

Energy Consumption Prediction in a Novel Automated Photovoltaic Design Platform

Tiago C. Pereira^{1,*}, João Murta-Pina^{1,2}, Rui Amaral-Lopes^{1,2}, Fernando Monteiro², Samuel Moraes³, and Francisco Oliveira⁴

¹NOVA School of Science and Technology, NOVA University of Lisbon, 2829-516 Caparica, Portugal

²Center of Technology and Systems (CTS), UNINOVA, 2829-516 Caparica, Portugal

³Engibase – Engenharia e Construção, Lda., 2830-271 Barreiro, Portugal

⁴Digitalmente, Lda., 3860-384 Estarreja, Portugal

Abstract. This paper describes a multi-step algorithm used to predict and typify the energy consumption profile of a prosumer, allowing the automation of the design of self-consumption photovoltaic (PV) power systems in a novel platform called PV SPREAD. The algorithm uses different methodologies to address various possible scenarios of data availability. In this paper, those scenarios are addressed using nonlinear autoregressive artificial neural networks (ANN) with external inputs (NARX) to predict energy consumption. Results reveal that the proposed algorithm successfully addresses data gaps in a hotel load profile used as a case study. The results also show the limitations of NARX when residential clients are analyzed.

1 Introduction

Renewable energies, especially solar, hydro and wind continue increasing worldwide [1]. Solar energy is considered to have the potentially shortest return of investment period among all other renewable energy sources (RES) [2]. Therefore, there is an increasing economic, and subsequently environmental interest from every sector, especially industrial and residential, to consider exploring photovoltaic (PV) power systems. For that, to improve the reliability and economic feasibility of such systems, its accurate sizing is necessary [3], especially if they are operated under self-consumption schemes.

Due to the multidisciplinary nature of the design of photovoltaic plants, a new ecosystem named PV SPREAD (an acronym for Stimulating PV Reliance, Advance, and Dissemination) [4] is proposed to build a climate of trust and rigour between designers/suppliers of PV installations and the customers.

Focusing on PV self-consumption schemes, according to the Portuguese Decree-Law 153/2014 [5], finding the optimal design of such systems involves calculating the balance between energy production and consumption for every timestep of fifteen minutes throughout

* Corresponding author: taa.pereira@campus.fct.unl.pt

the system lifecycle. Consequently, predicting the energy consumption of costumers that are interested in this type of power systems is essential for a design as accurate as possible.

There is a lot of research on time series prediction such as related to energy consumption [6]. Such methodologies are diverse where deep learning [7] and statistical methods [8] are some examples. During PV SPREAD application, 15 days of energy consumption (power profiles) and energy bills correspondent to a year are typically collected. Considering this acquired data, artificial neural networks (ANN) are a good choice to predict consumption throughout the year, because they have often shown good performances in modelling energy consumption behaviour, which can be increased if the amount of available data is higher than the 15 days.

Focusing on artificial neural networks, nonlinear autoregressive ANN with external input (NARX) have shown good results in real applications [5,7,8]. Regarding electric load forecasting using NARX, L. Ruiz et al. [6], J. Buitrago and Shihab Asfour [9] and M.U. Hashim et al. [10], demonstrated that the NARX architecture can achieve promising results if the external variables used to train the networks show good correlations. This architecture, as an example, allows the use of a client's electricity bill as an external variable to improve its energy consumption prediction if both have good correlations.

Therefore, this paper proposes a multi-step algorithm that forecasts energy consumption in specific customer cases that can occur during the design of self-consumption PV power systems, using mainly NARX.

This paper is organized as follows: the project in whose frame the algorithm is developed is presented in Section 2, while Section 3 describes the development of the algorithm and possible customer cases. Next, two case studies are presented in Section 4 to evaluate the algorithm. Finally, conclusions are drawn in Section 5.

2 The PV SPREAD Project

The algorithm proposed in this paper is developed in the frame of PV SPREAD, a Portuguese project funded by the Portugal 2020 programme. The objective of this project is to develop an ecosystem that supports the suppliers/designers of PV power plants in all its design stages, even before the contact with prospective clients, thus allowing them to expeditiously automating and optimising the elaboration of such designs. The ecosystem will be built by a set of hardware and software tools, with all the associated methodologies, to guide the user throughout the whole design of the PV power system, beginning with the characterization of the installation until the elaboration of such systems.

