High Spatial Resolution Simulation of Annual Wind Energy Yield Using Near-Surface Wind Speed Time Series

Christopher Jung

Environmental Meteorology, Albert-Ludwigs-University of Freiburg, Werthmannstrasse 10, Freiburg D-79085, Germany; christopher.jung@mail.unr.uni-freiburg.de; Tel.: +49-761-203-6822

Academic Editor: Frede Blaabjerg
Received: 15 February 2016; Accepted: 22 April 2016; Published: 6 May 2016

Abstract: In this paper a methodology is presented that can be used to model the annual wind energy yield ($AEY_{mod}$) on a high spatial resolution ($50 \, m \times 50 \, m$) grid based on long-term (1979–2010) near-surface wind speed ($U_S$) time series measured at 58 stations of the German Weather Service (DWD). The study area for which $AEY_{mod}$ is quantified is the German federal state of Baden-Wuerttemberg. Comparability of the wind speed time series was ensured by gap filling, homogenization and detrending. The $U_S$ values were extrapolated to the height 100 m ($U_{100m,emp}$) above ground level (AGL) by the Hellman power law. All $U_{100m,emp}$ time series were then converted to empirical cumulative distribution functions (CDF$_{emp}$). 67 theoretical cumulative distribution functions (CDF) were fitted to all CDF$_{emp}$ and their goodness of fit (GoF) was evaluated. It turned out that the five-parameter Wakeby distribution (WK5) is universally applicable in the study area. Prior to the least squares boosting (LSBoost)-based modeling of WK5 parameters, 92 predictor variables were obtained from: (i) a digital terrain model (DTM), (ii) the European Centre for Medium-Range Weather Forecasts re-analysis (ERA)-Interim reanalysis wind speed data available at the 850 hPa pressure level ($U_{850hPa}$), and (iii) the Coordination of Information on the Environment (CORINE) Land Cover (CLC) data. On the basis of predictor importance ($PI$) and the evaluation of model accuracy, the combination of predictor variables that provides the best discrimination between $U_{100m,emp}$ and the modeled wind speed at 100 m AGL ($U_{100m,mod}$), was identified. Results from relative $PI$-evaluation demonstrate that the most important predictor variables are relative elevation ($\Phi$) and topographic exposure ($\tau$) in the main wind direction. Since all WK5 parameters are available, any manufacturer power curve can easily be applied to quantify $AEY_{mod}$.

Keywords: annual wind energy yield ($AEY$); Wakeby distribution (WK5); least squares boosting (LSBoost); predictor importance ($PI$); wind speed extrapolation

1. Introduction

The world’s energy supply is facing multiple challenges. The depletion of conventional fuels is unavoidable [1,2], greenhouse gas emissions from the burning of fossil fuels most significantly contributes to global warming [3,4] and the emissions of air pollutants affect human health [3,5]. Although nuclear energy production enables the reduction of carbon dioxide (CO$_2$) emissions [6], nuclear power plants bear great short- and long-term risk of accidents [7]. In order to reduce and avoid negative impacts of the current use of energy resources on the environment and human health, alternative forms of energy utilization must be found.

Renewable energies provide a clean, environmentally friendly and health-compatible alternative to fossil energies and nuclear energy [2,5]. One major renewable energy resource is the kinetic energy contained in the atmosphere, commonly known as wind energy. The potential for wind energy...
utilization to play a key role in the future global energy mix is enormous. Wind energy could supply more than 40 times [8] the annual global electricity consumption. Consequently, the wind power generation capacity of the world is growing constantly with an average annual rate of about 30% over the last decade [2]. In the European Union, directive 2009/28/EC [9] aims to cover 20% of the primary energy demand by renewable energies in 2020, including wind energy. The leading wind power producer in Europe is Germany. Germany aims to supply at least 30% of the energy consumption in 2020 by renewable energies [4]. Yet, in some German federal states the utilization of wind energy is still far from being exhausted. For instance, the ministry of Environment, Climate Protection and the Energy Sector of the southwestern German federal state of Baden-Wuerttemberg plans to increase the share of wind energy in the energy mix from ~1% in 2015 to 10% in 2020 [10]. In order to achieve this political target, up to 1200 new wind turbines with an average output power between 2.5 MW and 3.0 MW must be installed in a period of only five years [10].

The first step in the onshore assessment of potential wind turbine sites is to quantify the site-specific atmospheric wind energy resource at the wind turbine hub height (~80–100 m) [11]. The wind resource is predetermined by the large-scale atmospheric circulation and modified by characteristics of surface roughness [12] and terrain [13]. As a result, the local wind resource can vary significantly over short distances [8]. In contrast to this, ground-based measurements of long-term wind speed at the landscape level are rare and only available for heights near the surface (10 m above ground level (AGL)). Because of the high spatiotemporal variability of the local wind resource [14,15], the low number of available near-surface wind speed measurement sites alone often limits the detailed assessment of the site-specific wind resource.

To overcome the problem of the low number of wind speed measurements and the strong influence of surface and terrain characteristics on the local wind resource, one option is highly resolved statistical modeling of wind speed at hub height. However, mapping of average wind speed alone is insufficient [16], since not only the central tendencies of wind speed distributions determine the wind resource. Therefore, fitting an appropriate theoretical wind speed distribution to empirical wind speed distributions is crucial [17]. Which theoretical distribution fits empirical wind speed distributions best is currently under discussion [18,19].

Due to the limited availability of wind speed measurement sites in Southwest Germany, a region with highly complex topography and mosaic-like land cover pattern, the goals of this study are (i) the quantification of the annual wind energy yield (\(AEY\)) on a high spatial resolution grid and (ii) the identification of the most important factors influencing the local wind resource.

2. Materials and Methods

2.1. Study Area and Wind Speed Measurements

The study area is the German federal state of Baden-Wuerttemberg (Figure 1). The low mountain ranges Black Forest (length ~150 km, width ~30–50 km, highest elevations >1400 m) and Swabian Alb (length ~180 km, width ~35 km, highest elevations >1000 m) are the most complex topographical features with the strongest impact on the wind resource over the study area [20]. The top of the Feldberg (1493 m) is the highest elevation in the study area. Approximately 38% (13,700 km\(^2\)) of the study area is covered with forests [21]. More details about land cover and topographical features in the study area are summarized in [20].

The wind speed database used in this study consists of time series of the daily mean wind speed measured from 1 January 1979 to 31 December 2010 at 58 meteorological stations by the German Weather Service (DWD). The height \((h_S)\) of wind speed measurements varies between 3 m AGL (stations Bad Wildbad-Sommerberg, Isny) and 48 m AGL (station Karlsruhe). Data preparation included gap filling, testing for homogeneity and detrending according to [20].

The median wind speed near the surface values \((U_{s})\) vary in the range 0.5 m/s (station Triberg) to 7.5 m/s (station Feldberg) (Table 1). To extend the database, four measurement stations located in the
bordering federal states Hesse and Bavaria were included in this study. Out of the 58 wind speed time series, 48 time series were put into a parameterization dataset (DS1). The remaining 10 time series, for which the original length was less than 10 years, belong to the validation dataset (DS2).

Figure 1. The study area Baden-Wuerttemberg in Southwest Germany and locations of German Weather Service (DWD) stations. Dots indicate parameterization dataset (DS1) stations; stars indicate validation dataset (DS2) stations.

Table 1. List of DWD stations and corresponding data features. DS1 stations are indicated by identification numbers (ID) 1–48; DS2 stations are indicated by ID values 101–110. \( \bar{U}_s \): median wind speed near the surface values; \( h_s \): height.

