Draft version December 21, 2022
Typeset using LaTeX **twocolumn** style in AASTeX631

Lab Report n°2: Polish Power System Analysis

Berretti Michele and Fontana Aleksandar

Abstract

Energy is a key element for the development of any nation. During the last decades power consumption has increased exponentially all over the world, also increasing awareness over global warning and climate change. To make sure that everyone in the world has access to clean and safe energy, we need to understand the energy consumption patterns and be able to predict the increasing demand and its impact around the world, today, in the future and in the past.

In this report we will exploit the different well-established techniques for time series analysis in order to compare the different results and achieve a better understanding of their application. Moreover, we will train a simple variant of the well-known ARIMA model for time series forecasting and an artificial neural network to make an approximated prediction of the upcoming power consumptions.

Keywords: Energy — Power Load — Power Demand — Power Consumption — Renewable Energies — Fourier Transform — Discrete Wavelet Transform — Continuous Wavelet Transform — Empirical Mode Decomposition — ARIMA — Artificial Neural Networks

1. INTRODUCTION

Electric energy is essential in modern society, everything regarding economic and social activities such as healthcare, industrial production, access to education and quality of life requires a stable power supply. Considering only the last decade, with the advent of electric vehicles and growing request of electronic devices, we witnessed an enormous rise in power supply demand. If this trend continues by 2030 the demand will increase over 50% Suganthi & Samuel (2012). Moreover, with the raising awareness over global warming and climate change, the energy production is slowly transitioning from fossil fuels to renewable energy, such as Solar Power, Wind Power, Hydro Power and Nuclear Power (both fission and in the not so far future fusion as well).

The main concern with such energy sources is the reliability (Blaabjerg et al. (2012), Borges (2012), Moeini-Aghtaie et al. (2017), Adefarati & Bansal (2019), Ourahou et al. (2020)). Indeed, solar power can only be produced during the day, wind turbines cannot be nearly as effective in non-windy days, hydro power can only be exploited in certain areas and nuclear energy is highly contested in many countries among fears (mostly unjustified) for nuclear wastes and nuclear disasters (Ritchie et al. (2022)). Therefore, to be able to implement in a stable framework such energy sources, it is mandatory to perform an in depth study of the energy consumption patterns over time at different scales in order to ensure

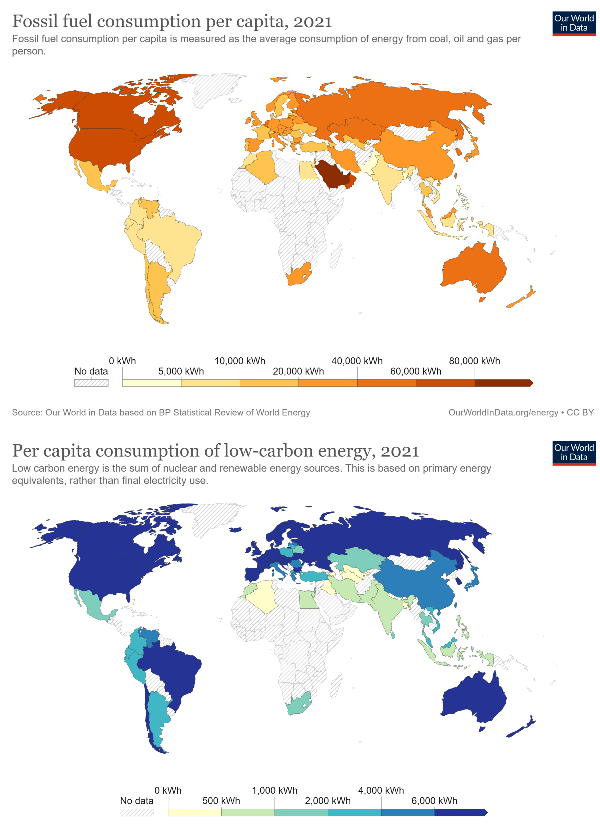

Figure 1. Credit: Ritchie et al. (2022)

a stable power supply, even during periods of global crysis where demand and production could rapidly change (e.g. Covid-19 Pandemic, Ukrainian War).

