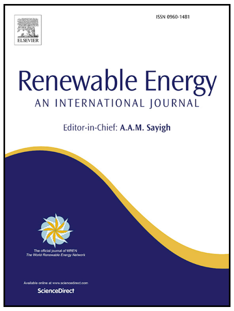
Accepted Manuscript



A novel method to optimize electricity generation from wind energy

E.E. Vogel, G. Saravia, S. Kobe, R. Schumann, R. Schuster

PII: S0960-1481(18)30375-6
DOI: 10.1016/j.renene.2018.03.064
Reference: RENE 9933

To appear in: *Renewable Energy*

Received Date: 21 October 2016
Revised Date: 25 February 2018
Accepted Date: 24 March 2018

Please cite this article as: Vogel EE, Saravia G, Kobe S, Schumann R, Schuster R, A novel method to optimize electricity generation from wind energy, *Renewable Energy* (2018), doi: 10.1016/j.renene.2018.03.064.

A novel method to optimize electricity generation from wind energy

E.E. Vogel[a,b,e], G. Saravia[a], S. Kobe[c], R. Schumann[c], R. Schuster[d]

[a] *Department of Physics, Universidad de La Frontera, Casilla 54-D, Temuco, Chile*

[b] *Center for the Development of Nanoscience and Nanotechnology, Chile*

[c] *Institut für Theoretische Physik, Technische Universität Dresden, D-01062 Dresden, Germany*

[d] *D-35759 Driedorf, Germany*

[e] *Corresponding author: eugenio.vogel@ufrontera.cl*

HIGHLIGHTS

- Data recognizer wlzip is used to anticipate favorable periods of wind energy
- The method can also be used to analyze wind energy production
- Data from all German turbines during 2010:2017 is used in this study
- A protocol for mixing wind energy with conventional sources is proposed
- Protocol indicators are tested on monthly basis during the eight-year period

---

Abstract

We present and discuss a new technique based on information theory to detect in advance favorable periods of wind activity (positive ramps) for electricity generation. In addition this technique could also help in the analysis of plant operation and management protocols design. Real data from wind power plants in Germany is used; this information is freely available in the internet with reliable registers every 15 minutes. A simple protocol to mix such wind energy production with electricity coming from conventional sources is proposed as a way to test the proposed algorithm. The eight-year period 2010-2017 is analyzed looking for different behaviors in wind activity. The first five years (2010-2014) are employed to calibrate the method, while the remaining three years (2015-2017) are used to test previous calibration without any further variation in the tuning possibilities described below.

Thus, the proposed protocol is tried on under different seasonal wind conditions. Both the algorithm and the general protocol could be adjusted to optimize performances according to regional conditions. In addition, this algorithm can also be used in retrospective studies to adjust productivity to operational conditions.

Preprint submitted to Elsevier *26 March 2018*

KEYWORDS: Wind energy, wind ramps, energy management, information theory, optimization

1 Introduction

Wind energy production (WEP) contributes in a moderate way to the total electricity generation in most of the countries in the world. We will concentrate here on Germany where 16% of the total electricity production was due to wind during 2017 according to recent statistics [1]. In spite of its present modest contribution WEP is basically free (except for low operation costs) and its installed capacity grows steadily in most countries. Then, it is possible to imagine that wind farms will play a very important role in the future electricity generation all over the world.

During the last five years the percentage of WEP has doubled in Germany. Thus, in 2017 more than 28 000 wind turbines onshore and 1000 wind turbines offshore have reached a productivity of about 105 TWh of electrical energy [1]. This development is part of a political program of the German government with the aim of transforming the future energy system to a higher use of renewable energies (RE) substituting for nuclear and fossil sources.

Because of the limited predictability of wind power the feed-in management into the national electricity system faces major challenges. According to the usual rules in Germany the generation of electricity by RE sources has priority which should be sustained by protocols that can guarantee such management. Consequently, the combination of other forms of energy (the production based on conventional sources and imports) with RE requires a reliable protocol to secure the power balance according to the required load. However, the availability of wind energy in Germany (or any country) varies enormously throughout the year, even from one week to next with abrupt changes within hours. At present, often power gradients of the order of 1 GW/h have to be managed. Coal and nuclear power plants are designed to work in a continuous operation regime for the purpose of ensuring the base load. Their flexibility is limited and complete shut down is undesired especially in the case of brown coal. Moreover, in the working regime a minimum power generation should not be undershot. The quality improvement of wind power prediction can contribute to reduce the shutoff times of the wind contribution during high wind periods in order to prevent overload of the power supply system. About 3500 GWh wind energy (several hundred million dollars) were lost in Germany during 2016 because of management problems [2].

There are basically three approaches to forecast wind activity intended for energy production [3–6]. First, those methods based on physical considerations to forecast the temporal development of the local wind speed [7–10]. Second, the methods based on time series with the assumption that the power output at any time depends on the previously observed values within a recent time range [11–14]; our present approach is along this line but we have replaced autoregressive models and neural network analysis by information recognition through data compressors techniques [15]. Third, hybrid approaches combining some of the previous methods by means of appropriate numerical algorithms [5,16].

Most of wind power prediction methods have been developed mainly for wind farms, namely, for local applications. In the present paper we follow a different approach by using the entire WEP for a country like Germany. From this point of view the work by Gonzalez-Aparicio and Zucker [17] is a bit related to our proposal in the sense that they looked at the data base for one country: Spain. However, their focus is more oriented towards the economical aspects of this problem. Our approach is to improve numerical and statistical methods to better mix wind power with other sources of energy.

The method we propose below is entirely new and it is based on the determination of the information content within a recent interval in the times series for WEP of a network of wind farms. To our knowledge, this is the first time information theory is applied to this problem. We propose to use the wind power production time series as the input to detect the onset of periods for good usage of wind energy, which are usually called positive ramps.

The information content determination is done by means of data compressor wlzip specially designed to recognize meaningful information within any sequence [15,18]. The tunable features of this algorithm have been adapted here to deal with the wind energy power data. This process will be fully discussed in Subsection 2.2. The use of the direct WEP data coming from the system under study ensures the response to the source of energy directly. Our approach is based on the actual data produced after the turbine operation, so the information content of this series reflect the real contributions of the turbines effectively connected to the system, with their efficiency and interconnection networks. This is advantageous for detecting changes of performance as compared to indirect information from time series previous to the turbine operation, like weather variables (wind velocity for instance). In addition, the application of this technique can be done in real time (hot) in parallel for different geographical places. In this way, networks can be locally optimized, favoring the saving of fuels where WEP is convenient.

Another new feature of the methodology introduced here is that performances can also be studied retrospectively in terms of the desired time spans: years, seasons, months, weeks. For the data analyzed below it turns out that Summer months (Northern Hemisphere) are hopeless, while during Winter months

WEP is high so the risk of overshoots is high, which could be prevented by detecting in advance a positive ramp. During Spring and Autumn months optimization is possible, which is precisely what it can be achieved by properly mixing WEP with other sources of energy.

As it will be discussed below, the anticipation for good periods of wind can be of a few hours and it could be adjusted to the season and local conditions. The main purpose of the present paper is to show the way this method can be applied to make better use of the electricity generated by wind turbines along two ways: anticipation of good productivity and seasonal analysis for future planning of WEP.

The method is based on information theory [19–21] which has been successfully used to detect phase transitions in magnetism [22–24,15], crisis in economical systems like stock markets [25] and pension funds [26], as well as in clinical variables like the blood pressure variations leading to hypertension diagnosis [27,28]. On the other hand, early results suggest that this method can also be adapted to seismology, in particular to finding indicators that can anticipate in a couple of years the approximate location of major earthquakes [29].

Energy data are public in Germany and can be obtained directly from the internet [30,31]. In any of these sources the entire WEP in all Germany is stored in registers every 15 minutes in an automatic and continuous way. In the present paper the data for the blue eight years: 2010-2017 is analyzed. The lustrum 2010-2014 is fully used to calibrate the several tunable capabilities of the information theory method presented here. Then, the three remaining years (2015-2017) are employed to retrospectively test the already calibrated method without further optimization, so to try its robustness.

We will begin next section by describing the way the data is handled and organized for the present study. Then we describe the methodology in a general way. Section 3 is for results and discussions. The first Subsection is devoted to the optimization of wlzip to the present problem; since this is the first time this method is applied to electricity production by wind turbines it requires calibration and tuning as any new instrument does. Then, in Subsection 3.2 we present an application of the method to anticipate good periods of WEP in combination with conventional sources. Subsection 3.3 goes onto yearly analyzes mostly intended to long run planning. In Section 4 we give the main conclusions of this work.

2 Methodology

2.1 Data organization

WEP data are updated every 15 minutes, namely, on the hour HH:00, then HH:15, HH:30, HH:45, (HH+1):00, and so on [30,31]. We will organize these data in yearly files beginning at 0:00 hours of that year and ending at 24:00 of December 31 that same year. This last register is the first register of next year and so on. Such sequence will be denoted by $P(t)$ and it represents the total instant power produced by all wind turbines connected to the generation of electricity in Germany. It is reported in megawatts (MW) with a production that at present reaches over 10 GW in the good periods. From this point of view $P(t)$ is stored in registers consisting of 7 or more digits: 5 of them correspond to the integer part, then we have the decimal point followed by two or more digits. However, the precision of the information is higher for the digits reflecting more energy than for the digits representing the smaller contributions, since there is no guarantee that the measurement in each wind turbine is done at the highest possible accuracy. So we will restrict ourselves to integer numbers in units of MW rounding up the decimal point in the usual way (equal or over .5 is approximated to the next integer). Examples are given in the third and fourth columns of Table 1, which will be fully explained below.

With the yearly data adjusted to five integer values in decimal numerical basis, registers are organized in files in the form of vectors: one entry per line. Then we have files with 35041 registers (lines) for years 2010, 2011, 2013 and 2014; the file for the leap year 2012 has 35137 lines. It should be noticed that these data cannot reflect local or regional variations of wind.