<table>
<thead>
<tr>
<th>ID</th>
<th>Station</th>
<th>( \bar{U}_s )</th>
<th>( h_s )</th>
<th>ID</th>
<th>Station</th>
<th>( \bar{U}_s )</th>
<th>( h_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Albstadt-Onstmettingen</td>
<td>1.7</td>
<td>10</td>
<td>30</td>
<td>Öhringen</td>
<td>2.3</td>
<td>16</td>
</tr>
<tr>
<td>2</td>
<td>Bad Säckingen</td>
<td>0.9</td>
<td>10</td>
<td>31</td>
<td>Schluchsee</td>
<td>1.6</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>Bad Wildbad-Sommerberg</td>
<td>0.7</td>
<td>3</td>
<td>32</td>
<td>Schönberg</td>
<td>0.6</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Baiersbronn-Obertal</td>
<td>0.9</td>
<td>10</td>
<td>33</td>
<td>Schwäbisch-Gmünd</td>
<td>0.7</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Beerfelden</td>
<td>1.7</td>
<td>10</td>
<td>34</td>
<td>Sipplingen</td>
<td>2.5</td>
<td>16</td>
</tr>
<tr>
<td>6</td>
<td>Dobel</td>
<td>2.3</td>
<td>10</td>
<td>35</td>
<td>Stimpfach-Weiptershofen</td>
<td>2.1</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>Dogern</td>
<td>1.7</td>
<td>10</td>
<td>36</td>
<td>Stötten</td>
<td>4.1</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>Donaueschingen</td>
<td>2.5</td>
<td>10</td>
<td>37</td>
<td>Stuttgart (Schnarrenberg)</td>
<td>2.5</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>Enzklösterle</td>
<td>1.0</td>
<td>10</td>
<td>38</td>
<td>Stuttgart-Echterdingen</td>
<td>2.5</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>Eschbach</td>
<td>2.4</td>
<td>10</td>
<td>39</td>
<td>Titisee</td>
<td>0.7</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>Feldberg</td>
<td>7.5</td>
<td>19</td>
<td>40</td>
<td>Trübberg</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>12</td>
<td>Freiburg</td>
<td>2.4</td>
<td>12</td>
<td>41</td>
<td>Uffenheim</td>
<td>1.3</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>Freudenstadt</td>
<td>3.7</td>
<td>34</td>
<td>42</td>
<td>Ulm</td>
<td>2.2</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>Friedrichshafen</td>
<td>3.1</td>
<td>10</td>
<td>43</td>
<td>Ulm-Wilhelmsburg</td>
<td>2.7</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>Gailingen</td>
<td>1.6</td>
<td>12</td>
<td>44</td>
<td>Waldachtal-Lützenhardt</td>
<td>1.2</td>
<td>6</td>
</tr>
<tr>
<td>16</td>
<td>Hinterzarten</td>
<td>1.2</td>
<td>6</td>
<td>45</td>
<td>Waldsee, Bad-Reute</td>
<td>2.2</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>Höchenschwand</td>
<td>1.2</td>
<td>6</td>
<td>46</td>
<td>Wallrumb</td>
<td>2.9</td>
<td>10</td>
</tr>
<tr>
<td>18</td>
<td>Hornisgründe</td>
<td>5.9</td>
<td>10</td>
<td>47</td>
<td>Weingarten</td>
<td>1.9</td>
<td>12</td>
</tr>
<tr>
<td>19</td>
<td>Isny</td>
<td>2.1</td>
<td>3</td>
<td>48</td>
<td>Würzburg</td>
<td>2.8</td>
<td>12</td>
</tr>
<tr>
<td>20</td>
<td>Kandern-Gupf</td>
<td>2.0</td>
<td>10</td>
<td>101</td>
<td>Bad Dürrheim</td>
<td>1.3</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>Karlsruhe</td>
<td>3.3</td>
<td>48</td>
<td>102</td>
<td>Bad Herrenalb</td>
<td>0.6</td>
<td>10</td>
</tr>
<tr>
<td>22</td>
<td>Klippeneck</td>
<td>3.9</td>
<td>16</td>
<td>103</td>
<td>Mülheim</td>
<td>1.3</td>
<td>6</td>
</tr>
<tr>
<td>23</td>
<td>Königsfeld</td>
<td>1.0</td>
<td>6</td>
<td>104</td>
<td>Neuhausen ob Eck</td>
<td>2.5</td>
<td>10</td>
</tr>
<tr>
<td>24</td>
<td>Konstanz</td>
<td>1.7</td>
<td>17</td>
<td>105</td>
<td>Pforzheim-Ispringen</td>
<td>2.6</td>
<td>12</td>
</tr>
<tr>
<td>25</td>
<td>Lahr</td>
<td>2.3</td>
<td>10</td>
<td>106</td>
<td>Söllingen</td>
<td>2.6</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>Laupheim</td>
<td>2.5</td>
<td>10</td>
<td>107</td>
<td>Stockach-Espasingen</td>
<td>1.5</td>
<td>12</td>
</tr>
<tr>
<td>27</td>
<td>Leipheim</td>
<td>2.3</td>
<td>10</td>
<td>108</td>
<td>Stuttgart-Stadt</td>
<td>1.7</td>
<td>26</td>
</tr>
<tr>
<td>28</td>
<td>Mannheim</td>
<td>2.7</td>
<td>22</td>
<td>109</td>
<td>Todtmoos</td>
<td>1.1</td>
<td>10</td>
</tr>
<tr>
<td>29</td>
<td>Münstertal</td>
<td>1.3</td>
<td>10</td>
<td>110</td>
<td>Weilheim-Bierbronnens</td>
<td>2.6</td>
<td>10</td>
</tr>
</tbody>
</table>
2.2. Wind Speed Extrapolation

All wind speed near the surface \((U_S)\) time series were extrapolated to 100 m AGL using the Hellman power law \([22–24]\). It was demonstrated by \([11]\) that the power law performs well compared to similar wind speed extrapolation methods. According to \([22]\), the accuracy of the power law increases when stratification effects and the influence of the wind speed are considered. Therefore, the Hellmann exponent \((E)\) was computed on a daily basis.

As has previously been done by \([20,25]\), daily mean wind speed at the 850 hPa pressure level \((U_{850hPa})\) and the height of the 850 hPa pressure level AGL \(h_{850hPa}\), both available from the European Centre for Medium-Range Weather Forecast \([26]\), were used to calculate daily, station-specific \(E\)-values:

\[
E = \frac{\ln \left( \frac{U_{850hPa}}{U_s} \right)}{\ln \left( \frac{h_{850hPa}}{h_s} \right)} \tag{1}
\]

After the \(E\)-values were determined, daily, station-specific \(U_S\)-values were extrapolated to 100 m AGL yielding \(U_{100m,emp}\):

\[
U_{100m,emp} = U_s \times \left( \frac{100m}{h_s} \right)^E \tag{2}
\]

2.3. Probability Distribution Fitting

Prior to the probability distribution fitting, \(U_{100m,emp}\) time series were transformed to empirical cumulative distribution functions (CDF\(_{emp}\)). Afterwards, 67 CDF were fitted to each CDF\(_{emp}\). The goodness of fit (GoF) of each CDF was quantified by calculating the coefficient of determination \((R^2)\) from probability plots \([19,27]\) and the Kolmogorov-Smirnov statistic \((D)\) \([28–30]\) to the fits. The \(D\)-values were obtained by measuring the largest vertical difference between CDF and CDF\(_{emp}\). The transformation of time series, fitting and GoF evaluation were done by EasyFit software (Version 5.5, MathWave Technologies, Dnepropetrovsk, Ukraine) and Matlab® Software Optimization Toolbox (Release 2015a; The Math Works Inc., Natick, MA, USA).

