SCADA-agnostic Power Modelling for Distributed Renewable Energy Sources

Ahlam Althobaiti, Anish Jindal, Angelos K. Marnerides
School of Computing & Communications, Lancaster University, Lancaster, UK
Email: a.althobaiti@lancaster.ac.uk, anishjindal90@gmail.com, angelos.marnerides@lancaster.ac.uk

***Abstract*—Distributed Renewable Energy Sources (DRES) are considered as instrumental within modern smart grids and more broadly to the various ancillary services contained within the energy trading market. Thus, the adequate power production profiling and forecasting of DRES deployments is of vital importance such as to support various grid optimisation and accounting processes. The variety of DRES installation companies in conjunction with the diversity of ownership on DRES machinery, controller firmware and Supervisory Control and Data Acquisition (SCADA) software leads to cases where centralised SCADA measurements are not entirely available or are provided under a subscription-based model. In this work, we consider this pragmatic scenario and introduce a SCADA-agnostic approach that utilises freely available weather measurements for explicitly profiling and forecasting power generation as produced in real wind turbine deployments. For this purpose, we leverage various machine learning (ML) libraries to demonstrate the applicability of our system and further compare it with forecasting outputs obtained when using SCADA measurements. Through this study, we demonstrate a viable and exogenous profiling solution achieving similar accuracy with SCADA-based schemes under much lower computational costs.**

***Index Terms*—Distributed renewable energy sources, machine learning, machine learning, SCADA, wind power.**

I. Introduction

The use of fossil fuels for power generation has led to the exponential growth of air pollution and in turn, climate change and global warming. According to the International energy agency, there was a 2.3% rise in energy consumption just in 2018, which caused CO_2 emissions to rise by 1.7% leading to an alarming value of 33.1 Gt of CO_2 in the air [1]. Hence, modern smart grid deployments adopt greener power generation solutions based on distributed renewable energy sources (DRES) including Photo Voltaic (PV) solar panels, wind turbines and bio fuel.

Nonetheless, as the DRES generated power output depends solely on intermittent environmental conditions (e.g., ample solar radiation, wind speed), there is always a level of uncertainty in terms of the power contribution that such deployments offer back to the main power grid. Under the objectives of a sustainable grid, it is therefore crucial to adequately profile and further forecast DRES power production. Grid optimisation routines rely on accurate DRES profiling, thus inaccurate and unavailable DRES profiling is highly likely to trigger resilience havoc with a number of severe consequences.

Power generation profiling for DRES has been the subject of investigation in a number of studies (e.g., [2], [3]). The majority of these studies engage with the assumption that measurements from Supervisory Control and Data Acquisition (SCADA) systems are always available and the various modeling components are restricted on explicitly utilising power generation values from such systems. However, the complex business and operational processes within large-scale power grids involve a diversity of ownership in terms of machinery (e.g., wind turbines) as well as control and measurement components (e.g., PLCs, SCADA) [4]. Thus, the acquisition of SCADA-based measurements is not always available, particularly for DRES installation owners that are not operators of either power transmission or distribution networks.

In this piece of work, we tackle the aforementioned scenario and propose a generic SCADA-agnostic DRES power profiling scheme. In general, the proposed DRES profiling system enables automated feature selection and tuning of machine learning (ML)-based regression models and can adapt to diverse measurement feeds. Through a proof-of-concept study explicitly on wind-turbine deployments, we demonstrate that our system is capable to adequately operate with the use of freely available third-party weather measurements. Thus, we introduce a SCADA-agnostic approach that can sufficiently aid or complement current profiling and forecasting practises. Through this study, we firstly build a ground-truth and confirm that under a SCADA-based approach and using real SCADA measurements from a wind turbine farm, we obtain accurate forecasting of power generation. Subsequently, we prove that a SCADA-agnostic approach produces comparable forecasting accuracy with lesser computational cost for the same DRES deployment.

The main contributions of this work are two-fold and summarised as follows:

- A generic DRES profiling system enabling adaptive feature selection as well as automated best-fit ML model tuning under low computational costs.
- A SCADA-agnostic cost-efficient approach relying strictly on freely available third-party weather measurements to model DRES deployments with a proof-of-concept evidence over real wind turbine deployment profiling.

The rest of this paper is organised as follows. Section II presents some related work on DRES power generation profiling. Section III describes the datasets and the methodology of the proposed approach whereas Section IV discusses the

evaluation conducted. Finally, Section V concludes and summarises this paper.

