Development of Neurofuzzy and Gaussian Regression Models for a Solar Photovoltaic System

William D. Chicaiza[1] [0000−0001−5473−9538], Yeyson A. Becerra-Mora[1,3] [000−0002−6718−4962], Juan M. Escaño[1] [0000−0003−1274−566X], Adolfo J. Sánchez[2] [0000−0003−3942−847X], and José Ángel Acosta[1] [000−0003−0040−338X]

[1] University of Seville, Dept. of System Engineering and Automatic Control, Seville, Spain
{wchicaiza,jescano,jaar}@us.es, yeybecmor@alum.us.es
[2] Munster Technological University, Dept. of Mechanical, Biomedical and Manufacturing Engineering, Cork, Ireland
adolfo.sanchezdelpozofernandez@mtu.ie
[3] Corporación Unificada Nacional, Dept. of Electronics Engineering, Bogotá, Colombia
yeyson_becerra@cun.edu.co

Abstract. This paper focuses on the development of two prediction models for a solar photovoltaic system that is part of a multimachine industrial manufacturing plant. These models are part of the set of models that form the digital twin of their physical counterparts, which will be used to perform control and optimization strategies to maximize the use of renewable energy sources within a Digital Twin (DT) architecture. The first model is based on a fuzzy neural network and the second one is a Gaussian regression model. The obtained models present a good performance in the prediction of the nonlinear dynamic over the entire operating range in the system.

1 Introduction

In recent years, renewable resources for the electricity generation has increased like photovoltaic fields and on/off-shore wind farms used to meet consumer demand. At the same time, the high price of energy is leading industries to adopt these solutions to supply energy to their different manufacturing processes, thus avoiding a total dependence on grid power. However, energy management within the industry is not always optimal, often preventing the full benefits of renewable sources from being realized. For this reason, optimization strategies are being implemented to maximize their use [6] and thus, make their installation cost-effective.

Current technological advances allow the development of digital replicas (virtual entities) of almost any product or system (physical entity) with a high level of detail, which is called Digital Twin (DT). The DT is a digital information construction of a system based on its features. It is composed of a set of high-fidelity or sufficient multiphysical models and simulators for each domain in which it is developed (fit for purpose), that simulates a specific aspect of the system in a single ecosystem. This allows multiscale tests and experiments to be performed virtually, avoiding the cost of physically performing them [4].

Control and optimization processes often use reduced (surrogate) models of highly nonlinear dynamical systems. For example, model-based predictive control employs various models (e.g., linear, nonlinear, grey-box, or black-box), as solving the optimization problem with many differential equations or distributed parameter models, with high accuracy, can be impractical, specially when the problem has to be solved in a limited time. Therefore, real-time computation requires reducing the model complexity to fit the purpose, saving unnecessary computational burden. In this context, the proposed models provide system evaluation, optimization, and prediction capabilities for the DT; furthermore, these models enable real-time decision-making and long-term planning to explore different scenarios quickly and easily once the DT is implemented.

AI techniques can be employed to develop these reduced models, which capture highly non-linear dynamics. One technique, neurofuzzy systems, has proven the effectiveness of nonlinear systems modelling, e.g., the model of a solar Fresnel plant in which the neurofuzzy model is based on an ANFIS network [10]; besides, it has been applied in the design of classifiers and estimators [11, 12]. The fuzzy neural network (ANFIS) combines the advantages of fuzzy logic (FL) and artificial neural networks (ANN), uses the FL to represent knowledge in an interpretable way and the learning capacity of an ANN to optimize its parameters based on input-output data. Another technique that can be used in this context is Gaussian Mixture Model (GMM) in conjunction with Gaussian Mixture Regression (GMR). GMM is employed to identify sub-populations or nonlinear patterns in the data, which can be beneficial for segmenting complex datasets into simple, more modellable components. On the other hand, GMR is used to perform non-linear regressions on data that exhibit a non-linear structure, allowing the prediction of outputs from inputs [2]. Furthermore, these modelling techniques have advantages, such as the ability to update and run quickly.

