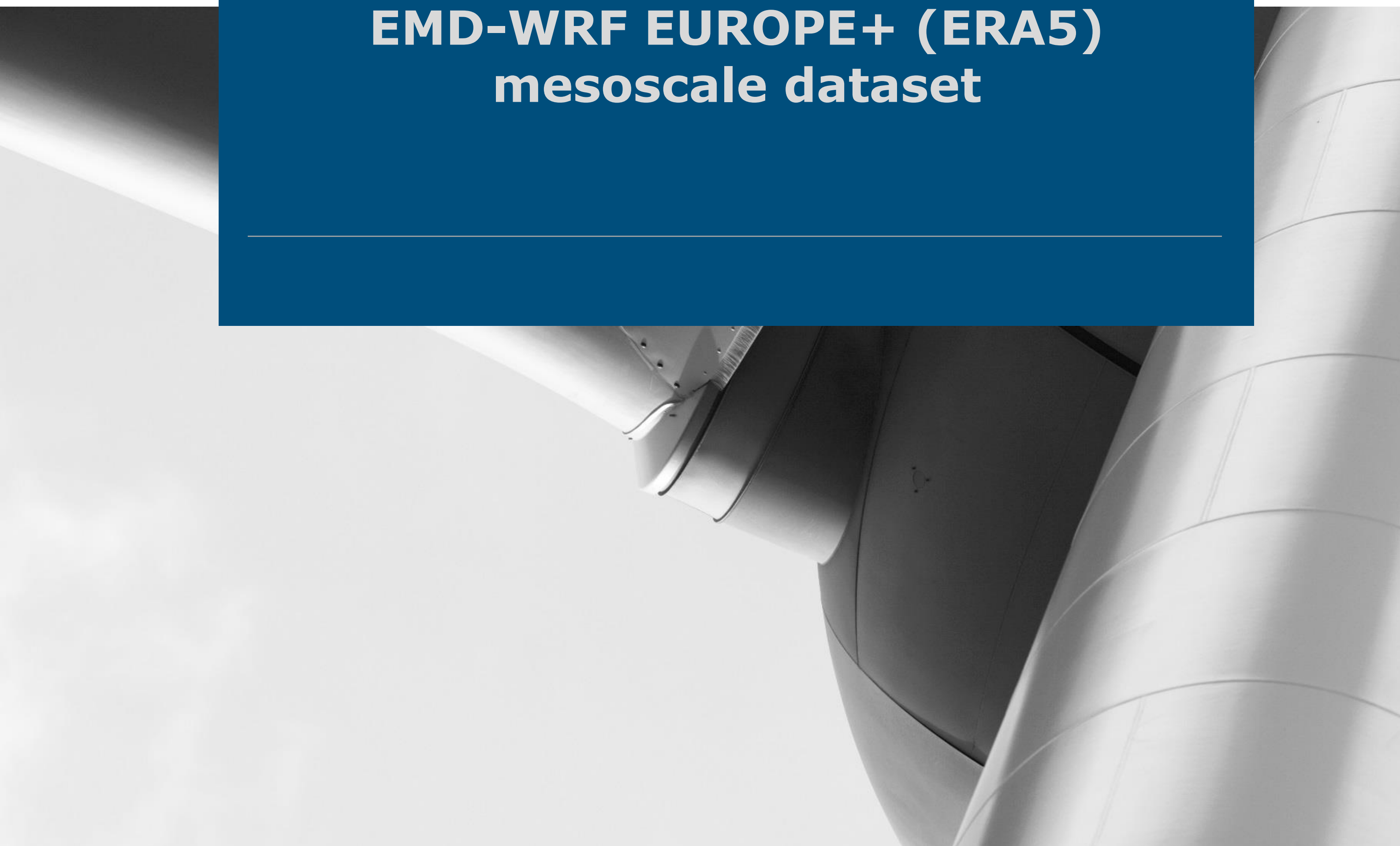
# Technical Note:

## Validation of EMD-WRF EUROPE+ (ERA5) mesoscale dataset



DATE

8 January 2020

PREPARED BY

EMD International A/S

Niels Jernes Vej 10

DK- 9220 Aalborg

T: + 45 96 35 44 44

E: emd@emd.dk

AUTHOR

Lasse Svenningsen

EMD International A/S

ls@emd.dk

CONTRIBUTORS

Morten Lybech Thøgersen

Thorkild Sørensen

REVIEW

Morten Lybech Thøgersen

EDITION

1.0

# Contents

# Executive Summary

In June 2019 EMD released a new pre-run mesoscale dataset named 'EMD-WRF Europe+ (ERA5)' or 'EMD-EUR+' for short. The dataset replaces EMD-ConWx which is downscaled from ERA-Interim, discontinued from August 2019. The EMD-EUR+ dataset is downscaled to 3km from ERA5 and the domain covers Europe plus Turkey and Iceland.

For 300 high-quality and tall masts within the modelling domain EMD-EUR+ shows significantly better performance than all the alternative datasets ERA5, MERRA2 and NEWA for all performance metrics considered. The evaluated metrics include wind speed correlation at several averaging periods, wind direction and shape of wind speed distribution. Compared to EMD-ConWx with a smaller domain size and 160 masts, EMD-EUR+ shows similar performance improvements for all metrics.

The highly profiled NEWA (New European Wind Atlas) dataset shows surprisingly poor performance on par with MERRA2, but poorer than ERA5 from which it is downscaled, and significantly poorer than EMD-EUR+.

# 1 EMD-WRF EUROPE+ (ERA5)

EMD-WRF Europe+ (ERA5), for short 'EMD-EUR+', is a new pre-run dataset released by EMD from June 2019. The dataset replaces the EMD-ConWx mesoscale dataset, which is downscaled from the discontinued global reanalysis ERA-Interim made by ECMWF[1]. EMD-EUR+ is downscaled from the new and much improved global reanalysis ERA5 also provided by ECMWF.

The EMD-EUR+ dataset and the parameters available is thoroughly documented in EMD's Knowledgebase[2] and includes data back to and including 1999 for **instant download in windPRO**. The full spatial domain may be **visualized in windprospecting.com** directly accessed from windPRO via the 'Data' menu. Annual subscription to EMD-EUR+ is 2,000€/year for first users and 800€ for additional users (prices from year 2019).

EMD-EUR+ is based on the significantly improved global reanalysis model ERA5 and a WRF modelling setup which is significantly improved and optimized compared to that used for EMD-ConWx. The remainder of this report provides a detailed description of the EMD-EUR+ dataset (section 1) and a thorough validation of the EMD-EUR+ data compared to alternative datasets (section 2).

## 1.1 Domain

The domain of the EMD-EUR+ is illustrated in figure 1. It covers the whole of Europe until the Ural mountains in Russia including both Iceland and entire Turkey.

