Ensemble Regression Model-Based Anomaly Detection for Cyber-Physical Intrusion Detection in Smart Grids

Anna Magdalena Kosek, Oliver Gehrke
Technical University of Denmark
Department of Electrical Engineering
Energy System Operation and Management
Email:{amko,olge}@elektro.dtu.dk

***Abstract*—The shift from centralised large production to distributed energy production has several consequences for current power system operation. The replacement of large power plants by growing numbers of distributed energy resources (DERs) increases the dependency of the power system on small scale, distributed production. Many of these DERs can be accessed and controlled remotely, posing a cybersecurity risk. This paper investigates an intrusion detection system which evaluates the DER operation in order to discover unauthorized control actions. The proposed anomaly detection method is based on an ensemble of non-linear artificial neural network DER models which detect and evaluate anomalies in DER operation. The proposed method is validated against measurement data which yields a precision of 0.947 and an accuracy of 0.976. This improves the precision and accuracy of a classic model-based anomaly detection by 75.7% and 9.2%, respectively.**

Keywords*—Data-driven modelling, machine learning, cyber-physical security, model-based anomaly detection, ensemble regression, power system.

I. Introduction

Power systems are critical infrastructures for industry, transportation, health care, water and food supply, telecommunication and financial systems. Cybersecurity in power grids is a topic of increasing concern [1], and a considerable effort is required to secure the infrastructure from cyber-attacks. This includes securing legacy systems and designing new systems with security in mind [2]. The discipline of cybersecurity analyzes threats, vulnerabilities and risks for computing systems and proposes defense mechanisms [3]. Initially, cybersecurity in power systems has focused on communication standards [4] including Advanced Metering Infrastructure [5], and SCADA security [6]. More recently, a new type of approach has been used which takes the cyber-physical nature of power systems into account, i.e. the interaction between the physical power system and the ICT infrastructure used in its operation (e.g. [7]).

Cybersecurity measures can be categorized along the time domain as preventive, real-time or post-mortem. Intrusion detection systems (IDS) gather and analyze the information from a computer network or system in order to discover malicious activities or violations of policy. Two general types of detection techniques are used in IDS: anomaly-based or signature-based. Current IDS focus on the analysis of software and network traffic, but do not usually take the physical component of a cyber-physical system into consideration. In this paper we investigate an intrusion detection method based on physical component models. Using the example of a photovoltaic (PV) generator as the potential target of a cyber-attack [2], we analyze operational data to detect anomalies in its operation which may be further classified as resulting from unauthorized or malicious control inputs.

In the data mining context, anomaly detection is concerned with identifying rare data instances or events that do not match an expected pattern. Applications include the detection of financial fraud, identification of manufacturing faults and monitoring of computers in data centers [8]. Three types of anomaly detection techniques can be distinguished: supervised, semi-supervised and unsupervised. Supervised methods use a fully labelled training set to train a classification method which distinguishes normal behaviour from different types of anomalies. Semi-supervised methods (so called model-based anomaly detection) use partially labelled data to create a model of normal behaviour and compare the model output to the observed network or system behaviour. Unsupervised methods assume that the total number of anomalies is small in comparison to the normal data points in the training set. Based on this assumption, statistical anomaly based techniques analyze operational data in order to distinguish between normal and anomalous operation through statistical inference tests.

In this work we investigate model-based anomaly detection for DERs. Anomaly detection with a single regression model has been used for PV fault diagnostics [9], wind turbine fault detection [10], and discovering cyber-attacks on SCADA [6]. In [9], the authors use redundant linear and non-linear models to detect different faults in PV operation. Anomaly detection in control systems with use of a linear model of the normal behaviour is investigated in [11]. Contextual anomaly detection was used in a cyber-physical IDS (Intrusion Detection System) to detect malicious voltage control actions [12]. Here, the proposed on-line method utilizes a model which is trained on data known to have no malicious control actions or sensor faults, therefore the trained normal model is accurate.

This paper continues the work presented in [12]. The proposed method recognises anomalies in pre-recorded data of DER operation; it therefore focuses on post-mortem analysis, detecting past occurrences of control events. The contribution of this paper is as follows: a) a novel model-based anomaly detection method using ensemble regression, in section II-A; b) a new method for selecting model training set to improve the anomaly detection performance, in section II-B; We further verify of the proposed method against the DER operation data, in section IV and perform quantitative comparison of the proposed method against single model anomaly detection, in section VI.

II. Cyber-physical intrusion detection system

The concept of a cyber-physical intrusion detection system (CP-IDS) was proposed in the SALVAGE project [13] as presented in figure 1 [12]. The CP-IDS uses data from the observed cyber-physical system and analyses it under three aspects: DER operation, power system vulnerability and cyber-security threat. The outcome of this analysis is passed to a cyber-physical analysis component. Here, all three aspects are combined into a joint cyber-physical security assessment.

