State Estimation with Exogenous Information for Grids with Large Renewable Penetration

Basel Morsy *, Mahmoud AlSadat† and David Pozo‡
Center for Energy Science and Technology,
Skolkovo Institute of Science and Technology, Moscow, Russia
Email: *Basel.Morsy@skoltech.ru, †Mahmoud.Alsadat@skoltech.ru, ‡D.Pozo@skoltech.ru

***Abstract*—State estimation is one of the fundamental power system tools used in the determination of system state variables based on measurements. Current approaches for state estimation problems rely only on direct measurement and pseudo-measurements related with electric parameters only. In this work, we introduce a methodology for the inclusion of weather data, linked with renewable production, as exogenous measured parameters into the state estimation problem. To test our proposed framework, a simulation environment was developed and validated in a 5-bus system. Statistical significance and stability of our proposed framework were investigated by running 100 simulation experiments. We also, explored statistical test analysis for detection of dishonest manipulation of renewable power injected. Our numerical test showed that state estimation with exogenous weather parameter measurement could enhance state estimation accuracy by up to 79%. We also showed that, with 99% of confidence, our framework is able to detect 81.25% of dishonest cases with small cases of false positives.**

***Keywords*—State estimation, Renewable generation, Parameter Error, Bad Data, Dishonest generator**

I. Introduction

A power system is fully characterized by the network topology and system states (voltage magnitudes and phase angles) at each bus [1]. In order to obtain a full description of the power system, which is critical for the operation and control of a power system, these parameters should be known with a high degree of accuracy. However, there is always a degree of uncertainty associated with these electric parameters arising from errors within data acquisition process from measurements devices (typically an SCADA system) [2] [3]. State estimation models enables the approximation of the state variables true values (the estimates) as well as detecting and eliminating bad data. Current state estimation models rely on the use of power system measurements to perform state estimation. Of these models, the weighted least square model (WLS) is the most investigated and used one [4] [5]. It relays on minimizing the squared error between the measured parameters and their equivalent calculated values based on the estimated states, giving high weights to more accurate measurements and lower weights to less accurate ones [2].

Non-Electrical Exogenous Parameters - An Extended State Estimation: Taking into consideration the increasing level of penetration of renewable energy sources, weather data can be a good candidate for enhancing the current state estimation models. This is motivated by the fact that these energy sources power output is highly dependant on weather parameters, which can be easily measured (e.g. wind speed, temperature and solar irradiance). For this exogenous measurements to fit into the current model, a relation between the measured parameter (wind speed in our case) and the power system states must be obtained, which will be discussed in details in the following sections. While such concept may seem very promising, there exist a number of challenges that must be investigated before arriving at a stable model.

Paper Contributions: The main contributions of this work is the proposition of an extended state estimation for including additional data from weather parameters. The new framework relies on the standard estate estimation method, therefore, implementation in real energy management systems should not be complicated. Also, we have proposed a simulation environment of synthetically generated measurements that would mimic the real data retrieval from SCADA systems. We have run large number of experiments for proving, empirically, the benefits of using exogenous data. Finally, by using classical hypothesis test, we can detect dishonest situations of renewable generation.

Paper Organization: Section II introduces the state estimation problem with exogenous weather data retrieved from renewable sources. Section III presents validation setup, quality metric definition, and numerical tests performed for an illustrative 5-bus case study. Finally, section IV presents a discussion and conclusions of this work.

II. Methodology

A. Static State Estimation

Static-state estimation does not take into considerations transients in the system governed by physical laws, instead it assumes a static steady state operation [1] [6]. In this paper, we assume static state estimation. It can be defined as follows.

Assume there are m measurements collecting information of the electric parameters of a power grid. We define $\boldsymbol{z}$ as the m-dimensional vector of measurements. We assume that each measurement has an associated an error $\boldsymbol{e}$ that follows a normal distribution with mean 0 and standard deviation $\boldsymbol{\sigma}$, i.e., for each measurement i, the error $e_i = \mathcal{N}(0, \sigma_i)$. In vectorial format, we have

$$\boldsymbol{z} = \boldsymbol{z^{true}} + \boldsymbol{e} \tag{1}$$

where $z^{true} \in \mathbb{R}^m$ is a vector of the true values of the measured electric parameters. Let the system states (bus voltages and phase angles) be defined by vector $x \in \mathbb{R}^{2n}$. Let $h(x) : \mathbb{R}^{2n} \rightarrow \mathbb{R}^m$, the vector function, where each element is the governing physical law that relates the corresponding measurement in z to the system states x. Then, the weighted least square (WLS) approach for SE problem is defined as follows [1], [4], [5]:

$$\min_{x} J(x) = \sum_{i=1}^{m} w_i(h_i(x) - z_i)^2 = [h(x) - z]^\top W[h(x) - z] \quad (2)$$

