Grothe, Oliver; Schnieders, Julius

Working Paper

Spatial Dependence in Wind and Optimal Wind Power Allocation: A Copula Based Analysis

EWI Working Paper, No. 11/05

Provided in Cooperation with:
Institute of Energy Economics at the University of Cologne (EWI)

Suggested Citation: Grothe, Oliver; Schnieders, Julius (2011) : Spatial Dependence in Wind and Optimal Wind Power Allocation: A Copula Based Analysis, EWI Working Paper, No. 11/05, Institute of Energy Economics at the University of Cologne (EWI), Köln

This Version is available at:
https://hdl.handle.net/10419/74402

Institute of Energy Economics
at the University of Cologne

EWI Working Paper, No. 11/05

Spatial dependence in wind and optimal wind power allocation: a copula based analysis

by
Oliver Grothe , Julius Schnieders

May 2011

Spatial dependence in wind and optimal wind power allocation: a copula based analysis

Oliver Grothe*,[a], Julius Schnieders[b]

[a] *University of Cologne, Department of Economic and Social Statistics, Albertus-Magnus-Platz, 50923 Cologne, Germany*
[b] *University of Cologne, Graduate School of Risk Management, Meister-Ekkehart-Strasse 11, 50923 Cologne, Germany*

Abstract

The investment decision on the placement of wind turbines is, neglecting legal formalities, mainly driven by the aim to maximize the expected annual energy production of single turbines. The result is a concentration of wind farms at locations with high average wind speed. While this strategy may be optimal for single investors maximizing their own return on investment, the resulting overall allocation of wind turbines may be unfavorable for energy suppliers and the economy because of large fluctuations in the overall wind power output. This paper investigates to what extent optimal allocation of wind farms in Germany can reduce these fluctuations. We analyze stochastic dependencies of wind speed for a large data set of German on- and offshore weather stations and find that these dependencies turn out to be highly nonlinear but constant over time. Using copula theory we determine the value at risk of energy production for given allocation sets of wind farms and derive optimal allocation plans. We find that the optimized allocation of wind farms may substantially stabilize the overall wind energy supply on daily as well as hourly frequency.

Key words: Wind power, Vine copula, Optimal turbine allocation
JEL classification: C32, C53, Q42, Q47
ISSN: 1862 3808

*Corresponding author
Email addresses: grothe@statistik.uni-koeln.de (Oliver Grothe), schnieders@wiso.uni-koeln.de (Julius Schnieders)

March 5, 2011

1. Introduction

Wind power is one of the world's largest and most accessible sources of renewable energy. However, intermittency presents a barrier for wind power to meet the world demand for electricity. Since wind shows sudden changes, wind power shows a high variability. In this paper, we analyze to what extent the variability may be reduced by optimally located power stations. Our analysis is based on models of wind speeds from different regions of Germany. Since wind speeds and their dependencies are highly non-Gaussian, nonlinear times series models and vine copula constructions are applied. Given the positions of wind farms, the models assess the lower quantiles of the distribution of the overall produced wind power. We maximize these quantiles with respect to the locations of wind turbines which are subject to certain constraints to obtain optimal allocation plans for wind energy production.

The study focuses on German on- and offshore data and identifies the optimal allocation of wind farms across the country. However, the proposed methodology may be applied to other regions as well. We focus on Germany for several reasons. Firstly, Germany incorporates one of the world's largest markets for wind energy. In 2007, nearly 25% of the world-wide wind capacity was installed in Germany (Windpower monthly, January 2008), followed by the USA with about 19%. Since wind energy accounts for over 9% of the whole electricity production in Germany, energy suppliers in this country are especially affected by the variability of wind power so that their need to smooth the wind power supply is crucial.

Although Germany exchanges power with foreign countries we do not include data from other countries in our study. We think, however, that a study similar to ours on a European level would be interesting. This would be of special relevance in the future, when interconnection capacities between countries are further extended and feed-in tariffs and distribution of renewable energies are jointly organized. At the moment, however, we believe that a focus on one country conforms more to the actual situation since the markets are still organized on national levels.

The allocation of wind energy production in Germany is highly imbalanced due to political reasons and feed-in tariffs to encourage the investment in renewable energies. Figure 1 depicts its spatial distribution (the data is provided by the four German transmission system operators (50Hz, Amprion, EnBw, Transpower)). The wind power production is mainly concentrated on the coastline and in eastern Germany. Note that in 2009 (and still) the amount of offshore wind power in Germany is negligible. Over time, production is highly volatile. Figure 2 shows the produced amount of wind power for 2009 on a daily basis. It varies between 1% and 70% of the installed capacity. This paper analyzes how and to what extent an optimal distribution of wind power production could smooth the overall wind power output. The necessary redistribution could be achieved by either installing new wind turbines or by repowering, i.e., replacing turbines by more powerful ones.

The possibility of smoothing wind power by geographical dispersion of wind farms or by interconnecting existing dispersed wind farms is studied

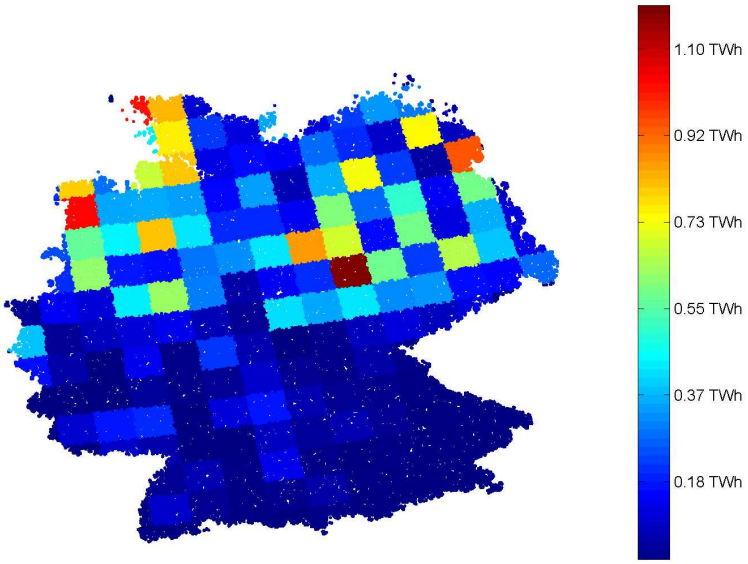

Figure 1: Produced wind energy in Germany in 2009. Dark red regions indicate high values of produced energy, dark blue regions indicate no produced energy. The figure is based on data of the German transmission system operators 50Hz, Amprion, EnBw and Transpower for the year 2009 and visualized by 50000 random points across the coordinates of Germany. The color of each rectangle corresponds to the total amount of wind energy produced in this rectangle in 2009. Apparently, wind energy is produced mainly at the coasts and in eastern Germany.

in several papers. It has a long history going back to Kahn (1979) who was the first to systematically analyze these effects for arrays of wind farms of different sizes. He used wind data from California and found that the reliability of the wind power output improved with the sizes of the arrays. For more recent studies showing that the interconnection of wind farms reduces the variability of their summed output, see among others Katzenstein et al. (2010), Archer and Jacobson (2007), Czisch and Ernst (2001), Giebel (2000)

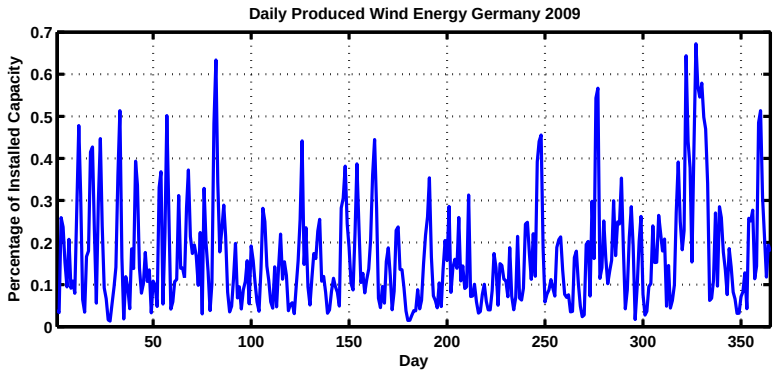

Figure 2: Overall wind energy produced in Germany in 2009 on a daily basis in percentage of installed capacity. The energy produced varies from nearly 70% to 1% of the installed capacity (data is provided by the four German transmission system operators 50Hz, Amprion, EnBw and Transpower).

Milligan and Porter (2005), and Drake and Hubacek (2007).

In particular, Archer and Jacobson (2007) find that at least 33% of the average power output of 19 interconnected wind farms in the central United States are online at a probability level of 12.5%. This is the average outage time (scheduled and unscheduled outages) of coal plants in the United States (see, e.g., Giebel (2000)). In contrast, 0% of the average power output of single farms may be online to this probability level. Drake and Hubacek (2007) go one step further. They analyze the average power and standard deviation of several allocations of capacity among 4 simulated wind farms in the UK to find the allocation with the least amount of wind power variability. They estimate the correlations of the wind speeds at single stations and use mean-variance portfolio theory to find optimal locations for wind power

capacity with respect to the variance of the overall power output.

The present paper analyzes optimal allocations of wind farms across Germany. However, when maximizing the part of the wind energy that may provide a stable baseload, the lower quantiles (value at risk) of the overall energy production are more important than its variance. Therefore, this paper focuses on the allocation of wind turbines by maximizing lower quantiles of the power supply instead of minimizing the variance. While the optimization of the variance only requires estimates of the marginal variances and covariances, the assessment of the quantiles of the overall power output is not trivial. It requires the modeling of the marginal distributions and the entire dependence structure of the wind speeds of the considered regions. The reason for this is the non-normality of the distributions of wind speeds as well as the nonlinearities of their dependence. Whenever multivariate data is not normally distributed, the quantiles of sums of margins may not be calculated from sums of variances and covariances.