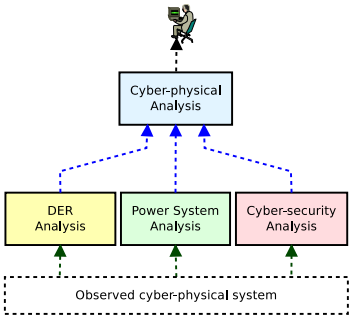

Fig. 1. Cyber-physical IDS architecture [12].

Work presented in [12] is an on-line power system and component analysis for the purpose of discovering malicious voltage control events. In this paper we focus on the DER analysis component of the CP-IDS and propose a method for off-line (post mortem) DER control detection. In this context, four operational states of a DER can be distinguished:

- Normal operation: a DER behaves as expected and its operation is not influenced by external set points. An internal DER controller may or may not govern the operation of the unit.
- Faulty operation: the operation of a DER deviates from normal due to a fault at the unit or in its electrical network environment.
- Verified control: a DER behaves as expected under a verified control scheme, or according to authorized external set points.
- Malicious control: a DER is operated under an unverified control scheme or according to unauthorized external set points.

In this paper we define a DER behaviour anomaly as either verified or malicious control, and consider faulty operation as part of normal DER operation to exclude it from the detection algorithm. After an introduction to model based-anomaly detection in section II-A we describe data cleaning and selection in section II-B and model training in section III. We present a model based anomaly detection method with ensemble regression models in section V and apply it to a PV plant data set in section IV. In section VI we compare this approach to several single model approaches and evaluate the method for selecting model training set.

A. Model-based anomaly detection

In the proposed model-based anomaly detection method, normal DER behaviour is modelled in the *DER model* component (figure 2). The output of the model is compared to sensor measurements (or target data) in the *Anomaly Detection* component. Differences between normal and observed DER behaviour can originate form several sources: sensor error, model error, DER fault, or malicious or verified DER control. The output of the model-based anomaly detection is either a label (class) or an anomaly score for every data input.

Fig. 2. Flow diagram of the model-based anomaly detection

Ensemble learning combines several models to produce a prediction to solve classification and regression problems [14]. The increased robustness and accuracy of ensemble methods over single model methods was reported in [15]. Ensemble learning consists of three steps: generation, pruning and integration. First several redundant models are generated, then the set of models is pruned by removing some of the generated models, finally the base model results are combined to create the ensemble prediction [14]. An overview of ensemble regression approaches for generation, pruning and integration are presented in [14]. The ensemble is evaluated by the degree of agreement between predictions represented by their overall spread. The ensemble prediction is usually evaluated in terms of an average of the individual predictions (mostly using equal weight averaging).

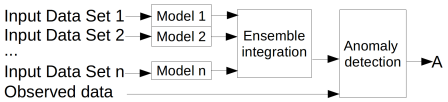

Fig. 3. Ensemble model-based anomaly detection architecture.

The proposed ensemble model-based anomaly detection (EM-AD) uses two or more DER normal behaviour models which produce the same output variables based on disjoint sets of inputs. The additional *Model merging* component calculates the final model output that is next compared to the observed output in the *Anomaly detection* component.

In this paper we apply the EM-AD method to a PV component and implement it as a proof of concept, using

historical time series of power and meteorological measurements obtained from a PV plant. The model building method is presented in section II-B.

B. Model building

The semi-supervised anomaly detection uses partially labelled data to train the normal model. Since the historical data has not been labelled, we use correlation analysis as a method for selecting a training set to improve the normal model and consequently enhance the anomaly detection performance. The chosen model building stages are as follows: data cleaning, aggregation, data scoring with correlation analysis, model data labelling and selection, removal of missing values, normalization, ANN model creation with supervised model training.

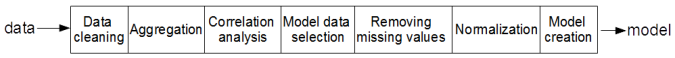

Fig. 4. The proposed model building method.

Figure 4 shows the data preparation and model building processes. The following section describes these processes in detail.

