Descriptive Statistical Analysis of Frequency control-related variables of Nordic Power System

[1]Acosta, Martha N.; [1]Andrade, Manuel A.; [1]Vazquez, Ernesto; [2]Sanchez, Francisco; [3]Gonzalez-Longatt, Francisco; [4]Rueda, Jose Luis

[1] School of Mechanical and Electrical Engineering - Universidad Autónoma de Nuevo León
[2] Centre for Renewable Energy Systems Technology (CREST) - Loughborough University
[3] Institutt for elektro, IT og kybernetikk - Universitetet I Sørøst-Norge
[4] Department of Electrical Sustainable Energy - Delft University of Technology (TU Delft)

M. N. Acosta, M. A. Andrade, E. Vazquez, F. Sanchez, F. Gonzalez-Longatt and J. L. Rueda, "Descriptive Statistical Analysis of Frequency control-related variables of Nordic Power System," 2020 IEEE Power & Energy Society General Meeting (PESGM), 2020, pp. 1-5, doi: 10.1109/ PESGM41954.2020.9282021.
https://doi.org/10.1109/PESGM41954.2020.9282021

Publisher's version: DOI: 10.1109/PESGM41954.2020.9282021

Descriptive Statistical Analysis of Frequency control-related variables of Nordic Power System

Martha N. Acosta
Manuel A. Andrade
Ernesto Vazquez
School of Mechanical and Electrical Engineering
Universidad Autónoma de Nuevo León
Nuevo León, México
martha.acostamnt@uanl.edu.mx

Francisco Sanchez
Centre for Renewable Energy Systems Technology (CREST)
Loughborough University
Loughborough, United Kingdom
F.Sanchez@lboro.ac.uk

F. Gonzalez-Longatt
Institutt for elektro, IT og kybernetikk
Universitetet I Sørøst-Norge
Prosgrunn, Norway
fglongatt@fglongatt.org

Jose Luis Rueda
Department of Electrical Sustainable Energy
Delft University of Technology
(TU Delft)
Delft, Netherlands
J.L.RuedaTorres@tudelft.nl

***Abstract*—This paper presents a descriptive statistical analysis (DSA) of time-series of electro-mechanical quantities related to the frequency control (e.g. kinetic energy (KE), electrical frequency and power demand) of the Nordic power system (NPS). The idea of the DSA is to identify main observables features and patterns between these variables. Historical data publicly available has been used in this research paper; pre-processing included evaluating and identify missing data, and it filled by using the linear interpolation. The DSA uses descriptive statistical indicators to obtaining observable features. The dispersion analysis is used to observes how affects the KE to the electrical frequency. The data is grouped by weeks, days and hours, and its correlation coefficient was calculated. A correlation analysis between the KE and the power demand was computed, and the linear regression was used to construct a prediction model.**

***Index Terms*—Correlation coefficients, dispersion analysis, statistical analysis, patterns, power system.**

I. Introduction

The increasing connection of renewable generation sources (RES) in the power system has raised concerns to the transmission systems operators (TSOs) since the rotational inertia contribution of these kinds of generation sources are smaller (or even non-existent) than those of synchronous generators [1]. The low amount of inertia in the power system affects the electrical frequency response, and the power system becomes weak and less tolerant of disturbances. Therefore, the capability to estimate and track the inertia available in the power system in real-time would allow TSOs taking control actions as well as have more precise operational planning scenarios[2]. This research paper presents a descriptive statistical analysis (DSA) of time-series of electro-mechanical quantities related to the frequency control, specifically is looking for obtaining observable features (mean, standard deviation, correlation coefficients and dispersion of the data) and identifying patterns of three electro-mechanical variables of the Nordic power system (NPS): kinetic energy (KE), electrical frequency and power demand. If the observable features and patterns of these variables are identified, a model could be constructed that allows estimating and predicting the KE of the system using the electrical frequency measurements and the power demand forecast [2]. The main contributions of this research paper are: (i) it presents a statistical analysis of the three electro-mechanical variables (see Section II), (ii) the use of the presented DSA allowed obtaining observable features time-series of electro-mechanical quantities related to the frequency control in the NPS: kinetic energy (KE), electrical frequency and power demand. Getting the main observable features and patterns that will allow constructing a prediction model to estimate the KE. This paper is organised as follows: in Section II, a description of the historical data used for the DSA is given. Moreover, the pre-processing of the time-series data and the missing data filling is presented. Furthermore, the methods applied to realise the DSA described. Section III presents the results of the DSA of the KE, power frequency and power demand frequency. Finally, Section IV presents the conclusions.