During this lab experience we studied the Polish Power System Load starting from 2008 up to 2017 with different Time-Series Analysis techniques well established in the field and presented the different results. Our main focus in this study was to detect the characteristic frequencies of the signal at different scales and study its features (e.g. stationarity).

Moreover, we trained a simple variant of the well-known time series forecasting algorithm ARIMA (AutoRegressive Integrated Moving Average, Hamilton (1994)) to predict the future loads of the system in absence of disrupting events that could radically change the consumption patterns. We then confronted the results of the ARIMA predictions against the validations set and the predictions carried out by an Artifical Neural Network we trained on the same training set of ARIMA.

2. DATASET

The dataset used in this lab experience has been downloaded by the PSE (Polskie Sieci Elektroenergetyczne, Polish Power Grid Company) through their website. Over the period that spans from January 1^{st} 2007 to December 31^{st} 2016 we obtained a single time series of the total load on the power grid with a sampling period of 15 minutes. We present a slice of the time series we are using in this lab experience in Fig. 2.

The raw data obtained by PSE is then organized in a Pandas Dataframe and resampled at various sampling frequencies with a moving average window of varying size depending on the resampled frequency in order to study the signal at different time scales.

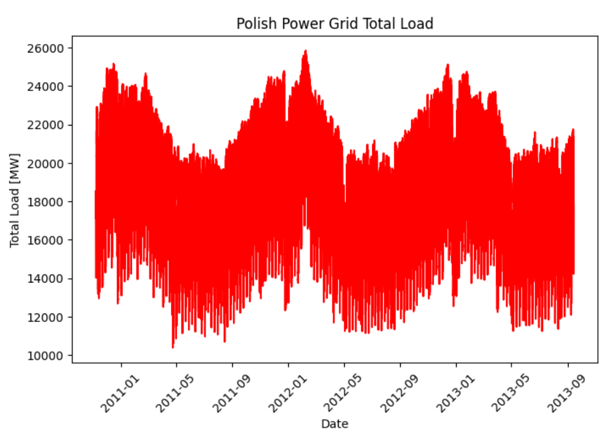

Figure 2. Sample of the dataset used in the lab experience.

3. METHODS

To carry out the intended analysis in this lab experience we applied different time series techniques well-established in the field.

3.1. *Fast Fourier Transform and Periodogram Estimation*

As a first approach to our dataset we decide to carry out the Fast Fourier Transform of the signal, represented in Fig. 3.

Anyways, it is well known that the periodogram obtained directly by the FFT (Fast Fourier Transform) of the time series is a biased estimator of the periodogram (Scargle (1982)). Therefore, in order to refine the estimation of the periodogram of the time series, we used the routine described in Chatfield (2003):

1. *Prewhitening*: A linear detrend is applied to the time series.

2. *Windowing*: To contain boundary effects due to the truncation of the signal a Hamming Window function is applied to the time series.

3. Finally, the FFT is computed.

Consequently, we estimated the periodogram of the time series using two different functions of the Python's **Scipy.Signal** library based on this routine we just described and compared the results.

3.2. *Continuous Wavelet Analysis and Scalogram Estimation*

The wavelet analysis overcome the main issue of the fourier transform as it can be used to analyze time series that contain non stationary power at many frequencies Daubechies (1990). Assuming we have a *Wavelet Function*, a function with zero mean and localized in both time and frequency spaces Torrence & Compo (1998), we can convolve the time series with different scaled and translated versions of such function to carry out the wavelet transform. This modified versions of the same wavelet function, called the *Mother Wavelet*, form a *Wavelet Basis* that can be either orthogonal or not. The first case would lead to a **Discrete Wavelet Transform** and the second one to a **Continuous Wavelet Transform** (Torrence & Compo (1998)). For the estimation of the scalogram we are interested only in the continuous wavelet transform and we will use the **Morlet Wavelet**.

