Impact of Wind Power Generation on European Cross-Border Power Flows

Marco Zugno, *Student Member, IEEE* Pierre Pinson, *Member, IEEE* and Henrik Madsen

***Abstract*—A statistical analysis is performed in order to investigate the relationship between wind power production and cross-border power transmission in Europe. A dataset including physical hourly cross-border power exchanges between European countries as dependent variables is used. Principal component analysis is employed in order to reduce the problem dimension. Then, nonlinear relationships between forecast wind power production as well as spot price in Germany, by far the largest wind power producer in Europe, and power flows are modeled using local polynomial regression. We find that both forecast wind power production and spot price in Germany have substantial nonlinear effects on power transmission on a European scale.**

***Index Terms*—wind power generation, power transmission, principal component analysis, regression analysis.**

I. Introduction

DRIVEN by the need to comply with stringent international agreements, which aim at reducing the environmental impact of energy production as well as energy dependence, the deployment of renewable energy in Europe has grown at an unprecedented pace in the recent years. Among renewable sources, wind power plays a central role both for its impressive technological development and for its expansion. Particular features of wind power, like its stochastic and non-dispatchable nature and its very low marginal cost, render it very different from the more conventional sources of energy.

Due to its low marginal cost, wind power production has the consequence of lowering market prices via the so-called "merit-order effect" [1]. This is because wind power enters the energy supply function from the left, or, in an alternative interpretation, reduces the load and thus shifts the intersection between supply and load to the left, thus pushing more expensive sources of energy out of the production schedule. Simulation with market models in [2] and statistical analysis in [3] confirm the price reduction effect of wind power. The latter work also shows that the driving variable of this mechanism is wind power forecast rather than actual production, since the former one is used when producers bid on the market.

The European transmission network is composed of five different synchronous zones, which in turn gather several interconnected national and international energy markets. These are organized with different rules and characterized by different generation portfolios; furthermore climate conditions vary widely across Europe. In these conditions significant price differentials are likely to develop between areas, and therefore also significant flows of power from areas with low power price to areas where energy is more expensive. Thus by influencing energy prices, wind generation also drives flows of power from areas with temporarily favorable conditions for wind power production to areas with higher price level, be it due to a high demand or an expensive generation mix [4]. Massive investments are planned in order to increase the transmission limits between countries of the EU in the years to come [5].

Among the challenges for a successful integration of high penetration of wind power are the variability of its power output and the limited accuracy of wind power forecasting [6], [7]. Both these problems can be addressed by aggregating the power output of wind farms distributed over a wide region. Indeed due to the lower variability of wind power production in Europe as compared to generation from a single region, more than 20% of the European demand could be covered by wind power without significant changes in the system [8]. Furthermore forecast errors can be drastically reduced by the so-called "smoothing effect" of aggregation, as discussed in [9]. Investigations of this type generally assume infinite transmission capacity, while as [10] points out, the interaction between wind power production and transmission constraints should be accounted for, if these phenomena are to be analyzed at a European scale.

In this context, modeling how wind power production interacts with the flow of energy in large international power systems is particularly appealing. Models of this type are needed when planning investments in new wind power or transmission capacity. From an operational point of view, they can help the process of scheduling cross-border power exchanges.

Power system models have been developed and simulated in the literature in order to study the effect of increasing penetration of wind power on European cross-border flow. Such models are simulated in [10] in order to study the congestion of individual interconnections in different scenarios of wind power penetration in Europe. A similar intent is pursued in [11], with the focus being on offshore wind power, and in [12], where copulas are employed to simulate wind power production at different locations. Besides, the effect of wind forecast errors on the uncertainty of cross-border flows has been investigated in [13].

As opposed to simulation using market and grid models, this work follows a top-down statistical approach based on historical data, along the lines of the method employed in [14]. Among the advantages of this approach is the relative

M. Zugno, P. Pinson and H. Madsen are with DTU Informatics, Technical University of Denmark, Kgs. Lyngby, Denmark (e-mail: {mazu,pp,hm}@imm.dtu.dk).

Manuscript received Month dd, yyyy; revised Month dd, yyyy.

simplicity, since no detailed modeling of the underlying physical structure of the power system is required. On the other hand, the reliance on historical data implies the impossibility of extrapolating the analysis out of the range of observations. In this paper a method for analyzing the impact of wind-related variables (external variables) on the European cross-border flows (dependent variables) is developed and employed. The focus is directed towards the effect of forecast wind power production in Germany, which, besides being the largest producer of wind energy in Europe, is centrally located and highly interconnected with the neighboring countries. As Germany—along with other countries in Northern Europe—is setting ambitious targets for installed wind power capacity already by 2020 [15] the presented methodology will allow to assess how future deployment of wind power in this important region will affect the European transmission grid, pinpointing its limitations and possible bottlenecks.

The methodology proposed in this work follows three steps. First, Principal Component Analysis (PCA) is employed in order to reduce the size of the problem, which would otherwise require the analysis of a large number of flows. PCA determines the most significant modes of the flow dataset, i.e. the directions in which it shows most of its variation. The dimension reduction is then performed by selecting a reduced set of modes, which account for a large fraction of the variance of the original dataset. At a second stage, local polynomial regression is applied on this basis in order to model the interaction between the external variables and the chosen modes of the flow dataset. The final step consists in mapping the results of the analysis back from the reduced basis to the original space, i.e. the individual cross-border flows.

This paper is structured as follows. The dataset used in this work is briefly introduced in Section II, and the choice of the explanatory and dependent variables is motivated. Section III describes the employed methodology. In Section IV we discuss the results of the application of this method on the available dataset. Finally, concluding remarks and possible future extensions of this work are provided in Section V.

II. Dataset

The dataset employed in this work spans a period of 3 years from January 2006 until the end of December 2008. Since during the winter daylight savings only one measurement is available for the duplicated hour, 26301 hourly observations are available in total both for dependent and explanatory variables.

A. Dependent variables

Physical hourly cross-border power flows between 34 European and bordering extra-European countries form the set of the dependent variables used in this work. This consists of 70 flows in 2006, 72 in 2007 after the addition of the tie-lines connecting Bulgaria-Macedonia and Estonia-Finland, and 74 in 2008 after the addition of the Norway-Netherlands and Greece-Turkey interconnections. Since the analysis to carry out needs data to be available for the whole period for all the interconnections, the set of physical flows is restricted to the original 70 interconnections established as of 2006. Furthermore, data are missing for significant parts of the period 2006–2008 in other two flows. The final dataset is therefore restricted to 68 cross-border interconnections. Given the low number of discarded flows compared to the total, such discard has a limited impact.

Finally, a data cleaning procedure indicated the presence of outliers stemming from phenomena of different nature. Exceptionally high or low values for most European flows were recorded during the UCTE system split on the 4th November 2006, see [16]. Similarly, unusual flows can be observed during the winter switch from daylight savings to solar time for most interconnections. Finally, a small number of local, single-hour spikes involving few adjacent flows is observed, possibly stemming from smaller technical failures. Such limited number of outliers is removed from the dataset leaving 26281 hourly observations.