Therefore, we apply nonlinear time series models and copula methods in this study. Copula functions capture the complete dependence structure of random variables. They may be applied if the random variables are highly non-normal as it is the case for wind speeds. In our case, the dimensionality of the data set is high (40 dimensions) and the dependency structure is heterogeneous, i.e., the kind of nonlinear pairwise dependence varies between dimensions. In such a setting, most multivariate copulas are unfavorable as they assume homogeneous dependency structures across dimensions. There-

fore, we use multivariate pair copula constructions as originally introduced by Joe (1996). Based on a hierarchical tree of 2-dimensional copulas, they allow for flexible modeling of heterogeneous dependence structures in higher dimensions.

The resulting model enables us to optimize the allocation of wind energy production with respect to certain constraints. The models are on daily frequency, while the results are evaluated on daily and on hourly frequencies. Firstly, we find the overall best allocation of wind energy production with respect to lower quantiles of the overall production. The result enables us to discuss the best case, i.e., to what maximal extent wind energy in Germany may be online for a given probability level. Secondly, we start from the status quo and optimize the wind power allocation when repowering and extending the installed wind capacity in Germany. This enables us to identify regions in which new wind power capacity would be most useful to improve the overall supply stability, i.e., the reliability of wind energy. The results show that repowering and expansion should be forced at the coasts and the offshore regions as well as in some regions in southern Germany and should be suppressed in most parts of eastern Germany.

The remainder of the paper is structured as follows. Section 2 presents the data set of wind speeds and the multivariate time series model which is based on multivariate vine copula constructions. The concept of these copulas is summarized in B. In section 3, the model for the wind data set is used to find optimal allocation and expansion plans for wind turbine positions. Section

4 concludes.

2. Data and wind speed models

In this section, the data sets of wind data used for our study and a time series model based on pair copula functions are presented. The data set (onshore and offshore data) was provided by the German weather service. It consists of daily and hourly means of wind speeds measured at 40 (daily) and 39 (hourly) German onshore and offshore weather stations[1] from 2005-01-01 to 2010-12-31, respectively. For most of the stations (36 stations), in particular the on-shore stations, longer time series are available. When possible we therefore use data beginning from 1980-01-01, but then explicitly indicate this. The offshore stations are Greifswalder Oie, Hallig Hooge, Helgoland, and UFS Deutsche Bucht. The resolution is $\frac{1}{10}$ m/s. Leap days are erased from the sample.

In the daily data, except for the offshore station UFS Deutsche Bucht, missing data are replaced by the means of hourly data from the same days, if at least one hour of data is available. If no hourly data is available, data are replaced by data of the same station and day of a randomly chosen

[1] Aachen, Augsburg, Bamberg, Berlin-Tempelhof, Bremen, Cuxhaven, Dresden-Klotzsche, Duesseldorf, Emden, Erfurt-Weimar, Fichtelberg, Frankfurt/Main, Goerlitz, Greifswalder Oie, Hallig Hooge, Hamburg-Fuhlsbuettel, Hannover, Helgoland, Hof, Hohenpeissenberg, Kahler Asten, Kempten, Konstanz, Leipzig-Halle, Lindenberg, Magdeburg, Meiningen, Neuruppin, Nuernberg, Potsdam, Rostock-Warnemuende, Saarbruecken-Ensheim, Schleswig, Schwerin, Straubing, Stuttgart-Echterdingen, UFS Deutsche Bucht (daily data only), Westermarkelsdorf, Wuerzburg, Zugspitze

year. An average of 0.60% of the data is missing before inserting means of hourly data and 0.28% after inserting hourly data. The longest time period of missing data in the time interval from the year 2005 on is 31 days for the station Greifswalder Oje. The length of the gaps is much smaller for most of the stations. Thee fourth longest gap is 9 days for the stations Emden and Erfurt-Weimar. The average length of missing data is 2.1 days with on average 4.9 gaps per station. For the hourly data, we replace missing data by the mean of data before and after the gap if the gap is not longer than 12 hours. For gaps longer than 12 hours, we use data from the same day of a randomly chosen year to maintain the intraday structure of the wind. On average 0.23% respectively 0.21% of the data is missing before and after inserting the means of adjoining data.

The data of the offshore station UFS Deutsche Bucht, which is located on an unmanned lightvessell in the north sea, is of rather poor quality. On the hourly frequency the amount of missing data in the considered time interval is around 12% and we do not use this station on this frequency. Even on daily frequency, there is a period of 8 weeks without data. Again, we replace missing daily data by the mean of hourly data from the same day, if available. To conserve the dependencies of the data, this time we do not replace the quite long remaining missing parts by data from other years as above, but replace it by daily means of data of NASA's QSCAT satellite. The QSCAT data consist of two measurements a day till 2009-11-23 for the respective area and are provided by Remote Sensing Systems (http://www.remss.com/). For

days where data of UFS are available, the original data and the satellite data fit very well with an UFS mean of 7.98 m/s and a satellite mean of 8.20 m/s (see, e.g., Beaucage et al. (2007) and the references therein for discussions regarding satellite wind data).

Table 2 (Appendix) shows descriptive statistics of the 39 (40) stations as well as geographical altitude of the station and wind detector height above ground. Figure 3 shows histograms of 6 representative stations (daily data). The data of all stations are heavily skewed. The mean values of the stations vary from 2 m/s to more than 8 m/s. The p-values of the Ljung-Box test and Engle's ARCH test are below 10^{-50}, indicating autoregressive structure and heteroskedasticity in the data.

In the next subsection, we present univariate time series models to clean the time series from these effects. The correlation structure of the filtered residuals is further analyzed in subsection 2.2. For hourly data (and other intraday frequencies) we find intertemporal cross correlation of the residuals. These vanish for daily frequencies but prohibit a generalization of the hourly models to the multivariate case. In subsection 2.3, we therefore combine only the daily models into a multivariate model. This is done by modeling the dependence of simultaneous residuals by pair copula constructions.

2.1. Univariate model

We model the univariate time series by a seasonal ARMA model, which was recently proposed for wind speed modeling by Benth and Benth (2010).

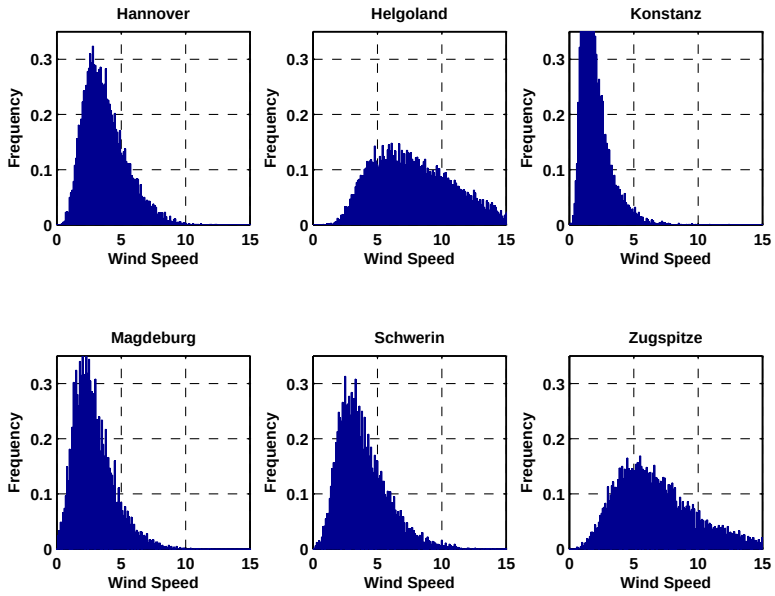

Figure 3: Histograms of the wind speeds in m/s of 6 representative stations for the daily mean of wind speeds for the longer data set starting 1980-01-01. All distributions are highly skewed to the right and the distributions are clearly different.

First, the skewness of the wind speed data (see figure 3) is removed by applying the Box-Cox transformation,

$$X = \frac{Y^{\lambda} - 1}{\lambda},$$

where Y is the time series of wind speed and the parameter λ is estimated by maximum likelihood (see, e.g., Box and Cox (1964)). The resulting time series X_t are modeled by ARMA(p,q) models with a seasonal functions S_t

and seasonal volatilities σ_t to account for heteroskedasticity.

$$X_t = S_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_k \epsilon_{t-j} + \varepsilon_t, \tag{1}$$

with

$$\varepsilon_t = \sigma_t \eta_t.$$

For daily data, where an increment in t corresponds to a day, we use seasonal functions of the form

$$S_t = a_0 + \sum_{k=0}^{1} a_{2k+1} \cos\left(\frac{(2k+2)\pi t}{365}\right) + \sum_{k=0}^{1} a_{2k+2} \sin\left(\frac{(2k+2)\pi t}{365}\right)$$

and

$$\hat{\sigma}_t^2 = c_0 + \sum_{k=1}^{3} c_k \cos\left(\frac{2k\pi t}{365}\right),$$

respectively, where $\hat{\sigma}_t^2$ is the average historical variance. For hourly data (as well as for 3, 4, 6 and 12 hour data), we add further seasonal intraday terms consisting of hourly (respectively 3,4,6,12-hourly) dummies. The models are estimated for each of the time series individually. The resulting standardized residuals η_t for each of the time series pass Engle's ARCH test, i.e., they show no significant heteroscedasticity.

2.2. Correlation analysis of the residuals

The optimal allocation of wind turbines relies on the dependence structure of wind speeds at different locations. We model these, as discussed in the next

subsection, by the dependence of the residuals η_t for the same t but different time series. For the justification of this approach, we have to ensure that the dependencies of concurrent residuals capture the complete dependence of the time series and that these dependencies are constant over time.

