REACTIVE POWER FORECASTING AT THE TRANSMISSION-DISTRIBUTION INTERFACES USING PHYSICS-BASED MACHINE LEARNING

Arnaud Rosseel
UMons - Belgium
arnaud.rosseel@umons.ac.be

Bashir Bakhshideh Zad
UMons – Belgium
bashir.bakhshidehzad@umons.ac.be

Zacharie De Grève
UMons - Belgium
zacharie.degreve@umons.ac.be

François Vallée
UMons - Belgium
francois.vallee@umons.ac.be

ABSTRACT

The power exchanges at the Transmission-Distribution (T-D) interfaces have a high implication for both Transmission System Operator (TSO) and Distribution System Operator (DSO). Historically, simple models were sufficient to characterize the behaviour of distribution networks and to predict power flows at T-D interfaces. With the growing share of distributed energy resources, active and reactive power flows find increasingly volatile behaviour such that the previous models would no longer remain efficient. This paper proposes a physics-informed Machine Learning model to improve the prediction of the power exchanges at T-D interfaces. The proposed model relies on the developed Inverse Load Flow (ILF) formulation that aims to bring additional physical insights to the pure data-driven approach. The ILF determines the equivalent model of the distribution network based on which the power exchanges at T-D interfaces can be predicted. The proposed ILF-based model is benchmarked against the classical Machine Learning methods on a synthetic distribution network. Simulation results confirm the superior accuracy of the proposed ILF-based model.

INTRODUCTION

In power systems, voltage levels are strongly linked to reactive power [1,2]. In order to ensure a safe operation of the power system and to avoid the excessive voltage fluctuations that could damage power system components, reactive power needs to be appropriately managed. For this purpose, reactive power exchanges at the Transmission-Distribution (T-D) interfaces are of a great importance for both Transmission System Operator (TSO) and Distribution System Operator (DSO). The TSO requires an accurate knowledge of reactive power exchanges at the T-D interfaces to ensure an effective voltage control in a short-term perspective and to be able to optimally invest on required assets for an efficient volt-var control in long term. From the DSO side, they are required to keep their reactive power exchanges within the predefined range to avoid eventual penalties and extra costs [3].

Prediction of reactive power exchanges at the T-D interfaces has nevertheless become increasingly complicated in recent years. Firstly, due to increasing penetration of renewable generations and their volatile nature, the active power flows are continuously changing within the power system that would influence the reactive power flows as well. In addition, the reactive power exchanges at the T-D interfaces have increasingly found a capacitive behaviour due to mainly new types of loads such as electric vehicles, heat pumps, etc. as well as the replacement of overhead lines with the underground cables at the distribution level [4,5]. Besides, the conventional power plants providing traditionally voltage control services are being phased out, which renders the voltage control task in transmission grids more complicated. In this regard, it becomes of a strategic importance for the TSO to better understand and predict the future reactive power behaviour at the T-D interfaces. However, the lack of detailed information available to the TSO on the topology of distribution networks makes the prediction of these reactive power exchanges extremely challenging [6], thereby hindering the TSO from scheduling the most appropriate voltage control strategies. Whether in the short-term operational management or in the long term for cost-effective investments, good accuracy reactive power predictions are needed.

Despite the importance of this topic and its underlying technical and scientific challenges, the literature on reactive power prediction at T-D interfaces is currently rather scarce. Overall, two categories of research can be identified in the literature. The first group relies on the model-based simulations of the distribution network considering the historical data [4]. It aims at predicting the future trends of reactive powers and obtaining the PQ behaviour at T-D interfaces. The application of the latter approach is however constrained by the lack of a complete and up-to-date distribution network topology and model. The second category employs the data-driven techniques and machine learning methods in the form of clustering and regression analyses in order to predict reactive power flows at T-D interfaces [7,8]. The latter techniques can provide acceptable results given that sufficient representative data would be available for the forecasting task. The main drawback of the second category is that they consist in black boxes and their provided results may not be physically interpretable.