II. Related Work

Measurement-based wind power profiling can be split into two categories; (i) single- or multi-measurement SCADA systems placed at a local wind farm, distribution or transmission operator, and, (ii) non-SCADA, sensor-based measurement systems deployed on individual devices (e.g., single/multiple wind turbines) or aggregation points of a given DRES deployment.

There has been a number of studies utilising single-measurement SCADA systems where a single parameter is utilised for profiling the generated wind power from a set of wind turbines. For instance, both studies in [5] and [6] strictly rely on wind speed timeseries and employ a spline regression model and Support Vector Machines (SVMs) respectively to model power generation of wind turbine deployments for power forecasting purposes. The authors in [7] propose a refined power curve modeling approach by utilising SCADA-based wind speed timeseries. Moreover, the work in [8] focuses on correlating wind speed with the output power of a given wind turbine such as to improve the examined modeling method [8]. Furthermore, the work in [3] models wind power generation using Artificial Neural Networks (ANNs) in synergy with six different parameters obtained from the local SCADA system.

Examples of wind power modelling based on conventional or customised non-SCADA sensor-based measurements are evidenced in studies such as [2] and [9]. In fact, the authors in [2] utilise statistical meta-features of raw measurements used in [9] to highlight the use of non-parametric methods (e.g., stochastic gradient boosted regression trees, randomized forests) for univariate wind power modeling.

Nonetheless, all of the above studies assume either the presence of locally (or regionally) deployed SCADA systems or sensor-based measurement components. To the best of our knowledge, none of the studies considered the pragmatic assumption that SCADA measurements as well as sensor components are not always present or available in all DRES deployments. Hence, an alternative SCADA-agnostic approach is required to confront such scenarios.

III. Data Description & Methodology

A. Data description

The herein reported proof-of-concept study focuses explicitly on profiling wind turbine deployments using third-party, freely available weather measurements under the assumption that SCADA or locally placed sensor-based measurement data is not available. However, in order to validate the performance of our exogenous-based wind power modeling, we utilise SCADA measurements gathered from a real deployment.

The used SCADA-based dataset was captured at the La Haute Borne wind farm, located in Meuse, France [1] and represents daily measurements gathered for the whole year of 2017[2]. The La Haute Borne wind farm consists of 4 Senvion MM82 wind turbines where measurements are obtained on 10-minute samples. Within each 10-minute sampling bin, there are 34 features related to various electro-mechanical (e.g., torque, rotor speed), power (e.g., apparent power, grid voltage) and environmental parameters (e.g., wind speed, outdoor temperature) explicit to a given wind turbine.

As already mentioned, our SCADA-agnostic scheme depends solely on third-party weather measurements that are freely available. For this purpose, we have extracted environmental measurements (e.g., wind direction) from the Dark Sky API [10] and Weather Online API [11] over the same observational period in which ground truth SCADA measurements were obtained for the La Haute installation. Moreover, we acquired wind and output temperature measurements from Weathernews [12] as observed by the Nancy-Ochey weather station which is geographically adjacent to the La Haute Borne wind farm. Both the third-party measurements as well as the SCADA-based measurements were processed within our generic DRES profiling system that we explain next.

B. DRES profiling system

This study relies on a system built to efficiently pre- and post-process DRES measurements such as to automatically identify the most suitable features within a best-fit ML model. The generic properties contained within our implemented system can serve the basis for close-to-real-time profiling of any type of DRES deployment (e.g., wind turbine/farm, solar PV panels etc.)[3].

As depicted in Fig 1, the first process within the implemented system is to pre-process diverse DRES measurements gathered either by conventional SCADA or sensor-based data acquisition deployments. Hence, the pre-processing module ensures that raw timeseries measurements of various features (e.g., wind speed, humidity, output power, etc.) are refined in terms of missing values, noisy timeseries and (re)sampling. Subsequently, the system normalises the pre-processed timeseries and feeds them directly to a feature selection software component that works in synergy with a ML component. Ultimately, the combination of the best statistical features alongside the best-fit model is chosen based on a repetitive auto-tuning process. We describe the mechanics of each individual stage and component by focusing on the proof-of-concept wind power modeling scenario as follows.

1) Data pre-processing: The quality of the aggregated data is a substantial factor for wind power modelling as missing or inconsistent data samples can affect the accuracy of power measurement estimation. Essentially, noisy data are inconsistent measurements caused by sensor reading errors or SCADA controller faults. In general, we have witnessed

[1]Explore – ENGIE France Renewable Energy Open Data, Available: https://opendata-renewables.engie.com/pages/home/

[2]The 2017 dataset is the most complete in comparison with all datasets for other years provided by ENGIE.