1) Data cleaning: Data cleaning detects and removes errors and inconsistencies in data in order to improve its quality [16]. In this paper, the observations of modelled phenomena are produced by sensors. Many errors can be hidden in raw sensor data; therefore the data needs to be cleaned before it can be used for modeling and analysis. The classification of data quality problems in data sources was proposed in [16]. According to this classification, the data can originate from a single measurement unit (single-source) or several measurement units (multi-source). Single source problems include schema and instance level issues. The schema level issues can be addressed with the mechanism of data storing and its integrity constrains. Instance-specific problems cannot be prevented at the schema level and include word misspellings and sensor errors. Multi-source data quality issues are consequences of integrating multiple sources of data. Resulting from conflicts due to different data models and representations, overlaps and contradictions can appear in the integrated data. In this work we consider both singe and multi-source data quality issues. In section IV-B of this paper we focus on single-source instance problems and will not cover sensor errors and multi-source instance level problems considering inconsistent timing.

2) Aggregation: The aggregation process targets two issues: model training time and missing values. Firstly the aggregation decreases the amount of the data that need to be processed to train the model, reducing the computation time for the model training. Secondly the column-wise aggregation based on mean or average on a matrix with some missing values (represented as NA) uses all available information from partially missing data samples. This process allows integrating the partially missing samples into the data set without removing them.

3) Correlation analysis: The analysis of data correlation serves two purposes: filtering data for the normal behaviour model and discovery of sensor faults. This data selection step is based on the assumption that the output of the model is correlated to one or more of its features. In this work we use the standard Pearson product-moment correlation coefficient of two variables. The proposed correlation analysis takes a defined subset of features and the model output and calculates its total correlation. Let $x_i^{(j,k)} = \{x_i^{(j)}, x_i^{(j+1)}, ..., x_i^{(j+k)}\}$ be a subset starting from sample $j \in N_0$ of size $k \in N_1$, where $k <= n$, of the i^{th} feature, and $y^{(j,k)} = \{y^{(j)}, y^{(j+1)}, ..., y^{(j+k)}\}$ is a variable that is the matching subset starting from sample j of size k of the output. The correlation Pearson product-moment correlation coefficient $corr(x_i^{(j,k)}, y^{(j,k)})$ is calculated as follows:

$$corr(x_i^{(j,k)}, y^{(j,k)}) = cov(x_i^{(j,k)}, y^{(j,k)})/(\delta x_i^{(j,k)} \delta y^{(j,k)}) \quad (1)$$

Where $cov(x_i^{(j,k)}, y^{(j,k)})$ is the covariance of variables x_i^k and y^k, and $\delta x_i^{(j,k)}$ and $\delta y^{(j,k)}$ are their respected standard deviations. The correlation calculated in equation 1 serves as a normality score for model data selection.

4) Model data selection: Samples of all features from the training set are evaluated based on the calculated correlation score. The proposed method allocates a sample $(x_i^{(j,k)} y^{(j,k)})$ into one of two groups: normal behaviour and suspicious behaviour. For a chosen $\alpha \in [0, 1]$, samples with $corr(x_i^{(j,k)}, y^{(j,k)}) > \alpha$ are allocated to normal behaviour group. If $corr(x_i^{(j,k)}, y^{(j,k)}) <= \alpha$ or if $corr(x_i^{(j,k)}, y^{(j,k)})$ does not exist, the samples are allocated to the suspicious behaviour group and are removed from the training set. Note that the correlation cannot be calculated if the standard deviation of $x_i^{(j,k)}$ or $y^{(j,k)}$ is zero. In this case the correlation is assigned a NA value. In sensor data this kind of feature can be observed for periods with long sensor failures.

5) Normalization: Vector normalization or scaling is usually performed before ANN model fitting. In this work normalisation was used. This is done to adjust values used for training by scaling them into the set $[0, 1]$. Large differences between values in the training set have an influence on the model weights which affects the model's ability to learn and aids generalization [17].

III. ANN MODEL CREATION

An artificial neural network (ANN) is a machine learning algorithm used to estimate unknown functions depending on several parameters. An ANN consists of interconnected neuronal nodes which perform simple calculations on outputs from neighbouring nodes in the previous layer. The result is passed to the next layer of the network.

We consider an ANN with $n \in N_1$ input variables $x = [x_0, x_1, x_2, ..., x_n]^\top$, where $x_n \in R$, and $x_0 = 1$ is a bias unit. The output variable of the considered ANN is $y \in R$. Let $a_i^{(j)}$ be the activation of neuron i in layer j, where $j \in 1, 2, .., l$, and l is the number of layers. $\Theta^{(j)}$ is a matrix of weights controlling the function mapping from layer j to layer $j + 1$. The considered hypothesis function approximated by the ANN is $h_\Theta(x) \in R$. Any layer L_j of the ANN consists of s_j neurons $a^{(j)} = [a_0^{(j)}, a_1^{(j)}, a_2^{(j)}, ..., a_{s_j}^{(j)}]^\top$. The size of the layer j can be different for every hidden layer. The input layer L_1 is of

size n, corresponding to the features vector. The output layer L_3 is of size 1 since the considered hypothesis function is $h_\Theta : R^n \rightarrow R$. The neural network architecture, including the number of inputs, outputs, layers and neurons in each layer, as well as the selection of the transfer function, describes an artificial neural network. Supervised learning methods for training ANN use the training examples $x_0, x_1, x_2, ..., x_n, y$ to calculate weight matrices $\Theta^{(1)}, \Theta^{(2)}, ..., \Theta^{(l-1)}$. The neural network architecture and the calculated weight matrices are jointly used for the approximation of an unknown function representing the relationship between input features and output variables. This way an artificial neural network can be trained to approximate transfer functions, especially unknown non-linear relationships.