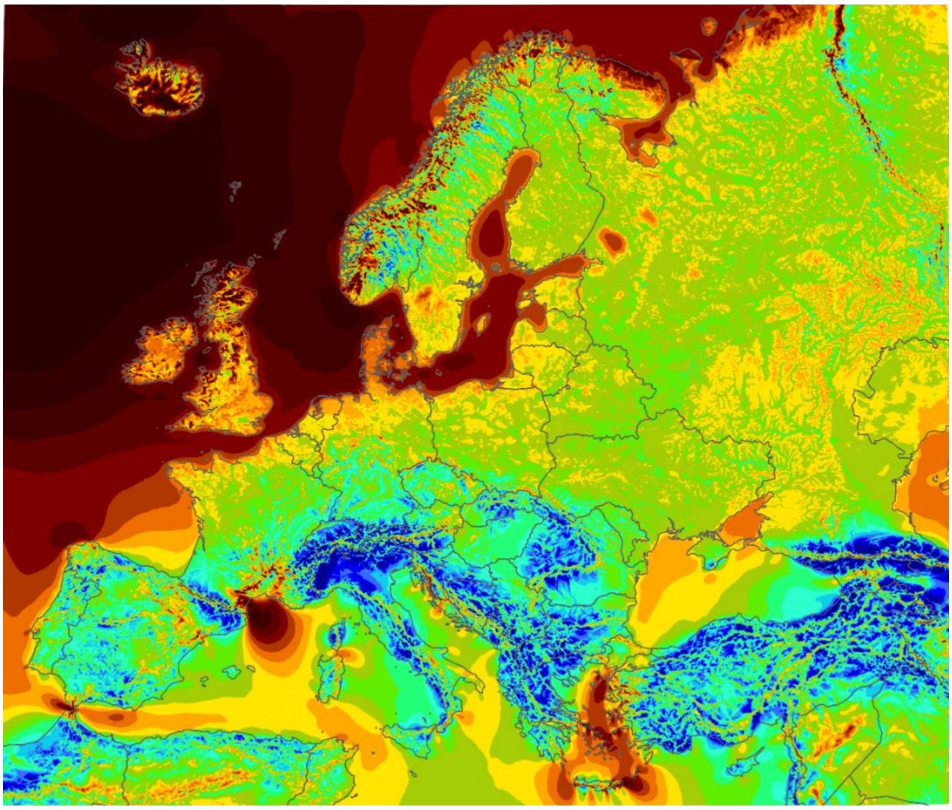


*Figure 1: Entire EMD-WRF Europe+ domain showing the average wind speeds at 100m agl.*

[1] ECMWF is: 'European Centre for Medium range Weather Forecasts' providing global forecast and reanalysis data.
[2] http://help.emd.dk/knowledgebase/ or http://help.emd.dk/mediawiki/index.php?title=EMD-WRF_Europe%2B

## 1.2 Parameters

The full list of 127 parameters available for EMD-EUR+ is listed in EMD's knowledgebase (see footnote 2, p. 4). The following main parameters are available at 13 heights between 10m and 4000m: wind speed mean, wind direction, wind speed stdev[3], pressure, temperature, relative humidity, cloud water content, cloud ice content. Additional parameters include e.g. visibility, sensible heat flux, total precipitation, direct shortwave irradiance, diffuse shortwave irradiance and the inverse Monin-Obukhov Length (for atmospheric stability classification).

# 2 Validation of EMD-WRF Europe+

This validation of EMD-EUR+ is based on EMD's internal database of more than 1000 masts of wind measurements conducted on 50m masts or higher. These masts are either from wind energy projects or research masts, and hence generally high-quality measurements. For the EMD-EUR+ domain 300 masts have been selected for the validation based on objective criteria such as measuring height, duration and data quality.

## 2.1 Alternative datasets

windPRO's elaborate on-line service makes several free global or regional datasets easily available, see the EMD's Knowledgebase[4] for additional information. Of these the main and widely used alternatives to EMD-EUR+ is the raw ERA5 data or the MERRA2 data, both global models described in more detail in [1][5]. Another alternative is the recently released New European Wind Atlas, NEWA[6], released as a main outcome of the highly-profiled high-budget EU funded NEWA project. Finally, since EMD-EUR+ replaces EMD-ConWx a comparison between the two is of interest to quantify the improvement of EMD-EUR+ relative to its predecessor, which is already a good and widely used dataset. The following gives a brief description of each of these main alternatives or 'competitors' to EMD-EUR+.

### 2.1.1 ERA5

The global model ERA5 is developed by ECMWF through the Copernicus Climate Change Service (C3S) released in batches from 2017 as a replacement to the popular ERA-Interim dataset which ended August 2019. ERA5 data has a horizontal resolution of ca. 31km with hourly values available at 100m agl. for wind speed and direction. More information on ERA5 data at the windPRO knowledgebase[7].

### 2.1.2 MERRA2

The global model MERRA2 is produced by NASA Goddard Space Flight Center at a resolution of ca. 50km and hourly values. Wind speed and direction are available at 50m agl. More information on MERRA2 is found at the windPRO knowledgebase[8].

---

[3] The standard measure of turbulence, however, the real parameter name is 'sqrtTKE' as the wind speed standard deviation inside WRF is derived as the square root of a Turbulent Kinetic Energy (TKE) budget.