This work focuses on an industrial manufacturing plant (physical entity) consisting of several computer numerical control (CNC) machines, a combined heat and power (CHP) generator, two renewable energy sources for electricity production and a battery bank for energy storage. The main purpose of this plant is to maximize the use of renewable energy in the manufacturing process. Given that the meteorological variables that can be predicted over the horizon are scarce (wind speed and global irradiance) as well as the renewable energy generation responds to a complex nonlinear model, it has been decided to use the above techniques. Therefore, this paper focuses on developing two models for the renewable energy system (solar). These models are part of the DT that will be used to perform control and optimization strategies within a DT architecture of the physical entity.

The rest of the document is structured as follows. Section 2 shows the initial treatment of the data for energy system. Section 3 presents the modeling of reduced models, as well as the learning process and its structures. Section 4 presents the results of the simulations that compare the results of the neuro-fuzzy and GMM/GMR models with the real values. Finally, the conclusions are presented in Section 5.

2 Data treatment

The techniques used in modelling are data-driven for its learning, validation, and updating; therefore an initial process is necessary to homogenizing sampling times, replace inconsistent data between samples by interpolation, removing outliers and filter each variable, as noted in [10]. Furthermore, variables that affect the process must be carefully chosen.

The data to be processed are from a 210 kW photovoltaic installation on the roof of the manufacturing plant, with a series-parallel configuration of 688 solar panels and three SMA inverters that can reach an annual energy production of approximately 160 MWh. Each solar module consists of 120 cells and the total PV installation has a surface area of $1,107\ m^2$. In addition, the SCADA of the photovoltaic plant collects data on the generated power, which is recorded every minute. In particular, this work uses historical SCADA data from May to August 2022, a total of 103 days.

2.1 Solar irradiance on a tilted surface

The photovoltaic modules in the installation have an angle of inclination ($\beta = 15°$) that is used to calculate the actual irradiance that reaches the modules based on the different irradiances measured by the pyranometers and the pyrheliometer. This irradiance is called solar radiation on the tilted surface and can be calculated with various models as described in more detail in [9]. These models differ mainly in the way diffuse radiation is calculated. In this case, one of them has been used, the isotropic sky model, which requires measurements of direct global radiation (G), the diffuse radiation (G_d), and reflected radiation (G_{refl}) which are provided by the instruments mentioned above. The total radiation on a tilted surface is given by

$$G_T = G_b R_b + G_d \frac{1 + \cos(\beta)}{2} + G \cdot \rho_g \frac{1 - \cos(\beta)}{2} \tag{1}$$

where ρ_g is the reflectivity coefficient of the ground, $G \cdot \rho_g = G_{refl}$ and R_b is the ratio of radiation over the tilted surface with respect to a reference plane. For a detailed description of the isotropic sky model, refer to [1].

After calculating the irradiance on the tilted surface for each day with the pre-treated data, this one is added as a variable to the final dataset for further analysis and selection of the involved variables in the power generation process presented in the next section.

2.2 Correlation analysis of I/O variables

The correlation coefficient measures the association between variables; the most commonly used is Pearson's correlation coefficient, which is a linear correlation coefficient (ρ) used in this work. The ρ values can be between $[-1, 1]$, where $\rho = -1$ stands for a complete negative correlation, $\rho = 1$ a complete positive correlation, and $\rho = 0$ indicates that the displayed variables over the axes are uncorrelated. The correlation coefficient matrix obtained with 6 variables is ${}^{PV}\mathbf{R} \in \Re^{6\times 6}$ for the photovoltaic installation.