For the first point, we look at the pairwise cross-correlations of the empirical innovations η_t, i.e., the correlations of $\eta_t^{\text{Region 1}}$ with $\eta_{t+l}^{\text{Region 2}}$, where the parameter l is an integer-valued lag parameter. Our approach of multivariate modeling is justified, when there is no cross-correlation for lags not equal to zero. Figure 4 shows the results for the pair Augsburg and Bamberg. The results are representative for the results of the other pairs. Shown are cross-correlations for $l = -15..15$ for the residuals of the time series on frequencies of 1 hour, 3 hours, 4 hours, 6 hours, 12 hours and 1 day. For frequencies of 1 day, there is no significant cross-correlation for lags not equal to zero. This justifies a multivariate model of the time series, where the dependence between dimensions is modeled by the dependence of concurrent residuals on a daily frequency as in copula-Garch models. However, on intraday-frequencies, there is significant cross-correlation for lagged residuals and the approach is not adequate since lagged residuals contain much information on the dependence structure which is not captured by the dependence of simultaneous residuals. Intuitively, this may be explained as follows. The innovation η_t corresponds to a random change in the wind speeds on the considered frequencies. Due to the finite velocity of the wind flow over the country, random changes in the wind in one region lead to delayed changes

in other regions, but not to simultaneous ones. On daily frequency the delay is small compared to the time unit, while on intraday frequencies the effect is not negligible.

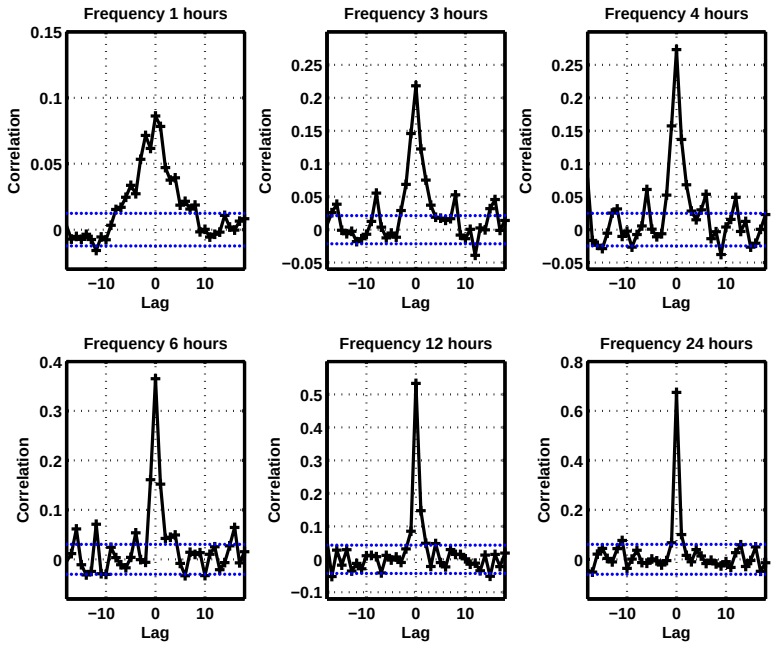

Figure 4: Lagged cross-correlation of the residuals η_t for the pair Augsburg and Bamberg for models on different frequencies of 1 to 24 hour grid. Shown are the values of the correlations and the 95% confidence intervals for the correlations in independent data. The copula approach for the multivariate models is only possible, when the cross-dependence of the residuals is negligible for lags not equal to zero. As shown in the figure, only for daily data this assumption is justified. For hourly data, actually most of the dependence is captured by lagged residuals and even on a 12 hour grid, the cross-correlation is significant.

Furthermore, it is important that the dependence between the wind speeds of different regions is constant over time. Otherwise, the optimal allocation of wind turbines would change over time and it would not be feasible to

re-allocate the wind turbines with the changing dependence. We look at pairwise rank correlation coefficients of wind speed residuals η_t at two locations at a time and use daily data. We choose rank correlation since it is robust to non-Gaussian data. Figure 5 shows 8 representative series of rank correlation coefficients estimated on a backward looking rolling window of 365 days window length on the extended data set beginning in 1980. The shown estimated values of the correlation vary in time but stay between certain levels. They may therefore be assumed to be constant in time. Additionally, the levels of the correlation of the different pairs are different, i.e., different stations show different dependence of the wind speeds.

Figure 6 shows a histogram of the values with all $36(36-1)/2 = 630$ pairwise rank correlations computed on the longer data set beginning in 1980. The lowest rank correlation is 0.04 for the pair Rostock-Warnemuende and Saarbruecken-Ensheim while the largest value is 0.894 for the pair Berlin-Tempelhof and Potsdam. In the second subfigure, the pairwise correlations are plotted versus the geographical distance of these pairs. The distances between two stations are calculated from values of longitude and latitude by the haversine assumption, i.e., assuming a spherical earth. As intuitively expected, the correlation tends to decrease with growing distance. The depicted fit of an exponential model according to $\rho \propto \exp(-distance/D)$ has a decay parameter D of 455 km and an intercept of $\rho = 0.88$ for zero distance. The results are consistent with the studies of Giebel (2000) and Katzenstein et al. (2010). Both studies find such an exponential relationship between correla-

tion of wind speeds and distance with decay parameters between 305 km and 723 km. Giebel (2000) analyzes wind speeds in Europe, while Katzenstein et al. (2010) concentrate on Texas. Note, however, that our study relates to the model residuals η_t, i.e., the not explained changes in wind speeds, while the cited studies use the wind speeds itselfs. The result of decreasing cor-

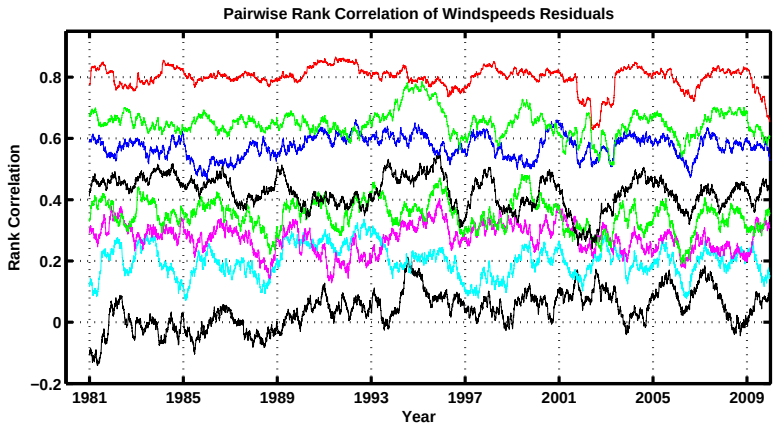

Figure 5: Pairwise rank correlation of the residuals η_t of the wind speeds of the stations (in ascending order) Saarbruecken-Ensheim and Rostock-Warnemuende, Zugspitze and Greifswald, Konstanz and Berlin-Tempelhof, Rostock-Warnemuende and Nuernberg, Stuttgart-Echterdingen and Magdeburg, Straubing and Kahler Asten, Schleswig and Berlin-Tempelhof, Kahler Asten and Frankfurt/Main, Magdeburg and Berlin-Tempelhof. The correlations are computed for daily data with a backward-looking moving window of length 365 days.

relations with growing distance suggests that geographical diversification of wind turbines over large distances results in a lower variance of the overall energy production. However, only in the case of Gaussian distributions and Gaussian dependence structure does the knowledge of the correlation directly allow the calculation of the lower quantiles. In the case of non-Gaussian dis-

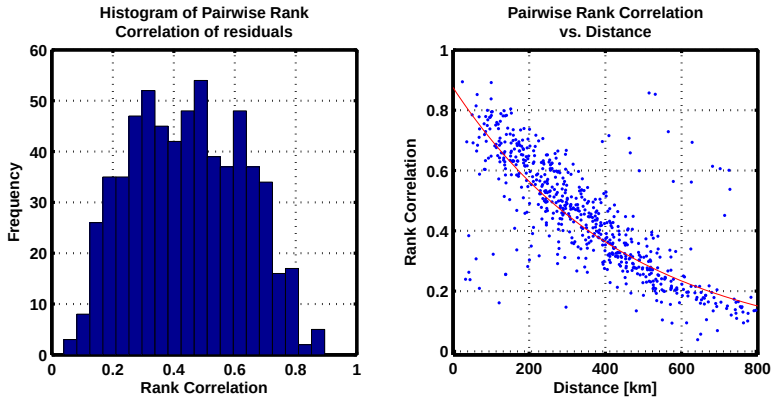

Figure 6: Histogram of the pairwise rank correlation of the residuals η_t of the wind speeds of all considered stations (left) of the longer data set of daily data beginning in 1980 (data of 36 stations). Scatter plot of the pairwise rank correlation of all stations against their geographical distance (right). Fitted is an exponential modell according to $\rho \propto \exp(-distance/D)$ with a decay parameter D of 455 km and an intercept of $\rho = 0.88$ for zero distance.

tributions, minimizing the variance of the energy production may not result in optimal quantiles.

2.3. Multivariate copula model and calibration

Since the dependence of the residuals is highly non-Gaussian, it is not entirely captured by correlation measures. An exact multivariate model, however, is possible if based on copula functions. Copulas are the most general dependence concept for random variables. A short introduction to copulas is given in B.1. Figure 7 shows the dependency function λ for the example of two pairs of stations. Originally introduced in Genest and Rivest (1993) for Archimedean copulas, the λ-function is defined by $\lambda(v) := v - K(v)$, where K is the distribution function of the (empirical) copula. Thus, it

describes the dependence structure of a given data set and can be used to identify the parametric copula family that provides the best fit. The empirical λ-function estimated from the wind data of these pairs is denoted by the black line. The green, red and blue lines indicate the theoretical λ-functions of data with Clayton, Gaussian and Clayton survival copula with adequate parameters. In both cases, the empirical dependence deviates from Gaussian dependence structure, while Clayton dependence structure and Clayton survival structure capture the respective dependencies adequately. This is also confirmed by a formal Goodness-of-Fit test, which is based on the λ-function (see Genest et al., 2009), where the P-values are given in the legend of figure 7 (larger P-values correspond to a better fit than smaller P-values). Thus, the dependence in our data set is non-Gaussian and heterogeneous, i.e., its type is varying across the dimensions. The heterogeneous structure together with the high dimensionality of our problem complicates the finding of an adequate copula function (see B.1 for a discussion). Solutions are pair copula constructions (PCC) which are described in detail in B.2. These constructions consist of multiple bivariate copulas and flexibly capture different pairwise dependence structures between different dimensions of the time series. Analogously to copula-GARCH models, the PCCs are embedded into the model by transforming the residuals η_t to the unit interval and modeling the dependence of these uniformly distributed residuals by PCC.

