Time Series Analysis and Forecasts with Holt-Winters-Method

Hermann Kraus, Gaby Seifert
Institute of Electrical Energy Systems
Friedrich-Alexander-University Erlangen-Nuremberg (FAU)
Erlangen, Germany
kraushermann@gmx.net, gaby.seifert@fau.de

Abstract

Within the scope of this work the public available measured data of the German TSO's are analyzed by stochastic time series methods. The main focus is on the following data types of the period 2010-2015: load, cross-border flows, infeed of photovoltaics and infeed of wind energy. The available data however are different in their format and textually structure. After the completion of data preparation and plausibility checks, the analysis of time series can be performed by disassembling the time series in their components: trend, season and random process. This procedure is used for all single year measurement series and measurement series for the whole period. Based on this procedure the forecast of the of time series in 2030 were made. Finally, the predicted time series are checked and compared with other studies.

This methodology offers an interesting way to evaluate data and generate forecasts for several years to support the system design in electrical energy systems.

I. Introduction

Within the change of the German power supply system and the growing number of decentralized power plants is the analysis and evaluation of information about the power network more importantly for future grid planning. The time series analysis is a good approach to get trends for the consume and power supply development and do forecasts for prospective years. Hereby measured data of the power system are used to extract the information with the mathematical methods of the time series theory, which are also used in other scientific fields.

The purpose of this work is to show a way to apply this time series analysis in the German electrical energy system based on measured data of the German transmission system operators.

II. Theoretical Background

With [1] and [2], time series are defined as a sequence of values $y_1, y_2, \dots, y_N$ with the set T = {1, 2, …, N} including the dates.

This empirical form has to be described by a theoretical model to extract all information the series contains. For this purpose, a time series can be considered as one possible realization of a stochastic process, which is defined as a sequence of random variables $Y_1, Y_2, \dots, Y_N$.

With the three types of the classic model of components (eq. (1.1-1.3)) the stochastic process is divided into trend T_t, season S_t and random process ε_t, which is represented by a white-noise-process [3].

As in [3] described, equation (1.1) shows the additive model, where the components are linked with a summation. This model fits for time series, which present a trend and a continuous season, for example **Fig. 1**.

$$Y_t = T_t + S_t + \varepsilon_t \qquad (1.1)$$

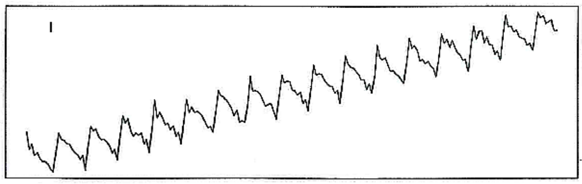

Fig. 1 Example for a time series, where the additive model can be used

The quasi multiplicative model (eq. (1.2)) describes a time series, which shows a trend and a season, that gets major peaks with increasing level. **Fig. 2** illustrates such a measurement series.

$$Y_t = T_t \cdot S_t + \varepsilon_t \qquad (1.2)$$

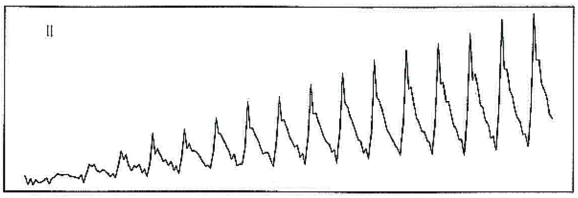

Fig. 2 Example for a time series, where the quasi multiplicative model can be used

The third type is the multiplicative model. The difference between the quasi multiplicative model is that the random process ε_t in Equation (1.3) is also multiplied. That causes a bigger dispersion in the measured series, while the peaks are growing up (see **Fig. 3**).

