

Research Article

Statistical Modeling of Near-surface Wind Speed: A Case Study from Baden-Wuerttemberg (Southwest Germany)

Jung C and Schindler D*

Chair of Meteorology and Climatology, Albert-Ludwigs-University of Freiburg, Germany

***Corresponding author:** Schindler D, Chair of Meteorology and Climatology, Albert-Ludwigs-University of Freiburg, Werthmannstrasse 10, D-79085 Freiburg, Germany**Received:** December 15, 2014; **Accepted:** January 29, 2015; **Published:** January 31, 2015**Abstract**

In this paper a methodology is presented that can be used to statistically model characteristics of near-surface wind speed in complex terrain at high spatial resolution. It was developed based on daily mean wind speed time series provided by the German Weather Service for 65 stations located in southwest Germany. After comprehensive preparation of the wind speed data that were measured in the period 1975 to 2010 including gap filling, homogenization, detrending and measurement height correction, 48 continuous distributions were fitted to the empirical distributions associated with the wind speed time series. The results of the evaluation of the goodness-of-fit demonstrate that the five-parameter Wakeby-distribution characterizes the statistical properties of measured wind speed better than all other tested distributions. Based on surface roughness, terrain-related parameters (curvature, topographic exposure) and ERA-Interim reanalysis wind speed data available for the 850 hPa pressure level, LSBoost-models were built to estimate station-specific Wakeby-parameters. The LSBoost-models were then used to model the Wakeby-parameters on a 50 m resolution grid in the entire study area as a function of the predictor variables. The area-wide availability of the Wakeby-parameters allows producing detailed wind speed quantile maps.

Keywords: Wind speed measurement; Ensemble methods; Wakeby-distribution; Hellmann power law

Introduction

A measure often used to quantify characteristics of near-surface wind fields is the absolute value of the horizontal wind vector commonly known as wind speed. While variations of wind speed at the earth surface are ubiquitous, there is often limited knowledge about the wind speed characteristics at a particular site. An inherent characteristic of wind speed is its high spatiotemporal variability [1,2]. Especially rough surfaces [3] and complex terrain [4] modify near-surface wind fields. Unfortunately, there are often too few wind speed measuring stations in place although wind speed measurements made near the ground are strongly influenced by land cover and terrain at and around the measuring sites [5-7]. The lack of near-surface wind speed measurements often limits knowledge about statistical wind speed properties and thus about the local wind resource.

On the other hand, there is great interest in detailed information on the characteristics of the near-surface wind speed field: forest administrations seek to locate areas prone to wind damage [8,9], insurance companies want to assess wind-induced loss potentials [10,11], local authorities work for the improvement of air quality by maintaining and/or enhancing natural ventilation in urban areas [12,13] and there is great potential to maximize the exploitation of wind power by providing detailed information on the local wind resource [7,14-16]. It was pointed out by [17] that for an accurate assessment of wind characteristics the choice of an appropriate statistical distribution function is crucial because the distribution of wind speed values determines the performance of wind turbines. An accurate assessment of the local wind resource on multi-year scales is

the basis for the selection of most appropriate sites for wind turbines which offer great potential to reduce CO₂-emissions [18].

There is ongoing debate whether there is a distribution function that describes empirical distributions of wind speed data best [19,20]. In many previous studies [1,2,4,6,19-29], the Weibull function was used to represent empirical wind speed distributions. However, results from other studies [30-33] demonstrate that wind speed distributions cannot always adequately be represented by the Weibull distribution, especially when wind speed strongly varies with wind direction [34].

Since the orography in the southwest of Germany is complex and the landscape is compartmentalized, the near-surface wind speed characteristics vary on small scales, i.e. from measuring station to measuring station. Thus, the goals of this study are (i) to evaluate which distribution function is most appropriate for describing the statistical properties of wind speed data measured near the ground in the southwestern German federal state Baden-Wuerttemberg, (ii) to develop a statistical model based on the best-fitting distribution that is able to provide reasonable estimates of near-surface wind speed quantiles in high spatial resolution (50 m resolution grid) in the study area.

Material and Methods

Study area

The study area is the German federal state of Baden-Wuerttemberg (southwest Germany). It has a surface area of 35752 km². Its orography is complex and includes the low mountain ranges Black Forest

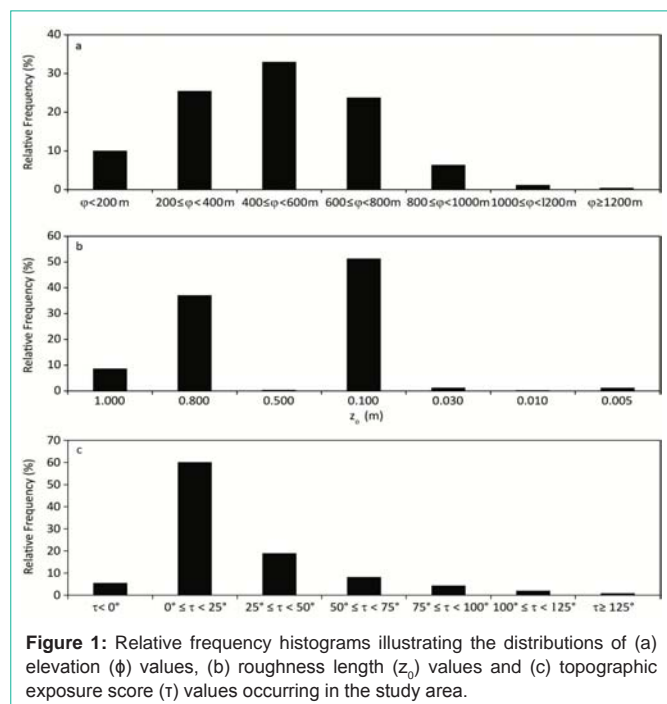


Figure 1: Relative frequency histograms illustrating the distributions of (a) elevation (ϕ) values, (b) roughness length (z_0) values and (c) topographic exposure score (τ) values occurring in the study area.

(length ~150 km, width ~30-50 km, highest elevations > 1400 m) and Swabian Alb (length ~180 km, width ~35 km, highest elevations > 1000 m) as well as the broad, flat Rhine Valley, which borders on France in the west. The elevation (ϕ) within the study area ranges from 85 m above sea level (a.s.l.) in the Rhine Valley to 1493 m a.s.l. at the top of the Feldberg, which is the highest mountain in Baden-Wuerttemberg. The ϕ -values in the study area mainly vary between 200 m and 800 m a.s.l. (Figure 1a). The land cover types within the study area might change on small scales (< 1 km) and often create a compartmentalized landscape. The land cover types available from the Corine Land Cover (CLC) 2006 dataset for Germany [35] and the roughness length (z_0) values associated with these land cover types are listed in Table 1. Rough surface types like urban areas and forests have higher z_0 -values (1.00 m and 0.80 m) than smoother surface types like agricultural areas (0.10 m) or natural grass land (0.03 m). The z_0 -values, which were interpolated from the original spatial resolution of the CLC-dataset of 100 m to a 50 m resolution grid, correspond to standard z_0 -values found in literature [36-38]. According to the CLC-data, the study area's surface is mainly covered by agricultural areas (51%), forests (38%) and artificial surfaces like urban areas, airports, road networks and rail networks (9%). The corresponding bimodal z_0 -value distribution is displayed in Figure 1b.

Table 1: Land cover types available from the Corine Land Cover 2006 dataset for Germany [35] and corresponding roughness length (z_0) values.

ID	Land cover type	z_0 (m)
1	Urban area	1.000
2	Forested area	0.800
3	Green urban area	0.500
4	Agricultural area	0.100
5	Natural grass land	0.030
6	Airports	0.010
7	Open spaces with little vegetation	0.010
8	Water bodies	0.005

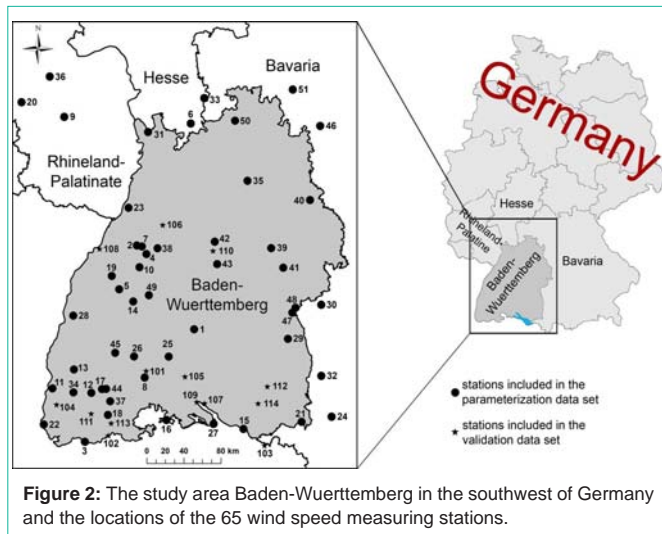
Table 2: List of indexed DWD-stations: stations with ID-values 1-51 are included in the parameterization data set (DS1); stations with ID-values 101-114 are included in the validation data set (DS2).

