windPRO

Technical Note

EMD-WRF Global On-demand Mesoscale Services
ERA5, ERA-Interim, MERRA2 and CFSR

Eurostars Project
windPROSPER

eurostars™

EMD International A/S
www.emd.dk

DATE
16 January 2018

PREPARED BY
EMD International A/S
Niels Jernes Vej 10
DK- 9220 Aalborg
T: + 45 96 35 44 44
E: emd@emd.dk

AUTHORS AND CONTRIBUTORS
Morten Lybech Thøgersen
Lasse Svenningsen
Thorkild Sørensen
Madalina Jogararu
EMD International A/S

Contents

Document History
2017-11-30: Second edition, amended with description of ERA5 data (first batch, 7 yrs, 2010-2016) – MLT
2017-05-05: First edition, with mesoscale service covering CFSR, MERRA2 and ERA-Interim – MJ

Acknowledgements
The current work is supported by the Danish Innovation Fund via the EUROSTARS framework.

1 Executive Summary

This technical note documents the quality of the EMD-WRF on-demand mesoscale products.
It describes and evaluates the different global reanalysis options, ERA5, ERA-Interim (ERA-I), MERRA-2 and CFSR, in terms of their long-term consistency and their correlation with observations. Both as a raw dataset and when used in downscaling in EMD-WRF mesoscale on-demand service. Results include more than 100 quality masts scattered globally.

ERA-5 is best, but covers only 7 years, so EMD-WRF ERA-I is the recommended long-term data product.
The results suggest that EMD-WRF ERA-I is the best long-term dataset for maximum correlation, if ERA-I's availability lag of three months is not an issue. ERA5 shows consistently higher correlations, but covers currently (2017) only seven years of data; not enough for a long-term dataset. ERA5 is released by ECMWF in batches during 2017-2019, and replaces ERA-I as the recommended dataset, when 20y are released (expected 2018-Q2).

EMD-WRF CFSR is the recommended dataset if immediate access to the most recent data is critical.
EMD-WRF CFSR has a quality very close to EMD-WRF ERA-I, with high correlations, but is available with only a few days delay to the current day. This study shows that EMD-WRF CFSR on-demand has a high level of long-term consistency although CFSR is combined of two separate datasets (CFSR and CFSv2). Correlation results show consistently high R^2 for all four EMD-WRF models, around 0.61-0.67 for hourly averages (0.75-0.81 for R).

2 Data Assimilation, WRF and Downscaling

Reanalysis datasets are used in climate and weather modelling as a source of meteorological information. These datasets are created by assimilating a large number of historical observations from many sources, like satellites, ground stations, weather balloons etc. The observations vary in space and time. They are typically assimilated by using the same model version throughout the entire assimilation period to improve consistency [1]. Datasets are generated globally, at different heights, surface grids and temporal resolutions, and contain various atmospheric observations. A number of reanalysis datasets are available, some with global coverage, and time span up to present time [2]. This study includes the four global datasets: ERA5, ERA-I, MERRA-2 and CFSR, all available in the EMD-WRF on-demand mesoscale product.

The WRF mesoscale model is a weather prediction system used to generate meteorological forecasts or hindcasts. WRF downscales the global datasets and the results can be used to generate spatial wind maps and as a valuable source of long-term time series wind data. WRF ingests several parameters, such as pressure and temperature fields from the global reanalysis dataset to generate initial and boundary conditions for the simulations. The WRF grid and output resolution is typically a few kilometres, ca. 3 km is often the preferred choice. Using the full windPRO modelling chain, it is possible to downscale the data from the global level to the mesoscale level (e.g. using EMD-WRF On-Demand Services) and further to the microscale level (e.g. using the windPRO scaler options, [3]) – see Figure 1.