[3]The complete DRES profiling system is available on Github at: https://github.com/Ahlam-Althobaiti/-DRES-Power-Modeling

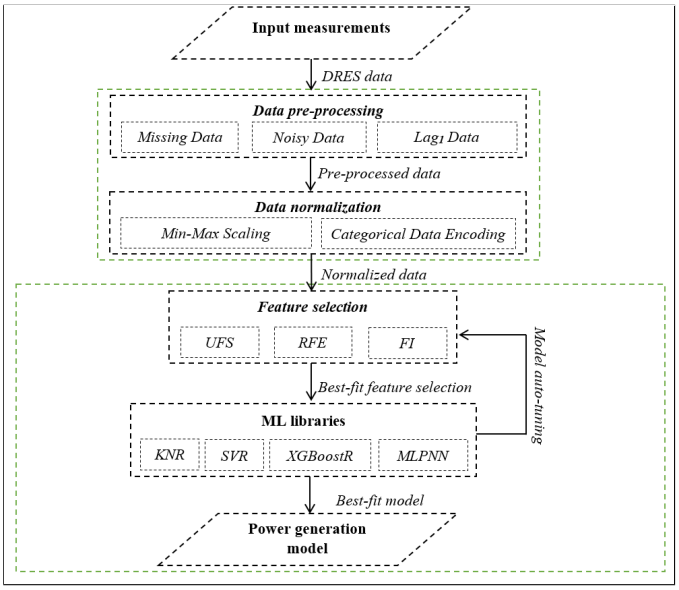

Fig. 1: Measurement-based DRES profiling system

missing samples resulted by turbine unavailability or electrical shut-down, icing events or out-of-range samples due to weather API object pull failures. Therefore, prior to considering the absolute power measurements, the SCADA-based measurement dataset is subjected to a filtering technique such as to remove all possible inconsistent and missing data. For instance, the generated power P_t^{avg} in kW should be as:

$$P_t^{min} \leq P_t^{avg} \leq P_t^{max} \tag{1}$$

where $t \in T$ is the coordinated time period, $P_t^{min} = 0$ and P_t^{max} = the wind turbine nominal power. Otherwise, the generated power P_t^{avg} can be considered inconsistent data and thus are filtered out. To be noted that these thresholds depend on the explicit wind turbines' specification and for other turbine models, these could differ.

Moreover, during the pre-processing stage, analyses on power measurements is performed through the utilisation of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) in order to build an underlying statistical ground-truth of the assessed timeseries. In simple terms, the ACF represents the correlation between the P_t^{avg} measurement of the $t \in T$ and the measurements at previous time lags. However, PACF is the correlation between P_t^{avg} and P_{t+k}^{avg} after removing the influence of the confounding variable:

$$P_{t-1}^{avg}, P_{t-2}^{avg}, ..., P_{t-k+1}^{avg} \tag{2}$$

2) Data normalization: Our DRES profiling system employs a min-max normalisation scheme such as to reconstruct the assessed timeseries in the range $[0, 1]$ with $n \times m$ vectors as given in Eq. (3). In this case, n is the number of the samples, m is the number of feature vectors and $t \in T$ is a time interval.

$$\bar{x_t} = \frac{x_t - x^{min}}{x^{max} - x^{min}} \tag{3}$$

where $\bar{x}_t$ represents the normalized value of x_t, x^{min} and x^{max} are the minimum and maximum value in each feature vector $z \in m$ respectively. The normalization procedure is only applicable to the numerical features. The dataset consists of a mixture of numerical as well as categorical features. The categorical features are transformed into numerical data as the proposed system takes only numerical input. The categorical data is encoded using the binary encoder function as follows:

1) An integer value is assigned to every unique category for a given categorical feature.
2) A new binary feature is created for each integer-encoded category.
3) New columns are created based on the majority of the bit encoding.

Unlike numerical features, binary encoded features only take binary values of 0 or 1, hence they do not need to be re-scaled or normalized.

3) Feature selection: The proposed DRES profiling system employs an automated feature selection process such as to obtain an adequate and effective set of attributes. Hence, the feature selection component is in charge of assessing the importance of the raw SCADA or third-party weather measurements. As evidenced in Fig. 1, the feature selection component works in synergy with the ML library component such as to identify the optimal set of features producing the best-fit ML-based power regression model. In more detail, the current prototype supports: i) Filter-based Univariate Feature Selection (UFS), ii) Wrapper-based recursive feature elimination (RFE) and iii) Ranking-based Feature Importance (FI).