2.2 Information recognizer

Data compressor wlzip was created to recognize repeated meaningful information in a sequence of data, which is different to the recognition of repeated random information done by usual data compressors like rar or bzip2 among others. In spite of been registered as intellectual property it is offered free of charge upon request by email (eugenio.vogel@ufrontera.cl) [23]. Actually wlzip compacts less than other compressors. However, compressions done by wlzip are based on exact matching of data structures representing properties of the system. Thus, a high degree of compression indicates repetitive information, namely a system that does not change significantly its properties within the time window under consideration. On the other hand, a very low degree of compression means lack of repetitive information, namely a system that is constantly and abruptly changing its properties; in the extreme situation it could be approaching chaos.

Table 1. The first column enumerates the instants sequence every 15 minutes; the second column is a five-digit random sequence for WEP; the third column corresponds to the actual quiet WEP sequence around noon of Saturday March 23, 2013; the fourth column gives the actual agitated WEP sequence during the morning of Sunday January 27, 2013. The fifth column repeats the data of the fourth column in quaternary basis. All WEP powers expressed in integer units of MW. The last row gives the mutability value for each column.

Instant	Random	2013.03.23	2013.01.27	2013.01.27
		Quiet	Agitated	Agitated
	Decimal	Decimal	Decimal	Quaternary
1	35743	00685	06664	001220020
2	34993	00699	06703	001220233
3	34823	00742	06818	001222202
4	35143	00757	07099	001232323
5	35173	00734	07221	001320221
6	35123	00728	08632	002103120
7	34383	00703	09432	002121000
8	34463	00653	09792	002132213
9	34213	00643	10151	002132213
10	34803	00588	10557	002210331
11	33953	00561	10731	002213223
12	33603	00567	11289	002300121
13	33143	00563	11583	002310333
14	33153	00576	11808	002320200
15	33773	00616	12084	002330310
16	32703	00630	12411	003001323
17	31683	00635	12845	003020231
18	30223	00634	13523	003103103
19	29803	00594	13496	003102320
20	30663	00612	13657	003111121
21	31273	00594	14133	003130311
22	31003	00621	14426	003201122
23	31193	00650	14403	003201003
24	30953	00646	14872	003220120
25	35743	00656	14930	003221102
26	34993	00652	15045	003223011
27	34823	00674	15384	003300120
28	35143	00676	15637	003310111
29	35173	00646	15992	003321320
30	35123	00633	15980	003321230
31	34383	00653	16184	003330320
32	34463	00704	16611	010003203
$\mu_{32\_3\_3}$	1.0	0.076	1.545	1.545

The dynamical application of wlzip requires the definition of a time window which will be kept constant through the study. This is one of the several calibration processes to be done below. To decide upon the length of the time window we have to pay attention to the properties of the system as well as to the urgency of obtaining a useful answer. Of course the longer the time window τ (measured in number of instants, or number of quarters of an hour for the present data) the better the precision achieved in the compression. However, shorter τ values will make the method more effective in terms of anticipation to use the information soon to make decisions. In next Section we will present evidence showing that $\tau = 32$ (8 hours) is an appropriate time window; this will be the first fixed calibrated index. At this point we anticipate this result to continue with the presentation of the methodology.

We are now in position of defining the relative mutability of a time series at any given time t. The instantaneous sequence consists of 32 values: the WEP at the present instant and the 31 precedent ones in the original file. Let us compress this partial vector obtaining its "weight" in bytes $w^*(t,\tau)$. This value has not absolute meaning and it can vary depending on τ. To define a parameter which oscillates around 1.0 we define the relative mutability by dividing previous value by $W(\tau)$ which is the weight of a fixed file, with 32 random registers with 5 digits similar to those of the series P(t). Then the relative mutability $\mu(t,\tau)$ is simply given by the ratio

$$\mu(t,\tau) = \frac{w^*(t,\tau)}{W(\tau)} \ . \tag{1}$$

To put previous equation in operational terms let us turn now to Table 1, whose first column is just the ordinal number of the 32 instants considered for the specific time window identified at the heading of each column. The second column gives a possible random sequence of weight $W(\tau) = W(32)$ to be used here as a reference. Since mutability is a relative indicator any random sequence will cope with this purpose. The first digit (3) is constant and irrelevant; the second digit presents some variations while third, fourth and fifth digits show high dispersion behaving randomly. Actually registers in this column are arbitrary and they have no real significance since all mutability values will be referred to this same sequence all the time. It has been chosen so a relative mutability value less than one tells of a monotonous time series, while a μ_r value larger than one identifies a more agitated sequence; the subindex r identifies the calibration adjustment which will be discussed below.

The third column of Table 1 copies the 32 values of a calmed eight hours period of the day 2013.03.23 (using the notation year.month.day: YYYY.MM.DD). This vector of 32 values is analyzed by yielding a weight $w^*(t,32)$; the mutability is then obtained by taking the ratio over $W(32)$ just defined in previous paragraph. This is the value for $\mu_{32\_3\_3}$ reported in the bottom line of Table 1

(0.076 in this case). The fourth column lists the 32 values of an agitated eight hours period of the day 2013.01.27. The corresponding μ_r value is given in the last line. The fifth column expresses in quaternary basis the same decimal information given in the fourth column for reasons to be discussed below. All power data are given in MW, with 5 integer digits. It can be noticed that zeroes to the left are explicitly included here to emphasize that the digits in these positions could also to be recognized by wlzip.

2.3 Use of the information recognizer

As any instrument wlzip needs calibration and tuning. One of these features was already mentioned: the time window needed for dynamic measurements. Other important adjustable knob is the numerical basis used to express the information to be recognized. We are accustomed to the decimal basis that is used worldwide nowadays. However, this is not necessarily the most appropriate basis for any numeric information recognition.

It is possible to gain precision if we translate the data into a lower numerical basis thus increasing the number of digits used to express the same information. An example of this is presented in the fifth column of Table 1, where we give the same information of the fourth column except that now this is expressed in quaternary basis, namely, a basis of four digits only: 0, 1, 2 and 3. In this way a certain power production is expressed now with more digits than in the decimal basis; the recognition of repetitions can be done now at intermediate precisions which were not available with decimal basis. We will define b as the number of digits present in the basis used for the compression. The corresponding mutability is denoted by μb. Since we will use quaternary basis in the rest of this work we could omit the suffix 4, namely, $\mu = \mu 4$.

One interesting feature of wlzip is that the information recognition can be focused on the digits bearing the significant changes. For instance it is clear from columns 3 and 4 of Table 1 that the first of the five digits varies very little. The variations of the last two digits have relatively low significance. The significant variations are in digits second and third for these results in decimal basis. But we will use quaternary basis in the applications so let us turn our attention to the fifth column of Table 1, where we realize that the significant changes in the data begin at the third digit. This initial position (3) for the meaningful information is the second calibration which will be 3 from now on. In the next Section we will justify that it is enough to look for only the three digits to the right of the initial position (third, fourth and fifth digits). The number of recognizable digits is the next calibration parameter; 3 in the present case. This establishes the notation for the mutability in the last row of Table 1: $\mu_{32\_3\_3}$.

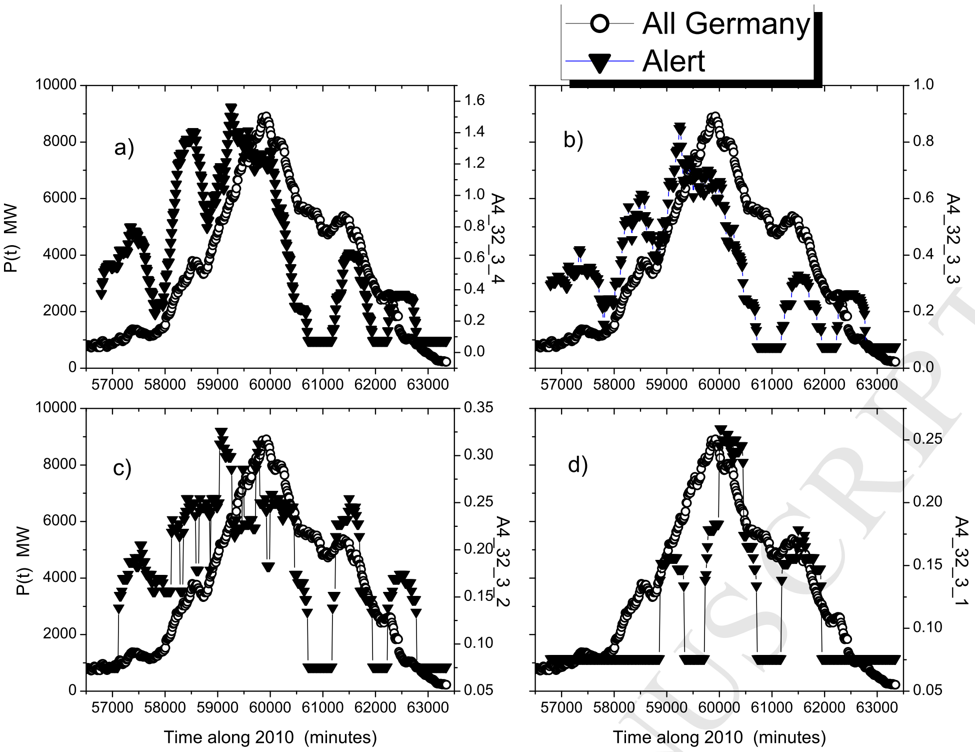


Fig. 1. Evolution of electricity generation by WEP during 100 hours around February 10 to February 13, 2010 (open circles). Alert function with 4, 3, 2 and 1 digit recognition are presented in Figs. a), b), c) and d), respectively (triangles).

To better illustrate the way wlzip works we have prepared a detailed treatment for two different sequences as an example in the Appendix at the end. More interested readers are referred to a broader presentation of the method [23].

Other important calibration feature has to do with the interval in which the time variation $\delta(P(t))/\delta t$ is calculated. We will settle for a four-fold variation with a separation of 8 instants each as it will be shown in Section 3.