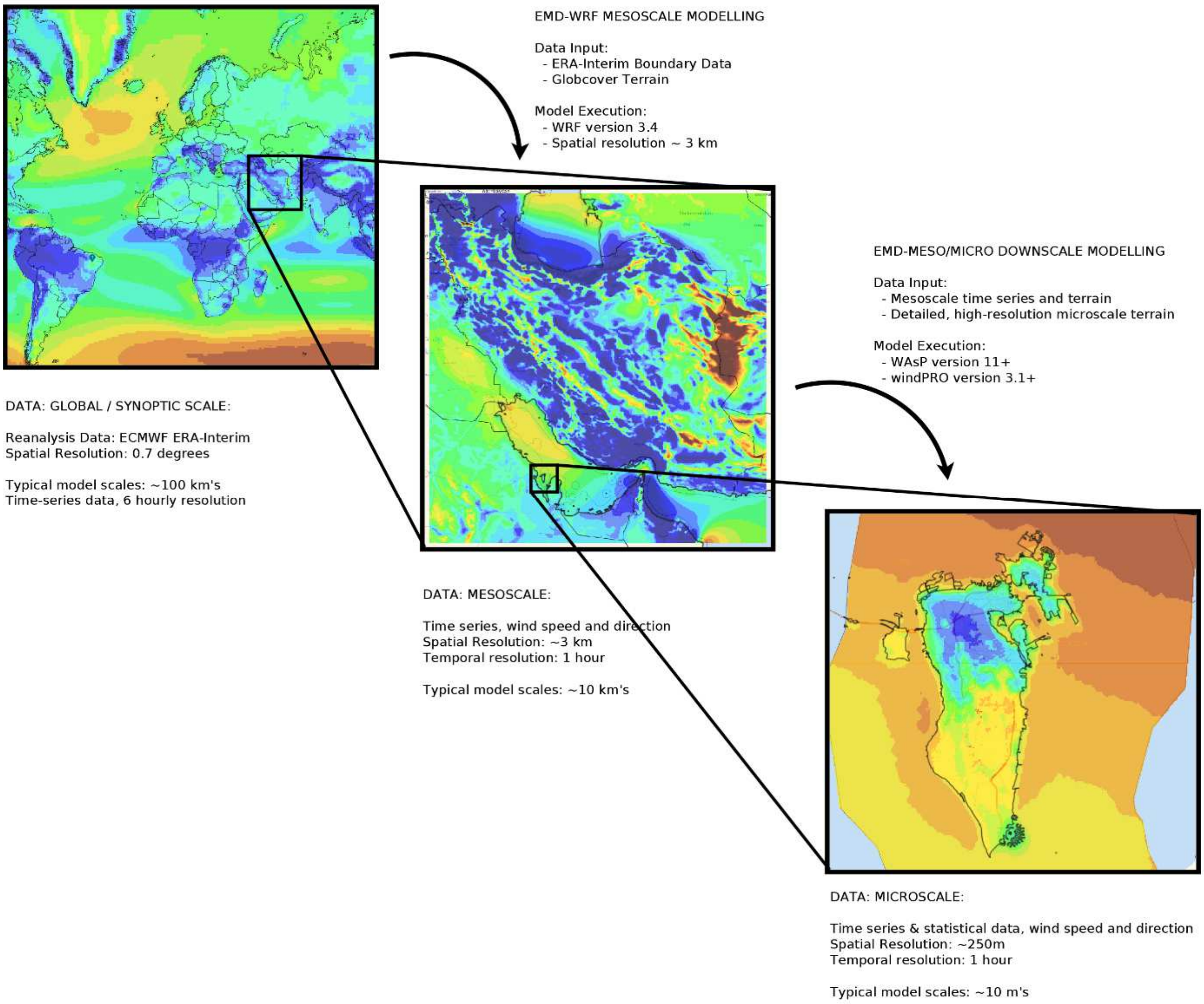
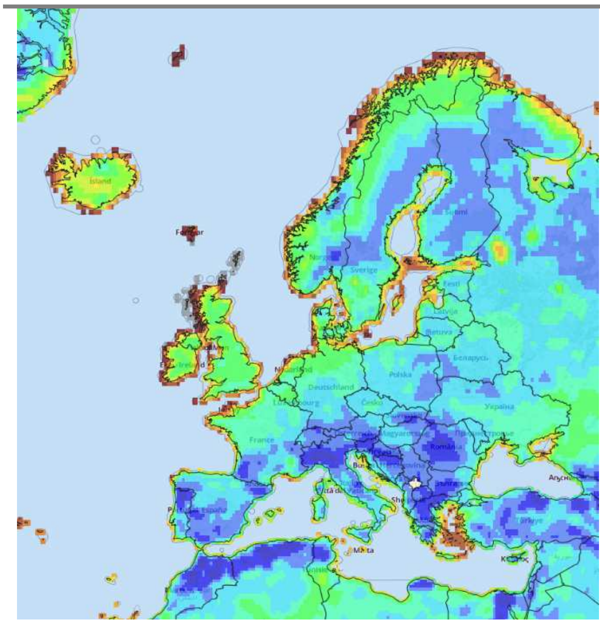


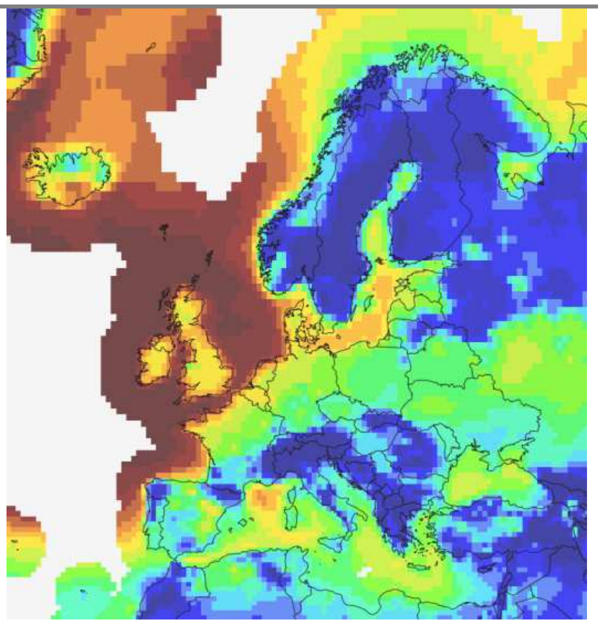
Figure 1: Example of EMD / windPRO downscaling chain.