II. Descriptive statistical analysis (DSA)

This section is dedicated to introducing the implemented DSA together with the data used in this paper.

A. Historic data

The historical data, publicly available of the NPS, is used in this paper. Essentially, three main electro-mechanical variables related to the frequency control are collected and analysed in this paper: the KE, the power demand (P_d) and the electrical frequency (f). Although a massive amount of data is available for illustrative purposes, this paper focuses on the data during the month of August 2019. Only one-month time series data was used for the DSA to determinate if the features and patterns that will be found give relevant information that indicates that an extensive data set is worth analysing. The historical time-series of the KE and the electrical frequency were obtained from Finland´s TSO company (Fingrid) database [3]. The KE raw data resolution is one sample per minute, while the electrical frequency data resolution is one sample every three minutes. Fingrid has no data about the power demand of the NPS; it is obtained from Nord Pool company database[4]. The power demand time-series has a resolution of one sample per hour. There is a clear problem with the granularity of the data.

B. Pre-processing data

Raw time-series has been collected and then imported from a .csv file into MATLAB®. The data structure selected was a matrix that contains seven columns: year (YY), month (MM), day (dd), day of the week (weekday or weekend), hour (hh), minute (mm), and variable values. Then, the matrix data structure was evaluated to identify potential missing data. The KE has 44,640 samples, of which 1.16% are missing samples. Meanwhile, the electrical frequency has 14,880, of which 0.14% are missing samples. The power demand has 744 samples, and it has no missing samples. Since the missing data rate, a simple process of filling the missing data was performed using the simplest principle, linear interpolation [5]. Due to the granularity of the data, time-series with different sample frequency, and to avoid adding noise or losing information, the time series were not resampled. The process used in this paper is: samples were taken from the variable with the highest sampling frequency corresponding to the time of the variable with lower sampling frequency. For instance, to perform the analysis between KE and the power demand, only the KE samples corresponding to each hour were taken to have the same resolution as the power demand, one sample per hour.

C. Methods

The techniques used to perform the DSA from the electrical frequency, KE and P_d of the NPS are presented [6]. Following subsections show the main statistical methods used in this paper. For the sake of simplicity, let consider a set of observed values of a random variable in the form of $\mathbf{x} = \{x_1, x_2, \ldots , x_N\}$.

The *standard deviation* (σ) is used to measure the time series data dispersion and is calculated as follows:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{\mathbf{x}})^2} \tag{1}$$

where x_i is the i-th observation of $\mathbf{x}$ and $\bar{\mathbf{x}}$ represents the mean value. The coefficient of variation (v) is used as a relative measure of the dispersion of the data concerning $(\bar{\mathbf{x}})$, and it is calculated as:

$$v = \frac{\sigma}{\bar{\mathbf{x}}} \times 100\% \tag{2}$$

The *correlation coefficient* (ρ): is a measure of the linear dependence between $\mathbf{x}$ and $\mathbf{y}$, and it is computed as follows:

$$\rho(\mathbf{x},\mathbf{y}) = \frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{x_i - \bar{\mathbf{x}}}{\sigma_{\mathbf{x}}}\right)\left(\frac{y_i - \bar{\mathbf{y}}}{\sigma_{\mathbf{y}}}\right) \tag{3}$$

where x_i is the i-th observed value, $\bar{\mathbf{x}}$ is the mean and σ_X is the standard deviation of $\mathbf{x}$. Meanwhile, y_i is the i-th observed value, $\bar{\mathbf{y}}$ is the mean and σ_Y is the standard deviation of $\mathbf{y}$.