All previous calibration procedures will be presented and discussed at the beginning of next Section. In any case, there is not a unique calibration; in what it will be presented below we show plausible ways to tune wlzip for the present application. The final selection of parameters may look a bit arbitrary, but this can be justified by the little variation there is in the results when parameters are varied. A discussion on alternative ways of dealing with some of the features involved in wind power ramp forecasting can be found in a recent review by Gallego-Castillo et al. [4].

3 Results

3.1 Calibration of wlzip

Tuning the sign. Large values of μ can mean variations to both increasing and decreasing periods of WEP (positive and negative ramps). To discriminate between these two regimes we combine μ with the time variation of the WEP function: when the time variation is positive and μ is high enough this is an anticipated signal for a positive period of electricity generation based on wind energy plants.

Let us consider the time variation $D = \delta P(t)/\delta t$, where $P(t)$ is a function of a discrete variable t. The range δt is measured by the number of q intervals of quarters of an hour. To look for more stable results we consider more than one $\delta P(t)$ difference in the definition above which now can be labeled as $D^{d,q}$, where d is the number of differences considered for the variation. Thus for $d = 4$, we can define a four-fold time variation $D^{4,q}$ in the following way

$$\begin{aligned} D^{4,q}(t) = \quad & P(t) - P(t-q) + P(t-1) - P(t-1-q) \\ & +P(t-2) - P(t-2-q) + P(t-3) - P(t-3-q)\ , \end{aligned} \tag{2}$$

where we could eventually divide this result by the number of intervals (4) but it is not necessary, since we will use its sign only.

With previous expression we define a multiplier M in such a way that $M^{4,q}(t) = +1$ when $D^{4,q} > 0$ and $M^{4,q}(t) = 0$ otherwise. In simple words, $M^{4,q}$ is the sign of the variation defined in previous equation. Let us define a ”treated” power sequence $Q(t)$ upon defining

$$Q^{4,q}(t) = M^{4,q}(t)P(t). \tag{3}$$

As it can be seen $Q(t)$ is exactly the same as $P(t)$ in the periods with positive variation while it is 0.0 otherwise. In a sense we will ignore the periods with negative tendency for the purposes of detecting the onset of a favorable period of WEP.

We are now certain that the maxima in the mutability function $\mu(t)$, for the sequence $Q(t)$, correspond to the moments when power generation is increasing at a large rate. This can be calibrated to recognize precursors of good periods for WEP. This is achieved by defining a function called alert $A(t)$ that for a time window of i instants and numerical basis b can be expressed in the

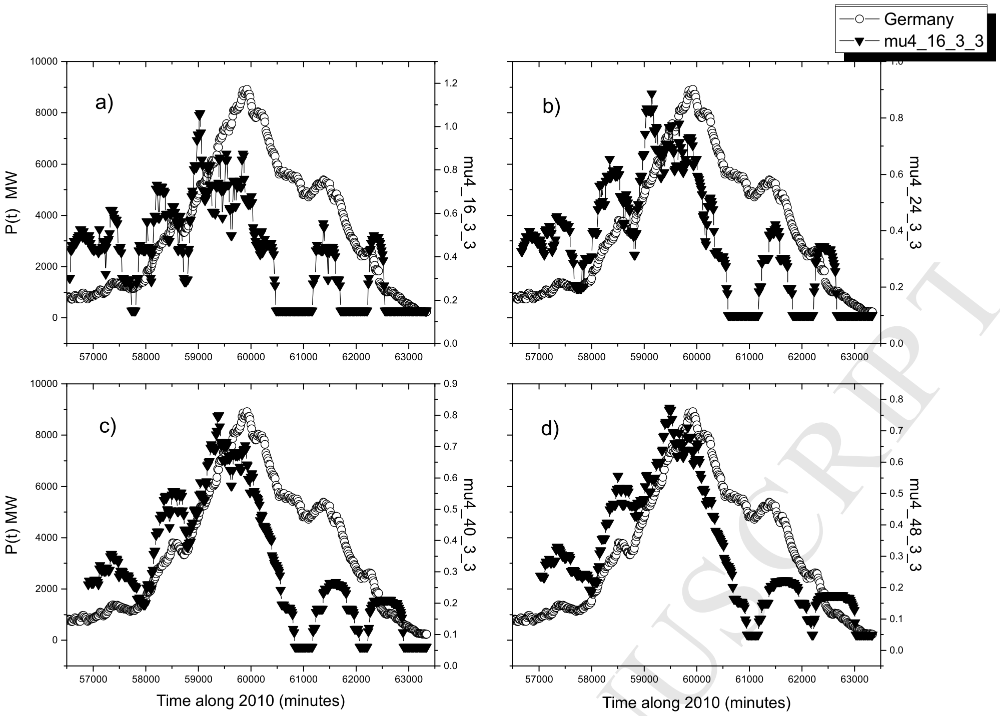


Fig. 2. Evolution of electricity by generation WEP during 100 hours around February 10 to February 13, 2010 (open circles). Alert function with ranges of 16, 24, 40, and 48 instants at intervals of 15 minutes each are presented in Figs. a), b), c) and d), respectively (triangles).

following way:

$$A_i^{4,q}(t) = \mu_i[Q^{4,q}(t)], \tag{4}$$

where we have dropped the suffix 4 in the mutability.

We need to decide about the time variation interval q. Upon looking at the data it is possible to realize that WEP can take a few hours to develop over 5 MW with strong positive slope. The time variation of WEP must consider this fact and it must reflect a stable tendency for a meaningful recent period of time. If this interval is too short (one hour say) quick variations can give erroneous behavior. If this interval is too large (a few hours say) the expected anticipation for a positive period could be lost. We have to settle for a value and we pick a two hours variation (q =8); some justification for this choice will be given below.

Tuning the field. As it can be seen from the data in column 5, the first digit 0 never changes in this sequence, while the second position changes very little (see the last entry of this column). This is the idea of the tuning mechanisms

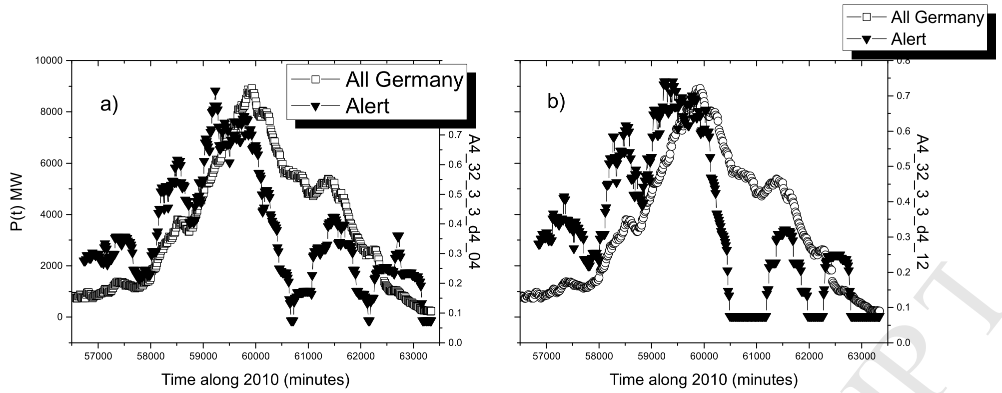


Fig. 3. Evolution of electricity generation by WEP during 100 hours around February 10 to February 13, 2010 (open circles). Alert function with time variations obtained with delays of 4, and 12 instants at intervals of 15 minutes each are presented in Figs. a), and b), respectively (triangles).

in information recognition: we can set wlzip to recognize s digits beginning at position r, which is also included in the count of s. Such mutability will be denoted as $\mu_{i\_r\_s}(t)$. The corresponding alert function will be labeled as $A^{4,q}_{i\_r\_s}(t)$, for a four-fold time variation. Which is the optimum s value?

Let us begin the data recognition from the third position ($r = 3$) including the 5 digits to its right ($s = 5$). We consider $i = 32$ and $q = 8$. In this way we calculated dynamic alert indicators like: $A^{4,8}_{32\_3\_5}(t)$, $A^{4,8}_{32\_3\_4}(t)$, $A^{4,8}_{32\_3\_3}(t)$, $A^{4,8}_{32\_3\_2}(t)$, and $A^{4,8}_{32\_3\_1}(t)$, thus progressively lowering the recognition field.

We now apply these variations to the calculation of alert to a period of 100 hours around February 10 to February 13, 2010. This period is appropriate because the increase of $P(t)$ is rather smooth as compared to other increases to be considered below and it has a very small precursor just under 57500 minutes; then it shows a more pronounced increase with a set back just over 58500 minutes followed by a vigorous increase over 59000 minutes. We want an indicator able of discriminating these behaviors. Results are shown in Fig. 1 for $s = 2$, 3, 4, and 5. (The case for $s = 1$ is quite similar to $s = 2$ so it has been omitted from the figure but it is included in the discussion below).

Fig. 1a) shows that $A^{4,8}_{32\_3\_4}(t)$ gives a too large response for the small precursor at 57500 minutes. On the other hand, the other two maxima are almost of the same height. Actually both features are even more so in the case of $A^{4,8}_{32\_3\_5}(t)$, which is not shown in the figure. This is an indication to move to lower s values. When we consider Fig 1b) we appreciate a better established role of the maximum over 59000 minutes for $A^{4,8}_{32\_3\_3}(t)$. However, as we go to Fig. 1c) we realize that $A^{4,8}_{32\_3\_2}(t)$ evidences the onset of saturation for the alert function and the maximum begins to shift to the right, thus losing anticipation. These comments can only be reinforced upon looking at function $A^{4,8}_{32\_3\_1}(t)$ in Fig.