ID	Station	ID	Station
1	Albstadt-Onstmettigen	34	Münstertal
2	Bad Herrenalb	35	Öhringen
3	Bad Säckingen	36	Pferdsfeld
4	Bad Wildbad-Sommerberg	37	Schluchsee
5	Baiersbronn-Obertal	38	Schömburg
6	Beerfelden	39	Schwäbisch Gmünd
7	Dobel	40	Stimpfach-Weiptershofen
8	Donaueschingen	41	Stötten
9	Dörrmoschel-Felsberghof	42	Stuttgart Schnarrenberg
10	Enzklösterle	43	Stuttgart-Echterdingen
11	Eschbach	44	Titisee
12	Feldberg	45	Triberg
13	Freiburg	46	Uffenheim
14	Freudenstadt	47	Ulm
15	Friedrichshafen	48	Ulm-Wilhelmsburg
16	Gailingen	49	Waldachtal-Lützenhardt
17	Hinterzarten	50	Walldürn
18	Höschenschwand	51	Würzburg
19	Hornisgrinde	101	Bad Dürkheim
20	Idar-Oberstein	102	Dogern
21	Isny	103	Lindau
22	Kandern-Gupf	104	Müllheim
23	Karlsruhe	105	Neuhausen ob Eck
24	Kempten	106	Pforzheim-Ispringen
25	Klippeneck	107	Sipplingen
26	Königsfeld	108	Söllingen
27	Konstanz	109	Stockach-Espasingen
28	Lahr	110	Stuttgart-Stadt
29	Laupheim	111	Todtmoos
30	Leipheim	112	Waldsee, Bad-Reute
31	Mannheim	113	Weilheim-Bierbronn
32	Memmingen	114	Weingarten
33	Michelstadt Vielbrunn		

The complexity of the terrain in the study area was further quantified by a distance limited topographic exposure score (τ) which was described by [39,40]. The τ -values were calculated by summing the vertical angles to the skyline at 100 m intervals for the eight main compass directions up to a distance limited to 1000 m for each grid cell. Lower τ -values indicate higher exposition of a site to wind. In the study area, close to 60% of all τ -values are in the range 0° to 25° (Figure 1c). The τ -values used in this study were provided by the Forest Research Institute of Baden-Wuerttemberg.

Wind speed data

Daily mean wind speed values provided by the German Weather Service (DWD) for 65 meteorological stations were used in the present study. The stations are listed and indexed in Table 2. The indexed DWD-stations are distributed in the entire study area as well as in the bordering federal states Rhineland-Palatinate, Hesse and Bavaria as shown in Figure 2. Although hourly mean wind speed values are also available from the DWD, daily mean wind speed values available for the period 1975-01-01 to 2010-12-31 were used here (i) to maximize the number of stations with multi-year wind speed records and (ii) to minimize the effects of terrain-induced wind speed variations on



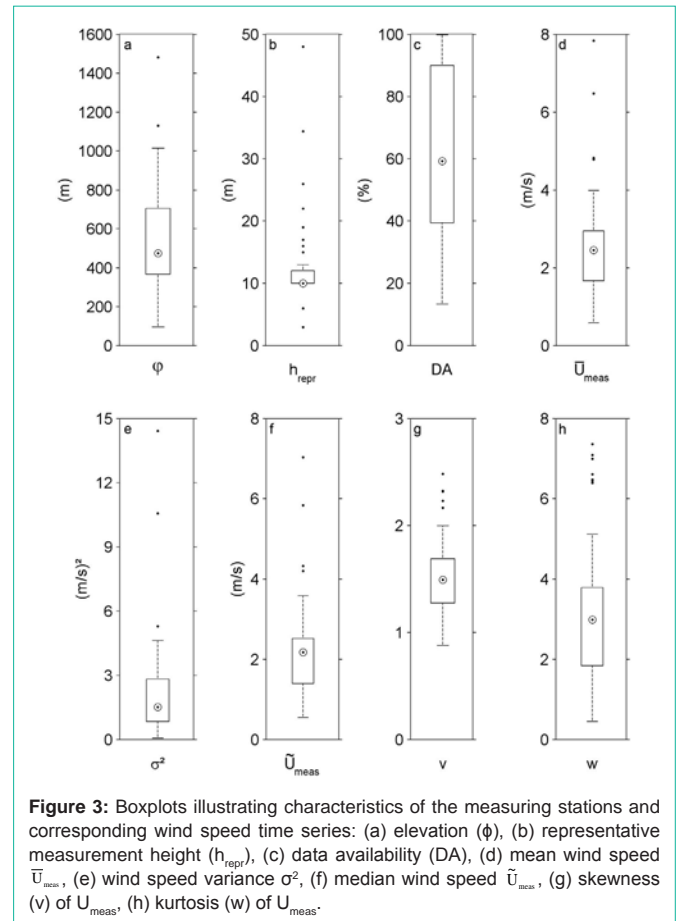
sub-daily scales. Wind speed data measured at sub-daily intervals might introduce local, diurnal and multi-modal behavior into time series [30,41-43] that is often not reflected by neighboring stations, especially in complex terrain. The potential to fill data gaps in multi-modal, but mostly incomplete wind speed time series using wind speed data from neighboring DWD-stations is limited. Moreover, it is important to note that until 2001-03-31 daily mean wind speed values provided by the DWD were calculated from three measurements made at the climatological standard times (7 a.m., 2 p.m. and 9 p.m.). Afterwards, 24 hourly mean wind speed values were used to calculate daily mean values.

Wind direction was not included in model development because of (i) the limited number of wind direction measurements in the study area and (ii) the complexity of the terrain in the study area which induces a spatiotemporal variability in wind direction that cannot be reproduced by the available measuring stations.

Out of the 65 wind speed time series 51 time series are combined in the parameterization dataset (DS1). This dataset was used for model building and parameterization. 15 DS1-time series, for which wind speed data are available for more than 85% of all days in the investigation period, served as reference time series (DS1_{ref}). The DS1_{ref}-data subset was used to complete and prepare all other wind speed time series. Wind speed values measured at 14 DWD-stations were combined to the validation dataset (DS2) and used to validate the model developed based on the data contained in DS1.

In the investigation period, wind speed (U_{meas}) was not always measured at the standard measurement height of 10 m above ground level (a.g.l.) defined by the World Meteorological Organization (WMO). Moreover, at a number of stations the measurement height changed within the investigation period. In these cases, the measurement height at which wind speed was measured at the end of the investigation period was defined as “representative” (h_{repr}) for the respective station.

In Figure 3 important characteristics of the measurement sites are summarized as boxplots. The stations are located in the ϕ -range 96 m to 1481 m a.s.l. (Figure 3a) with the ϕ -median being 475 m. Two stations (Feldberg, Hornisgrinde) are located above 1100 m. The h_{repr} -



median is 10 m a.g.l. (Figure 3b). However, at a few sites, wind speed measurements have been carried out far off the WMO-standard, e.g. 3 m a.g.l. at the stations Dobel and Isny or 48 m a.g.l. at the station Karlsruhe.

None of the analyzed wind speed time series was complete. Data availability (DA) ranges between 13% corresponding to about 4 years and 8 month (station Hornisgrinde) and 99.7% (Figure 3c). The median of station-specific mean values of U_{meas} (\bar{U}_{meas}) is 2.5 m/s (Figure 3d) and the associated median of the variance of U_{meas} (σ^2) displayed in Figure 3e is 1.5 (m/s)². Since wind speed values are non-normally distributed, the station-specific medians of U_{meas} (\tilde{U}_{meas}) are presented in Figure 3f in addition to \bar{U}_{meas} . The median of \tilde{U}_{meas} is slightly lower (2.2 m/s) than the median of \bar{U}_{meas} . The medians of skewness (v) and kurtosis (w) of U_{meas} are 1.5 (Figure 3g) and 3.0 (Figure 3h). The shape of the v -boxplot indicates that all empirical wind speed distributions are right-skewed which is supported by the \tilde{U}_{meas} -values being mostly lower than the \bar{U}_{meas} -values. The w -boxplot indicates that all empirical wind speed distributions are more peaked than the normal distribution.