The linear regression is a statistical technique for investigating and modelling the relationship between $\mathbf{x}$ and $\mathbf{y}$ and its model is:

$$\hat{y} = \beta_0 + \beta_1 x \tag{4}$$

where $\hat{y}$ is the predicted value from the fit, $\beta_0 = \bar{\mathbf{y}} - \beta_1\bar{\mathbf{x}}$ is the y-intercept and $\beta_1 = \sum_{i=1}^{N}(x_i - \bar{\mathbf{x}})(y_i - \bar{\mathbf{y}}) \big/ \sum_{i=1}^{N}(x_i - \bar{\mathbf{x}})^2$ is the slope. The suitability of the regression model can be expressed by computing the *goodness of fit* indicators: (i) the *sum of square due to error* (*SSE*) is a measure of the total deviation of $\hat{y}$ from the fit to the response values, (ii) the R^2 is the measure of how effective is the fit to explain the variations of the data, the adjusted R^2 is computed based on the residual degree of freedom and (iv) the *root mean square error* (RMSE) is an estimation of the standard deviation of the random components in the data.

D. Relationship between KE, electrical frequency and power demand

The variability of the electrical frequency in a power system depends mainly on two aspects: (i) the amount of KE stored in the rotating masses of the power system and (ii) the imbalance between the generation and the power demand. This relation can be represented with the swing equation [7]:

$$\frac{df}{dt} = \frac{f_0}{2H}P_m - P_e \tag{5}$$

where f_0 is the rated electrical frequency in p.u., H is the inertia constant in seconds. P_m and P_e are the mechanical power and electrical power in p.u., respectively. Since it is in terms of H, now the relationship between H and the KE most be defined. The inertia constant is commonly used as a measure of the stored energy in the rotating masses of the system and is calculated as follows [8]:

$$H = \frac{E_\mathrm{kin}}{S_B} = \frac{1}{2}J\omega_0^2\left(\frac{1}{S_B}\right) \tag{6}$$

where E_kin is the KE stored in a rotating mass in W·s. J is the moment of inertia expressed in kg·m^2, ω_0 is the rated speed in rad/s, and S_B is the rated apparent power in MVA. Substituting (6) in (5) the swing equation can be written in terms of KE as [9]:

$$\frac{df}{dt} = \frac{f_0}{2E_\mathrm{kin}}\left(P_m - P_e\right)S_B \tag{7}$$

From (7), to maintain the electrical frequency into its nominal value, P_m and P_e must always be equal. Following an imbalance, df/dt will depend on the amount of KE stored in the rotating masses. If the KE is low, the electrical frequency variation will be higher. Otherwise, if the KE is significant, the electrical frequency variations will be small. In the NPS, to maintain the electrical frequency into its nominal range (49.9 - 50.1 Hz) the active power reserves available are used. If the power demand is higher or lower than expected, an imbalance occurs between the power demand and the scheduled production a frequency deviation is produced and the *frequency-controlled reserves* (FCR) are activated to regulate up, and a new balance is reached. If the frequency deviation remains, the TSOs manually activated *the frequency restoration reserves* (FRR) within 15 minutes [10].

III. NUMERICAL RESULTS

This section is dedicated to present the DSA that was realised to the KE, electrical frequency and power demand. It is performed a data dispersion analysis from KE and electrical frequency. Then, the correlation and the correlation coefficients of KE and power demand are computed for data grouped weekly, daily and hourly.

A. *Dispersion analysis*

The objective of carrying out the dispersion analysis of the KE and electrical frequency data is to verify if the time-series data of these variables follows the behaviour described by (12). If it is assumed that when existing a power imbalance produced by the power demand variation, this imbalance will be corrected in the next 15 minutes by the activation of FFR. Therefore, within 15 minutes periods, the generators power dispatch remains constant, and the deviation of the electrical frequency will depend on the KE stored in the generators. For this analysis, it is taken from KE data set only the samples that correspond whit the observations of the electrical frequency every 3 minutes. Therefore, KE has the same number of samples as the electrical frequency, which are 14,880 samples. Once the data has the same resolution, it is grouped every 15 minutes, i.e., each group of data have five observations, and then mean, standard deviation and coefficient of variation are calculated. Table I presents the summary of the dispersion analysis results for the KE and electrical frequency of the NPS for August 2019.

