Data Documentation of DWD Weather Model Data for Energy System Simulation: 2015

Contributing Authors:

Gerrit Erichsen
gerrit.erichsen@tuhh.de

The author resides at:
Institute of Energy Systems
Hamburg University of Technology
Denickestraße 15, D-21073 Hamburg, Germany

January 2020

ABSTRACT

This is a short documentation of how the weather data of the *Deutscher Wetterdienst* (DWD) was obtained and prepared for the repository entry *"DWD weather model data for energy system simulation: 2015"*. Besides a short description a quality analysis was undergone comparing the obtained model data of the DWD to the weather station observation data of the DWD. Lastly, a very simplified categorization of its energy potential and demand is given. This document is meant to give a glimpse into the data set and not meant to be a holistic review of the DWD's weather model.

1. INTRODUCTION

Weather data is one of the crucial elements for energy system simulation, as it dictates the profile of volatile renewable energy (VRE) sources such as photovoltaic, solar thermal heat and wind power. A fixed scenario configuration may apply for one set of weather, but be insufficient for the next.
The weather data of the DWD is often used in publications such as [1–3], but not yet easily accessible and applicable. The data repository at hand aims to fix that. Additionally, it provides wind data for various heights above ground, which eliminates the need for estimating roughness lengths when extrapolating from usually available wind data.
The data was obtained from DWD and is made available under GeoNutzV [4] license. The targeted use of the data was the research project VEREKON (FKZ: 03ET7067A), an energy system analysis and is therefore mostly suitable for that.
In the following the data processing is documented, data quality is estimated, and a very rough categorization in terms of use for energy system analyses is given.

2. MATERIALS AND METHODS

The main material is the weather data received via Pamore [5]. From the available data, the final data set is derived. Additional material encompasses the climate data centre's observation data (CDC) [6] for comparison.

2.1. Pamore data

Pamore is a service of the DWD to access forecast and other data of their various weather models. For the solar, wind and temperature data the COSMO-DE model's [7] assimilation analysis data was requested.

Table 1: Variable names in COSMO and repository

Variable name COSMO	Variable Name repository	Changed for repository
ASWDIFD	ASWDIFD	✓
ASWDIR	ASWDIR	✓
T_2M	TMP	✓
U	WZU	✓
V	WMV	✓

2.1.1 DWD model run

The DWD operates its models on a regular basis, whereby the patterns of model runs changed over time as the DWD continuously improved their models, as can be seen from Figure 1, Figure 2 and Figure 3. Generally, the assimilation analysis runs before the main analysis followed last by the forecast runs. The COSMO forecast runs are executed at 0, 3, 6, 9, …, 21 o'clock UTC. For the COSMO-DE model the DWD's weather station's data on pressure and wind speed components at a height of 10 m are entered as a boundary condition as the COSMO-DE model's main target is an accurate mapping of storm cells [7]. The assimilation analysis data of the COSMO models can be interpreted as a historical log, according to personal correspondences.

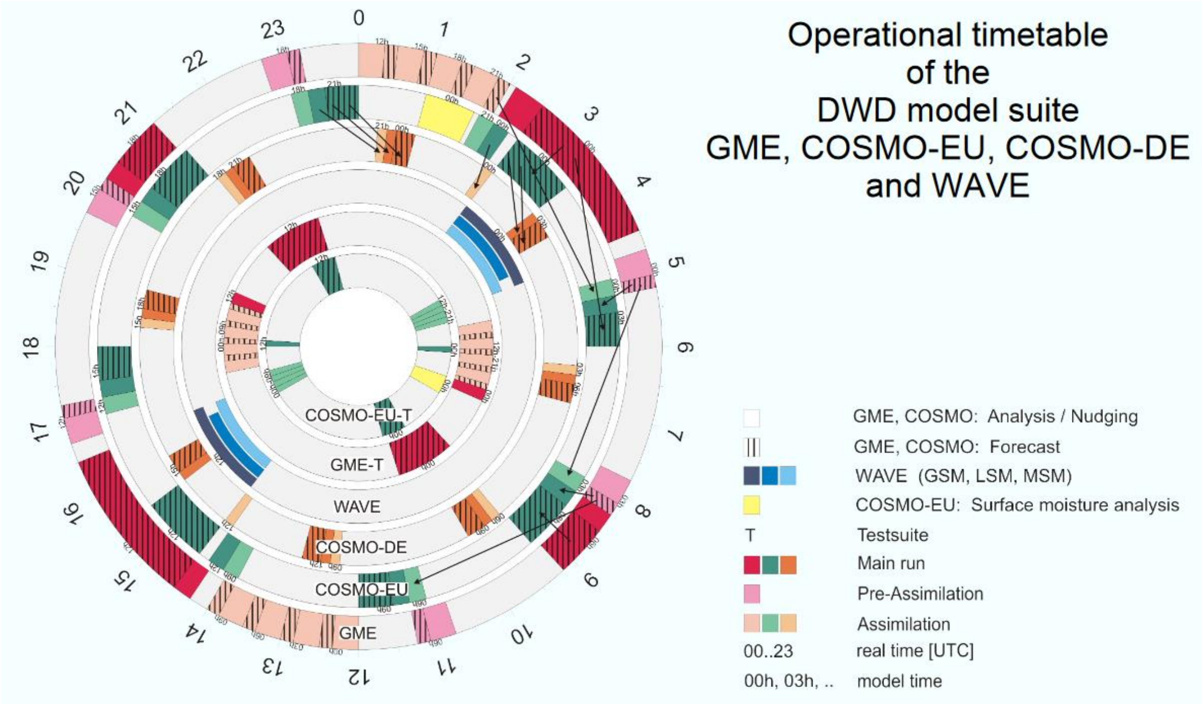

Figure 1: Model clock from 2009 [8]

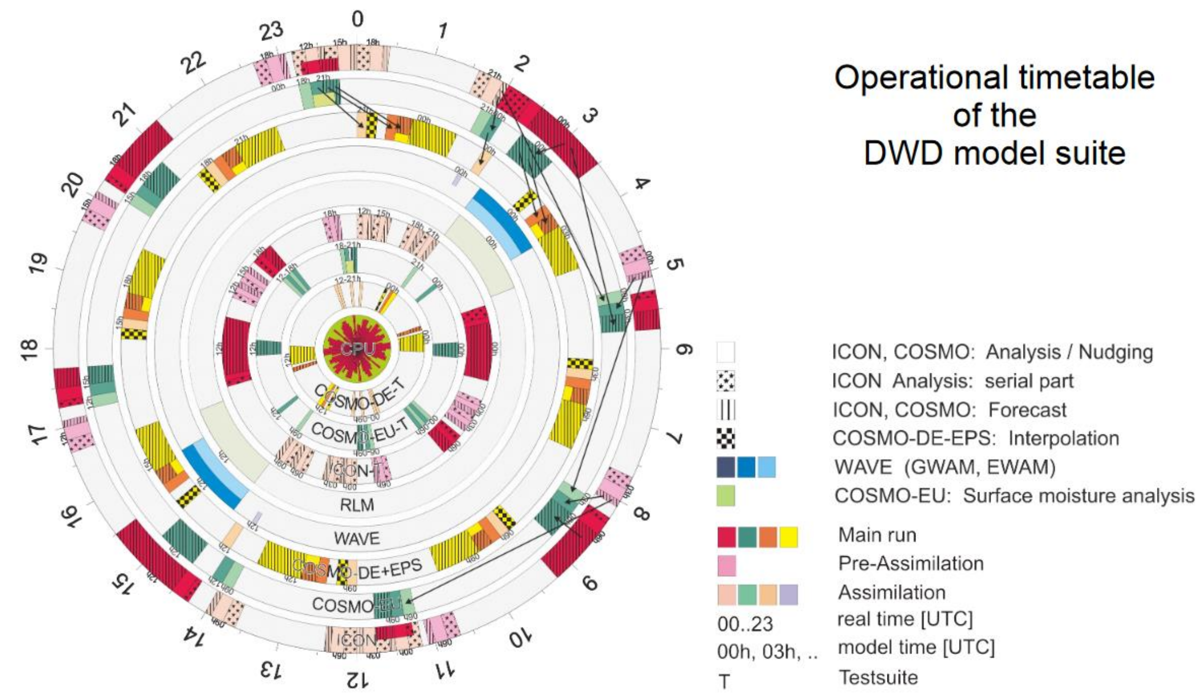

Figure 2: Model clock presented 2015 [9]

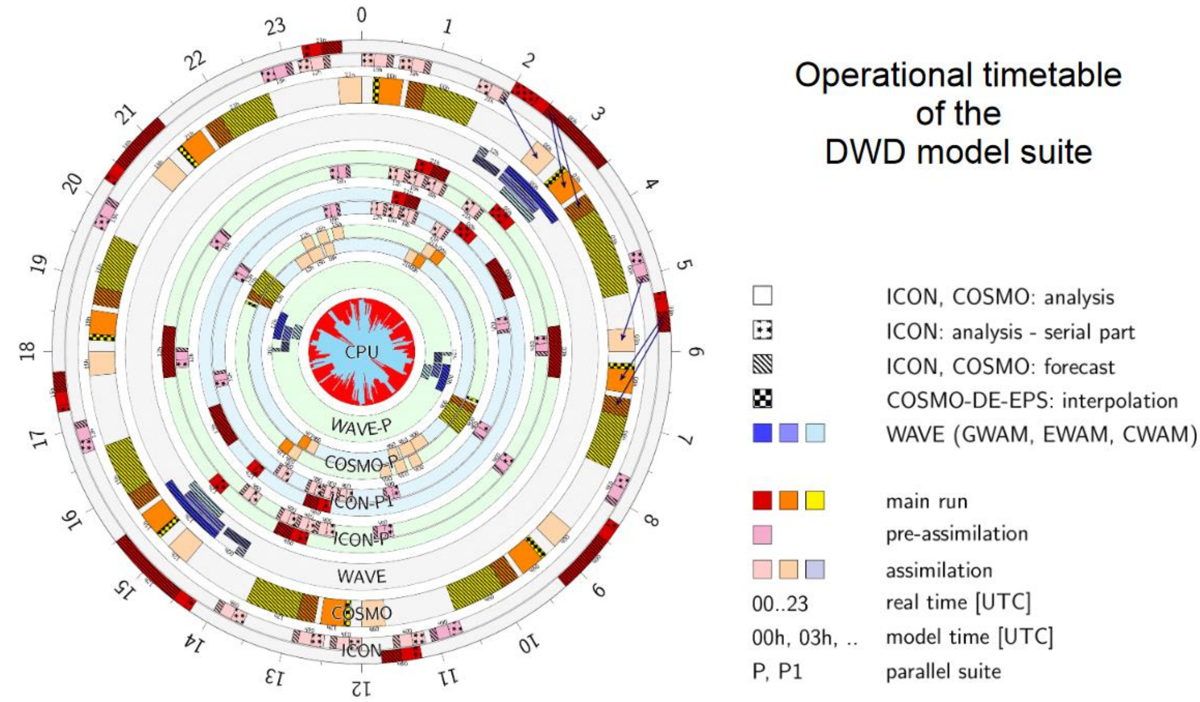

Figure 3: Model clock presented 2017 [10]

2.1.2 Assimilation analysis

The assimilation analysis is a re-run of the COSMO models, with data of wind speeds at 10 m height and air pressure at ground level from SYNOP weather stations as boundary conditions. The assimilation analysis reruns at the aforementioned regular intervals for a simulation time of 3 hours. The obtained data is then archived for "eternity" and accessible via the Pamore platform.