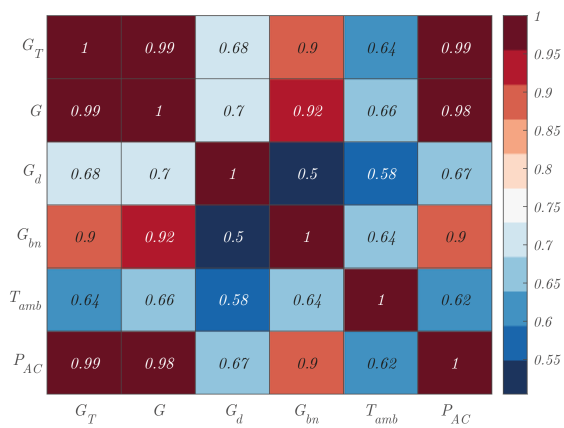

Fig. 1. Photovoltaic correlation coefficients matrix ${}^{PV}\mathbf{R}$.

Fig. 1 shows the correlation coefficient matrix of the renewable energy source. The variables are then ordered according to the degree of correlation of each input in relation to the desired output. This analysis indicates the variables that most affect the model output with respect to a threshold, $\rho \geq 0.5$.

Therefore, the input data used in the learning process to estimate the power generated (P_{AC}) by the renewable energy source (RES) for both methods are: G_T and T_{amb}. Which stand the irradiance on the tilted surface and ambient temperature, respectively. Since the estimation of G_T involves G, G_{bn}, G_d, those are omitted as model inputs. Thus, these variables (G_T ,T_{amb}) are stored in a matrix ${}^{Pv}\mathbf{X} \in \Re^{n\times m}$, where $n = 148320$ is the number of samples and $m = 2$ is the number of variables.

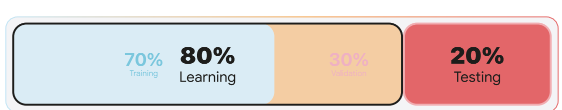

Fig. 2. Clustering of available data of the RES.

The different scales of variables can affect the data-driven learning process due to its nature, magnitude, and inconsistencies. This is solved by the normalization process, as noted in [3]. Thereby, each variable m in the data matrix $^{Pv}\mathbf{X}$ must be normalized to the range [0 1], given by

$$\mathbf{z}_{i,j} = \frac{x_{i,j} - x_{j,min}}{x_{j,max} - x_{j,min}} \quad (2)$$

The new normalized variables are stored in the matrix $^{Pv}\mathbf{Z}$ and organized into the sets: training, validation and testing as shown in Fig. 2.

3 Photovoltaic surrogate models

The large volume of data and the models complexity used in DT technology can create significant computational demands. One solution to alleviate this burden is the implementation of reduced-order models. Reduced-order models are a technique to simplify complex models by quickly capturing essential features of the phenomena. Often, it is not necessary to compute all details of a full-order model to meet real-time constraints. These models are especially useful in situations where computational time is a critical factor, such as in real-time decision making (e.g., [3, 10]). In addition, reduced-order models can also be used to optimize and control complex systems, as they allow for a compact representation of systems and their behavior. This can help improve the efficiency and performance of systems, reducing analysis and simulation time and costs [4].

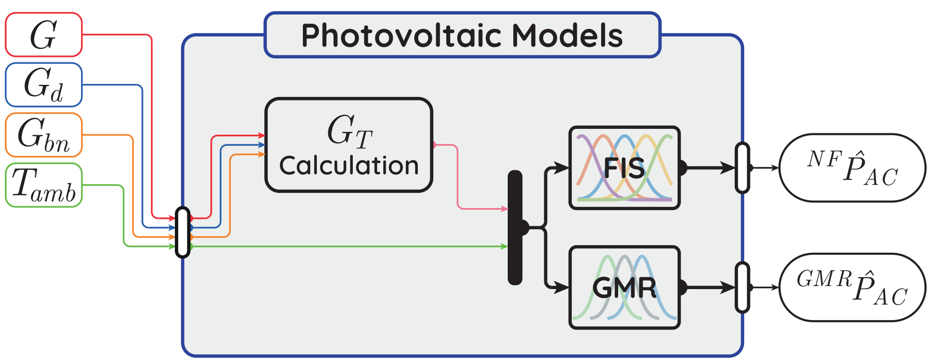

Fig. 3. Subrrogate models: Neurofuzzy and Gaussian Mixture Regression to predict the active power of a photovoltaic plant

3.1 Neurofuzzy Model

An artificial intelligence (AI) technique is employed to combine the advantages of fuzzy logic (FL) and artificial neural networks (ANN) in the design of the PV prediction model. This method is based on an adaptive neuro-fuzzy inference system (ANFIS), which was introduced in [7]. It can construct an input-output map based on human knowledge in the form of fuzzy "***if-then***" rules, enabling the handling of imprecise input data (uncertainty).

