Economics School of Louvain - ESL

Economics School of Namur - ESN

Probabilistic Approach of intermittent energy production

The case of wind turbines in Belgium

Author : Martial Toniotti

Thesis Director : Sébastien Van Bellegem

Thesis Reader : Jean-François Maystadt

Academic Year 2020-2021

Master in Economics: Econometrics – 120 credits – Focus : Research

Université catholique de Louvain - Economics School of Louvain - Place Montesquieu 3, 1348 Louvain-La-Neuve
Université de Namur – Economics School of Namur – Rempart de la Vierge 8, 5000 Namur

Probabilistic approach of intermittent energy production

The case of wind turbines in Belgium

Martial Toniotti

May 26, 2021

Abstract

Wind turbines are an intermittent energy source, meaning their production depends on a physical flux and is inherently random, as opposed to more traditional ways of producing electricity, that are dispatchable. Belgian data is used, analyzed with partial identification, and is modeled with a locally stationary wavelet (LSW) process to try to assess the inherent risk of wind energy, what can be done about it by considering off-shore and on-shore wind turbines as two different assets in a portfolio and see the potential of this energy source in terms of percentage of electrical consumption.

1. Introduction

Intermittent energy has today an increasing role in the production of electricity due to the ecological challenge. Decarbonizing electricity production is a very important policy outcome in the global fight against human-induced climate change. Such energy sources are wind energy, solar energy, but also less common like tidal energy. They are important to model properly and to assess, for the policymakers, as well as the private stakeholders in energy production, it being at the grid level or at the production level.

Intermittent energy is defined here as opposed to dispatchable energy. Given its dependence on the appearance and the intensity of natural phenomena it is subject to challenging sources of uncertainty. The purpose of this research is to provide an econometric framework that takes uncertainty into account and that helps to evaluate the probability of electricity production.

In classical econometric theory, (the parameters of) a model is identified under structural assumptions on the statistical model and the data design. In recent works, "credible inference" has been proposed in order to work with less structural assumptions, reflecting the uncertainty in the statistical model, but at the price of a lack of point-identification (e.g. Manski (2003)). The resulting model is "partially identified", meaning that only the parameters of interests are estimable up to set. The size of the set is of course a crucial aspect of the empirical analysis.

We will focus on the analysis of wind-turbines production in Belgium, on which rich data are publicly available (see next section for a description of the used data). Given the non-stationary nature of the wind speed, one originality here is to decompose this stochastic process as the sum of a trend and a locally stationary time series that will be modeled with locally stationary wavelets (LSW) (see Fryzlewicz et al. (2003) for example). The framework developed on wind-turbines production in Belgium can be used in other countries or on various intermittent energy sources (sunlight, tidal. . .).

The usual method to evaluate the production of intermittent energy sources, and in particular wind energy, is a method that is part of real-time network management. The management of an electrical network with given parameters to make decision rules whose goal is to be able to equalize demand and supply of electricity at any given moment. The intermittent nature of some energy sources is then seen as a parameter among many others in the management of the network by an operator. There exists a whole literature doing just that and describing the effect on the day-to-day operation of intermittent energy sources, notably in a unified way at the network level by Heussen et al. (2011) or by modeling intermittent production as a stochastic variable in Papaefthymiou et al. (2006). Those methods are part of engineering and have hardly any link with economics. Another part of the literature focuses on the inference of wind-turbine energy production from meteorological data, and thus making a link between wind velocity and energy production, what we call a « power curve ». A lot of different technics to estimate the power curve exist, a review of all of those can be found in Lydia et al. (2014). This approach is however limited when we try to aggregate at the wind farm level. Other methods are then used, see Kusiak et al. (2009).

But all those methods don't focus on the historical time-series data of production directly but use meteorological predictions. Here, on another hand, we use tools of analysis of non-stationary time-series. A method among many others for the analysis of those time-series is the use of locally stationary wavelet (LSW), as defined and used for meteorological data in Fryzlewicz et al. (2003). This method is used in different domains, some examples are shown in Van Bellegem and Von Sachs (2004).

In parallel to the use of LSW methods, this project's core use concepts of partial identification. As defined in its seminal work on the subject by Manski (2003).
Finally, a last way of approaching the problem of intermittent energy production exists. It is an application of portfolio management theory developed in Finance but applied to the different energy sources. In this approach pioneered by Shimon Awerbuch in Awerbuch (2006), each energy source is seen as a different asset in a portfolio, with its return and its inherent risk. This method has been greatly expanded since 2006, for example

for wind energy in Krohn et al. (2009).

The objective of this work is double, first, we want a way to model the instability of wind energy in a meaningful way and see how, for example by changing the number of off-shore and on-shore wind turbines, we can try to have an impact on this instability. Secondly, we want a way to assess the theoretical potential of wind energy in percent of consumption of electricity. We want this potential to be as meaningful and solid as possible to be able to see clearly what is possible and what is not with wind energy.

The next section will focus on the sources of data that we will use, what are their specificities, where they are from, and what time period they cover. Some comments on them will be presented.

Section 3 will present the theoretical framework used to interpret in a meaningful way what we observe in the data. The different parameters of interest will be stated and their identification, as well as their estimation strategies from the data, will be presented.

In Section 4, results for the parameters of interest will be presented in different ways, as well as different robustness checks made to our dataset.

Section 5 will discuss what are the implications of the results we have.

Finally, Section 6 will conclude and discuss some eventual extensions that can be done to this model, by adding several different things, like climate change, electrification of the vehicle fleet, but also extensions to other energy sources.

2. Datasets and their sources

Different sources of data are used, each is described hereafter.

2.1. Production of wind-turbines

The core of the analysis is around the time-series of wind-turbines production aggregated for the whole country. Data is from Elia (2021), the Belgian transmission system operator. It is available under an open data license. It is, for each 15 minutes between 2013 and 2020, the aggregated production for all wind turbines in Belgium (in MW / 15minutes, meaning we need to divide it by 4 to get it in MWh). We get too for every 15 minutes the aggregated capacity of wind turbines (in MW) connected to the network. The analysis focus on the time series of the capacity factor over time. It is computed with this data and we can see it in the following graph. We have 279016 datapoints once we have removed inconsistent data (moments where production was bigger than capacity, due to measurement error). Inconsistent data is about 20 data points on a weekend in 2014.

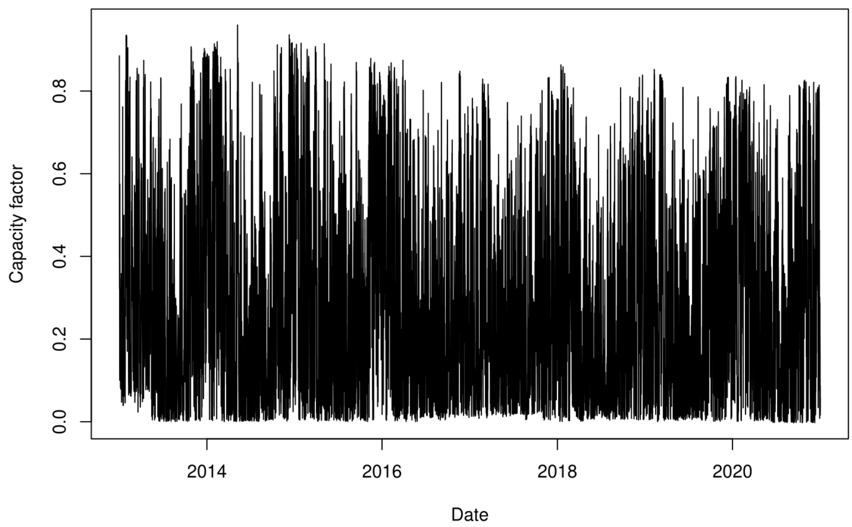


Figure 1: Capacity factor of Belgian wind energy

We consider throughout the analysis that what is limiting the capacity factor and the production at any given moment is the installed capacity and the meteorological conditions, which is not always the case. Sometimes, the transmission system operator can issue what is called decremental bids on wind production. It means that owners of the wind turbines are paid not to produce. It happens when wind overperforms compared to predictions or compared to load and that, in order to ensure grid-wide stability, the operator needs to tell owners of wind turbines to “turn off” their wind turbines. It is relatively a rare occurrence though. We have data on that with the same resolution as production (so every 15 minutes), and it happens only in 500 data points, or a cumulated 5 days during the 8 years of data. Of course, we cannot discard those data since it is not issued at random, but when wind turbines are overperforming. Analysis will be done with those limitations in mind, and during the robustness checks, it will be checked if the results we get are the same without those data points that we know are a bit problematic since capacity factor and production are underestimated compared to their physical bound (in which we are interested).

2.2. Total Load on the grid

Another dataset that will be used during our analysis is the dataset of the total load on the electrical grid in Belgium. Data is also from Elia (2021), and also available in 15 minutes intervals. Data is only available for 2014 until 2020, there is no way to get the data for 2013. Probably because it wasn't monitored in the same way before. Data is freely available under an open data license on Elia's website. We get, for each interval, the MW on the whole grid (not only on Elia network but on the whole network). To get the MWh consumed during this interval, since it is a fourth of an hour, we can simply

divide by 4.
This data will be used mainly to cap the expected production of wind energy. No matter the installed capacity of wind turbines, it is not possible to produce more than what is consumed at each moment. It is a bit different in real life because energy can be exported, but the general rule is that when there is too much wind in Belgium, there is also in the neighboring countries. In the following figure, you can see the evolution of the grid total load over time. We can see that 2020 was a special year (due to the covid pandemic and major disruption of production and habits in Belgium).

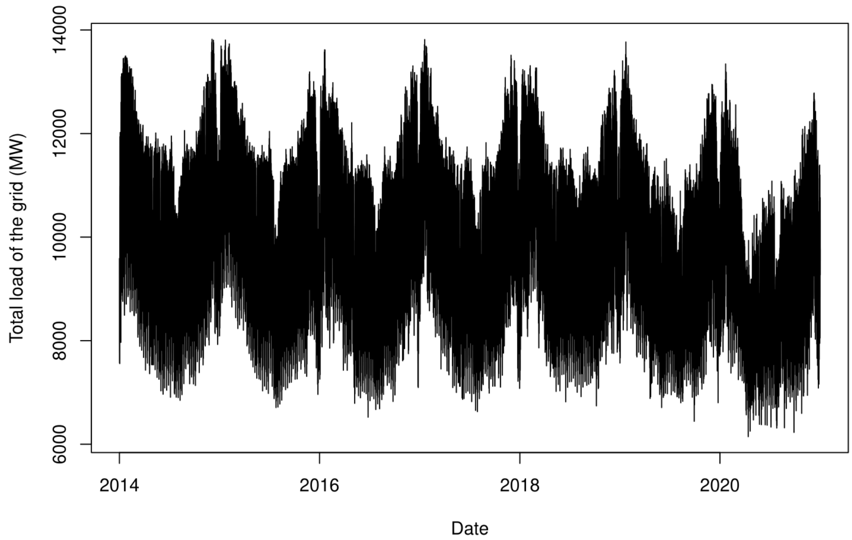
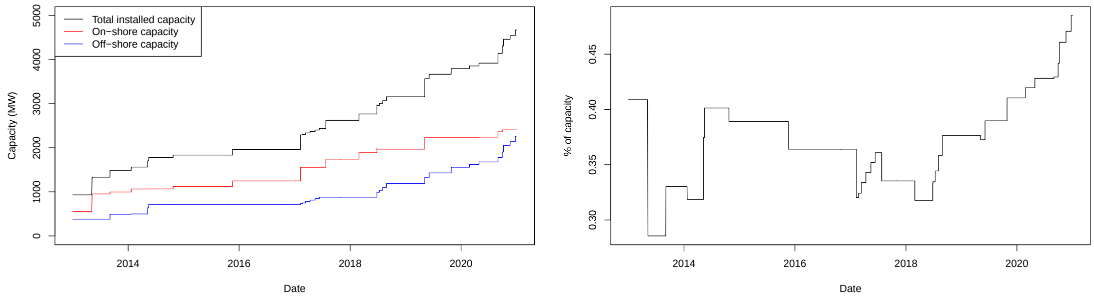


Figure 2: Load of the Grid (MW) over time (intervals of 15 minutes)

2.3. Date of opening of off-shore wind farms

As already stated, we get from the grid operator data about the installed capacity of wind turbines aggregated at the country level for every 15 minutes. It is aggregated also between onshore and offshore wind turbines. But we can, and we have, disaggregated capacity for each moment in on-shore and off-shore using the fact that off-shore wind farms are few (9 at the end of 2020). Information on when they got on the grid can be found on the website of the Belgian Offshore Platform (BOP (2020)), an association of all stakeholders interested in development of wind energy in the belgian part of the North Sea. With this disaggregated data we can then compute, for each 15 minutes, the capacity of on-shore and off-shore wind energy and the ratio of capacity that is off-shore in each moment. This ratio will be very important in the analysis.