B. Explanatory variables

For the reasons mentioned in Section I the focus of the analysis is directed towards wind power production in Germany. As [3] states, the driving variable to be considered when analyzing the effect of wind power is the production forecast rather than the actual production. Indeed the former is used when bidding wind power at the spot market, where the price is settled. Since both wind power forecasts and load affect the spot price, it is of interest to analyze their combined effect on the cross-border flow of power. Therefore the first explanatory variable to be considered in the analysis is the forecast wind power penetration

$$\widehat{r}_t = \frac{\widehat{W}_t}{L_t}\ , \tag{1}$$

where $\widehat{W}_t$ is the wind power production forecast and L_t the load, both aggregated for Germany as a whole. Both these variables are available in the considered dataset for the whole 2006–2008 period. It should be noticed that wind power in Germany developed constantly in the considered period. Indeed, the installed capacity grew from 18.4 GW at the beginning of 2006 to 23.9 GW at the end of 2008 [17]. Nevertheless, the impact on this study is limited owing to the fact that wind power penetration is employed rather than a scaling of the production with respect to the total installed capacity.

An alternative approach is the direct use of the spot price in Germany as an explanatory variable. The effect of wind power is then considered in an indirect way, under the assumption that wind power forecast, or penetration, is negatively correlated with the spot price, as shown in [3]. Although the core of the analysis presented in this paper considers the wind power penetration as explanatory variable, we provide an example using the electricity price in Section IV-B.

The time series of spot prices in the German electricity market (EEX) is characterized by sparse spikes reaching out to around €2500/MWh. As a way to solve this issue, among other benefits, logarithmic transformation is commonly employed when dealing with price time series, see e.g. [18].

The time series of logarithmic prices can be generated through the transformation

$$P_t^l = \log\left(1 + P_t\right) \ . \tag{2}$$

This way the distance between the sparse high prices is shrunk, while the relative distance between the denser low prices is increased. As a side effect, handling negative prices is not possible under the logarithmic transformation. In the dataset, 15 prices at the end of the year 2008 turned out to be negative. These values are discarded from the dataset, which is therefore further reduced to 26266 observations. Alternatively, a shifted log transformation [19] could be employed without requiring the exclusion of negative prices.

III. Methodology

Although the study of a power network could be performed by independently analyzing each single flowgate, several reasons point at other options. First, studying a wide power system like the European one would require the analysis of a large number of flows, with negative consequences on the dimensionality of the problem. Furthermore, due to the net structure of power systems several flows can be highly correlated due to e.g. loop flows. This means that part of the analysis carried out by independently considering each interconnection would be redundant. On top of that, noise can render less visible the object of the investigation. As Section III-A explains, PCA is used here in order to overcome these issues.

Once a reduced set of principal components is indicated by PCA, statistical regression is employed on the reduced basis in order to model the dependence structure between flows and external variables. As power systems are complex and nonlinear, their study requires nonlinear regression techniques. Local polynomial regression was successfully applied in [14] to perform an analysis similar to the one in this work, though only considering the Austrian power system. The same technique, which is introduced in Section III-B, is used in this work. The reader interested in a deeper presentation of local polynomial regression is referred to [20].

A. Principal component analysis

PCA is a technique that is often used when dealing with multivariate data in order to reduce the problem dimension, see [21]. Problem simplification is obtained by selecting a reduced basis of orthogonal variables, which account for most of the variance in the dataset. Let us denote with the vector $\mathbf{X_t}$ the values of the N physical hourly cross-border flows at time t. Let us also assume that T hourly values are available for each interconnection. The centered version $\widetilde{\mathbf{X}}_\mathbf{t}$ of the multivariate time-series of the power flows is given by

$$\widetilde{\mathbf{X}}_\mathbf{t} = \mathbf{X_t} - \bar{\mathbf{X}} \ , \tag{3}$$

where $\bar{\mathbf{X}}$ is the vector of the mean values of the N flows. The covariance matrix $\mathbf{C}$ of the flows can be computed as

$$\mathbf{C} = \frac{1}{T}\sum_{t=1}^{T} \widetilde{\mathbf{X}}_\mathbf{t} \widetilde{\mathbf{X}}_\mathbf{t}^T \ . \tag{4}$$

The eigenvectors of the covariance matrix $\mathbf{C}$ form a new orthogonal basis for the flow dataset. In practice, such eigenvectors represent modes of the dataset, i.e. groups of flows that often exhibit a similar behavior. By ordering the eigenvectors so that the corresponding eigenvalues are arranged in a decreasing fashion, one ensures that the higher the ranking of a vector in the new basis, the higher the fraction of total variance of the dataset it explains. This is a trivial result of the fact that these fractions and the eigenvalues are linearly proportional.

The principal components are obtained by selecting the first n eigenvectors in the new ordered basis. Although the choice of n is arbitrary, there are several criteria for this selection, e.g. the method of the average eigenvalue and the scree graph method, which are discussed in [21]. The latter method, which basically consists in a graphical discrimination between small and large eigenvalues, has been used in this work. As a consequence of this selection, only the n directions (or modes) of the flow space with the largest variance are retained in the analysis, while the remaining $N - n$ are discarded. This is done since the modes with larger variance carry most of the statistical information while the ones with smaller variance can often be associated with noise.

Let us denote with $\mathbf{Y_i}, i = 1, \ldots, n$ the principal components of the dataset. These form a reduced orthonormal basis for the original flows, and one can always write the centered flow observations $\widetilde{\mathbf{X}}_\mathbf{t}$ as a linear combination of the PCs $\mathbf{Y_i}$'s plus an error term ϵ_t, which has zero mean and finite variance

$$\widetilde{\mathbf{X}}_\mathbf{t} = \sum_{i=1}^{n} \alpha_{i,t} \mathbf{Y_i} + \epsilon_t \qquad \forall t \ , \tag{5}$$

In other words, we achieve a similar statistical representation of the original flow dataset through linear projection from the reduced space of the PCs. Generally speaking, the higher the fraction of the original variance is retained with the chosen PCs, the more accurate such representation will be. More precisely, it is to be underlined that the variance is a full description of the statistical information contained in a dataset only under the assumption of joint normality. When the original variables (flows) do not follow a multivariate normal distribution, the comparison between the variance of the original dataset and the variance of its projection on the PC space is an indicator of the amount of the retained statistical information up to moments of order 2. In the case that higher order moments of the residuals are large in comparison to the original signal, alternatives to PCA should be considered, e.g. Independent Component Analysis (ICA) [22].

The advantage of using PCA is now clearly visible, as it is possible to represent every multivariate flow observation $\widetilde{\mathbf{X}}_\mathbf{t}$, which is N-dimensional, with a set of n coefficients $\alpha_{i,t}$, where $n < N$.