According to \(D\)- and \(R^2\)-value evaluation, which will be presented in detail in the results section, the five-parameter Wakeby distribution (WK5) \([31]\) is clearly the best-fitting distribution. It can be defined by its quantile function \([20,25,31,32]\):

\[
U_{100m,distr}(F) = \varepsilon + \frac{\alpha}{\beta} \times \left[ 1 - (1 - F)^{\frac{1}{\beta}} \right] - \frac{\gamma}{\delta} \times \left[ 1 - (1 - F)^{-\frac{1}{\delta}} \right] \tag{3}
\]

where \(F\) is the cumulative probability with \(U_{100m,distr}(F)\) being the associated wind speed value. The four parameters \(\alpha, \beta, \gamma, \delta\) are distribution parameters and the fifth parameter, \(\varepsilon\), is the location parameter. WK5 can be interpreted as a mixed distribution \([33]\) consisting of a left and right part \([31,32]\). This enables WK5 to reproduce shapes of wind speed distributions that other distributions cannot reproduce \([25,31]\).

2.4. Predictor Variable Building

A total number of 92 predictor variables \((50 m \times 50 m)\) covering the study area were built by using the ArcGIS® 10.2 software (Esri, Redlands, CA, USA). All predictor variables originate from a digital terrain model (DTM), CORINE Land Cover (CLC) data \([34]\) or ERA-Interim reanalysis \(U_{850hPa}\) \([26]\).

The DTM was used to map \(\Phi, \tau\) \([35,36]\), curvature, aspect and slope. The \(\Phi\)-values were calculated by subtracting the mean elevation of an outer circle around each grid point from the grid point-specific elevation. Five different \(\Phi\) variants with outer-circle radii of 250 m, 500 m, 1000 m \((\Phi_{1000m})\), 2500 m \((\Phi_{2500m})\) and 5000 m \((\Phi_{5000m})\) were created. The \(\tau\)-maps were built for eight main compass directions (northeast \(22.5^\circ–67.4^\circ\), east \(67.5^\circ–112.4^\circ\), southeast \(112.5^\circ–157.4^\circ\), south \(157.5^\circ–202.4^\circ\), southwest \(202.5^\circ–247.4^\circ\), west
(247.5°–292.4°), northwest (292.5°–337.4°), north(337.5°–22.4°)) at 200 m radius intervals. This was done by summing angles up to a distance limited to 1000 m. Curvature, aspect and slope were calculated by using the Spatial Analyst Toolbox in ArcGIS.

Roughness length (\(z_0\)) was derived from CLC data with an original spatial resolution of 100 m × 100 m. Roughness length values were assigned to land cover types according to [20] yielding the local roughness length \(z_{0.1}\). Additionally, “effective” roughness length values \(z_{0,\text{eff}}\) for the eight main compass directions were calculated. This was done for four different radii around each grid point (100 m, 200 m, 300 m, 400 m). In the end, all \(z_0\)-values were interpolated to 50 m × 50 m resolution grids.

\(U_{850\text{hPa}}\) data (0.125° × 0.125° resolution) were included into model building because it represents large-scale airflow undisturbed by the surface [37]. The 0.01, 0.30, 0.50, 0.75 and 0.99 percentiles of \(U_{850\text{hPa}}\) time series covering the period from 01 January 1979 to 31 December 2010 were calculated \(U_{850\text{hPa},0.01}, U_{850\text{hPa},0.30}, U_{850\text{hPa},0.50}, U_{850\text{hPa},0.75}\) and \(U_{850\text{hPa},0.99}\) and mapped in ArcGIS®. A spline interpolation was applied to convert the \(U_{850\text{hPa}}\) layers to 50 m × 50 m resolution grids.

### 2.5. Wakeby Parameter Estimation and Modeling

The procedure applied to obtain the Wakeby parameters at every grid point in the study area comprised the following work steps: (1) estimating the Wakeby parameters of every CDF\(_{\text{emp}}\) based on L-moments [38,39]; (2) analyzing the obtained Wakeby parameters and identifying common characteristics of all distributions ; and (3) modeling target variables \(Y\) that enable the calculation of all WK5 parameters at every grid point in the study area. To make the WK5 parameter modeling more robust, the WK5 parameters estimated by L-moments were modeled indirectly according to [20,25].

Analyzing the estimated distributions led to the following parameter modeling and calculation approach: First, the estimated left-hand tail of WK5 \((Y_L)\), which is represented by \(\alpha\), \(\beta\) and \(\varepsilon\), was modeled:

\[
Y_L = \varepsilon + \frac{10}{\beta} \times \left[1 - (1 - 0.25)^\beta\right]
\]

The estimated location parameter \(\varepsilon\), which represents the lower bound of the distribution, was directly modeled. Because the L-moment-based WK5 parameter estimation showed that \(\alpha = 10\) at nearly all stations, it was set to this value. The use of a fixed \(\alpha\)-value enabled the subsequent calculation of \(\beta\).

Since \(Y_L\) affected WK5 parameter estimation up to \(F = 0.25\), exactly as described by [31,32], the percentiles \(F = 0.30\) (\(Y_{R1}\)), \(F = 0.50\) (\(Y_{R2}\)), \(F = 0.75\) (\(Y_{R3}\)) and \(F = 0.99\) (\(Y_{R4}\)) were modeled to build the right-hand tail of WK5 \((Y_R)\). A system of non-linear equations was solved at every grid point yielding \(\gamma\) and \(\delta\):

\[
\begin{align*}
\frac{\gamma}{\delta} \times \left[1 - (1 - 0.30)^{-\delta}\right] + [Y_{R1} - Y_L] &= 0 \\
\frac{\gamma}{\delta} \times \left[1 - (1 - 0.50)^{-\delta}\right] + [Y_{R2} - Y_L] &= 0 \\
\frac{\gamma}{\delta} \times \left[1 - (1 - 0.75)^{-\delta}\right] + [Y_{R3} - Y_L] &= 0 \\
\frac{\gamma}{\delta} \times \left[1 - (1 - 0.99)^{-\delta}\right] + [Y_{R4} - Y_L] &= 0
\end{align*}
\]

In order to calculate \(U_{100\text{m},\text{mod}}\), \(Y_L\) and \(Y_R\) were recombined yielding WK5 with modeled parameters (WK5\(_{\text{mod}}\)).

All \(Y\) were computed for every grid point by least squares boosting (LSBoost) [40]. This was done by using the Ensemble Learning algorithm LSBoost implemented in the Matlab®Software Statistics Toolbox (Release 2015a; The Math Works Inc.). LSBoost is basically a sequence of simple regression trees, which are called weak learners (\(B\)). The objective of LSBoost is to minimize the mean squared error (MSE) between \(Y\) and the aggregated prediction of the weak learners (\(Y_{\text{pred}}\)). In the beginning, the median of the target variables \((\tilde{Y})\) is calculated. Afterwards, multiple regression trees \(B_1, \ldots, B_m\)
are combined in a weighted manner [41] to improve model accuracy. The individual regression trees are a function of selected predictor variables (X):

\[ Y_{\text{pred}}(X) = \hat{Y}(X) + \nu \sum_{m=1}^{M} p_m \times B_m(X) \]  

with \( p_m \) being the weight for model \( m \), \( M \) is the total number of weak learners, and \( \nu \) with \( 0 < \nu \leq 1 \) is the learning rate [20,41].