(a) Capacity of each type over time

(b) Ratio of off-shore in the capacity

Figure 3: Off-shore and on-shore disaggregated capacity

3. Theoretical framework

In this part, the theoretical framework used will be described. First, the modeling of the wind turbines network in Belgium, then the decomposition of the capacity factor time-series. Finally, an expression of the variance of the capacity factor time-series in function to the different type of wind-turbines (off-shore and on-shore) will be given.

3.1. Modeling of the network

First, let's define, in a perfectly standard fashion, a parameter for time, t, that is a unique moment of a given day, a given year. Data will not repeat itself each year or each day. In other words, let's define

$$t = (m, d, y) \tag{1}$$

Where t is our time parameter, m is a specific moment (of 15 minutes) of the day, d is a specific day of the year, and y is a specific year. For simplification purposes, from now on, only t will be written to be used as the time parameter.

a wind-turbine

We can order all possible wind turbines in Belgium (off-shore and on-shore) in a list of J wind turbines, even if they are not built yet. A wind-turbine is defined by its capacity(in MW) c_j, its localisation g_j and a production in each time t, generated from a random variable ϕ_j.

a network

We can define networks. A network in our setting is a set of wind-turbine that are part of it and that produces some aggregated data. Mainly it has a well-known total capacity, that is the sum of the individual wind turbines that are part of it, and a production of electricity at each time t, defined as the sum of the production of all the wind turbines part of the network at that moment.

$$
\begin{aligned}
N &= (\kappa_N, \Phi_N) \\
where\ \kappa_N &= \sum_{j \in N} c_j \\
\Phi_N &= \sum_{j \in N} \phi_j
\end{aligned} \tag{2}
$$

The network N is defined by its capacity κ_N and a random variable Φ_N that produces a time series of production for each time t.

The complete network

We can now define a specific network, the complete network. It is defined as the network in which are all the possible wind turbines in Belgium. Meaning that all the wind turbines that can exist are in it, even if they are not built yet. This network is noted:

$$
\begin{aligned}
&(C, P) \\
where\ C &= \sum_{j} c_j \\
P &= \sum_{j} \phi_j
\end{aligned} \tag{3}
$$

The random variable P generates a time series $\{P\}_t$ that has a result for each t denoted P_t. We can also for this network define the capacity factor time series. The capacity factor is the ratio of the production at a moment over the installed capacity and gives a measure of the usage of the installed capacity. It is a unitless ratio.

$$
\{X\}_t = \frac{P_t}{C} \in [0, 1] \tag{4}
$$

Of course, this network is not built, so we cannot observe it. We have no idea what the values of P_t are, in fact, the only network we observe is the one defined in the next point.

The observed network

For all the wind-turbines J, we know with certainty if they are built or not for all time t.

$$
Z_{tj} = \begin{cases} 1\ if\ j\ built\ in\ t \\ 0\ if\ j\ not\ built\ in\ t \end{cases} \tag{5}
$$

We can then define the observed network (that of course changes with time, new wind-turbines are added),

$$
\begin{aligned}
&(C_t, R_t) \\
&where\ C_t = \sum_j Z_{tj} c_j (\leq C) \\
&R_t = \sum_j Z_{tj} \phi_j
\end{aligned}
\tag{6}
$$

We can in the same fashion as in the complete network case define the capacity factor of the observed network, Y_t. All the data we have are for the observed network, a network that changes with time. From that, we will try to draw conclusions on the complete network with some assumptions that will be stated.
We can also define in the same fashion the unobserved network, which is the complete network minus the installed network at each time t.

3.2. Partial identification of the capacity factor

One parameter of interest that we would like to have an identification set on is the capacity factor of the complete network, X_t. We can, in a first step, decompose it as the weighted (by the capacity of each network) average of the capacity factor of the installed network and of the uninstalled network.

$$\forall t, X_t(\in [0,1]) = \frac{C_t}{C} * Y_t + \frac{C - C_t}{C} * U_t \tag{7}$$

Where $\frac{C_t}{C}$ is the percentage of the network that is installed in time t, $\frac{C-C_t}{C}$ the percentage of the network that is uninstalled in time t, and Y_t and U_t respectively the capacity factor of the installed and uninstalled network.
We are interested in X_t but the only thing that we observe directly in its decomposition is $Y_t(= \frac{R_t}{C_t})$. If we want an identification region for X_t, we will need assumptions on the capacity factor of the unobserved network, U_t. The intuition behind them is that wind turbines are not put into random into the Belgian territory, the different companies building them have a profit motive in mind. So they built first the most profitable wind turbines in the population J of possible ones. Meaning the installed network is not a representative sample of the complete network.
All those assumptions permit us to bound the values that X_t can take and express it only depending on Y_t, that we observe. We call this set of possible values of X_t its identification region, denoted by $H[X_t]$. The idea of the computation of such a set can be found in Manski (1989) and was extended greatly in Manski (2003). After the assumption, the identification set for it is presented. Computation for the identification sets can be found in the appendix.

Missing-at-random assumption (MAR)

$$X_t = Y_t = U_t \tag{8}$$

When people use the current network as a predictor of future production of wind-turbines, this is the unconscious assumption that is made. But as stated before, it is assuming that companies building the network have built the wind-turbines at random and that the installed network is a good predictor of the complete network.
Due to the fact that the relation between the variables is here an equality, we won't have an identification set but a point-identification of our variable of interest. If we take back (7) and put the MAR assumption (8) in it, we get:

$$X_t = H_{MAR}[X_t] = Y_t \tag{9}$$

More productive installed network assumption

$$\forall t, Y_t \geq U_t \tag{10}$$

that gives the following identification set:

$$H_1[X_t] = [\frac{C_t}{C} * Y_t, Y_t] \tag{11}$$

This assumption is that, at each moment, the capacity factor (productivity of each unit of capacity) of the installed network is higher than the capacity factor of the uninstalled network. This assumption is a very strong one too, since it is for all time t. It incorporates in the model the rationality of the agents who installed the network, but it is not very credible to think that this holds true in every moment, due to the highly unstable and random nature of the wind-turbine production. Hence the next assumption.

More productive expectation of the installed network assumption

$$\mathbb{E}(Y_t) \geq \mathbb{E}(U_t) \tag{12}$$

and Its identification set:

$$H_2[\mathbb{E}(X_t)] = [\frac{\mathbb{E}(C_t * Y_t)}{C}, \mathbb{E}(Y_t)] \tag{13}$$

This assumption is that the capacity factor of the installed network is expected to be higher in each moment. It is more realistic than the last one, but still miss something: electricity availability is not the same at each moment.
We can only have an identification set for the expectation of X_t here, otherwise, it is the same as last assumption, but with expectation operator.

More productive expected during peak periods assumption

$$\mathbb{E}(Y_t | t \in \delta) \geq \mathbb{E}(U_t | t \in \delta) \tag{14}$$

giving:

$$H_3[\mathbb{E}(X_t)] = [\frac{\mathbb{E}(C_t * Y_t|t \in \delta)}{C} P(t \in \delta) + \frac{\mathbb{E}(C_t * Y_t|t \notin \delta)}{C} P(t \notin \delta), \quad (15)$$
$$\mathbb{E}(Y_t|t \in \delta)P(t \in \delta) + P(t \notin \delta)]$$

Where delta is the set of all t that are in peak period. The value of electricity highly depends on the moment, hence this change with a peak period. The rationale behind this expectation is that the agents building wind turbines may only be interested in the production in some subset where electricity is scarce (for example during the winter). It is still not "pure" economical rationality.
Once again, we only get an identification region for the expectation of X_t.

3.3. Modeling of the Observed capacity factor

The time series of the capacity factor of the observed network will be modeled. It can be decomposed like all time series, as a trend and a stochastic term (that also includes the seasonality).

$$Y_t = T_t + Z_t \quad (16)$$

Where T_t is the trend on the data and Z_t the stochastic term.

3.3.1. Trend modeling and estimation

The estimation of the trend will be modeled with a simple linear model using only time and will be estimated with an OLS estimator.

$$T_t = \beta t \quad (17)$$

More general techniques, that estimate the trend without assuming it is linear have been created, namely in Von Sachs and MacGibbon (2000), but won't be used here. The trend is assumed to be linear.

3.3.2. Stochastic term modeling and estimation

The stochastic term will be modeled by a locally non-stationary wavelet process (LSW process) using Haar wavelet and we will estimate its evolutionary wavelet spectrum (EWS), as defined in Nason et al. (2000) and Van Bellegem et al. (2008). An intuitive explanation of the modeling and the estimation is done to clarify it to the reader. More in-depth explanation of LSW process modeling can be of course found in the cited papers.

Modeling
The intuition behind LSW process modeling of time series is to extend the spectral analysis of stationary series to non-stationary series. As a recall, a stationary time series is

a time series where the covariance is, among other things, invariant over time, that is, if the time series is denoted by Y_t:

$$Cov(Y_s, Y_{s+u}) = \gamma_y(u) \tag{18}$$

Where $\gamma_y(u)$ does not depend on s. And the spectral analysis consist in modeling it using the Cramér representation, that is:

$$Y_t = \int_{-\pi}^{\pi} A(\omega)e^{i\omega t}d\zeta(\omega) \tag{19}$$

And generates the following auto-covariance function with a Fourier representation:

$$\gamma_y(u) = \int_{-\pi}^{\pi} e^{i\omega u}f(\omega)d\omega \tag{20}$$

Where in those two equations, $d\zeta(\omega)$ is an orthonormal incremetent process of expectation equal to 0 (see Priestley (1981)), and the spectrum of this stationary process is $f(\omega) = |A(\omega)|^2$.
But we have here a non-stationary process, which means we let $A(\omega)$ (and thus the spectrum too) vary with time. Also, we use Haar wavelet and not standard Fourier representation to reach better localization in time.
The representation of the time series becomes in our case:

$$Z_{t,T} = \sum_{j=-\log_2(T)} \sum_{k=0}^{T-1} W_j(\frac{k}{T})\psi_{jk}(t)\zeta_{jk} \tag{21}$$

Where $Z_{t,T}$ is our stochastic term in time t with T being the total number of term in the serie. j is the scale parameter of our wavelets with value between -1 and $-log_2(T)$ (scales are integers, meaning T must be a power of 2, our data set must have a power of 2 number of terms). k are the different temporal locations of the wavelet. ζ_{jk} is an orthonormal increment process (like in (19)) with expectation 0 and i.i.d in j and k. Together with $W_j(\frac{k}{T})$, a Lipschitz continuous function (one for each j), we have $\zeta_{jk}W_j(\frac{k}{T})$ the random amplitude of the wavelet. ψ_{jk} is the value of the Haar non-decimated discrete wavelet of scale j at point t. Each Lipschitz continuous function has the following regularity assumptions:

- $W_j(z)$ exist for $z \in (0, 1)$
- $$\sum_{j=-\infty}^{-1} W_j^2(z) < \infty \; uniformly \; in \; z \in (0, 1) \tag{22}$$
- for each j, the lipschitz function is bounded for the total variation norm by the Lipschitz constants L_j, which are uniformly bounded and $\sum_{j=-\infty}^{-1} 2^{-j}L_j < \infty$

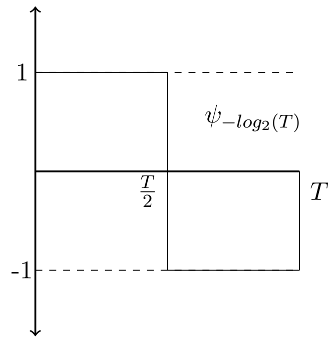


Figure 4: One example of non-decimated Haar wavelet of maximum scale

Data will be modeled with such wavelets, with j different sizes, and shifted around our rescaled time. Time is rescaled between 0 and 1 in this modeling, and the length of the data set needs to be a power of 2 to be perfectly modeled by the different wavelets (shifted and rescaled). It is the case because there is no "error term" in this modeling.
We will focus here (and in fact estimate) on what is called the Evolutionary wavelet spectrum (EWS). It is analog to the spectrum of a usual stationary process, but here defined for each scale j and over time. It can be interpreted loosely as the "intensity" of the covariance between X_t and X_{t-j} at each moment. It is defined for each j as

$$S_j(z) := |W_j(z)|^2 \tag{23}$$

For a more "in-depth" analysis of this object, one can read Nason et al. (2000).