It should be pointed out that PCA is often carried out on centered and standardized variables, i.e. by diagonalizing the correlation matrix $\mathbf{R}$ rather than the covariance matrix $\mathbf{C}$. The choice of using the covariance matrix is motivated by the fact that in this way the information on the magnitude of the flows is not lost in the division $r_{ij} = c_{ij}/(\sigma_{\widetilde{X}_i}\sigma_{\widetilde{X}_j})$. This is clearly an advantage in this case as all the dependent variables (flows) are measured in the same unit, and, as Section IV underlines, this choice allows a more intuitive interpretation of the principal components. Furthermore, as [23] points out, carrying out PCA on the covariance matrix can better isolate the strongest variations in a dataset with uniform units.

Finally, it is important to remark that data might present significant trends when longer datasets (i.e., spanning several years) are employed. For example, an increase in demand could introduce slow variations in power flows. In that case, simply centering the multivariate flow dataset as in (3) might reveal itself inadequate to remove such trends. To this end, one might filter the dataset of power flows so that low-frequency dynamics are discarded from the analysis.

B. Local polynomial regression

The model of Eq. (5), representing the power flows as a linear combination of the principal components of the dataset, can be modified in order to account for the effect of explanatory variables $\mathbf{u_t}$, e.g. the forecast wind power penetration $\widehat{r}_t$ and the transformed electricity price P_t^l. This is done by allowing the coefficients α_i of the principal components to vary as functions of $\mathbf{u_t}$

$$\widetilde{\mathbf{X}}_\mathbf{t} = \sum_{i=1}^{n} \alpha_i(\mathbf{u_t})\mathbf{Y_i} + \epsilon_t \qquad \forall t \ . \tag{6}$$

In this work local polynomial regression, see e.g. [20], is employed in order to study the functional forms of the $\alpha_i(\mathbf{u_t})$ coefficients. This technique allows to fit a curve or a surface (depending on whether $\mathbf{u_t}$ is formed by one or more explanatory variables) to these relationships by locally approximating them as low-order polynomials. Although in principle $\mathbf{u_t}$ could be of any dimension m, for practical applications this vector should be sized reasonably. For example it is not possible to visualize the coefficients $\alpha_i(\mathbf{u_t})$ if $m > 2$. Furthermore, the computational time increases with the dimension of this vector.

The first step of the technique consists in the definition of a grid in the space of the explanatory variable $\mathbf{u}$. The grid is formed here by choosing l equally spaced quantiles in each dimension of $\mathbf{u}$. Let us indicate with $\mathbf{u}_i$ the time series of the i-th explanatory variable, sorted in increasing order. We are interested in the quantiles with the following probabilities

$$p_k = \frac{k}{l} - \frac{1}{2l} \qquad k = 1, \dots, l \ . \tag{7}$$

Let us define $h_k = Tp_k + 1/2$, where T is the sample size. If h_k is an integer, the k-th quantile u_i^k is the h_k-th point in the sorted sequence $\mathbf{u}_i$. Otherwise, it can be estimated with the following linear interpolation

$$u_i^k = u_{i,\lfloor h_k \rfloor} + (h_k - \lfloor h_k \rfloor)\left(u_{i,\lfloor h_k \rfloor+1} - u_{i,\lfloor h_k \rfloor}\right) \ , \tag{8}$$

where the second subscript on the u's indexes a certain element of the vector $\mathbf{u}_i$.

A grid is then formed by considering the l^m combinations of points $(u_1^{k_1}, u_2^{k_2}, \dots, u_m^{k_m})$ where $k_i = 1, 2, \dots, l$ for $i = 1, 2, \dots, m$. Alternatively, equally spaced u_i^k could be used with no consequences on the remainder of the methodology presented here.

Weighted least-squares regression is then performed locally at each point of the grid. A set of q data points, where $1 \leq q \leq T$, is used for regression. These points, called *neighbors*, are the q closest points to the considered point of the grid. Naturally the neighbors selection procedure requires that a suitable metric ρ is defined on the explanatory variable space. Hereafter the Euclidean distance is adopted, after variables measured in different units are normalized. The ratio $h = q/T \leq 1$ between the considered number of neighbors and the total number of data points is referred to as *bandwidth*. High bandwidths increase the smoothness of the regression, with the trade-off of an increased bias of the regressed model. In this work a bandwidth $h = 0.2$ is used.

A weight function has to be defined in order to assign higher importance to the observations $(\widetilde{\mathbf{X}}_\mathbf{t}, \mathbf{u}_t)$, whose values of the explanatory variables are the closest neighbors of the grid point. Among the many possibilities, a weight function based on the tricube function is chosen here

$$w(z) = \begin{cases} (1-z^3)^3 & 0 \leq z < 1 \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{9}$$

One should notice that $w(z)$ is non-increasing for positive z. Let us name $\mathbf{u}_\#^j$ the j-th point of the grid and with $\overline{\mathbf{u}}^j$ its q-th furthest neighbor. Since $\rho(\overline{\mathbf{u}}^j, \mathbf{u}_\#^j)$ is at a maximum in the considered neighborhood, the weight function

$$f^j(\mathbf{u}) = w\left(\frac{\rho(\mathbf{u}, \mathbf{u}_\#^j)}{\rho(\overline{\mathbf{u}}^j, \mathbf{u}_\#^j)}\right) \tag{10}$$

is well defined. Indeed it assigns non-increasing weights to points with increasing distance from $\mathbf{u}_\#^j$. A weight of 1 is assigned to the grid point $\mathbf{u}_\#^j$, while $\overline{\mathbf{u}}^j$ and all the points outside the neighborhood have 0 weight.

Weighted least squares regression, see [24], is employed locally for each point $\mathbf{u}_\#^j$ of the grid. Each observation $(\widetilde{\mathbf{X}}_\mathbf{t}, \mathbf{u}_t)$ is weighted according to the weight function in (10). The output of weighted least squares regression is a local model for $\alpha_i(\mathbf{u_t})$, approximated as a first order polynomial of the explanatory variables. The model is determined according to a weighted least squares criterion, which ensures that the modeling error ϵ_t is minimized in a consistent way.

After this procedure is carried out for all the points in the grid, a curve or surface $\widehat{\alpha}_i(\mathbf{u_t})$ is fitted from the individual local approximations. Such a fit models the behavior of the coefficients $\alpha_i(\mathbf{u_t})$ of the principal components in the whole space of interest as functions of the explanatory variables. Conclusions could be drawn directly from the shape of the regression surface of the $\alpha_i(\mathbf{u_t})$'s, provided that the principal components can be easily interpreted. As an alternative, one can choose to map the analysis back to the original space of the non-centered flows, obtaining the models $\widehat{\mathbf{X}}(\mathbf{u}_t)$

$$\widehat{\mathbf{X}}(\mathbf{u}_t) = \sum_{i=1}^{n} \widehat{\alpha}_i(\mathbf{u}_t)\mathbf{Y_i} + \bar{\mathbf{X}}_\mathbf{t} \ . \tag{11}$$

This way regression curves or surfaces are obtained for each individual flow in the dataset.