Figure 1 shows conceptually the functionalities of the PV SPREAD ecosystem. The first step is to get all the data needed for the optimal system design through a mobile application with a computer-assisted personal interviewing (CAPI) methodology; after that, the data will be sent to the Cloud for optimization of the PV plant; and finally, a report with the optimal project and investment indicators, as the net present value (NPV), the internal rate of return (IRR) and the investment payback time (IPT), is produced.

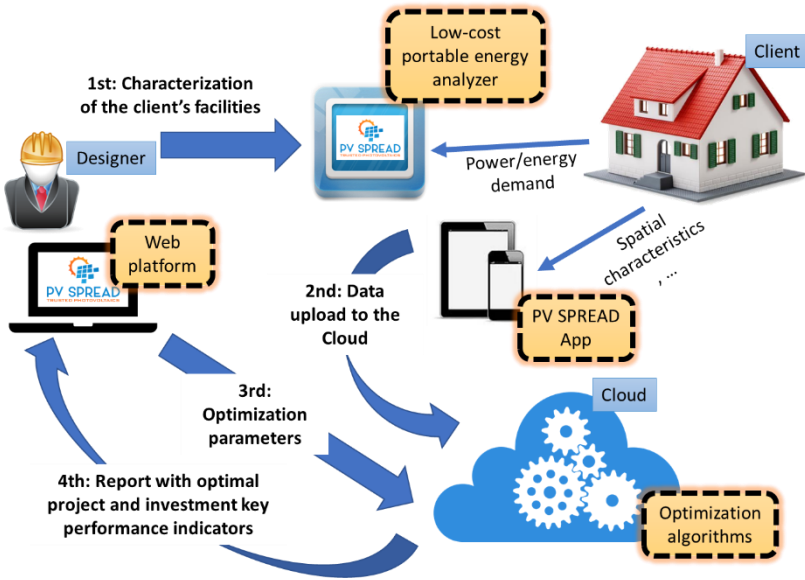


Fig. 1. PVSREAD ecosystem

The energy consumption forecast algorithm proposed in this paper will be processed in the Cloud to use the data acquired in both the energy analyser and the CAPI interview, to typify/forecast the clients' load profile whenever its necessary to design a self-consumption PV power system.

3 Proposed Algorithm

During the PV SPREAD project, a low budget energy meter was developed to obtain, if possible, at least 15 days of the power consumption profile. This is the minimum period that is considered to allow for typifying the profiles for the rest of the year, together with past, existing energy bills. The latter are obtained from the customer interview, so the algorithm normally has the fifteen days of energy consumption and a year of electricity bills as inputs, to predict the future years.

Considering the two types of costumers, residential (Res) and non-residential (NRes) (i.e. industries, business buildings), the amount of input data that the algorithm will be supplied depends on certain cases/possibilities that the designer will face. A total of six different cases is envisaged, and they are divided between the two types of costumers:

1. It is possible to get the fifteen days of energy consumption and one year of electricity bills.
2. One year of energy consumption profiles is available (which is typical for customers with remote metering), but those have some considerable flaws, due to system failures.
3. The customer doesn't have any information about energy consumption and there is no possibility to apply the energy meter. This is the case of, e.g., new buildings.

In this document, each case will be addressed as such: the type of client (Res or NRes) and the number of the specific case presented before (i.e. NRes2, Res1). In Figure 2, the main algorithm flowchart is presented.