The UFS technique is used to assign the importance scoring of each feature. Thus, each feature is linearly regressed and produces an estimated value that is scored against the original value under the F-score metric. Essentially, the F-score denotes how the regressed value of a given input behaves in terms of the averaged accuracy precision. Our current prototype supports both the univariate linear regression filtering of features as well as filtering through the ranking of correlations based on the Pearson correlation metric. Both filtering mechanisms are used interchangeably. By contrast with UFS, the RFE method recursively selects features by removing the less important features from the feature set using importance-based rankings. Our current prototype utilizes the Random Forest (RF) estimator for importance-based rankings and it has proven to be beneficial in occurrences of highly correlated features (e.g., wind speed and output power) [13]. Within the FI approach, a similar RF-based feature reduction is performed such as to isolate the most significant attributes. It is to be noted that both RFE and FI use RF to remove the least significant features; however the FI in contrast with the RFE is less robust as it is just based on a given threshold value and a single iteration.

4) ML-libraries component: The implemented DRES profiling system depends heavily on the collaborative functioning between the feature selection component and the ML-libraries component. The ML-libraries component is implemented under a pluggable fashion in which off-the-shelf or customised ML algorithms can inter-operate with the algorithms residing within the feature selection process. The synergy between the aforementioned components is orchestrated under a repetitive

feedback mechanism such as to identify the most optimal combination of features with an identified ML-based profiling model. Moreover, optimal hyper-parameters for the ML-based techniques employed are found by using a grid search technique with a k-fold cross-validation method [14]. In order to address aspects of non-linearity in the examined features as well as properties of non-stationary DRES measurements, we have implemented both supervised as well as unsupervised ML-based regression algorithms. In particular, the current prototype supports: i) K-nearest Neighbours Regression (KNR), ii) Support Vector Regressor (SVR), iii) Gradient Boosting Regressor (XGBoostR) and, iv) Multi-layer Perceptron Neural Network (MLPNN). We next describe the basic properties of each implemented algorithm.

KNR: The KNR model utilises feature vector similarity (or neighborhood) and predicts the value of new input samples. Thus, the value assigned to new input samples is based on the resemblance with training samples. In summary, KNR is decomposed into three main stages;

1) Calculation of the Euclidean distance between the new input data instance with each training samples given by:

$$D_t = \sqrt{\sum \mid x_t^{train} - x_t^{new} \mid^2} \tag{4}$$

where x_t^{train} and x_t^{new} represent the values of training sample and the new input data respectively.
2) k nearest samples are selected based on the closest Euclidean distance values.
3) Inserting the average of the k-nearest points as the predicted value of the new input instance.

SVR: The SVR model is a supervised scheme enabling the estimation of a fit function based on pre-computed training samples such as to map high-dimensional model inputs to the target output. Unlike other regression algorithms focusing on prediction error rate reduction, SVR fits any prediction errors within a a tolerable error (ϵ). Hence, describing the highest deviation from the targets, while keeping the fit function as flat as possible.

XGBoostR: The XGBoostR algorithm relies on the boosting idea is aiming to improve the regression stability of a weak learner that promote weak statistical hypotheses related to their input data instances. In general, a weak learner represents models holding slightly better performance than a random chance with respect to prediction error rates. XGBoostR depends on three components performing: i) loss function optimisation with respect to regression errors, ii) weak learner prediction for one decision at a time and, iii) weak learner additive model minimising the total loss function.

MLPNN: The MLPNN algorithm belongs in the category of supervised feed-forward artificial neural network (ANN) formulations and consists of more than one perceptrons. The input layer in MLPNN is used to receive input data, whereas the output layer is responsible for predicting the output value of a given input. Internally, the composition of the training model within MLPNN is performed by a back-propagation scheme. As within a traditional artificial neural network, hidden layers reside between input and output layers which work as computational engines. In particular, MPLNN exploits the correlation or dependencies between the variables used in the computed training to model the output value by tuning weight parameters such as to reduce prediction errors.

C. Evaluation Methodology

We conduct a thorough evaluation in order to assess the performance of the exogenous SCADA-agnostic wind power modeling in comparison with modeling performed using SCADA-based measurements. Our evaluation methodology is diagrammatically depicted in Fig. 2.