The WLS optimization problem (2) is a non-linear and non-convex for power systems[1]. This problem can be solved by first-order optimally conditions and using Newton-Raphson-type methods for solving first-order optimally conditions. The iterative solution method could be described as follows:

$$x^{k+1} = x^k - G^{-1}(x^k)g(x^k) \quad (3)$$

where,

$$G(x^k) = H^\top(x^k)WH(x^k) \quad (4)$$

$$g(x^k) = H^\top(x^k)W[h(x) - z] \quad (5)$$

and $H(x) = \frac{\partial h(x)}{\partial x}$.

Elements of $h(x)$ belong to a set of functions given as follows:

1) V_i: Voltage at bus i
2) P_i: Active power injection at bus i
3) Q_i: Reactive power injection at bus i
4) P_{ij}: Active power flow from bus i to bus j
5) Q_{ij}: Reactive power flow from bus i to bus j

From power flow equations derived in [5]:

$$P_i = V_i \sum_{j=1}^{n} V_j(G_{ij}\cos\delta_{ij} + B_{ij}\sin\delta_{ij}) \quad (6)$$

$$Q_i = V_i \sum_{j=1}^{n} V_j(G_{ij}\sin\delta_{ij} - B_{ij}\cos\delta_{ij}) \quad (7)$$

$$P_{ij} = V_iV_j(G_{ij}\cos\delta_{ij} + B_{ij}\sin\delta_{ij}) - V_i^2G_{ij} \quad (8)$$

$$Q_{ij} = V_iV_j(G_{ij}\sin\delta_{ij} - B_{ij}\cos\delta_{ij}) - V_i^2(B_{ij} - b_{s,ij}) \quad (9)$$

B. Assumptions

1) Steady state or quasi-steady state operation is investigated. Although a power system can never achieve a steady state due to the continuous changes in loads and generations, yet, it is reasonable to approximate the system as operating in steady state conditions for short periods of time [2].
2) We assume no topological changes are taking place.
3) Renewable energy bus (Wind turbine in this work) is modelled as load (PQ) bus of which active power load is negative of the generation and reactive power load is *zero*.
4) Exogenous parameter used is wind speed measurement for wind turbine bus.
5) Wind speeds less than 3 m/s are set to *zero* and wind speeds greater than 12 m/s are set to 12 to resemble the pitching mechanism taking place with wind turbines.

[1]There is the exception when power flow equations are linearized, i.e., the system is linear. Then, the WLS approach became convex

C. Extended State Estimation: Inclusion of Exogenous Parameters

In order to include exogenous parameters in the formulation of SE problem, it is necessary to find adequate analytical expression that relate these exogenous parameters to power system states. In this work, we are concerned with the inclusion of wind speed measurements in SE. In [7], the physical law that relates wind speed to generated power from a wind turbine is given by:

$$P_w = \frac{1}{2}\rho S v_w^3 C_P \quad (10)$$

where P_w is wind turbine generated power, ρ is air density, S is the surface area swept by the wind turbine rotor, v_w is wind speed and C_P coefficient of power of the wind turbine. Let $\frac{1}{2}\rho S C_P$ be a constant K, then:

$$v_w^3 = \frac{P_w}{K} \quad (11)$$

It is easy to extend equation (11) representing a single turbine, to a wind farm.

The extended state estimation would modify the vector function $h(x)$ to add another element that has to satisfy the constraint equation (11), i.e., $h(x) : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{m+1}$. Similarly, a new exogenous data point, *wind speed* v_w, can be added to the measurements vector $z \in \mathbb{R}^{m+1}$.