The predictor variable selection process comprised several steps. First, the most appropriate length of outer-circle radii for \( \tau \) and \( z_{0,\text{eff}} \) were determined by the correlation coefficient \( r \) between \( \tau \) respectively \( z_{0,\text{eff}} \) and \( Y \). Secondly, the importance of the remaining predictor variables was evaluated by predictor importance (\( PI \)) which quantifies the relative contribution of individual predictor variables to the model output [21]. The \( PI \)-values were determined by summing up changes in MSE due to splits on every predictor and dividing the sum by the number of branch nodes. All predictor variables with \( PI = 0.00 \) were sorted out.

After \( PI \)-evaluation, combinations of predictor variables were tested for their predictive power. Starting with one predictor variable, further predictor variables were added to the model and kept when the model accuracy measures \( R^2 \), mean error (ME), mean absolute error (MAE), MSE and mean absolute percentage error (MAPE) improved [42–44]. For model parameterization, DS1 data were used. Model validation was done with both DS1 and DS2 data.

Multicollinearity among the predictor variables was investigated by assessing the variance inflation and the condition index in combination with variance decomposition proportions according to [45].

2.6. Annual Wind Energy Yield Estimation

The relationship between wind speed and the electrical power output (\( P \)) of wind turbines is typically established by a power curve [46]. Power curve values are developed from field measurements and can be used for studies involving energy calculations [47]. There are three important points characterizing a typical power curve (Figure 2): (1) at the cut-in speed the wind turbine starts to generate usable power; (2) after exceeding the rated output speed the maximum output power (rated power) is generated; and (3) after exceeding the cut-out speed turbines cease power generation and shut down [46]. A standard 2.5 MW power curve [48] for onshore wind power plants was applied to calculate the \( AEY \).

![Figure 2. Power curve used to calculate empirical annual wind energy yield (\( AEY_{\text{emp}} \)) and modeled annual wind energy yield (\( AEY_{\text{mod}} \)) depending on wind speed in 100 m above ground level (AGL) (\( U_{100m} \)).](image-url)
The discrete $P$-values from the manufacturer power curve were interpolated by a spline to obtain a continuous power curve. The basic attributes of the applied power curve are: cut-in speed $U_{100m} = 3.0 \text{ m/s}$; cut-out speed $U_{100m} = 25.0 \text{ m/s}$; rated output speed $U_{100m} = 13.0 \text{ m/s}$ and; rated output power $P = 2580 \text{ kW}$. The empirical annual wind energy yield ($AEY_{\text{emp}}$) was calculated for each station in DS1 and DS2 following [49]:

$$AEY_{\text{emp}} = \left( \sum_{i=1}^{N} P(U_{100m,\text{emp},i})/Z_{1} \right)$$

with $N = 11,688$ being the total number of days in the investigation period and the number of years in the investigation period ($Z_{1}$).

The average electrical power output ($\overline{P}$) was calculated according to [19,50]:

$$\overline{P} = \int_{0}^{\infty} P(U_{100m,\text{mod}}) \times f(U_{100m,\text{mod}}) \, dU_{100m,\text{mod}}$$

The above equation describes the electrical power produced at each wind speed class multiplied by the probability of the specified wind speed class and integrated over all possible wind speed classes [50] with $f(U_{100m,\text{mod}})$ being the probability density of $U_{100m,\text{mod}}$. After $\overline{P}$ is calculated modeled annual wind energy yield ($AEY_{\text{mod}}$) can be computed by multiplying $\overline{P}$ with the respective number of days per year ($Z_{2}$):

$$AEY_{\text{mod}} = \overline{P} \times Z_{2}$$

2.7. Summary of the Methodology

The methodology for the quantification of $AEY$ in the study area is summarized in Figure 3. The basic steps are:

1. Extrapolation of near-surface wind speed time series to hub height;
2. Identification of a theoretical distribution that is capable of reproducing various shapes of empirical wind speed distributions;
3. Modeling the estimated parameters of the identified theoretical distribution, based on large-scale airflow, surface roughness and topographic features;
4. Mapping of distribution parameters in the study area; and
5. Calculation of the $AEY$ using a wind turbine-specific power curve.

![Figure 3. Schematic representation of the workflow applied to obtain annual wind energy yield (AEY).](image-url)
3. Results and Discussion

3.1. Distribution Fitting

According to results from the $D$-evaluation, WK5 fits 23 CDF$_{\text{emp}}$ best. As can be seen in Table 2, the $D$-value averaged over all stations for WK5 (0.02) is lower than the average $D$-value of all other theoretical distributions. Another well-fitting distribution is the four-parameter Johnson SB distribution ($D = 0.03$). The best fitting three-parameter distribution is the inverse Gaussian distribution ($D = 0.03$). In general, the performance of theoretical distributions defined by three or more parameters is better than the performance of two- and one-parameter distributions. In the case of eight theoretical distributions (Johnson SU, Log-Gamma, Log-Pearson 3, Nakagami, Pareto, Reciprocal, Phased Bi-Exponential, Phased Bi-Weibull) no fit to CDF$_{\text{emp}}$ could be achieved and therefore the parameter estimation procedure failed.

Table 2. Distributions ranked (RK) by Kolmogorov-Smirnov statistic ($D$)-values with their number of parameters (NP). $D$- and coefficient of determination ($R^2$)-values are averages over all meteorological stations.