3 Details: ERA5, ERA-I, MERRA-2 and CFSR

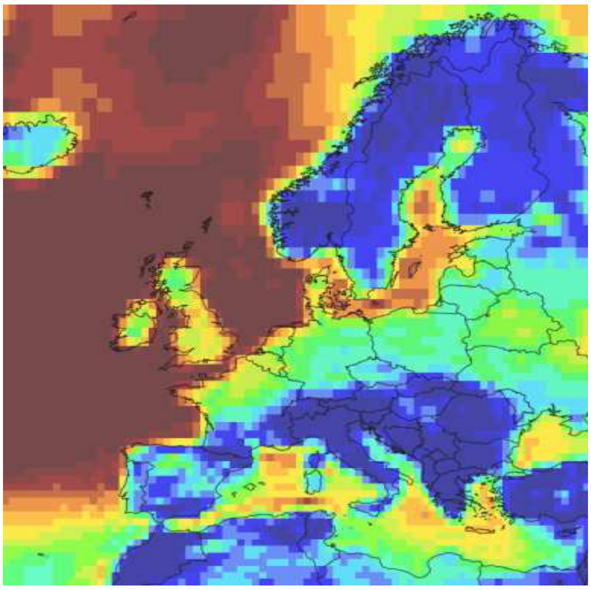
This chapter gives a brief introduction to the different global assimilation datasets that is used operationally as part of the EMD-WRF Mesoscale on-demand services. Generally, the newer datasets holds higher temporal and spatial resolutions, see Figure 2. Modern datasets also assimilates more data and are also able to use a wider range of different sensor-sources. A summary of the different datasets is given in the subsections below – with link to dataset references.



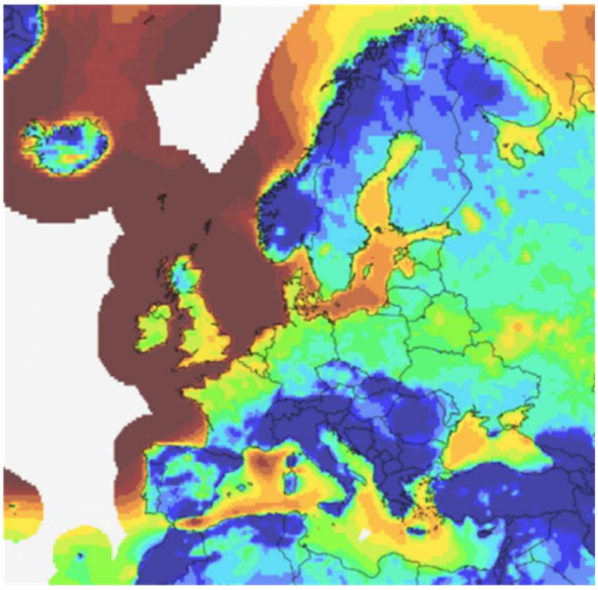
Wind Speed from CFSR



Wind Speed from MERRA-2



Wind Speed from ERA-Interim



Wind Speed from ERA5

Figure 2: Figures showing spatial resolution of the four global assimilation datasets. Figures originating from http://www.windprospecting.com.

3.1 CFSR

The CFSR (Climate Forecast System Reanalysis) datasets by NCEP is a third generation reanalysis product. A data assimilation algorithm based on GSI (Gridpoint Statistical Interpolation) is used to integrate a large number of observations in 6 hour time steps. Observations are mostly drawn from NCEP history archives, as well as other international centres like ECMWF, ESA, JMA, BOM etc [4]. NCEP has released two CFS time series products:

1) CFSR global dataset, which is an atmospheric, oceanic and land surface product with a time span from 1979 until 2011, when it was discontinued. The dataset is available at 0.3 degree global resolution (~38km) with 64 atmosphere levels.
2) CFSv2 dataset was released as a seamless extension to CFSR dataset. It is available from beginning of 2011 until present and is available with an increased global resolution of 0.2 degrees [5]. Apart from model improvements, the new dataset is available in near real time typical availability lag is one day to present.

When used in EMD-WRF downscaling the combined dataset of CFSR and CFSv2 is used to cover data up to present. Surface wind data from CFSR and CFSv2 are available as online-datasets in WindPRO [6] – and it is available at native grid resolutions, 0.2 and 0.3 degrees. Also, an extended CFSR dataset (further called CFSR-E) is made available, by extracting the surface wind vectors (u and v) at 0.5 degree global resolution, from both CFSR and CFSv2 datasets [7].

3.2 MERRA-2

MERRA 2 dataset is produced by NASA Goddard Space Flight Center. A large number of observations – actually millions at each 6 hourly interval - are assimilated by the GEOS-5 system (Goddard Earth Observation System Version 5), see Figure 3. MERRA 2 has a spatial resolution of 0.5 degrees (~ 50 km), and is available from 1980 until present [8], [9].

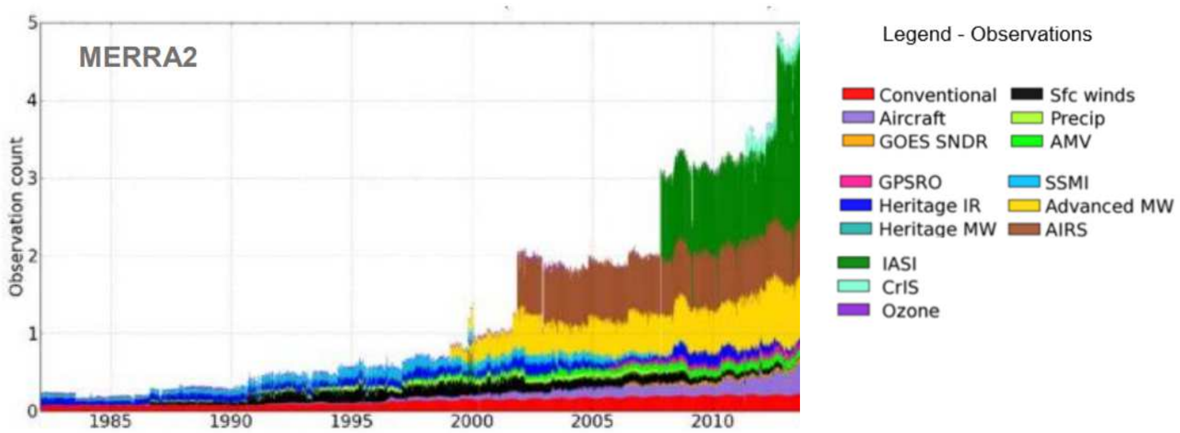


Figure 3: Observations assimilated in the MERRA2 datasets for the period 01.1980 until 12.2014. Units are millions per 6 hours. From Bosilovich et al [10].