2.1.3 Model Grid

The COSMO DE and EU grids at the DWD are rotated pole grids with their rotated north at 40 °N and 170 °W. Starting at model latitude 5 $°S_{rotated}$ and longitude 5 $°W_{rotated}$ centre grid points are generated going northeast to latitude 6.5 $°N_{rotated}$ and longitude 5.5 $°E_{rotated}$ in 0.025 ° resolution (translating to a resolution of roughly 2,8 km).

2.2. Resulting data set

The resulting data set is archived as an array of single files for each of the requested weather variables for the entire year and saved as an *.hdf5 file, which is binary file format designed for scientific data sets. For a brief overview of the data, check the section Appendix – Datastructure near the end of the document. The data is given at steps in hourly resolution. Timestamps could be generated starting at 01. January 0 o'clock UTC each year and the suggested interpretation is that, the variables start at these timestamps and are valid up until the timestamp of the next step.

Table 2: Variable translation and units

Variable in repo	Meaning	Unit
ASWDIFD	Average diffuse solar irradiation on the horizontal plane	W/m^2
ASWDIR	Average direct solar irradiation on the **horizontal** plane	W/m^2
TMP	Temperature at 2m above ground	°C
WZU	Zonal wind speed (along latitudes), positive towards east	m/s
WMV	Meridional wind speed (along longitudes), positive towards north	m/s

To convert the obtained data to the form described in Table 2 some recalculation was necessary as described in the following sections.

2.2.1 Temperature data adjustment

The temperature data was adjusted from the Kelvin temperature scale to a Celsius temperature scale, by simply using

$$\vartheta_{i,j}(t) = T_{\text{data},i,j}(t) - 273.15 \tag{2.1}$$

for value generation.

2.2.2 Recalculation of solar data

The solar data is archived as averaged values over every assimilation run, so a 1, 2, 3 hour average value is received via Pamore and recalculated to one hour average values in the following way

$$\begin{gathered} I_{i,j}(t_1) = \bar{I}_{\text{data},i,j}(t_1) \\ I_{i,j}(t_2) = \bar{I}_{\text{data},i,j}(t_2) \cdot 2 - \bar{I}_{\text{data},i,j}(t_1) \\ I_{i,j}(t_3) = \bar{I}_{\text{data},i,j}(t_3) \cdot 3 - \bar{I}_{\text{data},i,j}(t_2) \cdot 2, \end{gathered} \tag{2.2}$$

whereby $I_{i,j}(t)$ is the resulting data of the data set and $\bar{I}_{\text{data},i,j}(t)$ to the original average data of Pamore. The integer index of t refers to the duration over which the average of $I_{\text{data},i,j}(t)$ is calculated in hours.

2.2.3 Recalculation of wind data

The wind data, requested as U and V variables, are zonal and meridional wind speed along the model grid and need to be transformed to match the actual global grid. Therefor in a first step, the speed components for the grid's centre points are calculated via

$$\begin{gathered} u_{\text{m},i,j} = \frac{u_{i,j} + u_{i-1,j}}{2}, \forall\, i = 2 \dots n_{\text{Latitudes}}, j = 2 \dots n_{\text{Longitudes}} \\ v_{\text{m},i,j} = \frac{v_{i,j} + v_{i,j-1}}{2}, \forall\, i = 2 \dots n_{\text{Latitudes}}, j = 2 \dots n_{\text{Longitudes}} \end{gathered} \tag{2.3}$$

leaving out the grid's anchor points, as no values upstream of these exist. This does not affect the dataset at hand, as the grid is cut later anyway. From these values at the center of the grid cell, the globalized components are calculated via

$$\begin{gathered} u_{\text{g},i,j} = u_{\text{m},i,j} \cdot \cos\delta_{i,j} + v_{\text{m},i,j} \cdot \sin\delta_{i,j} \\ v_{\text{g},i,j} = -u_{\text{m},i,j} \cdot \sin\delta_{i,j} + v_{\text{m},i,j} \cdot \cos\delta_{i,j} \end{gathered} \tag{2.4}$$

with

$$\begin{gathered} \delta_{i,j} = \arctan\left(\frac{\cos\varphi_\text{N}\sin(\lambda_\text{N} - \lambda_{\text{g},i,j})}{\cos\varphi_{\text{g},i,j}\sin\varphi_\text{N} - \sin\varphi_{\text{g},i,j}\cos\varphi_\text{N}\cos(\lambda_\text{N} - \lambda_{\text{g},i,j})}\right) \\ \forall\, i = 1 \dots n_{\text{Latitudes}}, j = 1 \dots n_{\text{Longitudes}} \end{gathered} \tag{2.5}$$

for the appropriate angle correction, whereby φ_N is the reference latitude and λ_N the reference longitude of the rotated pole grid used in Pamore. $\varphi_{\text{g},i,j}$ and $\lambda_{\text{g},i,j}$ are used as latitude and longitude in the global pole grid, respectively.

The wind data is given in various heights, which correspond in their name to the main height levels of the Pamore model, which is translated to height in meters in Table 3. These heights are understood as above ground and the levels work as layers, which means that formally the wind speed is given for the entire layer. These layers' boundaries lie half between the corresponding level heights.

Table 3: Height levels of wind data and their corresponding heights

Pamore height level / File ending	Height in m above ground[1]
44	345.53
45	258.21
46	183.93
47	122.32
48	73.03
49	35.72
50	10

[1] Ground is defined individually for each cell in the COSMO model. Please look at [7] for further details.

2.3. Validation

To estimate, whether the generated data, can be used as a representative weather data set of the target year, the Pamore model data is compared to the weather observations of DWD's own weather stations, accessed via the Climate Data Centre. As judging metrics, the root mean square error (RMSE), mean absolute error (MAE) and mean bias error (MBE) were calculated.

2.4. Energy system assessment

For the use in energy system simulation, knowing whether the year is considered bright, warm and windy is a considerable advantage. Therefore, the total yearly sum of solar irradiation, the average wind speed and the average *Gradtagszahl*, indicating the demand for residential heating are calculated.

The *Gradtagszahl* is defined as

$$\overline{GTZ}_{20/12} = \frac{\sum_{i=1}^{n_{\text{Latitudes}}} \sum_{j=1}^{n_{\text{Longitudes}}} \sum_{d=1}^{n_{days/a}} 20\text{ °C} - \bar{\vartheta}_{i,j}(d)}{n_{\text{Grid}}} \; ; \forall\, \bar{\vartheta}_{i,j}(d) < 12\text{ °C} \quad \textbf{(2.6)}$$

whereby $n_{\text{Latitudes}}$ is the number of cells along the latitudes (362), $n_{\text{Longitudes}}$ the number of cells along the longitudes (330) and n_{Grid} the total number of cells (119,460).

3. RESULTS

The results were generated using some small python scripts, which were published in a git[2], so – next to reproducibility – it might also indicate how to work with dataset.

3.1. Validation results

As a simplification for this documentation the scatter plot of only those weather stations are shown, that have either the best or worst RMSE, MAE or MBE. To sum up all stations, a box plot for these values is given.

3.1.1 Temperature validation results

The temperature differences for 2015 show acceptable differences in RMSE and MAE, expecting to be 1 K in in difference between COSMO's and the weather station's observational data (as reference value) for most cases. The MBE indicates that COSMO tends to overestimate the actual temperature.

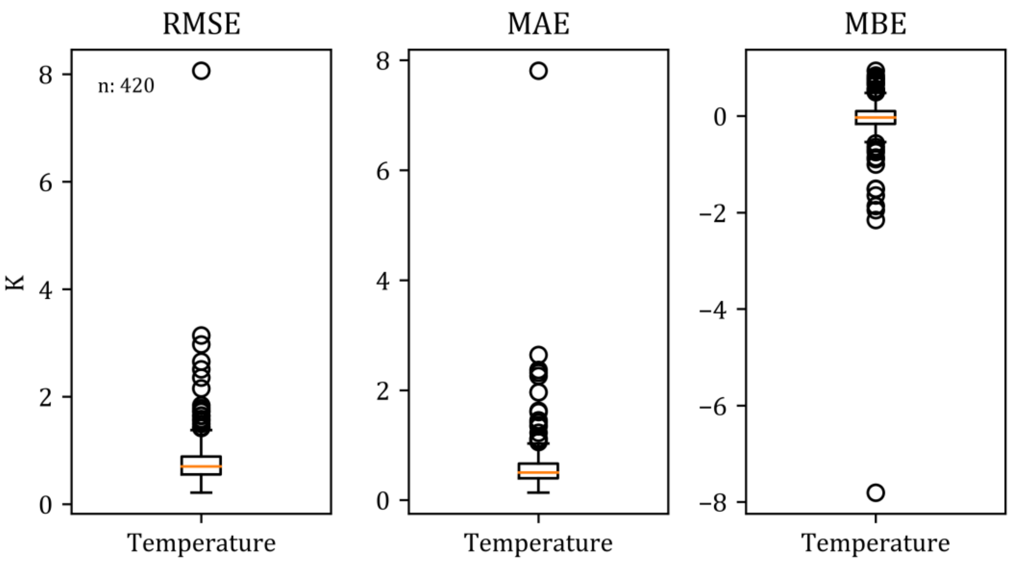

Figure 4: Overall statistical results for temperature. Temperature differences were accounted and are given in Kelvin. Size of dataset was n = 420.