The calibration of the entire model (univariate time series and dependence structure) to the data set of wind speeds involves both the estimation

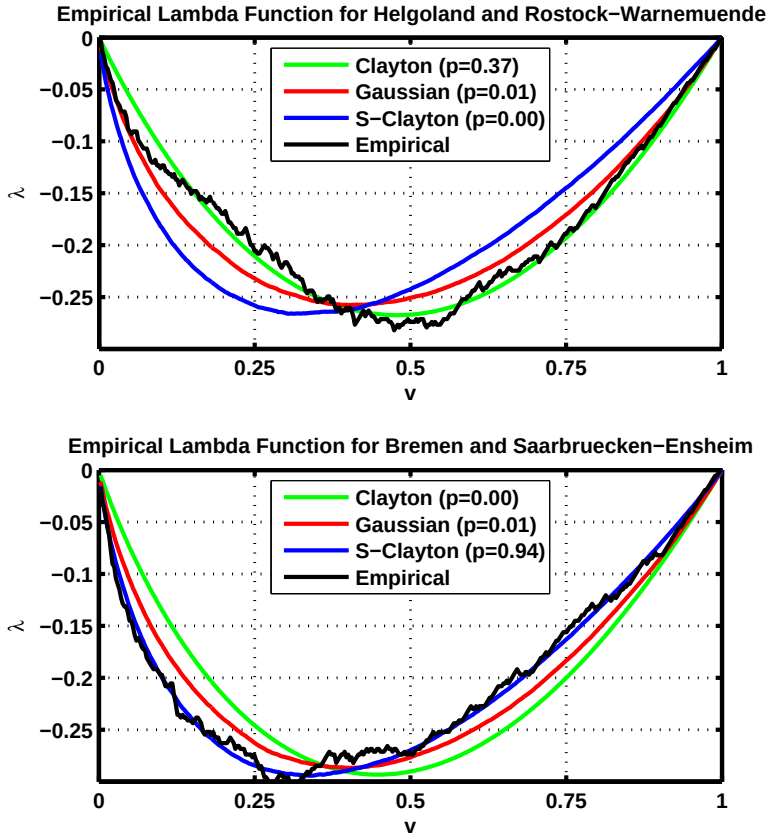

Figure 7: Plot of the λ-Function $\lambda(v) := v - K(v)$ as introduced by Genest and Rivest (1993) for the examples of the pairs Helgoland and Rostock-Warnemuende as well as Bremen and Saarbruecken-Ensheim. In both cases, the empirical line (black) deviates clearly from the Gaussian dependence structure (red line). Furthermore, the dependence structure in both cases is different. In the first case, the dependence fits very well to a Clayton copula (high dependence at the lower quantiles of the distributions). In the second case, the dependence is more of Clayton survival structure (high dependence at the upper quantiles of the distributions).

of the parameters of the model for the marginal time series as well as the estimation of the dependence structure, i.e., the PCC. Since the simultaneous estimation of the marginal models and the copula structure by maximum

likelihood becomes computationally very complex for higher dimensions, we use the inference for margins method and first estimate the model for the univariate time series and then use the corresponding residuals to compute the copula structure (see, e.g., Joe (2005) for a more detailed discussion). The copula structure itself is estimated according to the algorithm presented in Schnieders (2010). It estimates the structure of the PCC, i.e., the $40(40-1)/2$ best fitting bivariate copulas for any combinations of pairs of places, as well as their parameters. The resulting PCC with 780 bivariate copulas is partly pictured in a vine plot in figure 8 (see figure 12 in the Appendix for an introduction to vine plots). The entire PCC contains 343 Frank, 279 Gaussian, 100 Clayton survival and 58 Clayton copulas. The high proportion of non-Gaussian copulas reflects the non-normality of the wind data. The estimated vine structure and parameters of the time series models are available from the authors upon request. Using a Goodness-of-Fit test based on the Rosenblatt transformation (see Aas et al., 2009), we obtain a P-value of 0.8184 by bootstrapping (see Genest et al., 2009). Opposed to that, the P-value for a multivariate Gaussian copula is only 0.2016, clearly indicating the good fit of the vine approach. After the calibration of the model to our data set, we use the estimated parameters to simulate 1000 years of wind data. The optimizations in the next section rely on these simulated data sets.

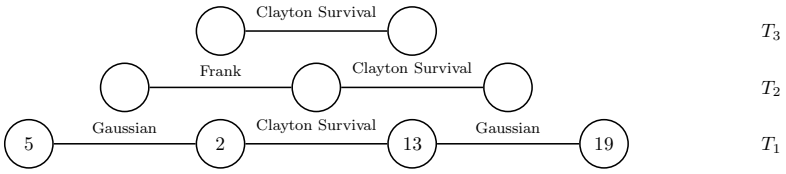

Figure 8: Excerpt of the estimated vine structure, containing the stations in Bremen (5), Augsburg (2), Hamburg (16) and Kempten (22). The names above the edges denote the copula family with the best fit to the respective pair of variables.

3. Optimal wind power allocation

In this section, we find optimal allocations of wind power production over Germany. Since our model is conceptually suited only for the daily frequencies (due to inter-temporal cross-dependence of residuals on higher frequencies), we conduct the optimization on simulated daily data, only. However, the derived solutions lead to large improvements on the hourly level as well. We show this by evaluating the derived optima also with the historical hourly data. Note that the amount of historical data is too small to use the historical data directly in an optimization and we depend on the use of simulated data.

To find optimal allocation plans of wind energy production, in subsection 3.1, the simulated 1000 years of wind speed data are transformed to wind power output. In subsection 3.2, the current allocation of wind power production is mapped onto the considered regions. In subsection 3.3 we explain the optimization setup in detail and subsection 3.4 contains the results. Optimal allocation plans are derived and compared to the distribution of the wind

energy production of 2009. Two scenarios are considered: an overall best allocation and an expansion of the German wind energy production by 40% as is reasonable for the next five to ten years, including offshore exploitation.

3.1. Simulation of wind power data

The model presented in section 2 is used to simulate 1000 years of daily wind data. The data is then transformed to the corresponding power output in two steps. In the first step, it is scaled to the hub height of modern wind turbines, and in the second step, the wind to power relationship of a benchmark wind turbine is used to transform wind speeds to power output. To scale it to the wind speed at the typical hub height of modern wind turbines (80 m), we follow the approach used in Katzenstein et al. (2010) based on Seinfeld and Pandis (2006). With growing altitude it assumes a vertical logarithmic profile of the wind velocity v leading to

$$v_{h_1} = v_{h_0} \cdot \left(\frac{\log(h_1) - \log(z_0)}{\log(h_0) - \log(z_0)} \right), \tag{2}$$

where h_0 and h_1 are the height of the measurements (see table 2 for the respective values of h_0) and the height of interest, respectively. The parameter z_0 corresponds to surface roughness length. According to Katzenstein et al. (2010) we use $z_0 = 0.03$ and $h_1 = 80\,\text{m}$. Having rescaled the wind speed data, we use a GE 1.5 MW turbine as a benchmark to convert the wind speed data into power output. We follow the approach of Archer and Jacobson (2007), and use a combination of third-order polynomials to determine the power

output as a function of the wind speed v_{h_1}:

$$P(v_{h_1}) = \begin{cases} 0 & v_{h_1} < 3\,\mathrm{m/s} \\ P_{\mathrm{lower}}(v_{h_1}) & 3\,\mathrm{m/s} \leq v_{h_1} \leq 8\,\mathrm{m/s} \\ P_{\mathrm{upper}}(v_{h_1}) & 8\,\mathrm{m/s} \leq v_{h_1} \leq 12\,\mathrm{m/s} \\ 1500 & 12\,\mathrm{m/s} \leq v_{h_1} \leq 25\,\mathrm{m/s} \\ 0 & 25\,\mathrm{m/s} \leq v_{h_1} \end{cases}, \tag{3}$$

where $P_{\mathrm{lower}}(v_{h_1}) = v_{h_1}^3 + 8v_{h_1}^2 - 53v_{h_1} + 60$ and $P_{\mathrm{upper}}(v_{h_1}) = -11.25v_{h_1}^3 + 307.5v_{h_1}^2 - 2520v_{h_1} + 6900$. Figure 9 shows a plot of this function.

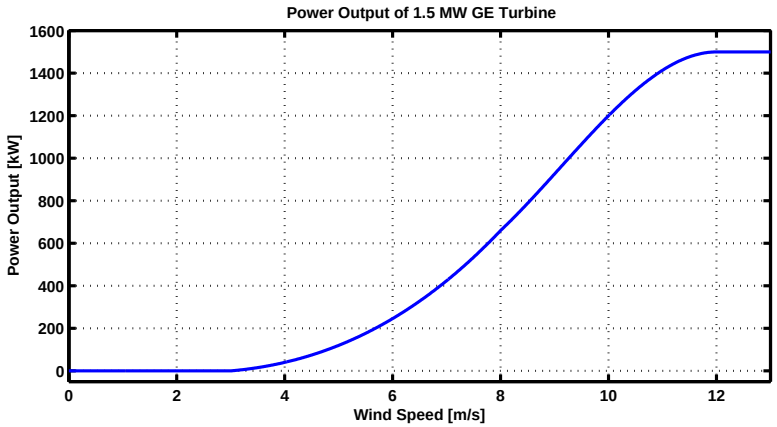

Figure 9: The power output as a function of the wind speed for a GE 1.5 MW turbine. The cut-in and cut-off wind speeds are 3 m/s and 25 m/s, respectively. The rated wind speed is 12 m/s.