On the other hand, wavelet analysis retains intact the temporal domain of the input signal and this could lead to peaks in the frequency spectrum caused by instantaneous peaks in intensity of the signal (due to noise) rather than an actual oscillatory phenomenon. So, we decided to exclude from the analysis oscillations with a period shorter than $\sqrt{2}$ the sampling period of the signal that could be caused by some kind of noise in the dataset. The Cone of Influence (COI) is the region of the wavelet spectrum where boundary effects become significant. For the Morlet Wavelet this is $\sqrt{2}$ the sampling period of the signal. The dimension of the COI at different scales provides a measure of the decorrelation time for a single peak in the time series. Hence, comparing the width of the peak in the power spectrum with this decorrelation time we can distinguish between peaks in the dataset and harmonic components (Torrence & Compo (1998)).

Finally, we can plot the power spectrum of the continuos wavelet transform, the *Scalogram*.

3.3. *Empirical Mode Decomposition (EMD)*

The Empirical Mode Decomposition is a method introduced by Huang et al. (1998) for analysing nonlinear and non-stationary data. The novelty of this method is its empirical base. This guarantees an highly adaptive, thus highly efficient, decomposition. Since such decomposition is based on the local characteristic time scale of the time series, it is applicable to nonlinear and non-stationary signals. As a consequence, any complicated dataset can be decomposed into a finite and often small number of *Intrinsic Mode Functions* (IMFs) that admit a well-behaved Hilbert Transform (Huang et al. (1998)). The result of such decomposition is presented in an energy-frequency-time distribution, the Hilbert Spectrum. The main advantage of this method is that it is no longer needed to perform the convolution of the signal with some kind of assumed harmonic base functions but instead the introduction of IMFs, functions based on local properties of the time series, allows to extract the actual instantaneous phase and frequency (Huang et al. (1998)).

3.4. *Time Series Forecasting*

As we previously stated, understanding the energy consumption pattern is vital for building an optimized and efficient power delivery infrastructure with the available energy sources. Accordingly, a key aspect of such analysis is providing a reliable forecast of future consumptions. Here we use a variant of the well-known ARIMA time series forecasting model, knwon as SARIMAX. This variant, as opposed to its base model, lets us implement seasonalities in the prediction, enabling much more reliable forecasts of periodic signals with a known period.

We then trained an Artificial Neural Network with eleven hidden layers, a REctified Linear Unit (RELU) activation function and three attention heads. The gradient optimization is performed with Adam.

The models are then trained, and will perform predictions, on two different datasets: one consinsting of the signal and one consisting of its Discrete Wavelet Coefficients. In both cases we performed a 70/30 split for training set and validation set. While the continuous wavelet is mostly used to construct the scalogram of the signal, the discrete wavelet transform (DWT) is a powerful tool when it comes to time series prediction. Once the DWT coefficients have been computed, we can use them to train a forecasting model and then use it to predict the coefficients. Through the Inverse DWT we can then retrieve the prediction of the reconstructed signal. This should result in a more robust prediction as we are predicting the single harmonics of the signal and the isolated noise rather than the whole noise and harmonics ensemble.

4. RESULTS & DISCUSSIONS

We now present the results obtained with the different techniques presented in the previous section.

4.1. *Fourier Transform & Periodogram*

We first present the power spectrum simply obtained by computing the FFT of the signal and plotting the results.

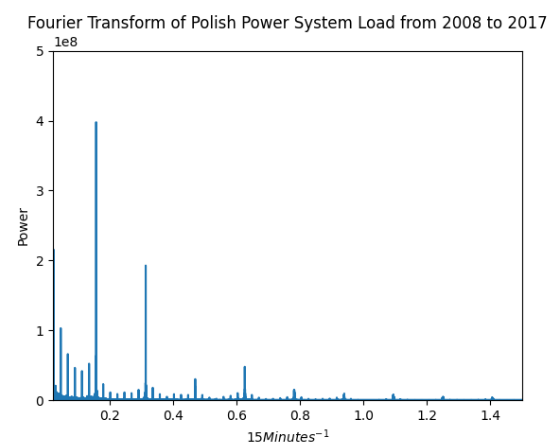

Figure 3. Fourier Transform of the signal at a sampling frequency of 15 minutes.

In Fig. 3 we can clearly detect 4 peaks, corresponding to the daily period and its sub-harmonics.

As previously anticipated, we now estimate the periodogram of the time series using a more refined technique presented in Chatfield (2003) and implemtend in python in two different functions: **Welch** and **Periodogram**.