IV. RESULTS

This section presents the results obtained from the application of the method proposed in Section III to the dataset described in Section II. Clearly, analyses of this type depend heavily on the availability and on the quality of large datasets, which are not always publicly available. Although datasets for power flows, wind power production and load are available for certain electricity markets, e.g., PJM [25] and the European markets [26], the collection of such datasets require the coordination of a number of entities so as to ensure consistency in terms of sampling frequency, sampling time and time-span. Obviously, this is a limitation for the readers willing to perform a similar study, but unable to interact with the entities owning the data.

A. Principal component analysis

The results of PCA applied to the dataset of the cross-border flows show that it is possible to express most of the original variance with a limited number of principal components (PCs). By using the scree-graph method, see [21], the 8 PCs with highest variance are selected. The criterion used ensures that further inclusion of PCs would not increase sensibly the cumulative fraction of variance explained by the set.

Table I summarizes the characteristics of the selected components. Its second column reports the fraction of total dataset variance explained by each of the PCs individually, while the third column shows the cumulative fraction of variance jointly explained by selecting the first i PCs. It is seen that almost $2/3$ of the total variance are explained by the first 4 PCs. Furthermore, the selected set of 8 PCs explains 82.11% of the original variance, despite the dramatic reduction in size of the problem from the original 68 flows to 8 modes. In relation to the discussion on the non-normality assumption in Section III-A, it is stressed that different statistical descriptions, such as the comparison of the interquantile ranges of the flows and their residuals, showed a similar behavior as the one illustrated in Table I for the variance.

TABLE I
INDIVIDUAL AND CUMULATIVE FRACTION OF VARIANCE EXPLAINED BY THE i-TH PRINCIPAL COMPONENT AND BY THE PRINCIPAL COMPONENTS UP TO THE i-TH ONE RESPECTIVELY

Principal component	Individual fraction of variance [%]	Cumulative fraction of variance [%]
1	28.89	28.89
2	18.83	47.73
3	10.55	58.28
4	7.69	65.97
5	5.53	71.50
6	4.83	76.32
7	2.97	79.29
8	2.81	82.11

The analysis of the structure of the PCs gives further insight into the characteristics of the European power system. Indeed when PCA is carried out on the covariance matrix, see the discussion in Section III-A, the structure of the PCs often offers a physical interpretation. Let us denote with $\mathbf{Y_i}$ the i-th PC. It is a vector of 68 elements

$$\mathbf{Y_i} = [Y_{i,1} \dots Y_{i,68}]^T \ , \tag{12}$$

where $Y_{i,j}$ represents the weight of the j-th individual flow in the i-th PC. Large weights of the same sign on a PC show that the corresponding flows tend to deviate from their mean in the same direction; conversely they tend to deviate in opposite directions if their weights have different signs.

Fig. 1 shows the structure of the first PC as an example. In this illustration the widths of the arrows in the map of Europe are scaled, so that each of them is proportional to the weight $Y_{1,j}$ of the respective j-th cross-border flow in the first PC. As one can see, the main contribution to this first PC is given by the simultaneous flow of power directly from Switzerland to Germany, directly from France to Germany and by the flow from France to Germany through Belgium and the Netherlands. The fact that this mode alone explains almost 30% of the variation of the dataset signals the importance of the power flow between France and Germany on a European scale.

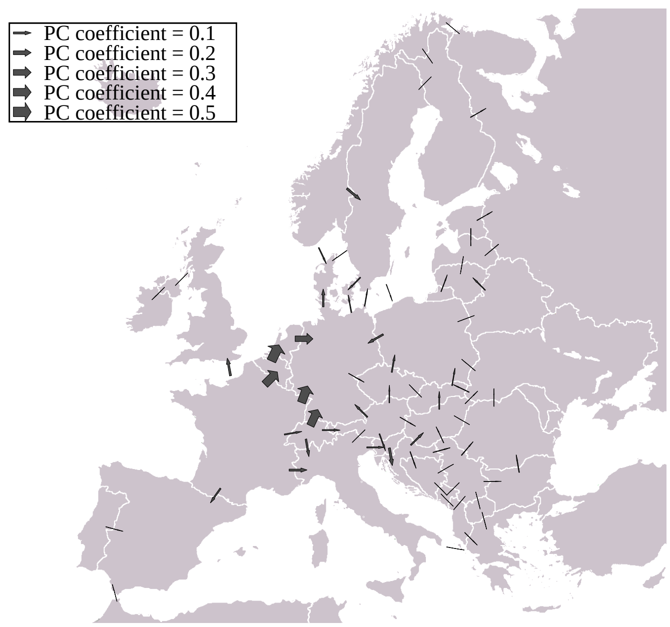

Fig. 1. Map of Europe showing the weight of each single centered physical flow $\tilde{\mathbf{X}}_\mathbf{i}$ in the first principal component $\mathbf{Y_1}$. The width of the lines in the figure is proportional to the coefficient of the respective flow in $\mathbf{Y_1}$

The second and third PC, not shown here for the sake of brevity, offer interesting interpretations, too. The second PC is mainly composed of power flowing from Germany to the Nordic region, and internally in Scandinavia from Sweden to Norway. A possible interpretation could be that cheap power is flowing from continental Europe to Norway so that water can be kept stored in the Norwegian hydro dams, or the other way around when continental Europe imports power from Norway. A second significant pattern in this principal component, although less important, is the flow of power through Switzerland in the North to South direction. Furthermore, the main trend in the third PC is the flow of power towards Italy from France, both directly and through

Switzerland, and from Germany, through Switzerland and to a lesser extent Austria.

B. Regression curves for principal components

Regression curves or surfaces modeling the behavior of the $\alpha_i(\mathbf{u}_t)$ coefficients in (6) as functions of the multivariate input $\mathbf{u}_t$ are readily obtained by using local polynomial regression.

Fig. 2 shows the curve modeling the behavior of the $\alpha_1(\widehat{r}_t)$ coefficient as a function of wind power penetration in Germany. The curve clearly represents the mean trend of the relationship between the two variables, and not a deterministic model of them. Therefore one should expect observations to be spread around this curve, due to their stochasticity and dependency on other variables not accounted by the model. Nevertheless one can draw some intuitive conclusions from this mean trend, also as a result of the interpretability of the first principal component. In Section IV-A it is underlined how the main trend in this mode is the power flow from France and Switzerland to Germany. As one can see in Fig. 2, the corresponding coefficient tends to decrease rather sensibly when wind power penetration in Germany increases. The implication is that the higher $\widehat{r}_t$ in Germany, the lower its import of power[1].