3.3 ERA-Interim

ERA Interim is a global atmospheric reanalysis dataset produced by ECMWF (European Center for Medium Range Weather Forecast). The atmospheric and data assimilation system, IFS (Integrated Forecasting System) is used to produce ERA Interim – the system includes a 4-dimensional variational analysis with a 12-hour analysis window. ERA has a spatial resolution of approximately 80 km, and a time range from 1979 until present [11].

3.4 ERA5

ERA5 is a climate reanalysis dataset developed through the Copernicus Climate Change Service (C3S) and processed/delivered by ECMWF. The dataset is intended to replace the ERA-Interim dataset from ECMWF shortly after the ERA5 dataset is complete. ERA5 dataset has several improvements compared to ERA-Interim:

- Newer modelling system
- More observations used in the assimilation
- Higher horizontal resolution (around 31 km compared to 79 km in ERA-Interim)
- Higher vertical resolution

Release of the ERA5 data follows a schedule from ECMWF, where the 7 years from 2010-2016 have been released on July 17th 2017, and the remaining part from 1979-2009 will follow in medio 2018. Details about ERA5 is available from the Copernicus C3S service [12].

3.5 Dataset Comparison Table

Figure 4 (below) shows an overview of the setup and configuration of the four different global data assimilation models – and the dataset availability (ERA5, ERA-Interim, MERRA2 and CFSR).

Parameter \ Dataset	ERA5	ERA-Interim	MERRA2	CFSR / CFSv2
Vertical levels	137	60	72	64
Horizontal resolution	~31 km	~80 km	~50 km	~38km/~25km
Upper modelling level	0.01hPa (~80 km)	0.1hPa (~60 km)	0.01hPa (~80 km)	0.26 hPa (~55 km)
Temporal resolution	1-hourly	6-hourly	1-hourly	1-hourly
Release schedule	Monthly*	Monthly	Monthly	Daily
Assimilation model	IFS Cycle 41r2	IFS Cycle 31r2	GEOS 5.12.4	Grid-Point Statistical Interpolation, GSI
Spatial grid type	Reduced Gaussian	Reduced Gaussian	Cubed sphere	Varies
Period available (now)	2010-2016	1979-present	1980-present	CFSR: 1979-2010 CFSv2: 2011-present
Period available (at completion)	1950-present	1979-present	1980-present	CFSR: 1979-2010 CFSv2: 2011-present
Delay in data delivery	3 months *)	3 months	1-2 months	1 day

Figure 4: Comparison of key parameters in the different reanalysis datasets in EMD-WRF Mesoscale On-Demand.

*) Note: A preliminary version of the ERA5 data – named ERA5T – will be a weekly release.

3.6 'Raw' Global Models vs Observations

Traditionally – and since the NCEP/NCAR Reanalysis version 1.0 (approx. year 2007) - the global reanalysis model data has been directly used as long-term reference sources for measure-correlate-predict (MCP) methods in the wind analyst community. To some degree, this is still done today (2017), even if downscaled mesoscale data is quickly taking over as the main source of data. The presented analyses show the hourly wind speeds from ERA5, ERA-I, MERRA2 and CFSR that have been compared to measured wind speeds from 107 tall meteorological masts around the globe. The masts have top sensor heights in the range from 60m to 140m. The R^2-correlation has been calculated for all sites - and ERA5 data shows a significant improvement over all other datasets – with an average R^2-correlation is increased with 0.10 and the variation/spread is also lower.

	Parameter	Dataset ->	ERA5	ERA-Interim	MERRA2	CFSR / CFSv2
Hourly	Mean Value		**0.65**	0.54	0.54	0.47
	Standard Deviation		**0.14**	0.17	0.15	0.16
	Minimum		**0.20**	0.10	0.17	0.08
	Maximum		**0.88**	0.81	0.84	0.80
Daily	Parameter	Dataset ->	ERA5	ERA-Interim	MERRA2	CFSR / CFSv2
	Mean Value		**0.83**	0.72	0.74	0.72
	Standard Deviation		**0.11**	0.18	0.15	0.14
	Minimum		**0.35**	0.17	0.27	0.18
	Maximum		**0.96**	0.93	0.96	0.93
Monthly	Parameter	Dataset ->	ERA5	ERA-Interim	MERRA2	CFSR / CFSv2
	Mean Value		**0.86**	0.78	0.76	0.74
	Standard Deviation		**0.12**	0.22	0.21	0.20
	Minimum		**0.34**	0.03	0.10	0.11
	Maximum		**0.99**	0.98	**0.99**	0.97

Figure 5: Wind Speed Correlation (R^2) at hourly, daily and montly averaging times. Data from 107 masts. Notes: ERA-I is interpolated to hourly values. CFSR/CFSv2 is from EMD CFSR-E dataset (0.5 deg). Green color-boldface color shows best dataset for the metric being considered.