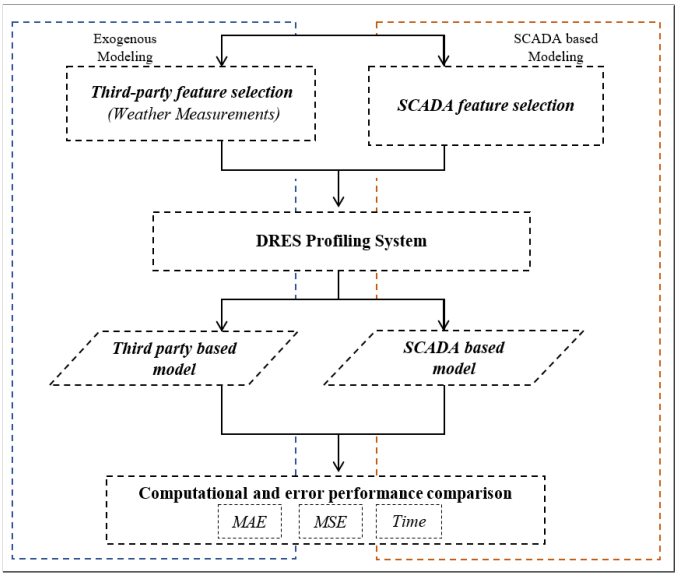

Fig. 2: Evaluation methodology.

Both SCADA and SCADA-agnostic data streams are passed through our DRES profiling system prototype to obtain the most optimal features with the best-fit regression models. Prior to the modeling as well as the feature selection phase, the correlation of individual generated power with its past measurements is extracted and integrated to the input measurements of the designed system. Subsequently, we perform a seasonality grouping for every type of measurements for better classification. Hence, we split our datasets in the four seasons of the year (i.e. spring, summer, autumn, and winter) for each wind turbine and re-sample the measurements to behave under hourly bins. The DRES profiling system assigns 70% of the feature samples to be used for training for any of the algorithms within the ML-libraries component and 30% for testing. Subsequently, the repetitive process between the feature selection component and the ML-library component takes places such as to identify the most optimal features for the best-fit model. The resulted models are assessed based on two error and one computational cost metric. The indices considered in this work in terms of prediction error are the mean absolute error (MAE) and the mean squared error (MSE), whereas for computation, we account the time taken to obtain a prediction. We briefly describe each metric as follows.

1) **MAE:** It depicts the mean of all absolute values of the difference between the actual and predicted power values defined as:

$$MAE = m^{-1} \sum_{t=1}^{m} \mid x_t - \hat{x}_t \mid \quad (5)$$

where $t \in T$, m is the test set length, $x_t, \hat{x}_t$ represent the actual power measurements and the estimated power measurements, respectively.

2) **MSE:** It depicts the mean of the squares of all differences between the actual and predicted powers defined as:

$$MSE = m^{-1} \sum_{t=1}^{m} (x_t - \hat{x}_t)^2 \quad (6)$$

3) **Computational complexity:** The time taken by the ML-based model within the DRES profiling system to produce prediction for the output power of a given wind turbine.

IV. Evaluation

A. ACF and PACF analysis

As a part of our pre-processing software component presented in Section III-B1, we utilize ACF and PACF analysis to test the correlation structure of the generated power measurements. Fig. 3 presents the result of ACF analysis. It can be observed that there is a high positive correlation with the lags outside of the 95% confidence interval.

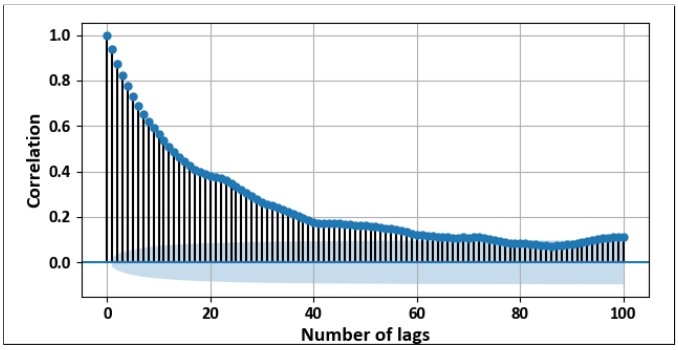

Fig. 3: Auto-correlation function (ACF) of the generated power.

We observe from ACF plot a high inter-correlation among the historical components of the generated power measurements. This can give rise to unreliable statistical inferences due to multi-collinearity. Therefore, we use the PACF plots, to only retain the relevant lags, in contrast to the complete ACF plot, and remove those which yield indirect correlations. In Fig. 4, the PACF plot shows that lag_1 has the highest positive correlation before it first intersects the confidence interval. Therefore, we utilize the lag_1 values of the generated power as a feature feed to the learning techniques in this study and it can be written as:

$$Lag_1(P_t^{avg}) = P_{t-1}^{avg} \quad (7)$$

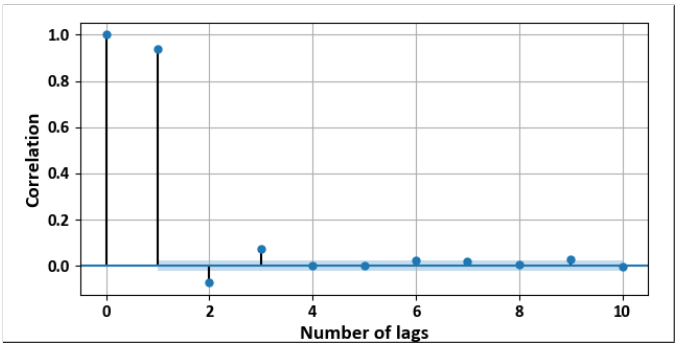

Fig. 4: Partial autocorrelation function (PACF) of the generated power.