Estimation
LSW processes are applicable to a lot of cases, as shown in Van Bellegem and Von Sachs (2004) for example. The estimation of EWS in specific values of j can help extract information from seemingly unreadable time-series that can help us understand the underlying phenomena.
The estimator we will use for the evolutionary wavelet spectrum is the smoothed wavelet periodogram. We will start by computing the empirical wavelet coefficients of our time serie $Y_{t,T}$, that is:

$$d_{j,k;T} := \sum_{t=0}^{T-1} Y_{t,T}\psi_{jk}(t) \tag{24}$$

Where $\psi_{jk}(t)$ is the value of our wavelet (here, a Haar Wavelet) of scale j shifted in position k in time t.
The square of this serie is the wavelet periodogram $I^j_{k,T}$.

$$I^j_{k,T} = (d_{j,k;T})^2 \tag{25}$$

But there are problems with this first "naive" estimator. One is that by proposition 4 in Nason et al. (2000), it is not an asymptotically converging estimator. It is biased. It needs to be corrected. The second problem is that it has non-vanishing variance. We need to smooth it to obtain consistency.

For the bias, it comes from the fact that information of other scale j is incorporated. it has been shown that we can compute a corrected empirical periodogram, $L^j_{k,T}$, by simply multiplying the vector $I^j_{k,T}$ by the inverse of a correction matrix, existing for each j, A_j. So:

$$L^j_{k,T} = A_j^{-1} I^j_{k,T} \tag{26}$$

For the smoothing problem, different techniques can be used, for example, Translation Invariant (TI) denoising, as defined in Coifman and Donoho (1995). But for our case, a more "classical" method will be used, local linear smoothing, using different sizes for the binwidth (on a data set of more than 250000 data points). The size of the binwidth can be large (and so an estimated EWS very smoothed), motivated by the fact that we are interested mainly in the big picture of the seasonality. Both the smoothed and unsmoothed versions of the corrected periodogram will be presented.

3.4. Off-shore ratio and Capacity factor variance

As explained before, the evolutionary wavelet spectrum (EWS) that we estimate can be roughly interpreted as the contribution to the variance of each scale j. In the fashion of Van Bellegem and Von Sachs (2004), we can try to take a specific scale j, and try to link this curve to other data we have. We observe a "point" variance with a specific scale of the estimated EWS. In our case, if the data can be properly modeled by a single scale j, we will use it. During the analysis, it will be shown that a specific scale is indeed where most of the information is.
Shimon Awerbuch, in its seminal paper Awerbuch (2006), proposes to see each energy source (gas, coal, solar panels, wind turbines) as different assets of a portfolio, on which we can do standard portfolio management just as we do in finance. Each asset having its own volatility and expected return (production in our context). We have here 2 different types of wind turbines we could model as different assets: off-shore and on-shore wind turbines. Each having its inner volatility.
Since we have the composition of the different portfolio we observe, we can see which portfolio, historically, have lower volatility and from that know which of on-shore and off-shore wind turbines, in the context of Belgium, have more volatility. We can model standard portfolio volatility with covariance between our two assets and get that the corrected empirical periodogram value, at time t (we rescale time to our original parameter t), for a specific value of j,g is:

$$L^g_t = w^2_{t,0}\sigma^2_{t,0} + w^2_{t,1}\sigma^2_{t,1} + 2w_{t,0}w_{t,1}\sigma_{t,0,1} \tag{27}$$

Where $w_{t,0}$ is the weight in our portfolio of off-shore wind turbines and $w_{t,1}$ is the weight in our portfolio of on-shore wind turbines. Different σ^2 are the variance of each asset and $\sigma_{t,0,1}$ are their covariance.

We have the weights and the value of the EWS at scale g, and we want to draw conclusions on the values of the variance and the covariance. We could compute those from data, but unfortunately disaggregated data for both types of wind turbines doesn't exist. We will try to estimate the variance and covariance parameters with our data. We adapt (27) and get:

$$L_t^g = w_{t,0}^2 \bar{\sigma}_0^2 + w_{t,1}^2 \bar{\sigma}_1^2 + 2 w_{t,0} w_{t,1} \bar{\sigma}_{0,1} + \varepsilon_t \tag{28}$$

Where The bar operator means the arithmetic mean and ε_t is a deviation from what we predict that we assume only depends on the meteorological conditions and not on the different weights.

With those results and knowing the return (here our serie of observed capacity factor) of our portfolio can be defined as:

$$Y_t = w_{t,0} Y_{t,0} + w_{t,1} Y_{t,1} \tag{29}$$

We can also estimate it from our data using OLS and following modeling:

$$Y_t = w_{t,0} \bar{Y_0} + w_{t,1} \bar{Y_1} + \delta_t \tag{30}$$

Where $\bar{Y_x}$ is the mean capacity factor of the network containing all installed wind-turbines of type x and δ_t is a deviation parameter from what we predict and that we assume only depends on the meteorological conditions, and not on the different weights.

First thing is that we cannot expect this to be normally distributed, it will have a seasonal impact, even if its mean is by construction is 0. Meaning we won't be able to interpret the standard deviation that we will observe.
Both the volatility and the return of our portfolio will be analyzed, but we will keep in mind that we have that non-normally distributed error term in both regressions.

3.5. External limitation of the production and capacity factor

Before this section, we model the capacity factor of the observed network in a way that suggests that the installed network produces at each moment as much as it is physically possible at this moment for it. But it is not always the case. It makes sense that wind energy shouldn't be bounded by anything else but the weather and the capacity. But in reality, it is not this simple.
Problem is, in practice, due to the way the electrical grid works, it is sometimes needed to limit wind energy production. For a grid to continue its operation, it needs to equalize production and consumption at each moment. It is done by doing prediction on intermittent energy production, consumption and fills the gap between the two with dispatchable energy production (typically gas). But sometimes, wind over-perform and thus needs to be limited. In this case, the grid operator (in our case Elia) issues decremental bids. It is a mechanism that will pay the owners of the wind turbines to not produce during a specific time.
It is pretty rare though. In our 8 years of data, it happens 500 times (500 times 15

minutes, about 5 days). In the following analysis, we keep those data points, because of course, the fact that decremental bids are active is not independent of the production of wind. In a robustness check, our analysis will be redone without those data points to see if the results change significantly.

Another thing to take into account, in our partial identification part, is that, of course, we cannot observe a production equal to more than what is consumed. It means, in practice, that the production of the network is bounded by the total load of the network, meaning:

$$R_t = \min(\sum_j Z_{tj}\phi_j, L_t) \tag{31}$$

where everything is defined just like before, and L_t is the total load of the network in time t.
Fortunately, we have access to the data of the Total load of the network, and we can compute in expectation, the value of L_t for a typical year. This is what will be used in our computations.

3.6. Parameters of interest

3.6.1. Estimator of the EWS, Variance, and Covariance of Off-shore and On-shore wind turbines

We are interested in the Estimator of the EWS to be able to measure the risk of wind energy in Belgium in a meaningful way and see what are the seasonal effects on it. We are also interested in the variance and covariance of the two types of wind turbines because portfolio mixing is possible between those two assets meaning that we can manage some of the risk by knowing the properties of those two types of wind turbines.

3.6.2. Identification set of expected wind production for a year

With the different assumptions we mentioned before, we can compute different identification sets for the expected wind production for a year. We are interested in this because it can tell us what is the potential of wind energy in Belgium, which is one of our objectives. The Problem is all of them depend on the capacity of the complete network, C, that, by construction, we don't observe. There are ways we could compute C for Belgium (see Hillring and Krieg (1998) for a methodology and Lejeune and Feltz (2008) for such a computation for the southern half of Belgium), and further work should be done in this direction to be able to be more precise in the definition of the potential of wind turbines. But for this work, we will compute such sets for different values of C, under different assumptions and see what conclusion we can draw from that.

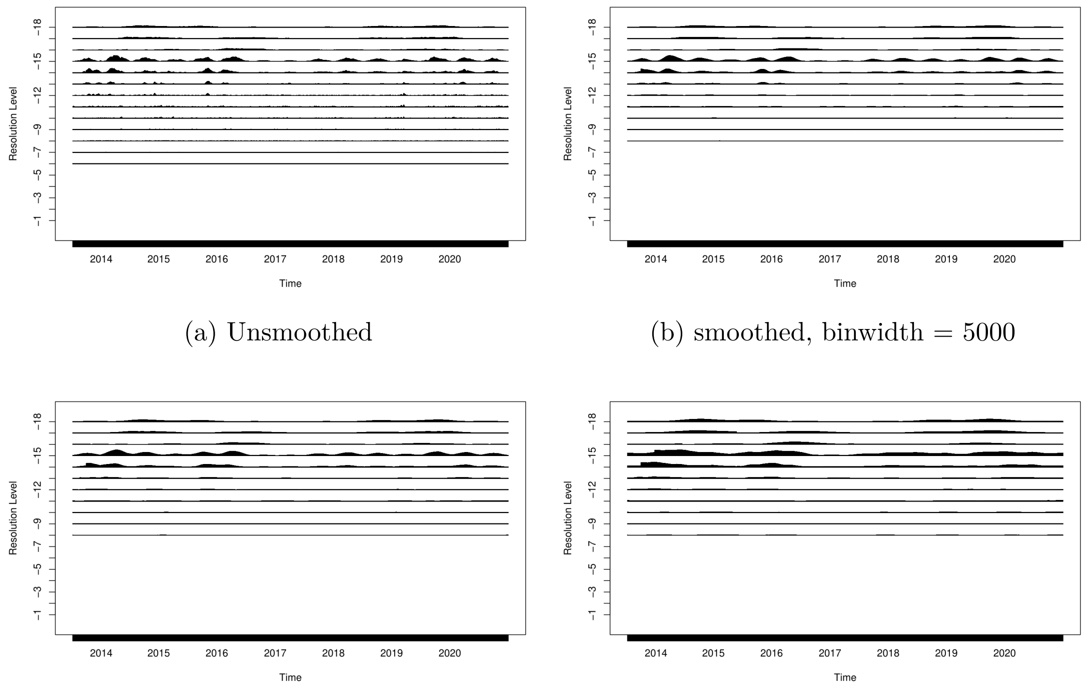
4. Results

This section will first present the main results we got from the data and then develop a number of robustness checks.

4.1. Main results

4.1.1. Wavelet decomposition

First, let's see the modeling of our dataset with locally stationary wavelets. The results presented follow the estimation methods explained in the last section, meaning that first we estimate the linear trend on our data and remove it. There is a negative trend. 4 results are presented. Without smoothing, with smoothing with a binwidth of 5000, 10000 and 20000. As explained before, the data must have a length equal to a power of two to be modeled perfectly by wavelets. The first months of 2013 are thus removed to arrive to a power of two (262144 data points). If line is not present, it is because it is insignificant in the estimation.



(a) Unsmoothed

(b) smoothed, binwidth = 5000

(c) smoothed, binwidth = 10000

(d) smoothed, binwidth = 20000

Figure 5: Estimated Evolutionary Wavelet Spectrum of the Capacity factor

First, we can see that a too big binwidth makes periodicity in the spectrum go away. For example, we can see that a binwidth of 20000 is probably too much considering it

makes the biannual periodicity of scale -15 go away.
Another thing to note is that there is a 6 months periodicity on some scales (namely -15). We can also see that the resolution level -15 is the most significant. We can see it "bleeds away" on -14 level. It is a very common occurrence when estimating EWS.

The clear seasonality that we can see shows us the impact of seasons on the risk of wind turbines. It was expected considering this energy source depends on physical flux.

4.1.2. Portfolio Analysis

For the rest of the analysis , we will focus on the -15 level estimation because this is where most of the information is, and we will assume in fact that all the information of our data can be summarized by this scale of the EWS. We can see what is the link between this variable and the weights of off-shore and on-shore wind turbines on the grid, computed following (27). We do that with an OLS regression as explained in last section and get:

estimated EWS	$\bar{\sigma}_0^2$	$\bar{\sigma}_1^2$	$\bar{\sigma}_{0,1}$	adjusted R^2
unsmoothed	0.0172 (0.0121)	0.01636 (0.0043)***	0.0224 (0.0146)	0.168
smoothed, binwidth=5000	0.0242 (0.0012)***	0.0198 (0.0004)***	0.0124 (0.0015)***	0.9471
smoothed, binwidth=10000	0.0256 (0.0010)***	0.0203 (0.0003)***	0.0106 (0.0012)***	0.9633
smoothed, binwidth=20000	0.0237 (0.0007)***	0.0195 (0.0002)***	0.0132 (0.0009)***	0.981

Table 1: Regression of weights on estimated EWS level -15.

It is presented with estimates and their standard error in parenthesis, with stars showing their significance if the standards errors (of the regression) were normal. Let's keep in mind they are not and thus the standards errors are not interpretable. For the unsmoothed estimated EWS, as explained in last section, it is not a converging estimator, meaning that standard errors are inflated without smoothing, making everything but one parameter insignificant.
Variation in weights of off-shore and on-shore is not yearly or daily, so the size of the binwidth is justified, it will not make go away the effect of the change in weights on the grid, even with 20000. What happens is that the smoothing makes other (seasonal...) variations go away and inflate the R^2. We have no way to know at which point the smoothing is perfect and at which point it becomes too much and starts to inflate the R^2.