The Figure 6 shows how ERA5 improves correlation (R) over MERRA2 – when considering the 107 local masts. On 93% of the sites we find an improvement (ΔR>0), and at 66% of the site the improvement in ΔR is larger than 0.05.

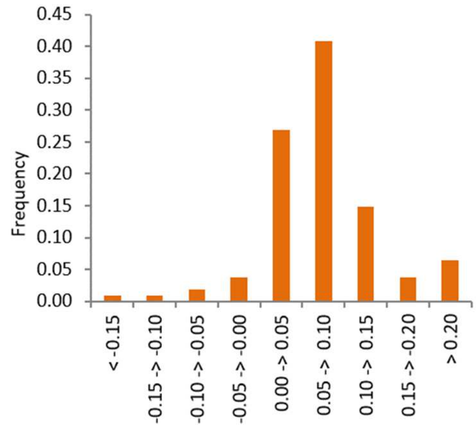


Figure 6: Improvement in correlation, ΔR, ERA5 and MERRA2 vs local masts.

4 'Downscaled' EMD-WRF vs Observations

This chapter quantifies the correlation of EMD-WRF Mesoscale On-Demand datasets – downscaled from ERA5, CFSR, MERRA2 and ERA-I – by comparing to local mast observations. The EMD-WRF datasets are validated against high quality observations from top anemometers on 107 masts around the world. Each mast typically covers at least a full year of data and the coefficient of determination (R^2-correlation) is calculated based on concurrent wind speed samples for different averaging periods (from hourly over daily to monthly). Each of the 107 mast locations is evaluated using the EMD-WRF on-demand mesoscale model – executed using each of the four global datasets as boundary data. The results are presented in the figures below, where the masts have been sorted according to R^2-correlation in the EMD-WRF ERA-I data. Figure 7 clearly shows that the red curve - representing EMD-WRF ERA5 is significantly better than the other datasets. It is also clear that the masts with moderate correlation holds the largest improvement (grey line, secondary axis). The histograms to the right show the improvement in R^2 from ERA-Interim to ERA5 – the improvement is obvious (positive number means improvement).

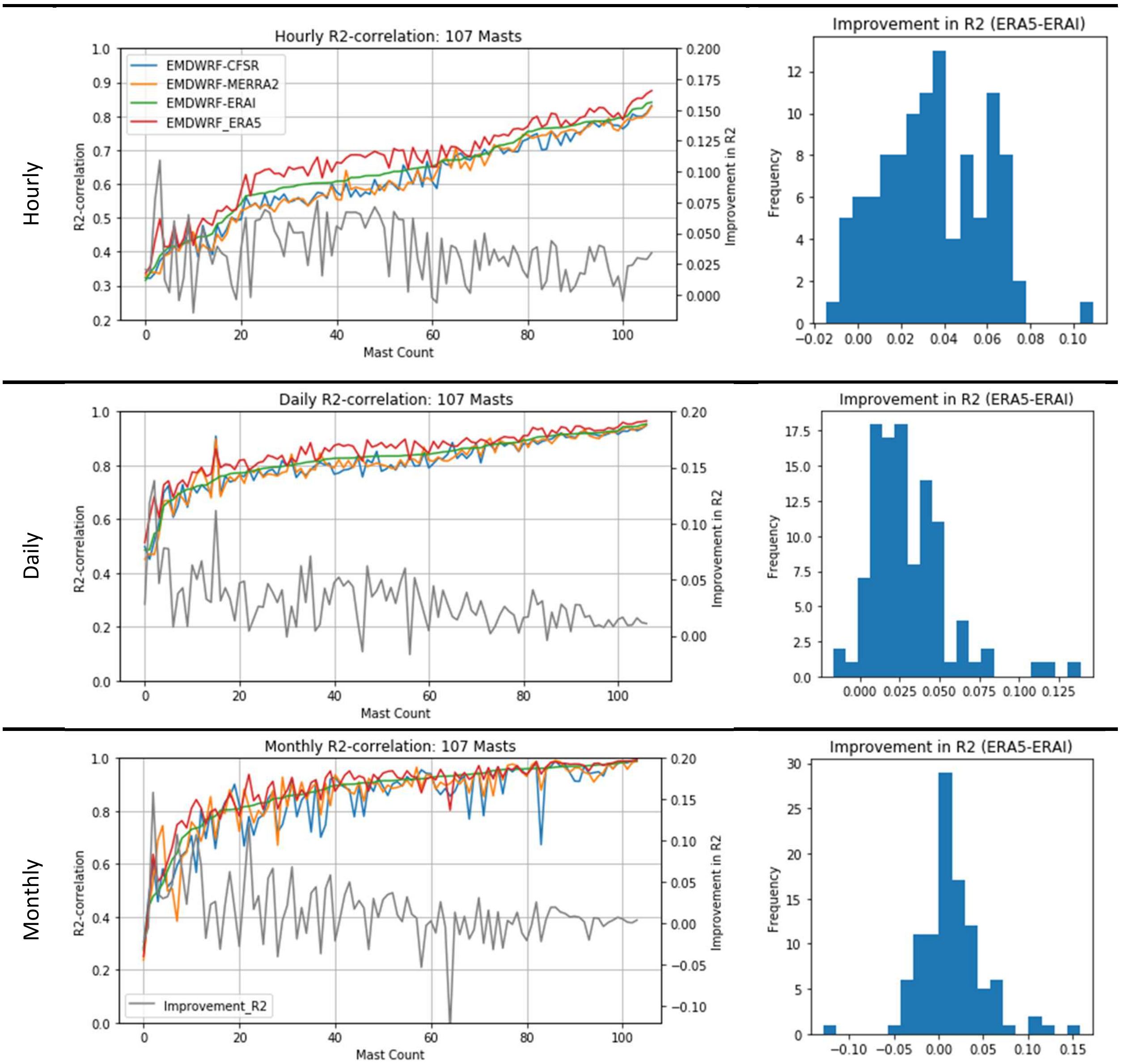


Figure 7: Correlation - R^2, from the different EMD-WRF configurations vs. local mast data.