In the current paper, a novel hybrid method combining data-driven and physics-based approaches is proposed. It takes advantage of the prediction power of the machine learning algorithms and the interpretability of the physic-based equivalent model of the distribution system in order to accurately forecast the reactive power exchanges at T-D interfaces. The proposed method tries to find an equivalent model of the distribution network seen from the transmission system based on the available historical data to the TSO. This equivalent model consists of a resistance and reactance (R^{eq} and X^{eq}) obtained via the developed Inverse Load Flow computation that can reproduce the active and reactive power losses in distribution system. Therefore, relying on an appropriate equivalent model of the considered distribution network, our proposed hybrid approach can map the predicted active and reactive load demands (P^d and Q^d) at the DSO side to the active and reactive power exchanges at T-D interface (P^g and Q^g) using the employed supervised learning algorithms. The performance of the proposed hybrid method is tested using Linear Regression (LR), K-Nearest Neighbours (KNN) and Random Forest (RF). It is benchmarked against the purely data-driven approaches from the literature.

This paper is organized as follows. Second Section describes the methodology in which, Inverse Load Flow is presented. Use Case is defined in the third Section as well as underlying assumptions. The fourth section presents the most important result obtained. Final Section discusses the results and concludes the presented paper.

METHODOLOGY

The proposed methodology of this paper consists of the developed Inverse Load Flow (ILF) formulation coupled with the employed Machine Learning (ML) algorithms, presented below.

ILF for Distribution System Modelling

The ILF formulation aims to model the entire distribution network with an equivalent line relying on the aggregated load demand of the system and the power exchanges at the T-D interface. The proposed ILF formulation is developed based on the power flow equations in hybrid form [9]:

$$P_i^g - P_i^d = \sum_{\langle i,j\rangle \in \mathcal{L}} P_{ij} \qquad \forall i \in \mathcal{B} \tag{1}$$

$$Q_i^g - Q_i^d = \sum_{\langle i,j\rangle \in \mathcal{L}} Q_{ij} \qquad \forall i \in \mathcal{B} \tag{2}$$

$$P_{ij} = g_{ij}v_i^2 - g_{ij}v_iv_j\cos(\theta_i - \theta_j) - b_{ij}v_iv_j\sin(\theta_i - \theta_j) \qquad \forall\langle i,j\rangle \in \mathcal{L} \tag{3}$$

$$Q_{ij} = -b_{ij}v_i^2 + b_{ij}v_iv_j\cos(\theta_i - \theta_j) - g_{ij}v_iv_j sin(\theta_i - \theta_j) \qquad \forall\langle i,j\rangle \in \mathcal{L} \tag{4}$$

where:

- P_i^g, Q_i^g, P_i^d and Q_i^d are active power and reactive power generation as well as active power and reactive power demand at node i, respectively;
- P_{ij} and Q_{ij} stand for active and reactive power flow of line $\langle i,j\rangle$;
- g_{ij} and b_{ij} are respectively the conductance and the susceptance of line $\langle i,j\rangle$;
- v_i and θ_i give the voltage magnitude and angle of node i;
- $\mathcal{B}$ and $\mathcal{L}$ represent the sets of buses and lines in the considered network.

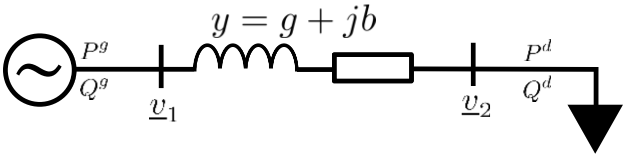

Figure 1 – 1-Line Network for Inverse Load Flow

Fig. 1 shows a 1-Line network with 2 buses. As it can be seen, P^g and Q^g are the net active and reactive powers at the first bus while P^d and Q^d are active and reactive power demand at the second bus. Each bus has a nodal voltage $\underline{v}$ and the two buses are connected by a line with a conductance g and a susceptance b.