[4] http://help.emd.dk/knowledgebase/ or https://help.emd.dk/mediawiki/index.php?title=Category%3AWind_Data

[5] http://help.emd.dk/knowledgebase/content/TechNotes/TechnicalNote6_EMD-WRF_On-Demand_20180116.pdf

[6] http://www.neweuropeanwindatlas.eu/ and https://map.neweuropeanwindatlas.eu/

[7] http://help.emd.dk/mediawiki/index.php?title=ERA5_Data

[8] http://help.emd.dk/mediawiki/index.php?title=MERRA2_Data

### 2.1.3 NEWA (New European Wind Atlas)

The regional dataset NEWA covers the European countries excluding Belarus, Ukraine and Russia, but including the entire Turkey. NEWA includes both mesoscale timeseries at ca. 3km and downscaled microscale wind and power statistics at 250m. In this validation we use the mesoscale timeseries at 100m agl. available as 30min data. NEWA only covers up till 2018 and is not updated forward in time.

### 2.1.4 EMD-ConWx

EMD-ConWx has for several years been a very popular mesoscale dataset made by EMD in collaboration with ConWx. The dataset covers most of Europe including most of Ukraine and the western parts of Russia (not to the Urals) and is based on ERA-Interim. The dataset ends with ERA-Interim August 2019 data – as ERA-Interim data is discontinued from ECMWF. More information on EMD-ConWx mesoscale data at the windPRO knowledgebase[9].

## 2.2 Validation parameters

300 measurement masts in the EMD-EUR+ domain has been selected based objective criteria like height of measurements close to 100m, high data quality and minimum one year of duration. The following performance metrics is used to evaluate each model dataset against each of the measured timeseries.

### 2.2.1 Wind speed correlation ($R^2$)

The coefficient of determination $R^2$ (or squared correlation coefficient) between model wind speeds and measured speeds captures how well the two timeseries follow each other in time. Put in statistical terms $R^2$ expresses the proportion of variance in the measurements that can be explained by the model data.

For correlation analysis we calculate $R^2$ for several averaging periods 10min, 1hour, 1day and 1month for two reasons. First, because models may predict the correct amount of variance but not the exact timing of the peaks, which will lead to low correlation for unaveraged data but higher correlations for longer averaging periods. Second, because wind measurements are 10min averages whereas the model data are hourly or half-hourly data, which makes it difficult to judge which period is most fair for comparison.

The 10min correlations are done by first interpolating the model data to 10min values, and then performing the correlation analysis. This limits so-called aliasing errors which occur if the measurements are directly down-sampled to hourly values.

The 1hour correlations are done by applying a 1hour averaging window to the 10min measurements, where the averaging acts as a low-pass filter which prevents aliasing. The 1day and 1month correlations are done by first applying daily or monthly averaging windows to both measurements and model timeseries and then performing the correlation analysis.

### 2.2.2 Wind direction error (MAE error)

For wind directions all model time series are interpolated to 10min values (via vector components) and then the direction error is calculated between corresponding model and measurement 10min samples. The errors are corrected for 360° periodicity (e.g. if model direction is 11° and measurement direction is 356°) and finally absolute values are taken for all the direction errors. The mean of these absolute errors is used as a robust measure of how well the modelled wind direction follows the measured, on average.

---

[9] http://help.emd.dk/mediawiki/index.php?title=EMD-ConWx_Meso_Data_Europe

### 2.2.3 Wind distribution error (CV error)

The wind distribution is critical to capturing the energy content of a wind resource and several measures can be used for this purpose. Some validation studies simply use the error on the mean wind speed, however, for validation of mesoscale data this is inappropriate as mesoscale data do not account for local microscale effects, which will often significantly affect the wind speeds. In this validation we focus on how well the shape of the wind speed distribution is captured. The shape is closely related to the ratio between the standard deviation and the mean of the timeseries, the so-called coefficient of variation ('CV' for short). For a Weibull type wind speed distribution, commonly adopted for wind resource assessments, the Weibull shape factor 'k' is solely determined by the CV of the wind speed time series [2]. Another important advantage of using CV errors is that microscale effects, typically modelled as scale factors, cancel out in the CV as it enters both in the numerator and denominator of the CV. Hence, the CV is mostly insensitive to microscale effects (if modelled as speed-up factors).