[2] https://collaborating.tuhh.de/ietge/dataevaluationdwd

The worst fits (Figure 5) can be observed at the Zugspitze station located near the top of Germany's highest mountain. Despite the fine resolution of the COSMO grid, the Zugspitze values are deemed explicable by the weather station's protruding location.

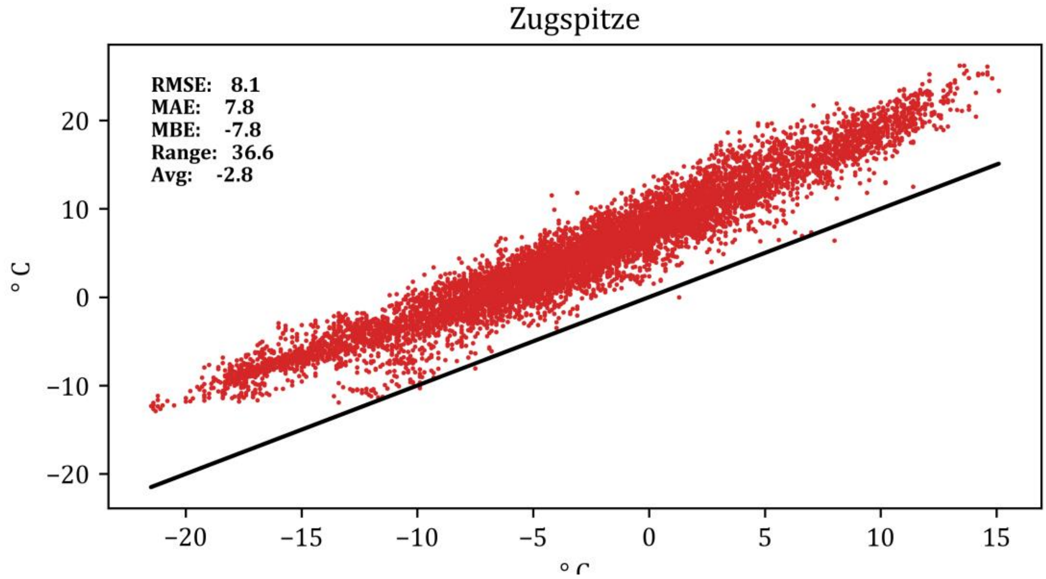

Figure 5: Worst fits of COSMO data and observation data. Scatter plots of observation data (x-axis) and model data (y-axis). Zugspitze has all, highest RMSE, MAE and |MBE|, among all weather stations.

The best fits (Figure 6) show Carlsfeld station as a very good fit of model and observation. Ueckermünde (best MBE) is a good representation of this data set, however, as RMSE and MAE lie near the median of all stations (see Figure 4).

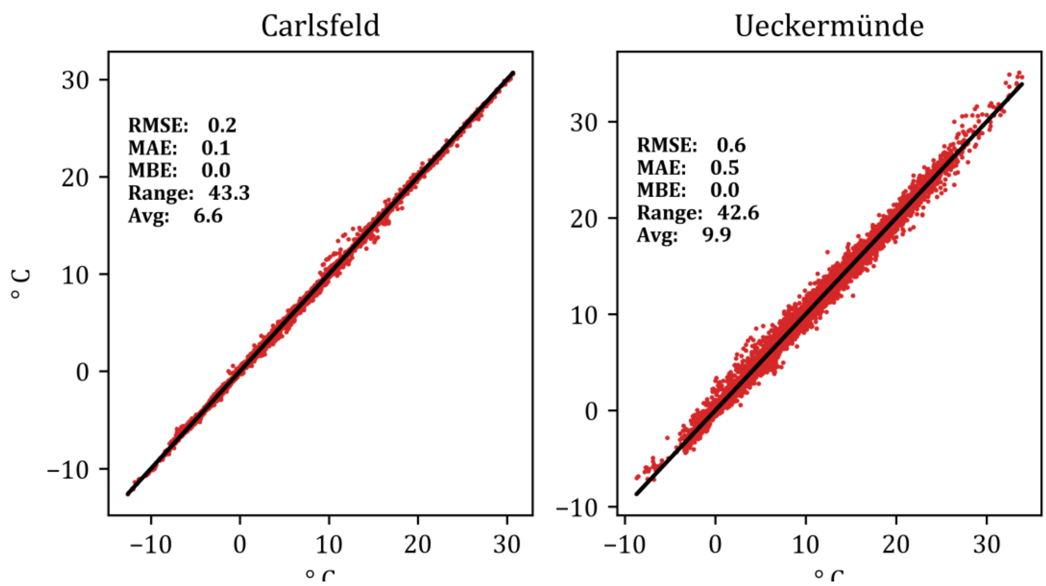

Figure 6: Best fits of COSMO data and observation data. The best in RMSE and MAE is Carlsfeld and in MBE Ueckermünde.

The statistics show such good overall agreement of model data and weather station observation data, that histograms were omitted for brevity.

3.1.2 Solar validation

The solar differences for 2015 show overall large differences in both diffuse horizontal irradiation (Figure 7) and global horizontal irradiation (Figure 8). The spreads are relatively large, compared to those of the temperature validation (Figure 4), which can be explained by the size of the observation stations for irradiation. The worse results for the global irradiation are likely due to the cell-wise resolution and respectively mean cloudiness and

mean direct irradiation for a cell, that might not correspond to the actual shadowing of a single point in that cell, i.e. the weather station.
The worst fit of observed to model solar data (see Figure 9), and the best fits (see Figure 10) look somewhat similar, which might indicate a systematic error. Hohenpreißenberg station is present in both worst and best fits, which rather goes to show the poor nature of the MBE as a quality measure on its own.

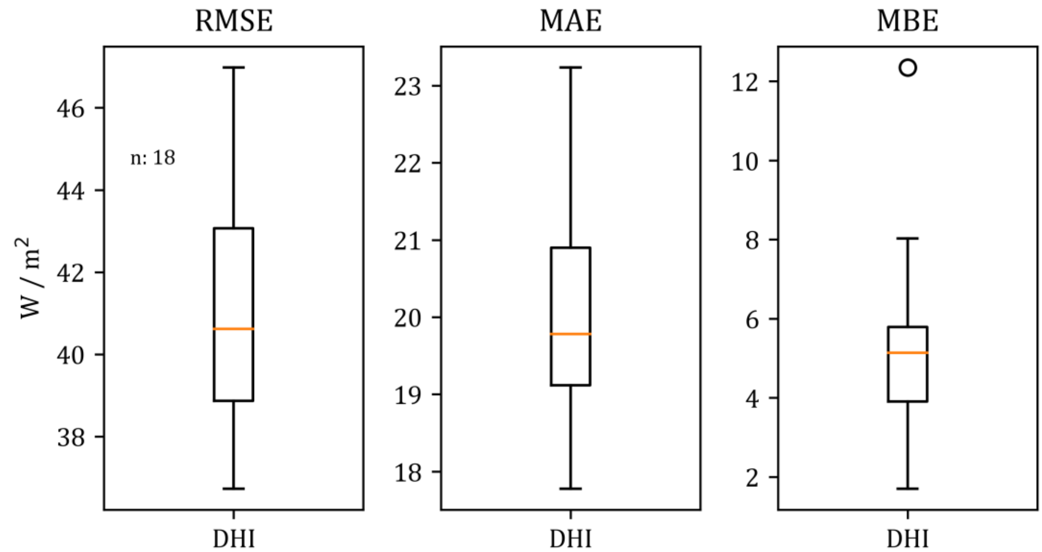

Figure 7: Overall statistical results for diffuse irradiation (ASWDIFD/DHI). Irradiation differences were accounted and are given in W/m². Size of dataset was n = 18.

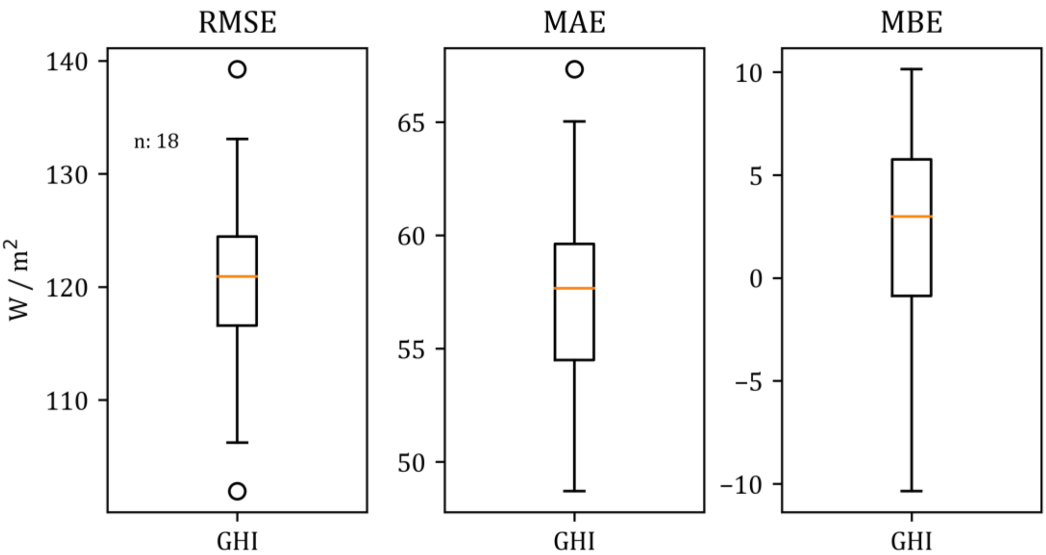

Figure 8: Overall statistical results for global irradiation (ASWDIFD+ASWDIR/GHI). Irradiation differences were accounted and are given in W/m². Size of dataset was n = 18.