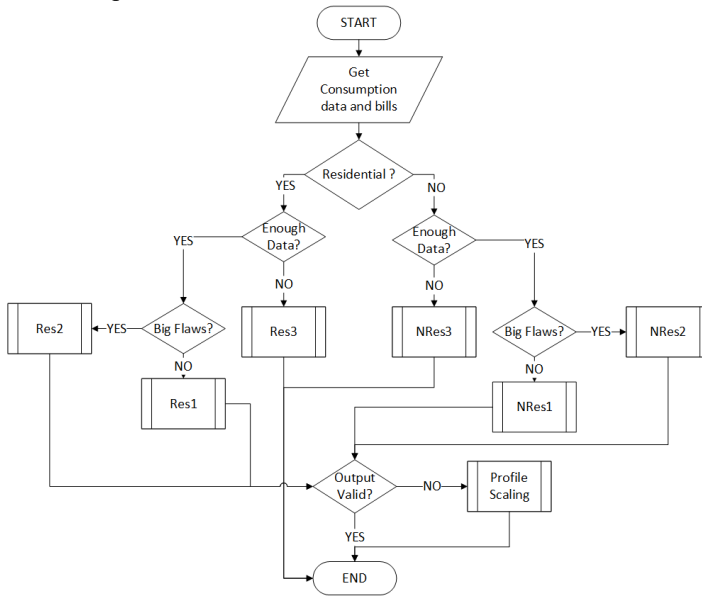


Fig. 2. Main algorithm flowchart

In Portugal, time-of-use (TOU) tariffs are the more common way of charging electric energy consumption, and each client can have very distinct TOU tariffs depending on the type of voltage they are supplied with. These tariffs vary with the time of the day and the day of the week. This information is also obtained during the interview and has an impact on the algorithm's output. For each case, the algorithm has a sub-algorithm designed to solve it.

The cases Res2 and NRes2 are solved by the same sub-algorithm. It starts by identifying the big flaws in the provided profile and, for each flaw, a NARX neural network is trained with the previous data and predicts the energy consumption corresponding to that specific failure. The difference between Res and NRes clients is not only related to the TOU tariffs but also the external variables used to train the ANN and forecast the profile failures. In NRes clients, for each fifteen minutes timestep, it is used the total energy consumed in the specific period of TOU tariff, provided by bills of the customer, the hour of the day (1-24), the day of the month and the external temperature profile. On the other hand, for Res clients, the external temperature is not used because of the poor correlation with energy consumption, especially in households with a low number of residents.

The cases Res1 and NRes1 are both solved with another sub-algorithm. In this case, the data of the fifteen days of real energy consumption obtained by the energy meter and the electricity bill is available. Because NARX networks are data-driven neural networks, its performance is influenced by the amount of data, but in this case, there is a need to extract the most information from the profile obtained and bills. Therefore, it is separated all days of the profile by day of the week and used a NARX network for each of them. This way, the specific day of the week network, will only focus on the pattern of that day, providing better performance if there is a different pattern in each day. The external variables are the same used in both cases Res2 and NRes2, respectively.

Case Res3 is solved using a sub-algorithm based on a stochastic model developed by Richardson et al. [11]. This model follows a bottom-up approach, where the individual domestic electricity loads are the building blocks. It uses stochastic occupancy profiles and information of the respective activity, performed by its building occupants when at home and awake to define the state of each load in the building. All the information needed for the model is obtained in the interview if this is the case.

Finally, in the case of NRes3, typical profiles will be assumed. A database will be built as the PV SPREAD application grows in clients, that will contain anonymously every type of profile for each type of industry/building. The typical profile will be a mean profile for the specific type of activity the client has (e.g., mean scaled profile for a paper factory, using other profiles of similar industries with similar overall power).

If the output of the profile is not valid for that specific client (i.e. load consumption exceeding the maximum installed, energy consumption not corresponding to energy bills) the profile will be scaled accordingly.

4 Case Studies

In this Section, two cases are studied, one for the case Res1 and NRes2. In Res1, a real household profile was used, where fifteen days of energy and electricity bills are used to predict the next year consumption profile. In NRes2, a real hotel profile is used to simulate a Non-Residential profile with flaws and predict the data missing in those failures. The predictions in these cases studies were evaluated using mean absolute error (*MAE*), defined as:

$$MAE = \frac{\sum_{i=1}^n |x_i - y_i|}{n} \quad (1)$$

where y_i is the predicted value, x_i the real value and n the number of timesteps evaluated.

Both cases were implemented with the MATLAB neural networks toolbox.

4.1 NRes2 Case

In this case, a real energy consumption profile from a hotel is used to test the proposed algorithm. Figure 3 presents the TOU tariff considered during this case study, which varies with the hour of the day, with the season (summer or winter) and with the specific day of the week. Periods are classified as super off-peak, off-peak, half-peak and peak periods, with increasing tariff.