3.2. Allocation of current wind power production

To compare optimal allocations of wind power to the current status, we first need to map the current wind energy production onto the regions of

our data sets. Note that there was (an still is) no offshore wind power in 2009. Therefore, the offshore stations are not considered. The 36 onshore stations are the same for the daily and the hourly data set, resulting in the same weights for both frequencies. The proportion of each German wind turbine in the overall wind energy production is added to the onshore regions weighted by its absolute distance to the respective region. A data set containing all German wind parks, their positions and their production in 2009 was provided by the transmission system operators (50Hz, Amprion, EnBw, Transpower). The result is a weight vector $\omega^{2009} \in \mathbb{R}^{40}$, with $\sum \omega(i) = 1$, where the i-th entry corresponds to the proportion of wind energy produced in region i. The vector is shown in the second column of table 3 (Appendix).

The region around Zugspitze shows with 1.1% the smallest proportion in the wind production. This low weight may be explained geographically since the region Zugspitze is a mountain range and the possibility for the installation of wind turbines is limited. The next smallest proportions between 1.2% and 1.4% are of the regions around Hohenpeissenberg, Konstanz, Kempten, Lindenberg and Augsburg which are in the south of Germany. Their low weights seem not to be caused geographically but by the political environment in these parts of Germany. The regions with the highest productions (between 4.0% and 4.7%) are Emden, Magdeburg, Cuxhaven, Bremen and Schleswig. In accordance to figure 1, these are located at the coasts and in eastern Germany.

3.3. Optimization setup

We now derive weight vectors $\omega \in \mathbb{R}^{40}$, where $\omega(i)$ is again the proportion of wind energy produced in region i, but where the overall power output is optimized on daily basis. We carry out the optimizations on the complete set of stations, i.e., on- and offshore, as well as on onshore data only. For each of the cases, we derive two different optimal allocations, an overall optimal allocation (*total optimization*) and the optimal allocation when starting at 2009 allocation and adding 40% of wind energy production (*repowering optimization*). We limit all weights to a maximum of 0.08, i.e., 8% of the overall produced wind energy. We think that this choice of an upper bound is reasonable for most of the regions and corresponds to about the triple of the weight of an uniform distribution of wind power over the country. Considering 4 offshore stations the bounds correspond to a maximum of 32% of offshore wind energy. In the pure onshore scenarios, the offshore weights are restricted to 0%. Additionally we limit the weights of the region Zugspitze to 1.5%, Fichtelberg to 3% and Kahler Asten to 5% due to their geographical characteristics.

For the first optimization we find weight vectors $\omega^\star \in \mathbb{R}^{40}$ such that

$$\omega^\star = \underset{\omega}{\operatorname{argmax}}\, Quantile_\alpha\left(\omega^\top \mathbf{X}\right), \tag{4}$$

under the constraint

$$\sum_{i=1}^{40} \omega(i) = 1,$$

where $Quantile_\alpha\left(\omega^\top \mathbf{X}\right)$ denotes the empirical α-quantile q_α of $\omega^\top \mathbf{X}$ and $\mathbf{X}$ contains all 40 simulated time series normalized to mean 1. Note that ω and $\omega^\star$ denote column vectors and $\omega^\top$ denotes the transposed vector of ω. Thus, the i-th entry of $\omega^\star$ is (on average) the optimal proportion of the produced wind energy in region i in the overall produced wind energy. The solution is optimal in the sense that the overall produced power output has the highest α-quantile q_α among all possible allocations ω. This means that in $(1-\alpha)\cdot 100\%$ of the cases the overall produced power lies above q_α. The quantile q_α is equivalent to the $\alpha \cdot 100\%$-value at risk.

For the second optimization, we assume that the overall produced wind energy is expanded by 40%. Thus, the optimization problem is:

$$\omega^\star = \underset{\omega}{\operatorname{argmax}}\, Quantile_\alpha\left((\omega + \omega^{2009})^\top \mathbf{X}\right), \tag{5}$$

under the constraint

$$\sum_{i=1}^{40} \omega(i) = 0.4.$$

Both optimizations, the total scenario and the repowering scenario, are done for the mean of the quantiles $q_{0.01}$, $q_{0.02}$, ..., $q_{0.12}$. The optimization

of the mean of the quantiles is done to account for possible sensitivity of the weights to the quantiles, i.e., to find weights that are valid for a wide range of quantiles. For the optimization Matlab's constraint optimization function with active-set algorithm is used, which is based on a quasi-Newton approximation to the Hessian of the Lagrangian.

3.4. Results of optimization

The resulting allocations of the optimizations are shown in table 3. The case with on- and offshore installation is additionally depicted in figure 10.

Inspection of the allocations of the total optimization reveals that, in an optimal pure onshore scenario, much more wind energy would be produced in the southern and western part of Germany (area around Saarbruecken-Ensheim, Frankfurt and Duesseldorf) than it is actually done as well as at the coasts of the North Sea (area around Emden and Cuxhaven) and the Baltic Sea (Rostock-Warnemuende, Westermarkelsdorf and Greifswald). On the contrary, regions that produce a high proportion of today's wind energy (e.g., Magdeburg, Neuruppin and Schwerin in the eastern part of Germany) should not contribute to the wind energy production at all or at least strongly decrease their proportion. In the scenario where offshore installation is possible, all 4 offshore regions receive the maximum possible weights. The onshore production would then be concentrated on the southern, western and northern parts of the country and less concentrated in the central and central eastern regions (see also figure 10).

The results for the repowering scenario, i.e., starting at the status quo and adding 40% of production, are of similar structure. Expanding and repowering should in particular be forced in the offshore regions (compare to figure 10). If we restrict to the onshore regions, repowering and installation of new wind turbines should be focused on the coasts of the North Sea and the Baltic Sea as well as mountainous regions in the south of Germany (Fichtelberg, Hohenpeissenberg).

The weights are derived by maximizing the daily model. We evaluate them on hourly and daily frequency by transferring the historical wind data to power data as discussed in section 3.1 and assuming an allocation according to the derived weights. The hourly data set does not include data from the offshore station UFS Deutsche Bucht. We therefore add the weights of this station equally to the nearest offshore stations Hallig Hooge and Helgoland (see table 3 for the resulting weights). We calculate the mean of the quantiles $q_{0.01}$, $q_{0.02}$, $\ldots$, $q_{0.12}$ of the resulting production as well as the 5% quantile and the 12% quantile. The 12% quantile is motivated by the (scheduled and unscheduled) averaged downtime of US coal plants of 12.5%, as discussed in Archer and Jacobson (2007).

The results are shown in table 1 together with the respective quantiles, i.e., value at risk numbers (VaR), of the 2009 allocation. In all cases, the quantiles of the optimized scenarios are higher than those of the 2009 allocation. In particular, for the total optimization with restriction to onshore regions, the 5% value at risk is increased by over 60% (from 6.1% to 10.1%)

for the daily data and by over 40% for the hourly data (8.0% to 11.5%). This can further be improved by allowing for offshore stations (6.1% to 19.3% daily and 8.0% to 20.3% hourly). For example, the last numbers mean that the proportion of the average power production that is online in at least 5% of the time would be 20.3% instead of 8.0%. In this case, 32% of the energy would be produced offshore and the onshore production would be optimally allocated. However, it becomes clear from the table that the larger effect comes from the inclusion of offshore regions than the optimal allocation onshore.

In the repowering scenario, i.e., when starting at the current allocation of wind power production and adding 40% of average production optimally, the improvement in stability is still impressive on hourly as well as on daily frequencies. For example, the 5% VaR of the production raises from 6.1% of the average production to 14.1% (daily) and 15.8% (hourly). In the total scenario this 5% VaR is 19.3%. In both the total and the repowering scenario, around 30% of the wind power is produced offshore, while in the total scenario, the onshore production is additionally optimized, which leads to the further improvement in VaR. Thus it seems, that first of all the offshore production raises stability (i.e., the quantiles) but that the additional optimal installation onshore evens out remaining fluctuations.