III. Case Study

A 5-bus system is investigated for different operating points. The SE is with and without the addition of exogenous parameters to assess the newly proposed model and evaluate the enhancement based on arbitrarily defined quality metrics defined in detail in the following subsections. Numerical simulations were performed in Julia 1.2 with the optimization suite JuMP [8], [9].

The system topology is shown in Figure 1. Demands are located buses {2,3,4}, wind generated power is located at bus 5, and bus 1 is a conventional generator bus. Constant K in equation (11) is calculated for our assumed wind turbine of 56 m diameter and 0.38 C_P to be 573.265, all values given are in SI units.

For simulation, we will use mimic the operation for 8 days on an hourly basis, which sums up to be 192 time instances, equivalently, 192 snap-shots of the system. The data used for demands and wind speed, hence wind power, were generated arbitrarily based on historical data. Details on input data can be found in the online appendix [10].

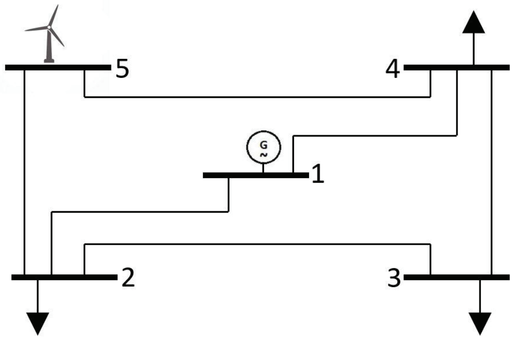

Fig. 1. Single-line diagram for the system investigated

A. Simulation Environment

In order to validate the proposed extended state estimation model, we developed a simulation environment to investigate the effect of inclusion of exogenous parameters. The simulation environment is composed of 4 sequential layers. Figure 2 shows a diagram of the simulation environment.

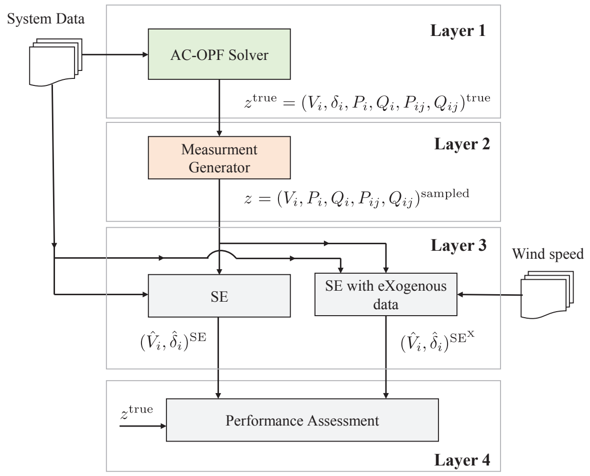

Fig. 2. Simulation environment for numerical experiments

The first layer resembles the physical system operation with 100% certainty of all information about it and satisfaction of all physical laws governing the power system. In this layer, the optimal power flow (OPF) is solve for the input grid data, and full and true information about the system electric parameters are provided. The second layer simulates the data that usually is obtained from the SCADA system, i.e., measurements. The true values of the physical system obtained in layer 1 are modified by adding noise normally distributed to mimic measurement error. Each time that the layer 2 is run, it would generate different samples, due to stochastic nature of the error. The layer 3 contain the two approaches for solving the state estimation problem. It here, we perform state estimation with and without exogenous parameters and save results. Finally, the layer 4 assesses the performance of each model according to some certain quality metrics that will be discussed later. Data from the ground true from the layer 1 is used in the fourth one.

In fact, our simulation environment resembles what happens in real cases, except that we do not have certain information about the physical power system, we only have measurements, that is the reason why it is important to perform state estimation.

In order to get statistically significant results, we will run 100 experiments, i.e., we run 100 times the previous steps. At the same time, each experiment compiles results from the 192 operating points. Each experiment is indexed by s.

B. Quality Metrics

We have chosen the sum of mean square error of voltages magnitude and phase angles to construct an arbitrarily defined index which we call *error index* (EI^s), the smaller it is, the closer the model to the real physical system operation.