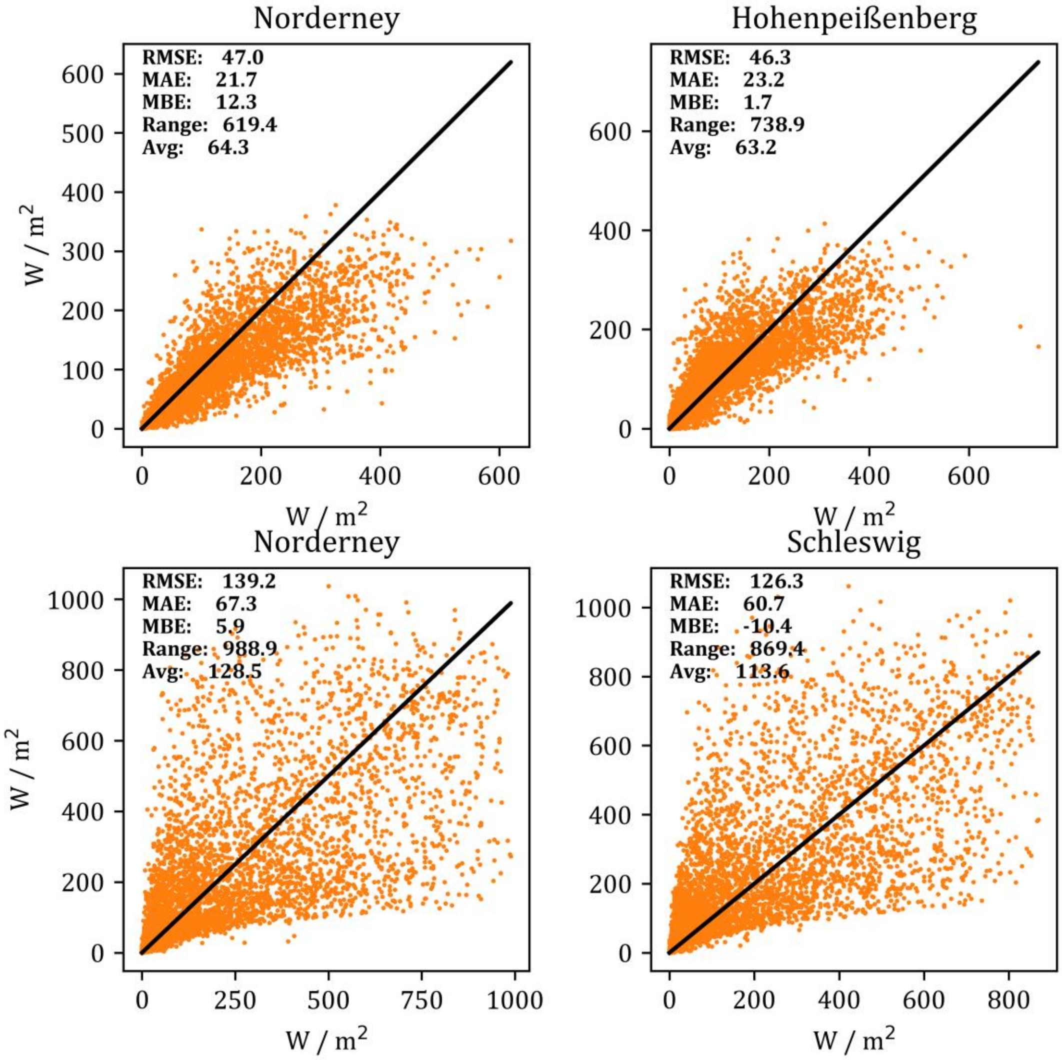

Figure 9: Worst fits of COSMO data and observation data. On top the diffuse irradiation and at the bottom the global irradiation scatter plots. Norderney has both highest RMSE and |MBE| in DHI, and both highest RMSE and MAE for GHI. Hohenpeißenberg has highest MAE in DHU and Schleswig has the highest |MBE| for GHI.

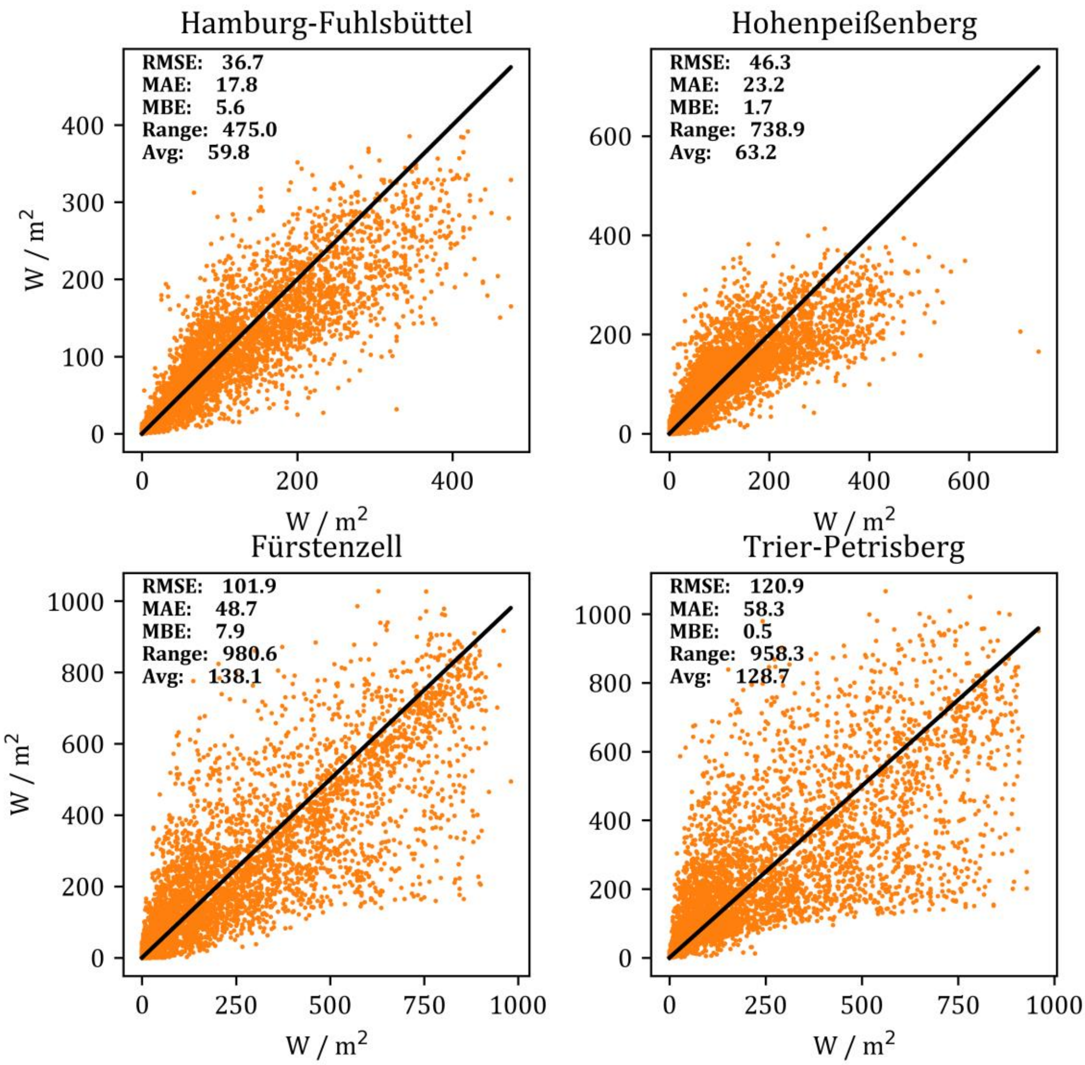

Figure 10: Best fits of COSMO data and observation data. On top the diffuse irradiation and at the bottom the global irradiation scatter plots. Hamburg-Fuhlsbüttel has both lowest RMSE and MAE in DHI, while Fürstenzell has both lowest RMSE and MAE for GHI. Hohenpeißenberg and Trier-Petrisberg both have the lowest MBE.

The histograms of the weather station with the worst RMSE (Figure 11) and the best RMSE (Figure 12) both show overall good agreement between observation and model data, which indicates a good fitness for energy system simulation.

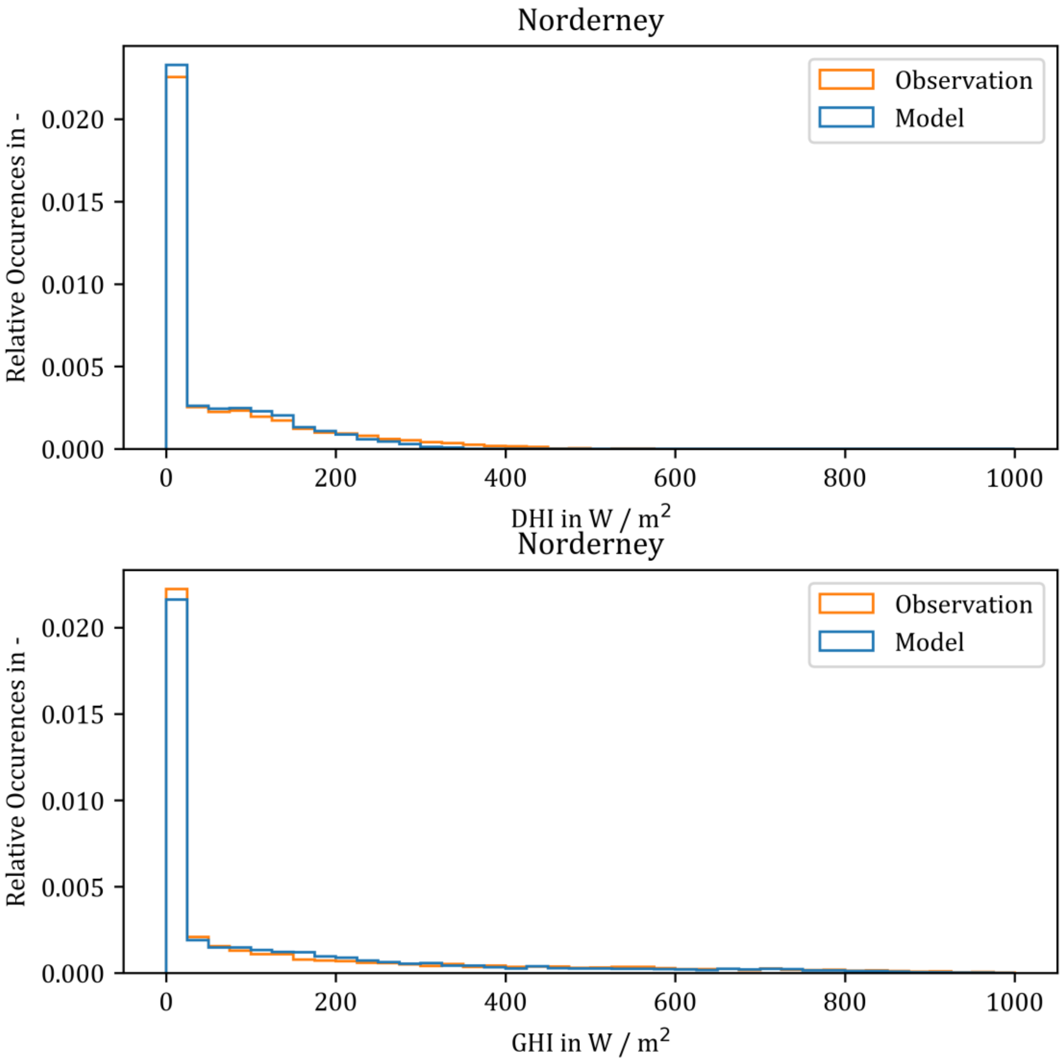

Figure 11: Histogram of worst RMSE fits between model data and weather station's observation data.