<table>
<thead>
<tr>
<th>RK</th>
<th>Distribution</th>
<th>$D$</th>
<th>$R^2$</th>
<th>NP</th>
<th>RK</th>
<th>Distribution</th>
<th>$D$</th>
<th>$R^2$</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wakeby</td>
<td>0.02</td>
<td>0.9992</td>
<td>5</td>
<td>35</td>
<td>Weibull</td>
<td>0.10</td>
<td>0.9768</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>Johnson SB</td>
<td>0.03</td>
<td>0.9991</td>
<td>4</td>
<td>36</td>
<td>Pert</td>
<td>0.11</td>
<td>0.9732</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Inv. Gaussian</td>
<td>0.03</td>
<td>0.9981</td>
<td>3</td>
<td>37</td>
<td>Rayleigh</td>
<td>0.12</td>
<td>0.9721</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Pearson 6</td>
<td>0.03</td>
<td>0.9984</td>
<td>4</td>
<td>38</td>
<td>Erlang</td>
<td>0.12</td>
<td>0.9866</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Pearson 6</td>
<td>0.03</td>
<td>0.9983</td>
<td>3</td>
<td>39</td>
<td>Normal</td>
<td>0.13</td>
<td>0.9492</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>Lognormal</td>
<td>0.03</td>
<td>0.9982</td>
<td>3</td>
<td>40</td>
<td>Rice</td>
<td>0.13</td>
<td>0.9653</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Dagum</td>
<td>0.03</td>
<td>0.9978</td>
<td>3</td>
<td>41</td>
<td>Logistic</td>
<td>0.13</td>
<td>0.9511</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Fatigue Life</td>
<td>0.03</td>
<td>0.9978</td>
<td>3</td>
<td>42</td>
<td>Hypersecant</td>
<td>0.14</td>
<td>0.9497</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>Gen. Extreme</td>
<td>0.04</td>
<td>0.9975</td>
<td>3</td>
<td>43</td>
<td>Uniform</td>
<td>0.14</td>
<td>0.9351</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>Burr</td>
<td>0.04</td>
<td>0.9974</td>
<td>4</td>
<td>44</td>
<td>Cauchy</td>
<td>0.15</td>
<td>0.9845</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>Log-Logistic</td>
<td>0.04</td>
<td>0.9971</td>
<td>3</td>
<td>45</td>
<td>Erlang</td>
<td>0.15</td>
<td>0.9909</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>Burr</td>
<td>0.04</td>
<td>0.9971</td>
<td>3</td>
<td>46</td>
<td>Chi-Squared</td>
<td>0.16</td>
<td>0.9908</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>Lognormal</td>
<td>0.04</td>
<td>0.9973</td>
<td>2</td>
<td>47</td>
<td>Error</td>
<td>0.16</td>
<td>0.9437</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>Bimodal Weibull</td>
<td>0.04</td>
<td>0.9985</td>
<td>5</td>
<td>48</td>
<td>Laplace</td>
<td>0.17</td>
<td>0.9412</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>Fatigue Life</td>
<td>0.04</td>
<td>0.9970</td>
<td>2</td>
<td>49</td>
<td>Chi-Squared</td>
<td>0.19</td>
<td>0.9920</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>Inv. Gaussian</td>
<td>0.04</td>
<td>0.9974</td>
<td>3</td>
<td>50</td>
<td>Gumbel Min</td>
<td>0.20</td>
<td>0.8976</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>Pearson 5</td>
<td>0.04</td>
<td>0.9958</td>
<td>3</td>
<td>51</td>
<td>Exponential</td>
<td>0.23</td>
<td>0.9863</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>Bimodal Normal</td>
<td>0.04</td>
<td>0.9956</td>
<td>5</td>
<td>52</td>
<td>Exponential</td>
<td>0.27</td>
<td>0.9833</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>Gen. Pareto</td>
<td>0.04</td>
<td>0.9976</td>
<td>3</td>
<td>53</td>
<td>Pareto 2</td>
<td>0.28</td>
<td>0.9841</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>Gen. Gamma</td>
<td>0.04</td>
<td>0.9954</td>
<td>4</td>
<td>54</td>
<td>Triangular</td>
<td>0.31</td>
<td>0.9132</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>Dagum</td>
<td>0.04</td>
<td>0.9957</td>
<td>4</td>
<td>55</td>
<td>Power Func.</td>
<td>0.31</td>
<td>0.9138</td>
<td>3</td>
</tr>
<tr>
<td>22</td>
<td>Pearson 5</td>
<td>0.05</td>
<td>0.9954</td>
<td>2</td>
<td>56</td>
<td>Levy</td>
<td>0.36</td>
<td>0.9777</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>Gen. Logistic</td>
<td>0.05</td>
<td>0.9947</td>
<td>3</td>
<td>57</td>
<td>Levy</td>
<td>0.39</td>
<td>0.9769</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>Log-Logistic</td>
<td>0.05</td>
<td>0.9952</td>
<td>2</td>
<td>58</td>
<td>Error Func.</td>
<td>0.70</td>
<td>0.9103</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>Gamma</td>
<td>0.06</td>
<td>0.9931</td>
<td>3</td>
<td>59</td>
<td>Student’s t</td>
<td>0.82</td>
<td>0.7991</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>Beta</td>
<td>0.06</td>
<td>0.9922</td>
<td>4</td>
<td>-</td>
<td>Johnson SU</td>
<td>No Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Gen. Gamma</td>
<td>0.06</td>
<td>0.9904</td>
<td>3</td>
<td>-</td>
<td>Log-Gamma</td>
<td>No Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Gamma</td>
<td>0.06</td>
<td>0.9910</td>
<td>2</td>
<td>-</td>
<td>Log-Pearson 3</td>
<td>No Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Frechet</td>
<td>0.06</td>
<td>0.9909</td>
<td>3</td>
<td>-</td>
<td>Nakagami</td>
<td>No Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Gumbel Max</td>
<td>0.07</td>
<td>0.9879</td>
<td>2</td>
<td>-</td>
<td>Pareto</td>
<td>No Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Weibull</td>
<td>0.07</td>
<td>0.9851</td>
<td>3</td>
<td>-</td>
<td>Reciprocal</td>
<td>No Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Frechet</td>
<td>0.07</td>
<td>0.9892</td>
<td>2</td>
<td>-</td>
<td>Phased Bi-Exp.</td>
<td>No Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Kumuaraswamy</td>
<td>0.09</td>
<td>0.9805</td>
<td>4</td>
<td>-</td>
<td>Phased Bi-Wei.</td>
<td>No Fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Rayleigh</td>
<td>0.10</td>
<td>0.9776</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A widely used theoretical distribution applied to empirical wind speed distributions is the two-parameter Weibull distribution [30,51–58]. However, in this study, the fit of the Weibull distribution is poor ($D = 0.10$) compared to many other theoretical distributions. These results are in accordance with similar studies where the GoF of various theoretical distributions to empirical
distributions was compared [59–61]. The Weibull distribution is not even the best-fitting two-parameter distribution, which is the lognormal distribution. The best GoF of a one-parameter distribution was achieved by the also widely used Rayleigh distribution [62,63]. However, compared to many distributions defined by more parameters the GoF of the Rayleigh distribution was rather poor ($D = 0.10$). An explanation for the poor fit of distributions with less than three parameters might be that their capacity for reproducing irregular shapes of empirical distributions is limited. Irregularly shaped empirical wind speed distributions often result from complex topography [64].

The evaluation of averaged $R^2$-values confirms results of the $D$-values evaluation. The best-fitting distribution is WK5 ($R^2 = 0.9992$), followed by Johnson SB ($R^2 = 0.9991$).

The superior fit of WK5 is in accordance to GoF measures of empirical near-surface (10 m AGL) wind speed distributions in the study area [20]. Based on the results presented in this study it is concluded that WK5 is a universal wind speed distribution for the study area.

### 3.2. Predictor Variable Selection and Importance

The screening of $r$-values showed that the most appropriate length of outer-circle radius was 1000 m for $\tau$ and 200 m for $z_{0\text{eff}}$. Table 3 lists the predictor variables used for all six least squares boosting models (LSBM) and their relative impact to the model outputs. From the large set of predictor variables, predictor selection finally reduced their number to 14.

Table 3. Relative importance of predictor variables used for final least squares boosting models (LSBM) in percent. The top three important predictor variables are highlighted in red.

<table>
<thead>
<tr>
<th>ID</th>
<th>Predictor variable</th>
<th>Symbol</th>
<th>$Y_L$</th>
<th>$\varepsilon$</th>
<th>$Y_{R1}$</th>
<th>$Y_{R2}$</th>
<th>$Y_{R3}$</th>
<th>$Y_{R4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wind speed at 850 hPa level ($F = 0.75$)</td>
<td>$U_{850\text{hPa},0.75}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>21.4</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>Wind speed at 850 hPa level ($F = 0.99$)</td>
<td>$U_{850\text{hPa},0.99}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>14.8</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Roughness length, local</td>
<td>$z_{0,\text{L}}$</td>
<td>1.9</td>
<td>0.1</td>
<td>0.4</td>
<td>-</td>
<td>-</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>Roughness length, effective, W</td>
<td>$z_{0\text{eff},W}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>Roughness length, effective, SW</td>
<td>$z_{0\text{eff},SW}$</td>
<td>-</td>
<td>-</td>
<td>0.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>Roughness length, effective, S</td>
<td>$z_{0\text{eff},S}$</td>
<td>-</td>
<td>-</td>
<td>7.8</td>
<td>2.2</td>
<td>1.5</td>
<td>6.2</td>
</tr>
<tr>
<td>7</td>
<td>Roughness length, effective, N</td>
<td>$z_{0\text{eff},N}$</td>
<td>-</td>
<td>-</td>
<td>1.1</td>
<td>-</td>
<td>6.9</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>Topographic exposure, NW</td>
<td>$\tau_{\text{NW}}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>9.8</td>
<td>-</td>
<td>5.2</td>
</tr>
<tr>
<td>9</td>
<td>Topographic exposure, W</td>
<td>$\tau_{W}$</td>
<td>-</td>
<td>21.4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>20.8</td>
</tr>
<tr>
<td>10</td>
<td>Topographic exposure, SW</td>
<td>$\tau_{\text{SW}}$</td>
<td>-</td>
<td>24.4</td>
<td>23.5</td>
<td>9.6</td>
<td>-</td>
<td>20.8</td>
</tr>
<tr>
<td>11</td>
<td>Topographic exposure, SE</td>
<td>$\tau_{\text{SE}}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.4</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>Relative elevation, 1000 m</td>
<td>$\Phi_{1000\text{m}}$</td>
<td>73.7</td>
<td>55.0</td>
<td>-</td>
<td>-</td>
<td>49.4</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>Relative elevation, 2500 m</td>
<td>$\Phi_{2500\text{m}}$</td>
<td>-</td>
<td>-</td>
<td>80.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>Relative elevation, 5000 m</td>
<td>$\Phi_{5000\text{m}}$</td>
<td>-</td>
<td>-</td>
<td>78.8</td>
<td>-</td>
<td>49.1</td>
<td>-</td>
</tr>
</tbody>
</table>