The mean square error of voltages for each node i and experiment s, MSEV_i^s, is defined by (12). Similarly, the mean square error of phase angles, $\mathrm{MSE}\delta_i^s$ is defined and they are calculated as follows.

$$\mathrm{MSEV}_i^s = \frac{1}{T} \sum_{t=1}^{T} \left(\hat{V}_{i,t}^s - V_{i,t}^s \right)^2 \tag{12}$$

$$\mathrm{MSE}\delta_i^s = \frac{1}{T} \sum_{t=1}^{T} \left(\hat{\delta}_{i,t}^s - \delta_{i,t}^s \right)^2 \tag{13}$$

where $\hat{V}$ and $\hat{\delta}$ represent the estimates for voltage and phase angles, respectively. Then, the *error index*, (EI^s), is defined as follows:

$$\mathrm{EI}^s = \sum_{i=1}^{I} \left(\mathrm{MSEV}_i^s + \mathrm{MSE}\delta_i^s \right). \tag{14}$$

Note that EI^s is a vector with the size equal to the number of experiments performed. Statistical information about EI^s, MSEV_i^s, and $\mathrm{MSE}\delta_i^s$ vectors and its median, denoted by $\mathrm{med}_s\left[\,\cdot\,\right]$, for all experiments will be used in the next subsections.

C. Input Data

The input demands and wind energy generation are given to the system as a time-series of 192 periods. Detailed data information is provided in the online appendix [10].

An additional case was considered in the numerical tests. In this case, the wind power generation data was assumed to be higher than the maximum possible capacity given the wind speed at this case. This, actually, resembles dishonest renewable generation (DRG). This case was assumed to be at the last two days of the period considered within our study. Power generation at wind energy bus is then given as follows in Figure 3.

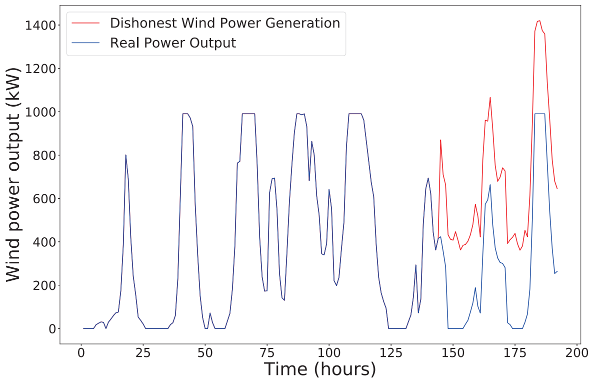

Fig. 3. Generation at wind turbine bus with DRG

D. Results

The aforementioned input data was fed into the system, using the work flow discussed in the simulation environment description. The SE performance with and without inclusion of exogenous parameters were assessed subject to the different quality metrics discussed previously (mean squared error and error index). The following results were found.

1) No DRG Case: As data were provided for the first 6 days with no DRG (144 hours, consequently 144 snap-shots of the system), all fundamental procedures were then performed 144 times, one time at each snap-shot of the system, including SE. The following Figure 4 and Figure 5 show the value for median of the MSE for bus voltages, $\text{med}_s\left[\text{MSEV}_i^s\right]$, and phase angles, $\text{med}_s\left[\text{MSE}\delta_i^s\right]$. It can be seen that for bus voltages and phase angles, the MSE values for SE with exogenous parameters is lower than that of SE without exogenous parameters for all buses.

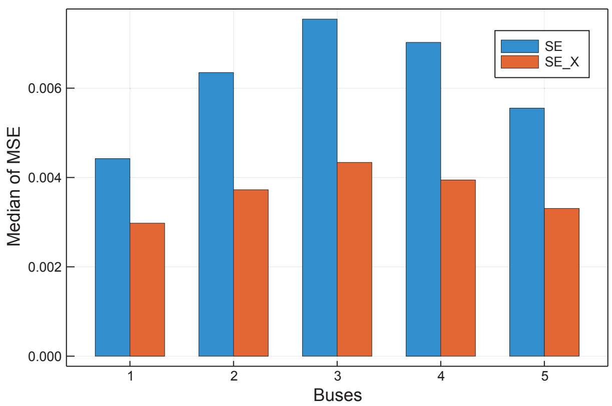

Fig. 4. Median of MSE of bus voltages

The overall assessment of both SE models is also done by evaluation of the median of the *error index* for both of them considering the 100 run experiments. As shown in table I the value for error index for SE with exogenous parameters is less