$$Y_t = T_t \cdot S_t \cdot \varepsilon_t \qquad (1.3)$$

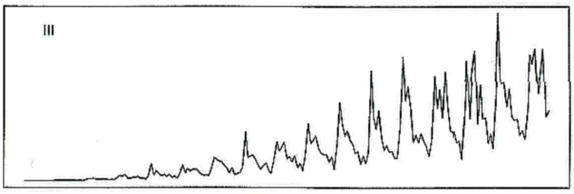

Fig. 3 Example for a time series, where the multiplicative model can be used

III. Data Preparation and Plausibility Checks

With §17 StromNZV the German TSO's are obligated to publish determined measured data on their web-sites [4]. For further investigations, the main focus was set on the four available data types within the period 2010-2015: load, cross-border flows, infeed of photovoltaics and infeed of wind energy. In the wind data of TenneT there is an additional determination in wind onshore and wind offshore. These data sets have some kind of differences in format and textually structure for each TSO. First aspect is that not every data type is published for the period 2010-2015. Second aspect is that 50Hertz and Amprion are providing their data in yearly measurement series, TenneT and TransnetBW in monthly measurement series. Further there are also differences between TSO's data in considering the time shift from summer time to winter time and the other way around. Finally, each TSO documents the cross-border flows in different ways [4-7].

All these arguments show the need for data preparation and plausibility checks. So, all data are rewritten in form of yearly time series: the time shifts are considered in each data file and the cross-border flows for each TSO are summed to a total flow.

In later implementations, it is also necessary to determine substitute values for data gaps and outliers by interpolating them with surrounding values. Especially the measurement series of the load had many data gaps. **Fig. 4** shows Amprion Load 2010 with the original data set and **Fig. 5** the time series after replacement value creation.

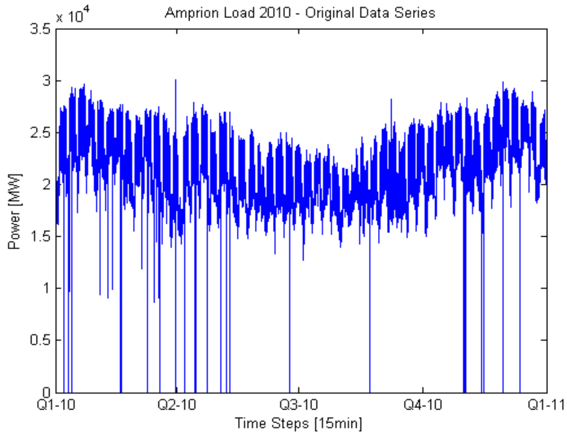

Fig. 4 Amprion Load 2010 with original data set

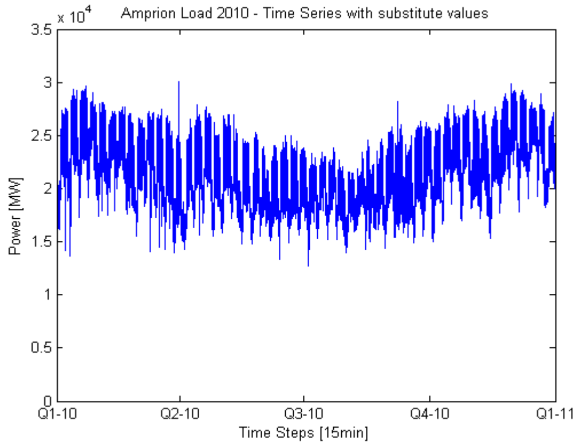

Fig. 5 Amprion Load 2010 with substitute values

To predict the time series for later years it's necessary to build time series including values of the whole period 2010-2015. With these composed measurement series it is then possible to see long-standing trends. **Fig. 6** shows the distribution functions per year of Amprion load for the period 2010-2015. It's visible, that the Amprion load curves shift from 2010 to 2015 to lower values.

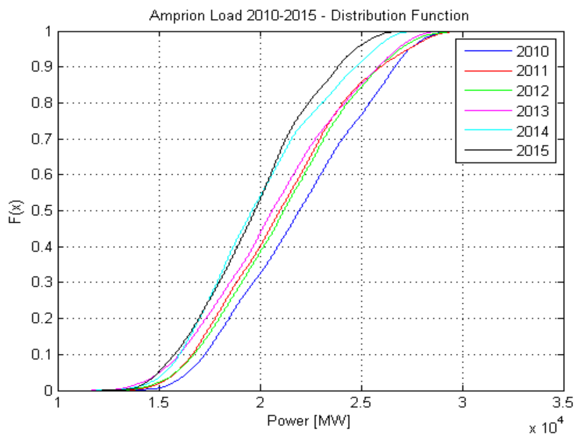

Fig. 6 Distribution functions of Amprion load curve for the period 2010-2015

IV. Time Series Analysis

In the next step the analysis of the time series can start. At first the trend component is determined with simple moving averages. In this method, the arithmetic mean of sections with equal lengths d is calculated and the temporal center of sections $s + \frac{(d+1)}{2}$ is referred, like in [3] shown.