The main wind directions in the study area are west and southwest. It is therefore reasonable that southwesterly and westerly oriented $\tau$- and $z_{0\text{eff}}$-predictor variables have a distinct impact to the model outputs. The highest $PI$-values for any roughness length predictor variable are found for the LSBM output $Y_{R2}$ and the western sector ($PI = 9.2\%$). However, the $PI$-value for $Y_{R1}$ and the southwestern sector is relatively low ($PI = 0.6\%$). The topographic exposure for the southwestern sector, respectively the western sector, is one of the most important predictor variables for modeling $\varepsilon$, $Y_L$, $Y_{R3}$ and $Y_{R4}$.

It is important to note that $U_{850\text{hPa}}$ was not used to model the left-hand tail of WK5, which represents $U_{100\text{m},\text{mod}}$-values. Low wind speed values mostly occur when the atmosphere is stably stratified [22]. Thus, the influence of $U_{850\text{hPa}}$ on $U_{100\text{m},\text{mod}}$ is rather small.

When modeling $Y_{R3}$ and $Y_{R4}$, the large-scale airflow becomes more important $PI = [21.4\%, 14.8\%]$ because high $U_{100\text{m},\text{mod}}$-values usually occur when the atmosphere is neutrally stratified [22].

Results from $PI$-evaluation indicate the fundamental role of relative elevation in wind turbine site assessment. The high $PI$-values for $\Phi$ indicate the great importance of $\Phi$ for model outputs. The highest $PI$-value is $80.5\%$ for $\Phi_{2500\text{m}}$ when modeling $Y_{R1}$. In contrast, the absolute elevation ($\psi$)
was never used as predictor variable. This is reasonable because sites with high \( \psi \)-values are not necessarily exposed to high wind speeds.

3.3. Wind Speed Mapping

Median \( U_{100\text{m,mod}} \)-values (\( \tilde{U}_{100\text{m,mod}} \)) are shown in Figure 4. In large parts (75%) of the study area, \( \tilde{U}_{100\text{m,mod}} \)-values are in the range between 3.0 m/s and 4.0 m/s. In only 0.2% of the study area, \( \tilde{U}_{100\text{m,mod}} \)-values are above 4.9 m/s. Due to the complex topography, high and low \( \tilde{U}_{100\text{m,mod}} \)-values can occur within small distances (<500 m). For example, in the Black Forest, which is characterized by narrow, forested valleys, \( \tilde{U}_{100\text{m,mod}} \)-values are very low. However, there are many exposed mountaintops in close proximity to these valleys where \( \tilde{U}_{100\text{m,mod}} \)-values are high. Beside narrow, forested valleys, lowest \( \tilde{U}_{100\text{m,mod}} \)-values (<3.1 m/s) occur in large cities. In the entire study area the effect of topographic exposure on the modeling results is evident by predominantly higher \( \tilde{U}_{100\text{m,mod}} \)-values at sites exposed to the West and Southwest.

![Figure 4. Median of modeled wind speed in 100 m AGL (\( \tilde{U}_{100\text{m,mod}} \)) in the study area. The legend values indicate highest class values.](image)

3.4. Annual Wind Energy Yield

In Figure 5, the empirical AEY per wind speed class (\( \Delta \text{AEY}_{\text{emp}} \)), the modeled AEY per wind speed class (\( \Delta \text{AEY}_{\text{mod}} \)), the probability density distributions of WK5\(_{\text{mod}} \) and the probability density distributions fitted to \( U_S \)-values (\( U_S,\text{distr} \)) are presented as a function of wind speed classes (intervals of 0.1 m/s) for the stations Hornisgrinde (Figure 5a) and Laupheim (Figure 5b).

It is clear that percentiles \( F = [0.30–0.99] \) from the right-hand tail of WK5\(_{\text{mod}} \) contribute more to AEY and are thus more important for the total amount of AEY\(_{\text{mod}} \). In Laupheim the mode of \( U_{100\text{mod}} \) is 2.3 m/s, whereas highest \( \Delta \text{AEY}_{\text{mod}} \) is obtained at 8.0–8.1 m/s. Even at the top of the Hornisgrinde, which is one of the windiest places in the study area, the \( U_{100\text{mod}} \) mode value at 4.2 m/s is clearly lower than the wind speed class assigned to the highest \( \Delta \text{AEY}_{\text{mod}} \)-value (9.0–9.1 m/s). Overall, the \( \Delta \text{AEY}_{\text{mod}} \)-curves fit \( \Delta \text{AEY}_{\text{emp}} \)-values obtained for both stations well.
Figure 5. $\Delta AEY_{\text{emp}}$ and $\Delta AEY_{\text{mod}}$ as a function of wind speed classes ($U$) (intervals of 0.1 m/s) as well as the probability of wind speed near the surface fitted to a WK5 distribution ($U_{850\text{hPa}}$) and modeled wind speed in 100 m AGL ($U_{100\text{m,mod}}$)-classes for stations: (a) Hornisgrinde; and (b) Laupheim.

The map of $AEY_{\text{mod}}$ (Figure 6) shows similar patterns like the $U_{100\text{m,mod}}$-map. By applying the power curve to $U_{100\text{m,mod}}$, the mean $AEY_{\text{mod}}$-value in the study area is 3.4 GWh/yr. The highest $AEY_{\text{mod}}$-value (13.6 GWh/yr) occurs at the top of the Feldberg. Only in 3% of the study area is $AEY_{\text{mod}}$ higher than 5.0 GWh/yr. In 31% of the study area $AEY_{\text{mod}}$ is lower than 3.0 GWh/yr with a tendency towards lower $AEY_{\text{mod}}$-values in the southeast, which is mainly due to low $U_{850\text{hPa}}$-values in this part over the study area. In contrast, generally higher $AEY_{\text{mod}}$-values were calculated in the northeast where $U_{850\text{hPa}}$-values are highest at the landscape level. The spatial $AEY_{\text{mod}}$-pattern indicates that the local wind resource is mainly determined by terrain features and surface roughness.

Figure 6. $AEY_{\text{mod}}$ in the study area. The legend values indicate highest class values.
This is underlined by the map extract shown in Figure 7. In the topographically structured Black Forest region, it appears that highest and lowest \( \text{AEY}_{\text{mod}} \) values occur over horizontal distances shorter than 500 m. This finding is in good accordance to a previous study regarding gust speed in the same area [25]. The main wind direction can be inferred from highest \( \text{AEY}_{\text{mod}} \) values over southwest-facing slopes.