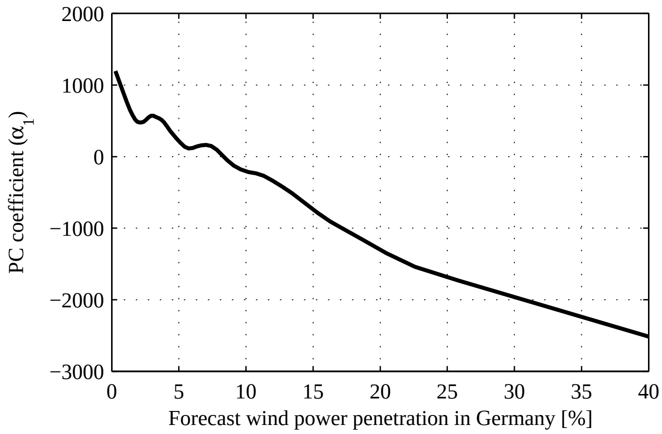

Fig. 2. Regression on the coefficient $\alpha_1(\mathbf{u}_t)$ of the first principal component relative to wind power penetration in Germany

Similar conclusions can be drawn from the regression on the coefficients of the other PCs, which are not shown here. For instance, the value of α_2 rises as $\widehat{r}_t$ increases, indicating an increased flow of power from Germany to the Nordic region and, less markedly, to Switzerland.

Regression curves modeling the behavior of the PC coefficients as a function of the logarithmic spot price in Germany can be obtained in exactly the same fashion. Fig. 3 shows the α_2 coefficient modeled as a function of this independent variable. The results can again be interpreted quite intuitively. Indeed, the regression curve shows that high values of flow from Germany to Scandinavia are in average achieved with low spot price level at the EEX market. The coefficient then decreases as prices rise in Germany, and its sign changes when $\log(1+P_t)$ approaches 4.

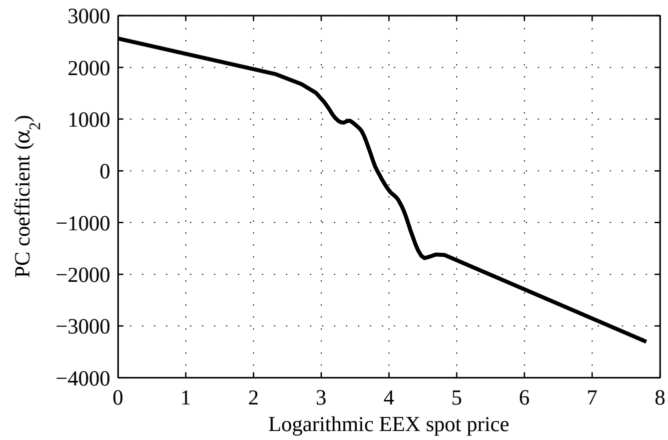

Fig. 3. Regression on the coefficient $\alpha_2(\mathbf{u}_t)$ of the first principal component relative to the logarithm of spot price in Germany

The approach is easily extendable to the case where the independent variable $\mathbf{u}_t$ is multivariate. Fig. 4 shows the regression surface modeling $\alpha_1(\mathbf{u}_t)$ as a function of $\mathbf{u}_t = [\widehat{r}_t, h_t]$, where h_t is the day-time. The latter variable appears to influence the coefficient, too, as higher flow values are obtained during hours where consumption peaks. Not surprisingly the decreasing trend relative to wind power penetration is confirmed at every hour of the day.

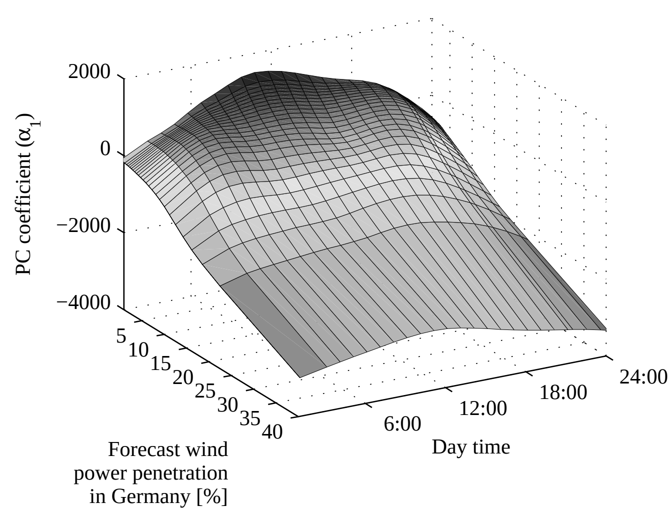

Fig. 4. Regression on the coefficient $\alpha_2(\mathbf{u}_t)$ of the first principal component relative to wind power penetration in Germany and day-time

C. Regression curves for power flows

So far only intuitive conclusions based on the structure of the PCs have been drawn, since the regression was carried

[1]It is to be noted that at this time only qualitative conclusions can be drawn. For example, one cannot distinguish when Germany is importing or exporting. Therefore, the statement could be rephrased to "the higher $\widehat{r}_t$ in Germany, the higher its export of power".

out on their coefficients α_i. By applying (11) it is possible to perform similar analyses on the space of interest, i.e. the original space of power flows. Indeed regression curves or surfaces for the coefficients sum up to curves and surfaces for each single flow.

Fig. 5 shows the regression for the flow between the Danish DK1 area, i.e. the Jutland peninsula and the Funen island, and Norway (NO). The surface models the relationship between this flow and wind power penetration in Germany as well as day-time. It is seen that, as wind power penetration increases, DK1 passes from importing power from Norway to exporting power. Therefore the statistical model confirms the intuitive economic reasoning according to which the flexible Norwegian hydro plants withhold their production when energy prices are low due to significant wind power penetration, and increase their production when prices are high.

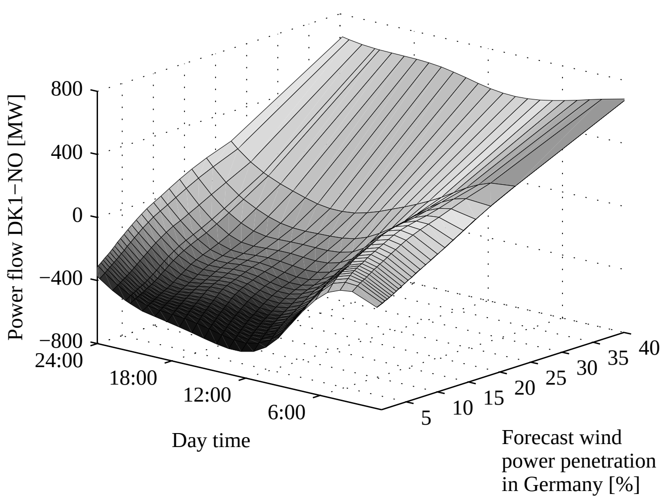

Fig. 5. Regression curve modeling the relationship between the power flow from the continental part of Denmark (the Jutland peninsula and the Funen island, area code DK1) and Norway (NO) as function of wind power penetration in Germany and day-time

Models for the total net power flow of a country or control area can be determined as the signed sum of the models for individual flows. This can help shed some light on the behavior of individual power systems as a reaction to increased wind power production. Fig. 6 shows the example of the Austrian power system. It is seen that Austria is on average a net power exporter for low levels of wind power penetration in Germany, while it is a net importer with high wind power penetration. This is partly caused by the flexibility of the Austrian generation portfolio, which is largely dominated by hydro plants. Once more, the statistical model shows how the market pushes hydro power producers to provide arbitrage services as wind power production increases the volatility of market prices.