4. Conclusion

In this paper, the possibility of smoothing the German wind power output by optimally allocating wind energy capacity across the country is investi-

			VaR Mean	VaR 0.05	VaR 0.12
daily	Status Quo		7.2	6.1	11.8
	onshore	Total	11.9	10.1	18.5
		Repower	10.0	8.2	15.4
	on-/offshore	Total	21.7	19.3	31.4
		Repower	16.0	14.1	23.4
hourly	Status Quo		9.1	8.0	13.8
	onshore	Total	13.3	11.5	20.2
		Repower	11.8	10.3	17.8
	on-/offshore	Total	22.6	20.3	32.7
		Repower	17.7	15.8	25.9

Table 1: Evaluation of the derived allocation vectors using daily and hourly wind data from 2005-01-01 to 2010-12-31. Shown are value at risk numbers of the wind power production in percent of the average wind power in the considered scenarios. They are compared to the respective numbers assuming the allocation of wind power production of 2009 (Status Quo). The actual used weight vectors are shown in table 3. On both frequencies, the optimal weights lead to a clear improvement of the value at risk.

gated. The aim is to find allocations of wind power production maximizing the lower quantiles, i.e., the value at risk of the overall wind power output over time. The optimization is model based. Since the distributions of wind speeds are highly non-Gaussian, nonlinear time series with copula models are used to assess and maximize the quantiles. The models are suited for modeling the wind on daily frequencies. The resulting optimal allocations are then evaluated using historical wind data of daily and hourly frequency. The results show that the current allocation of wind power production in Germany is far from optimal. There is not sufficient capacity installed offshore, at the coasts and in the mountainous regions, whereas too much capacity is located at the eastern part of Germany. The installation of offshore wind

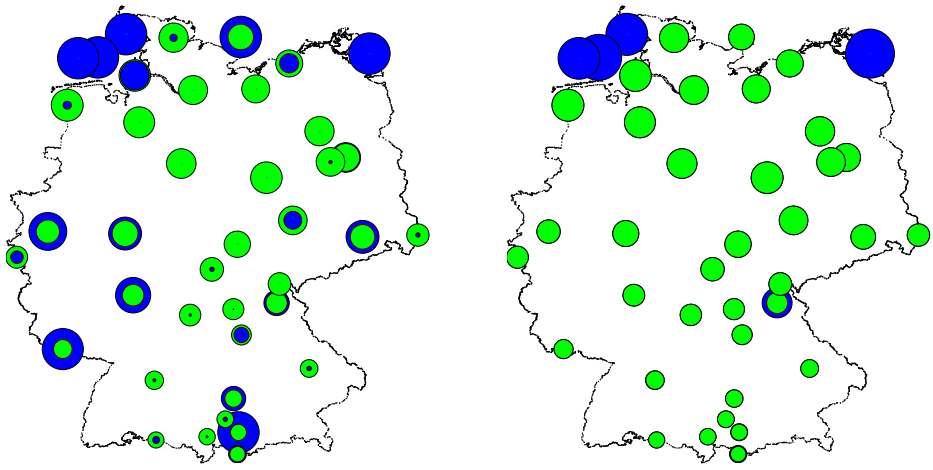

Figure 10: Graphical illustration of the overall optimal allocation (left) and the repowering scenario (right). The status quo is represented by the green circle with sizes according to the weight of the corresponding region. In the left figure, the blue circles denote the proportion of wind energy produced in the given regions according to the optimal allocation, while the blue circles in the right figure denote the areas where the repowering should be focused on. See table 3 for the numerical results.

parks should in particular be forced.

Settling these gaps would strongly reduce the shortfall risk and increase the proportion of wind energy production which is online at a certain probability by more than 150%. Our analysis shows that increasing the mean wind power by 40%, e.g., by installing new capacity or repowering existing turbines, may increase these certain proportions by more than 100%. For this, the expansion of wind power capacity should be forced in the offshore regions and at the coasts as well as in the southern parts of Germany.

References

References

Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics 44 (2), 192–198.

Archer, C. L., Jacobson, M. Z., 2007. Supplying baseload power and reducing transmission requirements by interconnecting wind farms. Journal of Applied Meteorology and Climatology 46 (11), 1701–1717.

Beaucage, P., Lafrance G., Lafrance J., Choisnard J., Bernier M., 2011. Synthetic aperture radar satellite data for offshore wind assessment: A strategic sampling approach. Journal of Wind Engineering and Industrial Aerodynamics 99 (1), 27–36.

Bedford, T., Cooke, R. M., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. Annals of Mathematics and Artificial Intelligence 32 (1-4), 245–268.

Bedford, T., Cooke, R. M., 2002. Vines: A new graphical model for dependent random variables. The Annals of Statistics 30 (4), 1031–1068.

Benth, J. S., Benth, F. E., 2010. Analysis and modelling of wind speed in New York. Journal of Applied Statistics 37 (6), 893–909.

Berg, D., Aas, K., 2009. Models for construction of multivariate dependence: A comparison study. The European Journal of Finance 15 (7), 639–659.

Box, G. E., Cox, D. R., 1964. An analysis of transformations. Journal of the Royal Statistical Society Series B 26 (2), 211–252.

Cherubini, U., Luciano, B., Vecchiato, W., 2004. Copula Methods in Finance. John Wiley & Sons, New York.

Clayton, D., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65 (1), 141–151.

Czisch, G., Ernst, B., 2001. High wind power penetration by the systematic use of smoothing effects within huge catchment areas shown in a european example. American Wind Energy Association, Washington, DC.

Drake, B., Hubacek, K., 2007. What to expect from a greater geographic dispersion of wind farms?–a risk portfolio approach. Energy Policy 35 (8), 3999–4008.

Fischer, M., Köck, C., Schlüter, S., Weigert, F., 2009. An empirical analysis of multivariate copula models. Quantitative Finance 9 (7), 839–854.

Genest, C., Rivest, L.-P., 1993. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 88 (423), 1034–1043.

Genest, C., Quessy, J.-F., Rémillard, B., 2006. Goodness-of-fit Procedures for Copula Models Based on the Probability Integral Transformation. Scandinavian Journal of Statistics 33 (2), 337–366.

Genest, C., Rémillard, B., Beaudoin, D., 2009. Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics 44 (2), 199–213.

Giebel, G., 2000. On the benefits of distributed generation of wind energy in Europe. Ph.D. Thesis.

Joe, H., 1996. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. Distributions with Fixed Marginals and Related Topics, 120–141.

Joe, H., 1997. Multivariate models and dependence concepts. Chapman & Hall, London.

Joe, H., 2005. Asymptotic efficiency of the two-stage estimation method for copula-based models. Journal of Multivariate Analysis 94 (2), 401 – 419.

Kahn, E., 1979. The reliability of distributed wind generators. Electric Power Systems Research 2 (1), 1–14.

Katzenstein, W., Fertig, E., Apt, J., 2010. The variability of interconnected wind plants. Energy Policy 38 (8), 4400–4410.

Milligan, M., Porter, K., 2005. Determining the capacity value of wind: a survey of methods and implementation. Windpower, Denver, Colorado, USA.

Nelsen, R., 2006. An introduction to copulas. Springer, New York.

Schnieders, J., 2010. Pair-copula constructions: Model selection and goodness-of-fit. Preprint.

Seinfeld, J. H., Pandis, S. N., 2006. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. Wiley.

A. Tables

Number	and Region	*mean*	*med*	*std*	*skew*	*kurt*
1	Aachen [202, 16.0]	3.1 / 3.3	2.7 / 2.8	1.7 / 2.0	1.2 / 1.2	4.3 / 4.7
2	Augsburg [462, 10.0]	2.9 / 2.9	2.5 / 2.4	1.5 / 2.0	1.6 / 1.4	6.5 / 5.7
3	Bamberg [239, 10.0]	2.2 / 2.2	2.1 / 2.0	0.9 / 1.4	1.0 / 1.0	4.3 / 3.9
4	Berlin-Tempelhof [48, 10.0]	3.8 / 3.7	3.5 / 3.5	1.5 / 1.8	1.0 / 0.8	4.2 / 4.1
5	Bremen [4, 10.0]	4.1 / 4.1	3.8 / 3.8	1.7 / 2.1	0.9 / 0.8	4.2 / 3.7
6	Cuxhaven [5, 26.1]	5.3 / 5.3	5.0 / 4.9	2.0 / 2.5	0.9 / 0.7	3.8 / 3.5
7	Dresden-Klotzsche [227, 10.0]	4.1 / 4.0	3.8 / 3.6	1.7 / 2.0	1.0 / 1.0	4.0 / 4.0
8	Duesseldorf [37, 10.2]	3.9 / 3.8	3.7 / 3.5	1.7 / 2.1	0.8 / 0.8	3.7 / 3.8
9	Emden [0, 9.7]	4.4 / 4.3	4.1 / 4.1	1.8 / 2.2	0.8 / 0.7	3.9 / 3.6
10	Erfurt-Weimar [316, 10.0]	4.2 / 4.1	3.7 / 3.5	2.0 / 2.3	1.1 / 1.2	4.2 / 4.8
11	Fichtelberg [1213, 23.8]	9.4 / 9.0	8.6 / 8.3	4.2 / 4.5	0.7 / 0.6	3.1 / 2.9
12	Frankfurt/Main [112, 10.0]	3.3 / 3.2	3.0 / 2.8	1.5 / 2.0	1.2 / 1.3	4.9 / 5.3
13	Goerlitz [238, 13.0]	3.9 / 3.8	3.4 / 3.3	1.9 / 2.2	1.0 / 1.0	3.6 / 3.9
14	Greifswalder Oie [12, 3.0]	6.9 / 6.7	6.3 / 6.2	2.9 / 3.3	0.8 / 0.7	3.3 / 3.4
15	Hallig Hooge [4, 10.0]	7.6 / 7.5	7.3 / 7.2	2.9 / 3.4	0.6 / 0.5	3.0 / 3.2
16	Hamburg-Fuhlsbuettel [11, 10.0]	3.9 / 3.9	3.7 / 3.7	1.6 / 2.0	0.7 / 0.6	3.4 / 3.2
17	Hannover [55, 10.0]	3.8 / 3.7	3.5 / 3.4	1.5 / 1.9	0.9 / 0.8	3.8 / 3.8
18	Helgoland [4, 10.0]	8.5 / 8.3	8.0 / 7.9	3.3 / 3.8	0.5 / 0.5	2.8 / 2.9
19	Hof [565.1, 16.0]	3.1 / 3.1	2.9 / 2.9	1.3 / 1.6	1.1 / 0.9	4.6 / 4.0
20	Hohenpeissenberg [977, 40.5]	5.3 / 4.7	4.4 / 3.8	3.0 / 3.2	1.6 / 1.8	6.1 / 7.2
21	Kahler Asten [839, 27.3]	6.1 / 6.1	5.8 / 5.8	2.1 / 2.5	0.8 / 0.7	3.4 / 3.6
22	Kempten [705, 5.0]	2.1 / 2.1	1.9 / 1.9	0.8 / 1.1	1.7 / 1.2	7.7 / 6.0
23	Konstanz [443, 17.0]	2.2 / 2.2	1.8 / 1.7	1.0 / 1.4	1.9 / 1.8	7.8 / 7.7
24	Leipzig-Halle [131, 10.0]	4.3 / 4.2	3.9 / 3.7	1.8 / 2.2	1.0 / 1.2	3.9 / 4.8
25	Lindenberg [98, 10.4]	3.5 / 3.5	3.2 / 3.1	1.5 / 1.7	1.3 / 1.4	5.2 / 5.7
26	Magdeburg [76, 18.0]	2.5 / 2.5	2.2 / 2.3	1.2 / 1.4	1.2 / 1.1	4.7 / 4.9
27	Meiningen [450, 18.0]	3.2 / 3.1	3.0 / 2.9	1.4 / 1.9	0.8 / 0.8	3.9 / 3.7
28	Neuruppin [38, 18.0]	2.9 / 2.9	2.7 / 2.7	1.3 / 1.7	0.8 / 0.8	3.6 / 3.8
29	Nuernberg [314, 10.0]	3.0 / 2.9	2.8 / 2.6	1.3 / 1.7	1.3 / 1.2	5.5 / 5.0
30	Potsdam [81, 37.7]	4.2 / 4.2	4.0 / 4.0	1.6 / 1.8	0.9 / 0.8	3.9 / 4.3
31	Rostock-Warnemuende [4, 22.0]	4.9 / 4.8	4.2 / 4.1	2.4 / 2.8	1.4 / 1.4	5.1 / 5.7
32	Saarbruecken-Ensheim [320, 10.0]	3.5 / 3.5	3.2 / 3.2	1.6 / 2.1	1.1 / 1.0	4.6 / 4.6
33	Schleswig [43, 16.6]	4.0 / 3.9	3.7 / 3.6	1.5 / 1.9	1.0 / 0.8	4.3 / 3.6
34	Schwerin [59, 22.0]	3.8 / 3.8	3.5 / 3.5	1.7 / 1.9	1.1 / 0.9	4.5 / 4.0
35	Straubing [371, 10.0]	2.6 / 2.5	2.3 / 2.2	1.2 / 1.6	1.5 / 1.3	6.3 / 5.4
36	Stuttgart-Echterdingen [371, 10.0]	2.6 / 2.5	2.2 / 2.0	1.3 / 1.8	1.3 / 1.3	5.5 / 5.3
37	UFS Deutsche Bucht [0, 10.0]	8.0 / –	7.7 / –	3.2 / –	0.5 / –	2.9 / –
38	Westermarkelsdorf [3, 10.0]	6.1 / 6.0	5.7 / 5.6	2.7 / 3.1	0.9 / 0.7	3.6 / 3.4
39	Wuerzburg [268, 10.0]	3.1 / 3.1	2.7 / 2.6	1.6 / 2.0	1.4 / 1.4	6.3 / 5.5
40	Zugspitze [2964, 16.0]	7.2 / 7.3	6.6 / 6.6	3.2 / 4.0	1.0 / 1.1	4.1 / 4.8