It makes sense that the covariance is positive. Due to the small size of Belgium. A

negative covariance would have meant that when there's a lot of wind in the country, there is less wind in the sea, and the opposite. Such negatively correlated grids are very interesting to make intermittent energies less risky.
Also, we can see that the variance of the off-shore network is consistently higher than the variance of the on-shore network. In the robustness checks, other methods to try to confirm our results will be presented.

Let's now do the same analysis on the capacity factor itself. doing like in (30), we get:

	$\bar{Y}_0$	$\bar{Y}_1$	adjusted R^2
Capacity factor	0.4038 (0.0076)***	0.1881 (0.0045)***	0.5763

Table 2: Capacity factor portfolio decomposition estimation

Once again, those results are subject to caution due to the fact standard errors of the regression are not supposed to be normal. But what we see is in line with previous observations: off-shore wind turbines have a significantly higher capacity ratio than on-shore wind turbines. Disaggregated data would be better to estimate that, but it is in any way in line with previous observations in other settings. Now we have more information on our two assets. Implications of those results in an optimal portfolio approach will be discussed in next section.

4.1.3. Results on partial identification

Before presenting the results of partial identification, let's discuss a bit the production. In the next graph is presented the percentage of electricity each year produced by wind turbines. It is computed from the disaggregated data, by summing production in every 15 minutes over the whole year and dividing it by the total consumption over the whole year.
We can, from our analysis, compute the average load (energy consumed) every fifteen minutes between 2014 and 2020. We can also compute the average capacity ratio for every fifteen minutes between 2014 and 2020. From that, we can, for any given installed capacity, compute the percentage of electricity produced by wind turbines, assuming it is only bounded by the load (meaning that whenever we can, we will

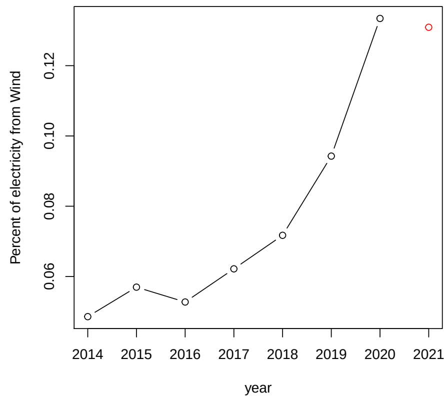
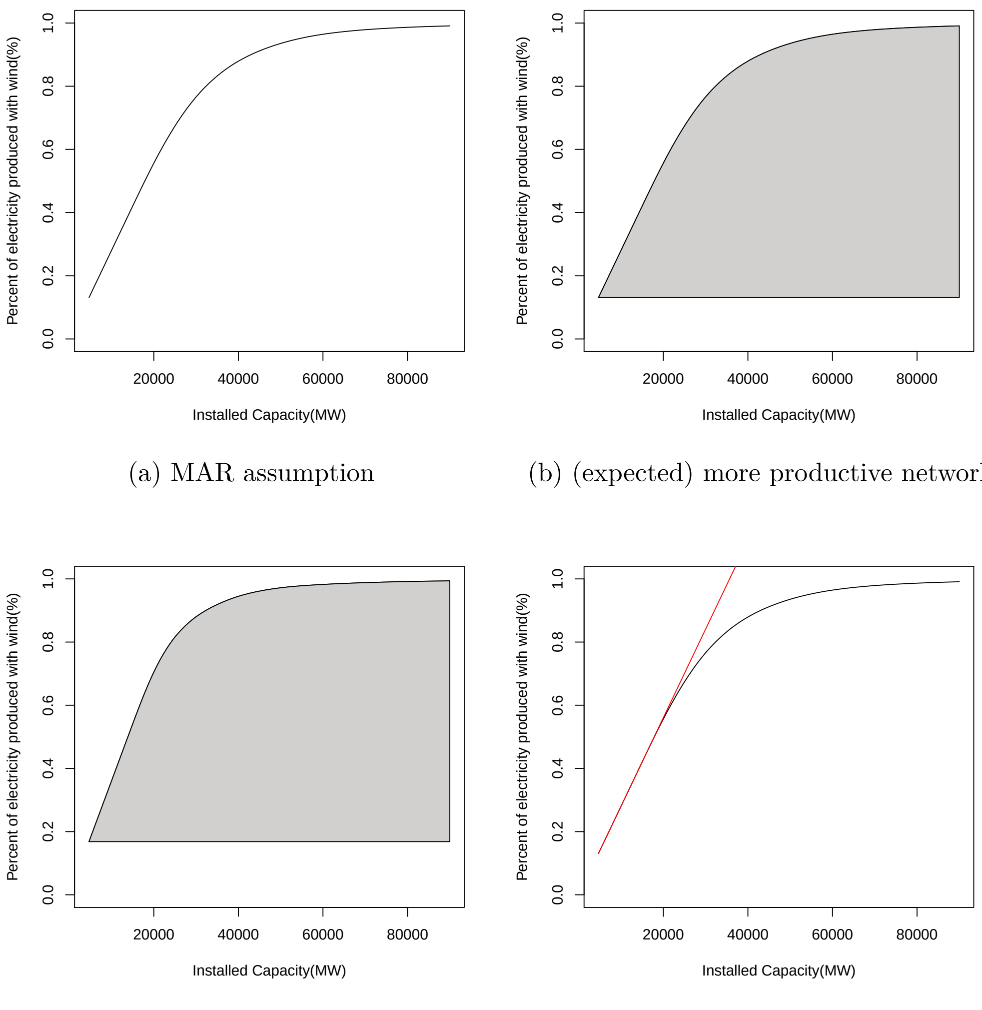


Figure 6: Percentage of electricity produced by wind turbines (black) and prediction with current network (red)

use wind turbines). If we do that for the current capacity and an average year, we get the red dot on the graph. It is what, on average, would be the percentage of electricity produced by wind turbines with the current network. We can see that our prediction is not far from what is observed in 2020. A robustness check will do that for each year in our dataset later.
Now will be presented the computed identification region for the percentage of electricity produced by wind turbines over an average year, for a lot of different values of the network. The graphs presented will show on the x-axis the size of the network (from the 4670 currently installed MW to 90000MW), and on the y-axis the computed identification region of the expected percentage of electricity that will be produced with it.



(a) MAR assumption

(b) (expected) more productive network

(c) more productive in peak period

(d) Intertemporal Batteries (in red)

Figure 7: Identification region of the expected percentage of electricity produced by wind depending on the installed capacity

In grey is the identification region of the expected percentage of electricity produced by wind turbines for each value of the installed capacity. We have no idea what is the max possible installed capacity, so we cannot pin down the exact potential, but we can see on the graphs that, without batteries, the marginal addition of new installed capacity will at some point not help to produce more electricity. An analysis that would smooth production(over a day, a month, a year) could not see that. It is only possible because we have very high-definition data. We see that at around 55 % of electricity produced by wind, we go from a linear increase of electricity produced with increasing capacity to an asymptotic increase to 100 %. The intuition is that at some point, the new wind turbines will produce electricity, but only when 100 % of demand is already served, making them worthless except in some marginal situations. Keep in mind that the MAR assumption is very optimistic, the second graph with its identification region is more realistic. Finally, we have no idea where we will fall in this identification region, but we know both borders of it are “corner cases” that won’t happen.
The third graph shows the percentage of electricity only during the peak periods (here defined to be December and January), where we can see that it is steeper than the second graph. It is indeed good news if the point of the planners is to produce more electricity during this period.
The last graph show what would happen if the linear relation between additional capacity and additional production would continue. It is what would happen if we had a lot of batteries to smooth the consumption over the production. With this graph, we can see that the utility to have batteries is increasing with the installed capacity.
Data for production (in TWh/year) and percentage of grid demand with and without batteries are presented in the table below for different installed capacity.

	Now (4670MW)		10000MW		20000MW		40000MW		60000MW		90000MW	
	%	prod	%	prod	%	prod	%	prod	%	prod	%	prod
MAR	13.1	11.3	28.0	24.2	55.6	48.1	87.9	76.0	96.4	83.4	99.1	85.6
Battery			28.0	24.2	56.1	48.5	112.1	96.9	168.2	145	252.2	218

Table 3: Percent of electricity and production (in TWh/year) with and without batteries

We can see that, beginning at about 4 times the current network (20000MW), having batteries start increasing the production of wind turbines. How is it so ? Because without those batteries, at some moments, more than 100 % of the electric demand will be satisfied by wind turbines, hence the additional production will be lost. Note that in theory Belgium could export it, but the problem is that neighboring countries have similar wind regimes, and will probably be in the same situation as Belgium when Belgium is in a position to export electricity.

Data presented here is if the network stays with the same off-shore and on-shore ratio. Since we don’t have disaggregated data, we cannot, without smoothing assumptions, know what would be the result if we only added off-shore or on-shore wind turbines. The most probable scenario is that if we add a significant amount of off-shore wind turbines, the percentage of electricity produced with wind would climb more than with the same significant amount of capacity on-shore since the capacity factor of off-shore seems to be higher. But we cannot be sure without disaggregated data.

Now some robustness checks will be presented.

4.2. Robustness checks

4.2.1. For Portfolio analysis

Aggregate per Ratio

We've seen in our results that the variance of the off-shore wind turbines seems to be higher than the one of the on-shore wind turbines. Here, we will aggregate the mean observed L_t^{-14} for each value of $w_{off-shore}$. That is for each different portfolio we observe. We get the following graph, with, in red, a simple linear regression of it:

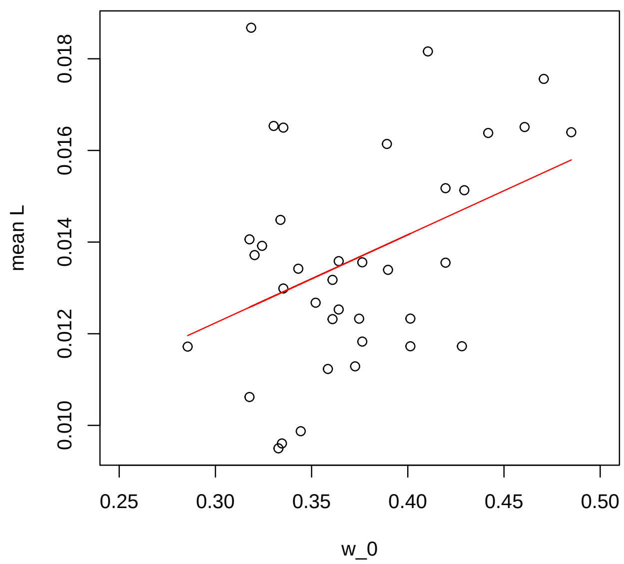


Figure 8: Mean L^{-14} per portfolio vs w_0

So, what we observe is that, apparently, our risk seems to go up when w_0 goes up in our portfolio. We only have data for $w_0 < 0.5$. There are two explanations for that. Either $\bar{\sigma_0}^2 > \bar{\sigma_1}^2$, making the risk go up when w_0 goes up. Or $\bar{\sigma_{0,1}} > 0$, making the risk go up when we mix more our portfolio (so when we get near $w_0 = w_1 = 0.5$). More will be said later about the covariance between our two assets, but we estimated it to be positive before, making it a part of the positive trend we observe. Also, we observe indeed that the mean variance of off-shore is higher than on-shore. With available data, we cannot disentangle the two effects using this aggregated per portfolio approach. We need to either observe a portfolio with more than half of off-shore or to simply observed directly the variance and covariance with disaggregated data.

Does Decremental bids have an effect?
To assess if the data points with decremental bids (= wind energy production is limited due to the lack of need of more electricity) have an effect on our regression, we will redo our analysis without them. On our censored sample, we reestimate the EWS with binwidth 10000, then we rerun the regression on the 14th scale to get estimates for $\bar{\sigma_0}^2$,$\bar{\sigma_1}^2$,$\bar{\sigma_{0,1}}$. We get:

estimated EWS	$\bar{\sigma_0^2}$	$\bar{\sigma_1^2}$	$\bar{\sigma_{0,1}}$	adjusted R^2
smoothed, binwidth=10000	0.05876 (0.0106)***	0.02490 (0.0003)***	-0.01657 (0.0012)***	0.9625

Table 4: Regression of weights on estimated EWS level -15 (without decremental bids)

We can see that the results of the regression are very different. It supports the analysis that tells us that removing decremental bids would bias significantly the data due to the fact that having the presence of a decremental bid is linked with very good performance of wind energy. We can see that our basic result that $\bar{\sigma_0}^2 > \bar{\sigma_1}^2$ still hold here. Let's keep in mind that the results we get in our analysis with the decremental bids are also biased due to the fact that data during those times are censored.
The most strange thing that we observe here is that the covariance between the two assets is negative. It is surprising considering the small size of Belgium, we would expect wind over the north sea to be somewhat the same as the wind in Belgium. Once again this result is biased and thus it probably is not the case that covariance is negative, but, as stated multiple times before, disaggregated data could really help us compute this parameter. The sign is of importance because if the covariance is negative, it means that there is room for diversification to manage the inherent risk of wind turbines.