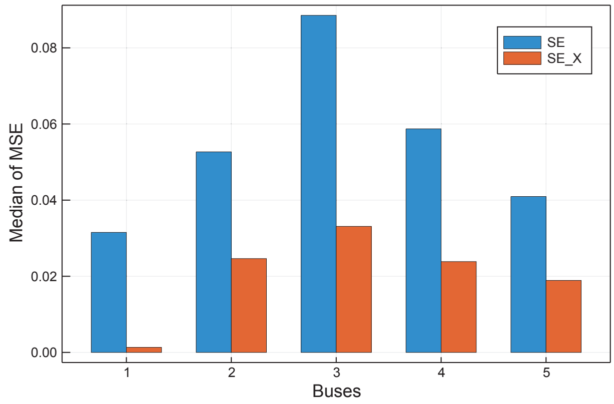

Fig. 5. Median of MSE of phase angles

than that for SE without exogenous parameters. The values are namely $\text{med}_s\left[\text{EI}_s^{\text{X}}\right] = 0.0546$ and $\text{med}_s\left[\text{EI}_s\right] = 0.2650$. This means a decrease in error index by 79% caused by inclusion of exogenous information of wind speed in the state estimation.

2) Case Analysis with DRG Presence : In case that wind turbine bus generates more active power than the available capacity concluded from wind speed, which can happen in case of DRG, the inclusion of exogenous parameters is expected to give worse results than not including them. It would also be the case when wind speed measurements are considered as bad data, both cases are equivalent. As shown in Figure 6 and Figure 7, the median MSE for bus voltages and phase angles is larger for SE with exogenous parameters inclusion. This case of presence of spoiled data, can be detected and identified. In the next subsections we provide further details on it.

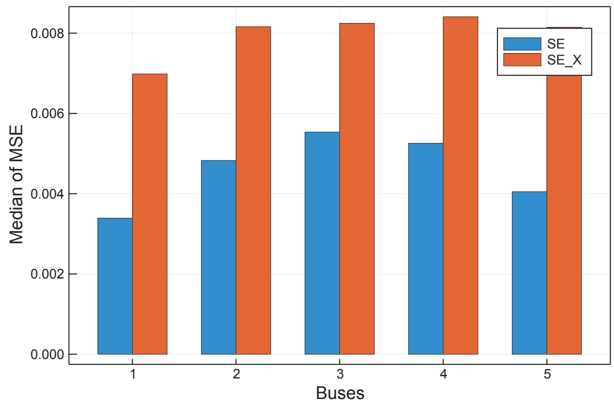

Fig. 6. Median of MSE of bus voltages in case of DRG

As shown in Table I, the error index for SE with exogenous parameters is much higher than that of SE without. $\text{med}_s\left[\text{EI}_s^{\text{X}}\right] = 0.1409$ while $\text{med}_s\left[\text{EI}_s\right] = 0.0001$, which indicates that the inclusion of exogenous parameters is not quiet useful when there is DRG.

To further elaborate on the overall assessment of the two models, Figure 8 shows the results of the 100 experiments