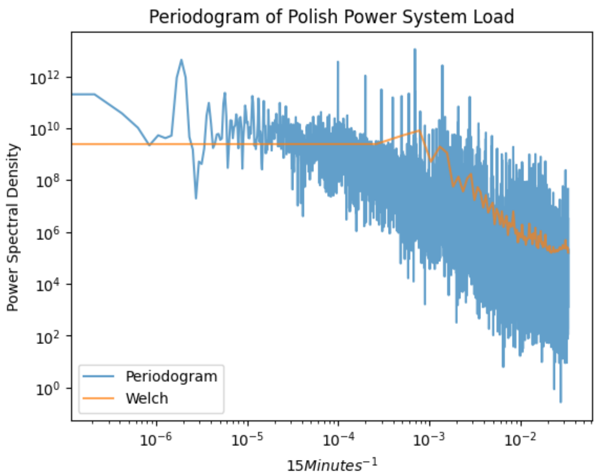

Figure 4. Periodogram obtained with both the Welch function and the Periodogram function in Python.

As we can clearly see in Fig. 4, both functions provide a poor representation of the characteristic frequencies of the signal. This is due to mainly two issues: (1) the signal used is too noisy and (2) the signal is non-stationary. The first issue could be solved by using a resampled version of the signal with the noise smoothed out. The second issue is the point of failure of the Fourier Transform and the very reason why we will study the signal with the continuous wavelet transform and the scalogram.

4.2. *Continuous Wavelet Transform & Scalogram*

As we have already noted in the previous sub-section, the signal we are analysing is non-stationary and therefore the Fourier Transform cannot be used to detect the frequencies of the time series. Consequently, we resort to the continuous wavelet analysis and we present the scalogram of the dataset in Fig. 5.

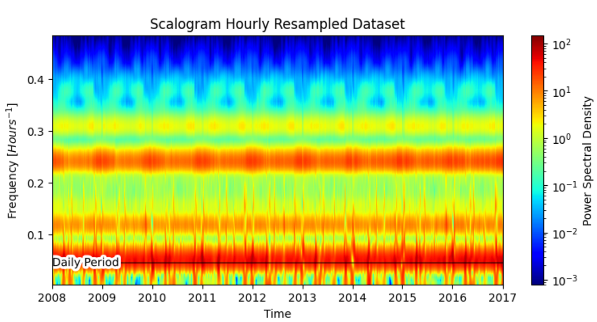

Figure 5. Scalogram of the time series. The four peaks in the power spectra corresponds to the ones we detected with the Fourier Transform. Zoomed image in appendix.

The scalogram makes it obvious to see that the signal is actually non-stationary. Indeed, the peaks in frequency have a slight, but obvious, variation in intensity over time.

We can see how, just like in the Fourier Transform Spectrum, we detected the daily period and its subharmonics.

4.3. *Empirical Mode Decomposition*

We can now proceed with the empirical mode decomposition of the signal in order to retrieve the IMFs and the trend of the signal. The latter will be particularly relevant to us as in order to perform the forecasting of the time series we will need to detrend the training set.

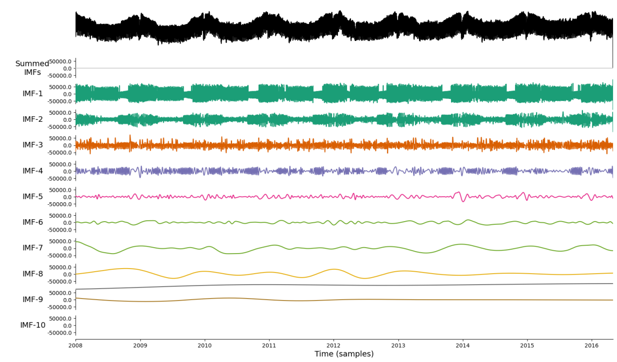

Figure 6. The original signal, its IMFs and its trend obtained through Empirical Mode Decomposition. Zoomed image in the appendix.

Once we obtained the trend and the IMFs of the signal we can plot the Hilbert spectrum of the time series in Fig. 7.