4.2.2. For Partial identification

Predict expected percentage per year in historical data
If our approach of estimating the expected percentage of electricity is correct, it should predict it too for historical data. That is what we do here. We compute yearly production vs our expected yearly production. In red, our prediction, in black, the actual production. A big part of the difference can be attributed to change in the off-shore / on-shore ratio. Years where there are a lot of off-shore, it is underestimated (2014, 2015, and 2020) and overestimated when it is the opposite. Without disagregated data, it is not possible to

correct for that without making big smoothing assumptions. It is way off in 2020 due to the fact that the structure of the electricity demand in 2020 was very different from a “normal” year due to the disruption due to the pandemic.

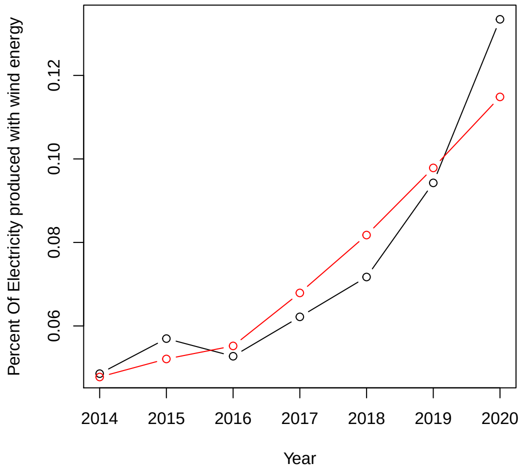


Figure 9: Percentage of electricity produced by wind turbines, estimation (red) vs actual (black)

Is the decreasing capacity factor assumption in historical data ?

The key assumption that we present is that the capacity factor of the installed network should be higher than the capacity factor of the uninstalled network because the rational builders of wind turbines will first build wind turbines where it is the most profitable, installing later the less productive one, making the current capacity factor at each moment an overestimation of the capacity factor of the complete network. We use different variations of this assumption (notably, that it holds in expectation but not at each moment). If this assumption is correct, we should already see the decreasing capacity factor of wind turbines in action in our data. We will try to show that we can see it.

Let’s measure the expected capacity factor for each different capacity that we observe. Let’measure the capacity factor free of seasonal interference because we don’t observe the capacities for full years but only some moments, and using raw capacity factor will systematically bias our results. We take only networks that have been online for at least a month otherwise unnecessary noise, due to the change in meteorological conditions, will be in our regression. We can do a weighted (by the time it has been online) linear regression of each network capacity by its expected deviation from the seasonal mean at each moment.

Here is the graph. It shows deviation from seasonality (since we remove seasonality to avoid disturbances) for each network, each network bubble is bigger with the amount of time it has been online (and thus the weight it has on the regression.) We don't have a very negative trend, but we have a trend anyway, and a very significant one if the error term was normal (which we can assume here). The regression tells us that the deviation from the seasonal mean will decrease by $9.898 * 10^{-7}$ per unit of capacity added. The standard error of this measure is $5.119 * 10^{-8}$, meaning we have a p-value for this parameter of $< 2 * 10^{-16}$. This result seems to support the fact that our assumption is correct. Let's keep in mind that since capacity is increasing with time, that this decreasing effect could be due to climate change (if wind speeds were to decrease in Belgium over time).

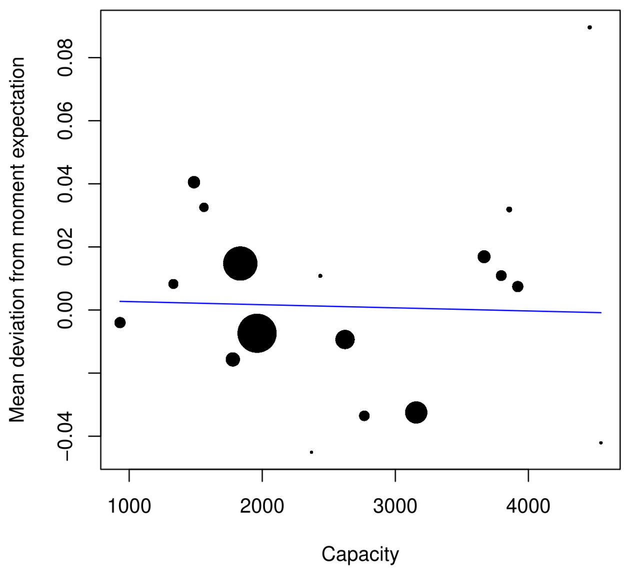


Figure 10: Capacity and mean deviation from seasonality for each network (black) and regression (blue)

5. Discussion

5.1. Wavelet analysis

From the estimated wavelet spectra that we presented in the last section (each with a different smoothing), we can see that almost everything is in the same -14 scale of the haar wavelet. We can see that the same structure "bleeds away" on lower scales, which is a pretty common occurrence when analyzing wavelet spectrum.

We can see that there is a clear 6-month seasonality on this scale. Let's keep in mind we have not, prior to the wavelet spectrum estimation, unseasoned the data. Data is

untrended, for a technical reason already stated before. It makes sense that we observe this seasonality in there. Seasonality seems to go by pair of 2 times 6 months, where the censored sinusoidal curve we observe has a different amplitude. This amplitude should be investigated further and we have there probably a good parameter to characterize a year of wind production.
Once again, even if it has been stated before, let's keep in mind the smoothing process, that we must do because the raw EWS is not consistent, smooth away some seasonality in the data. In fact, when we observe the curve with the binwidth equal to 20000, we see that it is indeed too large and masks the 6-month seasonality that we spoke about before. Caution needs to be applied when analyzing the smoothed data.

We have now a way to model the seasonality of the wind energy capacity factor. It seems to permit us to summarize well with a single wavelet scale that varies over time in intensity. But of course, since this seasonality is due to meteorological conditions and overall the weather, we cannot act on it to minimize it per se. It is more a fact of nature that we need to take into account for wind energy in Belgium (more analysis is needed to show if this 6 month seasonality also appears in other countries). This analysis thus completes the first objective of this work that was given at the beginning. We have now a way to measure the risk of wind energy, in the case of Belgium, depending on the month and year. If we had disaggregated data, we could think of ways to tackle it, but for now, we can only observe it. One parameter that we can act on is the number of on-shore and off-shore wind turbines. Let's discuss that now.

5.2. Portfolio analysis

Before speaking about our results, let's re-explain briefly the ideas developed by Shimon Awerbuch in Awerbuch (2006), because it is needed to understand the differences with what is said here.
What is developed in this cited paper is the idea that we can see different sources of energy (wind, gas, oil) as different assets in a portfolio where the return of the assets is the energy they produce. It is assumed that the point of view is over a year. The main source of uncertainty that is to take into account in this approach is the volatility in the price of fossil fuels, which, of course, will induce a significant cost to society due to the costly adjustment that will have to take place due to this volatility in price. He then develops that renewable energy can help with that because, even if they cost more per kWh of energy produced (meaning their mean return is lower), their cost is basically constant, since there is no significant variability in the price of wind turbines over time. At least, not at the scale of the variability on the price of oil or gas. We thus have a risky but rewarding asset versus a stable but less profitable asset, meaning there is room for portfolio mixing.

From this original paper, we try to add two things. The first is that, in fact, renewable energy has, too, inherent risk, be it not of the same kind. The second one is that wind turbines are not homogenous and cannot be averaged to a single asset like we would with oil (due to the fact that the price of oil is global). There exist at least two different types of wind turbines that are very different. We could even speak of a country as a continuous space of possible wind turbines that would have specific returns and risks. Let's continue to speak about those two points.

The inherent risk that intermittent energy sources face is not a risk on the price like dispatchable energy sources have. It is a risk on the possibility of producing directly. For a gas plant, we can know that it will be able to produce electricity, and which amount throughout the year. The only thing we don't know is at which price the gas will be bought. Here, since the production for intermittent energy sources (which wind energy is part of) depends by definition on a physical flux such as wind speed, for example, it is the ability to produce that is in question. If there is no wind, it is impossible to produce more without increasing capacity, no matter the price you are willing to make.
The difference between those two risks is one of time-perspective. From the point of view of an average year, it is possible to adapt capacity of wind turbines. But we don't live in a theoretical year smoothed production, we live in a specific moment. This is the limit of the Awerbuch approach that here we try to extend, by using high definition data and measuring this risk of wind turbines at each moment.

In our analysis, we tried to identify parameters for the two different types of wind turbines that we have installed capacity data for: on-shore and off-shore. But there is no reason to stop there, we could do a way more precise portfolio analysis if we had information on each wind farm in Belgium. But here, with the data we have, we can identify parameters for off-shore and on-shore wind turbines. What seems to hold from our results is that the risk of off-shore wind turbines is higher than for on-shore (in our results, that $\bar{\sigma_0}^2 > \bar{\sigma_1}^2$) and that the capacity factor is higher for offshore. The second result is coherent with the observations made in other settings, for example in the UK, as seen in Department for business and industrial strategy (2020) recently (but earlier digests of the same government agency show exactly the same). Actually, it makes sense. If it wasn't the case, there would be no reason to build off-shore wind turbines at all.
With our results, we can see that we can do some portfolio-mixing with the 2 assets to get minimal risk for each level of return. Let's keep in mind 3 things though.
First, the price of building an off-shore wind turbine is of course higher than building an on-shore one for obvious construction reasons. For this reason, when we compare a MW of off-shore and on-shore capacity, we cannot expect them to be two different assets with the same price. It is out of the scope of this work to be interested in the price of those since it is a complex matter (there are permits, electrical materials to adapt the produced power for the grid).
Secondly, we cannot build as many wind turbines as we want, especially off-shore. The Belgian territory in the north sea where those can be built is not that big and a significant part of it is already took by wind turbines. Actually, until 2019, no new project was possible, until the government extended the area where building wind turbines is possible. We can see that in the following map. Sources for the location of offshore wind turbines are from Royal Belgian Institute for Natural Sciences (2020).

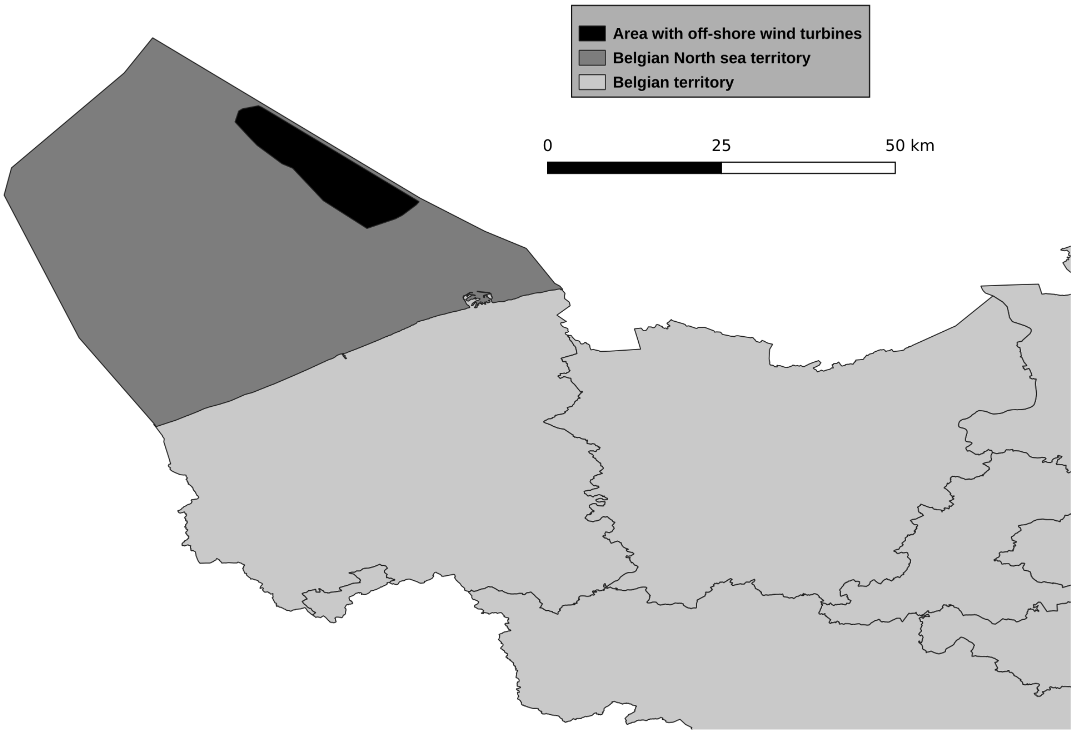


Figure 11: Off-shore wind turbines in the Belgian North Sea territory

Lastly, it would be way better if the point is to use portfolio diversification techniques on wind turbines to have access to disaggregated data of production, in high definition, for off-shore and on-shore wind turbines.