D. Sensitivity of Results to Non-Stationarity

As mentioned in the last paragraph of Section III-A, non-stationarity of data might be an important practical issue when performing analyses based on PCA. Indeed, the latter methodology as defined in Section III-A is based on the

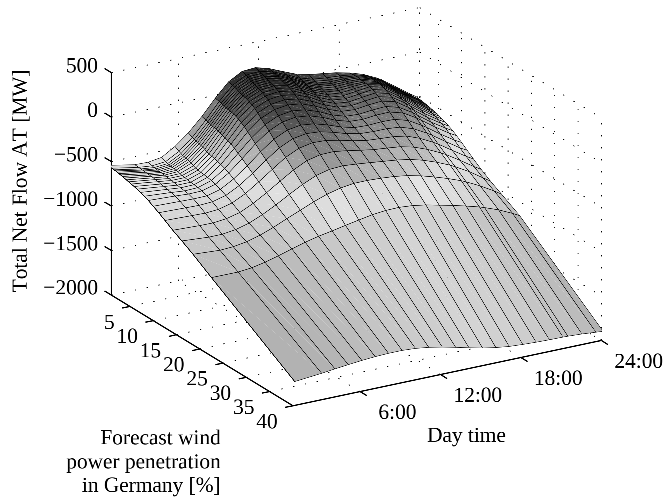

Fig. 6. Regression curve modeling the relationship between the net power flow of Austria as function of wind power penetration in Germany and day-time

stationarity assumption, as data are centered by subtracting the mean from the dataset. When considering long datasets spanning several years, low frequency dynamics ought to be removed from the dataset by employing a low-pass filter.

Some relevant figures for the considered period 2006–2008 in Germany are reported in Table II. Such table is useful to get a preliminary assessment of the presence and of the magnitude of trends. As one can see, despite a constant increase in installed wind power capacity in the period, the total annual wind power output in Germany declined between 2007 and 2008. Besides, the growth in electricity demand almost stopped between these two years. As a result, the ratio between these two quantities is not monotonic in the considered period.

TABLE II
TOTAL ANNUAL WIND ENERGY PRODUCTION, TOTAL ANNUAL LOAD AND THEIR RATIO IN GERMANY DURING THE PERIOD 2006–2008

Quantity	2006	2007	2008
Total wind energy production [TWh]	34.31	42.36	41.67
Total load [TWh]	489.03	496.59	497.61
Ratio	0.0702	0.0853	0.0837

Table II illustrates that there is no steady, slow-dynamics increase in the ratio between total demand and wind power production as one may expect intuitively in light of the constant increase in installed wind power capacity. The significant swings in the total annual wind power production, e.g., between 2006 and 2007, might compromise the stationarity assumption, hence calling for a more detailed analysis of the models on a year-to-year basis.

A measure to validate the models (11) obtained for the flows is the Normalized Root Mean Squared Error (NRMSE), which can be calculated for each interconnection j as follows:

$$\mathrm{NRMSE}_j = \frac{\sqrt{\frac{\sum_{t=1}^{T}\left(\widehat{X}_j(\mathbf{u}_t) - X_{j,t}\right)^2}{T}}}{\max_t\{X_{j,t}\} - \min_t\{X_{j,t}\}} \, . \tag{13}$$

The value in the numerator in (13) is the Root Mean Squared Error (RMSE), i.e., the average squared deviation of the flow model $\widehat{X}_j(\mathbf{u}_t)$ from the actual observations. The term in the denominator is a scaling factor normalizing all the flows to the range of their observations.

In order to assess the level of non-stationarity in the dataset, NRMSE is calculated for each flow in the dataset, first by employing data for the whole period 2006–2008, then considering data for one single year at a time. This can be done simply by modifying the time indices in the sum in (13).

The results of this analysis are shown in Fig. 7. The average value of NRMSE across the flows is roughly 15-20%. More interestingly, NRMSE seems to be rather stable across the considered years for each interconnection. This indicates that non-stationarity in the dataset is not so critical for the application considered. Indeed, one would expect that, if data are highly non-stationary, the performance of the models swing significantly across subsets of the dataset.

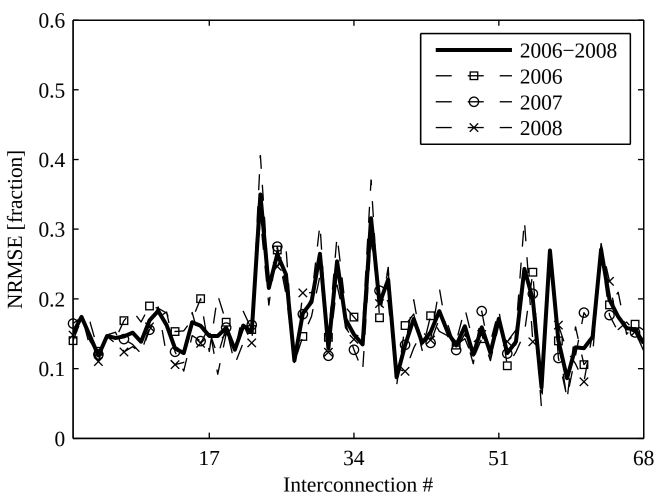

Fig. 7. Normalized Root Mean Squared Error (NRMSE) for every interconnection for the whole 2006–2008 period, and for each year in the period individually

E. Regional Analysis

The models shown up to this point consent an analysis of the effect of explanatory variables on each individual flow or country. We now group the results obtained for individual interconnections in different geographical regions: North, South, West and East of Germany.

Table III reports the average value of the models for the flows obtained for Northern Europe for values of wind power penetration in Germany from 0% to 25% with step increases of 5%. Spline interpolation of the models at the desired wind power penetration levels was required, due to the definition of the grid in the explanatory variable space, which is based on quantiles. For the sake of consistency, all the flows are directed in the South to North direction (so that negative flows indicate flow of power southwards). The pattern emerging from the analysis is quite clear. At low levels of wind power penetration, Germany is importing power from the Nordic countries through any available interconnections, where power flows in the North to South direction indeed. This North-South flow tends to drop, though, as wind power penetration increases in Germany. For example, the average flow in the two interconnections between Denmark and Sweden is reversed at a penetration as low as 10%, also due to a likely increased production from the Danish wind turbine fleet. The same trend is seen for all the flows shown in Table III. When penetration reaches 25%, power is flowing in the South to North direction in all the interconnections except the one between Germany and DK1, where the percentage of installed wind power on the total production capacity is even higher than in Northern Germany.