The hotel bills provide the amount of energy consumed in each specific period for each month of the year. With that information, it is possible to create an external variable that has both the period and total amount of energy consumed for each timestep. Because this is an NRes client, the external temperature will also be used as an external variable for the NARX networks. Both load profile and external temperature are shown in Figure 4.

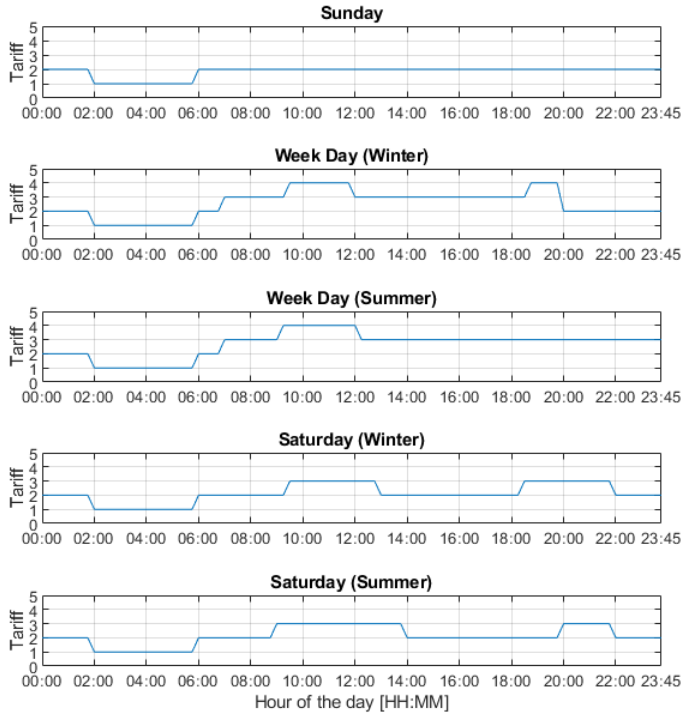


Fig. 3. TOU tariffs for case NRes2

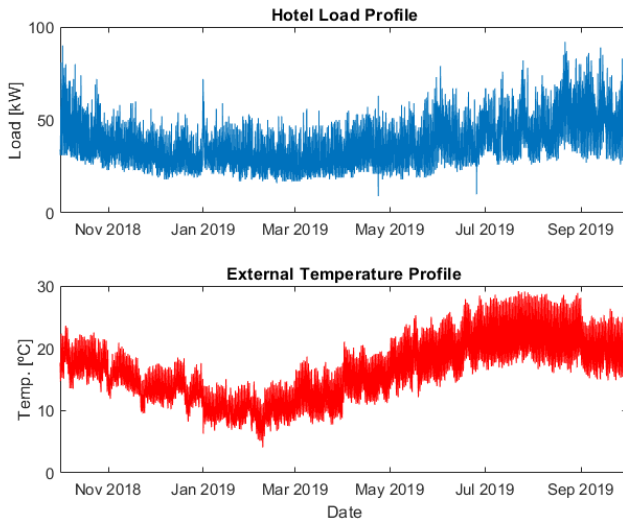


Fig. 4. Hotel load profile and External Temperature profile

Flaws were created randomly in this case to test the algorithm. For each flaw, a NARX network will train with previous profile values and forecast the specific flaw. Figure 5 shows the load profile used as input for the algorithm.

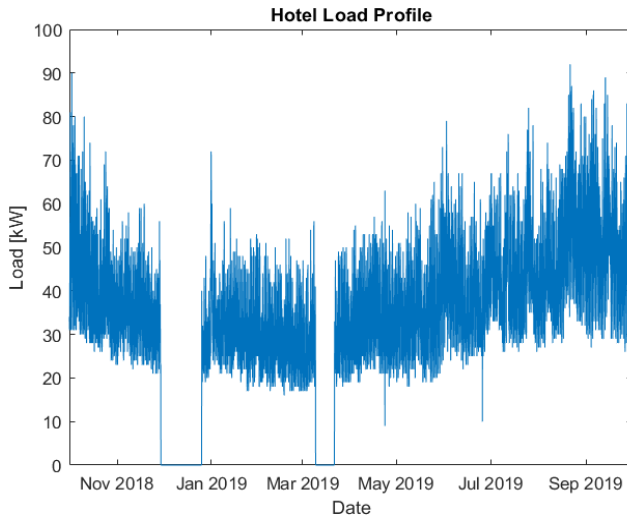


Fig. 5. Hotel load profile with flaws, that may be caused by communications equipment failure

After identifying which type of case is requested, the algorithm will guide the data for the correspondent sub-algorithm. This sub-algorithm can be described by the flowchart presented in Figure 6.