$$y_{\left(s+\frac{(d+1)}{2}\right)} = \frac{1}{d}\sum_{t=s+1}^{s+d} y_t \qquad (3.1)$$

Second the trend component is subtracted from the component model and the season component is determined in followed steps:

1. All dates, that are repeating after one season, are collected in one line of a matrix.
2. For each line the arithmetic mean value is calculated.
3. The average values are combined to a new time series and are centered to zero.

After subtracting the season component from the remaining time series, the random process remains.

In **Fig. 7** the disassembly of a time series for the Amprion load curve in the period 2010-2015 is shown. In the top left corner the original time series is represented, in the top right corner the trend component, in the bottom left corner the season component and on the bottom right corner the remaining random process.

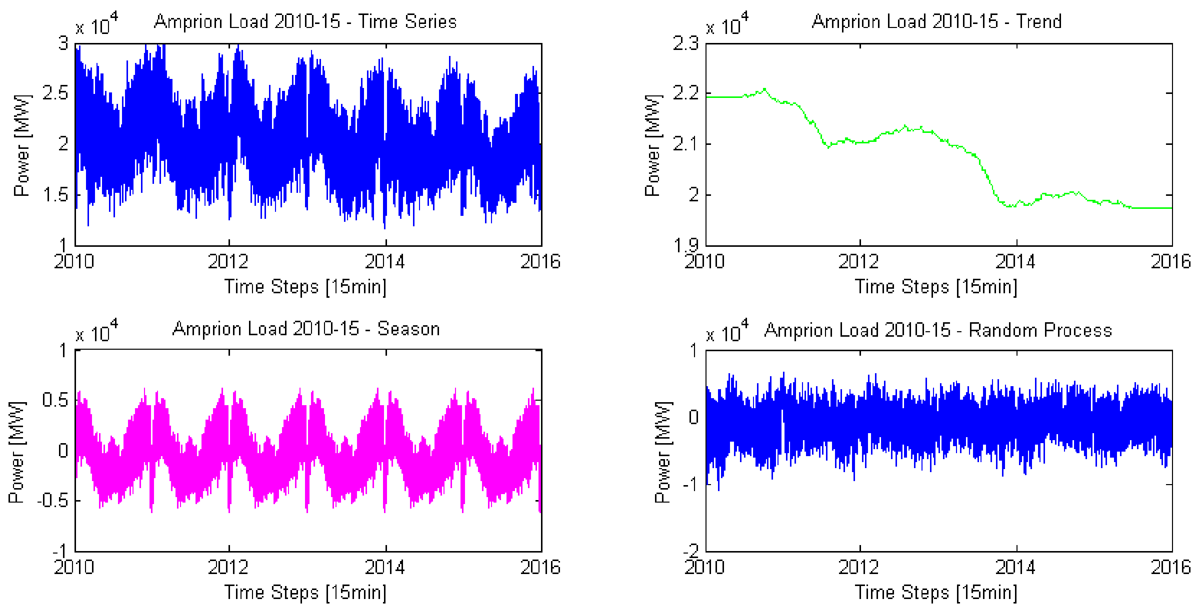

Fig. 7 Disassembly of Amprion load curve for the period 2010-2015

V. Forecasts for the year 2030

In the last step time series are predicted for the year 2030. The forecast is carried out with the Holt-Winters-Algorithm in additive form (eq. (4.1a)-(4.1d)) or in multiplicative form (eq. (4.2a)-(4.2d)), which are presented in [1]. In this work the additive Holt-Winters-Method was used for the load, however for the infeed of photovoltaics and wind energy the multiplicative method was utilized.