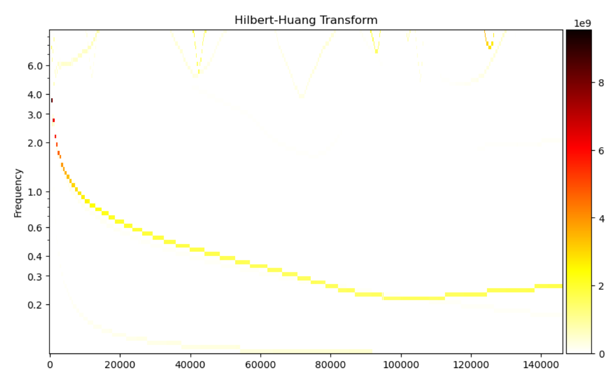

Figure 7. Hilbert Spectrum

4.4. *Time Series Forecasting*

Finally, once the patterns of our time series have been thoroughly analysed, we can proceed with the forecasting of our signal.

As previously mentioned, we made forecast with two different models and two different approaches for each.
The results on the validation set are here presented:

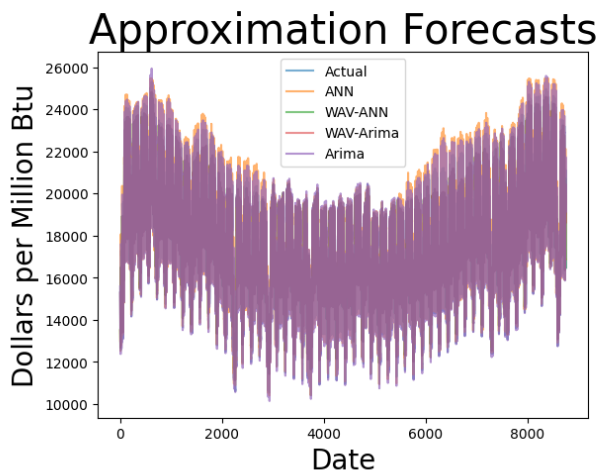

Figure 8. Forecasts of Polish Power System Total Load with ARIMA, ARIMA+Wavelets, Artificial Neural Network (ANN) and ANN+Wavelet.

The above plot can only be considered as a qualitative study of the accuracy of the predictions. Indeed, all the different models seems to more or less reproduce a faithful version of the validation set.
In order to better visualise the performance of the different models here we plot the differences between the original values of the validation set and the predicted values for the two different models.

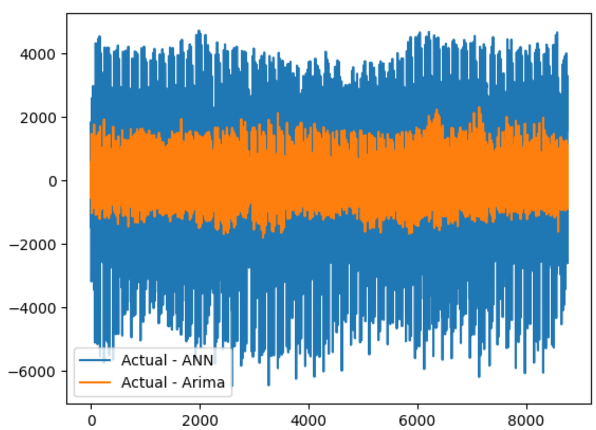

Figure 9. Differences between the original values of the validation set and the predicted values for the different models. ARIMA and ANN (top), ARIMA+WAV and ANN+WAV (bottom).

Surprisingly enough, the ARIMA model seemed to provide the best forecasts and as expected, the prediction on the wavelet coefficients notably improved the quality of the predictions as we can see in the figure below.

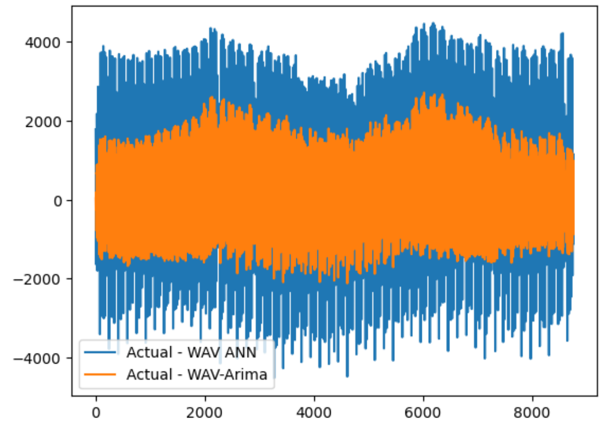

Figure 10. Difference between predictions on the validation set with ARIMA+WAV and ANN+WAV