B. SCADA-based wind power modelling

As discussed in Section III-C, our evaluation methodology firstly targets to compose a ground truth profiling model using SCADA-based measurements from the La Haute Borne wind farm. Hence, a total of 28 year-wide SCADA-based features were initially scrutinised by the feature selection component within the DRES profiling system presented in Section III-B. The iterative feature selection process within the DRES profiling system has demonstrated that the RFE technique produced the best set with a total of 13 SCADA-based features for SVR. The filtered set of features is composed by a range of mechanical (e.g., pitch angle and generator converter speed), power (i.e., apparent power and the lag_1 feature) and weather (i.e., wind speed) features. Whereas, FI for KNNR, XGBoostR and MLPNN with a total of 3 features including a the converter torque, apparent power and the lag_1 measurements. Hence, these ML techniques covered all exogenous as well as intrinsic factors related to the wind-turbines behaviour in terms of power generation.

Under the combination of the selected features with the various ML-based regression components of the DRES profiling system, we have witnessed improved regression models in all of the ML-based algorithms. As illustrated by Fig. 5, the designed feature selection schemes positively impacts the performance of KNNR, SVR, XGBoostR and MLPNN, reducing MAE errors to around 0.000847, 0.00027, 0.000621 and 0.005623 kW respectively, where the MAE for these (based on all features) are approximately 0.008232, 0.000302, 0.000975, 0.013537 kW. Similar trends are also observed for the MSE. In parallel, the SVR model under the RFE-based feature selection produced an extremely low MAE and MSE; MAE0.00027 kW and MSE ≈ 0 kW^2, for almost all sampling bins.

C. SCADA-agnostic wind power modelling

Following the same pre-processing, normalization and feature selection performed within the DRES profiling system (as with the SCADA-based profiling), we have produced regression models using third-party weather features. Besides the lag_1 of the power measurements, there was the identification of more two weather features within the core learning process. Our process utilised measurements from the three third-party

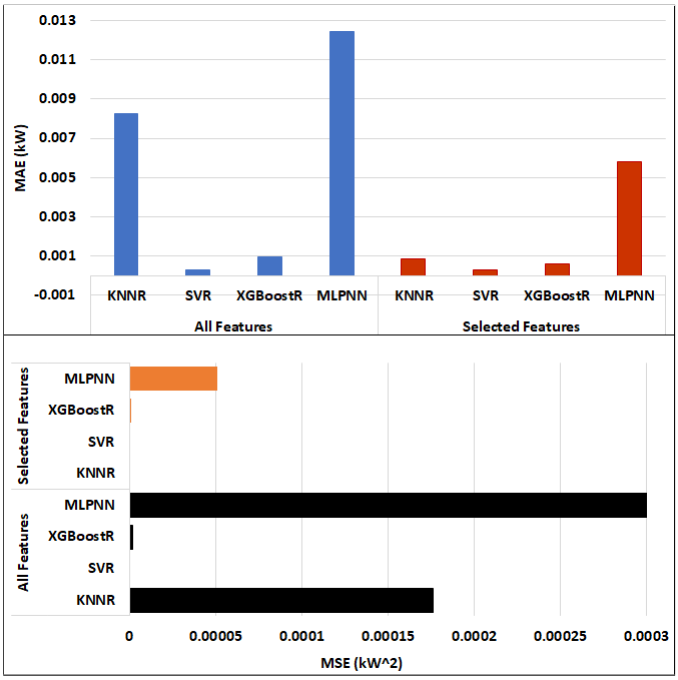

Fig. 5: Prediction errors for wind power modelling based on SCADA measurements.

data providers including measurements such as wind direction and gust.

We observe that the prediction results for wind power regression based on the freely available third-party weather features varied slightly from the SCADA-based profiling. Nonetheless, the conducted experiments indicate no major difference in the obvious pattern for the estimated power curves as depicted in Fig. 6. Moreover, the performance analysis of the resulted ML-based regression models in relation to the MAE and MSE respectively shows that SVR outperforms the rest of formulations.