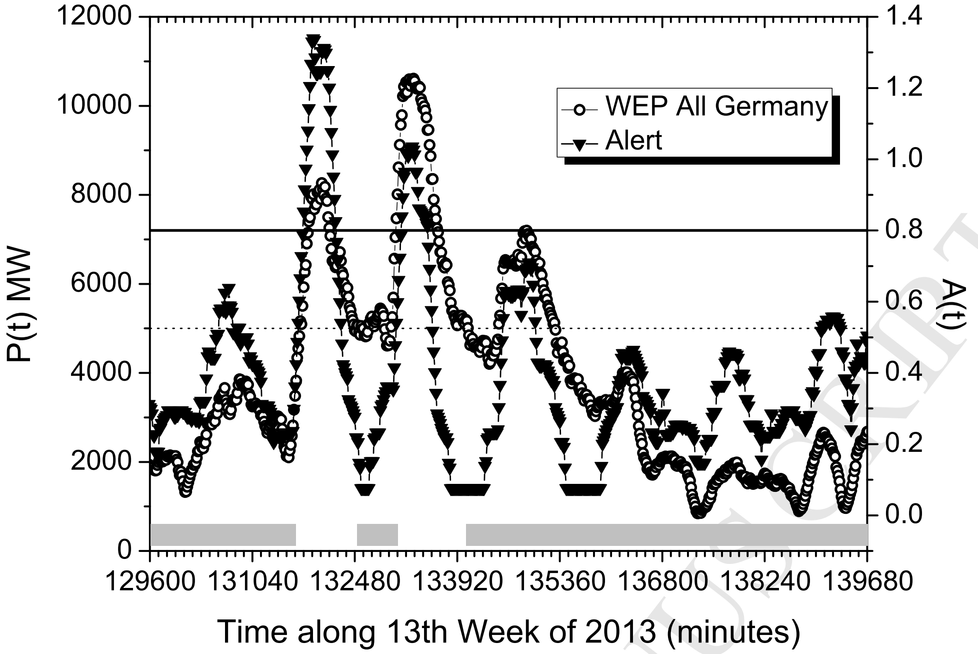


Fig. 4. Several short moderate/negative periods for WEP during the last week of March 2013. Not all of them can be conveniently used to feed the electricity network replacing conventional sources. If the protocol proposed in the text is used only the two white periods shown on the gray band just over the abscissa axis would have been used for a total of about 32 useful hours.

1d). This analysis shows that an optimization is possible and that $A^{4,8}_{32\_3\_3}(t)$ combines the right contrast of the maxima, the sensitivity to the changes in wind power generation and a reasonable anticipation to an incoming positive period of WEP.

At this point we want to emphasize that previous choice (and others coming below) are not unique but represent plausible values for the first time this method is used in this field. A true optimization using actual wind farm data is far beyond the present scope of this paper.

From previous analysis we settle from now on to the precision $r = 3$ and $s = 3$ for information recognition on the WEP data expressed in quaternary basis.

Tuning the time window. Let us now vary the time window i. Results for $A^{4,8}_{16\_3\_3}(t)$, $A^{4,8}_{24\_3\_3}(t)$, $A^{4,8}_{40\_3\_3}(t)$, and $A^{4,8}_{48\_3\_3}(t)$ are shown in Figs. 2a), 2b), 2c) and 2d), respectively. As it can be seen $A^{4,8}_{16\_3\_3}(t)$ tends to give a discrete response indicating low accuracy; in addition the discrimination between the weak increase at time 57500 minutes with respect to the second one near 58500 is very poor. On the other extreme, $A^{4,8}_{48\_3\_3}(t)$ presents a clear delay with respect to $A^{4,8}_{32\_3\_3}(t)$ given already in Fig. 1b). We have settled for a time

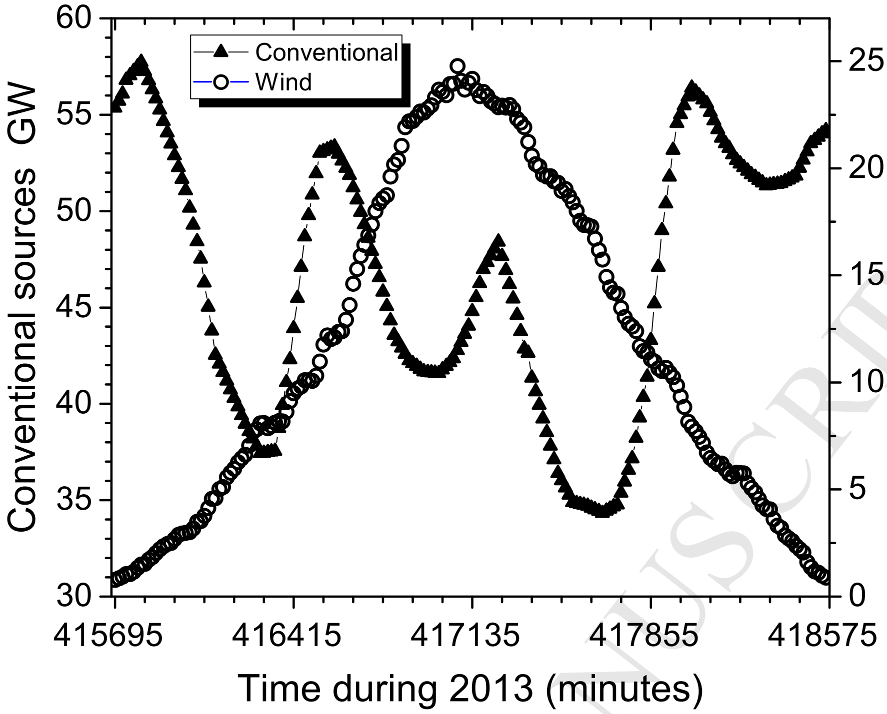


Fig. 5. Generation of electricity by conventional sources (filled triangles) and WEP (open circles) from noon of October 16, 2013 to noon of October 18 2013.

window of $i = 32$ instants (8 hours).

Tuning the anticipation. We analyze now the value of the interval q in the time variation. In Fig. 3a) we present the results for $q = 4$ where we see an acceptable behavior. However, the case $q = 8$ already presented in Fig. 1 has a more continuous variation and the periods with negative tendency are better recognized (flat minima on the right-hand side). In the case of $q = 12$ presented in Fig. 3b) we appreciate a lower contrast among the maxima of alert on the left-hand side, while the anticipation is slightly lost. Then, $q = 8$ looks like a reasonable value which we use from now on.

Summary on Tuning. Previous interval of 4 days during February 2010 was chosen because it shows an almost continuous increase of WEP along one and a half day, which is somewhat extended as compared to most increases in WEP. One can think that what it works for this extended period with several variations should work even better for more sudden continuous increases of WEP. So, for the rest of the paper we consider exclusively values for $A^{4,8}_{32\_3\_3}(t)$ which we will simply denote as $A(t)$ from now on.

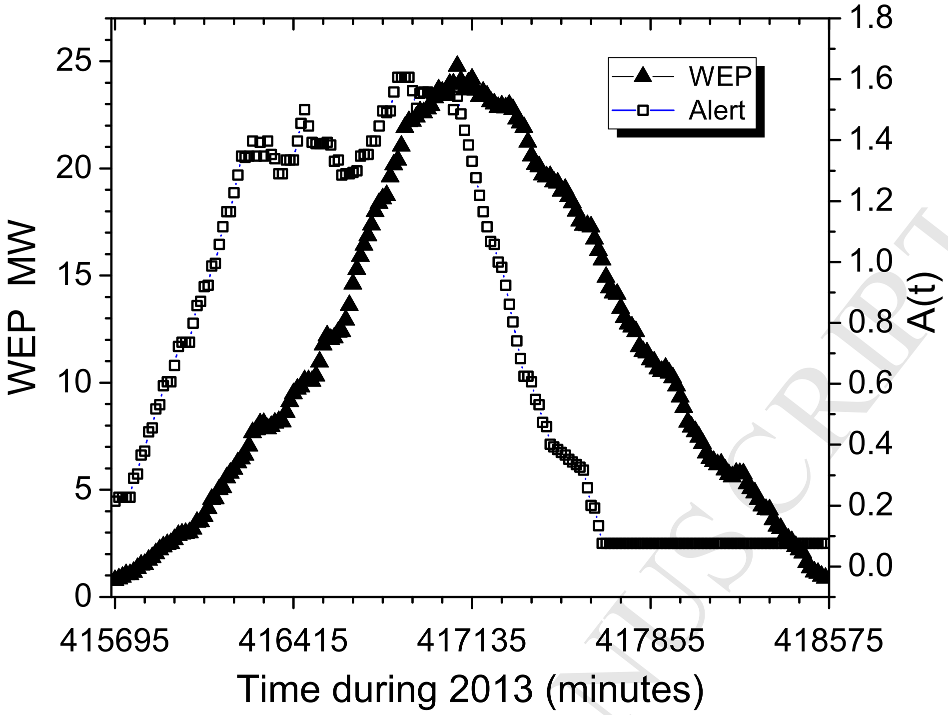


Fig. 6. Treated WEP in the way described by Eq (3) and related discussions (filled stars) and the corresponding Alert $A(t)$ function (open squares) from noon of October 16, 2013 to noon of October 18 2013.

3.2 Operation

Protocol. The function alert $A(t)$ defined above will be now the main indication for the operation of a fictitious plant that will combine WEP with conventional sources.

We propose here a very simple initial operational protocol which can be defined in terms of the following cyclic three steps: 1) When alert $A(t)$ overcomes a critical value A_C, namely when $A(t) > A_C$, WEP supplies energy and conventional sources lower their production accordingly. 2) WEP is used through the network while power produced in this way overcomes a preestablished minimal power P_{min}. 3) When WEP goes under P_{min} the plant working on conventional sources is back in full operation. 4) The process continues indefinitely in this way alternating steps 1 through 3.