$$L_t = (1-\alpha)(L_{t-1} + b_{t-1}) + \alpha(y_t - S_{t-s}) \qquad (4.1a)$$

$$b_t = (1-\beta)b_{t-1} + \beta(L_t - L_{t-1}) \qquad (4.1b)$$

$$S_t = (1-\gamma)S_{t-s} + \gamma(y_t - L_t) \qquad (4.1c)$$

$$\hat{y}_{t,h} = L_t + hb_t + S_{t+h-s} \qquad (4.1d)$$

$$L_t = (1-\alpha)(L_{t-1} + b_{t-1}) + \alpha\left(\frac{y_t}{S_{t-s}}\right) \qquad (4.2a)$$

$$b_t = (1-\beta)b_{t-1} + \beta(L_t - L_{t-1}) \qquad (4.2b)$$

$$S_t = (1-\gamma)S_{t-s} + \gamma\left(\frac{y_t}{L_t}\right) \qquad (4.2c)$$

$$\hat{y}_{t,h} = (L_t + hb_t) \cdot S_{t+h-s} \qquad (4.2d)$$

L_t represents the Level, b_t the slope and S_t the seasonal component of the measurement series. The forecasted series is calculated by the summary of the three values with h-steps in the future. With the parameters α, β and γ the Holt-Winters-Method can be optimized by minimizing the difference between predicted series and original series. For this procedure it is possible to make forecasts within the period 2010-2015 and compare it with the measured data.

To estimate initial values for the parameters, the structure of the time series is essential. If the time

series has a low trend, then β should be set with a low value. Has the time series seasonal fluctuations, which are each year very similar, the γ parameter should also start with a small value. α has to be chosen small, if past values should be included to calculate the level of the time series in the forecast.
Within the scope of this work a prediction for time series of the year 2015 was done from time series of the year 2011 and the forecasted data were compared with the original data from the year 2015.
The initial values of level, trend and season are given with followed equations (4.3a-4.3d), as in [3] described:

$$L_t = \frac{1}{s}\sum_{t=1}^{s} y_t \qquad (4.3a)$$

$$b_t = 0 \qquad (4.3b)$$

$$S_t = y_t - L_t \quad \text{(additive)} \qquad (4.3c)$$

$$S_t = \frac{y_t}{L_t} \quad \text{(multiplicative)} \qquad (4.3d)$$

If you want to do forecasts with special goals in the future, e.g. 15 GW Wind Offshore in 2030 set from the government, is it possible to use another form of the Holt-Winters-Algorithm than the additive and multiplicative form. For orientation are the points of current and future power helpful to describe a special function, which describes the development better than the additive or multiplicative form. A way to find out the suitable function is for example the regression analysis.

Fig. 8 and **Fig. 10** show the time series of PV/Wind infeed from whole Germany for the year 2015, **Fig. 9** and **Fig. 11** their forecasts for the year 2030. By comparing the measured series from the year 2015 with the predicted series for the year 2030, it is clearly visible, that the infeed of PV/Wind and the difference between summer and winter time is much greater.

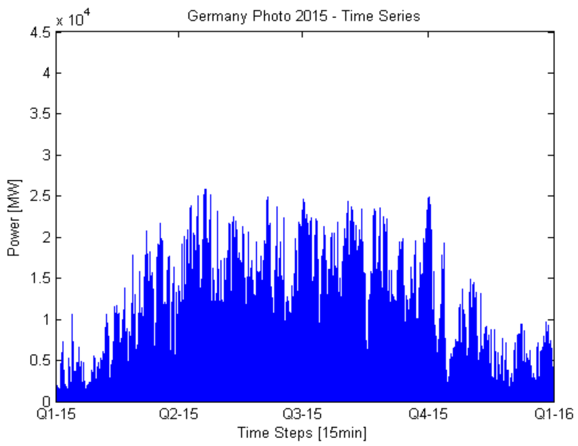

Fig. 8 Time series of Germany PV for the year 2015

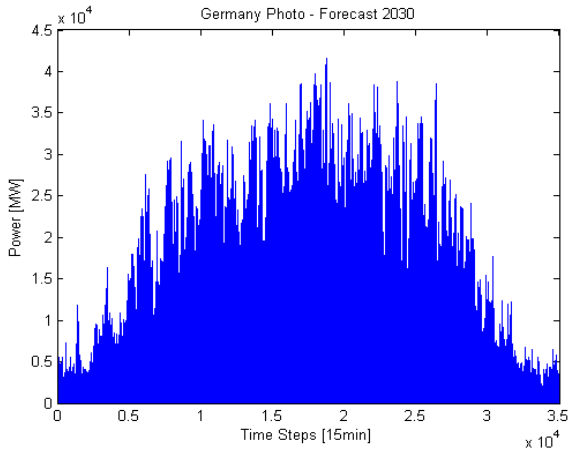

Fig. 9 Forecast of Germany PV for the year 2030 (35040 Time Steps)