Even if the covariance is positive, there is room for diversification here and to choose a portfolio on the efficiency frontier of return/risk. This approach could be extended with a high definition differentiation of wind turbines in different assets depending on which wind farm they are part of. This approach could also be extended to other renewable energy such as solar panels, for which high-definition data is available.

All of that said, even if only the intuition is presented here do the unreliable way in which the variance and returns are computed, that we could manage some of the inherent risk of wind energy production by using a portfolio approach.

5.3. Wind potential in Belgium

In our results, we get different identification region for the percentage of the electricity we can produce with a given capacity of wind turbines, under some assumption and the assumption that the ratio of off-shore and on-shore wind turbines stay more or less the same. As explained before, if the ratio change it becomes difficult to have an identification region because we don't have high definition data for off-shore and on -shore, and the only way to get a solution is to use smoothing solutions like "off-shore will always produce more", which are not in the spirit of the analysis here trying to focus as much as possible on not smoothing the result, because such smoothing leads to the underestimation of the inherent risk of wind energy.

The current identification region is based on the assumption that, on average, the most productive wind turbines already have been built. It stems from economic rationality, that, first, most productive wind turbines will be built. We have proved in a robustness check that, indeed, the capacity factor of the network seems to go down with capacity with time (This decreasing capacity factor only works if the ratio of on-shore and off-shore stays the same, once again). As we can see in the graph below, that shows all the wind farms in Belgium (data from OpenStreetMap contributors (2021)), area where the population density is too high (around Brussels, in the Walloon Brabant), there are no wind turbines, because the price of putting them there would be too expensive compared to their production (due to a high price of soil in those areas).
The point of showing that is to say that wind turbines are indeed installed following some sort of economic rationality by the different firms in this market. Other assumptions based on economic rationality could be used, and further research in this way could be interesting to help build a smaller identification set for the production of wind turbines in the future. One key thing to understand here is that we don't have point identification precisely because of economic rationality. If the agents were building wind turbines at random, our installed capacity would give us meaningful information on the complete network.

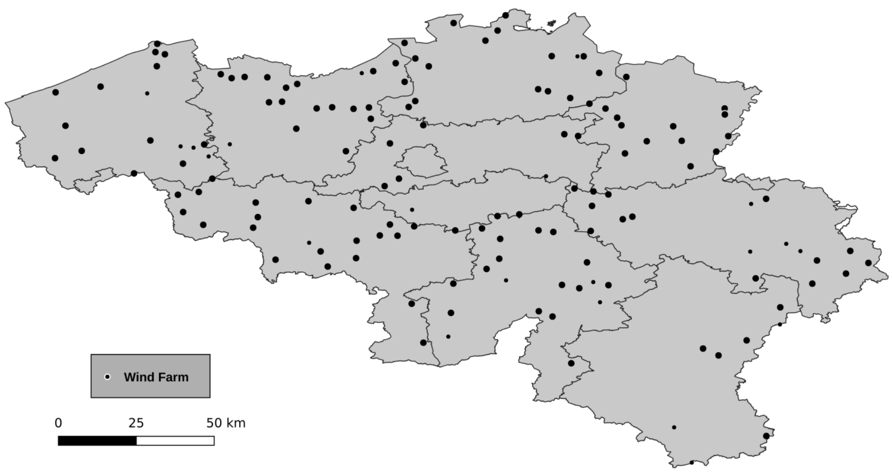


Figure 12: Wind Farms in Belgium (2021)

What our results show too is that if it is possible to get to 100 % of the electricity produced by wind turbines, it is pretty wasteful if no energy is exported or if no batteries are installed because at some point the marginal productivity of a MW of wind energy capacity becomes small due to the inability to use the additional production if consumption is already 100 % served. Of course no one proposes to use wind energy as the only source of energy, but it is important to keep that in mind even though this problem of saturation is far away, as we estimate it to start around 20000MW installed, so about 4 times the current network. In that, the conclusion of this work is pretty standard for the literature but here based on high-definition data of historical production; a stable and sustainable electrical grid must rely on a variety of different energy sources and not put everything in the same type of energy, even more strongly if those energy sources are intermittent and not dispatchable.

Another thing that we can see, is that the usage of batteries becomes increasingly more useful to make production go up as the installed capacity increases. More should be done to prove that it is also the case with solar panels (even though the same intuition that batteries permit to smooth production over longer time periods). With the capacity of intermittent energy sources increasing rapidly in Belgium, more should be investigated in ways to store electricity, be it for small periods of time.

The curve that we compute for the expected part of electricity produced by wind turbines could be extended in more dimension with other intermittent energy sources, giving a convex surface that could help policymakers in what is the most interesting energy source in each situation. An intermittent energy source in our setting is not good or efficient by itself but must be compared to the electricity demand it can actually satisfy. A very productive intermittent energy source that only works when no one wants electricity has no inherent value, at least in the absence of electricity storage and/or export of electricity.

We also assume that climate is stable over time in Belgium and that thus wind speeds won't change, which is a pretty strong assumption to make. Another possible extension for this model is to add exogenous change of the wind speed due to climate change, that could be based on the different climate models made for Belgium, see Termonia et al. (2018) and De Troch et al. (2014).
the last assumption that we make is that consumption of electricity will stay comparable in the future to what it was during the period 2014-2020. It is of course not the case, and additions to this model should be done with proper modeling of the electricity consumption. It was purposely ignored here to focus on the intermittent production side of things. Consumption, on another hand, is difficult to predict in long term due to the shocks that could happen on it for example, does the car park electrify, at which rate? How does evolve technology and efficiency, and so on and so on.

We here don't make estimations or assumptions on C, the total installable capacity of wind turbines in Belgium, because, even though some papers tried to compute it in some settings (Justus et al. (1976), Lejeune and Feltz (2008) and Hillring and Krieg

(1998)), we don't have yet a solid method to do it at the scale of a whole country, and no idea on how to do that for offshore wind-turbines (that are, as said multiple times, about 40% of the current installed capacity in Belgium).

6. Conclusion

We wanted to measure the two types of uncertainty of intermittent energy, for the case of wind energy production in Belgium. The first one is the uncertainty around the availability of the physical flux on which the intermittent energy is based, here wind speed. a LSW process is used to model that. The second source of uncertainty is about the characteristics of the future installed capacities (here, we use partial identification and assumptions stemming from economic rationality to be able to assess this uncertainty).
In this thesis, starting with data of wind energy production in Belgium between 2013 and 2020 for each interval of 15 minutes, we model the evolution of installed wind turbines as different "networks" that need to be analyzed in relation with a future network, the "complete" network. we then develop modeling of the possible capacity factor of the future wind energy network in Belgium using locally stationary wavelets and partial identification.
From that, we estimate the risk of each type of asset (off- and on-shore) and their expected capacity factor. We argue that we can do some portfolio mixing to get to any level of risk and return for our wind turbines network but that disaggregated data is needed.
Then, we compute under different assumptions the identification set of the production of the future network, in percentage of electricity consumption, over different capacities from the current capacity (a bit less than 5000MW) to 90000MW. This permits us to see the potential of wind energy in Belgium and also see that the usage of batteries (or any way to store electricity between time periods) would really help achieving higher production for wind turbines. We could not see that without high-definition data because smoothed data over a day, a week, a month or a year would make this source of uncertainty go away.
Further extensions are needed to make this model more credible in predicting risk and return of wind energy in Belgium, for example around the forecast of consumption of electricity, climate change, limitations of off-shore capacity... Those extensions are purposedly not mentioned here to stay clear.

Contents

References

Awerbuch, S. (2006). Portfolio-based electricity generation planning: policy implications for renewables and energy security. *Mitigation and adaptation strategies for Global Change*, 11(3):693–710.

BOP (2020). Projects. Belgian Off-Shore Platform, https://www.belgianoffshoreplatform.be/fr/projets/.

Coifman, R. R. and Donoho, D. L. (1995). Translation-invariant de-noising. In *Wavelets and statistics*, pages 125–150. Springer.

De Troch, R., Giot, O., Hamdi, R., Saeed, S., Tabari, H., Taye, M. T., Termonia, P., Van Lipzig, N., and Willems, P. (2014). Overview of a few regional climate models and climate scenarios for belgium. *Scientific and technical publication*, (65).

Department for business, e. and industrial strategy (2020). Digest of uk energy statistics for 2020. Available at : https://www.gov.uk/government/statistics/digest-of-uk-energy-statistics-dukes-2020.

Elia (2021). Wind power generation. data retrieved from Elia grid data available at, https://www.elia.be/en/grid-data/power-generation/wind-power-generation.

Fryzlewicz, P., Van Bellegem, S., and Von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics*, 55(4):737–764.

Heussen, K., Koch, S., Ulbig, A., and Andersson, G. (2011). Unified system-level modeling of intermittent renewable energy sources and energy storage for power system operation. *IEEE Systems Journal*, 6(1):140–151.

Hillring, B. and Krieg, R. (1998). Wind energy potential in southern sweden—example of planning methodology. *Renewable Energy*, 13(4):471–479.

Justus, C., Hargraves, W., and Yalcin, A. (1976). Nationwide assessment of potential output from wind-powered generators. *Journal of applied meteorology*, 15(7):673–678.

Krohn, S., Morthorst, P.-E., Awerbuch, S., et al. (2009). The economics of wind energy. *European Wind Energy Association*, 3.

Kusiak, A., Zheng, H., and Song, Z. (2009). Models for monitoring wind farm power. *Renewable Energy*, 34(3):583–590.

Lejeune, P. and Feltz, C. (2008). Development of a decision support system for setting up a wind energy policy across the walloon region (southern belgium). *Renewable Energy*, 33(11):2416–2422.

Lydia, M., Kumar, S. S., Selvakumar, A. I., and Kumar, G. E. P. (2014). A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, 30:452–460.

Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, pages 343–360.

Manski, C. F. (2003). *Partial identification of probability distributions*. Springer Science & Business Media.

Nason, G. P., Von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):271–292.

OpenStreetMap contributors (2021). Planet dump retrieved from https://planet.osm.org . `https://www.openstreetmap.org`.

Papaefthymiou, G., Schavemaker, P., Van der Sluis, L., Kling, W., Kurowicka, D., and Cooke, R. (2006). Integration of stochastic generation in power systems. *International Journal of Electrical Power & Energy Systems*, 28(9):655–667.

Priestley, M. B. (1981). *Spectral analysis and time series: probability and mathematical statistics*. Number 04; QA280, P7.

Royal Belgian Institute for Natural Sciences, Directorate Natural Environment, G. (2020). Windmill locations in the belgian part of the north sea. Available at: `https://www.geo.be/catalog/details/d409cd5d-3a54-4722-b255-2b4d4374d8ae?l=en`.

Termonia, P., Van Schaeybroeck, B., De Cruz, L., De Troch, R., Caluwaerts, S., Giot, O., Hamdi, R., Vannitsem, S., Duchêne, F., Willems, P., et al. (2018). The cordex. be initiative as a foundation for climate services in belgium. *Climate Services*, 11:49–61.

Van Bellegem, S. and Von Sachs, R. (2004). On adaptive estimation for locally stationary wavelet processes and its applications. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(04):545–565.

Van Bellegem, S., Von Sachs, R., et al. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Annals of Statistics*, 36(4):1879–1924.

Von Sachs, R. and MacGibbon, B. (2000). Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scandinavian Journal of Statistics*, 27(3):475–499.

A. Mathematical Appendix

A.1. Identification Set Computation

Missing-at-random assumption (MAR)

It is straightforward from (7) and (8).

More productive installed network assumption

We know that $U_t \in [0,1]$ by definition. Using (10), we know that for each t,

$$0 \leq U_t \leq Y_t \tag{32}$$

We multiply each term by $\frac{C-C_t}{C}$, it still holds

$$0 \leq \frac{C-C_t}{C}U_t \leq \frac{C-C_t}{C}Y_t \tag{33}$$

If we add $\frac{C_t}{C} * Y_t$ to each term and take advantage of the definition of X_t given in (7):

$$\begin{aligned} \forall t, \frac{C_t}{C} * Y_t + 0 \leq \frac{C_t}{C} * Y_t + \frac{C-C_t}{C} * U_t \leq \frac{C_t}{C} * Y_t + \frac{C-C_t}{C} * Y_t \\ \frac{C_t}{C} * Y_t + \frac{C-C_t}{C} * 0 \leq X_t \leq \frac{C_t}{C} * Y_t + \frac{C-C_t}{C} * Y_t \\ \frac{C_t}{C} * Y_t \leq X_t \leq Y_t \end{aligned} \tag{34}$$

Hence

$$H_1[X_t] = [\frac{C_t}{C} * Y_t, Y_t] \tag{35}$$

More Productive expectation of the installed network assumption

It is the same computation as the last assumption, but working with the expectation operator.