The fuzzy neural network uses the training and validation data sets in the learning process to capture the dynamic behavior (nonlinear) of the power generated for the PV system. In this process, the ANFIS looks at the normalized RMSE (nRMS) of the training and validation sets to not overfit only the first set, which would cause inappropriate values for the obtained FIS. In this way, it searches for a middle ground where learning is general in both groups according to [3].

The subtractive clustering (SC) method is initially applied in ANFIS learning, which estimates the number of clusters per input that defines the number of membership functions (MFs) to compose its fuzzy set and the number of rules. Additionally, it estimates the initial parameters of these membership functions. A hybrid learning method combining least squares and gradient descent is then applied [10]. In this process, a gradient descent to determine the mean (c_{ij}) and standard deviation (σ_{ij}) of Gaussian MFs denoted as *antecedent parameters* and least squares to estimate the coefficients (g_{0j}, g_{ij}) of each first-order polynomial function referred to as *consequent linear parameters* at each epoch or sweep.

In this case, a cluster influence range of 0.7 and 250 epochs was used, which generated a fuzzy set per input (A_{1j} , B_{2j}) consisting of two Gaussian MFs and two rules (j). The nRMSE index obtained in the learning process for the ANFIS present small errors, $nRMSE_{Trn} = 0.038$ and $nRMSE_{Chk} = 0.040$ on the training and validation sets, respectively.

Table 1. Obtained parameters during the learning process

Antecedent Parameters MFs				
Inputs (x)	G_T		T_{amb}	
	A_{1j}		B_{2j}	
j	σ_{1j}	c_{1j}	σ_{2j}	c_{2j}
1	0.4355	-0.1361	0.3338	0.5201
2	0.3209	0.4441	0.3327	0.4329
Consequent Parameters $\hat{P}_{AC,j}$				
j	g_{1j}	g_{2j}	g_{0j}	
1	-0.6053	0.1841	-0.3279	
2	0.7298	0.0505	0.5723	

Once the ANFIS training is completed, a fuzzy inference system (FIS) is obtained. The FIS can be considered a gray box model [8], as the rules that define the behavior of the system can be extracted from it. The FIS that predicted the active power generated by the photovoltaic installation contains 2 rules of type Takagi-Sugeno (TS):

$$
\begin{aligned}
&\mathit{Rule}\ 1: \\
&\qquad \texttt{IF}\ G_T\ \texttt{is}\ A_{11}\ \texttt{and}\ T_{amb}\ \texttt{is}\ B_{21}\ , \\
&\qquad \texttt{THEN}: \hat{P}_{AC,1}(x) = g_{01} + g_{11}G_T + g_{21}T_{amb} \\
&\mathit{Rule}\ 2: \\
&\qquad \texttt{IF}\ G_T\ \texttt{is}\ A_{12}\ \texttt{and}\ T_{amb}\ \texttt{is}\ B_{22}\ , \\
&\qquad \texttt{THEN}: \hat{P}_{AC,2}(x) = g_{02} + g_{12}G_T + g_{22}T_{amb}
\end{aligned}
$$

Each rule has *antecedent* and *consequent* parameters. Both parameters were adapted in the learning process and are shown in the Table 1. The output of each rule is a linear combination of input variables added to a constant term ($\hat{P}_{AC,j}$). The final output of the FIS is the weighted average of each output of the rule, where, ($\bar{w}$) is the ratio of the j_{th} rule's firing strength to the sum of all rules' firing strengths.