The results of Figure 7 are summarized in the table in Figure 8. The summary shows the differences between the four datasets with EMD-WRF on-demand ERA-5 and ERA-Interim showing best correlation results overall.

	Parameter Dataset ->	ERA5	ERA-Interim	MERRA2	CFSR / CFSv2
Hourly	Mean Value	**0.67**	0.64	0.61	0.61
	Standard Deviation	**0.12**	**0.12**	0.13	**0.12**
	Minimum	**0.34**	0.32	0.33	0.32
	Maximum	**0.88**	0.84	0.83	0.83
Daily	**Parameter Dataset ->**	**ERA5**	**ERA-Interim**	**MERRA2**	**CFSR / CFSv2**
	Mean Value	**0.86**	0.83	0.81	0.81
	Standard Deviation	**0.08**	0.09	0.10	0.09
	Minimum	**0.51**	0.49	0.45	0.45
	Maximum	**0.96**	0.95	0.95	0.95
Monthly	**Parameter Dataset ->**	**ERA5**	**ERA-Interim**	**MERRA2**	**CFSR / CFSv2**
	Mean Value	**0.89**	0.87	0.86	0.84
	Standard Deviation	**0.12**	0.13	0.14	0.14
	Minimum	0.25	0.27	0.24	**0.28**
	Maximum	**0.99**	**0.99**	**0.99**	**0.99**

Figure 8: Summary of R^2-correlation for the four different reanalysis datasets used in EMD-WRF Mesoscale On-Demand. Green color-boldface color shows best dataset for the metric being considered.

The results from the different analyses are shown in the box and whisker plot, Figure 9 – where the mean value is plotted with a green triangle, the box represents the median, 25% and 75% percentiles and the bars the min/max values. It is clear, that the ERA-5 represents the best datasets in terms of correlation and has the most narrow distribution (illustrated as the box size).

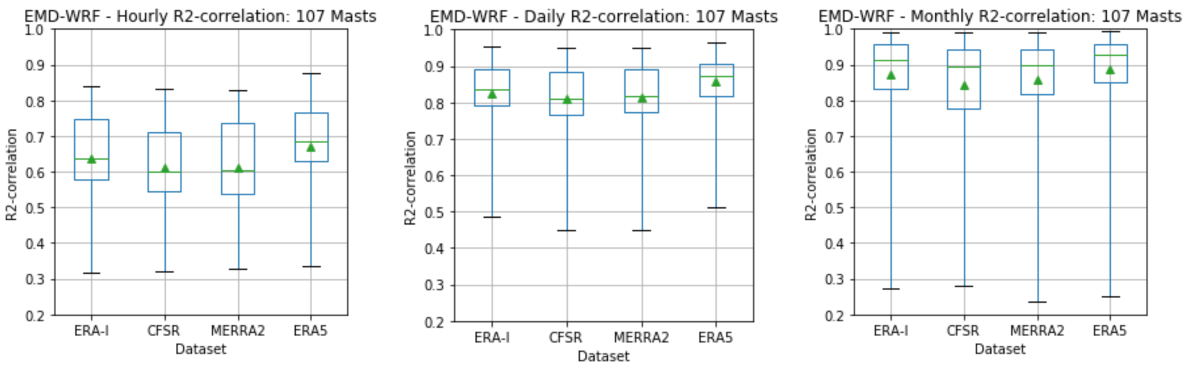


Figure 9: Box and whisker plots – R^2-correlation for the for EMD-WRF Mesoscale On-Demand Datasets.

4.1 Analysis of Regional Differences

The 107 masts have been split into regional datasets, covering four regions: Europe, America(s), Asia and Africa. The focus is to investigate if any regional differences exist – and if some reanalysis datasets perform better or worse at particular locations. Each sub-dataset holds at least 17 mast, with the main part in Europe.

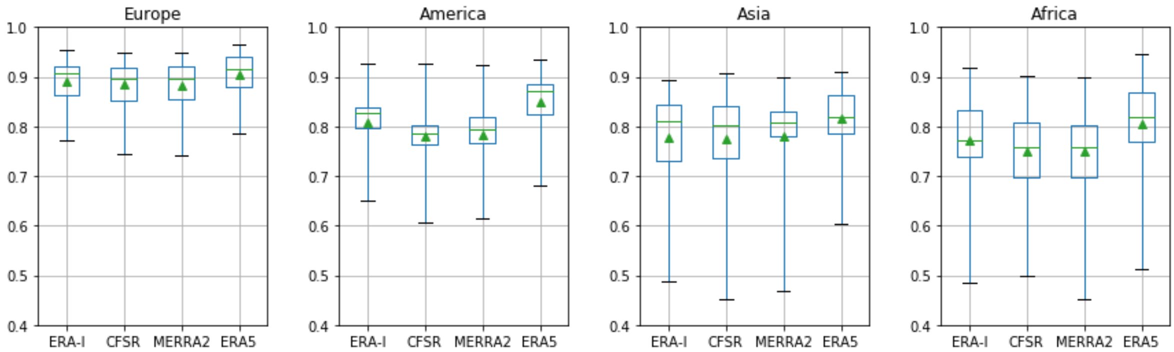


Figure 10: Daily R2-correlations from Four Different Regions: Europe, Americas, Asia and Africa.