As evident, the SVR technique has a minimum MAE and MSE, where MAE is $0.003487\ kW$ and MSE $\approx 0.0\ kW^2$. Meanwhile, the MAE for KNNR, XGBoostR and MLPNN are 0.008863, 0.004184 and $0.003757\ kW$, and MSE are 0.00018, 0.000062 and $0.00004\ kW^2$ respectively. Hence, the error performance shows a slightly higher error rate than SCADA-based but arguably to be of minimal importance for large-scale accounting and optimisation processes as required by the main grid. In parallel, under the scenario of a windfarm owner or third-party company with no access to SCADA measurements, we highlight that the approximate generation and potentially financial forecasting is not necessarily affected on a macroscopic scale. In addition, the actual SCADA-agnostic estimation is of minimal financial cost in comparison with a subscription-based SCADA-based approach as it usually happens.

As depicted in Fig. 7, the computational cost for producing a reasonable regression model is far smaller using a SCADA-

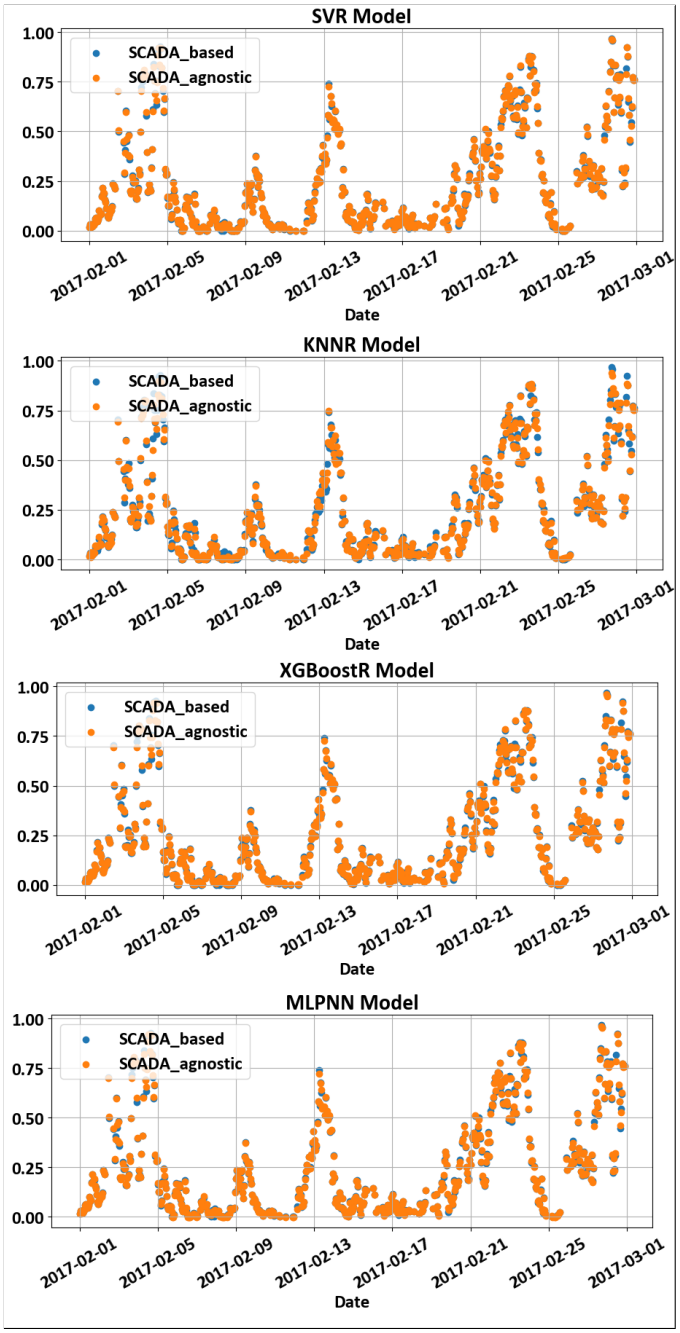

Fig. 6: SCADA-based and SCADA-agnostic power curve.

agnostic approach in comparison to SCADA-based[4] approach. We also witness that the KNNR model can act as a good approach for real-time use, however with some minimal trade-off with respect to their error rate performance. For long-term estimation processes, we observe that the MLPNN alongside the SVR formulation promotes slightly more accurate SCADA-agnostic wind power profiling.

In general, the simplicity of utilising just three freely available features in comparison to expensive SCADA-based

[4]On a 64-bit Windows operating system with Intel Core i7 ($7th$ Gen) CPU with 2.70 GHz clock cycle and 12 GB RAM.