Data preparation

Since time series of wind speed measured near the ground are known to have problems with temporal homogeneity, spatial representativity and completeness, a comprehensive data preparation process preceded the statistical analysis and model building. Inhomogeneities and temporal discontinuities in wind speed time

series can arise from station relocation [44,45], instrument change [44], changes in measurement height [44-47], changes in sampling frequency [45] and changes in the surroundings of the measurement sites [5,45,48]. It is critical that all wind speed data are consistent before analysis. Especially, adjusting the actual measurement height to a common height is of great importance [44,46,47].

Data preparation started with completing the 15 wind speed time series contained in $DS1_{ref}$ by using the ensemble learning method bagging implemented in the Matlab® Statistics Toolbox (The Math Works Inc., Natick, MA, Release 2014a). Bagging stands for bootstrap aggregating and enhances the predictive capabilities of machine learning algorithms used in regression [49,50].

Before gap filling started, it was made sure that the Bagging models are able to reproduce the data of the incomplete time series with high accuracy. The coefficient of determination (R^2) which was calculated between the available measured data and the modeled parts of the incomplete time series was always higher than 0.90 during the gap filling process. Up to six wind speed time series measured at neighboring DWD-stations were used to mutually fill gaps in $DS1_{ref}$ -time series.

Following gap filling, the completed $DS1_{ref}$ -time series were tested for homogeneity. The search for inhomogeneities started with the analysis of the DWD-metadata describing station history. Unfortunately, the documentation of station history was often poor and only of limited use because the metadata provide only basic information on data quality, station relocations and changes in measurement height.

In a second step, inhomogeneities were searched and identified by using the RHtestsV4 software package implemented in R [51] by [52]. The methodology is based on the penalized-maximal-F-test [53,54] and quantile-matching [55]. The penalized-maximal-F-test is used to control whether there are shifts in the trend component during the measurement period, while considering annual cycle, linear trend and lag-1 autocorrelation. After the detection of shifts in the trend component, quantile-matching adjusts the empirical distributions of all inhomogeneous time series segments to the empirical distribution of the last homogeneous time series segment.

After completing the $DS1_{ref}$ -data homogenization process, the linear trend was removed from the wind speed time series because the proposed statistical model does not consider the trend component. The 15 completed, homogenized and detrended $DS1_{ref}$ -time series were then used to fill the gaps in all other 50 wind speed time series with $DA < 85\%$. Bagging was applied again to fill the data gaps by using up to six $DS1_{ref}$ -time series. Then the completed time series were tested for homogeneity and detrended as described above.

The Hellmann power law [56] was used to extrapolate wind speed to the WMO-standard wind speed measurement height of 10 m a.g.l. at all stations where necessary. Since the application of the Hellmann power law requires wind speed data from a second height a.g.l., use was made of the area-wide available ERA-Interim reanalysis wind data [57] provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). The reanalysis data is interpreted as an indicator of the kinetic energy resource available from higher parts of the troposphere that has a determining influence on the near-surface wind speed field.

In the study area, the reanalysis data have a spatial resolution of 0.125° (~13 km). After pre-testing the predictive power of the two horizontal ERA-Interim wind vector components (u , v) available at the 700 hPa, 850 hPa and 950 hPa pressure levels, which are affected only to a limited extent by surface properties, it turned out that the 850 hPa level horizontal wind vector component data (u_{850hPa} , v_{850hPa}) were the most informative predictors for the target variables modeled in this study. Thus, u_{850hPa} - und v_{850hPa} -values available for 00 UTC, 06 UTC, 12 UTC and 18 UTC in the period 1979-01-01 to 2010-12-31 were used to calculate daily mean values of wind speed (U_{850hPa}). As done by [58], the absolute height a.g.l. associated with daily U_{850hPa} -values was derived from the ERA-Interim geopotential height layer by converting geopotential height (gpdm) to geometric height (h_{850hPa}) in meters a.g.l. In order to make the U_{850hPa} -data usable as predictor variable for the final wind speed model, it was interpolated on a 50 m resolution grid. It was shown by [59] that U_{850hPa} changes only little on scales comparable to the size of the study area, so the interpolation was thought of being justified.

Based on daily U_{meas} - and U_{850hPa} -values at the nearest ERA-Interim grid points to the positions of the 65 DWD-stations, the median value of the Hellmann exponent (\tilde{E}) was calculated for all stations. The equation used to calculate \tilde{E} is:

$$\tilde{E} = \frac{\ln(U_{850hPa}/U_{meas})}{\ln(h_{850hPa}/h_{repr})} \quad (1)$$

Once station-specific \tilde{E} is known, it can be used to calculate daily, station-specific U_{10m} -values:

$$U_{10m} = U_{meas} \times \left(\frac{10m}{h_{repr}} \right)^{\tilde{E}} \quad (2)$$

The values of \tilde{E} determined for the 65 stations ranged between 0.06 (station Hornisgrinde) and 0.66 (station Triberg). At all grid points in the study area, \tilde{E} varies between 0.06 and 0.70.

Effects of atmospheric stability on the Hellmann power law [47] were not considered because the data that is needed to adequately take atmospheric stability into account were not available.

Probability distribution fitting

To provide probabilistic estimates of the long-term spatial variability of the near-surface wind speed pattern in the study area, 48 cumulative distribution functions (CDF) were fitted to the empirical cumulative distribution functions (CDF_{emp}) derived from the wind speed time series. The EasyFit software (MathWave Technologies, Dnepropetrovsk, Ukraine, version 5.5) was used to carry out the estimation of CDF-parameters as well as to calculate CDF and evaluate the goodness-of-fit. Basic information on the fitted distributions, the number of CDF-parameters and the fitting methods implemented in the EasyFit software are summarized in Table 3 according to [60]. Depending on the distribution, the EasyFit software applies maximum likelihood estimation (MLE), least square estimation (LSE), method of moments (MOM) and method of L-moments (LMOM) to fit CDF to CDF_{emp} .

For each of the 65 stations the goodness-of-fit of the 48 CDF was evaluated by (i) applying the Kolmogorov-Smirnov (KS) Test [24,34,61,62] and (ii) analyzing probability-probability (P-P) plots

Table 3: Distributions with abbreviation (number of distribution parameters) and method used by the EasyFit software to fit CDF to CDF_{emp} [60].

Distribution	Abbreviation (no. parameters), fitting method	Distribution	Abbreviation (no. parameters), fitting method
Beta	BE4, MLE	Laplace	LA2, MOM
Burr	BU3, MLE	Levy	LE1, MLE
	BU4, MLE		LE2, MLE
Cauchy	CA2, MLE	Log-Gamma	LG2, MOM
Chi-Squared	CH1, MOM	Logistic	LO2, MOM
	CH2, MLE		LL2, LSE
Dagum	DA3, MLE	Log-Logistic	LL3, MLE
	DA4, MLE		LN2, MLE
Erlang	EL2, MOM	Lognormal	LN3, MLE
	EL3, MLE		Log-Pearson
Error	ER3, MLE	Nakagami	NA2, MOM
Error Function	EF1, MOM	Normal	NO2, MLE
Exponential	EX1, MOM	Pareto (1 st kind)	PAF2, MLE
	EX2, MLE	Pareto (2 nd kind)	PAS2, MLE
Fatigue Life	FL2, MLE	Pearson Type 5	PFi2, MLE
	FL3, MLE		PFi3, MLE
Frechet	FR2, LSE	Pearson Type 6	PSi3, MLE
	FR3, MLE		PSi4, MLE
Gamma	GA2, MOM	Pert	PE3, MLE
	GA3, MLE	Phased Bi-Exponential	PBE4, LSE
Generalized Extreme Value	GE3, LMOM	Phased Bi-Weibull	PBW6, LSE
Generalized Gamma	GG3, MLE	Power Function	PF3, MLE
	GG4, MLE		RA1, MOM
Generalized Logistic	GL3, LMOM	Rayleigh	RA2, MLE
Generalized Pareto	GP3, LMOM	Reciprocal	RE2, MLE
Gumbel Max	GMX2, MOM	Rice	RI2, MLE
Gumbel Min	GMN2, MOM	Student's t	ST1, MOM
Hyperbolic Secant	HS2, MOM	Triangular	TR3, MLE
Inverse Gauss	IG2, MOM	Uniform	UN2, MOM
	IG3, MLE	Wakeby	WK5, LMOM
Johnson SB	JSB4, MOM	Weibull	WE2, LSE
Johnson SU	JSU4, MOM		WE3, MLE
Kumaraswamy	KU4, MLE		

which plot CDF and CDF_{emp} against each other [20,63]. The decision to use the KS-Test for goodness-of-fit evaluation was made because the KS-Test weights central tendencies more than other commonly used goodness-of-fit tests like the Anderson-Darling test [64]. The Anderson-Darling test weights observations in the distribution tails more than the KS-Test.