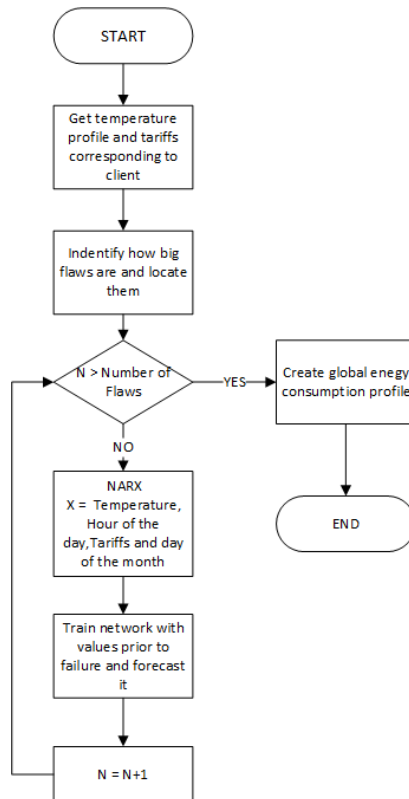


Fig. 6. NRes2 case sub-algorithm flowchart

Before initializing the ANN training, the network must have its parameters specified. The inputs are the known load consumption and external inputs. The delay of its inputs, external input delay and feedback delay. External input delay is determined by analysing the cross-correlation between each external variable with the feedback variable (i.e. determining the best delay that shifts the temperature profile to have the best correlation with the load consumption profile). The feedback delay is the energy consumption delay and it was always one because the cross-correlation was determined with a non-shifted load consumption profile. The network has two layers, one hidden layer and one output layer. Each layer has a specific number of neurons with weights, bias, and a transfer function. The number of neurons used in the hidden layer was 12, which was iteratively determined and chosen the one with better training results. The weights and bias are determined during the training and the transfer functions are determined depending on the type of problem, in this case, functions that give nonlinearity to the model. A schematic of the NARX network used for the first failure in the hotel profile is presented in Figure 7.

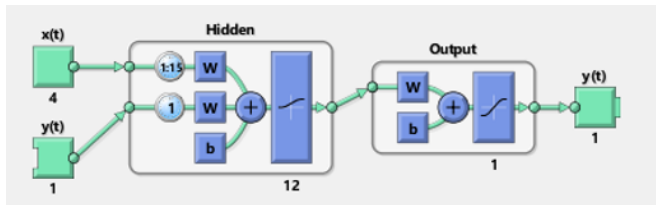


Fig. 7. First flaw MATLAB's NARX networks schematic

The network training was done with Levenberg-Marquardt backpropagation algorithm using mean squared error as a performance function. The training dataset was divided into three datasets, 70% for training, 20% for validation and the last 10% for testing. The training ends if the performance threshold is met or its error increases eight times in a row.

To forecast the load consumption during the timesteps where the profile fails, the network is converted into a closed loop as shown in Figure 8.

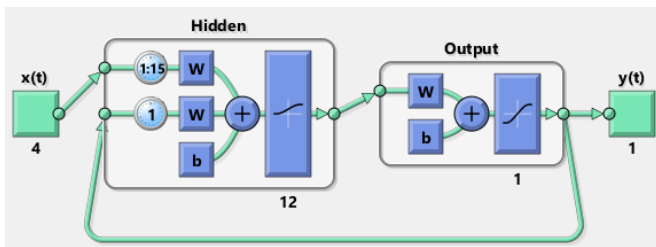


Fig. 8. Diagram of the first flaw closed-loop MATLAB's NARX networks

In this configuration, the network will predict multi-step ahead, by feeding back each fifteen minutes prediction of load and its respective known external variables, to produce the next flaw forecast. The objective of this study is to obtain the forecasted results at first try simulating a real application of the PV SPREAD project.