$$\hat{P}_{AC} = \sum_j \bar{w_j} \hat{P}_{AC,j} \tag{3}$$

3.2 Learning based on GMM/GMR

An alternative way to estimate the real output (P_{AC}) is using an unsupervised learning algorithm like GMM and its regressor GMR. GMM is a probabilistic algorithm used to organize the data into clusters K, which is based on the weighted sum of probability density function $\mathcal{P}(x_i^n, x_o^n; \mu_k, \sigma_k, \pi_k)$. The underlying idea is to adjust a finite set of Gaussian distributions defined by mean and covariance (i.e. clusters) to a nonlinear system. From the dataset, inputs x_i and output x_o are established to feed the model; therefore, a mean and covariance for every Gaussian distribution can be defined as

$$\mu_k = \begin{bmatrix} \mu_k^{x_i} \\ \mu_k^{x_o} \end{bmatrix}, \quad \Sigma_k = \begin{bmatrix} \Sigma_k^{x_i} & \Sigma_k^{x_i x_o} \\ \Sigma_k^{x_o x_i} & \Sigma_k^{x_o} \end{bmatrix}.$$

Irradiance of a tilted surface G_T and ambient temperature T_{amb} for a single data-point are denoted as x_i^n and the power generated P_{AC} for a single data-point is denoted as x_o^n.Thus, the probability that a data-point fits into a specific cluster is given by:

$$\mathcal{P}(x_i^n, x_o^n; \mu_k, \sigma_k, \pi_k) = \sum_{k=1}^{K} \pi_k \mathcal{P}(x_i^n, x_o^n | k) \tag{4}$$

where π_k is the prior probability and $\mathcal{P}(x_i^n, x_o^n|k)$ is the probability density function (PDF). In order to know the parameters (π_k, μ_k, σ_k) for every Gaussian distribution, K-means algorithm [2] is used to initialize such parameters and EM algorithm [5] to optimize them. Thereby, GMM learns the optimal parameters from the dataset.

A power generated estimate $\hat{P}_{AC}$ is calculated with GMR. Optimal parameters (π_k, μ_k, σ_k) coming from GMM and the inputs x_i are employed to yield an estimate (see Fig. 3). The regressor is defined as

$$\hat{\mathcal{P}}_{AC} = \sum_{k=1}^{K} \gamma_k(x_i)(\mu_k^{x_o} + \Sigma_k^{x_o x_i}(\Sigma_k^{x_i})^{-1}(x_i - \mu_k^{x_i})) \tag{5}$$

where γ_k is described in [2] as a nonlinear weighting term to measure the influence of each Gaussian distribution.

$$\gamma_k(x_i) := \frac{\pi_k \mathcal{P}(x_i^n|k)}{\sum_{i=1}^{K} \pi_i \mathcal{P}(x_i^n|i)} \in \mathbb{R}_{[0,1]} \tag{6}$$

3.3 Digital Twinning

Synchronization/twinning between a physical and its virtual asset is an essential feature of the DT Model. It allows the digital twin to reflect the state and behavior of its physical counterpart and thus, achieve interaction and convergence between both entities. The act of synchronizing the states of a physical and its virtual entity [4], where all virtual parameters reflect the same values as the physical parameters, is referred to as *digital twinning*. However, the synchronization/twinning between the two entities is subject to a *twinning rate*, which is the time required (near real-time or real-time) for the bidirectional exchange of information between the physical entity and its virtual counterpart to operate simultaneously.

Furthermore, the convergence between the physical and virtual twin throughout its lifecycle depends on the granularity of the models that compose it. Since physical parameters and properties change over time, these models must be adapted/updated so that the digital twin can reflect the current state of its physical counterpart. In this context, the *twinning time* of a given digital twin is the updating rate of its constituent models. Hence, knowing the twinning time of a given model is essential to defining the twinning rate of the virtual entity, as noted in [4, 10].