As will be demonstrated in the Results and Discussion section, the five-parameter Wakeby-distribution [65] shows a better performance than all other distributions. The Wakeby-distribution (WK5) is therefore chosen to model statistical properties of the empirical wind speed distributions. The Wakeby-distribution can be defined by its quantile function as [66-70]:

$$x(F) = \varepsilon + \frac{\alpha}{\beta} \left[1 - (1-F)^\beta \right] - \frac{\gamma}{\delta} \left[1 - (1-F)^{-\delta} \right] \quad (3)$$

where F is the non-exceedance probability with x(F) being the F-associated quantile value, α , β , γ and δ are parameters and ε is the location parameter. This parameterization exhibits WK5 being a generalization of the Generalized Pareto distribution for $\alpha = 0$ or $\gamma = 0$. To be a valid quantile function the conditions $\gamma \geq 0$ and $\alpha + \gamma \geq 0$ must

hold. The quantile function is defined for the domain $\varepsilon \leq x < \infty$ if $\delta \geq 0$ and $\gamma > 0$, $\varepsilon \leq x \leq \varepsilon + \alpha/\beta - \gamma/\delta$ if $\delta < 0$ and $\gamma = 0$ [67-69].

Geographic data

For WK5-model building, the predictive power of the predictor variables listed in Table 4 was tested. In addition to ϕ , the orographic features aspect (η), curvature (φ) and slope (σ) were deduced from a digital terrain model (DTM) on a 50 m resolution grid for the entire study area. They were calculated using the ArcGIS® 10.2 software Spatial Analyst extension. For all geospatial data sets the same coordinate system (Gauß-Krüger, reference ellipsoid Bessel 1841) was defined.

Following [71], “effective” z_0 - and φ -values were calculated for different directional sectors. This was done to account not only for local surface roughness and terrain characteristics at each grid point itself but also for the surface roughness and terrain characteristics in the upwind fetch. The effective roughness length ($z_{0,eff}$) and effective curvature (φ_{eff}) were calculated for the four direction sectors (NE (0°-90°), SE (91°-180°), SW (181°-270°), NW (271°-360°)) and for four different distance ranges (d_1 : 50-250 m; d_2 : 251-500 m; d_3 : 251-1000 m; d_4 : 501-1000 m) for all grid points.

Table 4: Predictor variables available for this study. The • marks the predictor variables that were used in the final LS Boost-models to reproduce Y_L (BM_{Y_L}), Y_{R1} ($BM_{Y_{R1}}$), Y_{R2} ($BM_{Y_{R2}}$) and Y_{R3} ($BM_{Y_{R3}}$).

ID	Predictor variable	Symbol	Y_L	Y_{R1}	Y_{R2}	Y_{R3}
1	Aspect	η				
2	Curvature, local	ϕ				
3	Curvature, effective, NE, d_1	$\phi_{\text{eff,NE},d1}$	•	•	•	•
4	Curvature, effective, SE, d_1	$\phi_{\text{eff,SE},d1}$			•	
5	Curvature, effective, SW, d_1	$\phi_{\text{eff,SW},d1}$				
6	Curvature, effective, NW, d_1	$\phi_{\text{eff,NW},d1}$		•	•	
7	Elevation	ϕ	•	•	•	•
8	Latitude	λ				
9	Longitude	ψ				
10	Roughness length, local	Z_0				•
11	Roughness length, effective, d_1	$Z_{\text{0eff},d1}$				
12	Roughness length, effective, d_2	$Z_{\text{0eff},d2}$				
13	Roughness length, effective, d_3	$Z_{\text{0eff},d3}$				
14	Roughness length, effective, d_4	$Z_{\text{0eff},d4}$				
15	Roughness length, effective, NE, d_1	$Z_{\text{0eff,NE},d1}$	•	•	•	•
16	Roughness length, effective, SE, d_1	$Z_{\text{0eff,SE},d1}$	•		•	
17	Roughness length, effective, SW, d_1	$Z_{\text{0eff,SW},d1}$	•	•	•	•
18	Roughness length, effective, NW, d_1	$Z_{\text{0eff,NW},d1}$			•	
19	Roughness length, effective, NE, d_2	$Z_{\text{0eff,NE},d2}$				
20	Roughness length, effective, SE, d_2	$Z_{\text{0eff,SE},d2}$				
21	Roughness length, effective, SW, d_2	$Z_{\text{0eff,SW},d2}$				
22	Roughness length, effective, NW, d_2	$Z_{\text{0eff,NW},d2}$				
23	Slope	σ				
24	Topographic exposure	τ		•	•	•
25	Median wind speed (850 hPa level)	$\tilde{U}_{850\text{hPa}}$	•			•

Here, “effective” means that the z_0 - and ϕ -values available on the 50 m resolution grid were averaged over d_1 - d_4 because surface roughness and terrain characteristics in the immediate vicinity of the DWD-stations strongly affect local wind conditions. The local, isotropic z_0 -values listed in Table 1 were used for the $z_{0,\text{eff}}$ -calculations. The consideration of different directional sectors in the $z_{0,\text{eff}}$ - and ϕ_{eff} -calculations introduces a minimum of directionality into the proposed model that otherwise has no directional dependence.

Wind speed model building

The proposed statistical wind speed model uses orographic features, surface roughness and reanalysis data to predict WK5-parameters on a 50 m resolution grid in the entire study area. The model is developed by making use of the Ensemble Learning algorithm for least squares boosting (LSBoost) implemented in the Matlab® Statistics Toolbox. LSBoost uses of a sequence of regression trees called weak learners (B) with the aim to minimize the mean-squared error (MSE) between target variable Y and the aggregated prediction of the weak learners (Y_{pred}) based on methods described in [72]. LSBoost starts with an initial guess of the aggregated prediction of the median of the target variable (\tilde{Y}) as a function of the predictor variables (X). Then it combines multiple regression models B_1, \dots, B_m in a weighted manner to improve its overall predictive performance [73]:

$$Y_{\text{pred}}(X) = \tilde{Y}(X) + v \sum_{m=1}^M \rho_m B_m(X) \tag{4}$$

with ρ_m being the weight for model m, M is the total number of weak learners, v with $0 < v \leq 1$ being the learning rate.

The LSBoost-models were parameterized using DS1-data and the corresponding values of surface roughness, orographic features and reanalysis data at and in the surroundings of the wind speed measuring stations. The predictive performance of the final LSBoost-models (BM) was evaluated using DS1- and DS2-data. Although finding the optimal number of weak learners is a trial and error process, a low weak learner number ($M < 100$) should already be sufficient to reasonably approximate Y. Otherwise, a significant improvement of the prediction result will not be achieved by increasing the number of weak learners. LSBoost was chosen to model the WK5-parameters because (i) it is insensitive to outliers, (ii) its stability is maintained during the adjustment process as only simple regression models are added, (iii) irrelevant predictor variables are sorted out and (iv) no data transformation is necessary. Thus, it can be used to reasonably model low quality data [74].

Before the WK5-model building process started, the strength of collinearity among the available predictor variables was tested by evaluating the variance inflation factor (VIF) and the conditions index (CI) in combination with variance-decomposition proportions (VD) according to [75]. By applying these tests, collinearity as quantified by $VIF > 2$ and/or $CI > 30$ with $VD > 0.5$ was not identified among the predictor variables used in the final LSBoost-models.

Since WK5 is a five-parameter distribution, it can be fitted to a large number of shapes. With the choice of appropriate parameters, WK5 can mimic most of the commonly used distributions [65,67]. However, the determination of its parameters might not always be stable. In order to enhance the stability of the WK5-parameter estimation, the WK5-parameters were modeled separately following [70] with a similar but simplified approach. First, the left part of the right-hand side of equation 3 (Y_L) which is represented by α , β and ε was modeled for $F = 0.25$:

$$Y_L = \varepsilon + \frac{\alpha}{\beta} \left[1 - (1 - 0.25)^\beta \right] \tag{5}$$

The separation of Y_L from the right part of the right-hand side of equation 3 (Y_R) is justified because Y_L is quasi-constant for $F \geq 0.25$, i.e. distinct changes of Y_L are expected only for empirical values located in the 1st quartile [70].