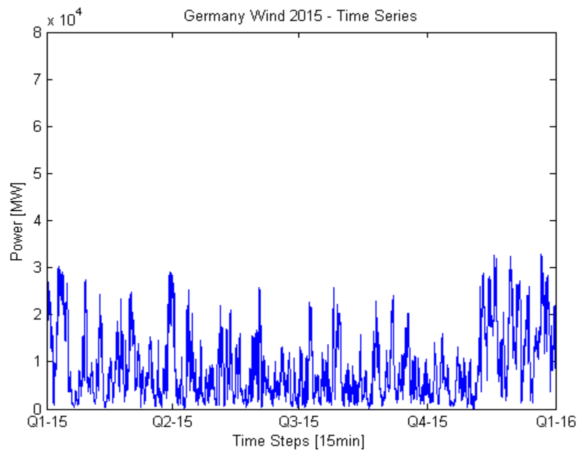

Fig. 10 Time series of Germany Wind for the year 2015

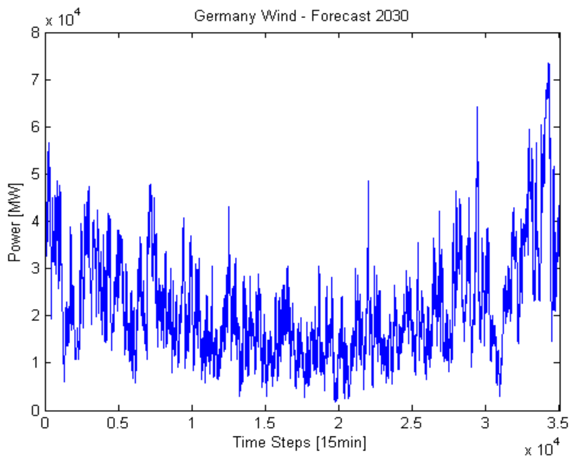

Fig. 11 Forecast of Germany Wind for the year 2030 (35040 Time Steps)

VI. Plausibility Checks of the forecasted Time Series

To validate the predicted time series for the year 2030, a comparison with other studies and forecasts for this year is performed. For best plausibility checks the Scenario B of the network development plan (NEP) of the German TSO's is considered.
The values of NEP for installed photovoltaic (56.3 GW) and wind (88.8 GW) power are in the same range as the installed photovoltaic (59.4 GW) and wind (91.9 GW) capacity of the predicted time series [8]. The values of the time series forecast are given as the infeed of PV and wind, so they have to be converted in installed capacity to compare it with the NEP values. The converting factor for PV is 0,7 and for wind 0,8, which were determined with the ratio of maximum infeed and installed capacity from the asset master data of the period 2010-2015 [9]. To calculate the installed capacity, the infeed values have to be divided with the converting factor.

In **Fig. 12** the time series of the residual load for Germany in the year 2030 is illustrated. This load curve results by subtracting the wind and PV feed from the load ($Load - Wind - PV$). **Fig. 13** shows the distribution function of the residual load. It demonstrates, that the consume of Germany is covered for $0.13 \cdot 8760\text{h} = 1140\text{h}$ of the year 2030 only by wind and PV Power Plants.

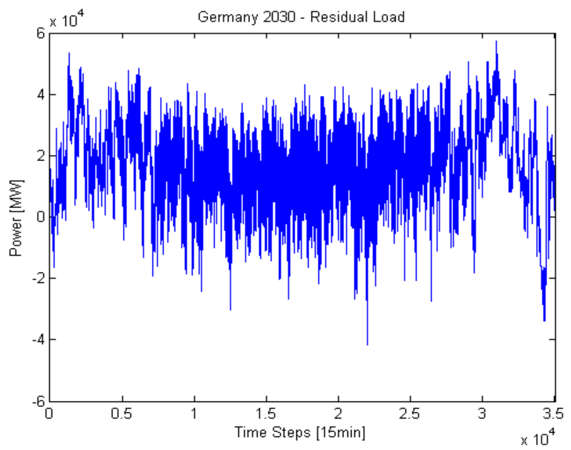

Fig. 12 Residual load curve for Germany in the year 2030 (35040 Time Steps)