A. ANN model training

The next step of data processing is the creation of an ANN model from the data by supervised training. The ANN training method chosen for training is called feed-forward training method (or forward propagation). Let's consider an ANN with $l \in N_1$ layers, $n \in N_1$ inputs, one output, a single set of model features $x = [x_0, x_1, x_2, ..., x_n]^\top$ and the output variable $y \in R$. Each layer L consists of S_l neurons. The neuron activation function is the sigmoid function, as defined in equation 2.

$$g(z) = 1/(1 + e^{-z}) \tag{2}$$

The Cyberenko theorem proves that the sigmoid function fulfills the universal approximation theorem which states that a single layer feed-forward artificial neural network can approximate continuous functions. The sigmoid activation function is therefore used to add non-linearity to the artificial neural network.

The forward propagation algorithm takes the vector x as an input and assigns it to the first layer $a^{(1)}$, therefore $a^{(1)} = x$. Neurons $a^{(2)}, a^{(3)}, ..., a^{(l)}$ can be constructed with the following vectorised equations: $\forall j \in [2, l] \quad a^{(j)} = g(\Theta^{(j-1)} a^{(j-1)})$.

The matrices $\Theta^{(1)}, \Theta^{(2)}, ..., \Theta^{(l-1)}$ are model weights. $\Theta^{(j)}$ is a matrix of weights controlling the function mapping from layer j to layer $j + 1$, for any $j \in (2, l)$, additionally $\Theta^{(j)} \in R^{s_{j+1} \times s_j + 1}$. Because the layer L_l is an output layer, $h_\Theta(x) = a^{(l)}$, therefore the hypothesis in the forward propagation algorithm is as follows:

$$h_\Theta(x) = g(\Theta^{(l-1)} g(\Theta^{(l-2)} ... g(\Theta^{(1)} x))) \tag{3}$$

Forward propagation takes the features $x_1, x_2, ..., x_n$ and modifies them with matrices $\Theta^{(1)}, \Theta^{(2)}, ..., \Theta^{(l-1)}$ and the sigmoid function g to create better suited features $a^{(1)}, a^{(2)}, ..., a^{(n)}$. In order to calculate the matrices $\Theta^{(1)}, \Theta^{(2)}, ..., \Theta^{(n-1)}$, the cost function J with least-squares fitting is described as follows:

$$J(\Theta) = \frac{1}{2n} \sum_{i=1}^{n} (h_\Theta(x^{(i)}) - y^{(i)})^2 \tag{4}$$

In feed-forward ANN, the problem of over-fitting can be solved with regularization [18] which is used to minimise the $\Theta^{(1)}, \Theta^{(2)}, ..., \Theta^{(l-1)}$ weights of the model. The cost function J with regularization is as follows:

$$J_R(\Theta) = \frac{1}{2n} \sum_{i=1}^{n} (h_\Theta(x^{(i)}) - y^{(i)})^2 + \frac{\lambda}{2n} \sum_{k=1}^{l-1} \sum_{i=1}^{S_l} \sum_{j=1}^{S_{l+1}} (\theta_{i,j}^{(k)})^2 \tag{5}$$

where S_l is a number of units without a bias unit in the layer, λ is a regularization parameter called weight decay, and $\theta_{i,j}^{(k)}$ is an element of the matrix $\Theta^{(k)}$. Ripley [18] suggests to use $\lambda = 10^{-4} - 10^{-2}$ as a regularization parameter for least-squares fitting.

By minimising the cost function $J_R(\Theta)$, the ANN model weights $\Theta^{(1)}, \Theta^{(2)}, ..., \Theta^{(l-1)}$ can be computed. The Broyden–Fletcher–Goldfarb–Shannon (BFGS) algorithm [19] is used for solving the unconstrained nonlinear optimization problem of minimising the cost function $J_R(\Theta)$. While the BFGS algorithm is not guaranteed to converge, the Hessian matrix can be inspected in order to check if a secure local minimum has been found. There are many solutions to the optimisation problem and the weights are initialised at random at the start of the process, therefore the results might differ.