Next, Y_R , which is represented by γ and δ , was modeled for $F = 0.50$ (Y_{R1}), $F = 0.75$ (Y_{R2}) and $F = 0.99$ (Y_{R3}). Knowing Y_{R1} , Y_{R2} and Y_{R3} , the following system of nonlinear equations can be solved to determine γ and δ :

$$\begin{cases} \frac{\gamma}{\delta} [1 - (1 - 0.50)^{-\delta}] + [Y_{R1} - Y_L] = 0 \\ \frac{\gamma}{\delta} [1 - (1 - 0.75)^{-\delta}] + [Y_{R2} - Y_L] = 0 \\ \frac{\gamma}{\delta} [1 - (1 - 0.99)^{-\delta}] + [Y_{R3} - Y_L] = 0 \end{cases} \tag{6}$$

After modeling the two parts of the right-hand side of equation 3 separately, Y_L and Y_R were recombined to yield the Wakeby-distribution with modeled parameters ($WK5_{\text{mod}}$). Following [33,76,77], the predictive performance of $WK5_{\text{mod}}$ was evaluated in DS1 and DS2 by the mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and R^2 .

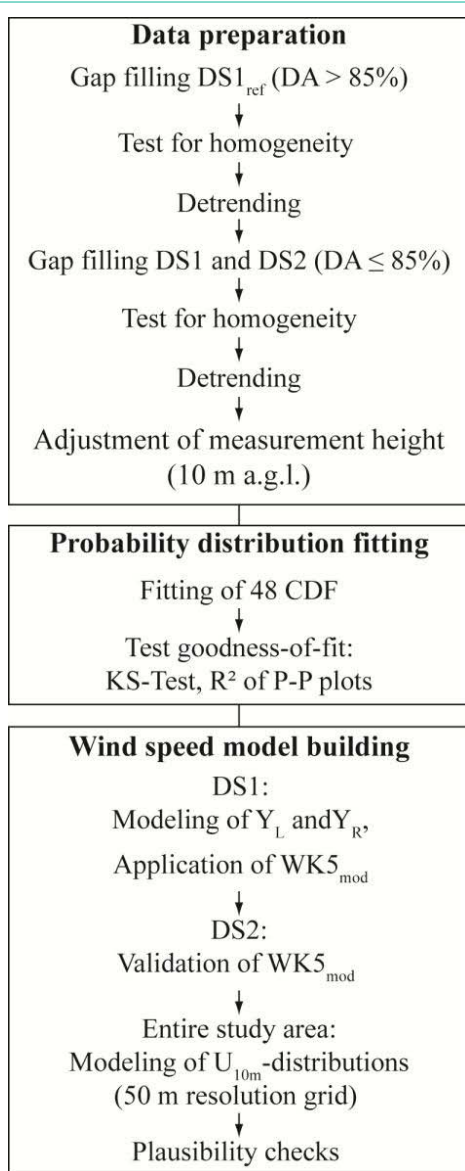


Figure 4: Schematic representation of the workflow including data preparation, probability distribution fitting and wind speed model building.

The workflow including data preparation, probability distribution fitting and wind speed model building is summarized in Figure 4.

Results and Discussion

Probability distribution fitting

The absolute frequency of all CDF which ranked at least once best after evaluating the goodness-of-fit by the KS-Test statistic D is shown in Figure 5. It is obvious that WK5 could be fitted most often (at 26 stations) best to CDF_{emp} . Moreover, based on the evaluation of D, WK5 always ranked between #1 and #10 after being fitted to all 65 CDF_{emp} . These findings are in accordance with [20] who also report a good performance of WK5 in comparison to other distributions. Only the Johnson SB distribution [78] could also be fitted best to a noteworthy number of CDF_{emp} (at 15 stations). Since the varieties of the Weibull-distribution (WE2, WE3) never ranked #1, they are not shown.

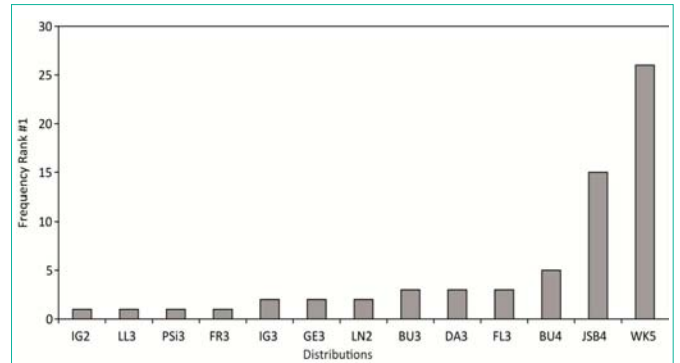


Figure 5: Absolute frequency of CDF which ranked at least once best (Rank #1) after being fitted to CDF_{emp} .

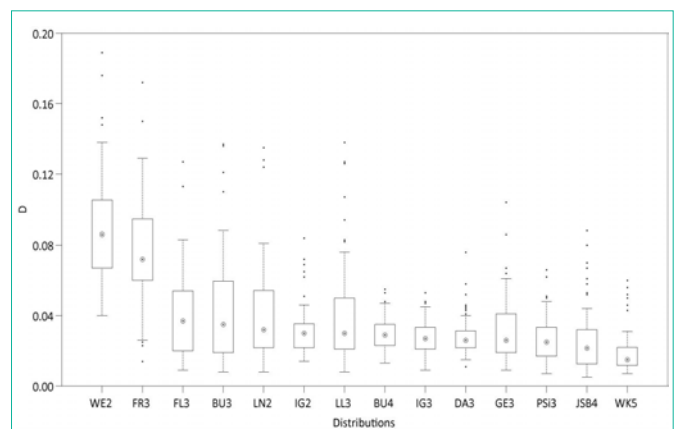
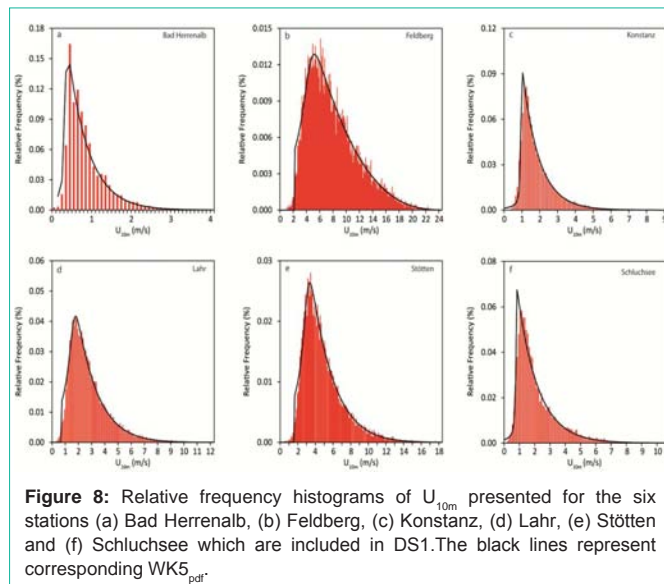
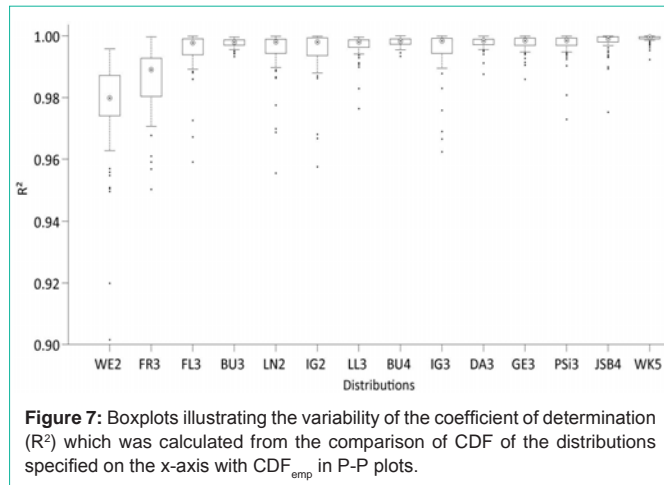


Figure 6: Boxplots illustrating the variability of the KS-Test statistic D for CDF that ranked at least once best after being fitted to CDF_{emp} , as well as a boxplot for D derived from fitting WE2 to CDF_{emp} .

To illustrate the findings of the KS-Test, Figure 6 shows boxplots of D for all distributions that at least once ranked #1 plus the boxplot derived from the D-values determined from fitting WE2 which was used in numerous previous studies as mentioned in the Introduction (results obtained for WE3 are very similar to the results obtained for WE2 and therefore not shown). It is obvious that the D-median of 0.015 obtained from fitting WK5 to CDF_{emp} is the lowest. In comparison to that, the D-median for WE2 is 0.086. In addition and in comparison to most of the other displayed distributions, the variability of D (e.g. as illustrated by the interquartile range) determined from WK5-fitting is also low.