Regarding the first flaw, the training stopped at a mean squared error of 13.1 kW. Comparing the real load profile and the first flaw prediction the MAE was 7.36 kW. The second flaw, training mean squared error was 9.86 kW and its prediction MAE 8.10 kW. Figure 9 shows the results for the first and second flaw and Figure 10 presents the real and the predicted hotel load. The overall MAE was 3.04 kW, which means, on average the forecasted profile is 3.04 kW higher or lower than the real profile. Considering the low amount of data used to forecast, the algorithm showed good results. Due to the scale of the

profile, this error will have a negligible impact on the design of a self-consumption PV power system for the hotel. This type of evaluation will vary a lot between the type of client and its profile. The PV SPREAD application will also develop risk models that also take prediction errors into account that influence the application's output.

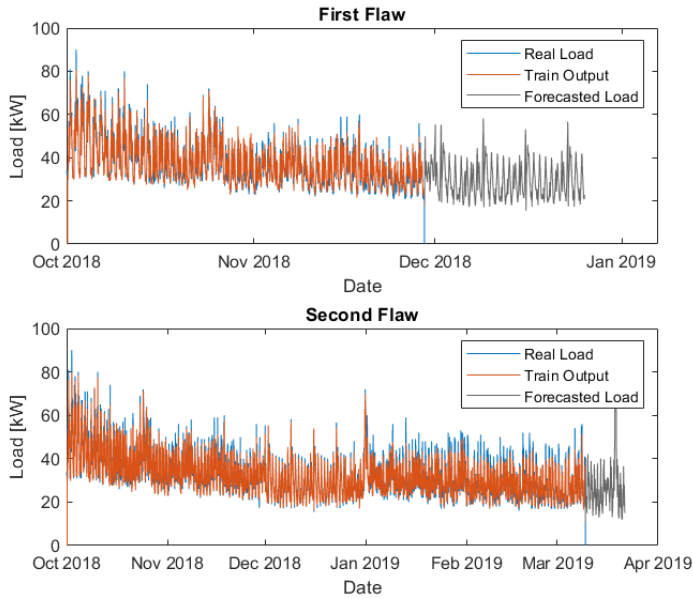


Fig. 9. NRes2 Results

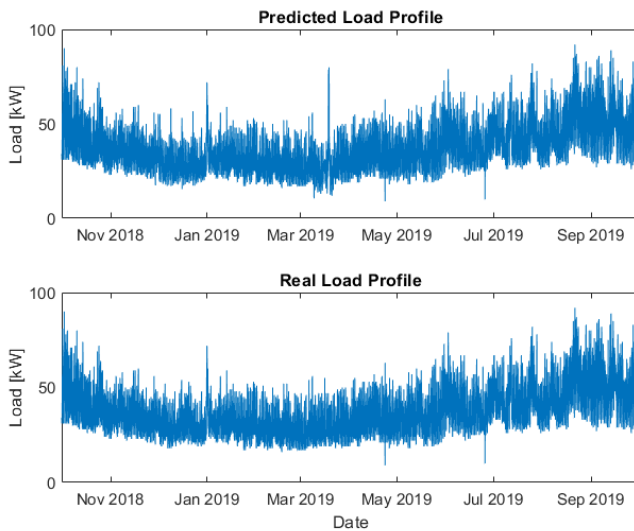


Fig. 10. Predicted load profile VS Real load profile

4.2 Res1 Case

In this case, a real load profile from a household was used. Figure 11 presents the TOU tariff considered during this case study.

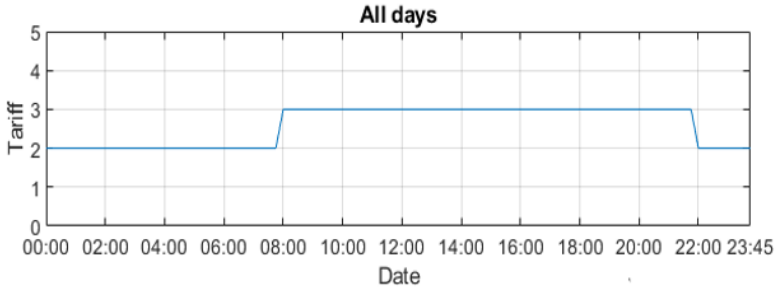


Fig. 11. TOU tariffs for case NRes2

This TOU tariff was used for all days of the week, where 2 represents the off-peak period and 3 the peak period. The load profile is shown in Figure 12.