![Figure 7. \( \text{AEY}_{\text{mod}} \) in the Southern Black Forest region. The legend values indicate highest class values.](image)

Figure 8 shows \( r \)-values which were calculated between \( \text{AEY}_{\text{mod}} \) and various predictor variables. The \( r \)-values confirm the results of the \( PI \)-evaluation. The highest and lowest \( r \)-values are obtained for the most important predictor variables. The highest absolute \( r \)-values are \((r = 1 - 0.59)\) for \( \tau_{\text{SW}} \) and \((r = 10.58)\) for \( \Phi_{2500m} \). The correlation between \( \psi \) and \( \text{AEY}_{\text{mod}} \) is relatively weak \((r = 0.08)\). This is due to the fact that some highly elevated Black Forest valleys are sites with the lowest \( \text{AEY}_{\text{mod}} \) values.

The correlation between \( U_{850hPa,0.75} \) and \( \text{AEY}_{\text{mod}} \) \((r = 0.08)\) is also relatively low since the influence of the large-scale airflow on \( \text{AEY}_{\text{mod}} \) is superimposed by influences of local terrain features and surface roughness. All correlations are highly significant with significance values \((p) p \leq 0.0000\).
The exemplary functional relationships between classes of four important predictor variables and \( A\dot{E}Y_{mod} \) are shown in Figure 9. The variability of \( A\dot{E}Y_{mod} \)-values as a function of \( U_{850hPa,0.75} \) (Figure 9a) is lower than the variability of the other displayed predictor variables. This is interpreted to mean that the variability of \( U_{850hPa,0.75} \) is of minor importance for explaining the spatial \( A\dot{E}Y_{mod} \)-patterns in the study area. Due to their high roughness, \( A\dot{E}Y_{mod} \) is lower over forests and cities (Figure 9b). Areas that are exposed to the southwest (\( \tau_{SW} < 2^\circ \)) show higher \( A\dot{E}Y_{mod} \)-values (median: 3.9 GWh/yr) than sheltered areas (\( \tau_{SW} > 18^\circ \)) (median: 1.5 GWh/yr) (Figure 9c). The strongest functional relationship is between \( \Phi_{2500m} \) and \( A\dot{E}Y_{mod} \) (Figure 9d). The assigned median \( A\dot{E}Y_{mod} \)-values increase from 1.6 GWh/yr at \( \Phi_{2500m} < -150 \) m to 4.7 GWh/yr at \( \Phi_{2500m} > 150 \) m.

**Figure 8.** Correlation coefficient (\( r \))-values calculated between \( A\dot{E}Y_{mod} \) and various predictor variables.

**Figure 9.** Boxplots of \( A\dot{E}Y_{mod} \) as a function of: (a) 0.75 percentile of the wind speed at the 850 hPa pressure level (\( U_{850hPa,0.75} \)); (b) local roughness length (\( z_0 \)); (c) topographic exposure in southwest direction (\( \tau_{SW} \)); and (d) relative elevation with outer circle radius of 2500 m (\( \Phi_{2500m} \)). Boxplot style: red lines indicate medians, boxes indicate interquartile ranges, whiskers indicate 1.5-times interquartile ranges. The legend values indicate highest class values.
3.5. Model Validation

The MAPE-values indicate that $U_{100m,\text{mod}}$ was simulated accurately (Table 4). They are always below 6% for both DS1 and DS2. The $R^2$-values are mostly 0.97 for DS1 percentiles and about 0.95 for DS2 percentiles. The largest downward bias is $ME = -0.30$ m/s for $F = 0.99$.

Table 4. Performance measures coefficient of determination ($R^2$), mean error ($ME$), mean absolute error ($MAE$), mean squared error ($MSE$) and mean absolute percentage error ($MAPE$) calculated from the comparison of empirical and modeled cumulative probabilities ($F$) associated with $U_{100m}$-time series included in DS1 and DS2.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$F$</th>
<th>$R^2$</th>
<th>$ME$ (m/s)</th>
<th>$MAE$ (m/s)</th>
<th>$MSE$ (m/s)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS1</td>
<td>0.10</td>
<td>0.97</td>
<td>0.05</td>
<td>0.11</td>
<td>0.02</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.97</td>
<td>0.04</td>
<td>0.11</td>
<td>0.02</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.97</td>
<td>0.01</td>
<td>0.12</td>
<td>0.02</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>0.97</td>
<td>0.00</td>
<td>0.13</td>
<td>0.03</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.97</td>
<td>0.00</td>
<td>0.14</td>
<td>0.03</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.97</td>
<td>-0.01</td>
<td>0.16</td>
<td>0.04</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.97</td>
<td>-0.01</td>
<td>0.20</td>
<td>0.06</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.97</td>
<td>-0.01</td>
<td>0.25</td>
<td>0.09</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>0.97</td>
<td>-0.09</td>
<td>0.33</td>
<td>0.17</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>0.98</td>
<td>0.00</td>
<td>0.38</td>
<td>0.23</td>
<td>3.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$F$</th>
<th>$R^2$</th>
<th>$ME$ (m/s)</th>
<th>$MAE$ (m/s)</th>
<th>$MSE$ (m/s)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DS2</td>
<td>0.10</td>
<td>0.95</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.01</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>0.20</td>
<td>0.95</td>
<td>-0.03</td>
<td>0.10</td>
<td>0.01</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.96</td>
<td>-0.06</td>
<td>0.10</td>
<td>0.02</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>0.40</td>
<td>0.97</td>
<td>-0.07</td>
<td>0.12</td>
<td>0.02</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>0.50</td>
<td>0.97</td>
<td>-0.07</td>
<td>0.14</td>
<td>0.02</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>0.60</td>
<td>0.96</td>
<td>-0.08</td>
<td>0.17</td>
<td>0.04</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.95</td>
<td>-0.10</td>
<td>0.21</td>
<td>0.07</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.95</td>
<td>-0.12</td>
<td>0.25</td>
<td>0.10</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>0.95</td>
<td>-0.24</td>
<td>0.36</td>
<td>0.17</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>0.94</td>
<td>-0.30</td>
<td>0.46</td>
<td>0.42</td>
<td>4.0</td>
</tr>
</tbody>
</table>

The model performance for DS2 is only marginally worse than for DS1. This indicates the portability of LSBM to other data sets.

Performance measures from the comparison of modeled cumulative distribution functions ($CDF_{\text{mod}}$) with $CDF_{\text{emp}}$ associated with $U_{100m}$-time series included in DS2 are shown in Table 5.

Table 5. Performance measures from the comparison of modeled cumulative distribution function ($CDF_{\text{mod}}$) with empirical cumulative distribution function ($CDF_{\text{emp}}$) associated with $U_{100m}$ time series included in DS2.

<table>
<thead>
<tr>
<th>Station</th>
<th>$D$</th>
<th>$R^2$</th>
<th>Station</th>
<th>$D$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad Dürrheim</td>
<td>0.05</td>
<td>0.9977</td>
<td>Söllingen</td>
<td>0.06</td>
<td>0.9973</td>
</tr>
<tr>
<td>Bad Herrenalb</td>
<td>0.05</td>
<td>0.9973</td>
<td>Stockach-Espasingen</td>
<td>0.08</td>
<td>0.9918</td>
</tr>
<tr>
<td>Müllheim</td>
<td>0.05</td>
<td>0.9982</td>
<td>Stuttgart-Stadt</td>
<td>0.09</td>
<td>0.9924</td>
</tr>
<tr>
<td>Neuhausen ob Eck</td>
<td>0.02</td>
<td>1.0000</td>
<td>Todtmoos</td>
<td>0.07</td>
<td>0.9968</td>
</tr>
<tr>
<td>Pforzheim-Ispringen</td>
<td>0.04</td>
<td>0.9991</td>
<td>Weilheim-Bierbronnien</td>
<td>0.06</td>
<td>0.9978</td>
</tr>
</tbody>
</table>

It appears that the GoF measures for modeled WK5 parameters are better than for many statistical distributions that were directly fitted to $CDF_{\text{emp}}$ (compare Table 2).
In Figure 10, $A_{\text{EY}}^{\text{emp}}$ is plotted against $A_{\text{EY}}^{\text{mod}}$. Related performance measures for DS1 (Figure 10a) and DS2 (Figure 10b) are $R^2 = \{0.98, 0.97\}$, $ME = \{-0.16 \text{ GWh/yr}, -0.23 \text{ GWh/yr}\}$, $MAE = \{0.32 \text{ GWh/yr}, 0.31 \text{ GWh/yr}\}$, $MSE = \{0.16 \text{ GWh/yr}, 0.13 \text{ GWh/yr}\}$ and $MAPE = \{10.0\%, 17.1\%\}$. Thus, it can be concluded that the calculated $A_{\text{EY}}^{\text{emp}}$-values were modeled with sufficient accuracy.