This protocol is a simplification of a gradual shut out of conventional sources in balance with WEP production. The main purpose here is to illustrate the detection of the onset of a favorable ramp. Step 3 is the simplest possible way to return to conventional sources and can be readily replaced by any other

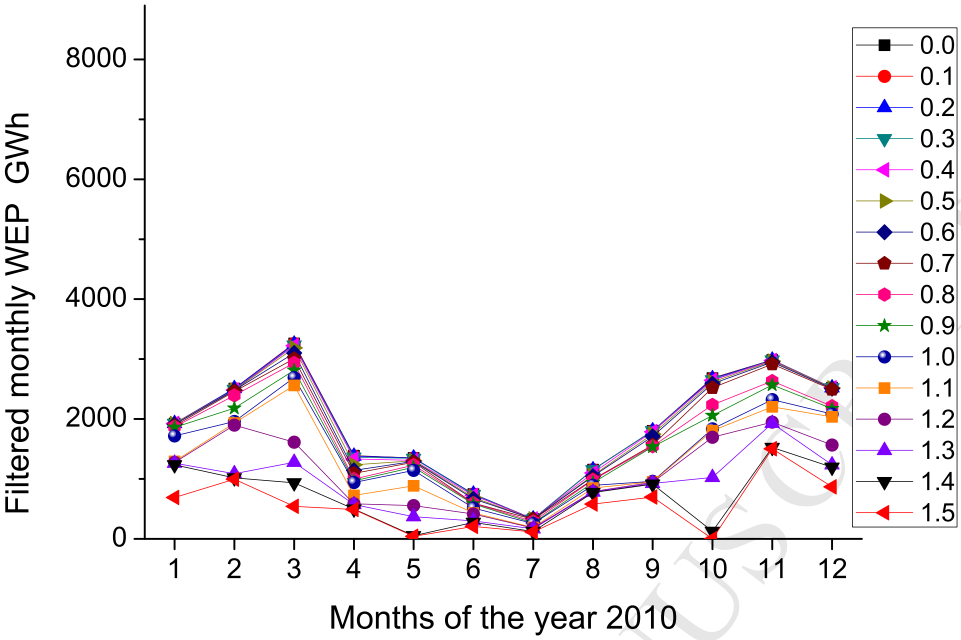


Fig. 7. Electricity generated by wind turbines in all over Germany during 2010 filtered according to the A_C values given in the inset.

established method for the same purpose. What is new in our proposal is the way to achieve Step 1 by means of information theory.

The value of P_{min} can be defined in terms of practical terms. Upon looking at the actual data for the entire WEP in all over Germany a sensitive P_{min} can be 5 GW, value which we will use for illustrative reasons only. However, this value can be adjusted according to seasons, local conditions and evolution of the productivity.

Example of administration Let us do an exercise to appreciate the way previously proposed mechanism can help to save energy. We use the data for entire Germany and we choose to illustrate the protocol during a rather poor week for WEP, namely the 13th week of year 2013 going from Monday March 25 to Sunday March 31. The generated electric power is given by the function $P(t)$ in Fig. 4 by means of open circles. The solid downward triangles give the values of $A(t)$ calculated as described above; this function is to be read on the scale to the right of Fig. 4. What should be the value of A_C to make appropriate use of the scarce WEP during this week? We pick the value $A_C = 0.8$ for the purposes of the present exercise only.

So now we invoke the protocol for $A_C = 0.8$ and $P_{min} = 5$ GW. The result is shown by the bar just over the abscissa axis: gray means conventional sources period, white means partial replacement of energy generation by means of

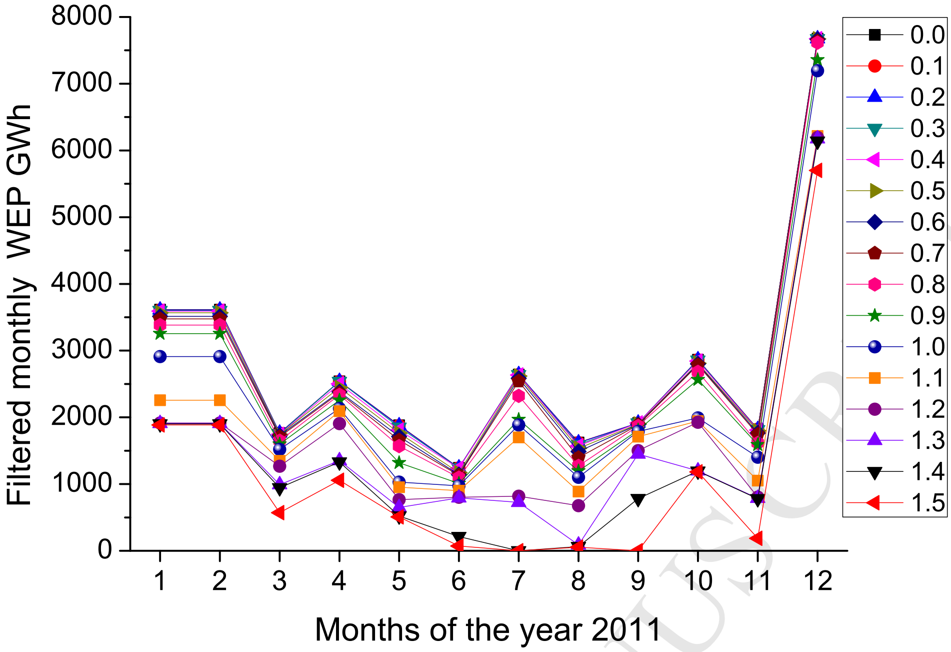


Fig. 8. Electricity generated by wind generators in all over Germany during 2011 filtered according to the A_C values given in the inset.

wind power plants. For this example we find about 32 hours during this week where electricity generated by wind had a real significance. This can change a bit according to the parameters defining the protocol but the point here is that a protocol is feasible to make use of the electricity generated in this way even during unfavorable periods.

Anticipation The next point is to establish the degree of anticipation of a protocol like the one just presented above. To do this job we combine previous data with the actual production of electricity both by wind and by conventional sources [31].

When the energy data is examined it is found that overshoots between 5 and 15 % occur during days with high WEP. So energy is lost during periods with the most favorable condition for wind energy. This happens for about 15 to 25 days during a year which means that energy from conventional sources could have been saved. This is a clear indication that protocols still have not been optimized to handle favorable periods of WEP. In the next example we show a way the previously defined protocol could have helped to avoid using conventional sources thus saving energy.

Let us pick the overshoot that occurred on Friday, October 17, during the 42nd week of 2013. In Fig. 5 we present the total energy produced by conventional sources (filled triangles) and WEP (open circles) from noon October 16 to

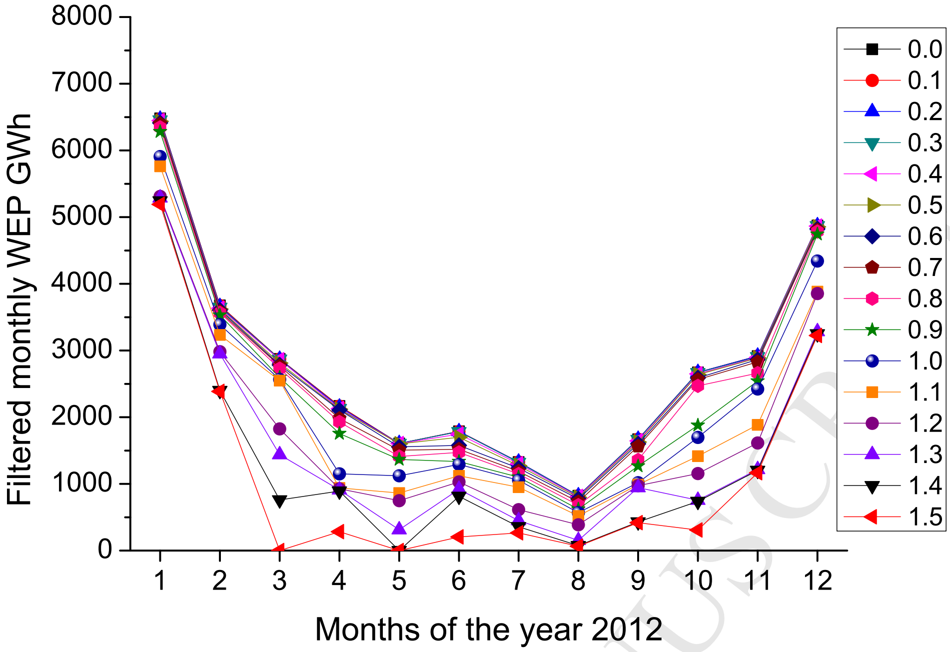


Fig. 9. Electricity generated by wind power plants in all over Germany during 2012 filtered according to the A_C values given in the inset.

noon October 18 [31]. Could the incoming favorable period for WEP have been anticipated in a better way? The answer is yes and it is contained in Fig. 6 where open squares give the function $Q(t)$ defined in Eq (3) and solid stars give the corresponding alert $A(t)$ function defined by Eq. (5). This last indicator goes over $A_C = 0.8$ when conventional sources continue to be used at normal pace as seen from Fig. 5. It is clear that a protocol similar to the one described above would have had the anticipation to save at least part of the energy generated excessively.

3.3 Yearly outcomes

The method based on information theory proposed above can also help to analyze the production of wind energy on seasonal bases. Eventually different strategies can be defined for the different months or even weeks along the year if the tendencies are known.