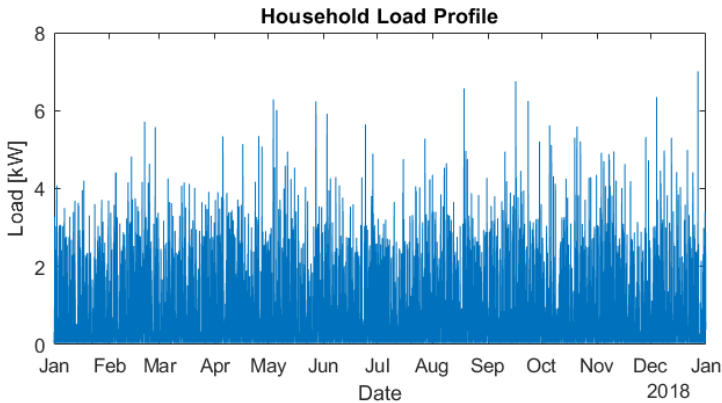


Fig. 12. Household load profile

To simulate the Res1 case, only the first fifteen days of the load profile and electricity bill was used. The sub-algorithm responsible for this case used the same type of NARX network as NRes2 case, but with a different methodology. The flowchart representing the sub-algorithm used for this case is presented in Figure 13.

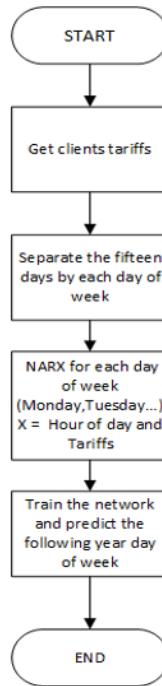


Fig. 13. Res1 case sub-algorithm flowchart

Each NARX network is trained with each day of the week, using the hour of day and tariff. After training, each network is converted into a closed loop to predict the next days of the year (e.g., the Monday network will predict all Mondays throughout the year). Then a global load profile is created using the predictions for each day.

Each network parameter was determined as explained before, using 10 neurons in the hidden layer instead of 12, which was the number that allowed better training results in this case. The overall MAE was 0.441 kW and the result can be seen in Figure 14. These results are extremely dependent on the type of load profile. Residential loads are known for their random behaviour due to its residents' consumption habits. In this case, this load has a poor correlation with the chosen external variables. Also, using a low amount of data as input limits NARX neural networks performance. These limitations are evident when comparing both NRes2 and Res1 cases. Where in case NRes2 there is more data and a better correlation between external variables and the load profile.

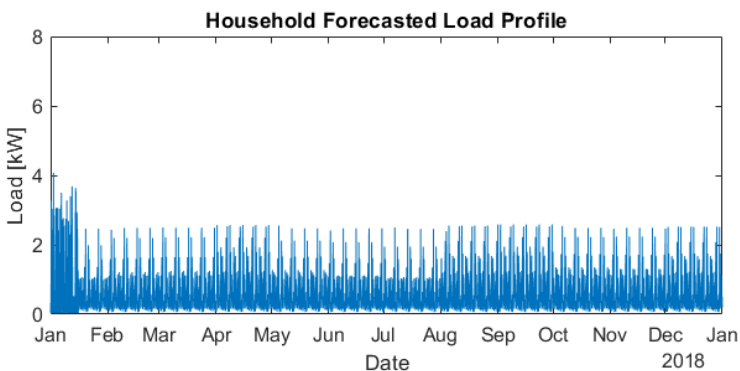


Fig. 14. Res1 Forecasted Load Profile

In this case, the predicted load was not matching the client’s total energy consumed presented in its electricity bills. Therefore, as shown in Figure 2, the predicted load profile is scaled considering the amount of energy consumed in each month and each different TOU tariff period. The scaled profile is presented in Figure 15. Its overall MAE was 0.466 kW, which is higher than the predicted profile. Considering both MAE, the chosen profile for PV design would be the predicted one. But initially, during the PV SPREAD project, comparing the real profile or typical profile with the predicted one will not be possible, so to evaluate if the predicted profile is valid, the client’s characteristics must be considered. Therefore, in a real case, if the predicted profile needs to be scaled, its scaled version will be the one used for designing the self-consumption PV power system.