The boxplots presented in Figure 7 are based on R^2 -values which were calculated from P-P plots produced for all distributions that at least once ranked #1 plus WE2. These boxplots are similar to the boxplots presented by [20] and further confirm the WK5-capabilities to reasonably characterize CDF_{emp} . The median of WK5-related R^2 -values is 1.000; the interquartile range is represented by $R^2 > 0.999$.

To demonstrate the flexibility of WK5 in fitting U_{10m} -distributions, Figure 8 shows relative frequency histograms of U_{10m} from six stations contained in DS1 that were fitted using WK5-probability density functions ($WK5_{pdf}$). At station Bad Herrenalb (Figure 8a), the general wind speed level was generally very low and varied in the range 0 m/s to 4 m/s. The corresponding empirical frequency distribution is right-skewed with relative frequency values peaking below 1 m/s. It is



obvious that WK5 fits the relative frequency distribution reasonably well. In contrast to that, Figure 8b shows the relative frequency distribution from station Feldberg. At this station which is located on the top of the Feldberg, \bar{U}_{10m} was highest. The wind speed range on the x-axis includes U_{10m} -values up to 24 m/s. The relative frequency distribution is also skewed to the right and peaks in the U_{10m} -range of 5 m/s to 6 m/s. The relative frequency distributions shown in Figure 8c (station Konstanz) to Figure 8f (station Schluchsee) exemplarily represent U_{10m} -ranges between the rather low U_{10m} -values recorded at station Bad Herrenalb and the high U_{10m} -values from station Feldberg. The relative frequency distributions are also right-skewed without exception and more peaked than the normal distribution.

The Wakeby-distribution has the ability to fit almost all quantiles of the presented relative frequency distributions with sufficient accuracy. Especially, wind speed values located in the 2nd to 4th quartile are fitted well which is crucial for wind energy applications [2,20]. However, the limited potential of WK5 to fit the lowest levels of wind speed is obvious. This behavior of WK5 can be attributed to the facts that (i) its domain of definition is bounded by ϵ at the lower tail and (ii) the lowest wind speed values are most affected by the immediate

Table 5: Mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and coefficient of determination (R^2) calculated from the comparison of CDF_{mod} with CDF_{emp} of the wind speed time series included in DS1 and DS2.

Data set	Quantiles	ME (m/s)	MAE (m/s)	RMSE (m/s)	R^2
DS1	0.01	-0.03	0.15	0.26	0.75
	0.25	-0.02	0.14	0.21	0.96
	0.50	0.00	0.14	0.19	0.98
	0.75	0.01	0.18	0.24	0.98
	0.99	0.04	0.31	0.40	0.98
DS2	0.01	-0.03	0.17	0.22	0.59
	0.25	0.03	0.10	0.13	0.92
	0.50	0.09	0.16	0.20	0.93
	0.75	0.11	0.27	0.34	0.90
	0.99	0.15	0.78	0.96	0.80

surroundings of the measurements sites, which cannot be properly reproduced by WK5.

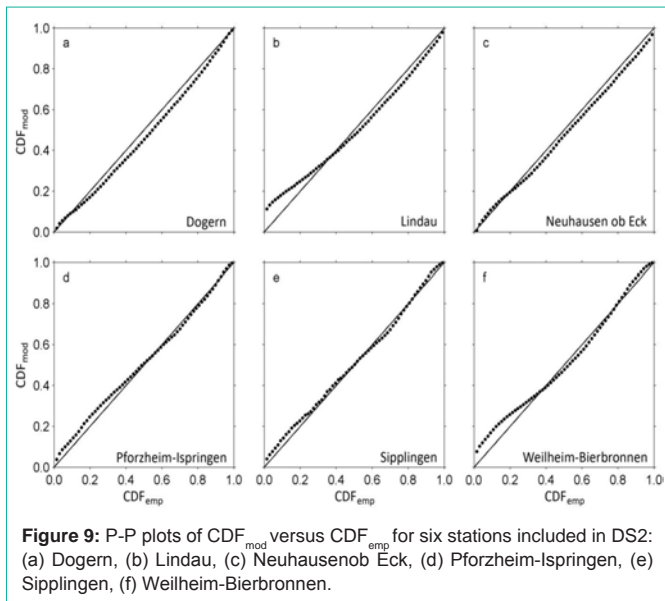
Wind speed model development

The selection of the final predictor variable combinations was a trial and error process. It started with keeping $M = 20$. The final BM-configuration was chosen after comparing the predictive performance of several configurations with different combinations of predictor variables. The predictor variable combination which simultaneously gave the lowest MAE-values in DS1 and DS2 was selected as predictor variable combination to build BM. After the final predictor variable combinations were certain, M was increased to further minimize MAE. The final number of weak learners varied between $M = 30$ for modeling $F = 0.75$ and $M = 97$ for modeling $F = 0.99$.

The predictor variables having the strongest univariate linear association ($|R| > 0.5$) with any of Y_{L1} , Y_{R1} , Y_{R2} and Y_{R3} were τ , $\phi_{eff,NE,d1}$ and $\phi_{eff,SE,d1}$. The predictor variable combinations which reproduced Y_{L1} , Y_{R1} , Y_{R2} and Y_{R3} best are summarized in Table 4. They were fed into the final LSBoost-models $BM_{Y_{L1}}$, $BM_{Y_{R1}}$, $BM_{Y_{R2}}$ and $BM_{Y_{R3}}$ to reproduce the WK5-parameters associated with the wind speed time series contained in DS1 and DS2.

The combinations of predictor variables that were used in this study to model near-surface wind speed distributions are similar to combinations of predictor variables used in previous studies. The regression equation-based statistical wind field model reported by [79] estimates mean annual wind speed as a function of elevation, latitude, longitude, surface shape and surface roughness in Germany on a 200 m resolution grid. Without accounting for surface roughness, [80] used generalized additive models to estimate maximum daily gust speed (98 percentile) in Switzerland on a 50 m resolution grid based on landform, elevation, curvature and slope.

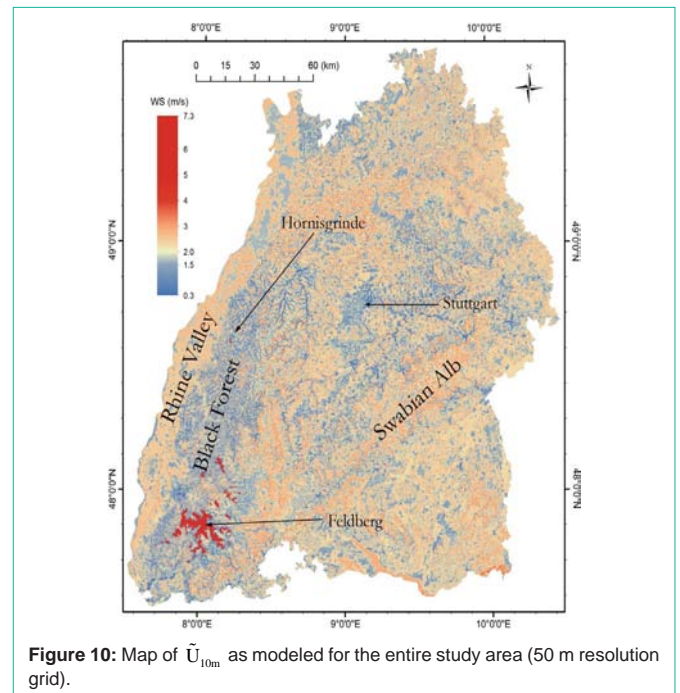
To examine the modeling accuracy of $WK5_{mod}$, it was used to estimate CDF_{emp} . The quality of $WK5_{mod}$ -estimated CDF_{emp} (CDF_{mod}) was evaluated by ME, MAE, RMSE and R^2 . The results of this evaluation are given in Table 5 for various quantiles. As can be expected, the prediction results are in most cases better for DS1 than for DS2. They demonstrate that there is a tendency that with increasing quantiles the absolute prediction error increases, both in DS1 and DS2. This is mainly attributable to increasing wind speed represented by increasing quantiles. However, the prediction errors are generally small and, except for $F = 0.99$, in the typical range of wind speed measuring accuracy. Except for $F = 0.01$ associated



with DS1, R^2 is at least 0.80, which further indicates that $WK5_{mod}$ reasonably reproduces the analyzed quantiles. To illustrate the modeling abilities of $WK5_{mod}$ for individual stations, Figure 9 shows CDF_{mod} plotted against CDF_{emp} of six stations contained in DS2 which were not part of the model parameterization process. The closer the data to the 1:1 line, the better the agreement between CDF_{mod} and CDF_{emp} . It is clear, that if distinct deviations of the presented data from the 1:1 line occur, then they occur in the low quantile range (e.g. Figures 9b,9f). The low quantiles correspond to low wind speed values which are most probably a direct result of small-scale surface and terrain characteristics that are not always correctly reproduced by $WK5_{mod}$. Furthermore, its domain of definition causes $WK5$ to be prone to overestimate the frequency of lower wind speeds [20].