TABLE III
MODELED AVERAGE CROSS-BORDER FLOWS NORTH OF GERMANY AS FUNCTIONS OF GERMAN WIND POWER PENETRATION

Average Flow	Wind Power Penetration [%]					
[MWh/h]	0	5	10	15	20	25
DE-DK1	-803	-485	-410	-465	-372	-199
DE-DK2	-276	-86	-39	-63	-24	55
DE-SE	-276	-102	-46	-61	-13	67
DK1-NO	-636	-255	-119	-83	40	179
DK1-SE	-228	-54	14	23	64	116
DK2-SE	-510	-117	44	76	210	356

TABLE IV
MODELED AVERAGE CROSS-BORDER FLOWS SOUTH OF GERMANY AS FUNCTIONS OF GERMAN WIND POWER PENETRATION

Average Flow	Wind Power Penetration [%]					
[MWh/h]	0	5	10	15	20	25
DE-CH	759	1390	1733	2096	2435	2680
CH-IT	3064	2715	2917	3206	3277	3213
DE-AT	212	423	565	728	930	1136
AT-CH	422	545	618	697	774	836
AT-IT	158	156	161	166	168	168

Let us now consider the flows directed southwards from Germany, included in Table IV. It is clear that the higher the wind power penetration in Germany, the more this country exports to its direct neighbors to the South: Switzerland and Austria. As opposed to the situation in the Nordic region, the trend seems to stop at the direct neighbors. Indeed, there is no clear pattern in the interconnection between Switzerland and Italy, which is somewhat stable at 3000 MWh/h. Furthermore, although the power flow between Austria and Italy increases, this trend is marginal due to the low capacity of this line. A possible explanation for this phenomenon is the high installed hydro power capacity in Switzerland and Austria, which confers extra flexibility to their power systems in comparison to e.g. Denmark, which is similarly located in the middle between Germany and the Nordic region. Finally, it appears that there is a loop-flow in the power transit from Germany to Switzerland through Austria, as the flow from the last to the second country is positively correlated with German wind power penetration.

Clear trends emerge as well from the analysis of power flows to the West of Germany in Table V. At null wind power penetration in Germany, the Netherlands imports power both from this country and from France through Belgium. When

TABLE V
MODELED AVERAGE CROSS-BORDER FLOWS WEST OF GERMANY AS FUNCTIONS OF GERMAN WIND POWER PENETRATION

Average Flow [MWh/h]	Wind Power Penetration [%] 0	5	10	15	20	25
DE-NL	1583	2131	2316	2498	2580	2582
NL-BE	-352	132	313	568	810	953
BE-FR	-1284	-852	-685	-452	-173	55
DE-FR	-1919	-1571	-1510	-1368	-1175	-1029
DE-CH	759	1390	1733	2096	2435	2680
FR-CH	463	760	916	996	1084	1174

penetration reaches 5%, the Netherlands only import power from Germany on average, as the direction of the flow to Belgium is reversed. In the same fashion, the French export of power to Belgium drops, even turning into a slight import when wind power penetration reaches 25% in Germany. This country imports power from France in all the cases shown in Table V, but this import is gradually halved from about 2000 MWh/h to 1000 MWh/h at the extreme columns in the table. This fact disproves the belief that there is a loop flow carrying power from the North to the South of Germany through France, at least on a country level. According to Table V, the loop flow appears to be a bit souther than that. Indeed the power flow from France to Switzerland appears to increase as German wind power penetration rises, while at the same time the direct import of Switzerland from Germany is heightened, as we already commented on earlier.

The situation in Eastern Europe is summarized in Table VI. The trends in this region are more complex, also due to the huge number of cross-border flows. Proceeding in the analysis from North to South, we see that there is an increasing power export from Germany to Poland as wind power penetration rises in the former country. Nevertheless, most of the extra power imported by Poland is exported in turn to Slovakia and the Czech Republic. The latter country, a net exporter of energy on average, sees an increasing part of its export shifting to Austria and, to a lesser extent, to Slovakia from Germany, as wind power penetration rises in this country. The increased power imported from North by Slovakia is then exported in the South direction through Hungary and Croatia to Slovenia. Finally it is seen that Slovenia imports less and less power from Austria, and exports more and more power to Italy as the German wind power penetration increases.

TABLE VI
MODELED AVERAGE CROSS-BORDER FLOWS EAST OF GERMANY AS FUNCTIONS OF GERMAN WIND POWER PENETRATION

Average Flow [MWh/h]	Wind Power Penetration [%] 0	5	10	15	20	25
DE-PL	414	414	473	582	660	700
PL-CZ	779	819	885	951	947	888
PL-SK	312	330	375	425	452	451
CZ-DE	877	1033	1026	958	902	852
CZ-AT	589	655	720	799	848	874
CZ-SK	784	680	736	834	865	850
SK-HU	801	897	987	1078	1137	1163
HU-HR	488	614	686	759	826	873
HR-SI	113	381	474	531	597	641
SI-IT	365	477	525	561	577	575
SI-AT	-176	-44	-23	-19	19	68

V. CONCLUSION

In this work a statistical method for analyzing the impact of wind power in Germany on European cross-border power flows is presented and applied.

The problem dimension is successfully reduced by applying Principal Component Analysis (PCA). Besides, PCA indicates the most important modes of physical power flow in the European system. These modes are the flows carrying power from France to Germany, both directly and through Belgium and the Netherlands, the flow from Germany to Scandinavia and the one from Germany in the South direction. This confirms the centrality of Germany in the study of the European power system.

Local polynomial regression is employed on the PCs both with respect to forecast wind power penetration and spot price in Germany. It is shown that both the external variables have a remarkable impact on the flows. Indeed an increase in forecast wind power penetration causes a fall in the German import of power (or rise in the export), while rising spot prices have the opposite effect. Furthermore, especially in the case of EEX spot price, non-linearities are evident in these relationships.

From a global perspective, it is seen that variations of wind power penetration in Germany have significant effects on power flows in Europe. Indeed import and export patterns between countries change significantly, and loop flows originate. Furthermore, while some of the interconnections benefit from an increasing forecast wind power penetration in Germany, i.e. the ones linking the main average exporters to Germany (France and, at least at low wind power levels, Scandinavia), the stress on other interconnections, e.g. the ones linking Germany to the South of Europe, increases as more and more wind power is produced in Germany. Analyses like the one presented in this paper can contribute to the state-of-the-art by quantitatively assessing such global phenomena, whose understanding is currently limited to a qualitative or intuitive level.

Clearly, the study presented in this paper could be performed thanks to the availability of datasets for wind power production, consumption and power flows, which is in general not straightforward, with the notable exceptions of the PJM market [25] and ENTSO-E [26]. It is hoped that in the near future, convinced by the results of data-driven research studies like this one, TSOs and market operators will strengthen their effort to make more and more datasets of this type available to researchers worldwide.

Possible applications of the methodology proposed in this paper are related to both long- and short-term problems. Long-term problems that could benefit from analyses of this type include decisions on investment in new wind power capacity and in grid expansion. As far as the short-term problems are concerned, this methodology could support the process of scheduling cross-border flows as well as the assessment of risk in interconnected power systems.

The described methodology has been employed for analysis only. In the future though, such a modeling approach could also be used in connection with power system models or forecasting tools. Such tools would have to be calibrated with the results of data-driven analysis, as the one shown in this paper, for simulating European power flows as a function of appropriate explanatory variables, which could include the nominal power capacity of a certain country, market prices, etc.