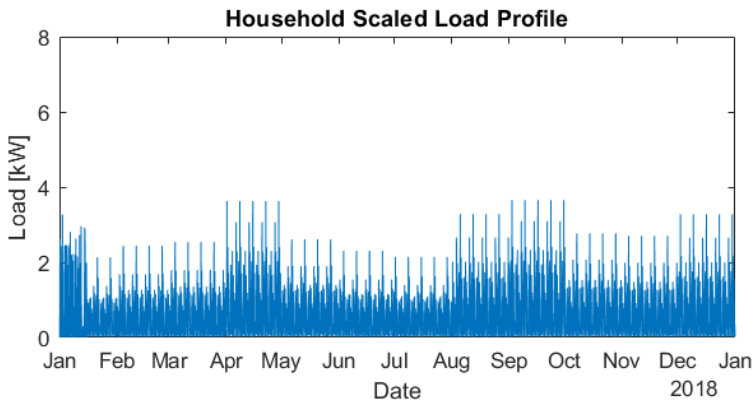


Fig. 15. Res1 scaled load profile

5 Conclusion

In this paper, an ecosystem that is intended to create a climate of trust and rigour between designers/suppliers of PV installations and the general public is presented. To design a PV self-consumption power plant, it is mandatory to know the client's energy consumption profile. Therefore, an energy consumption prediction algorithm is proposed. This algorithm considers every possible case that the designer will face when using the PV SPREAD application.

The proposed algorithm uses different methodologies to predict load diagrams, depending on the available information obtained during the interview with clients. In this document, two cases are studied, where the algorithm uses NARX neural networks to solve them. The two cases belong to two different types of clients. NRes2 corresponds to a hotel Res1 to a residential client. In the NRes2 case, the algorithm was successful in predicting the load profiles correspondent to random failures. In Res1, it is demonstrated the NARX networks limitations and the need for a validation sub-algorithm that scales the predicted profile to match the client's characteristics.

In future work, the algorithm will be adapted and optimized to use a higher amount of data, a consequence of the increase in users of PV SPREAD application. With this adaptation, the algorithm is expected to improve its performance.

This work was financed by the European Union, within the frame of the Portugal 2020 programme, in the framework of project PV SPREAD (reference LISBOA-01-0247-FEDER-039846)

References

1. G. Masson and I. Kaizuka, *Trends in Photovoltaic Applications 2019* (IEA, 2019).
2. S. Chowdhury, U. Sumita, A. Islam, and I. Bedja, *Energy Policy*, **68**, 285–293 (2014)
3. C. V. T. Cabral, D. O. Filho, A. S. A. C. Diniz, J. H. Martins, O. M. Toledo, and L. V. B. M. Neto, *Solar Energy*, **84**, 9, 1628–1636 (2010)
4. “*PV SPREAD Trusted Photovoltaics*.” Available at: <https://pvspread.com/>.
5. Ministério Do Ambiente Ordenamento Território Energia, “Decreto-Lei n.º 153/2014,” *Diário da República - I Série*, **202**, 5298–5311 (2014) Available at: <https://dre.pt/application/file/58428682>.
6. L. Ruiz, M. Cuéllar, M. Calvo-Flores, and M. Jiménez, *Energies*, **9**, 9, 684 (2016)
7. W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, *IEEE Trans. Smart Grid*, **10**, 1, 841–851 (2019)
8. S. Chen, J. Wu, Y. Pan, J. Ge, and Z. Huang, *Energy and Buildings*, **223**, 110182 (2020)
9. J. Buitrago and S. Asfour, *Energies*, **10**, 1, 1–2 (2017)
10. M. U. Hashmi, V. Arora, and J. G. Priolkar, *ICIC 2015*, 1418–1423 (2015)
11. I. Richardson, M. Thomson, D. Infield, and C. Clifford, *Energy and Buildings*, **42**, 10, 1878–1887 (2010)