Since $WK5_{mod}$ is well able to reproduce CDF_{emp} of all stations included in DS1 and DS2, BM_{Y1} , BM_{YR1} , BM_{YR2} and BM_{YR3} were used to model $WK5$ -parameters based on the selected predictor variable combinations. This basically opens the possibility to model all wind speed quantiles in the entire study area. As an example, $WK5_{mod}$ -modeled \tilde{U}_{10m} -values are shown in Figure 10 in a detailed map at a resolution of $50\text{ m} \times 50\text{ m}$. As expected, highest \tilde{U}_{10m} -values (7.3 m/s) are found on the tops of the highest elevations in the study area ($\phi > 1100\text{ m}$) like the Feldberg region in the southern part of the Black Forest and in the vicinity of the Hornisgrinde (5.8 m/s) which is the highest elevation in the northern part of the Black forest. The lowest \tilde{U}_{10m} -values are found in a number of narrow, forested valleys in the Black Forest where \tilde{U}_{10m} drops down to 0.3 m/s. In and around Stuttgart, being the capital and the largest urban area of Baden-Wuerttemberg, \tilde{U}_{10m} is mostly below 1.5 m/s.

Although the proposed $WK5$ -model seems to be well able to capture main features of the near-surface wind speed field in the study area, it is far off from being fully developed. Future versions of the model should be built on a monthly basis to better account for seasonal variations of wind speed in Germany [81]. Although $WK5_{mod}$ already accounts to a minor degree to directional variations in wind speed, more needs to be done to include wind direction in the modeling process. First of all, even before model building, more stations



where wind speed and wind direction are measured simultaneously with high quality should be established in all parts of the study area. Thus, a more complete description of the statistical properties of the near-surface wind field, especially in the low mountain ranges, can be achieved by modeling joint frequency distributions of wind speed and wind direction instead of modeling the frequency distribution of wind speed alone [4,18,41,82].

Conclusions

The proposed statistical model is able to simulate near-surface wind speed quantiles in a large area with complex terrain with sufficient accuracy. The parameters of the Wakeby-distribution were chosen to be modeled by LSBoost-models because the Wakeby-distribution demonstrated its superior abilities in fitting all empirical wind speed distributions available for this study. The LSBoost-models reasonably reproduce the parameters of the Wakeby-distribution as a function of the predictor variables (i) roughness length, (ii) elevation, curvature and topographic exposure and (iii) ERA-Interim wind speed at the 850 hPa level. These predictor variables are available in similar versions in many parts of the world. Therefore, it is assumed that the model can be transferred to other wind speed data sets.

Since the modeled Wakeby-distribution parameters are available at every grid point in the entire study area, detailed maps of wind speed quantiles (not only the median) can be produced. Due to their high spatial resolution, these maps provide useful basic information on the near-surface wind field for applications like urban planning, regional planning, air pollution control or civil engineering. There is also potential to use the methodology for wind site assessment, because during the model building process the Hellman exponent is calculated in the entire study area. It can be used to extrapolate modeled 10 m a.g.l. wind speed fields to heights where wind turbines harvest kinetic energy contained in the wind. A major task for the future will be the refinement of the proposed model towards the

monthly-based inclusion of wind speed and wind direction measured on sub-daily scales.

Acknowledgement

We thank (i) the German Weather Service for providing wind speed data, (ii) the Forest Research Institute of Baden-Wuerttemberg for providing topographic exposure score values.

References

- Garcia A, Torres JL, Prieto E, de Francisco A. Fitting wind speed distributions: a case study. *Sol Energy*. 1998; 2: 139-144.
- Celik AN. Assessing the suitability of wind speed probability distribution functions based on wind power density. *Renew Energ*. 2003; 28: 1563-1574.
- Wieringa J. Roughness-dependent geographical interpolation of surface wind speed averages. *Quart J R Met Soc*. 1986; 112: 867-889.
- Kalthoff N, Bischoff-Gauß I, Fiedler F. Regional effects of large-scale extreme wind events over orographically structured terrain. *Theor Appl Climatol*. 2003; 74, 53-67.
- Wieringa J. Does representative wind information exist? *J Wind Eng Ind Aerod*. 1996; 65: 1-12.
- Petersen EL, Mortensen NG, Landberg L, Højstrup J, Frank HP. Wind Power Meteorology. Part I: Climate and Turbulence. *Wind Energ*. 1998; 1: 25-45.
- Landberg L, Myllerup L, Rathmann O, Lundtang Petersen E, Hoffmann Jørgensen B, Badger J, et al. Wind Resource Estimation – An Overview. *Wind Energ*. 2003; 6: 261-271.
- Schindler D, Grebhan K, Albrecht A, Schönborn J. Modelling the wind damage probability in forests in Southwestern Germany for the 1999 winter storm 'Lothar'. *Int J Biometeorol*. 2009; 53: 543-554.
- Schindler D, Grebhan K, Albrecht A, Schönborn J, Kohnle U. GIS-based estimation of the winter storm damage probability in forests: a case study from Baden-Wuerttemberg (Southwest Germany). *Int J Biometeorol*. 2012; 56: 57-69.
- Klawa M, Ulbrich U. A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Nat Hazards Earth Syst Sci*. 2003; 3: 725-732.
- Donat MG, Pardowitz T, Leckebusch GC, Ulbrich U, Burghoff O. High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over Germany. *Nat Hazards Earth Syst Sci*. 2011; 11: 2821-2833.
- Vardoulakis S, Fisher BEA, Pericleous K, Gonzalez-Flesca N. Modelling air quality in street canyons: a review. *Atmos Environ*. 2003; 37: 155-182.
- Hang J, Li Y. Ventilation strategy and air change rates in idealized high-rise compact urban areas. *Build Environ*. 2010; 45: 2754-2767.
- Palma JMLM, Castro FA, Ribeiro LF, Rodrigues AH, Pinto AP. Linear and nonlinear models in wind resource assessment and wind turbine micro-siting in complex terrain. *J Wind Eng Ind Aerod*. 2008; 96: 2308-2326.
- Uchida T, Ohya Y. Micro-siting technique for wind turbine generators by using large-eddy simulation. *J Wind Eng Ind Aerod*. 2008; 96: 2121-2138.
- Wan C, Wang J, Yang G, Li X, Zhang X. Optimal micro-siting of wind turbines by genetic algorithms based on improved wind and turbine models. *Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference Shanghai, P.R. China*, 2009.
- Bhumralkar CM, Mancuso RL, Ludwig FL, Renne DS. A practical and economic method for estimating wind characteristics at potential wind energy conversion sites. *Sol Energy*. 1980; 25: 55-65.
- IPCC (Intergovernmental Panel on Climate Change). *Renewable Energy Sources and Climate Change Mitigation. Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. 2012.
- Carta JA, Ramirez P, Velazquez S. A review of wind speed probability distributions used in wind energy analysis. Case studies in the Canary Islands. *Renew Sust Energ Rev*. 2009; 13: 933-955.
- Morgan EC, Lackner M, Vogel RM, Baise LG. Probability distributions for offshore wind speeds. *Energy Convers Manage*. 2011; 52: 15-26.
- Rehman S, Halawan TO, Husain T. Weibull parameters for wind speed distribution in Saudi Arabia. *Sol Energy*. 1994; 53: 473-479.
- Lun IYF, Lan JC. A study of Weibull parameters using long-term wind observations. *Renew Energ*. 2000; 20: 145-153.
- Seguro JV, Lambert TW. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *J Wind Eng Ind Aerod*. 2000; 85: 75-84.
- Sulaiman MY, Akaak AM, Wahab MA, Zakaria A, Sulaiman ZA, Suradi J. Wind characteristics of Oman. *Energy*. 2002; 27: 35-46.
- Ramirez P, Carta JA. Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study. *Energy Convers Manage*. 2005; 46: 2419-2438.
- Toure S. Investigations on the eigen-coordinates method for the 2-parameter Weibull distribution of wind speed. *Renew Energ*. 2005; 30: 511-521.
- Cellura M, Cirricione G, Marvuglia A, Miraoui A. Wind speed spatial estimation for energy planning in Sicily: Introduction and statistical analysis. *Renew Energ*. 2008; 33: 1237-1250.
- Tar K. Some statistical characteristics of monthly average wind speed at various heights. *Renew Sust Energ Rev*. 2008; 12: 1712-1724.
- Zaharim A, Razali AM, Abidin RZ, Sopian K. Fitting of statistical distributions to wind speed data in Malaysia. *Eur J Sci Res*. 2009; 26: 6-12.
- Jaramillo OA, Borja MA. Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case. *Renew Energ*. 2004; 29: 1613-1630.
- Kiss P, Janosi IM. Comprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe. *Energy Convers Manage*. 2008; 49: 2142-2151.
- Zhou J, Erdem E, Li G, Shi J. Comprehensive evaluation of wind speed distribution models: A case study for North Dakota sites. *Energy Convers Manage*. 2010; 51: 1449-1458.
- Celik AN, Kolhe M. Generalized feed-forward based method for wind energy production. *Appl Energy*. 2013; 101: 582-588.
- Tuller SE, Brett AC. The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis. *J Appl Meteorol*. 1984; 23: 124-134.
- CLC (CORINE Land Cover). Umweltbundesamt, DLR-DFD. 2009.
- Davenport AG. The interaction of wind and structures. In: Plate EJ, editor. *Engineering Meteorology*. Amsterdam: Elsevier Scientific. 1982; 527-572.
- Panofsky HA, Dutton JA. *Atmospheric Turbulence*. New York: Wiley Interscience. 1984.
- Blackadar AK. *Turbulence and Diffusion in the Atmosphere*. Berlin: Springer. 1997.
- Wilson JD. Determining a topex score. *Scot Forest*. 1984; 38: 251-256.
- Quine CP, White IMS. The potential of distance-limited topex in the prediction of site windiness. *Forestry*. 1998; 71: 325-332.
- Carta JA, Ramirez P, Bueno C. A joint probability density function of wind speed and direction for wind energy analysis. *Energy Convers Manage*. 2008; 49: 1309-1320.
- Akdag SA, Bagiorgas HS, Mihalakakou G. Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean. *Appl Energy*. 2010; 87: 2566-2573.
- Kollu R, Rayapudi SR, Narasimham SVL, Pakkurthi KM. Mixture probability distribution functions to model wind speed distributions. *J Energy Environ Eng*. 2012; 3: 27.
- Wan H, Xiaolan L, Wang L, Swail VR. Homogenization and trend analysis of