By simplifying equations (1)-(4) for the 1-Line network shown in Fig. 1, the system of equations (5)-(8) is obtained. In the latter, there exists 4 non-linear equations with 6 unknown parameters namely, $g, b, v_1, \theta_1, v_2, \theta_2$. In order to have a solvable system of equations, it is assumed that the node 1 is the slack node with $v_1 = 1\,pu$ and $\theta_1 = 0$. It leads to the following system of equations having 4 equations with 4 unknown parameters:

$$P^g = g - gv_2\cos(\theta_2) + bv_2\sin(\theta_2) \tag{5}$$

$$Q^g = -b + bv_2\cos(\theta_2) + gv_2\sin(\theta_2) \tag{6}$$

$$P^d = gv_2^2 - gv_2\cos(\theta_2) - bv_2\sin(\theta_2) \tag{7}$$

$$Q^d = -bv_2^2 + bv_2\cos(\theta_2) - gv_2\sin(\theta_2) \tag{8}$$

The ILF formulation (5)-(8) receives the 4 input parameters P^g, Q^g, P^d, Q^d and determines the line parameter b and g. The ILF formulation is adopted to obtain the equivalent line that can represent the link between P^g, Q^g (the power exchange at T-D interface) and P^d, Q^d (the aggregated load demand of studied distribution network). The obtained equivalent line (b and g) is then converted to resistance and reactance R^{eq}, X^{eq} as follows. Fig. 2 illustrates the mapping between the inputs and outputs of the ILF formulations

$$Z = \frac{1}{Y} = \frac{1}{g + bj} = R^{eq} + jX^{eq} \tag{9}$$

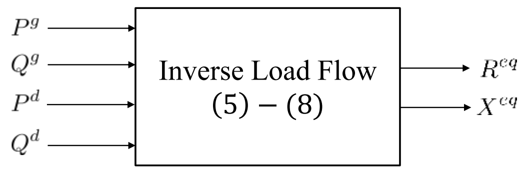

Figure 2 – Representation of Inverse Load Flow inputs and outputs

Machine Learning Algorithms

Machine Learning (ML) is generally employed in order to extract the hidden links and patterns from the available data without any explicit knowledge on the mathematical relations governing the system at hand. In this paper, we combine the exploration capability of machine learning with the insights from the system physics (obtained through the equivalent model) to improve the performance of the prediction task. To this end, the supervised learning as a regression task is formulated to predict R^{eq} and X^{eq} for the eventual active and reactive power predictions at T-D interface. Three supervised learning algorithms have been considered in this work namely, Linear Regression, K-Nearest Neighbors and Random Forest.

To simply explain ML, a generic nomenclature is described below. With $\boldsymbol{X} = \{X^1, \dots, X^n\}$, the independent variables vector of n dimensions and y, the target such as:

$$Dataset = \{(\boldsymbol{X_1}, y_1), (\boldsymbol{X}_2, y_2), \dots, (\boldsymbol{X_m}, y_m)\} \qquad (10)$$

Where $\boldsymbol{X} \in R^n$ and $y \in R$. The ML regression goal is to get $\hat{y}$, the predicted target as a function of $\boldsymbol{X}^{new}$ as close as possible of y^{new}. In other words, regression can be seen as an optimization problem where error $\epsilon = f(\hat{y}, y^{new})$ is minimized.

Linear Regression (LR)

LR is a simple supervised learning algorithm. It consists in a linear function that minimizes least squares error as follows:

$$f(w_1, \dots, w_n, b) = y = \boldsymbol{w}.\boldsymbol{X} + b + \epsilon \qquad (11)$$

Where $\boldsymbol{w}$ represents the regression coefficient vector and b is the intercept. As explained, linear regression is an error minimization:

$$Min\left(\sum_{i=1}^{m}(y_i - \hat{y}_i)^2\right) = \sum_{i=1}^{m}\left(y_i - (\boldsymbol{w}.\boldsymbol{X}_i + b)\right)^2 \qquad (12)$$

All the parameters $\boldsymbol{w}$ and b are computed via this minimization and allow to find the best coefficient vector according to the dataset. $\hat{y}$ is then predicted by replacing $\boldsymbol{X}$ by $\boldsymbol{X^{new}}$ in the linear equation.