Table 2: Descriptive statistics of the wind speeds of all 40 considered weather stations for the time 2005-01-01 to 2010-12-31. The first number in each column refers to daily data, the second to hourly data. Brackets behind the names contain the respective absolute altitudes of the stations and wind detector heights in m. As to be expected, the mean values of the stations deviate clearly with values below 2 m/s (Kempten) and over 8 m/s (Fichtelberg, Helgoland). Kempten is located in a shielded valley in southern Germany, while Fichtelberg lies exposed in the forelands of Bavaria and Helgoland lies exposed in the north sea. All wind distributions are skewed to the right and show excess kurtosis, i.e., are heavier tailed than the Gaussian distribution. Both data sets are independently provided by the German wheather service.

Number and Region		2009	On- and Offshore		Only Onshore	
			Total	Repower	Total	Repower
1	Aachen	2.2	0.7 / 0.7	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
2	Augsburg	1.4	2.8 / 2.8	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
3	Bamberg	2.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
4	Berlin-Tempelhof	3.7	4.1 / 4.1	0.0 / 0.0	2.4 / 2.4	0.0 / 0.0
5	Bremen	4.5	0.0 / 0.0	0.0 / 0.0	3.5 / 3.5	0.0 / 0.0
6	Cuxhaven	4.6	3.9 / 3.9	0.0 / 0.0	8.0 / 8.0	1.3 / 1.3
7	Dresden-Klotzsche	2.8	5.1 / 5.1	0.0 / 0.0	7.6 / 7.6	0.0 / 0.0
8	Duesseldorf	2.5	6.7 / 6.7	0.0 / 0.0	7.7 / 7.7	2.5 / 2.5
9	Emden	4.7	0.3 / 0.3	0.0 / 0.0	8.0 / 8.0	3.5 / 3.5
10	Erfurt-Weimar	3.3	0.0 / 0.0	0.0 / 0.0	0.9 / 0.9	0.0 / 0.0
11	Fichtelberg	2.1	3.0 / 3.0	2.1 / 2.1	3.0 / 3.0	2.1 / 2.1
12	Frankfurt/Main	2.2	5.8 / 5.8	0.0 / 0.0	7.4 / 7.4	0.0 / 0.0
13	Goerlitz	2.3	0.1 / 0.1	0.0 / 0.0	1.1 / 1.1	0.0 / 0.0
14	Greifswalder Oie	0.0	8.0 / 8.0	11.2 / 11.2	– / –	– / –
15	Hallig Hooge	0.0	8.0 / 12.0	8.2 / 12.2	– / –	– / –
16	Hamburg-Fuhlsbuettel	3.9	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
17	Hannover	4.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
18	Helgoland	0.0	8.0 / 12.0	9.9 / 14.0	– / –	– / –
19	Hof	2.4	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
20	Hohenpeissenberg	1.2	8.0 / 8.0	0.1 / 0.1	8.0 / 8.0	10.0 / 10.0
21	Kahler Asten	3.1	5.0 / 5.0	0.0 / 0.0	5.0 / 5.0	3.9 / 3.9
22	Kempten	1.3	0.0 / 0.0	0.0 / 0.0	0.4 / 0.4	0.3 / 0.3
23	Konstanz	1.2	0.2 / 0.2	0.0 / 0.0	0.2 / 0.2	0.0 / 0.0
24	Leipzig-Halle	3.8	1.5 / 1.5	0.0 / 0.0	2.0 / 2.0	0.5 / 0.5
25	Lindenberg	1.3	0.1 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
26	Magdeburg	4.7	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
27	Meiningen	2.6	0.1 / 0.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
28	Neuruppin	4.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
29	Nuernberg	1.9	1.1 / 1.1	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
30	Potsdam	3.8	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
31	Rostock-Warnemuende	3.4	1.7 / 1.7	0.0 / 0.0	8.0 / 8.0	0.9 / 0.9
32	Saarbruecken-Ensheim	1.7	7.9 / 7.9	0.0 / 0.0	8.0 / 8.0	5.7 / 5.7
33	Schleswig	4.0	0.2 / 0.2	0.0 / 0.0	7.2 / 7.2	0.4 / 0.4
34	Schwerin	3.6	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
35	Straubing	1.5	0.1 / 0.1	0.0 / 0.0	1.5 / 1.5	0.5 / 0.5
36	Stuttgart-Echterdingen	1.6	0.0 / 0.0	0.1 / 0.1	0.6 / 0.6	0.0 / 0.0
37	UFS Deutsche Bucht	0.0	8.0 / –	8.1 / –	– / –	– / –
38	Westermarkelsdorf	3.1	8.0 / 8.0	0.0 / 0.0	8.0 / 8.0	8.1 / 8.1
39	Wuerzburg	2.2	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0	0.0 / 0.0
40	Zugspitze	1.1	1.5 / 1.5	0.3 / 0.3	1.5 / 1.5	0.3 / 0.3

Table 3: Percentages of the wind energy production for the status quo and the optimization scenarios. Note that in the repower scenario the vector of 2009 data and the repower vector sum up to 140% of production, while in the total optimization the total vector sums up to 100%. Daily and hourly weights differ in the offshore stations 38, 18 and 15.

B. Copulas

In this section we briefly discuss multivariate copula constructions. A more detailed introduction to copula modeling is given in Joe (1997), Cherubini et al. (2004) and Nelsen (2006).

B.1. Copula functions

Let $X_1, \ldots, X_d$ be continuous random variables with joint distribution $F_{X_1,\ldots,X_d}$ and marginal distributions $F_{X_i}(x_i)$ for $i = 1 \ldots d$. Then there exists a uniquely defined distribution function $C : [0,1]^d \longrightarrow [0,1]$ with uniform margins such that

$$F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = C\left(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\right) \qquad (x_1, \ldots, x_d) \in \mathbb{R}^d.$$

C is called copula of $X_1, \ldots, X_d$ and captures the complete dependence structure of $X_1, \ldots, X_d$.