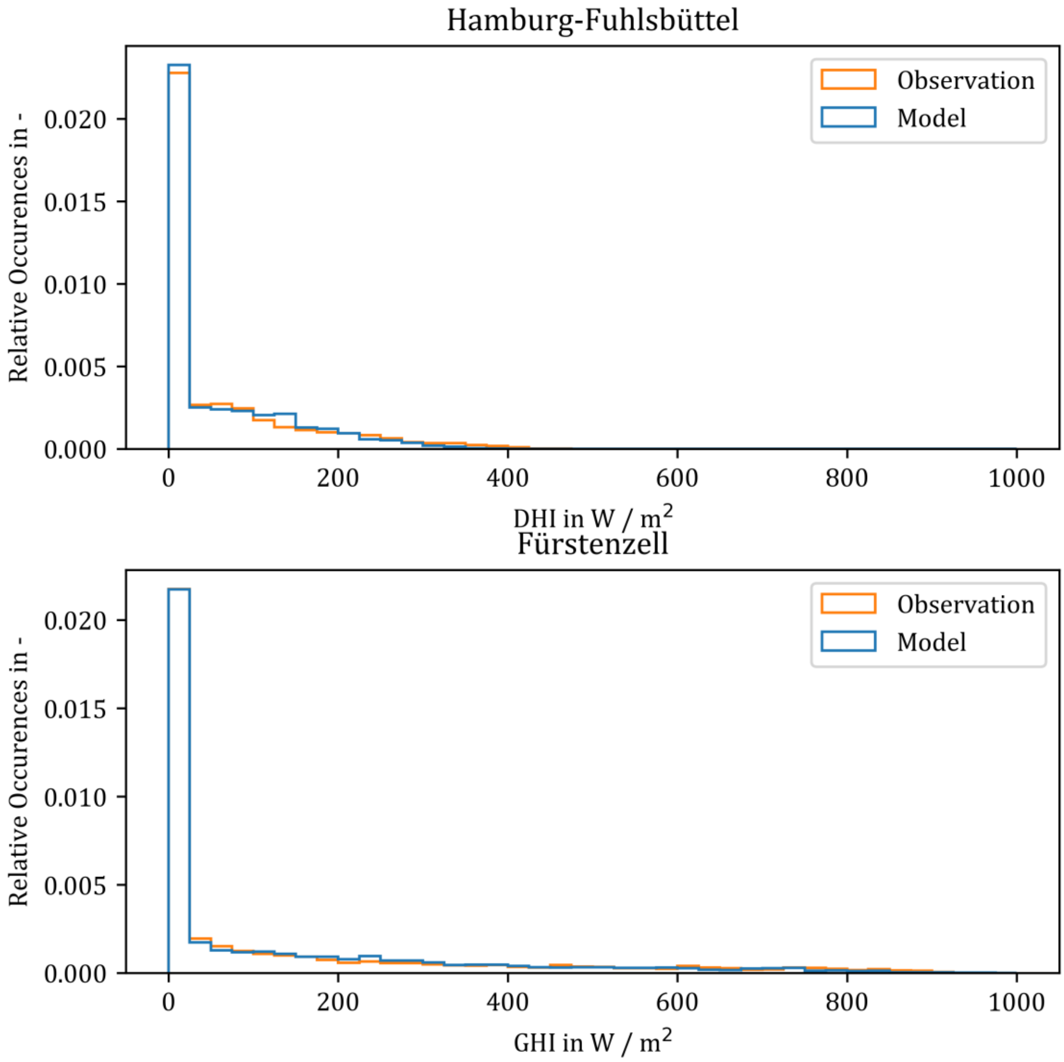

Figure 12: Histogram of best RMSE fits between model data and weather station's observation data for solar data.

3.1.3 Wind validation

The wind data show a low MBE in general for the wind speed (see Figure 13) and an acceptable MBE for the wind direction (see Figure 14), which is only acceptable as the CDC data is resolved in 10° steps. The RMSE and MAE results indicate a massive scattering between model and observation data, which is shown in both Figure 15 and Figure 16. These figures are split into worst and best fit according to the statistical indicators of that station's wind speed. The results for the Brocken weather station (Figure 15) are a good example of the problem resulting from the model's grid resolution, as the Brocken weather station is in a protruding position of its cell's orography.

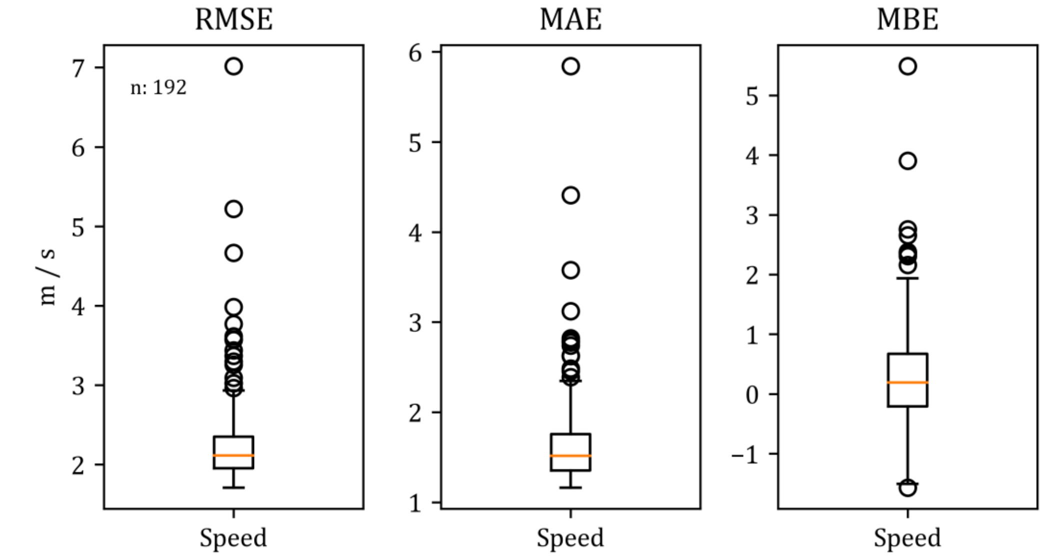

Figure 13: Overall statistical results for wind speed. Speed differences were accounted and are given in m/s. Size of dataset was n = 192.

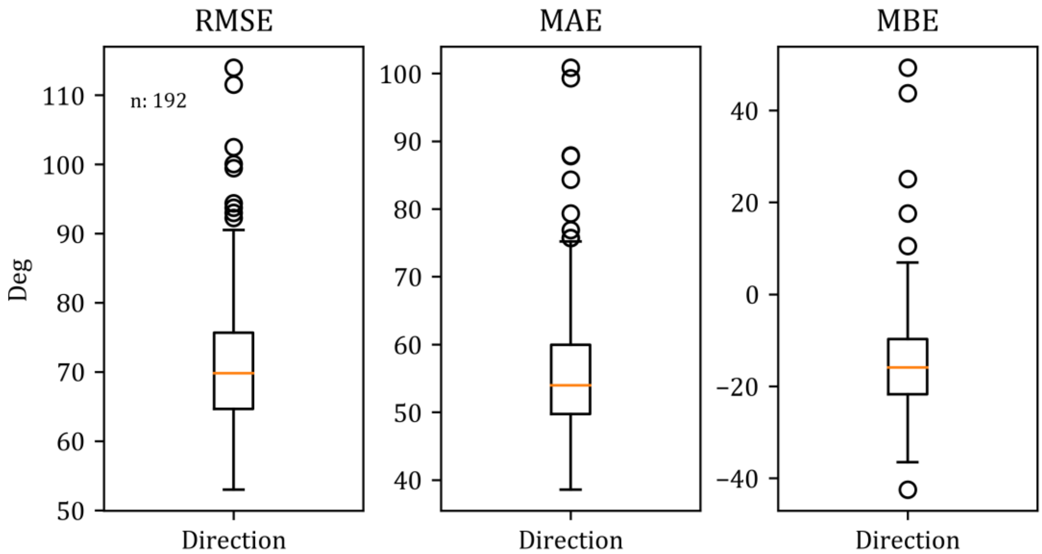

Figure 14: Overall statistical results for wind direction. Direction differences were accounted and are given in ° (degrees). The data resolution of the weather stations was 10°. Size of dataset was n = 192.