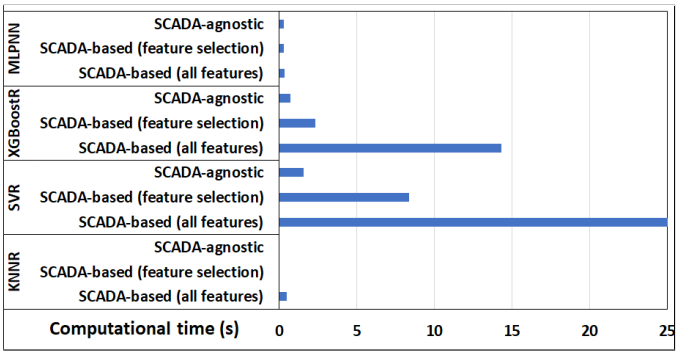

Fig. 7: Computational time comparison.

monitoring and measurement components could effectively pave the path towards new directions on real-time and low-cost DRES power profiling.

V. Conclusion

The increasing utilisation of DRES in the modern smart grid engages a complex energy trading model with vague policies in terms of hardware and software ownership in DRES deployments. Hence, it is not uncommon for independent DRES deployment owners to not have a complete control or access of their installations through SCADA systems managed by third-party providers or main grid operators. In this work, we propose a SCADA-agnostic DRES profiling system and exhibit its applicability on a proof-of-concept study over a real wind turbine installation. We demonstrate that by simply utilising freely available third-party weather data with available regression models, we can reasonably match up to a great scale the regression accuracy performance of models utilising SCADA measurements. Moreover, the proposed SCADA-agnostic profiling is achieved with a minimal set of weather features in contrast to the SCADA-based approach and under a lower computational cost. Thus, paving the path towards independent and cost-efficient power generation profiling serving a range of envisaged smart grid applications such as Virtual Power Plants.

Acknowledgment

The authors are grateful to Weathernews for providing data. This work has received funding from the EU's Horizon 2020 research and innovation programme for "EASY-RES" project under grant agreement No 764090 and Taif University.

References

[1] International Energy Agency, "Global energy and co2 status report/the latest trends in energy and emissions," 2018, accessed: 2019-12-06. [Online]. Available: https://www.iea.org/geco/electricity/.

[2] O. Janssens, N. Noppe, C. Devriendt, R. Van de Walle, and S. Van Hoecke, "Data-driven multivariate power curve modeling of offshore wind turbines," *Engineering Applications of Artificial Intelligence*, vol. 55, pp. 331–338, 2016.

[3] F. Pelletier, C. Masson, and A. Tahan, "Wind turbine power curve modelling using artificial neural network," *Renewable Energy*, vol. 89, pp. 207–214, 2016.

[4] K. Leahy, C. Gallagher, P. O'Donovan, and D. T. O'Sullivan, "Issues with data quality for wind turbine condition monitoring and reliability analyses," *Energies*, vol. 12, no. 2, p. 201, 2019.

[5] Y. Wang, Q. Hu, D. Srinivasan, and Z. Wang, "Wind power curve modeling and wind power forecasting with inconsistent data," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 16–25, 2018.

[6] T. Ouyang, A. Kusiak, and Y. He, "Modeling wind-turbine power curve: A data partitioning and mining approach," *Renewable Energy*, vol. 102, pp. 1–8, 2017.

[7] E. Taslimi-Renani, M. Modiri-Delshad, M. F. M. Elias, and N. A. Rahim, "Development of an enhanced parametric model for wind turbine power curve," *Applied Energy*, vol. 177, pp. 544–552, 2016.

[8] Y. Zhao, L. Ye, W. Wang, H. Sun, Y. Ju, and Y. Tang, "Data-driven correction approach to refine power curve of wind farm under wind curtailment," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 1, pp. 95–105, 2017.

[9] T. Jin and Z. Tian, "Uncertainty analysis for wind energy production with dynamic power curves," in *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*. IEEE, 2010, pp. 745–750.

[10] "Dark sky api," 2020, accessed: 2020-01-22. [Online]. Available: https://darksky.net/dev

[11] "World weather online," 2020, accessed: 2020-01-22. [Online]. Available: https://www.worldweatheronline.com/developer/api/

[12] "Weathernews," 2020, accessed: 2020-01-22. [Online]. Available: www.weathernews.fr

[13] B. F. Darst, K. C. Malecki, and C. D. Engelman, "Using recursive feature elimination in random forest to account for correlated variables in high dimensional data," *BMC genetics*, vol. 19, no. 1, p. 65, 2018.

[14] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and svm-based data analytics for theft detection in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005–1016, 2016.