Moreover, we noted that predicting the wavelet coefficients and the reconstructing the signal removed the bias towards lower values that we would not have spotted if it wasn't for this kind of analysis in the Artificial Neural Network prediction as we can see in Fig. 11

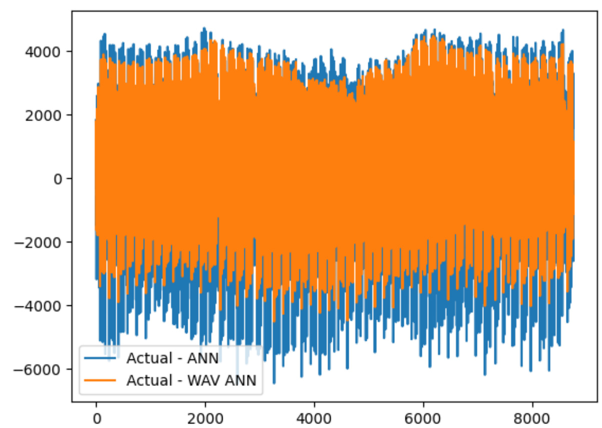

Figure 11. Difference between ANN prediction and ANN+WAV prediction.

5. CONCLUSION

We have carried out a detailed analysis of the Polish Power System Total Load in order to test the different most exploited time series techniques in the field. Indeed we had the opportunity to visualise the advantages of each method and the different results and applications.
We then performed a couple of very simple prediction models in order to test some of the basic forecasting techniques and the various pre/post-processing steps

needed to provide a valid training set and qualitatively compare the results between different forecasts.

REFERENCES

Adefarati, T., & Bansal, R. 2019, Applied Energy, 236, 1089, doi: https://doi.org/10.1016/j.apenergy.2018.12.050

Blaabjerg, F., Ma, K., & Zhou, D. 2012, in 2012 IEEE International Symposium on Industrial Electronics, 19–30, doi: 10.1109/ISIE.2012.6237053

Borges, C. L. T. 2012, Renewable and Sustainable Energy Reviews, 16, 4008, doi: https://doi.org/10.1016/j.rser.2012.03.055

Chatfield, C. 2003, The Analysis Of Time Series - An Introduction, 6th edn., Chapman Hall/CRC Texts in Statistical Science (Chapman and Hall/CRC)

Daubechies, I. 1990, IEEE Transactions on Information Theory, 36, 961, doi: 10.1109/18.57199

Hamilton, J. D. 1994, Time series analysis, 1st edn. (Princeton University Press). libgen.li/file.php?md5=46573ee323bfcecad694f61564ca8e1c

Huang, N. E., Shen, Z., Long, S. R., et al. 1998, Proceedings of the Royal Society of London Series A, 454, 903, doi: 10.1098/rspa.1998.0193

Moeini-Aghtaie, M., Farzin, H., Fotuhi-Firuzabad, M., & Amrollahi, R. 2017, IEEE Transactions on Power Systems, 32, 368, doi: 10.1109/TPWRS.2016.2549747

Ourahou, M., Ayrir, W., EL Hassouni, B., & Haddi, A. 2020, Mathematics and Computers in Simulation, 167, 19, doi: https://doi.org/10.1016/j.matcom.2018.11.009

Ritchie, H., Roser, M., & Rosado, P. 2022, Our World in Data. https://ourworldindata.org/nuclear-energy

Scargle, J. D. 1982, ApJ, 263, 835, doi: 10.1086/160554

Suganthi, L., & Samuel, A. A. 2012, Renewable and Sustainable Energy Reviews, 16, 1223, doi: https://doi.org/10.1016/j.rser.2011.08.014

Torrence, C., & Compo, G. P. 1998, Bulletin of the American Meteorological Society, 79, 61, doi: 10.1175/1520-0477(1998)079⟨0061:APGTWA⟩2.0.CO;2