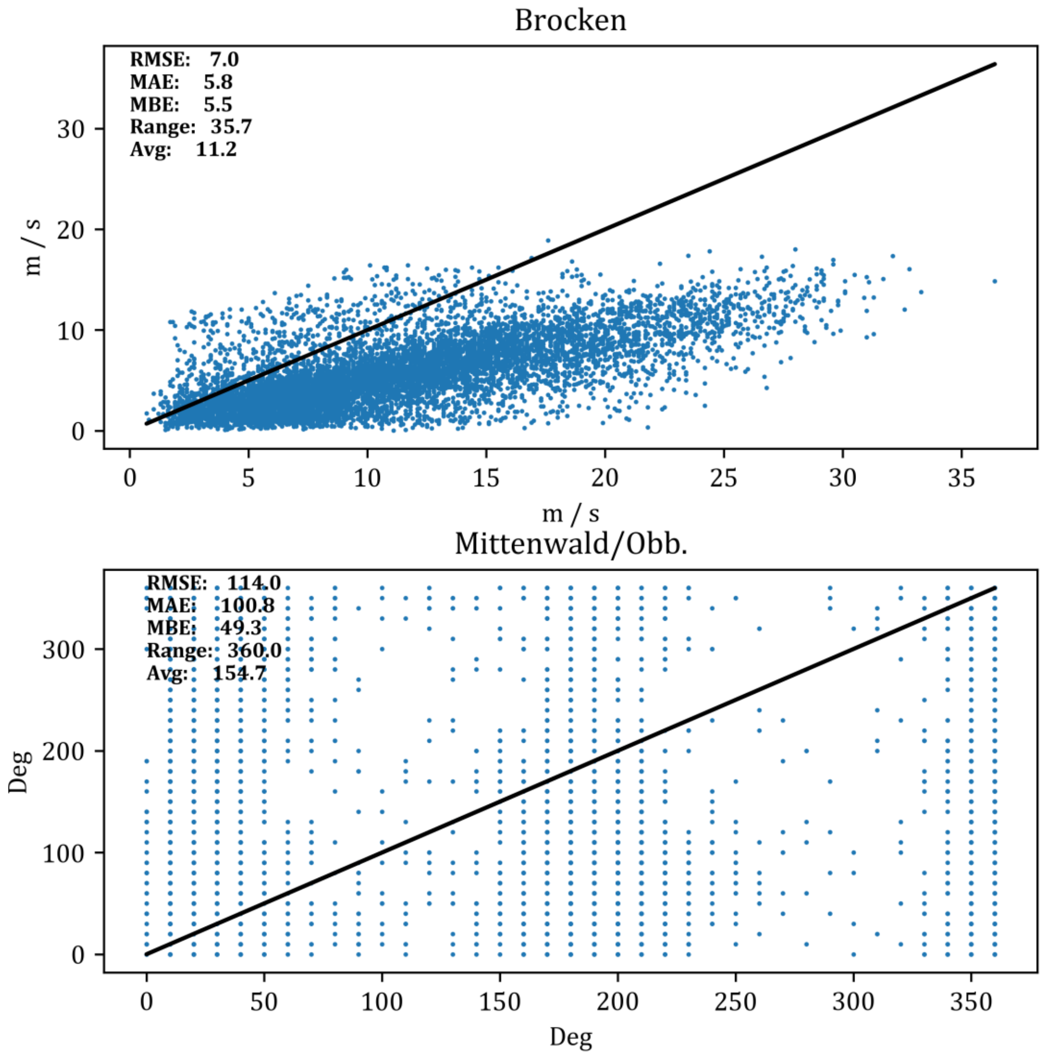

Figure 15: Worst fits of COSMO data and observation data for wind speed, which coincidentally are one of the worst direction results. On top wind speed and at the bottom the wind direction scatter plots. Brocken has all, highest RMSE, MAE and |MBE|, among all weather stations for wind speed, while Mittenwald/Obb. Shows highest values in all categories for wind direction.

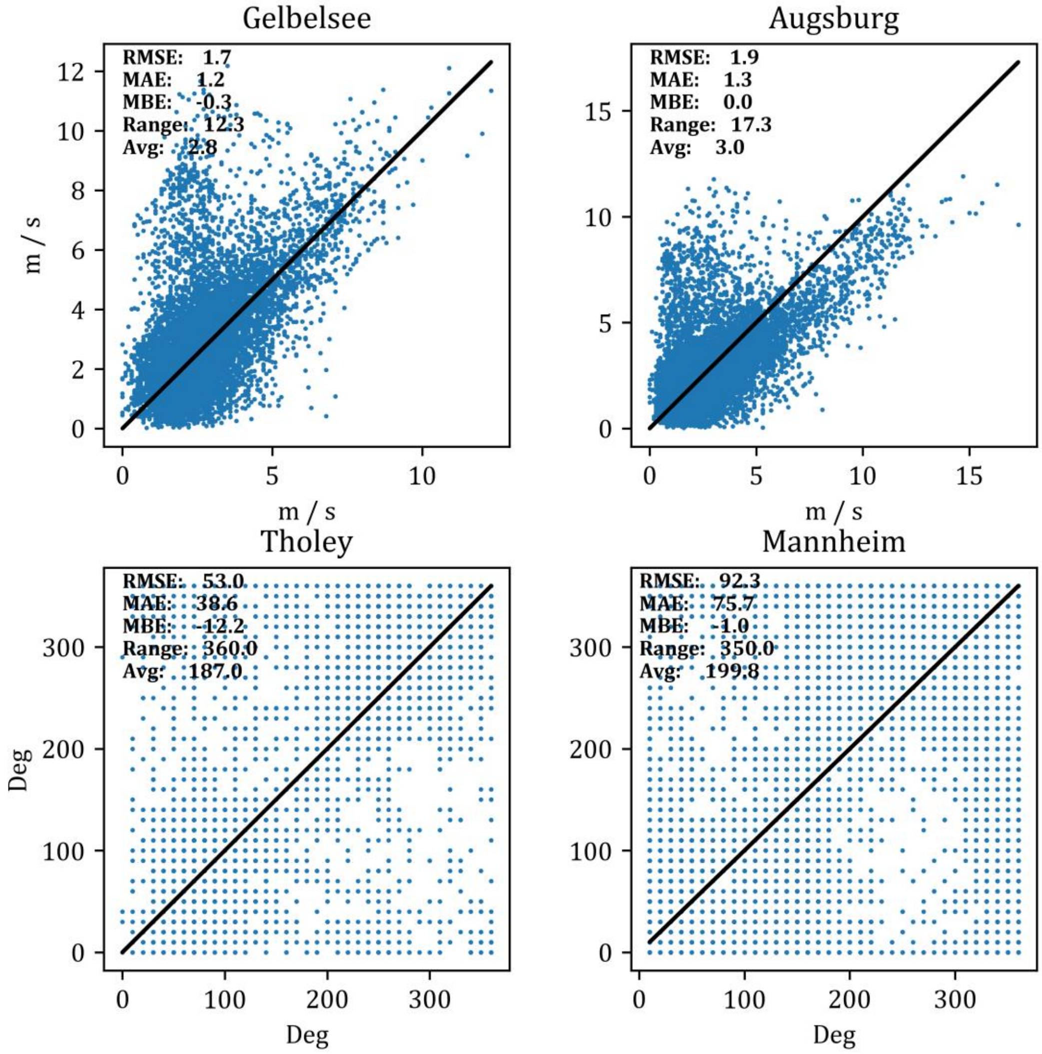

Figure 16: Best fits of COSMO data and observation data for wind speed, which coincidentally are one of the worst direction results. On top wind speed and at the bottom the wind direction scatter plots. Gelbelsee shows lowest RMSE and lowest MAE for wind speed, while Tholey shows the same for wind direction. Augsburg and Mannheim show lowest |MBE| for wind speed and wind direction, respectively.

Expectedly the wind speed histogram of the Brocken weather station (Figure 17) confirms the bad fit of wind data for points with complex orography, while the wind speed histogram for the weather station with the best RMSE (Figure 18) shows a better fit, than the scatter plot (Figure 16) would suggest. Judging by the overall statistics (Figure 13), the wind data is suitable for energy system simulation.

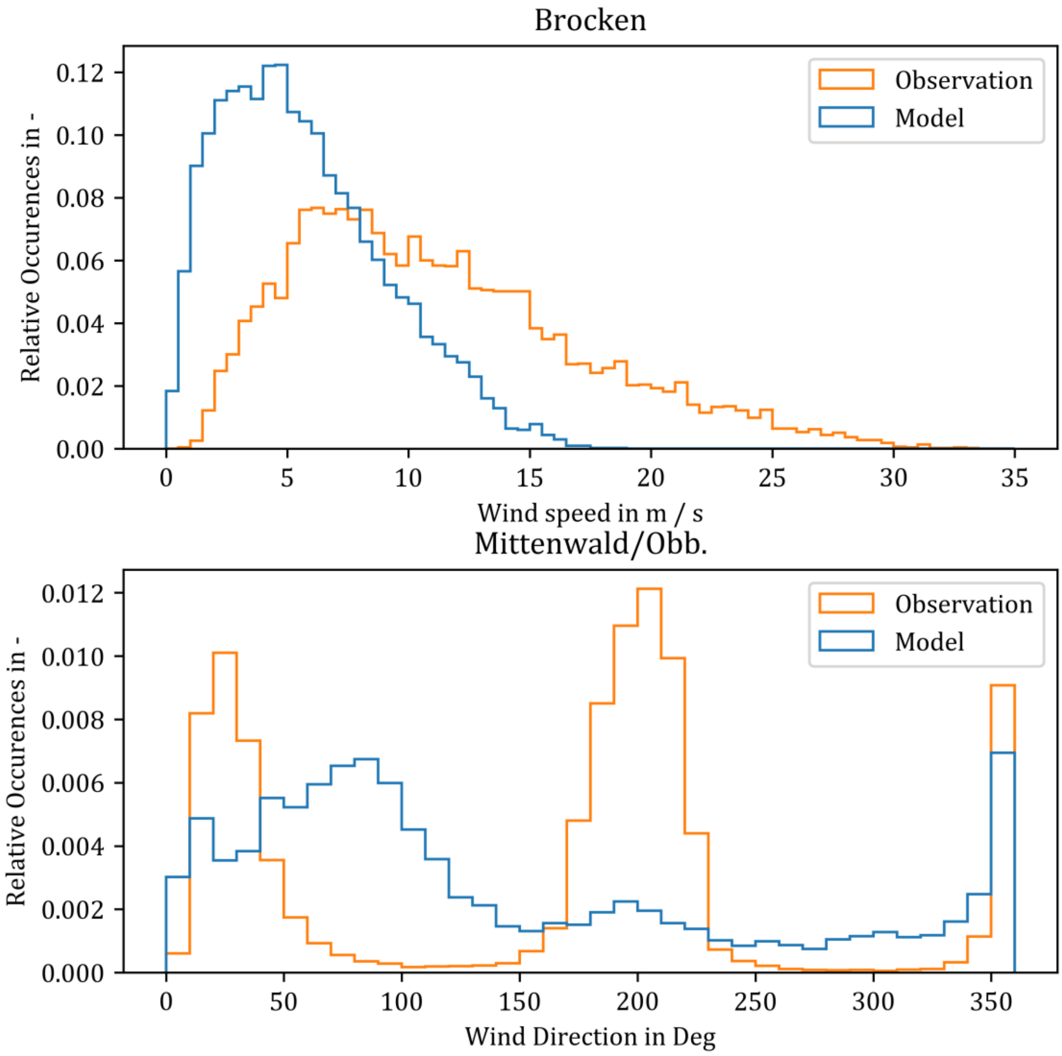

Figure 17: Histogram of worst RMSE fits between model data and weather station's observation data for wind data.