IV. PV MODEL ENSEMBLE GENERATION

The DER modeling process presented in section II-B is applied to a data set from a single PV inverter at the SYSLAB laboratory at the Technical University of Denmark. Two different models, a meteorological and a neighbourhood model, are presented in sections IV-C and IV-D and used in the EM-AD (section V).

A. Data sources

The data used for this study has been recorded from a 10kWp PV array located in Risø, Denmark in October 2014. The active power consumption data is recorded from the inverter in 1 second intervals. Meteorological data at the same time resolution - irradiation, temperature, wind speed and direction - is obtained from a meteorology mast about 600 m away from the PV site.

The data cleaning and preparation procedure presented in section II-B is used to pre-process the data before the ANN model can be trained. Since all presented models use the same data for training, the process of data cleaning is identical.

B. Data preparation

In this paper we focus on single-source instance problems removing discovered sensor errors and multi-source instance level problems considering inconsistent timing. Easily observed sensor failures result in missing data or false measurements. In the considered data set the observed false measurements were either constant values or inconsistent values, for example negative solar irradiation. Inconsistent value errors were present in the data set due to a sensor logging error. The threshold between consistent and inconsistent values is determined manually and inconsistent values are removed with the first filtering step.

TABLE I. DATA PROPERTIES OF SOLAR IRRADIATION (12.10.2014).

Data	Min.	1st Q	Median	3rd Q	Max.	NA's
Recorded	-31.450	-0.001	-0.001	0.014	31.450	0
Filtered	-0.001	-0.001	-0.001	0.015	0.087	4543

In the second filtering step, a 3rd order Butterworth low pass filter has been used to remove high frequency components appearing in the data due to sensor errors. The properties of the recorded and filtered data is presented in table I. Both the first and the third quadrant did not change significantly after the filtering process. The median remained the same, therefore it can be concluded that mostly outliers have been removed from the data set. This filtering process removed around 5% of the irradiation data set. Once the data is clean and uniform it can be aggregated to 1-minute values. The resulting time-series is then randomly divided into 3 sets: a training set D_t of size 14841, a validation set D_v of size 14901, and a cross-validation set D_{cv} of size 14898 samples.

C. Meteorological model

The meteorological model is based on the assumption that meteorological data can be used to model the yield of a PV panel. The available meteorological data (solar irradiation, wind speed, wind direction, ambient temperature) was used to construct the PV production model. The input significance analysis of the linear model based on the same inputs and outputs as the presented model, using test statistics under the null hypothesis, shows that all inputs are significant.

1) Model data selection with correlation analysis: The correlation between irradiation and yield data from the training set D_t was calculated using equation 1 with $k = 44640$ corresponding to a single day (see figure 5). The data set was extended with the normality score which is equal to the daily correlation. All data points with a score larger or equal to 0.2 were included in the normal behaviour model.

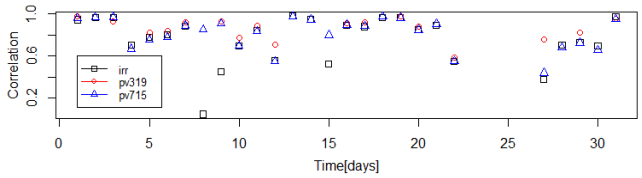

Fig. 5. Correlation analysis for irradiation, active power production from PV319 and PV715 with active power from PV117.

All data points for which the correlation could not be calculated due to zero standard deviation, were excluded from the training set. Based on correlation analysis, training data was selected for the ANN meteorological model, excluding data recorded on the 8th and 23-26th of October (figure6).

From the initial size of 14841, correlation analysis decreases the size of the training set D_t to 12480 rows. After excluding all rows where any of the data points is NA (removed by data cleaning), the size of the actual training set becomes 11739 rows. In the next step of the data preparation process, the training set is normalized.

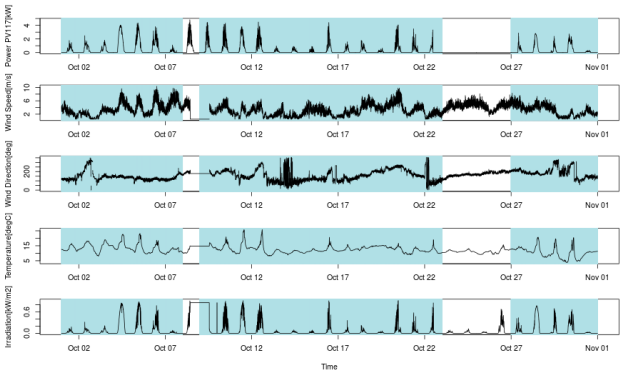

Fig. 6. Meteorological model normal training data set.

