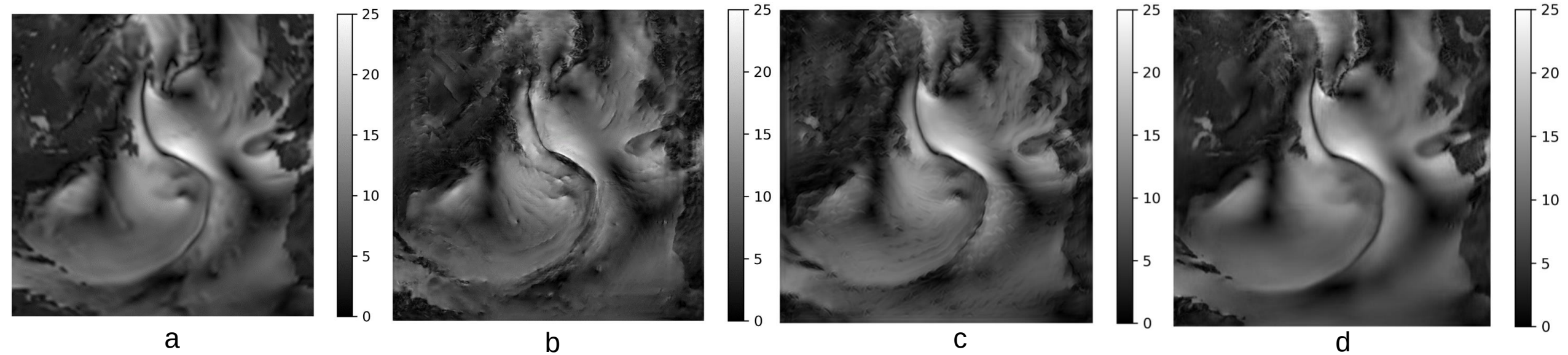


Fig.9. Wind magnitude (m/s, 10 m above surface), 00:00 1 January 2010, obtained in different models: (a) Bicubic interpolation; (b) SRGAN; (c) U-Net; (d) fully-convolutional NN

Metrics	Bicubic	SRGAN	U-Net	Fully-convolutional NN
RMSE, m/s	1,435	1,902	1,769	1,290
RMSE-95, m/s	1,898	3,031	2,793	1,972
PSNR	35,159	30,505	31,555	33,929

Results – Summary

- Results show that, according to the used quality metrics, deep learning methods of statistical downscaling are not able to outperform simple method of bicubic interpolation.
- For example, RMSE-95 and PSNR values shown by bicubic interpolation are better than the ones obtained by two types of CNNs or SRGAN.
- The most obvious quality metric is RMSE. For the bicubic interpolation, its value is much lower than for SRGAN or U-Net and it is just slightly lower than for simple fully-convolutional neural network.
- Nevertheless, visual qualitative analysis shows that results obtained in SRGAN model looks much more similar to the original high-resolution (HiRes) data. SRGAN downscaled output has mesoscale details that are not in LoRes data but easily distinguishable in original HiRes data. And on the contrary, the bicubic interpolation is the worst method visually because mesoscale information is absolutely lost.

Metrics	Bicubic	SRGAN	U-Net	Fully-convolutional NN
RMSE, m/s	1,435	1,902	1,769	1,290
RMSE-95, m/s	1,898	3,031	2,793	1,972
PSNR	35,159	30,505	31,555	33,929