TABLE I. DISPERSION ANALYSIS OF THE *KE* AND ELECTRICAL FREQUENCY FOR AUGUST 2019.

Variable	**Mean**		**Standard deviation (σ)**		**Coefficient of variation (v)**	
	Max	***Min***	***Max***	***Min***	***Max***	***Min***
KE [GW·s]	206.600	127.600	3.130	0.000	1.854%	0.000%
f [Hz]	50.115	49.876	0.071	0.002	0.140%	0.004%

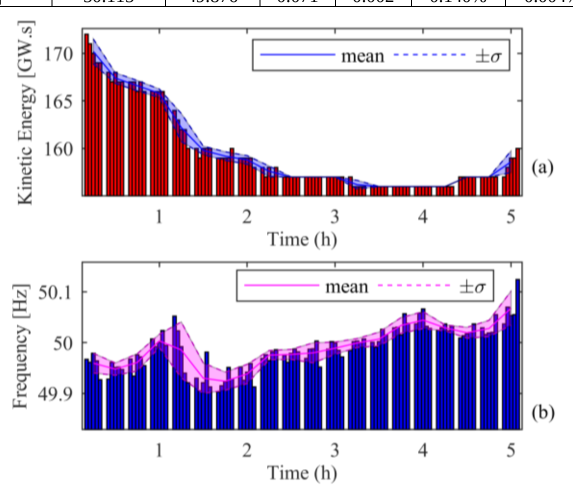

Figure 1. The mean and standard deviation of (a) KE and (b) electrical frequency for the first five hours of August 2019.

In Table I, the maximum standard deviation of KE is 3.130 GW·s and its maximum deviation concerning to the mean is 1.854%. These two indicators remain low within 15 minutes periods. Therefore, the KE tend to be close to the mean, and it can be concluded that the KE remains constant within 15 minutes periods and its variations depend on the change of power setpoint in the generators. Meanwhile, mean values of the electrical frequency are not into its nominal range, and this means that the among of KE is not enough to maintain the electrical frequency into its nominal values. Fig. 1 represents the standard deviation of KE and the electrical frequency in 15 minutes periods for the first five hours of August 2019.

The variation of the electrical frequency for 15 minutes periods does not depend exclusively on the variation of the KE. Observing hours 2:30, 2:45, 3:00, 3:30, 3:45, 4:00, 4:15, 4:30and 4:45, in Fig. 1 (a) the standard deviation is zero, i.e., the KE values remain constant within these periods, while the standard deviation of the electrical frequency for the same time periods is different from zero (see Fig. 1 (b)), which indicates that other factors like load-self regulation and the change in the power demand are causing these deviations.

B. *Correlation analysis*

The correlation analysis was used to determinate the degree of dependency of KE and the power demand of the NPS. Besides, the regression coefficients were computed to investigate whether it is possible to obtain a reliable prediction model. The time-series data of KE must be in the same resolution of the power demand to perform the correlation analysis, and since the power demand times-series data has a resolution of one sample per hour, to having the KE in the same resolution, it is taken the KE samples that correspond for each hour. Therefore, KE has 744 samples, equal to P_d.

All month considering weekends and without weekends: In this analysis, the measurements of all month were taken. Fig. 2 presents the time-series data grouped by weeks; it can be observed that KE and P_d follow two kinds of patter depending if it is weekdays or weekend days. P_d and KE present its lowest values between 0:00 to 5:00 hours; then they start to increase until reaching their maximum values around 10:00 hours and remains around these values until 21:00 hours when they begin to decrease to the minimum amount, each weekday follow this behaviour. On the other hand, on weekend days P_d and KE present its lowest values between 0:00 to 5:00 hours, then they start to increase until they reach their maximum value. However, the hours of the maximum values are different on Saturday and Sunday. These difference in the patterns are expected since the human activities are differents during the weekdays and weekend days, i.e., on weekdays the P_d consumption is defined by scheduled activities such as work hours, school hours, business opening and closing hours. Meanwhile, on weekends there are no specific activities planned [11].