Figure 10 outlines the results as a box and whisker plot - where the mean value is plotted with a green triangle, the box represents the median, 25% and 75% percentiles and the bars the min/max values. It is clear, that the ERA-5 represents the best datasets in terms of correlation, however also that it is not always the model with the least spread (visualized as the size of the box). In general – the correlation of the ERA5 based data is better than the other models, but in some cases, it may be of value to also use and investigate alternative data sources.

4.2 Conclusions

The main conclusions of this section are:

- ERA5 as input to WRF (or on its own) is a significant improvement over previous reanalysis datasets
- The standard deviation / spread is smaller – so the probability of larger errors is smaller when using ERA5 compared to other reanalysis datasets
- On sites where moderate correlation is found with previous modelling – these benefit the most from the improved ERA5 dataset
- Until a longer period of ERA5 data become available, ERA-Interim is the preferred choice for long-term correction analysis

5 Temporal Consistency: CFSR-E & EMD-WRF CFSR

CFSR is currently the only dataset with a near daily update frequency. However, it is also a dataset 'merged' from two distinct sources (CFSR and CFSv2) – even if they are closely related. So, it is an important question to ask whether the combined dataset of CFSR and CFSv2 is consistent over time? Can this dataset really be used as a trusted long-term reference? This question is addressed in the following analysis, which compares the consistency over time of both the raw CFSR-E reanalysis and the downscaled dataset EMD-WRF CFSR.

5.1 Trendlines

The trendline-analysis is made by evaluating the slope of the trends over a time span of 20 years, from 1996 until 2015 including both years using a 12 months moving average following the method proposed in [13]. The analysis is performed for four grid nodes – each one in different regions of the world: Europe, South Korea, South Africa and South America. The 12 months moving average is applied to minimize the effect of short-term fluctuations and smoothen the data emphasizing long term cycles or trends. After averaging, the time series is normalized using the mean of the complete series. Finally, a linear best fit is used to quantify the slope of the trend in the averaged and normalized time series.

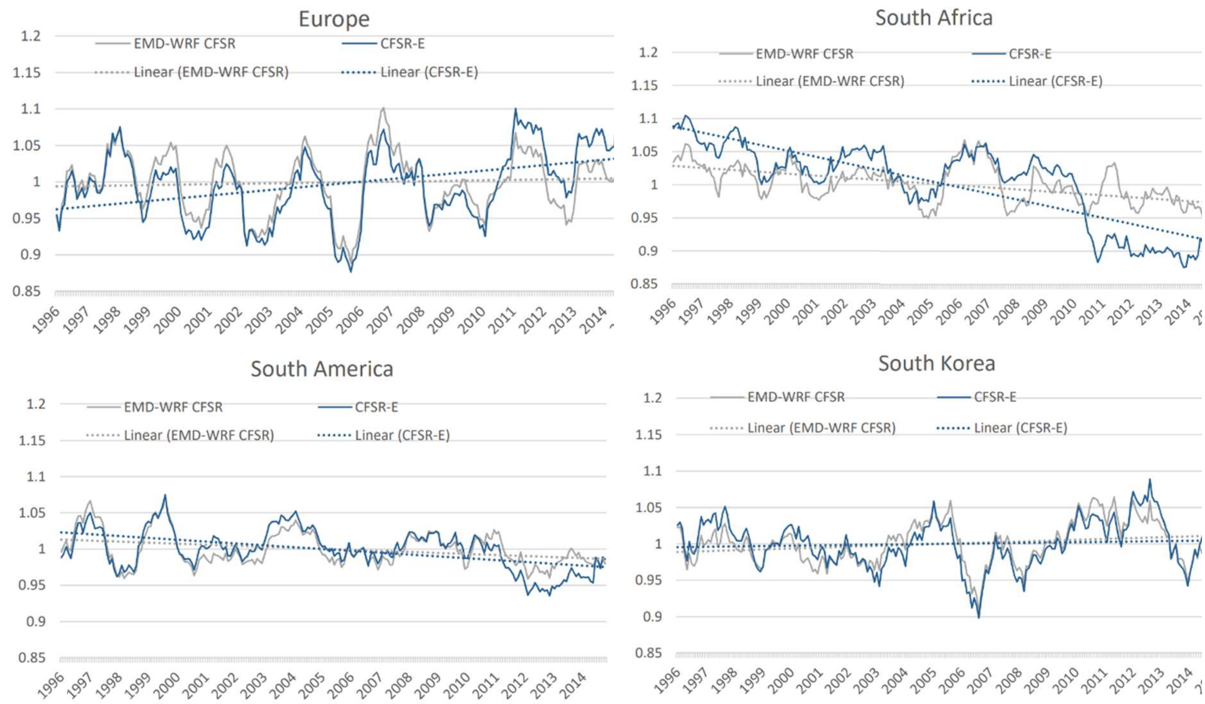


Figure 11: Comparison of 12 months moving average - CFSR-E and EMD-WRF CFSR in four parts of the world.