2) Model training: The proposed ANN network consists of 5 input neurons (representing solar irradiation, wind speed, wind direction, ambient temperature and time of day), one hidden layer with 10 neurons and a bias unit. The network has a single output neuron (PV117 power production) and 71 weights. The used transfer function g is sigmoid (as in equation 2) and the regularization parameter λ is set to 0.0006. The package *nnet* (Feed-Forward Neural Networks and Multinomial Log-Linear Models) [18] for the R scripting language has been used for the creation of the supervised learning ANN model. The *nnet* package calculates the Θ parameters of a single-hidden layer neural network, as described in section III-A. The model output for each input is presented in figure

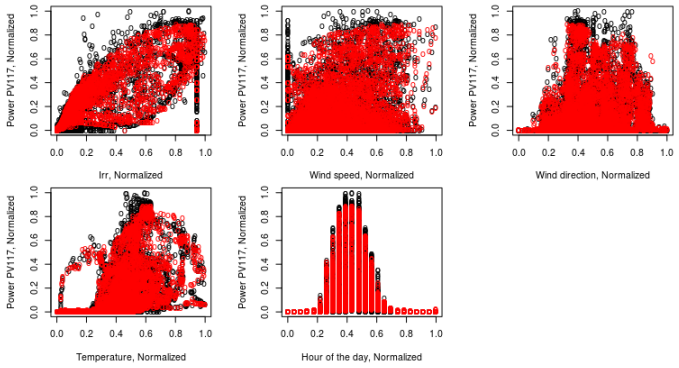

Fig. 7. Actual and predicted PV117 power consumption mapped to the inputs of the ANN meteorological model.

7. The root mean square error (RMSE) is calculated in order to evaluate the model. The RMSE for the training set is 1.13 versus 1.116 for the validation set and 1.118 for the cross-validation set. The small difference in RMSE between the validation and cross-validation sets indicates that the model generalises well.

D. Neighbourhood model

The proposed model uses data from two neighboring PV systems to take advantage of correlations between the three systems. In the SYSLAB laboratory the distance between

PV117 and PV715 is $630m$, compared to $340m$ between PV117 and PV319. In this investigation PV117 is being modelled. The correlation between the active power productions of PV117, PV319 and PV715 is calculated by using data from the training set D_t and equation 1 with $k = 44640$, corresponding to a single day (see figure 5). All days with a correlation larger

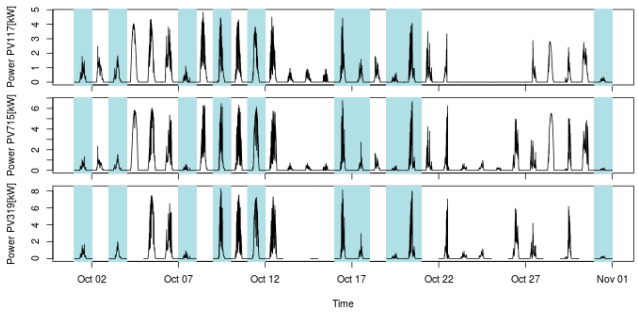

Fig. 8. Neighbourhood model training data with days selected by the correlation analysis in blue.

or equal to 0.2 have been included in the model. In this case, the 2nd, 4th, 8th, 13-15th, 21st, 23-26th and 30th of October have been removed from the training set. After the correlation analysis, the size of the training set D_t decreased from 14841 to 8201 samples.

The proposed ANN network consists of 3 input neurons (PV715 and PV319 power production and time of day), one hidden layer with 8 neurons and a bias unit. The network has a single output neuron (PV117 power production). The used transfer function g is sigmoid (as in equation 2) and the regularization parameter λ is 0.0006. Similarly to the meteorological ANN model, the R package *nnet* is used to find the Θ parameters of the model, as described in section III-A. The output of the model compared to its input with use of data from the validation set is presented in figure 9. The RMSE for the training set is 0.96, versus 0.570 for the

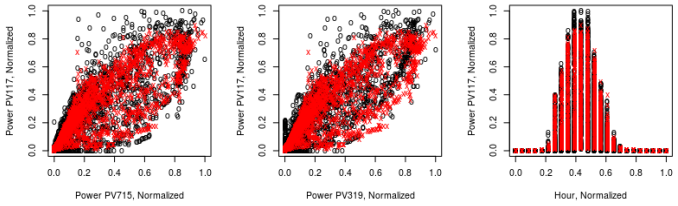

Fig. 9. Actual and predicted PV117 power consumption mapped to the model inputs for ANN neighbourhood model.

validation set and 0.572 for the cross-validation set. The small difference in RMSE values between the validation and cross-validation sets indicates that the model generalises well.