Fig. 3 shows the correlation between KE and P_d of the NPS and the slopes of the linear regression for August 2019. The correlation coefficient, ρ, for this data set considering weekends is 0.8438 and without weekend days is 0.8536. ρ is higher when weekend days are despised than when they are admitted. The increase ρ when weekends are not considering indicates that the KE and P_d on weekends have no defined patterns, unlike weekdays, where the power demand follows the same pattern every 24 hours.

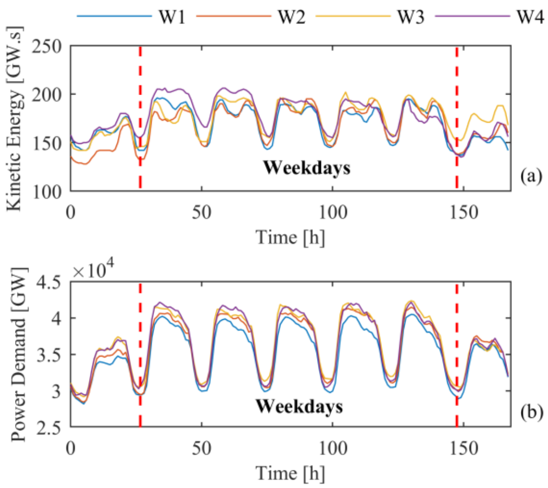

Figure 2. (a) KE and (b) P_d grouped by weeks.

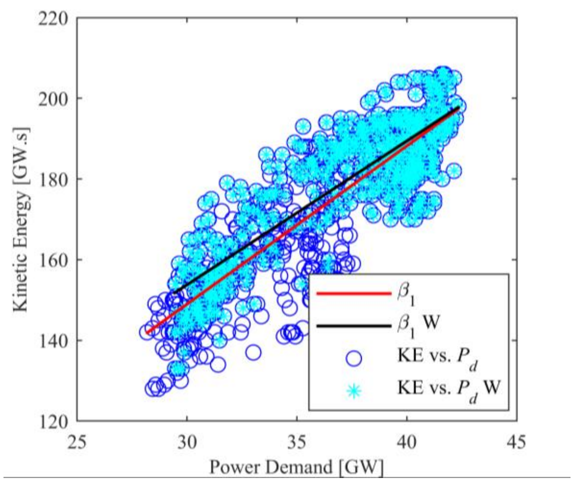

Figure 3. Correlation between KE and P_d: considering weekends (dark blue marks and red line) and without weekends (blue marks and dark line) of August 2019.

Since this value of correlation shows that exist a dependency between the P_d and KE, now it is essential to investigate whether this dependency allows making a reliable prediction model. It is known that the RMSE value is related to the degree of complexity of the model, i.e., to the degree of the linear regression model [12]. The degree of the linear regression model was chosen as one since the RMSE reduction between one and nine degrees is 3.5% and it is not worth adding complexity to the model if the decrease of RMSE is not significant. The results of performing the linear regression for KE and power demand data are summarised in Table II.

TABLE II. INDICATORS OF THE LINEAR REGRESSION OF KE AND POWER DEMAND FOR AUGUST 2019.

Indicator	***Values considering weekends***	***Values without weekends***
SSE	7.206×10^4	3.838×10^4
R^2	0.712	0.729
Adjusted R^2	0.711	0.728
RMSE	9.855	8.543

From Table II, the values of *SSE* indicate that the linear regression model has a high error component due to the fit only can represent 72.9% of the total variation in the data about the average. Moreover, when it is not considered the weekend measurements in the prediction model, the *SSE* and *RMSE* indicators decrease 46.7% and 13.3%, respectively. These results lead to conclude that it is advisable to consider creating two separate prediction models, one to represent the behaviour of the weekend days and another to describe the weekdays.