Figure 11 illustrates the results for the 'raw' CFSR-E and the EMD-WRF CFSR in the four parts of the world. The moving average series of the two datasets show consistent fluctuations throughout the years, except for the last part of the dataset. In the beginning of 2011 there is a clear separation between raw CFSR-E and EMD-WRF CFSR, driven by a discontinuity in the raw CFSR-E. This discontinuity is clearly reflected in the linear trend lines, which have a significantly larger slope for the raw CFSR-E data.

The observed discontinuity in the raw CFSR-E coincides with the time where the two CFSR datasets (CFSR and CFSv2) are merged. Thus, the temporal inconsistency of the raw CFSR-E must relate to the differences in native models, grid resolutions or data assimilations. However, it is very important to note that where the raw CFSR-E dataset is based on surface winds, the EMD-WRF CFSR dataset is driven by e.g. pressure and temperature fields. Therefore, the inconsistency seen around 2011 in the raw CFSR-E dataset is not observed in EMD-WRF on-demand CFSR mesoscale dataset.

The slopes of the linear trends in the two datasets are consistent within the same hemisphere, but with very different magnitudes – and with much smaller slopes for the EMD-WRF. Europe and South Korea show an increasing trend, while a negative trend is observed in South America and South Africa.

6 Bibliography

[1] Earth System Research Laboratory, Physical Sciences Division, "U.S. Departament of Commerce; National Oceanic & Athmosferic Administration," 2014. [Online]. Available: http://www.esrl.noaa.gov/psd/data/gridded/reanalysis/. [Accessed March 2017].

[2] National Center for Atmospheric Research Staff; D. Dick; D. Shea; J. Fasullo; J. Walsh, "The Climate Data Guide: Atmospheric Reanalysis: Overview & Comparison Tables," December 2016. [Online]. Available: https://climatedataguide.ucar.edu/climate-data/atmospheric-reanalysis-overview-comparison-tables. [Accessed January 2017].

[3] EMD International A/S, "WindPRO Quick Guides - Scaler Options," 2017. [Online]. Available: http://help.emd.dk/knowledgebase/default.aspx. [Accessed 01 12 2017].

[4] S. Suranjana, M. Shrinivas , P. Hua-Lu , W. Xingren, W. Jiande, N. Sudhir, P. Tripp, R. Kistler, J. Woollen, D. Behringer, H. Liu, D. Stokes, R. Grumbine, G. Gayno, J. Wang, Y.-T. Hou, H.-Y. Chuang, H.-M. H. Juang , J. Sela, M. Iredell, R. Treadon, D. Kleist, P. Van Delst, D. Keyser , J. Derber, M. Ek, J. Meng, H. Wei, R. Yang , S. Lord, H. Van Den Dool, A. Kumar, W. Wang, C. Long, M. Chelliah, Y. Xue, B. Huang, J.-K. Schemm, W. Ebisuzaki, R. Lin , P. Xie, M. Chen, S. Zhou, W. Higgins, C.-Z. Zou, Q. Liu, Y. Chen, Y. Han, L. Cucurull, R. R. Reynolds, G. Rutledge and M. Goldberg, "The NCEP Climate Forecast System Reanalysis," 12 April 2010.

[5] NCAR; Research Data Archive Computetional & Information Systems Lab, "NCEP Climate Forecast System Reanalysis (CFSR)," 2017. [Online]. Available: http://rda.ucar.edu/pub/cfsr.html. [Accessed 26 March 2017].

[6] M. L. Thøgersen, "WindPRO Documentation on Online Data: Wind Data," EMD International A/S, 26 5 2013. [Online]. Available: http://help.emd.dk/mediawiki/index.php?title=Category%3AWind_Data. [Accessed 01 12 2017].

[7] M. L. Thøgersen, "WindPRO Documentation on Online Data: CFS and CFSR Data," 26 5 2013. [Online]. Available: http://help.emd.dk/mediawiki/index.php?title=CFS-_and_CFSR_Data. [Accessed 1 12 2017].

[8] M. L. Thøgersen, L. U. Svenningsen, T. G. Sørensen and S. A. Coasta, "Is MERRA2 able to replace MERRA as a trusted wind reference dataset?," in *Brasil Wind Power 2016*, Rio De Janeiro, 2016.

[9] National Aeronautics and Space Administration Goddard Space Flight Center, "Modern-Era Retrospective analysis for Research and Applications, Version 2," December 2015. [Online]. Available: https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. [Accessed 11 2016].

[10] M. e. a. Bosilovich, "MERRA-2: Initial Evaluation of the Climate - Technical Report Serieson Global Modeling and Data Assimilation – Volume 43," NASA/TM–2015-104606/Vol. 43, 2015.

[11] P. Berrisford, D. Dee, P. Poli, R. Brugge, M. Fuentes, K. Fielding, P. Kållberg, S. Kobayashi, S. Uppala and A. Simmons, "The ERA-Interim archive Version 2.0," European Centre for Medium-Range Weather ECMWF, 2011. [Online]. Available: http://www.ecmwf.int/en/research/climate-reanalysis/era-interim. [Accessed 11 2016].

[12] ECMWF, "Copernicus Climate Change Service - About ERA5," [Online]. Available: http://climate.copernicus.eu/products/climate-reanalysis. [Accessed 30 11 2017].

[13] S. Liléo, "Long-term correction of wind measurements State-of-the-art, guidelines and future work," Elforsk 13:18, 2013.

Appendices – Available upon request

Appendices are available upon request, please contact the EMD Mesoscale Modelling team through support@emd.dk.