K Nearest Neighbors (KNN)

KNN is a non-parametric, instance-based learning method that finds the k training examples closest to the new observation according to a distance metric (e.g., Euclidian distance [10]). This latter is calculated between new input $\boldsymbol{X^{new}}$ and all input dataset $\boldsymbol{X}$ with equation above:

$$d_j = \sqrt{\sum_{i=1}^{n}\left|X_i^j - X_i^{new}\right|^2} \qquad (13)$$

The average of the corresponding y to the k nearest $\boldsymbol{X}$ is then calculated.

Random Forest (RF)

A decision tree is a model of decisions and their possible consequences. Each internal node represents a feature, each branch represents a decision based on that feature, and each leaf node represents a predicted outcome. Random Forest is an ensemble model made of several decision trees. RF is built by training various decision trees in parallel on different subsets of the training dataset [11]. The final prediction of the RF is made by averaging all decision trees forecasts.

PERFORMANCE EVALUATION

The overall framework to evaluate the performance of the developed approach is shown in Fig. 3. It consists of 3 main stages that are described below.

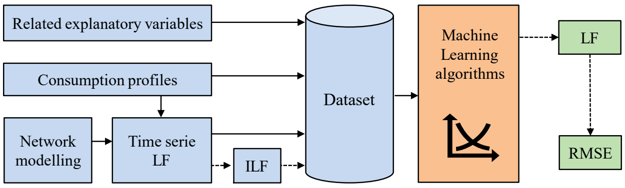

Figure 3 – Overall representation of the proposed ILF-based methodology for power forecasting at T-D interface

The first stage (illustrated in blue) aims to create the dataset from network model and Synthetic Load Profiles (SLPs) that will be introduced in next section. Then, created time series power flows are computed to get P^{net} and Q^{net} which represent active and reactive power at the T-D interface. The ILF is used to determine equivalent resistance (R^{eq}) and reactance (X^{eq}) based on the T-D interface and aggregated powers.

The second stage (shown in orange) focuses on the forecasting task with the three employed ML algorithms previously described. In this step, ML algorithms are trained to extract relations and patterns present between input features (aggregated demand, weather and calendar data) and equivalent resistance/reactance of the network. Once the training is completed, in the test phase, the trained ML algorithms predict R^{eq} and X^{eq} for new (unseen) observations (inputs). The power exchanges at T-D interface (P^g and Q^g) are eventually computed by Load Flow (LF) calculation considering the predicted line parameters and the corresponding aggregated demand.

Finally, in the last stage (shown in green), the predicted power exchanges at T-D interface are analyzed and evaluated through the relative root mean square error (rRMSE) index described below.

Stage 1: Database Creation

The proposed methodology requires access to active and reactive power exchanges at T-D interface. In practice, these measurements are not always available. Therefore, a synthetic distribution network and synthetic load profiles are considered to construct the required data for subsequent forecasting task.

Synthetic Distribution Network

The considered distribution network includes 4 feeders and 7 nodes as shown in Fig. 4. It is a 20 kV distribution network. Each node of the system represents a LV network that is modelled as a load (of residential or industrial type).

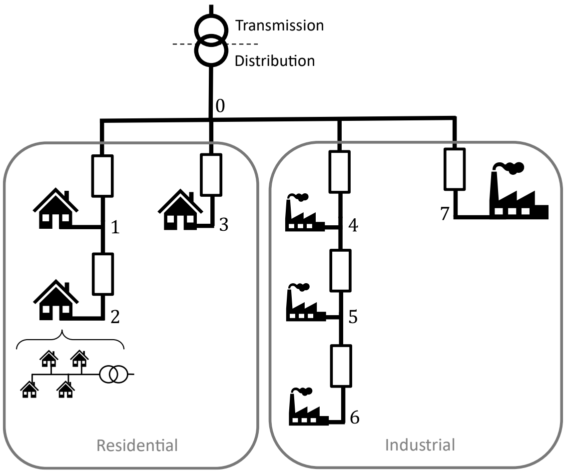

Figure 4 – Studied 4-Feeder Distribution Network

For the sake of simplicity, all the system lines have been modelled with identical parameters (given in Table 1) while having different lengths. The first two feeders are related to urban residential load type while the two last ones represent industrial customers.