## 2.3 Results: EMD-EUR+ vs ERA5/MERRA2/NEWA

The three performance-metrics described in sections 2.2.1-2.2.3 have been applied to EMD-EUR+ and the three alternative datasets of ERA5, MERRA2 and NEWA described in sections 2.1.1 to 2.1.3. The performance of EMD-ConWx is handled separately in section 2.4 as this dataset covers a smaller domain with fewer validation masts.

The performance of the four datasets is evaluated against 300 masts scattered around the domain of EMD-EUR+. For 10min correlations against the 300 masts the results are presented in Figure 2, as probability density distributions calculated using a kernel density estimator. Notice how the EMD-EUR+ has the highest peak at the highest $R^2$ slightly above 0.7, indicating superior performance relative to the other datasets. ERA5 has its main peak at a similar $R^2$ but with a much lower amplitude and a larger part of the probability distributed at lower $R^2$, mainly between 0.4-0.6.

Surprisingly, the recent NEWA dataset does not perform as well as the ERA5 it is downscaled from and supposed to add value to. NEWA performs significantly poorer than EMD-EUR+ with the peak $R^2$ value more than 0.1 below the $R^2$ peak value of EMD-EUR+.

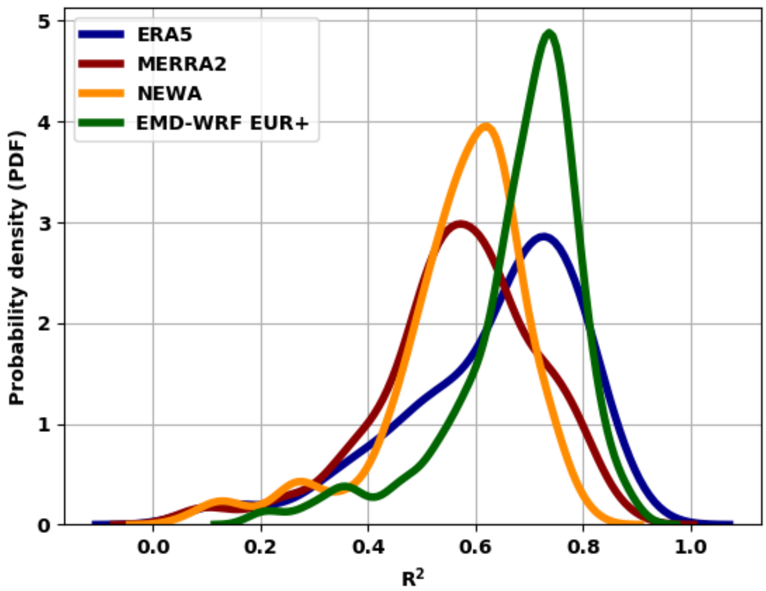



*Figure 2: Probability density of 10min correlations for the four datasets. Based on 300 masts using a 'kernel density estimator.*

Table 1 summarizes the statistics for the four datasets using the three performance metrics. For each metric and dataset, the mean across the 300 masts is shown together with a '±' followed by the standard deviation across the masts to indicate the spread. To make the results easy to read, the best performing dataset is highlighted in boldface green text and the worst as highlighted by red text. For all metrics EMD-EUR+ shows the best performance. NEWA and MERRA2 shows approximately similar and poorer performance.

*Table 1: Performance statistics for the four datasets based on 300 masts. 'Mean' values are averages across the 300 masts, and 'std' represent the spread across the masts.*

| Statistics 300 masts | MERRA2 | NEWA | ERA5 | EMD-WRF EUROPE+ |
|---|---|---|---|---|
| **Wind speed, 10min ($R^2$)** mean($R^2$) ± std($R^2$) | 0.57 ± 0.15 | 0.56 ± 0.13 | 0.63 ± 0.17 | **0.68 ± 0.12** |
| **Wind direction (MAE)** mean(MAE) ± std(MAE) | 47° ± 17° | 47° ± 14° | 42° ± 18° | **39° ± 16°** |
| **Wind distribution (CV error)** mean(CV error) ± std(CV error) | -7% ± 10% | 6% ± 11% | -3% ± 10% | **1% ± 9%** |
| **Highlight legend:** worst performance, second worst performance, **best performance** | | | | |

Figure 3 shows mean $R^2$ values across the 300 masts for four different averaging periods for each of the four datasets. Again EMD-EUR+ shows significantly higher $R^2$ values than the other datasets, in general ca. 0.05 above ERA5 and 0.1 above MERRA2 and NEWA.