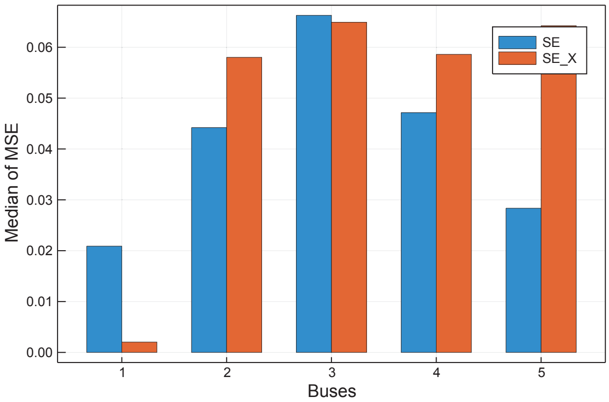

Fig. 7. Median of MSE of phase angles in case of DRG

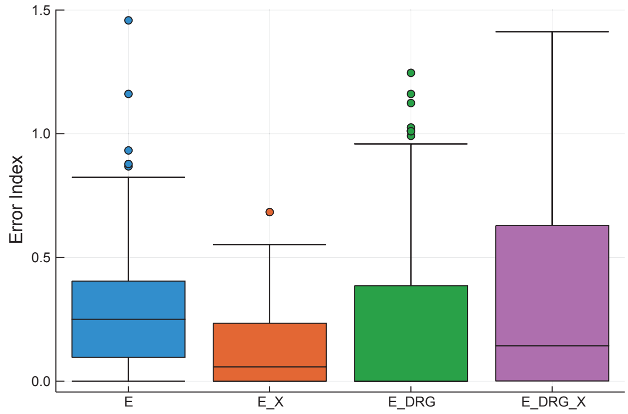

Fig. 8. Error index box plot

TABLE I
COMPARISON OF ERROR INDICES

	$\text{med}_s\left[\text{EI}_s\right]$	$\text{med}_s\left[\text{EI}_s^X\right]$
Without DRG	0.2650	0.0546
With DRG	0.0001	0.1409

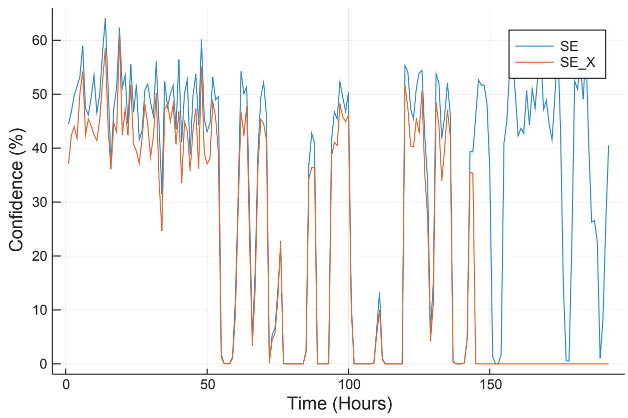

Fig. 9. Confidence percentage of null hypothesis

performed. As shown in the figure, the first two box plots shows the case of no dishonest renewable generation (DRG) and the inclusion of exogenous parameters enhanced the performance of SE indicated by a lower error index median value. The median value is chosen to be representative as 50% of the experiment population lies there. It is also noticed that the inter-quartile range of our introduced model is less than that of the conventional SE model, which indicates a higher stability. On the other hand, it can be seen that in case of DRG the value of the error index of our introduced model is higher than that of the conventional SE, and this is plausible as the DRG introduces inconsistency between the supposed renewable generation and the actual one we have. It can also be seen that the inter-quartile range is larger for SE with exogenous parameters which indicates less stability than conventional SE.

3) Confidence of Absence of Bad Data: We consider a null-hypothesis, that is there is no bad data. We tested this hypothesis using the classical χ^2-test as done in [11] and [1]. As shown in Figure 9 the period from hour 0 to hour 144, where there is no DRG, the two models are consistent in terms of the confidence level of the null hypothesis. However, in the last 48 hours, where there is DRG, we can see that the null hypothesis is totally rejected for the SE with exogenous parameters, which means that we are almost sure that there exists bad data during this period of time. It can be seen that the conventional SE could not detect this anomaly of DRG because no exogenous parameters were included, which means that the model does not distinguish between power generated from renewable and non-renewable sources (source blind).

4) Detection of DRG Using Normalized Residuals Test: Using the conventional Largest Normalized Residual statistical Test (LNRT), [1], source of measurements classified as bad data can be identified. In the presented case study, we artificially generated bad data for the DRG for the last 48 hours or our case study (see Figure 3).

We have run LNRT for each time instance and for all experiments. Figure 10 shows the values of normalized residuals for each measurement in a particular experiment and time instance, 160. In this case, we can observe that measurements 10 and 40 are the largest residuals and way above of the limits for the 99% confidence level that we have set. In particular, measurements 10 and 40, which are corresponding to injected active power in bus 5 (wind turbine bus) and wind speed measurement respectively. We identified a large inconsistency between the wind speed and wind power generated that is reported. In this case, we label it as “detected dishonest renewable generation”.