V. EM-AD for a PV plant

The architecture of the EM-AD is presented in figure 10. Sensor data of solar irradiation, wind speed, wind direction, ambient temperature, hour of day and power consumption of two neighbouring PVs (PV319 and PV715) are used as input. The proposed ensemble regression is composed of two regression models. The models were generated from disjoint parameter sets and a contextual parameter (hour of day), creating redundant heterogeneous ANN regression models of active power production as presented in sections IV-C and IV-D. The ensemble model set was not pruned because the set contains only two models. The ensemble integration is usually calculated as a linear combination of the predictions [14]. Here the ensemble power prediction P' is calculated from predictions for each model P^N and P^M as follows: $P' = \alpha P^N + \alpha P^M$, where $\alpha = 1/2$ corresponds to equal weight averaging.

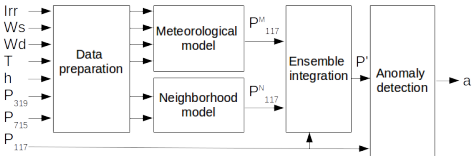

Fig. 10. Architecture of the proposed PV ensemble regression model anomaly detection (EM-AD)

The ensemble prediction P' is weighted with the anomaly score in the anomaly evaluation component. The anomaly score is based on the correlation analysis for both ANN models as presented in figure 5. Partial anomaly scores a_M and a_N are calculated for both models as in equation 6.

$$a = \begin{cases} 1 & corr \geq 0.2 \\ 10 & corr < 0.2 \end{cases} \tag{6}$$

The anomaly score a_s combines the partial scores for the models a_M and a_N and is calculated as $a_s = 1/(a_M \cdot a_N)$. The anomaly score a_s is multiplied by the difference between the ensemble prediction P' and measured power P to calculate the anomaly $a = a_s \cdot (P' - P)$. The chosen anomaly threshold is $\epsilon = 0.1$, therefore only observations with $a > \epsilon$ are considered.

VI. Results

In the considered scenario one month of the historical active power production of a single PV plant is analysed. In the analysed period of time the PV should have not been controlled, the considered anomalous cyber event is curtailment of the PV active power production to zero. The cyber event is being discovered by the proposed EM-AD, in this section we present the anomaly detection results for the EM-AD and compare it to other anomaly detection techniques.

The degree of agreement between the ensemble predictions is given by their overall spread ($s = P^N - P^M$) with the first and third quadrant at -0.011 and 0.006, respectively, and a standard deviation of 0.511. This indicates that the models generally agree in their predictions. Table II presents nine approaches for model-based anomaly detection which were performed using the October 2014 PV data set. The evaluated models are M (meteorological), N (neighbourhood), MN (joint model with inputs from M and N), EMN (ensemble of M and N). The used training sets are: cor (correlated days for the data set), full (entire data set). Two anomaly detection

TABLE II. RESULTS CONFUSION MATRIX AND STATISTICAL MEASURES

Model	Train	AD	TP	TN	FP	FN	ACC	PPV	NPV	FNR	TPR	TNR	FPR	FDR
M	full	M-AD	1198	37445	4531	1466	0.866	0.209	0.962	0.550	0.450	0.892	0.108	0.791
M	cor	M-AD	1633	41305	671	1031	0.962	0.709	0.976	0.387	0.613	0.984	0.016	0.291
N	full	M-AD	1543	39242	2734	1121	0.914	0.361	0.972	0.421	0.579	0.935	0.065	0.639
N	cor	M-AD	1736	41660	316	928	0.972	0.846	0.978	0.348	0.652	0.992	0.008	0.154
MN	full	M-AD	764	38723	3253	1900	0.885	0.190	0.953	0.713	0.287	0.923	0.077	0.810
MN	cor	M-AD	764	41756	220	1900	0.953	0.776	0.956	0.713	0.287	0.995	0.005	0.224
EMN	full	M-AD	1447	38730	3246	1217	0.900	0.308	0.970	0.457	0.543	0.923	0.077	0.692
EMN	cor	M-AD	1709	38614	3362	955	0.903	0.337	0.976	0.358	0.642	0.920	0.080	0.663
EMN	cor	EM-AD	1709	41880	96	955	0.976	0.947	0.978	0.358	0.642	0.998	0.002	0.053

methods are used: M-AD (model based anomaly detection) and EM-AD (ensemble regression model anomaly detection). The confusion matrix is a compilation of instances of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), evaluated from a population of results. To measure the correctness of the anomaly detection we calculate eight significant measures: accuracy (ACC), precision (PPV), negative predictive value (NPV), false negative rate (FNR), sensitivity (TPR), specificity (TNR), false positive rate (FPR) and false discovery rate (FDR).