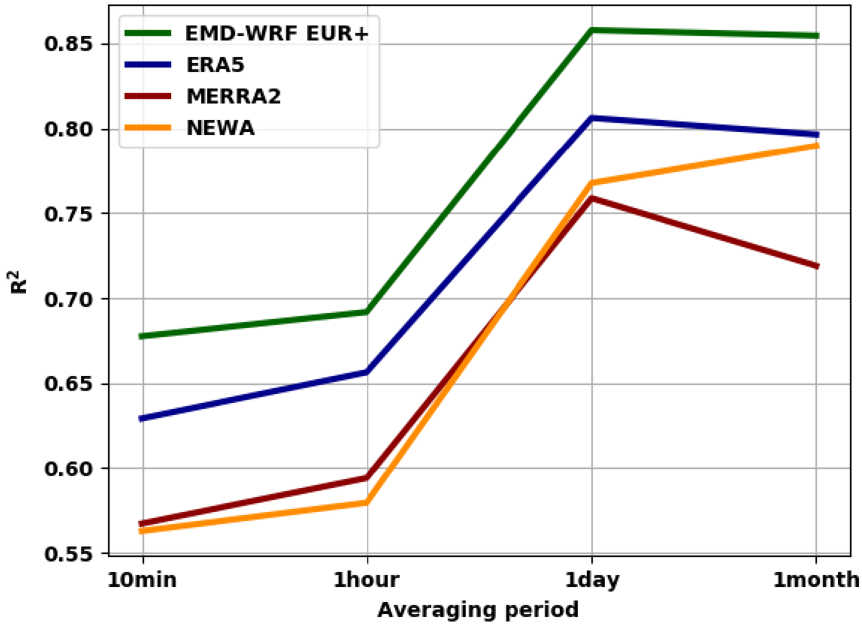



*Figure 3: Mean correlations across the 300 masts for four different averaging periods for the four datasets.*

## 2.4 Results: EMD-EUR+ Vs EMD-Conwx

EMD-WRF Europe+ is an improved replacement for EMD-Conwx. The magnitude of improvement is quantified for the three performance metrics as mentioned above – but using 160 masts available for the smaller EMD-Conwx domain. Figure 4 below shows the probability density plots for 10min $R^2$ values. Again EMD-EUR+ has a higher peak at a higher $R^2$.

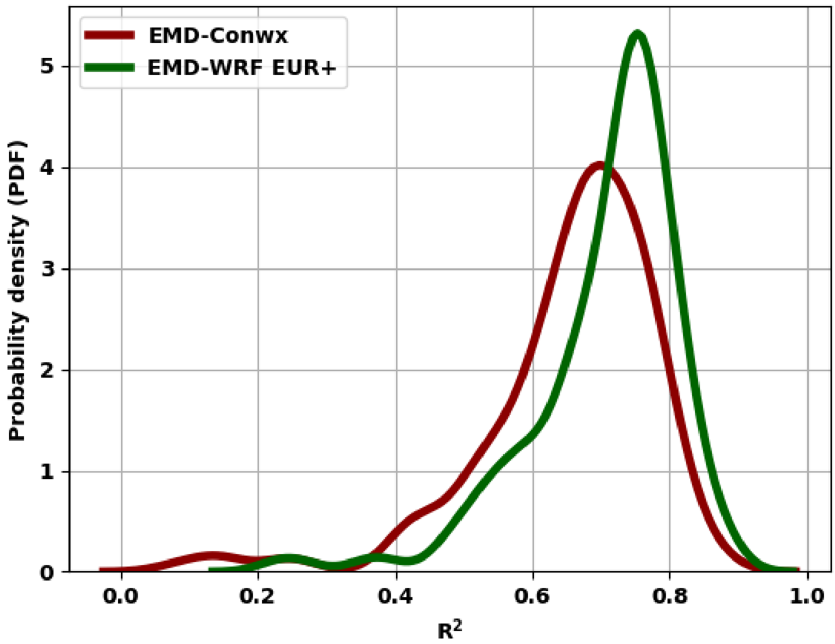



*Figure 4: Probability density of 10min correlations for the two datasets. Based on 160 masts using a 'kernel density estimator'.*

Table 2 summarizes the performance statistics comparing the performance of EMD-EUR+ and EMD-ConWx. Again EMD-EUR+ shows a consistent improvement on all parameters. $R^2$ values are on average improved by 0.06, and wind direction errors (MAE) reduced by around 10%.