After running LNRT for all time instances, we can construct the confusion matrix for the DRG detection. It is showed in Table II. At time instances when the estimated wind speed is zero, the normalized residuals were undefined. Other than these times, the normalized residuals were calculated and for a confidence level of 99%. The proposed LNRT could detect 81.25% of the cases of dishonest renewable generator

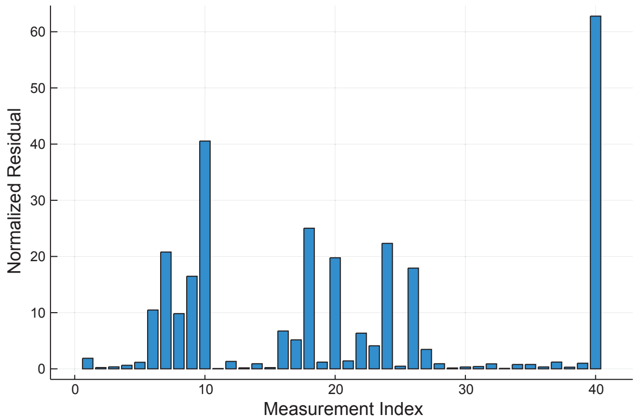

Fig. 10. Normalized residuals of the time instance 160

behaviours (true positive).

TABLE II
CONFUSION TABLE FOR DRG DETECTION

		Actual	
		DRG	No DRG
Predicted	DRG	39	2
Predicted	No DRG	3	124

IV. Conclusion and Discussion

As discussed in the results section, adding exogenous measurements improved the accuracy of state estimation by about 79%. This accuracy improvement was during the first 6 days where there was no inconsistency in the measurements (no spoiled data from dishonest generation is added). While during the last 2 days, DRG case, the state estimation with exogenous measurements was much worse than the state estimation without exogenous parameters. Additionally, we have presented statistical hypothesis test that allows us to detect bad data in the state estimation outcomes, and further, identified what measurements are responsible to introduce bad data. We have used these tests to identify possible dishonest behaviors from renewable generators.

Although the presented framework showed an enhancement in the state estimation performance when tested on a 5-bus network, it would be necessary to extend numerical analysis on more complex networks. This includes networks of larger number of buses as well as different topologies.

Another important aspect to consider is the number of available measurements. In this work, we considered that all the power system measurements are used in the state estimation. Where in reality this is not the case always. Also, in order to further investigate the effect of the exogenous parameter on state estimation performance, we need to consider fewer number of power system measurements. We also need to consider the instantaneous ratio between renewable generation and conventional generation, as this would give an estimate of how critical and effective the inclusion of exogenous parameters is.

Finally, we envision that machine learning tools can also be considered in this problem. These tools can help with regards to two main aspects. First, machine-learning-based classification and regression models could improved our current framework with regard to bad data detection and identification. A second aspect is related to finding relations between relevant weather data and power system states of which a mathematical model is hard to obtain (e.g. temperature).

Acknowledgement

This work was supported by Skoltech NGP Program (Skoltech-MIT joint project).

References

[1] A. G. E. Ali Abhur, *Power System State Estimation*. CRC Press, 2004.
[2] F. C. SCHWEPPE and J. WILDES, "Power system static-state estimation, part i: Exact model," *IEEE Transactions on Power Apparatus and Systems*, 1970.
[3] O. A. N. Vempatis, et al, "Generalized state estimation," *Proceedings of the 20th International Conference on Power Industry Computer Applications*, 1997.
[4] P. Zarco and A. G. Exposito, "Power system parameter estimation: A survey," *IEEE TRANSACTIONS ON POWER SYSTEMS*, 2000.
[5] J. John J. Grainger, William D. Stevenson, *Power System Analysis*. McGraw-Hill, Inc., 1994.
[6] B. F. W. A. J. Wood, *Power generation, operation and control*. John Wiley and Sons, Ltd, 1996.
[7] T. Ackermann, *Wind Power in Power Systems*. John Wiley and Sons, Ltd, 2005.
[8] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V.B., "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, Jan. 2017.
[9] M. Lubin and I. Dunning, "Computing in Operations Research Using Julia," *INFORMS Journal on Computing*, vol. 27, no. 2, pp. 238–248, Apr. 2015.
[10] B. Morsy, M. AlSadat, and D. Pozo, "State estimation with exogenous information for grids with large renewable penetration," 2020. [Online]. Available: https://sites.skoltech.ru/paco/?attachment%20id=496
[11] T. P. Vishnu, V. Viswan, and A. M. Vipin, "Power system state estimation and bad data analysis using weighted least square method," in *2015 International Conference on Power, Instrumentation, Control and Computing (PICC)*, Dec 2015.