Let us consider the electricity generated by means of WEP during the first five years of the present decade: 2010, 2011, 2012, 2013 and 2014. For each year we have the power generation $P(t)$. To this series we can instantly calculate $A^{4,8}_{32\_3\_3}(t) = A(t)$ in the way described in previous subsection. From all the WEP we can filter the electricity generated according to the previously defined

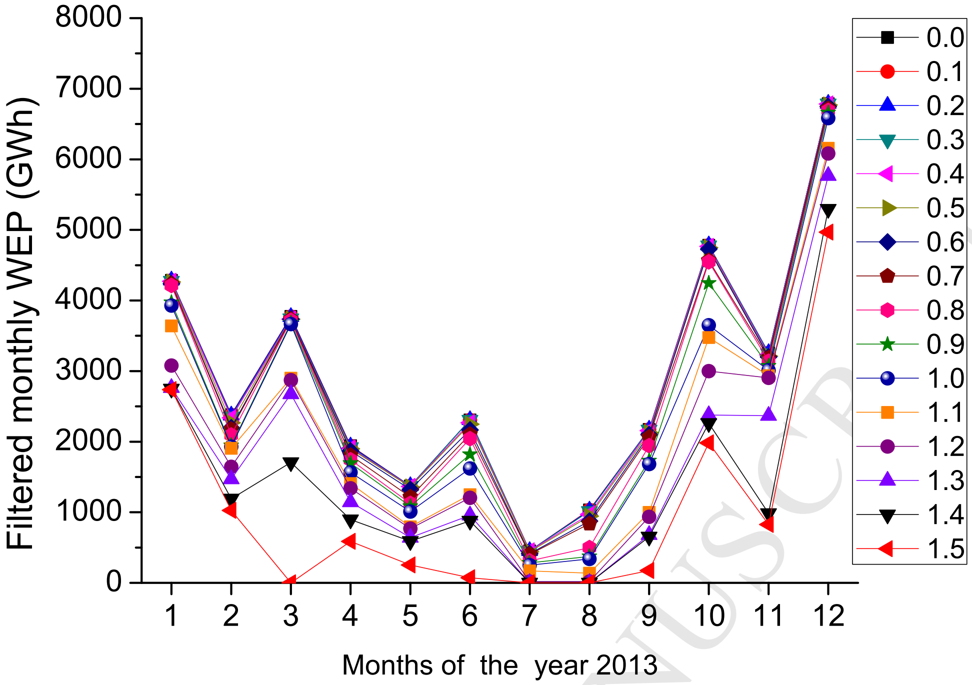


Fig. 10. Electricity generated by wind turbines in all over Germany during 2013 filtered according to the A_C values given in the inset.

protocol, with $P_{min} = 5$ GW and with values of A_C in the range $[0.0, 1.5]$ with increments of 0.1. The value $A_C = 0.0$ means no filtering so every Wh produced by any of the interconnected wind turbines is accounted for.

As A_C increases some small contributions are left out of consideration. For large values of A_C only favorable periods of WEP contribute to the filtered power. Electricity generated in this way is added up during each month as a way to appreciate the variations within a calendar year.

Results for years 2010 through 2014 are presented in Figs. 7 to 11, respectively. Several comments can follow from these results. WEP presents clear fluctuations along the year. Winter months tend to be the most productive ones while the opposite is the tendency for the Summer months. However, huge variations are possible as it can be appreciated from the error bars in Fig. 12, where we present the average filtered WEP for $A_C = 1.0$ over the five years under consideration for this calibration approach.

Moreover, filtered WEP changes from one year to next as it can be appreciated in Fig. 13 for the selected values of the critical parameter A_C given in the inset. The dominant fact is the gradual growth due to the installation of more turbines. In any case, some WEP energy is left out of consideration as A_C

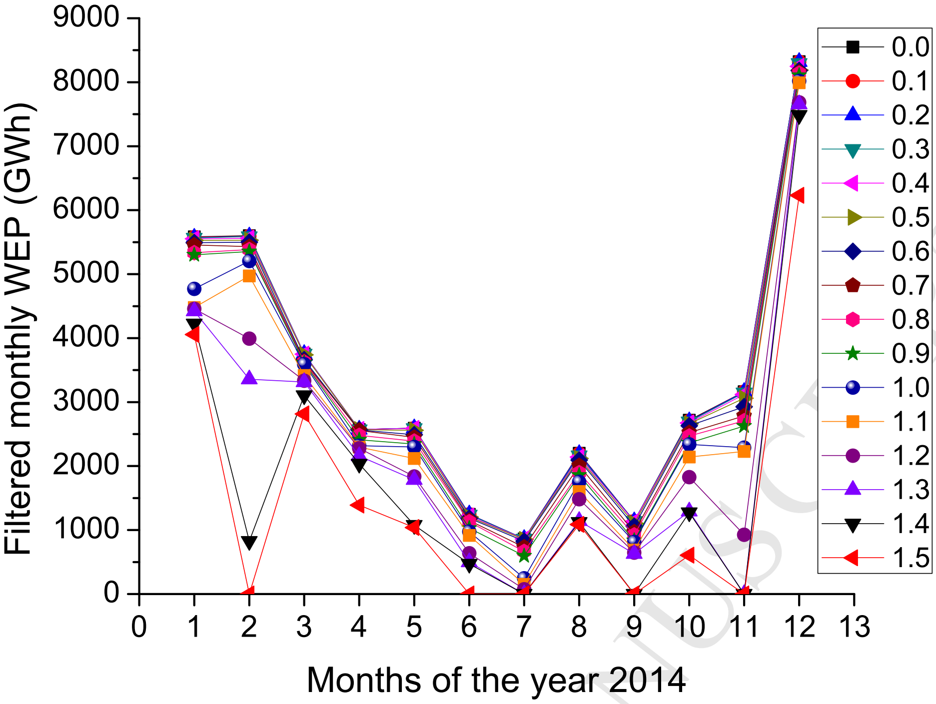


Fig. 11. Electricity generated by wind generators in all over Germany during 2014 filtered according to the A_C values given in the inset.

increases. However this is compensated by the lower operational costs as it can be seen from Fig. 14 where we present the number of connections according to the protocol for the same A_C values of Fig. 13. As is can be seen the number of connections for $A_C = 0.8$ is more than twice the number of connections for $A_C = 1.2$ to gain about 25 % of energy only.

3.4 Direct application to recent years

In previous sections we have used the five-year period 2010-2014 to tune wlzip to make the best use of the available wind energy in the long run. In the present section we just use the best set of tuning parameters to apply them to the most recent years not covered in previous period. The purpose of this exercise is to see if the main results obtained by this method are robust enough as time evolves.

The already optimized tuning parameters for wlzip are listed next. Time window for the dynamical recognition: last 32 instants (8 hours). Numerical basis: quaternary. Sign of the time derivative: two hours delay ($q = 8$ in Eq. (2)). Digits recognition: third, fourth and fifth digits ("_3_3" notation). Then we

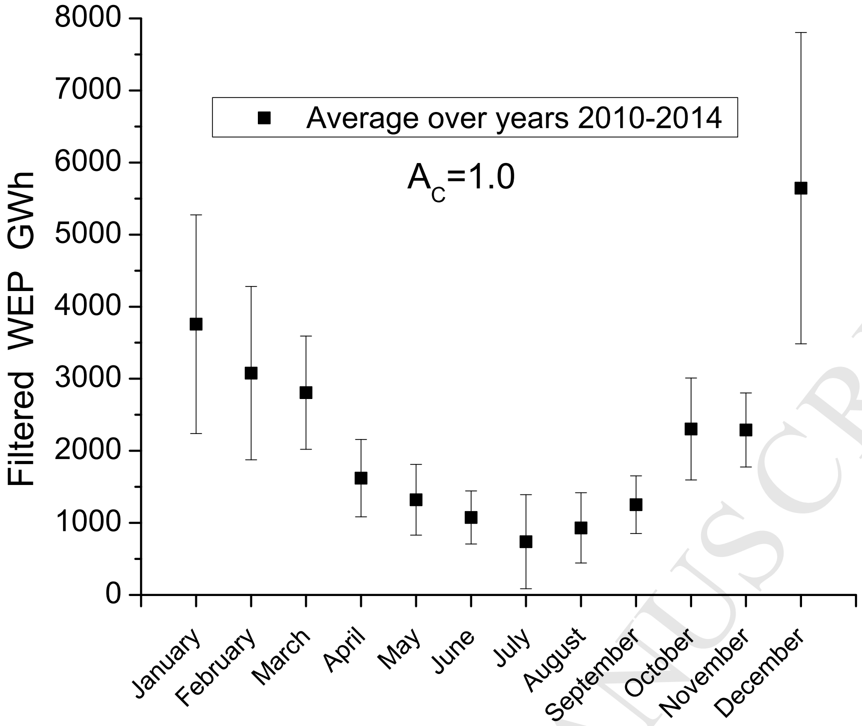


Fig. 12. Average yearly electricity generated by wind power plants in all over Germany filtered for $A_C = 1.0$.

have the options in the protocol (subsection 3.2) where we set for the intermediate one, namely $A_C = 1.0$ and $P_{min} = 5$ GW. We now just apply these options to the data of the years 2015, 2016 and 2017.

Fig. 15 shows the active time of connection of the system according to the protocol by means of dark rectangles. The duration of the connection can be read as the interval on the abscissa axis, while the ordinate gives the average power generated in that interval (the area of the rectangle is the total power generated in this way). Years 2015, 2016 and 2017 are piled up on the same plot to appreciate general seasonal trends. Months are only approximate upon dividing each year in 12 equal periods.

November, December and January are the most reliable months for good wind energy generation, which confirms the tendency already established during the previous five years. Along the same way, May, June, July and August present short and weak intervals of usable wind energy. The other months are erratic and it is precisely here where algorithms as the one presented here can help to anticipate good periods.

Fig. 16 shows the yearly trend for the functioning of the protocol. The bars on the left show the total wind energy generated during each year, namely, they

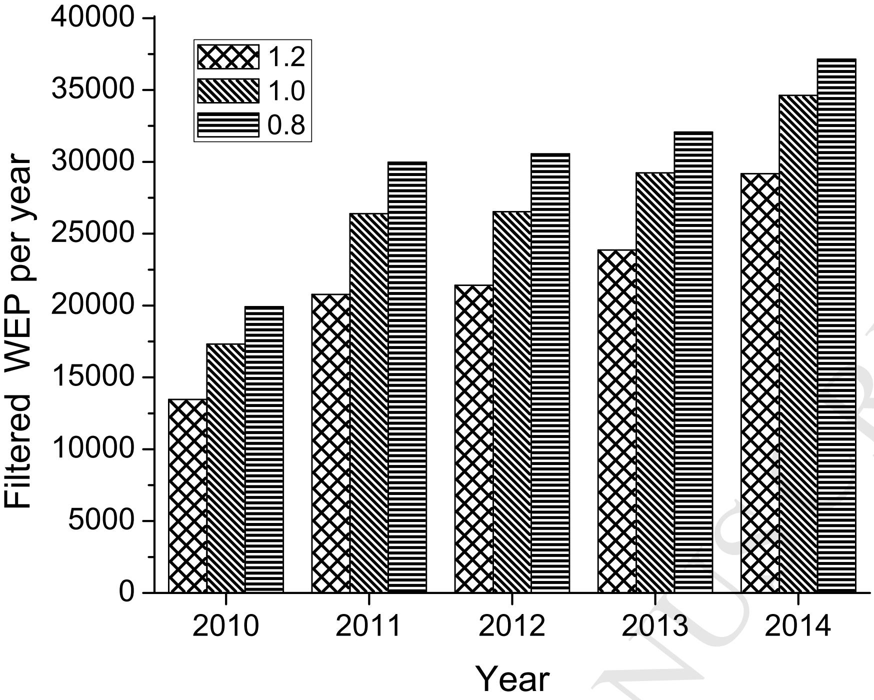


Fig. 13. Filtered WEP according to the protocol defined in the text for the A_C values given in the inset.

represent the addition of all the corresponding dark areas for each year in Fig. 15. The increasing tendency already observed in Fig. 13 for previous years still holds. This reflects the investment of resources in the form of more wind turbines connected to the system; weather variations only slightly modulate this nearly steady increase in used wind power.