Table 1- Parameters of studied network

r [Ω/km]	x [Ω/km]	c [nF/km]	i_{max}[kA]
0.1188	0.32	11.0	0.645

Synthetic Load Profiles

SLPs are load profiles calculated using aggregated load databases and external variables such as weather and calendar data. A SLP is a load profile whose summation over all the time steps (quarter-hourly) over a year equals to one:

$$\sum_{t=1}^{\mathcal{T}} P_{t,\%} = 1 \qquad (14)$$

Two types of SLPs are collected from Synergrid [12], namely residential and industrial profiles. The former is used for the 3 first nodes of the system while the latter is applied to the 4 last nodes. In order to construct SLP for each node, it is necessary to first estimate the annual consumption of each node (C_{tot}). This is then multiplied by the SLP corresponding to the node type. The result is a time series representing the annual consumption of a LV network.

$$P_t = P_{t,\%} * C_{tot} \quad \forall t \in \mathcal{T} \qquad (14)$$

SLPs provide the active power load demands. The reactive power load demands are also required as an input of the proposed model. In this regard, for each time step and for each load, a Power Factor (PF) is randomly generated between 0.8 and 1 using which the reactive power of the load is calculated as follows.

$$Q_t = P_t * \sqrt{\left(\frac{1}{PF^2}\right) - 1} \; \forall t \in \mathcal{T} \qquad (15)$$

Load Flow and Inverse Load Flow calculations

In order to obtain the active and reactive power exchanges at T-D interface, time series power flow calculations are performed on Python (PandaPower library [13]) for each time step considering the studied synthetic network and the created load demands (the SLPs). Finally, the Inverse Load Flow calculation is conducted for each time step in order to compute R^{eq} and X^{eq} required for the forecasting task via employed ML algorithms.

Stage 2: Forecasting

First, database is divided into a training set (80%) and a test set (20%). Then, for each algorithm with hyperparameters, these are optimized through a grid search combined with a 5-fold cross validation.

For each method, explanatory variables used were:

- aggregated load: P_{agg}^{L} and Q_{agg}^{L}, respectively active and reactive aggregated network load;
- weather data: Temperature, heating inertia and low cloud cover;
- calendar data: Hour of the day and day of the week.

Stage 3: Performance Evaluation

Several KPIs can be used to analyse the forecasting performance of the proposed framework. Dealing with a regression task, the relative Root Mean Square Error (rRMSE) is adopted:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad rRMSE = \frac{RMSE}{\sigma} \qquad (16)$$

where y_i is the i^{th} target, $\hat{y}_i$ the prediction of the i^{th} target, σ the standard deviation of y and n, the number of time steps of y.

Benchmark: Classical pure data-driven Prediction

The proposed physics-based machine learning model is benchmarked against the classical data-driven prediction methods (coming from literature). The latter is called hereafter Direct Prediction (DP), which infers that it does not include the ILF analyses, and it directly maps the aggregated load demands to the power exchanges at T-D interface. The framework for performance evaluation of DP model can be easily implemented by eliminating both ILF and LF steps shown in Fig. 3.

RESULTS

Prior to presenting the simulation results, the relationships between aggregated load demands and power exchanges at T-D interface are illustrated in Fig .5. While a strong linear correlation between P_{net} and P_{agg}^{L} is observed, Q_{net} finds wider variations due to the assumption adopted on the power factor range (varying between 0.8 and 1).

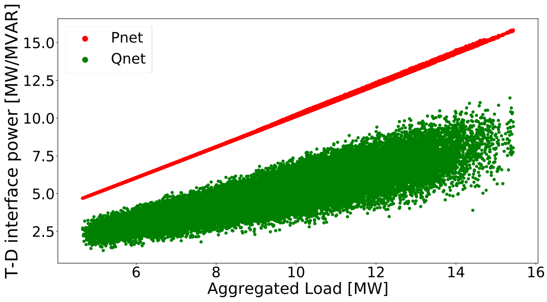

Figure 5- Illustration of active and reactive power exchanges at T-D interface in function of the aggregated load demand P_{agg}^{L}

Table 2 presents the relative RMSE obtained for the prediction of P_{net} and Q_{net} using the proposed ILF-based model and the classical DP by 3 employed ML algorithms namely, LR, KNN and RF.