- Canadian near-surface wind speeds. *J Climate*. 2010; 23: 1209-1225.
45. Lindenberg J, Mengelkamp H-T, Rosenhagen G. Representativity of near surface wind measurements from coastal stations at the German Bight. *Meteorol Z*. 2012; 21: 99-106.
 46. Powell MD, Houston SH, Reinhold TA. Hurricane Andrew's landfall in South Florida. Part I: Standardizing measurements for documentation of surface wind fields. *Weather Forecast*. 1996; 11: 304-328.
 47. Thomas BR, Kent EC, Swail VR. Methods to homogenize wind speeds from ships and buoys. *Int J Climatol*. 2005; 25: 979-995.
 48. Pryor SC, Barthelmie RJ, Young DT, Takle ES, Arritt RW, Flory D, et al. Wind speed trends over the contiguous United States. *J Geophys Res*. 2009; 114, D14105.
 49. Breiman L. Bagging predictors. *Mach Learn*. 1996; 24: 123-140.
 50. Breiman L. Random Forests. *Mach Learn*. 2001; 45: 5-32.
 51. Dean CB, Nielsen JD. Generalized linear mixed models: a review and some extensions. *Lifetime Data Anal*. 2007; 13: 497-512.
 52. Wang XL, Feng Y. RHTestsV4 User Manual. Climate Research Division Atmospheric Science and Technology Directorate Science and Technology Branch, Environment Canada Toronto, Ontario, Canada. 2013.
 53. Wang XL. Accounting for autocorrelation in detecting mean-shifts in climate data series using the penalized maximal t or F test. *J Appl Meteorol Clim*. 2008a; 47: 2423-2444.
 54. Wang XL. Penalized maximal F-test for detecting undocumented mean shifts without trend-change. *J Atmos Ocean Tech*. 2008b; 25: 368-384.
 55. Wang XL, Chen H, Wu Y, Feng Y, Pu Q. New techniques for detection and adjustment of shifts in daily precipitation data series. *J Appl Meteorol Clim*. 2010; 49: 2416-2436.
 56. Touma JS. Dependence of the wind profile power law on stability for various locations. *JAPCA J Air Waste Ma*. 1977; 27: 863-866.
 57. Dee DP, Uppala SM, Simmons AJ, Berrisford P, Polia P, Kobayashi S, et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q J Roy Meteor Soc*. 2011; 137: 553-597.
 58. Smits A, Klein Tank AMG, Können GP. Trends in storminess over the Netherlands, 1962-2002. *Int J Climatol*. 2005; 25: 1331-1344.
 59. Pryor SC, Barthelmie RJ. Long-term trends in near-surface flow over the Baltic. *Int J Climatol*. 2003; 23: 271-289.
 60. Choi JS, Hong S, Chi SB, Lee HB, Park CK, Kim HW, et al. Probability distribution for the shear strength of seafloor sediment in the KR5 area for the development of manganese nodule miner. *Ocean Eng*. 2011; 38: 2033-2041.
 61. Poje D, Cividini B. Assessment of wind energy potential in Croatia. *Sol Energy*. 1988; 41: 543-554.
 62. Dorvlo ASS. Estimating wind speed distributions. *Energ Convers Manage*. 2002; 43: 2311-2318.
 63. Soukissian T. Use of multi-parameter distributions for offshore wind speed modeling: The Johnson SB distribution. *Appl. Energ*. 2013; 111: 982-1000.
 64. Anderson TW, Darling DA. Asymptotic theory of certain "Goodness of Fit" criteria based on stochastic processes. *Ann Math Stat*. 1952; 23: 193-212.
 65. Houghton JC. Birth of a parent: the Wakeby distribution for modeling flood flows. *Water Resour Res*. 1978; 14: 1105-1109.
 66. Guttman NB, Hosking JRM, Wallis JR. Regional precipitation quantile values for the continental United States computed from L-moments. *J Climate*. 1993; 6: 2326-2340.
 67. Park J-S, Jung H-S, Kim R-S, Oh J-H. Modelling summer extreme rainfall over the Korean peninsula using Wakeby distribution. *Int J Climatol*. 2001; 21: 1371-1384.
 68. Su B, Kundzewicz ZW, Jiang T. Simulation of extreme precipitation over the Yangtze River Basin using Wakeby distribution. *Theor Appl Climatol*. 2009; 96: 209-219.
 69. Öztekin T. Wakeby distribution for representing annual extreme and partial duration rainfall series. *Meteorol Appl*. 2007; 14: 381-387.
 70. Öztekin T. Estimation of the parameters of Wakeby distribution by a numerical least squares method and applying it to the annual peak flows of Turkish Rivers. *Water Resour Manag*. 2011; 25: 1299-1313.
 71. Masters FJ, Vickery PJ, Bacon P, Rappaport EN. Toward objective standardized intensity estimates from surface wind speed observations. *B Am Meteorol Soc*. 2010; 91: 1665-1681.
 72. Cherkassky V, Ma Y. Another look at statistical learning theory and regularization. *Neural Netw*. 2009; 22: 958-969.
 73. vanHeijst D, Potharst R, van Wezel M. A support system for predicting eBay end prices. *Decis Support Syst*. 2008; 44: 970-982.
 74. Friedman J. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001; 29: 1189-1232.
 75. Belsley DA, Kuh E, Welsh RE. *Regression Diagnostics*. New York: John Wiley & Sons. 1980.
 76. Willmott CJ, Matsuura K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res*. 2005; 30: 70-92.
 77. Hyndman RJ, Kohler AB. Another look at measures of forecast accuracy. *Int J Forecasting*. 2006; 22: 679-688.
 78. Johnson NL. Systems of frequency curves generated by methods of translation. *Biometrika*. 1949; 36.
 79. Gerth WP, Christoffer J. *Windkarten von Deutschland*: *Meteorol Z*. 1994; 3: 67-77.
 80. Etienne C, Lehmann A, Goyette S, Lopez-Moreno J-I, Beniston M. Spatial predictions of extreme wind speeds over Switzerland using generalized additive models. *J Appl Meteorol Climatol*. 2010; 49: 1956-1970.
 81. Walter A, Keuler K, Jacob D, Knoche R, Block A, Kotlarski S, et al. A high resolution reference data set of German wind velocity 1951-2001 and comparison with regional climate model results. *Meteorol Z*. 2006; 15: 585-596.
 82. Ettoumi YF, Sauvageot H, Adane A-E-H. Statistical bivariate modelling of wind using first-order Markov chain and Weibull distribution. *Renew Energ*. 2003; 28: 1787-1802.