The bars on the right reflect the number of connections needed according to the protocol. Values are close to the year 2012 of previous period. This indicator is rather constant and around 100 connections per year for the parameters defined above.

Except for small variations or fluctuations the general trend observed in previous five years prevails which is a good indication for the robustness of the method. Improvements and optimizations are still possible. However, this should be done in situ, with local data for the particular wind farms under consideration.

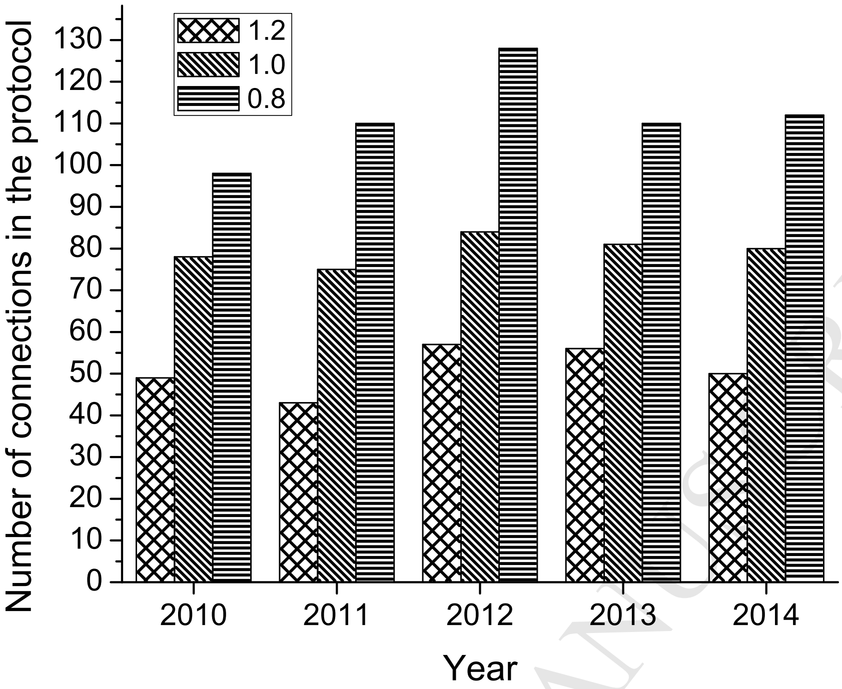


Fig. 14. Number of connections in the protocol for the A_C values given in the inset.

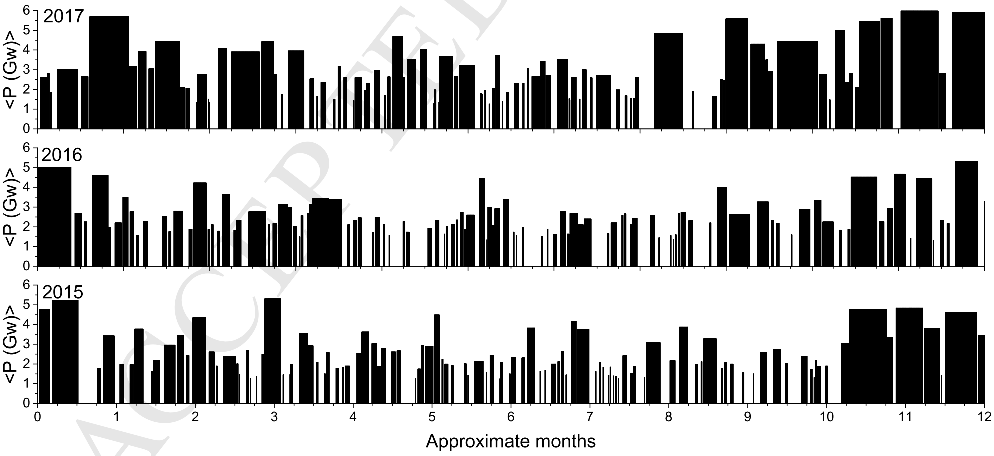


Fig. 15. Dark areas show the usable wind energy according to the parameters given in Subsection 3.4. The connected time is read directly for each interval on the abscissas. The average power for the connected time is given as ordinate. Years 2015, 2016 and 2017 are shown on a common axis showing approximate months to appreciate seasonal variations.

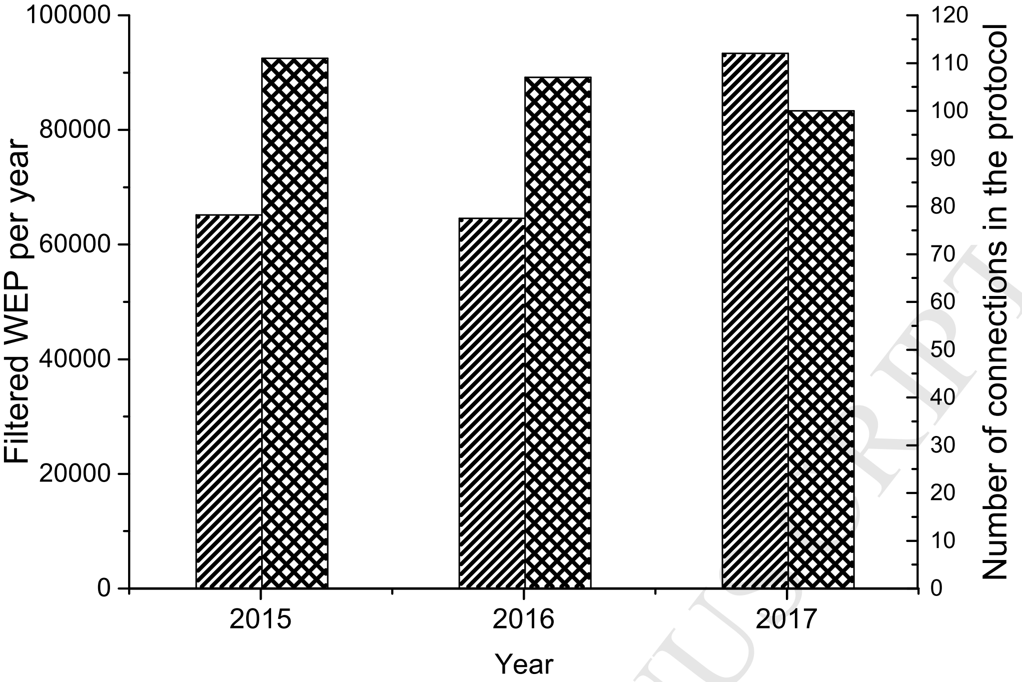


Fig. 16. Yearly accumulation of usable wind energy (left) and number of needed connections (right) for the parameters given in Subsection 3.4.

4 Conclusions

The variability present in the wind power can be recognized by information theory in a dynamical way. The mutability $\mu(t)$ of the recent WEP productivity is a valid indicator for the changes in the productivity pattern. When this is combined with the time variation of WEP we can define function alert $A(t)$ which focuses on the increases of WEP only.

A calibration procedure can allow to determine the threshold value for $A(t)$, namely A_C, which announces a period of high wind energy generation. We have presented a way to do this using the data for all Germany. Similar procedures could be established for local plants. Moreover, seasonal corrections can also be contemplated.

In this way the information content of the time series giving the actual electricity generated by wind turbines can be used to predict its favorable periods. This is similar to what has been done in the case of economical variables [25,26] and biomedical data [27,28]. In a way, this instrument can be thought of like a ”thermometer” measuring the positive agitation prior to a potent period of WEP.

It is possible to calibrate a protocol according to different local conditions and productivity levels. As explained above wlzip allows tuning of several knobs to optimize its performance. The numerical basis can be chosen in accordance with the range of oscillations of the data: for relatively small oscillations a low numerical basis should be used (In the case of blood pressure data a binary basis was used [27,28]); we settled for a quaternary basis for the data adding all WEP sources in Germany. The field can be tuned so as to begin the data recognition over the most sensitive digit subtle to change within the time window of operation; the third digit in quaternary basis turned out to be appropriate for the present study. The number of significant digits over which the data recognition is to be performed can also be adjusted to each problem; three digits in quaternary basis are enough in the present case. In the case of dynamical analysis like this one the time window over which the data recognition is to be performed is the most delicate choice to balance both precision (long time windows) and anticipation (short time windows).

The examples analyzed in this paper show that adjustment is possible to get an alert indication to partially shut down conventional sources and make use of electrical energy generated by wind.

This method can also be used retrospectively to analyze the performance of the turbine network over weeks, months or years. Some conclusions can be drawn from the monthly variation through the five-year period based on figures 7 through 14 above.

During Winter time the choice of A_C is only slightly critical as even high values for A_C lead to high WEP production. The danger here is that energy can be lost by overshoots which could be anticipated by an appropriate protocol.

During Summertime the choice of A_C is not critical as any value leads to low WEP production for most of the years. Actually these months (particularly July and August) can be better invested in maintenance and installation rather than operation.

In Spring and Autumn seasons the choice of A_C is critical to make better use of the scarce and at times short periods of WEP. Values of A_C less or equal to 0.9 should be used to obtain better results for the filtered energy although costs will increase as lower values of A_C are used (see Fig. 14).