The proposed EM-AD with the correlation training selection approach achieves an accuracy of 0.976, which improves the accuracy by 0.4-11.1% for single model AD, 2.3-9.2% for joint model AD, and 7.3-7.6% over the method without correlation ensemble integration. The precision of the proposed method is 0.947, which improves the precision by 23.8-73.8% for single model AD, 10.1-58,6% for joint model AD, and 61-63.8% over the method without correlation ensemble integration. EM-AD with correlation training data selection additionally keeps low values for FNR of 0.358, FPR of 0.002 and FDR of 0.053. While the specificity has improved only by 10.6% at best, totalling to 0.998, the sensitivity is 0.642 which presents an improvement of up to 35.5% over other presented methods.

VII. CONCLUSION

This paper proposes a novel ensemble model anomaly detection method with non-linear regression models and anomaly scores based on correlation analysis (RM-AD) used for cyber-physical intrusion detection in smart grids. The models are presented and evaluated and the ensemble integration and anomaly detection methods are described in detail. A proof-of-concept RM-AD analysing a data set from a PV plant is presented and compared to other M-AD approaches. Future work will include automatic ensemble model set generation, an investigation into whether a larger ensemble can improve the prediction accuracy, and alternative interpolation methods for missing data.

This research was conducted as part of the SALVAGE project (Cyber-physical security for low-voltage grids) funded by ERA-Net Smart Grids.

REFERENCES

[1] "Security Guideline for the Electricity Sector: Physical Security," North American Electric Reliability Corporation(NERC):Critical Infrastructure Protection Committee, Tech. Rep., June 2012.

[2] A. McIntyre, "Renewable systems interconnection study: Cyber security analysis," Sandia Natl. Lab., Tech. Rep., February 2008.

[3] J. Moteff, "Risk management and critical infrastructure protection: Assessing, integrating, and managing threats, vulnerabilities and consequences," in *Library of Congress Washington DC Congressional Research Service*. DTIC Document, 2005.

[4] G. N. Ericsson, "Cyber security and power system communication — essential parts of a smart grid infrastructure," *Power Delivery, IEEE Trans.*, vol. 25, no. 3, pp. 1501–1507, 2010.

[5] F. M. Cleveland, "Cyber security issues for advanced metering infrastructure (AMI)," in *PES GM - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*. IEEE, 2008, pp. 1–5.

[6] D. Yang, A. Usynin, and J. W. Hines, "Anomaly-based intrusion detection for SCADA systems," in *5th Intl. Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies*, 2006, pp. 12–16.

[7] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proc. of the IEEE*, vol. 100, no. 1, pp. 210–224, 2012.

[8] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[9] M. Sanz-Bobi, A. M. San Roque, A. de Marcos, and M. Bada, "Intelligent system for a remote diagnosis of a photovoltaic solar power plant," in *Journal of Physics: Conference Series*, vol. 364, no. 1. IOP Publishing, 2012, p. 012119.

[10] A. Zaher, S. McArthur, D. Infield, and Y. Patel, "Online wind turbine fault detection through automated SCADA data analysis," *Wind Energy*, vol. 12, no. 6, p. 574, 2009.

[11] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM symposium on information, computer and communications security*. ACM, 2011, pp. 355–366.

[12] A. M. Kosek, "Contextual anomaly detection for cyber-physical security in Smart Grids based on an artificial neural network model," in *Joint Workshop on Cyber-Physical Security and Resilience in Smart Grids (CPSR-SG2016), CPSweek 2016*, 2016.

[13] SALVAGE project - Cyber-physical security for low-voltage grids. http://salvage-project.com. Accessed: 2016-01-03.

[14] J. Mendes-Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: A survey," *ACM Comput. Surv.*, vol. 45, no. 1, pp. 10:1–10:40, Dec. 2012.

[15] N. García-Pedrajas, C. Hervás-Martínez, and D. Ortiz-Boyer, "Cooperative coevolution of artificial neural network ensembles for pattern classification," *Evolutionary Computation, IEEE Transactions on*, vol. 9, no. 3, pp. 271–302, 2005.

[16] E. Rahm and H. H. Do, "Data cleaning: Problems and current approaches," *IEEE Data Eng. Bull.*, vol. 23, no. 4, pp. 3–13, 2000.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[18] B. D. Ripley, *Modern applied statistics with S*. Springer, 2002.

[19] J. D. Head and M. C. Zerner, "A Broyden— Fletcher —Goldfarb —Shanno optimization procedure for molecular geometries," *Chemical physics letters*, vol. 122, no. 3, pp. 264–270, 1985.