More productive expected during peak periods assumption

First let's decompose the expectation of X_t on the peak-period parameter:

$$\mathbb{E}(X_t) = \mathbb{E}(X_t|t \in \delta)P(t \in \delta) + \mathbb{E}(X_t|t \notin \delta)P(t \notin \delta) \tag{36}$$

And we know, by (35) applied for the expectation conditionally on $t \in \delta$:

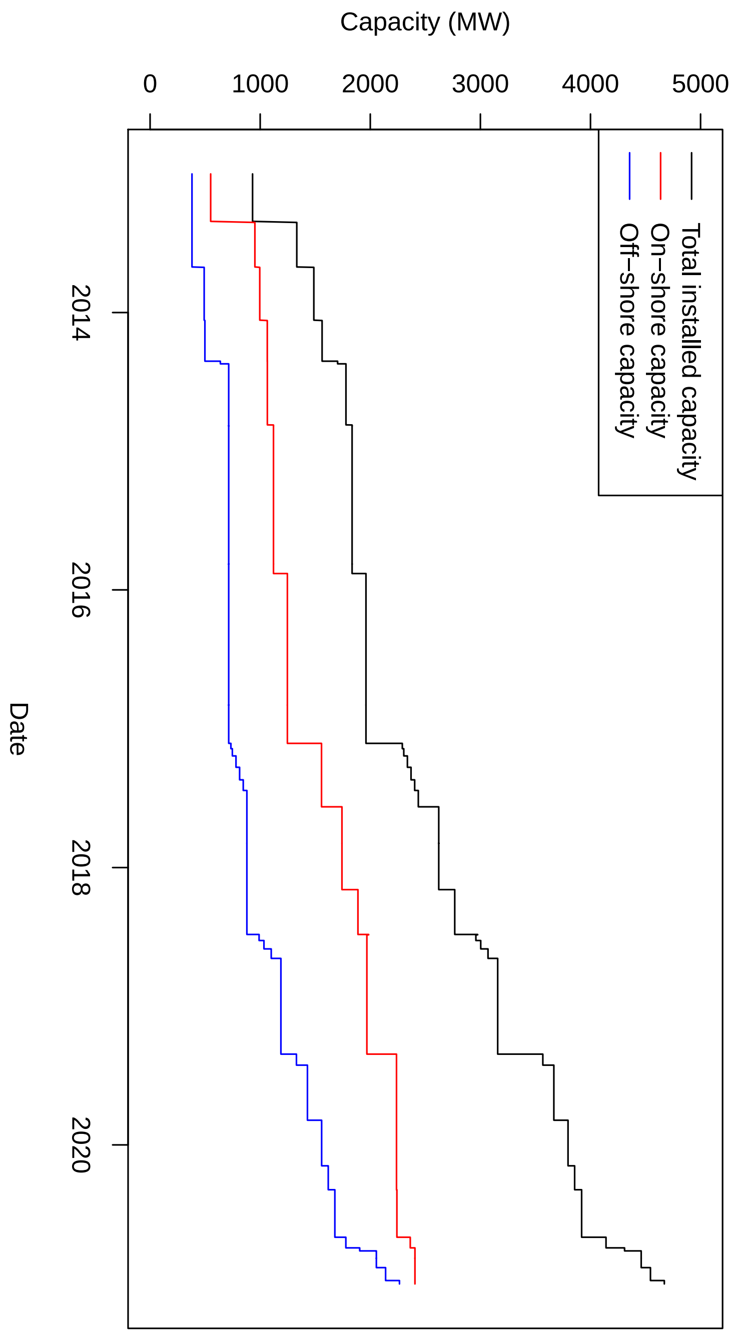
$$H_3[\mathbb{E}(X_t|t \in \delta)] = [\frac{\mathbb{E}(C_t * Y_t|t \in \delta)}{C}, \mathbb{E}(Y_t|t \in \delta)] \tag{37}$$

Hence, adding that $\mathbb{E}(X_t|t \notin \delta) \in [\frac{\mathbb{E}(C_t * Y_t|t \notin \delta)}{C}, 1]$ (during off peak period, the capacity ratio need to at least shows that we don't produce less on average), the identification region:

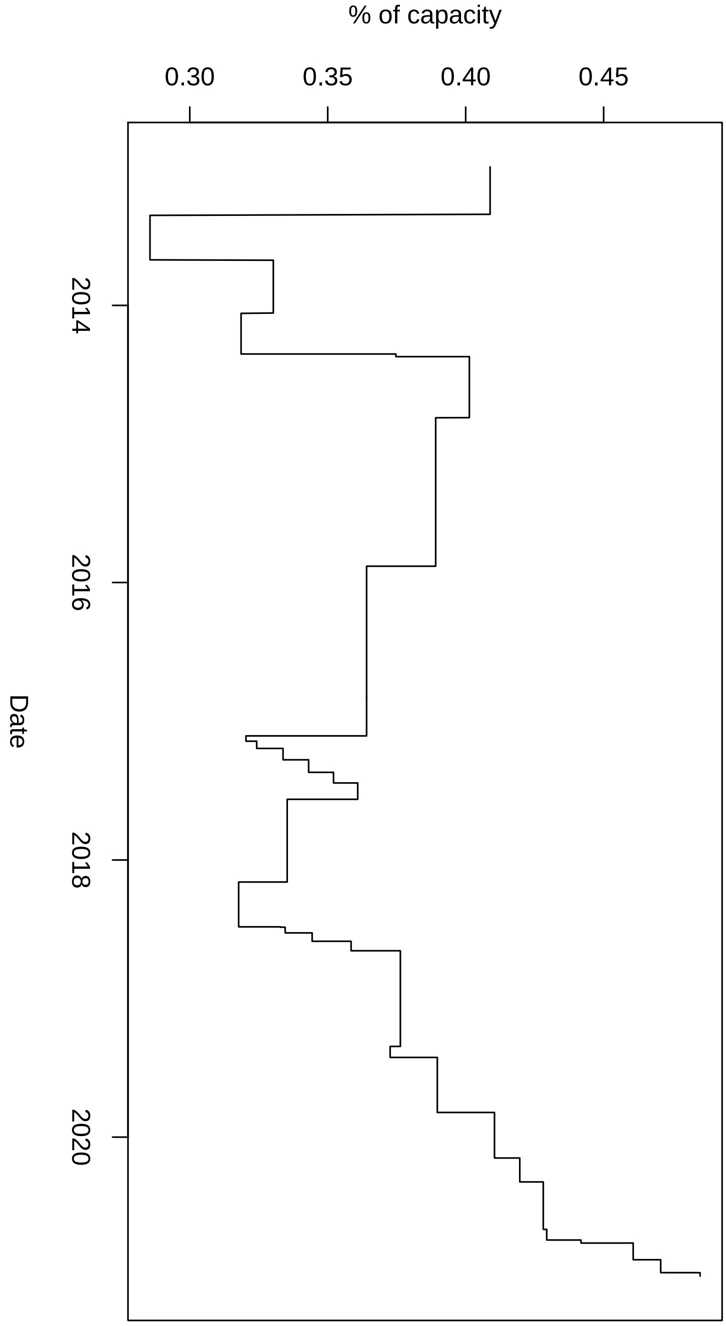
$$H_3[\mathbb{E}(X_t)] = [\frac{\mathbb{E}(C_t * Y_t|t \in \delta)}{C} P(t \in \delta) + \frac{\mathbb{E}(C_t * Y_t|t \notin \delta)}{C} P(t \notin \delta), \\ \mathbb{E}(Y_t|t \in \delta)P(t \in \delta) + P(t \notin \delta)] \quad (38)$$