![Figure 10](image-url)

Figure 10. $A_{\text{EY}}^{\text{emp}}$ plotted against $A_{\text{EY}}^{\text{mod}}$ for: (a) DS1; and (b) DS2.

4. Conclusions

A methodology is presented that allows assessing the statistical $A_{\text{EY}}$ on a high spatial resolution (50 m × 50 m) grid in an area with mosaic-like land cover pattern and complex topography. It was found that highest and lowest $A_{\text{EY}}$ occurs in highly textured terrain within very small distances (<500 m). The results of this study therefore emphasize the need to assess $A_{\text{EY}}$ at very small spatial scales. This is demonstrated in particular by the great importance of the predictor variables relative elevation and topographic exposure in the main wind direction.

Since the methodology allows for the calculation of all WK5 parameters, the $A_{\text{EY}}$ for any manufacturer power curve can be estimated. The methodology is easily portable to other heights above ground level as well as to other study areas. The only requirements for the portability are the availability of the following: (i) near-surface wind speed time series as measured in meteorological networks; (ii) a DTM; (iii) a land cover data set; and (iv) wind speed data not influenced by local topography or land use.

The proposed modeling approach is a useful first step in the exploration of the most appropriate wind turbine sites based on the local wind resource. The produced model outputs and maps are valuable starting points for further in-depth wind turbine site assessment.

**Conflicts of Interest:** The author declares no conflict of interest sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

**Nomenclature**

- $A_{\text{EY}}$ Annual wind energy yield
- $A_{\text{EY}}^{\text{emp}}$ Empirical annual wind energy yield
- $A_{\text{EY}}^{\text{mod}}$ Modeled annual wind energy yield
- $B$ Regression tree
- $B_m$ Regression tree $m$
- $D$ Kolmogorov-Smirnov statistic
- $E$ Hellmann exponent
- $f$ Probability density
- $F$ Cumulative probability
\( h_{850\text{hPa}} \)  
Height above ground level of the 850 hPa pressure level

\( h_s \)  
Measurement height of \( U_s \)

\( M \)  
Total number of weak learners

\( \text{MAE} \)  
Mean absolute error

\( \text{MAPE} \)  
Mean absolute percentage error

\( ME \)  
Mean error

\( \text{MSE} \)  
Mean squared error

\( N \)  
Number of days in the investigation period

\( p \)  
Significance value

\( P \)  
Electrical power output

\( \overline{P} \)  
Average electrical power output

\( PI \)  
Relative predictor importance

\( p_m \)  
Weight for model \( m \)

\( r \)  
Correlation coefficient

\( R^2 \)  
Coefficient of determination

\( U \)  
Wind speed

\( U_{100\text{m}} \)  
Wind speed in 100 m AGL

\( U_{100\text{m},\text{distr}} \)  
Wind speed in 100 m AGL fitted to a WK5 distribution

\( U_{100\text{m},\text{emp}} \)  
Empirical wind speed in 100 m AGL

\( U_{100\text{m},\text{mod}} \)  
Modeled wind speed in 100 m AGL

\( \tilde{U}_{100\text{m},\text{mod}} \)  
Median of modeled wind speed in 100 m AGL

\( U_{850\text{hPa}} \)  
Wind speed at the 850 hPa pressure level

\( U_{850\text{hPa},0.01} \)  
1st percentile of the wind speed at the 850 hPa pressure level

\( U_{850\text{hPa},0.30} \)  
30.th percentile of the wind speed at the 850 hPa pressure level

\( U_{850\text{hPa},0.50} \)  
50.th percentile of the wind speed at the 850 hPa pressure level

\( U_{850\text{hPa},0.75} \)  
75.th percentile of the wind speed at the 850 hPa pressure level

\( U_{850\text{hPa},0.99} \)  
99.th percentile of the wind speed at the 850 hPa pressure level

\( U_s \)  
Wind speed near the surface

\( \tilde{U}_s \)  
Median wind speed near the surface

\( U_s,\text{distr} \)  
Wind speed near the surface fitted to a WK5 distribution

\( v \)  
Learning rate

\( X \)  
Predictor variables

\( Y \)  
Target variables

\( \tilde{Y} \)  
Median of target variables

\( Y_L \)  
Left-hand side of WK5

\( Y_{\text{pred}} \)  
Aggregated prediction of predictor variables

\( Y_R \)  
Right-hand side of WK5

\( Y_{R1} \)  
30.th percentile of WK5

\( Y_{R2} \)  
50.th percentile of WK5

\( Y_{R3} \)  
75.th percentile of WK5

\( Y_{R4} \)  
99.th percentile of WK5

\( z_0 \)  
Roughness length

\( z_{0\text{eff}} \)  
Effective roughness length

\( z_{0\text{eff},E} \)  
Effective roughness length in east direction

\( z_0 \)  
Local roughness length

\( z_{0\text{eff},N} \)  
Effective roughness length in north direction

\( z_{0\text{eff},NE} \)  
Effective roughness length in northeast direction

\( z_{0\text{eff},NW} \)  
Effective roughness length in northwest direction

\( z_{0\text{eff},S} \)  
Effective roughness length in south direction

\( z_{0\text{eff},SE} \)  
Effective roughness length in southeast direction

\( z_{0\text{eff},SW} \)  
Effective roughness length in southwest direction

\( z_{0\text{eff},W} \)  
Effective roughness length in west direction
Number of years in the investigation period
Number of days per year
Parameter of WK5
Parameter of WK5
Parameter of WK5
Parameter of WK5
Empirical annual wind energy yield per wind speed class
Modeled annual wind energy yield per wind speed class
Location parameter of WK5
Topographic exposure
Topographic exposure in east direction
Topographic exposure in north direction
Topographic exposure in northeast direction
Topographic exposure in northwest direction
Topographic exposure in south direction
Topographic exposure in southeast direction
Topographic exposure in southwest direction
Topographic exposure in west direction
Relative elevation
Relative elevation with outer circle radius of 1000 m
Relative elevation with outer circle radius of 2500 m
Relative elevation with outer circle radius of 5000 m
Absolute elevation
Above ground level
Theoretical cumulative distribution function
Empirical cumulative distribution function
Modeled cumulative distribution function
CORINE Land Cover
Coordination of Information on the Environment
Parameterization dataset
Validation dataset
Digital terrain model
German Weather Service
European Centre for Medium-Range Weather Forecasts re-analysis
Goodness of fit
Identification number
Least squares boosting model
Least squares boosting
Number of parameters
Rank of distribution according to $D$-evaluation
Wakeby distribution
Modeled Wakeby distribution

References


35. Wilson, J.D. Determining a topex score. *Scott. For.* **1984**, *38*, 251–256.


© 2016 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).