Acknowledgment

The authors gratefully acknowledge Austrian Power Grid AG (APG) for the support through the project "Impact of Stochastic Generation on EU Cross-border Flows". The European Network of Transmission System Operators for Electricity (ENTSO-E) is also acknowledged for its role in providing the dataset. Furthermore, we would like to thank the Editor of this Journal and three anonymous referees for providing constructive comments that undoubtedly contributed to improving the quality of this manuscript. Finally, we thank Peter Meibom for his valuable comments on this work.

References

[1] P. Morthorst, S. Ray, J. Munksgaard, and A. F. Sinner, "Wind energy and electricity prices," European Wind Energy Association, Tech. Rep., 2010. [Online]. Available: http://www.ewea.org/fileadmin/ewea_documents/documents/publications/reports/MeritOrder.pdf

[2] P. Giabardo, M. Zugno, P. Pinson, and H. Madsen, "Feedback, competition and stochasticity in a day ahead electricity market," *Energy Econom.*, vol. 32, no. 2, pp. 292–301, 2010.

[3] T. Jónsson, P. Pinson, and H. Madsen, "On the market impact of wind energy forecasts," *Energy Econom.*, vol. 32, no. 2, pp. 313–320, 2010.

[4] H. Abildgaard, D. Klaar, B. Kriszak, J. Rodriguez, and W. Winter, "European Wind Integration Study (EWIS) - Reference study towards a successful integration of wind power into European electricity grids," in *Proc. CIGRE Sess.*, Paris, France, 2008.

[5] "ENTSO-E pilot ten years network development plan," ENTSO-E, Tech. Rep., 2010. [Online]. Available: https://www.entsoe.eu/fileadmin/user_upload/_library/SDC/TYNDP/TYNDP-final_document.pdf

[6] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renew. Sust. Energy Rev.*, vol. 12, no. 6, pp. 1725–1744, 2008.

[7] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, "The state-of-the-art in short-term prediction of wind power : A literature overview, 2nd edition," ANEMOS.plus, Tech. Rep., 2011. [Online]. Available: http://www.prediktor.dk/referenc.htm

[8] G. Giebel, "A variance analysis of the capacity displaced by wind energy in europe," *Wind Energy*, vol. 10, no. 1, pp. 69–79, 2007.

[9] U. Focken, M. Lange, K. Mönnich, H.-P. Waldl, H. G. Beyer, and A. Luig, "Short-term prediction of the aggregated power output of wind farms–a statistical analysis of the reduction of the prediction error by spatial smoothing effects," *J. Wind Engin. Industr. Aerodyn.*, vol. 90, no. 3, pp. 231–246, 2002.

[10] F. van Hulle, P. Kreutzkamp, and S. Uski-Joutsenvuo, "Enhancing cross border exchange to facilitate wind power integration at European scale," in *Proc. German Wind Energy Conf. (DEWEC)*, Bremen, Germany, nov 2008.

[11] J. Tande, M. Korpås, L. Warland, K. Uhlen, and F. van Hulle, "Impact of TradeWind offshore wind power capacity scenarios on power flows in the European HV network," in *Proc. 7th Int. Wind Integration Workshop*, Madrid, Spain, may 2008.

[12] S. Hagspiel, A. Papaemannouil, M. Schmid, and G. Andersson, "Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid," *Appl. Energy*, vol. 96, pp. 33–44, 2012.

[13] N. A. Cutululis, P. Sørensen, G. Giebel, M. Korpas, and L. Warland, "Uncertainty on predicted cross border flows caused by wind forecast errors," in *Proc. 7th Int. Wind Integration Workshop*, Madrid, Spain, may 2008.

[14] B. Klöckl and P. Pinson, "Effects of increasing wind power penetration on the physical operation of large electricity market systems," in *Proc. CIGRE/IEEE PES Jt. Symp. Integr. Wide-Scale Renew. Resour. Power Deliv. Syst.*, 2009.

[15] "Energy concept for an environmentally sound, reliable and affordable energy supply," Federal Ministry of Economics and Technology (BMWi), Tech. Rep., 2010. [Online]. Available: http://www.bmwi.de/English/Navigation/Service/publications,did=367764.html

[16] "Final report system disturbance on 4 November 2006," UCTE, Tech. Rep., 2007. [Online]. Available: https://www.entsoe.eu/resources/publications/former-associations/ucte/other-reports/

[17] DEWI, "Statistics archive," Online, 2012. [Online]. Available: http://www.dewi.de/dewi/index.php?id=47&L=%5C%5C%5C%27

[18] A. León and A. Rubia, *Modelling prices in competitive electricity markets*. Wiley & Sons, 2004, ch. 8, pp. 177–189.

[19] R. Weron, *Modeling and Forecasting Electricity Loads and Prices–A statistical Approach*. Wiley, 2006, ch. 4.

[20] W. Cleveland and S. Devlin, "Locally weighted regression: An approach to regression analysis by local fitting." *J. Amer. Statist. Assoc.*, vol. 83, no. 403, pp. 596–610, 1988.

[21] A. C. Rencher, *Multivariate statistical inference and applications*. Wiley-Interscience, 1998, ch. 9.

[22] A. Hyvärinen, J. Karhunen, and O. Erkki, *Independent Component Analysis*. Wiley, 2001.

[23] D. S. Wilks, *Statistical methods in the atmospheric sciences, 2nd edition*, ser. International Geophysics. Elsevier Academic Press, 2006, vol. 91, ch. 11, pp. 469–471.

[24] H. Madsen, *Time series analysis*. Chapman & Hall/CRC, 2008, ch. 3.

[25] PJM Electricity Market, Website, 2012. [Online]. Available: http://www.pjm.com/home.aspx

[26] ENTSO-E, Website, 2012. [Online]. Available: http://www.entsoe.net/default.aspx

Marco Zugno (S'11) received the M.Sc. degree in Electrical Engineering from the Technical University of Denmark (DTU) in 2008 and the M.Sc. degree in Automation Engineering from the Università degli Studi di Padova, Padua, Italy, in 2009. He is currently pursuing the Ph.D. degree at the Informatics and Mathematical Modeling department of the Technical University of Denmark. His research interests include electricity market modeling, stochastic programming and hierarchical optimization.

Pierre Pinson (M'11) received the M.Sc. degree in Applied Mathematics from the National Institute for Applied Sciences (INSA Toulouse, France) in 2002 and the Ph.D. degree in Energetic from Ecole des Mines de Paris in 2006. He is currently with the Informatics and Mathematical Modeling department of the Technical University of Denmark as an Associate Professor. His research interests include among others forecasting, uncertainty estimation, optimization under uncertainty, decision sciences, and renewable energies.

Henrik Madsen received the Ph.D. degree in statistics at the Technical University of Denmark, Kgs. Lyngby, Denmark, in 1986. He was appointed Professor in mathematical statistics in 1999. He is an elected member of the International Statistical Institute (ISI), and he has participated in the development of several ISO and CEN standards. His research interests includes analysis and modeling of stochastic dynamics systems, signal processing, time series analysis, identification, estimation, grey-box modeling, prediction, optimization, and control, with applications mostly related to energy systems, informatics, environmental systems, bioinformatics, biostatistics, process modeling, and finance.