B. Figures

B.1. Figure 3

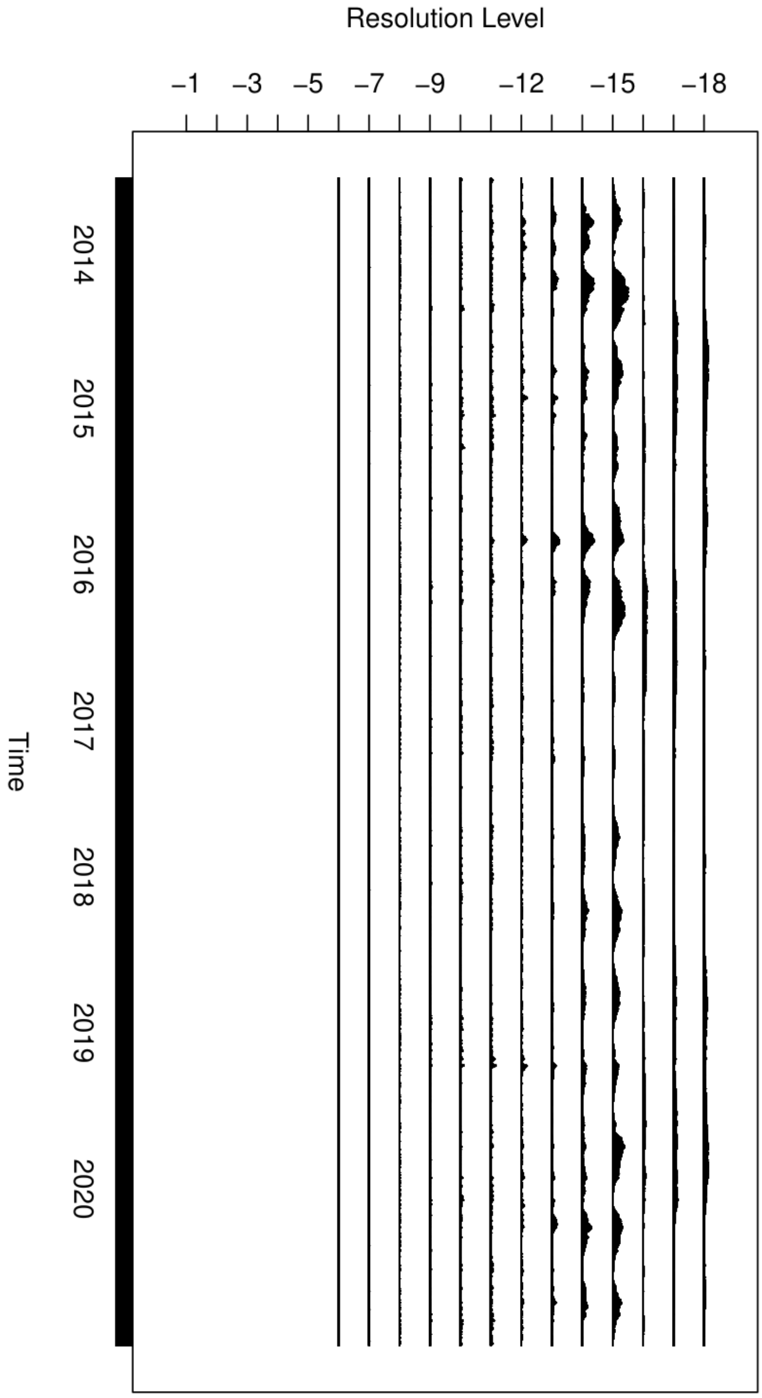


Capacity of each type over time

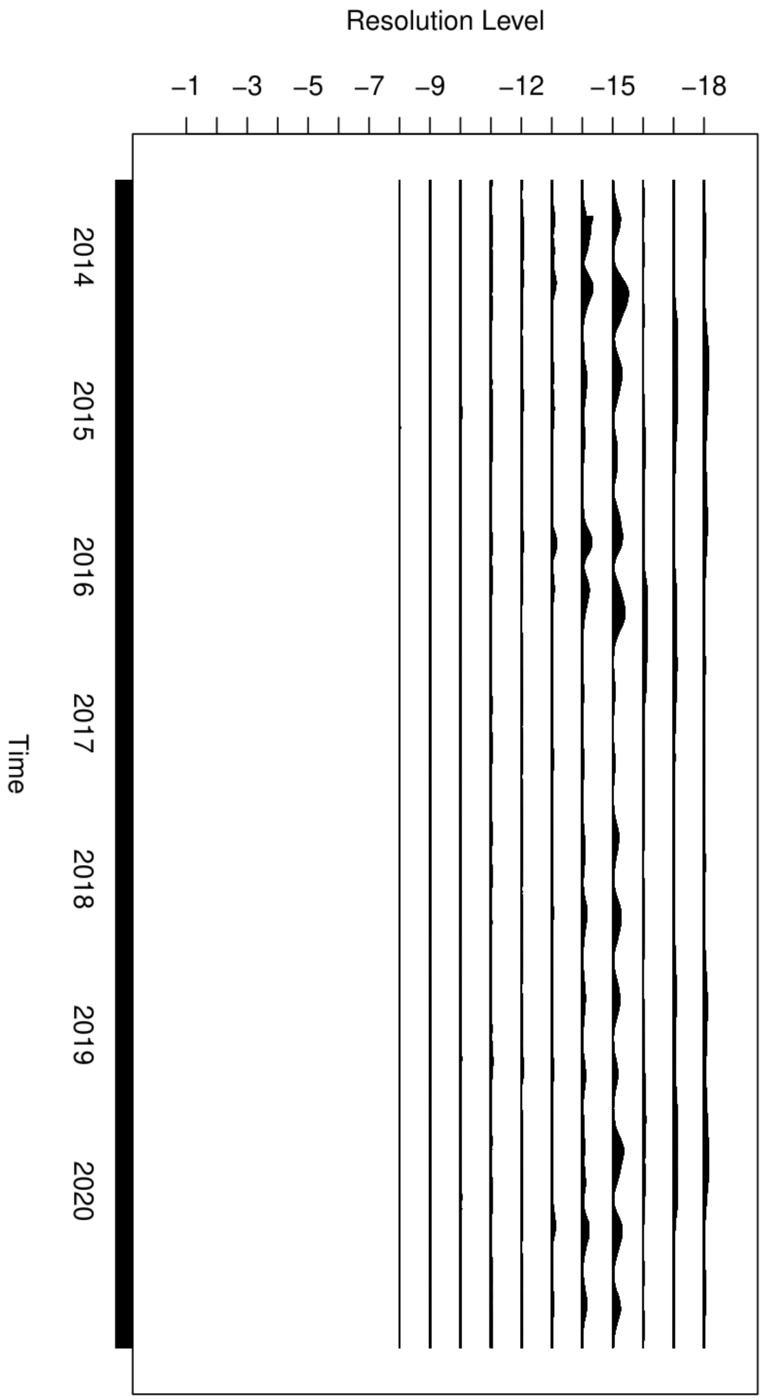


Ratio of off-shore in the capacity

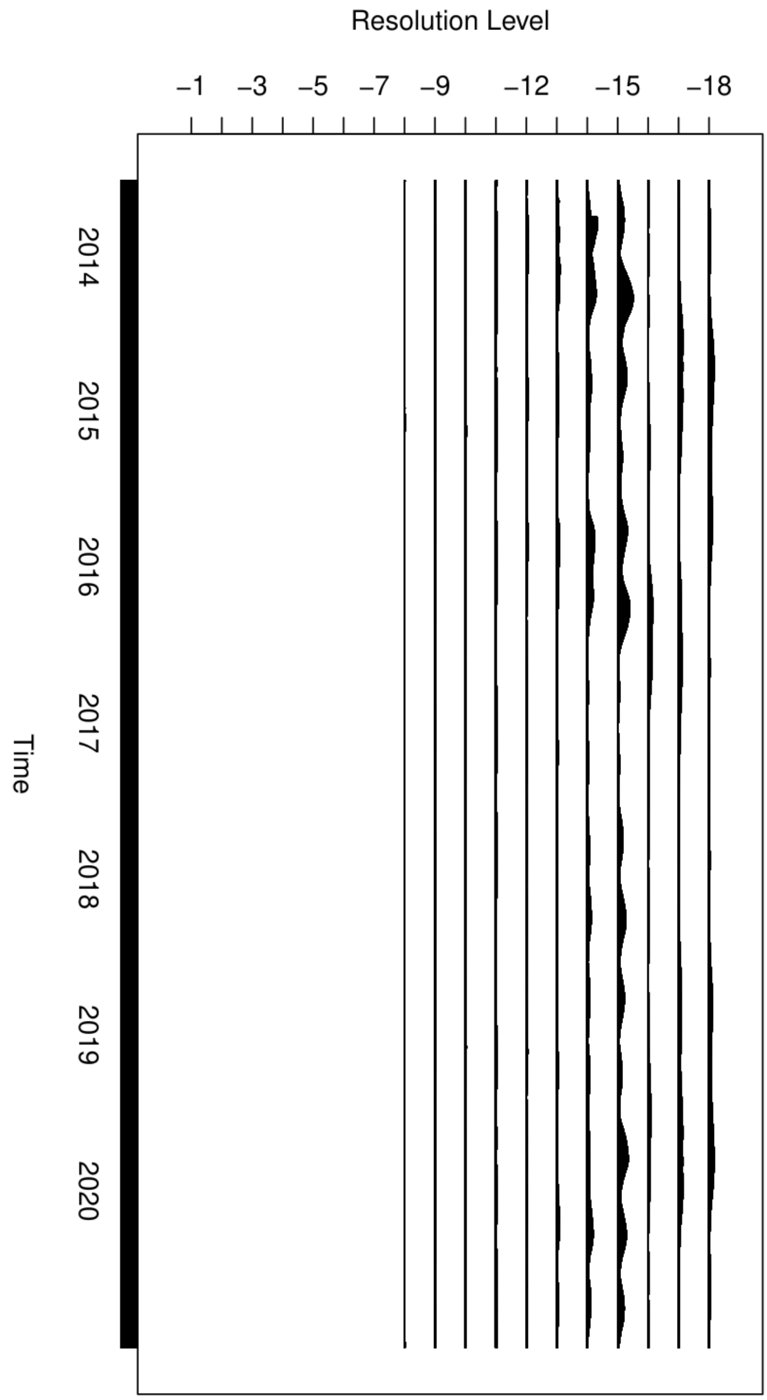
B.2. Figure 5



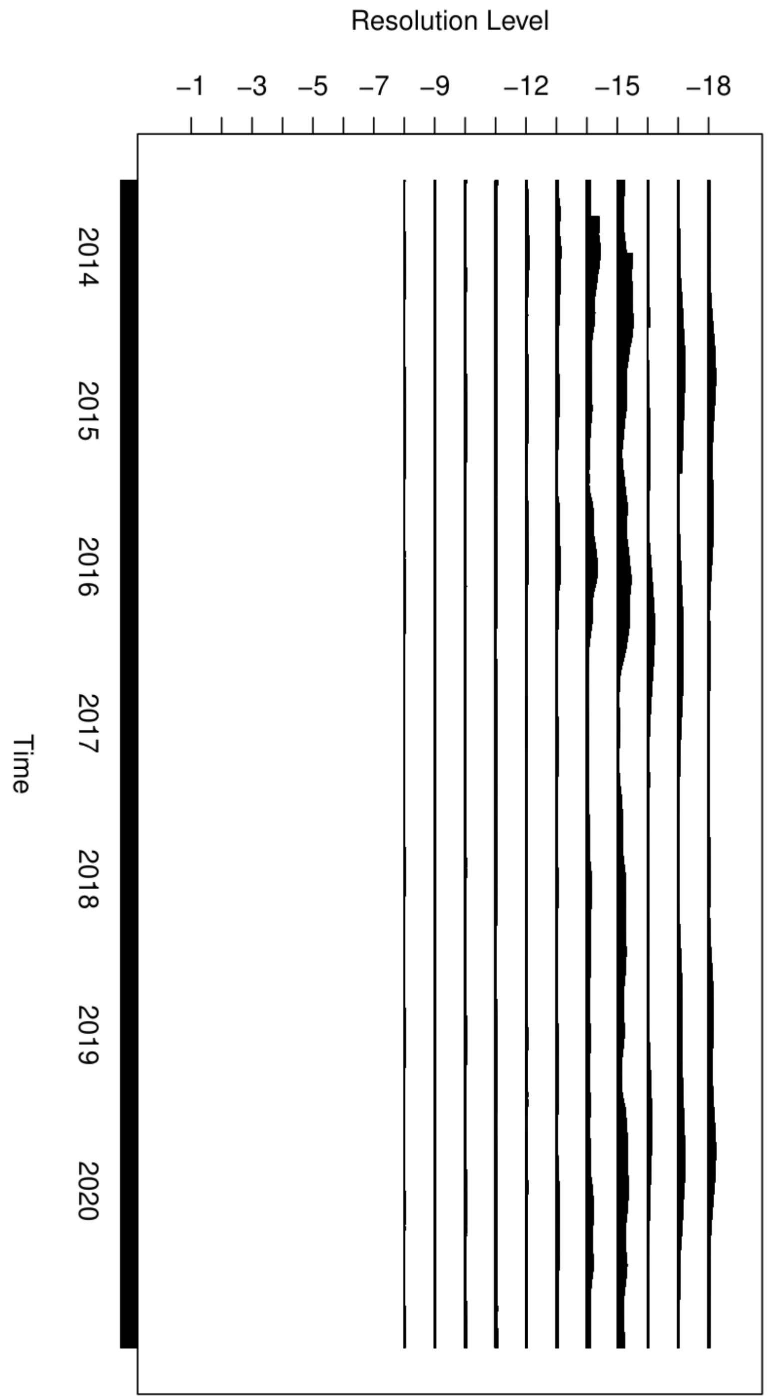
Unsmoothed estimated evolutionary wavelet spectrum of the capacity factor



Smoothed estimated evolutionary wavelet spectrum of the capacity factor, with a bin-width of 5000

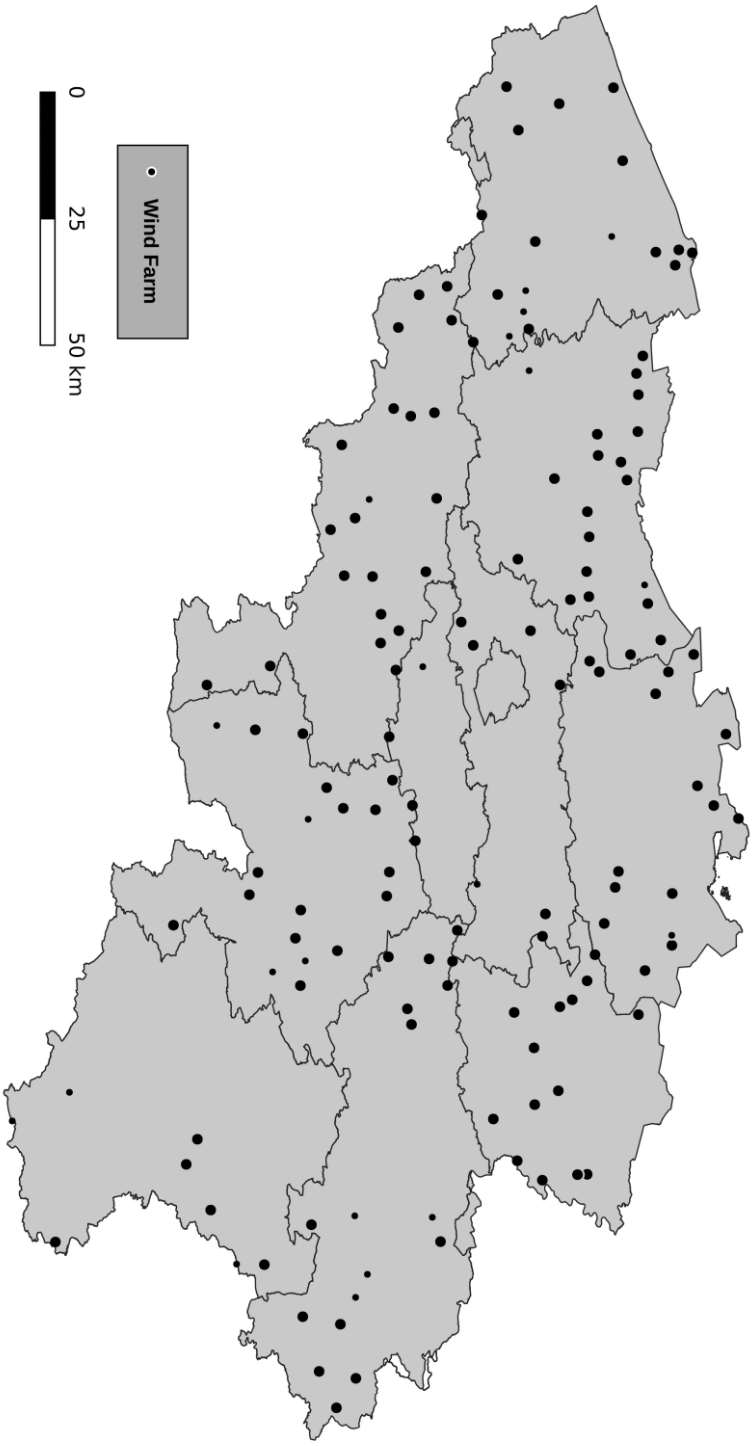


Smoothed estimated evolutionary wavelet spectrum of the capacity factor, with a bin-width of 10000



Smoothed estimated evolutionary wavelet spectrum of the capacity factor, with a bin-width of 20000

B.3. Figure 12



Wind Farms in Belgium (2021)