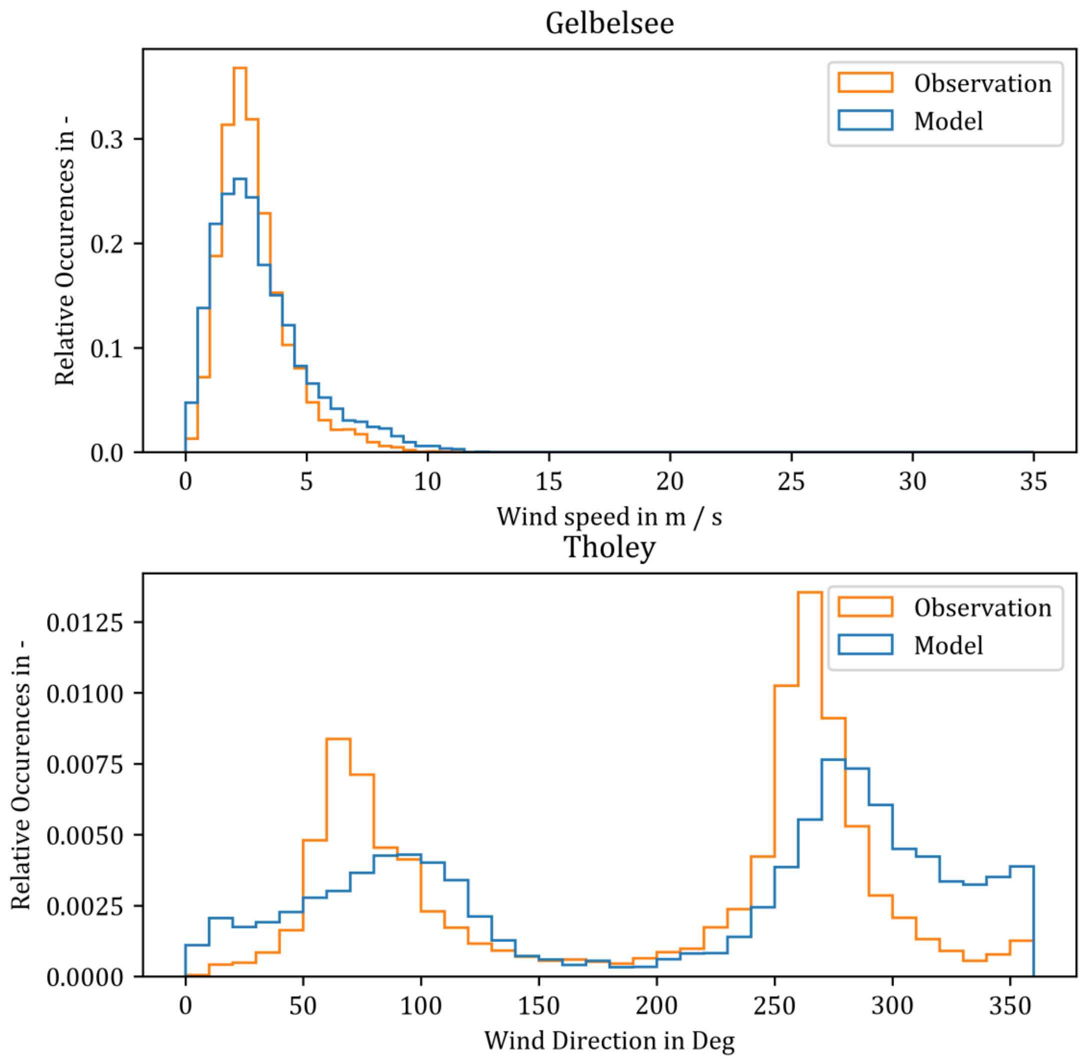

Figure 18: Histogram of best RMSE fits between model data and weather station's observation data for wind data.

3.2. Assessment results

Here the indication results of what using 2015 weather model data might cause in energy system simulation are given. What can be inferred from Table 4 is that 2015 is a rather average year with high average wind speeds and may be used as a somewhat representative year for energy system simulation.

Table 4: Results across all datasets for energy system simulation assessment. The documented year is highlighted in magenta.

Type	Temperature	Solar	Wind
Unit	GTZ in K	Mean annually sum in kWh/m²a	Mean wind speed at 122.32 m in m/s
2007	2988		7.74
2008	3198		7.63
2009	3156		7.37
2010	3790		7.10
2011	2924	1101	7.52
2012	3322	1066	7.43
2013	3496	1052	7.35
2014	2750	1039	7.30
2015	3054	1049	7.70
2016	3247	1004	7.11
2017	3132	1024	7.39
Mean	3187	1048	7.42

In Figure 19 to Figure 21 the distribution of energy system indicators are given. The temperature and wind speed distribution thereby match the orography of the map somewhat closely, which seems sensible to some extent.

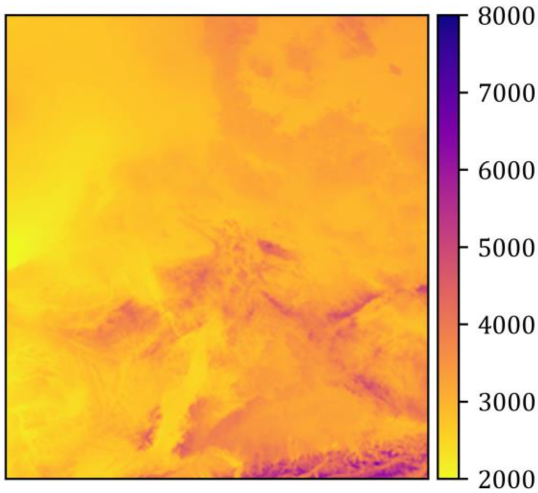

Figure 19: GTZ distribution 2015 in the data set in K. Every pixel represents a ~2.8x2.8 km grid cell.

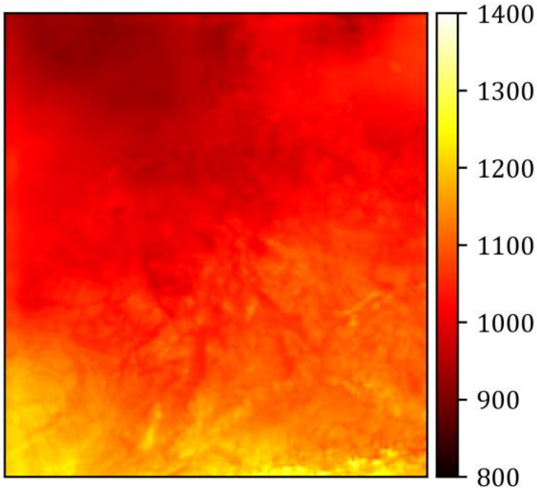

Figure 20: Annual GHI sum distribution 2015 in the data set in kWh/m²a. Every pixel represents a ~2.8x2.8 km grid cell.

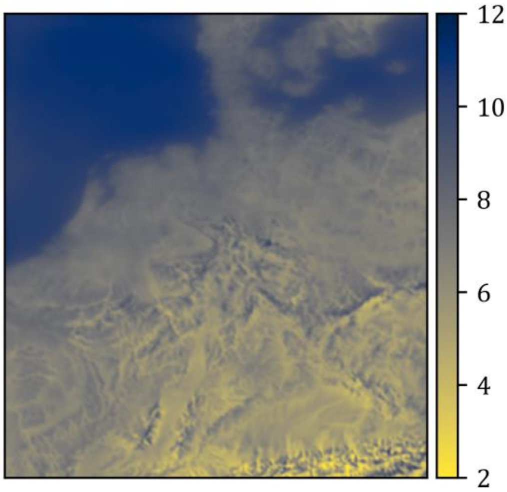

Figure 21: Mean wind speed at 122.32 m distribution 2015 in the data set in m/s. Every pixel represents a ~2.8x2.8 km grid cell.

4. DISCUSSION

Before discussing the observation and model data comparison, the general methodology is addressed. Assuming, that a ~2.8x2.8 km grid cell has the same results, as a single point in that cell (namely the weather station) is somewhat dependant of the orography of that cell, which the Zugspitze and the Brocken stations highlight prominently. Therefore, the user of the dataset has to keep in mind, that the data rather tells the weather of a region than a place.

Given that disclaimer, the solar and wind data still seem somewhat problematic, especially the wind direction and global irradiation data. It is left as an open question, whether that is a problem of the model, the measuring instruments, or the averaging process for generating an hourly profile.

The yearly grid-wide data (Table 4), their distribution (Figure 19 - Figure 21), and the MBE results on the other hand, indicate that the model data is – taken as a whole – rather accurate. Another case in point is an appraisal of the model by DWD staff, stating that the model can be taken as a good approximation of reality. The appraisals were gathered via e-mail.

5. CONCLUSION

Overall, the given data set can be interpreted as the weather that was and is applicable for energy system simulation.