*Table 2: Performance statistics comparing EMD-Conwx and EMD-WRF EUROPE+.*

| Statistics<br>160 masts | EMD-Conwx | EMD-WRF EURope+ |
|---|---|---|
| **Wind speed, 10min ($R^2$)**<br>mean($R^2$) ± std($R^2$) | 0.65 ± 0.13 | **0.71 ± 0.11** |
| **Wind direction (MAE)**<br>mean(MAE) ± std(MAE) | $42^o$ ± $16^o$ | **$38^o$ ± $16^o$** |
| **Wind distribution (CV error)**<br>mean(CV error) ± std(CV error) | 4% ± 14% | **2% ± 9%** |
| **Highlight legend:** worst performance, **best performance** | | |

Finally, Figure 5 shows the $R^2$ at different averaging periods for EMD-EUR+ and EMD-ConWx. Where the latter seems to have a surprising drop for monthly averages indicating a possible bias in seasonal variation of EMD-ConWx.

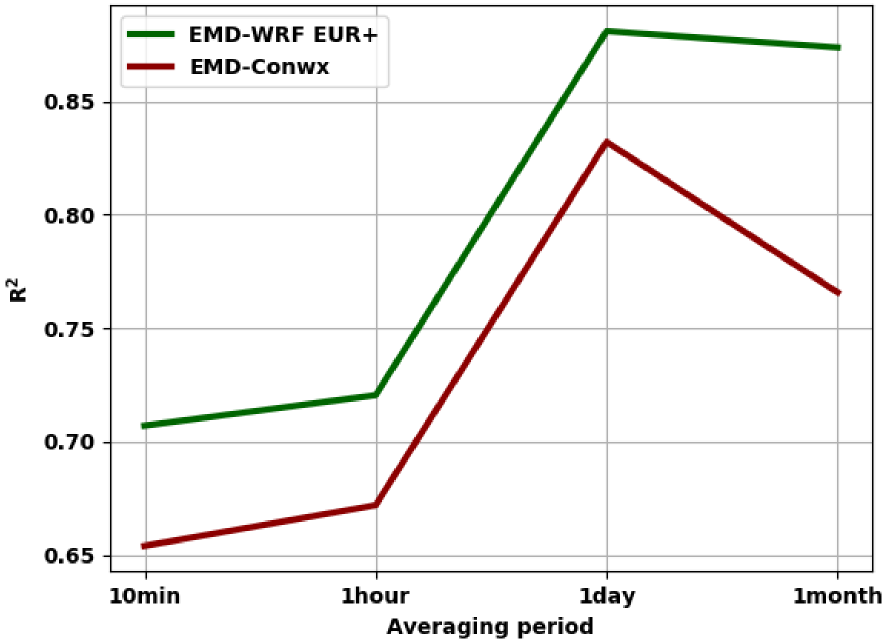



*Figure 5: Mean correlations across the 160 masts at different averaging periods.*

# 3 Acknowledgements

EMD-WRF Europe+ is developed as part of the windPROSPER project financed by Innovation Fund Denmark within the EUROSTARS framework.

Access and processing of NEWA data is attributed to the RECAST project also funded by Innovation Fund Denmark.

# 4 References

[1] Thøgersen, M. L. et al., 2018, windPRO Technical Note: EMD-WRF Global On-demand Mesoscale Services ERA5, ERA-Interim, MERRA2 and CFSR, EMD International, Aalborg, Denmark.

[2] Justus, C. G, Hargraves, W. R, Mikhail, A. and Graber, D., 1978, Methods for Estimating Wind Speed Frequency Distributions, Journal of Applied Meteorology, American Meteorological Society, USA.