1) Data grouped by week: In this study, the data were grouped by weeks, and its correlation coefficients were computed. From Fig. 2, in each week, both KE and Pd have the same periodic variations over 24 hours, and graphically it can be concluded that both follow the same pattern every 24 hours. The correlation coefficient calculation was carried out to validate the above conclusion. The correlation coefficients without weekends have values between 0.85 and 0.92, as shown in Fig. 4. These coefficients show a strong correlation between KE and power demand, and it can be concluded that the four weeks of August follows the same pattern. Meanwhile, the correlation coefficients considering weekends have values between 0.84 and 0.89. The correlation decreases around 3% when weekends are included, and it indicates that weekday and weekend days have not the same patter.

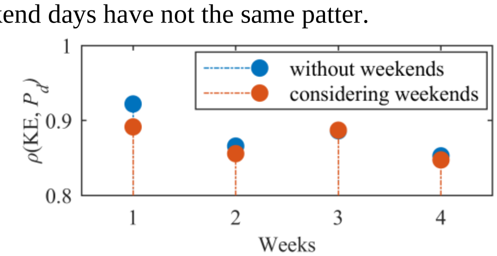

Figure 4. The correlation coefficients $\rho(KE, P_d)$ of the four weeks.

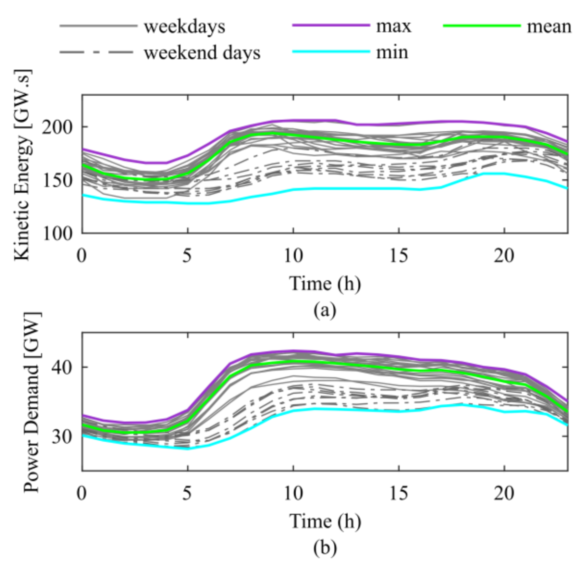

Figure 5. (a) KE and (b) P_d grouped by days considering weekends.

2) Data grouped by day: in this study, the data were grouped by days, and its correlation coefficients were computed. Fig. 5 presents the data of KE and the P_d of August 2019, including the measurements of weekends. In this figure, it is shown that two groups of data are formed: (i) weekdays around the mean and (ii) weekend days below the mean. Graphically it could be concluded that these two variables follow different patterns, a pattern representing weekdays and another pattern describing the weekend days.

The correlation coefficient between KE and P_d are presented in Fig. 6. The lowest values of the correlation coefficients, around 0.6, are those that correspond to the weekend days and it is expected since on weekends there are not defined patterns. Otherwise, weekdays have the highest correlation coefficient values, around 0.9. Therefore, these variables follow the same pattern in weekdays.

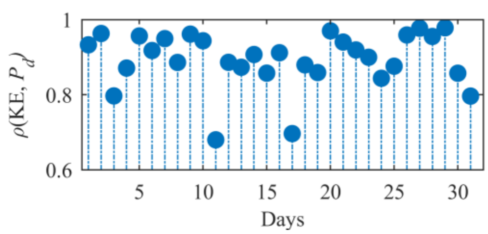

Figure 6. The correlation coefficient $\rho(KE, P_d)$ of days of August 2019.

3) Data grouped by hours: The correlation coefficient for the data grouped by hours is shown in Fig. 7. The range of values of the correlation coefficients that were considered with strong correlation is between 0.76 and 0.91, which correspond to the time range between 5:00 and 18:00 hours. The KE and power demand follow the same pattern and between 5:00 and 18:00 hours the hours are the same for all the month without considering the measures of weekend days.