6. REFERENCES

[1] NITSCH, Joachim ; PREGGER, Thomas ; NAEGLER, Tobias ; HEIDE, Dominik ; TENA, Diego Luca de ; TRIEB, Franz ; SCHOLZ, Yvonne ; NIENHAUS, Kristina ; GERHARDT, Norman ; STERNER, Michael ; TROST, Tobias ; OEHSEN, Amany von ; SCHWINN, Rainer ; PAPE, Carsten ; HAHN, Henning ; WICKERT, Manuel ; WENZEL, Bernd: *Langfristszenarien und Strategien für den Ausbau der erneuerbaren Energien in Deutschland bei Berücksichtigung der Entwicklung in Europa und global*. 2012

[2] PALZER, Andreas: *Sektorübergreifende Modellierung und Optimierung eines zukünftigen deutschen Energiesystems unter Berücksichtigung von Energieeffizienzmaßnahmen im Gebäudesektor*. Fraunhofer-Institut für Solare Energiesysteme. Dissertation

[3] KNORR, Kaspar ; ZIMMERMANN, Britta ; KIRCHNER, Dirk ; SPECKMANN, Markus ; SPIECKERMANN, Raphael ; WIDDEL, Martin ; WUNDERLICH, Manuela ; MACKENSEN, Reinhard ; ROHRIG, Kurt ; STEINKE, Florian ; WOLFRUM, Philipp ; LEVERINGHAUS, Thomas ; LAGER, Thomas ; HOFMANN, Lutz ; FILZEK, Dirk ; GÖBEL, Tina ; KUSSEROW, Bettina ; NICKLAUS, Lars ; RITTER, Peter: *Kombikraftwerk 2*. 2014

[4] BUNDESMINISTERIUM FÜR UMWELT, NATURSCHUTZ UND NUKLEARE SICHERHEIT: *Verordnung zur Festlegung der Nutzungsbestimmungen für die Bereitstellung von Geodaten des Bundes* (in Kraft getr. am 19. 3. 2019). In: *Bundesgesetzblatt* 2013 (2019-03-19), Nr. 1, S. 547. URL http://www.gesetze-im-internet.de/geonutzv/index.html

[5] DEUTSCHER WETTERDIENST: *Pamore*. URL https://www.dwd.de/DE/leistungen/pamore/pamore.html – Überprüfungsdatum 2020-01-18

[6] DEUTSCHER WETTERDIENST: *Climate Data Center*. URL ftp://ftp-cdc.dwd.de/ – Überprüfungsdatum 2020-01-18

[7] BALDAUF, M. ; FÖRSTNER, J. ; KLINK, S. ; REINHARDT, T. ; SCHRAFF, C. ; SEIFERT, A. ; STEPHAN, K.: *Kurze Beschreibung des Lokal-Modells Kürzestfrist COSMO-DE (LMK) und seiner Datenbanken auf dem Datenserver des DWD* : *Version 2.3*. Offenbach, 2014

[8] HANISCH, Thomas: *The NWP system at DWD* (12th Workshop on Meteorological Operational Systems). Reading, UK, 2009

[9] HANISCH, Thomas: *The new NWP forecast system of the DWD based on ICON / ICON-EU and COSMO-DE* (15th Workshop on Meteorological Operational Systems). Reading, UK, 2015

[10] HANISCH, Thomas: *Recent developments of the NWP forecast system at DWD based on ICON / ICON-EU and COSMO-DE* (16th Workshop on Meteorological Operational Systems). Reading, UK, 2017

7. Nomenclature

Table 5: Nomenclature

Symbol	Unit	Description
$ASWDIFD$	W/m²	File indicator for diffuse solar irradiation on the horizontal plane
$ASWDIR$	W/m²	File indicator for direct solar irradiation on the horizontal plane
d	-	Day in a year (time step)
DHI	W/m²	Same as AWDIFD
GHI	W/m²	Global solar irradiation on the horizontal plane
GTZ	K	Gradtagszahl (indicating heating demand)
I	W/m²	Solar irradiation (any)
i	-	Index variable for a grid cell
j	-	Index variable for a grid cell
n	-	Number of elements (count)
t	-	Time step
T	K	Temperature
TMP	°C	File indicator for temperature
u	m/s	Zonal wind speed

v	m/s	Meridional wind speed
WMV	m/s	File indicator for meridional wind speed
WZU	m/s	File indicator for zonal wind speed
δ	°	Angle between rotated and global polar grid
ϑ	°C	Temperature
λ	°	Longitude
φ	°	Latitude

Table 6: Terms and abbreviations

Term/Abbreviation	Description
CharL	Energy system model of the research project of this data acquisition
COSMO	Weather model, originally developed by the DWD
DWD	Deutscher Wetterdienst (German Meteorological Service)
HDF5	Binary file format for large(, scientific) data sets
Pamore	Model data access program of the DWD
VEREKON	Research project
VRE	Volatile Renewable Energy

ACKNOWLEDGEMENT

The research data acquisition was funded by the BMWi (FKZ: 03ET7067A).

Supported by:

Federal Ministry
for Economic Affairs
and Energy

on the basis of a decision
by the German Bundestag

APPENDIX – DATASTRUCTURE

Every file shares the same structure and therefore only differs in data type and unit of data (see Table 2 (repeated at bottom of this page) for those). Here some of the meta data is listed.

Grid

The grids were cut to 362 x 330 points polar grids with 8760 time steps (one for each hour of the year starting at 01.01.2017 0:00 UTC). They originate from rotated and pole-shifted polar grids of the COSMO model. The COSMO DE and EU grids at the DWD are rotated pole grids with their rotated north at 40 °N and 170 °W. Starting at model latitude 5 $°S_{rotated}$ and longitude 5 $°W_{rotated}$ centre grid points are generated going northeast to latitude 6.5 $°N_{rotated}$ and longitude 5.5 $°E_{rotated}$ in 0.025 ° resolution (translating to a resolution of roughly 2,8 km). The cut grid now starts in the top left at 56.0032 °N and 1.48992 °E and goes to the bottom right at 47.1331 °N and 15.069 °E.

Data structure

In the files the latitudes and longitudes are organised as 362 x 330 2D arrays containing the corresponding cell's latitude or longitudes, respectively, as a float (also known as single precision or 4 Byte floating point) and accessible via `/latitude` or `/longitude` respectively. The actual data is organised as a 362 x 330 x 8760 3D array containing the variables value as a float and accessible via its accessor name, which in the repository correspond to the file indicators (e.g. `/ASWDIR`). Should someone rename the file, the accessor name is repeated in the metadata. The resulting file size hence is around 4 GB.

Available Metadata

Every file contains the following metadata:

Accessor (in hdf5 file)	Content
/creation_date	Date of file creation
/author	Who created that file
/datasource	Where the data is from (DWD)
/datatype	Accessor/Variable name of data itself
/datatype_description	Description of the available data
/unit	Unit of the data
/timeframe	File's date range
/steptime	Time between steps in h
/license	License of the file
/comment	Any additional info (e.g. bodged data points)
/level	Height level (only filled with information in wind data files)

Copy of Table 2:

Variable in repo	Meaning	Unit
ASWDIFD	Average diffuse solar irradiation on the horizontal plane	W/m^2
ASWDIR	Average direct solar irradiation on the **horizontal** plane	W/m^2
TMP	Temperature at 2m above ground	°C
WZU	Zonal wind speed (along latitudes), positive towards east	m/s
WMV	Meridional wind speed (along longitudes), positive towards north	m/s

APPENDIX – FILE CREATION AND DATA RETRIEVAL CODE SNIPPETS

The data files were created using Matlab® 2018b and hdf5® as the target file format. Which results in an older version of that file format. See Listing 1 for the creation code. The files are therefore readable with Matlab, the `h5py` library for python, the C and Fortran libraries of the hdf5 group[3] or the C++ library in our own energy system model `CharL`[4].

Listing 1: Example Matlab code used to create the files (in this case, a WZU file is created)

```
h5create(pathoutU, strcat('/',typeU), [nX nY nTime],...
                          'Datatype','single');
h5create(pathoutU, '/latitude', [nX nY], 'Datatype','single');
h5create(pathoutU, '/longitude', [nX nY], 'Datatype','single');
h5write(pathoutU, '/latitude', latitudes_cut);
h5write(pathoutU, '/longitude', longitudes_cut);
h5write(pathoutU, strcat('/',typeU), dataCutU);
h5writeatt(pathoutU, '/','creation_date', creation_date);
h5writeatt(pathoutU, '/','author', author);
h5writeatt(pathoutU, '/','datasource', data_source);
h5writeatt(pathoutU, '/','datatype', typeU);
h5writeatt(pathoutU, '/','datatype_description', descriptionU);
h5writeatt(pathoutU, '/','unit', unit);
h5writeatt(pathoutU, '/','timeframe', timeframe);
h5writeatt(pathoutU, '/','steptime', steptime);
h5writeatt(pathoutU, '/','license', 'GeoNutzV');
h5writeatt(pathoutU, '/', 'comment', comment);
h5writeatt(pathoutU, '/', 'level', levelTextF);
```

The commands used for retrieving the COSMO data via Pamore were used for each variable individually and month-wise, due to line-limits per request set by the Pamore servers. Examples for these commands are given in the following listings. Please note, that the `-lv` command in the wind data requests did not have the `44` for every year in the overall dataset. For e.g. 2017 it does not.

Listing 2: Command to retrieve temperature data. The dates (highlighted) were set for each month.

```
pamore -G -F -d 2017010100 -de 2017020100 -dinc 1 -tflag best -ee T_2M%105
-model lmk_ana
```

Listing 3: Command to retrieve direct irradiation on the horizontal data up until 21st March 2017. The dates (highlighted) were set for each month.

```
pamore -G -F -d 2017010100 -de 2017020100 -dinc 1 -tflag best -ee
ASWDIR_S%1 -model lmk_ana
```

Listing 4: Command to retrieve diffuse irradiation on the horizontal data up until 21st March 2017. The dates (highlighted) were set for each month.

```
pamore -G -F -d 2017010100 -de 2017020100 -dinc 1 -tflag best -ee
ASWDIFD_S%1 -model lmk_ana
```

[3] https://portal.hdfgroup.org/display/HDF5/HDF5
[4] https://collaborating.tuhh.de/iet/CharL

Listing 5: Command to retrieve zonal wind data. The dates (highlighted) were set for each month.

```
pamore -G -F -d 2017010100 -de 2017020100 -dinc 1 -tflag best -ee V%hi -lv
44,45,46,47,48,49,50 -model lmk_ana
```

Listing 6: Command to retrieve meridional wind data. The dates (highlighted) were set for each month.

```
pamore -G -F -d 2017010100 -de 2017020100 -dinc 1 -tflag best -ee V%hi -lv
44,45,46,47,48,49,50 -model lmk_ana
```

The forecast data on the other hand, were acquired for the forecast run from 6 o'clock at each day and variable individually, as the following listings show.

DISCLAIMER

This document was written rather hastily and by resuing a lot of paragraphs as well as using automated plot generation. Please note the author, if you are unable to reproduce the results or discover any inaccuracies. However, the quality of the data is very well tested, so please do not let this document discourage you from using them.

HOW TO CITE

If you are unsure about the way to cite the dataset, here is a suggestion:
In text: "DWD weather data supplied by <citation>"
In references: "2020 – Gerrit Erichsen - DWD Weather Model Data for Energy System Simulation: 2015 – doi"
Citing the documentation should also work by citing the repository, as only the repository in its entirety has a DOI.
In text: "[…] considering the data documentation of <citation> […]"
In references: Same as above.