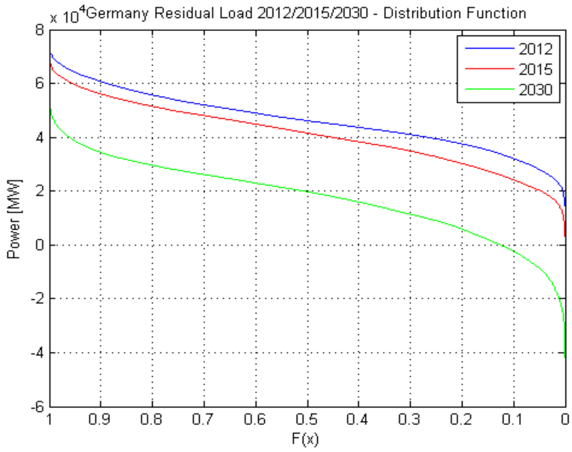

Fig. 13 Residual load curve for Germany in the year 2030 – Distribution function

VII. Conclusion

Due to the application of time series analysis methods in other scientific fields there is a solid basis to integrate it in the electric engineering. This statement can be confirmed with the results of this work. The used data was the public available data of the German TSO's. Four types for the period 2010-2015 were extracted from their homepages: load, cross-border flows, infeed of photovoltaics and infeed of wind energy. Due to the differences the data is provided, there were several data preparations and plausibility checks required. After that the analysis of the time series could be done. With the disassembly in their components the trend and the season of the data type easily could be determined. Based on the Holt-Winters-Algorithm the forecasts for the year 2030 were possible. By comparing the forecasts and the residual load of the year 2030 with the NEP study, it was possible to verify the predicted time series.
The comparison showed that the forecasted data was in the range of the results from other studies, but instead of individual data like maximum of installed capacity or load, the time series analysis gives a series for a whole period back. So, in future this method offers an interesting way to evaluate data and generate forecasts for several years to support the system design in electrical energy systems.

VIII. References

[1] Stier, Winfried: Methods from Time Series Analysis (German title: Methoden der Zeitreihenanalyse). Springer publishing house Berlin Heidelberg New York, 2001. – ISBN 978–3–540–41700–1

[2] Schlittgen, Rainer; Streitberg, Bernd H. J.: Time Series Analysis (German Title: Zeitreihenanalyse). 9., slightly revised version. Oldenbourg scientific publishing house 2001. – ISBN 978–3–486–25725–0

[3] Schlittgen, Rainer: Applied Time Series Analysis with R (German title: Angewandte Zeitreihenanalyse mit R). 3., updated and expanded version. De Gruyter Oldenbourg publishing house, 2015. – ISBN 978–3–110–41398–4

[4] Federal ministry of justice and consumer protection in coorporation with juris GmbH: Regulation about the connection to the electricity supply system (Electricity Network Access Regulation) (German title: Verordnung über den Zugang zu Elektrizitätsversorgungsnetzen (Stromnetzzugangsverordnung - StromNZV)). Last amended by Art. 5 G f. 29-8-2016 | 2034. 2005

[4] Amprion GmbH: Official Website. http://www.amprion.net/. – Retrieved on 23-01-2017

[5] 50Hertz Transmission GmbH: Official Website. http://www.50hertz.com/. – Retrieved on 23-01-2017

[6] TenneT TSO GmbH: Official Website. http://www.tennet.eu/. – Retrieved on 23-01-2017

[7] TransnetBW GmbH: Official Website. https://www.transnetbw.de/. – Retrieved on 23-01-2017

[8] 50Hertz Transmission GmbH, Amprion GmbH, TenneT TSO GmbH, TransnetBW GmbH: Possible scenarios for the network development plan electricity 2030 (German title: Szenariorahmen für die Netzentwicklungspläne Strom 2030). Design of the German transmission system operators. 2017

[9] Amprion GmbH, 50Hertz Transmission GmbH, TransnetBW GmbH, TenneT TSO GmbH: Netztransparenz.de – information platform of the German TSO's (German title: Netztransparenz.de - Informationsplattform der deutschen Übertragungsnetzbetreiber). https://www.netztransparenz.de/. – Retrieved on 31-01-2017