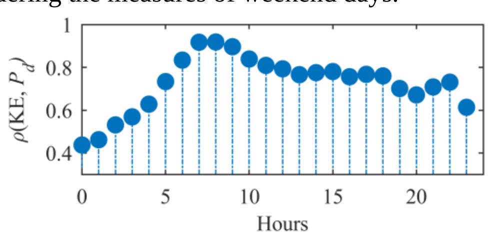

Figure 7. The correlation coefficient $\rho(KE, P_d)$ per hour without weekends.

IV. CONCLUSION

Based on the methods applied for the DSA, it is concluded that the KE values remain constant within 15 minutes periods since the maximum coefficient of variation does not exceed 2%, which means that the data are close to the mean. Besides, for the same 15 minutes periods, when the KE remains constant, i.e., the standard deviation is zero, the electrical frequency continues to vary. These variations manifest that the electrical frequency variation not only depends on the KE change; it also depends on load self-regulation and the difference in the power demand. The power demand and the KE have a strong correlation and follow the same pattern over 24 hours in weekdays. Otherwise, on weekend days, these variables do not follow the same pattern over 24 hours and therefore have a weak correlation. It suggests that it is necessary to create two separate models to represent the data: weekdays model and weekend days model. Although in this paper, it just analysed the electromechanical variables of one month, it is obtained relevant information about the behaviour of the variables between, minutes, hours, days and weeks. It was found that depends on the time scale it can be extracted specific information of the power system and hence create a particular model that related these variables.

ACKNOWLEDGEMENT

Ms Martha N. Acosta wants to acknowledge the financial support given by CONACYT (Mexico) and the support of the University of South-Eastern Norway.

REFERENCES

[1] F. Sanchez and F. Gonzalez-Longatt, “Optimization of Frequency Controller Parameters of a BESS by considering Rate of Change Constraints,” in *2019 IEEE Milano PowerTech, Powertech 2019*, 2019.

[2] H. R. Chamorro, C. A. Ordonez, J. C.-H. Peng, Francisco Gonzalez-Longatt, and V. K. Sood, “Coherency Estimation in Power Systems: A Koopman Operator Approach,” in *Computational Intelligence and Optimization Methods for Control Engineering*, J. S. Sáez, Maude, Josée Blondin, Panos M. Pardalos, Ed. Springer, Cham, 2019, pp. 201–225.

[3] “Fingrid.” [Online]. Available: https://www.fingrid.fi/en/. [Accessed: 10-Sep-2019].

[4] “Nord Pool.” [Online]. Available: https://www.nordpoolgroup.com/. [Accessed: 09-Oct-2019].

[5] L. Ehrlinger, T. Grubinger, B. Varga, M. Pichler, T. Natschlager, and J. Zeindl, “Treating Missing Data in Industrial Data Analytics,” in *2018 Thirteenth International Conference on Digital Information Management (ICDIM)*, 2018, pp. 148–155.

[6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. Elsevier, 2012.

[7] P. Kundur, N. J. Balu, and M. G. Lauby, *Power system stability and control*. New York: McGraw-Hill, 1994.

[8] P. M. Anderson and A. A. Fouad, *Power system control and stability*. New York: IEEE Press, 2003.

[9] H. R. Chamorro, A. C. Sanchez, A. Pantoja, I. Zelinka, F. Gonzalez-Longatt, and V. K. Sood, “A network control system for hydro plants to counteract the non-synchronous generation integration,” *Int. J. Electr. Power Energy Syst.*, vol. 105, pp. 404–419, Feb. 2019.

[10] ENTSOE, “Nordic Balancing Philosophy.”

[11] J. Torriti, “Understanding the timing of energy demand through time use data: Time of the day dependence of social practices,” *Energy Res. Soc. Sci.*, vol. 25, pp. 37–47, Mar. 2017.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., vol. 27, no. 2. New York: Springer New York, 2009.