The direct application of the parameters which optimize the WEP in one period to next period shows the same general trend. This fact indicates the robustness of the method put forward in this paper. Further optimizations and updates are always possible, however this has to be done in situ for the local data of the particular wind farms of interest. We have used here a general data bank just to present the method and its possibilities.

It is very likely that previous conclusions should be revised if the turbines are split according to location: offshore or onshore; valley or hill; etc.. However it is clear that once the local data sequence is provided it is possible to determine a protocol that can optimize that particular performance.

5 Appendix

The purpose of this appendix is to show the way wlzip actually works in the present case. First column in Table 2 enumerates the 32 instants of the interval used in the compression. The second (fourth) column lists the power generated during a quiet (agitated) period labeled Q (A); quaternary basis is used here. The third (fifth) column is the map created by wlzip with the information in the vector of 32 entries immediately to its left.

The map is created by very simple rules which we illustrate here for the case $\mu_{32\_3\_3}$:
1) Consider the first register: detect the digit position # 3 from left to right and detect the 3 digits from here to the right (third, fourth and fifth digits).
2) Write the truncated register on the map file (column to the right) and indicate its position relative to the beginning of the interval (zero in the initial case, to indicate this is the beginning of this series).
3) Go to next register and consider the digits at the preselected positions:
a) If the digits coincide with those of immediately previous register, add a comma to this register in the map file and then write the number of times this register has repeated so far. If this register repeats immediately again, keep on increasing the counter after the comma. In the example Q of Table 1, all 32 registers are the same under the truncation $\mu_{32\_3\_3}$ so at the end the map file exhibits the value of the truncated register followed by ",32" to indicate it repeated 32 consecutive times. (This period was chosen precisely to illustrate this extreme situation). In the A file the digits of the first register repeat themselves 3 times at the preselected positions then to the right of the truncated register ",3" is written in the first entry of the fifth column. Several other repetitions are also shown along the fifth column.
b) If the digits do not coincide with any previously stored register at their corresponding positions write a new line in the map file writing the truncated register followed by its position in the original file. This is the case of registers 00123 (position 3), 00132 (position 4) etc. for the A column.
c) If the digits coincide with those of one previously stored register p positions before, we just go back to the position such register was stored and add p to the right. This happens with the value 00330 towards the end of the file. This procedure is done at any time a non consecutive coincidence is found.

The weight $w*$ of the map files lead to the mutability values according to Eq. (1). The corresponding values for the present examples are given in the bottom row. Further details and examples can be found in the already quoted literature in the Methodology section.

Table 2. The way wlzip works is illustrated here for two very different periods of time: a quiet period (Q) and an agitated period (A). Columns 1 gives the sequence of consecutive instants; Column 2 gives the produced power for Q in quaternary basis; Column 3 gives the recognized information for Q starting at position 3 and for a total of three digits to the right (boldface characters); Column 4 gives the produced power for A in quaternary basis; Column 5 gives the recognized information for A starting at digit # 3 and for a total of three digits to the right (boldface characters). The corresponding mutability values for each case according to Eq. (1) are given in the last row.

Instant	Q b4	Q 32_3_3	A b4	A 32_3_3
1	000022231	00**002** 0,32	001220020	00**122** 0,3
2	000022323		001220233	00**123** 3
3	000023212		001222202	00**132** 4
4	000023311		001232323	00**210** 5
5	000023132		001320221	00**212** 6
6	000023120		002103120	00**213** 7,2
7	000022333		002121000	00**221** 9,2
8	000022031		002132213	00**230** 11
9	000022003		002132213	00231 12
10	000021030		002210331	00232 13
11	000020301		002213223	00233 14
12	000020313		002300121	00300 15
13	000020303		002310333	00302 16
14	000021000		002320200	00310 17,2
15	000021220		002330310	00311 19
16	000021312		003001323	00313 20
17	000021323		003020231	00320 21,2
18	000021322		003103103	00322 23,3
19	000021102		003102320	00330 26 5
20	000021210		003111121	00331 27
21	000021102		003130311	00332 28,2
22	000021231		003201122	00333 30
23	000022022		003201003	
24	000022012		003220120	
25	000022100		003221102	
26	000022030		003223011	
27	000022202		003300120	
28	000022210		003310111	
29	000022012		003321320	
30	000021321		003321230	
31	000022031		003330320	
32	000023000		003301003	
		$\mu_{32\_3\_3} = 0.048$		$\mu_{32\_3\_3} = 0.955$

Acknowledgements. Partial support from the following two Chilean sources is acknowledged: Fondecyt under contract 1150019, and Financiamiento Basal para Centros Científicos y Tecnológicos de Excelencia (Chile) through the Center for Development of Nanoscience and Nanotechnology (CEDENNA).

References

[1] Statistisches Bundesamt, Bruttostromerzeugung in Deutschland fr 2015 bis 2017, https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/Energie/Erzeugung/Tabellen/Bruttostromerzeugung.html.

[2] EEG in Zahlen 2016; http://www.bundesnetzagentur.de.

[3] A. Tascikaraoglu, M. Uzunoglu, A review of combined approaches for prediction of short-term wind speed and power, Renewable and Sustainable Energy Review 34 (2014) 243-254.

[4] C. Gallego-Castillo, A. Cuerva-Tejero, O. Lopez-Garcia, A review on the recent history of wind power ramp forecasting, Renewable and Sustainable Energy Review 52 (2015) 1148-1157.

[5] Y. Ren, P.N. Suganthan, N. Srikanth, Ensemble methods for wind and solar power forecasting - A state-of-the-art review, Renewable and Sustainable Energy Review 50 (2015) 82-91.

[6] A.M. Foley, P.G. Leahy, A, Marvuglia, E.J. McKeogh, Current Methods and Advances in Forecasting of Wind Power Generation, Renew. Energy 37 (2012) 1-8.

[7] Song Li, Lalit Goel, Peng Wang, An ensemble approach for short-term load forecasting by extreme learning machine, Applied Energy 170 (2016) 22-29.

[8] Y. Feng, S.M. Ryan, Day-ahead hourly electricity load modeling by functional regression, Applied Energy 170 (2016) 455-465.

[9] M.G. De Giorgi, A. Ficarella, M. Tarantino, Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods, Energy 36 (7) (2011) 3968-3978.

[10] D. Ambach, Short-term wind speed forecasting in Germany, J. Appl. Statistics 43 (1-4) (2016) 301-369.

[11] E. Erdem, J. Shi, ARMA based approaches for forecasting the tuple of wind speed and direction, Applied Energy 88 (2011)1405-1414.

[12] L. Xiao, J. Wang, R. Hou, J. Wu, A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting, Energy 82 (2015) 524-549.

[13] I. Colak, S. Sagiroglu, M. Yesilbudak, Data mining and wind power prediction: A literature review, Renew. Energy 46 (1) (2012) 241-247.

[14] G. Sun, C. Jiang, P. Cheng, Y. Liu, X. Wang, Y. Fu. Y. He, Short-term wind power forecasts by a synthetical similar time series data mining method, Renew. Energy 115 (2018) 575-584.

[15] V. Cortez, G. Saravia, E.E. Vogel, Phase diagram and reentrance for the 3D Edwards-Anderson model using information theory, J. Magn. Magn. Mater. 372 (2014)173-180.

[16] Hiu Liu, Hong-qi Tian, Yan-fei Li, Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction, Applied Energy 98 (2012) 415-424.

[17] I. Gonzlez-Aparicio, A. Zucker, Impact of wind power uncertainty forecasting on the market integration of wind energy in Spain, Applied Energy 159 (2015) 334-349.

[18] G. Saravia, E.E. Vogel, ”Word length zipper: WLZIP” software registered at Chilean Intelectual Property System, RPI # 240108, 2014.(Freely available upon request).

[19] D.G. Luenberger, Information Science, Princeton University Press, Princeton, NJ, Oxford, 2006.

[20]] T.M. Cover, J.A. Thomas, Elements of Information Theory, second ed., John Wiley & Sons, New York, 2006, p. 24.

[21] J.G. Roederer, Information and its role in Nature, Springer, Heidelberg, 2005.

[22]] E.E. Vogel, G. Saravia, F. Bachmann, B. Fierro, J. Fischer, Phase transitions in Edwards-Anderson model by means of information theory, Physica A 388 (19) (2009) 4075-4082.

[23] E.E. Vogel, G. Saravia, L.V. Cortez, Data compressor designed to improve recognition of magnetic phases, Physica A 391 (4) (2012) 1591-1601.

[24] V. Cortez, Series de tiempo en vidrios de espin de tres dimensiones; Electronic Civil Engineer Thesis, Universidad de La Frontera, Chile, 2009.

[25] E.E. Vogel, G. Saravia, Information theory applied to econophysics: stock market behaviors, European Physical Journal B 87 (8) (2014) 1-15.

[26] E.E. Vogel, G. Saravia, J. Astete, J. Daz, F. Riadi, Information theory as a tool to improve individual pensions: The Chilean case, Physica A 424 (2015) 372-382.

[27] D. Contreras, Diagnóstico de hipertensión arterial (HTA) a partir del análisis de Holters mediante compresores de información; Medical Physics Magister Thesis, Universidad de La Frontera, Chile, 2015.

[28] D.J. Contreras, E.E. Vogel, G. Saravia, B. Stockins, Derivation of a measure of systolic blood pressure mutability: a novel information theory-based metric from ambulatory blood pressure tests, Journal of the American Society of Hypertension 10 (3) (2016) 217-223.

[29] E.E. Vogel, G. Saravia, D. Pastén, V. Munñoz, Time-series analysis of earthquake sequences by means of information recognizer, Tectonophysics 712-713 (2017) 723-728.

[30] https://www.eex-transparency.com

[31] https://www.energy-charts.de/power.htm

Highlights

- Data recognizer wlzip is used to anticipate favorable periods of wind energy
- The method can also be used to analyze wind energy production
- Data from all German turbines during 2010:2017 is used in this study
- A protocol for mixing wind energy with conventional sources is proposed
- Protocol indicators are tested on monthly basis during the eight-year period