Copula functions may be defined in a parametric way. An important example for a parametric family of copulas is the Gaussian copula (see Joe (1997)). The Gaussian copula is the copula of multivariate Gaussian distributed random variables and it is completely determined by the pairwise rank correlations of the variables. It is defined by

$$C_{\mathbf{\Sigma}}(u_1, \ldots, u_d) = \Phi_{\mathbf{\Sigma}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right),$$

where $\Phi_{\mathbf{\Sigma}}$ is the distribution function of the multivariate normal distribu-

tion with covariance matrix $\mathbf{\Sigma} = (\rho_{ij})_{i,j=1\ldots d}$, which is a positive definite correlation matrix and Φ^{-1} denotes the quantile function of a standard normal distribution. Another example for a parametric family of copulas is the Clayton family (see Clayton (1978)), given by

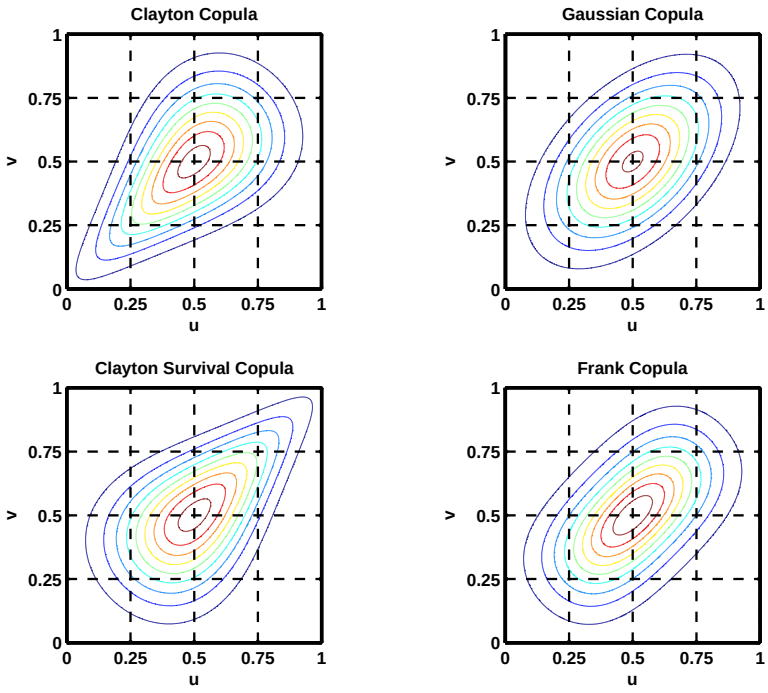

Figure 11: Contour plots of Clayton, Gaussian, Clayton survival and Frank copulas with linear correlation $\rho = 0.5$. The dependence given by the Gaussian copula is symmetric while the Clayton and Clayton survival copula show asymmetric dependence structure. In the case of the Clayton copula, U and V are more dependent for smaller values, and in the case of the Clayton survival for larger values. The Frank copula shows less dependence for more extreme values of U and V than for values in the middle range, whereas the Gaussian copula shows the same strength of dependence for all ranges.

$$C_\theta(u_1, \ldots, u_d) := \left(\sum_{i=1}^{d} u_i^{-\theta} - (d-1) \right)^{-1/\theta},$$

where $\theta > 0$. Given the bivariate Clayton copula, we may also define its survival copula, the Clayton survival copula:

$$C_\theta(u_1, u_2) := u_1 + u_2 - 1 + \left((1-u_1)^{-\theta} + (1-u_2)^{-\theta} - 1 \right)^{-1/\theta}.$$

The fourth example of copulas considered in this paper is the Frank copula, given by

$$C_\Psi(x_1, x_2, \ldots, x_d) = \Psi^{-1} \left(\sum_{i=1}^{n} \Psi(F_i(x_i)) \right),$$

where Ψ is a generator function with

$$\Psi(x) = -\ln \left(\frac{e^{-\alpha x} - 1}{e^{-\alpha} - 1} \right).$$

Figure 11 shows contour plots of the densities of uniformly $[0, 1]$-distributed random variables U and V with Clayton copula, Gaussian copula and Clayton survival copula, respectively. The dependence given by the Gaussian copula is symmetric while the Clayton and survival Clayton copula show asymmetric dependence structure. In the case of the Clayton copula, U and V are more dependent for smaller values, in the case of the survival Clayton for larger values.

In section 2, the latter two copulas were found to capture the dependence of the wind speeds of the pairs Helgoland and Rostock-Warnemuende (Clay-

ton) and Bremen and Saarbruecken-Ensheim (Survival Clayton) Thus, for Helgoland and Rostock-Warnemuende the dependence gets stronger for low wind speeds whereas for Bremen and Saarbruecken-Ensheim it gets higher for higher wind speeds. The Frank copula shows less dependence for extreme values of U and V than for values around 0.5.

The construction of proper copula functions for the case $d > 2$ is not trivial and topic of recent research (see, e.g., Fischer et al. (2009)) among many others. The introduced copulas above and many multivariate families of copulas suffer from the drawback that the pair-wise dependence structure of their one-dimensional margins are of the same type. However, for wind speeds and in many other cases the pairwise dependence structure varies between the dimensions. More flexible concepts of copulas are pair copula constructions which are constructed from bivariate copulas.

B.2. Pair-copula constructions

Pair copulas, originally introduced by Joe (1996), provide a flexible way to extend bivariate copula theory to the multivariate case (see Bedford and Cooke (2001, 2002) and Berg and Aas (2009) for a more detailed introduction). The main idea of pair-copulas is to decompose multivariate copulas into a cascade of bivariate copulas. Let F be a joint distribution function with marginals $F_1, F_2, \ldots, F_d$ and denote the corresponding densities by f and $f_1, f_2, \ldots, f_d$, respectively. Then the multivariate density $f(x_1, x_2, \ldots, x_d)$

may be uniquely (up to relabeling) decomposed by iteratively conditioning:

$$\begin{aligned} f(x_1, x_2, \ldots, x_d) &= f(x_d \,|\, x_1, x_2, \ldots, x_{d-1}) \cdot f(x_1, \ldots, x_{d-1}) \\ &= \ldots = f(x_1) \cdot \prod_{i=2}^{d} f(x_i \,|\, x_1 \ldots x_{d-1}). \end{aligned} \tag{6}$$

To obtain a pair-copula construction for $f(x_1, x_2, \ldots, x_d)$, (6) can be expressed in terms of the marginal densities and a product of $d(d-1)/2$ bivariate pair-copulas C.

These pair-copulas C are arbitrary bivariate copulas. The decomposition is not unique anymore. Depending on the variables we condition on in the PCC, we obtain different decompositions of the same multivariate density. Since the number of different decompositions increases sharply with dimension d, Bedford and Cooke (2001, 2002) introduced a graphical model, called *regular vines*, to describe these structures. A d-dimensional vine is represented by $d-1$ trees T_j, $j = 1, \ldots, d-1$, which have $d+1-j$ nodes and $d-j$ edges. Each edge of a tree corresponds to a pair-copula density. The edges of tree T_j become the nodes in tree $j+1$. Two nodes in tree T_{j+1} are joined by an edge if the corresponding edges in tree T_j share a node. The whole decomposition is defined by the marginal distributions and the $d(d-1)/2$ bivariate pair-copulas, that do not necessarily need to belong to the same class of copulas.

There are different approaches of building such vines. In this study we concentrate on the D-vine approach. A D-vine is a regular vine for which no

node in any tree is connected to more than two edges. See figure 12 for an illustration of a 4-dimensional D-Vine structure. For D-vines one gets the following expression for the conditional densities in (6)

$$\begin{aligned} f(x_i \,|\, x_1, \dots x_{i-1}) &= \prod_{j=1}^{i-2} c_{j,i|j+1,\dots,i-1}(F(x_j \,|\, x_{j+1}, .. \\ &\quad .., x_{i-1}), F(x_i \,|\, x_{j+1}, \dots, x_{i-1})) \\ &\quad \cdot c_{i-1,i}(F(x_{i-1}), F(x_i)) \cdot f_i(x_i), \end{aligned} \tag{7}$$

where $c_{j,i|j+1,\dots,i-1}$ is the density (derivative) of the copula of the conditional distribution of x_j and x_i given $x_{j+1}, \dots, x_{i-1}$. Replacing the conditional densities in (6) with (7) finally leads to the density of a D-vine PCC:

$$\begin{aligned} f(x_1, \dots, x_d) &= \prod_{i=1}^{d-1} \prod_{j=1}^{d-i} c_{j,i+j|j+1,\dots,i+j-1}(F(x_j \,|\, x_{j+1}, .. \\ &\quad .., x_{i+j-1}), F(x_{i+j} \,|\, x_{j+1}, \dots, x_{i+j-1})) \\ &\quad \cdot \prod_{i=1}^{d} f_i(x_i). \end{aligned} \tag{8}$$

It is easy to check that formula (8) consists only of 1-dimensional conditional marginal distribution functions and 2-dimensional conditional copula densities. Thus, formula (8) decomposes the joint density of $X_1, \dots, X_d$ in a product of bivariate copula functions. These copulas may be of arbitrary and mixed type. For this reason the decomposition (8) enables for a very flexible modeling of dependence structures based on 2-dimensional copulas.

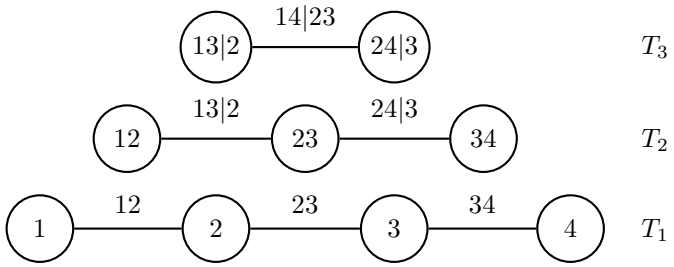

Figure 12: D-Vine structure for a 4-dimensional random vector. The bottom row, tree T_1, indicates the 4 dimensions of the vector denoted by the nodes 1 to 4. The nodes are connected by edges 12, 23 and 34 denoting pairwise copulas describing the (unconditional) dependence of the respective dimensions. In the next row, tree T_2, the edges 13|2 and 24|3 denote the copula of the conditional distribution of 1 and 3 given 2 respectively 2 and 4 given 3. In tree T_3 the edge 14|23 denotes the copula of the conditional distribution of 1 and 4 given 2 and 3. Altogether, the dependence of the 4 dimensions is captured by $4 \cdot 3/2 = 6$ 2-dimensional copulas.