Table 2 – rRMSE of power predictions at T-D interface obtained by the proposed ILF-based model and the classical DP using different ML algorithms

	P_{net} [%]		Q_{net} [%]	
	DP	ILF	DP	ILF
LR	1.710	1.508	6.545	5.870
KNN	13.464	0.921	20.628	3.565
RF	2.303	1.055	6.410	4.038

Considering the results of Table 2, it can be noticed that the proposed ILF-based model outperforms the classical DP for both active and reactive power predictions and using the three employed ML algorithms. The accuracy of the proposed model is rooted in its physics-based nature that simplifies the prediction task by going through the inverse load flow analysis. As a result, the prediction space is reduced to the mapping between the aggregated load demand and the parameters of the equivalent model in contrast to the classical direct prediction that maps the aggregated demand to the power exchanges at T-D interface.

In Table 2, it is also observed that regardless of the prediction model (ILF-based or DP), the reactive power predictions lead to larger RMSEs as they present more nonlinear characteristics with respect to active power (see Fig. 5).

CONCLUSION

This work proposes a novel physics-based machine learning model to predict the active and reactive power exchanges at T-D interface. The proposed model relies on an inverse load flow calculation to determine an equivalent resistance and reactance that can link the aggregated load demands to the power exchanges at the T-D interface. By leveraging the insights from the equivalent model, it is shown that the proposed physics-based model can improve the accuracy of the classical data-driven prediction methods found in literature.

The current paper presents a proof of concept of the proposed physics-based machine learning model for forecasting active and reactive powers at T-D interface. The future work will focus to add PV production with Synthetic Production Profiles. In addition, the proposed model needs to be upscaled considering real-life distribution grids.

REFERENCES

[1] T. Cutsem, C. Vournas, 1998, *Voltage Stability of Electric Power Systems,* Springer New York, NY, US.
[2] P. Kundur et al., 2004, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions", *IEEE TPWRS.* vol. 19, no.3, pp. 1387-1401
[3] ENTSO-E , Aug. 2022, ENTSO-E Network Code on Demand Connection, [online], available: https://www.entsoe.eu/network_codes/dcc/
[4] C. G. Kaloudas et al., 2017, "Assessing the Future Trends of Reactive Power Demand of Distribution Networks", *IEEE TPWRS.* vol. 32, no. 6, pp. 4278-4288
[5] M. Tasky, M. Regula, A. Otcenasova, 2021, "Impact of Changes in a Distribution Network Nature on the Capacitive Reactive Power Flow in the Transmission Network in Slovakia", *Energies.* Vol.14, 5321
[6] A. G. Givisiez, K. Petrou, L. F. Ochoa, 2020, "A Review on TSO-DSO Coordination Models and Solution Techniques", *Electric Power Systems Research,* vol.189
[7] P.-D. Dapoz et al., 2019, "Mixed Regression Clustering Techniques for the Medium Term Prediction of Reactive Power within Transmission Grids", *ISGT-Europe,* pp. 1-5
[8] E. F. Arruda et al., 2015, « Disaggregated active and reactive demand forecasting using first difference measured data and neural networks", *Proceedings ICED*
[9] C. Coffrin, P. Van Hentenryck, 2014, "A Linear-Programming Approximation of AC Power Flows", INFORMS Journal on Computing, Vol.26(4), pp.718-734.
[10] E. L. Codjo et al., 2021, "Machine Learning-Based Classification of Electrical Low Voltage Cable Degradation", MDPI Energies, vol 14, no10, pp.2852.
[11] B. Bakhshideh Zad and al., 2022, "Supervised Learning-Assisted Modelling of Flow-Based Domains in European Resource Adequacy Assessments", Applied Energy, 325.
[12] Synergrid, accessed on 15 July 2022, "*Synthetic Load Profiles*", Available: http://www.synergrid.be/index.cfm?PageID=16896
[13] L. Thurner et al., 2018, *"Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems"*, IEEE TPWRS, vol. 33, no. 6, pp. 6510-6521.