Table 2. Twinning time

Models	**NF**	**GMR**
t_{tw}(min)	2.79	0.023
$t_{tw}/sample$ (ms)	13.68	0.011
t_{tw}/day(s)	1.97	0.016

To compare the NF and GMR models updating performance, twinning of both models is run with operational data whose sampling time (Ts) is of 1 minute, inherent in the historical data. Both models use the same 85 days of actual data operation, which results in 122400 samples, while the number of daily samples is 1440. The performance indexes for twinning are the total twinning time (t_{tw}), which is the total time that it takes to update the model, the twining time per sample, $t_{tw}/sample$, and twinning time per day of operation, t_{tw}/day.

Table 2 presents the twinning time results of each model, highlighting that the elapsed time to estimate the generated power in GMM/GMR is shorter than NF. However, both models can be daily updated at night due to the reduced twinning time to estimate.

4 Simulations and results of the two approaches

The validation process compares the output of NF and GMR with the validation data set corresponding to the photovoltaic installation, the validation setup runs for 18 days: from July 25^{th} to 31^{th} and August 1^{th} to 11^{th}. These validation set are not used in the learning process of NF and GMM/GMR.

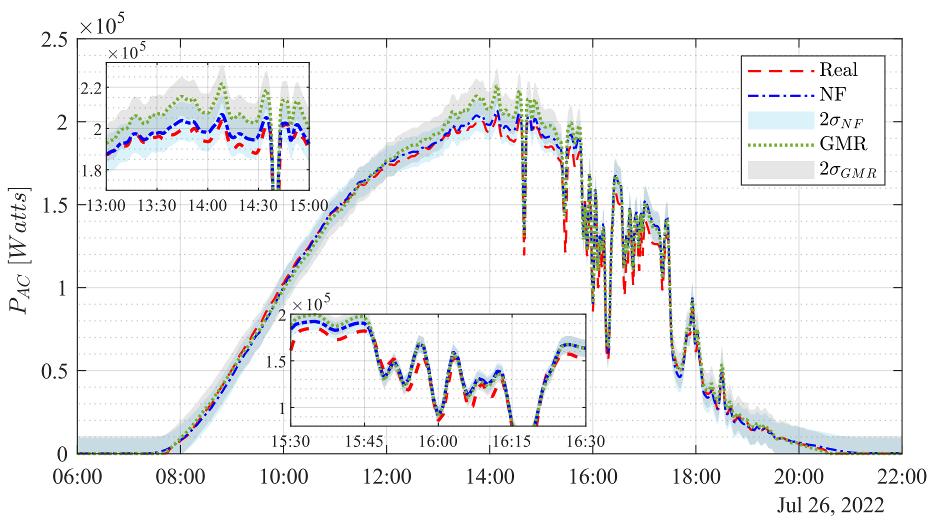

Fig. 4. Validation results. Real data (red line), NF model (blue line) and GMM/GMR model (green line), outputs. Blue and gray shaded areas depict the standard deviation with 95% confidence interval of each model.

Hence, four error indexes were used to compare the output of both models with the actual predicted output data: the arithmetic error mean ($\bar{E}$), the standard error deviation (σ_E), Root Mean Square Error ($RMSE$), and the coefficient of determination [4] R^2. The error indices are given by Equations (7)

$$\bar{E} = \frac{\sum_{i=1}^{N}(x_{i,j} - \hat{x}_{i,j})}{N} \tag{7a}$$

$$\sigma_E = \sqrt{\frac{\sum_{i=1}^{N}(E_{i,j} - \bar{E})^2}{N}} \tag{7b}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_{i,j} - \hat{x}_{i,j})^2}{N}} \tag{7c}$$

$$R^2 = 1 - \frac{x_{i,j} - \hat{x}_{i,j}}{x_{i,j} - \bar{x}_{i,j}} \tag{7d}$$

where $N = 25920$ stands for the number of samples in the validation set, $x_{i,j}$ is the real output and $\hat{x}_{i,j}$ is the output of the predicted variable obtained by each model and $\bar{x}_{i,j}$ is the mean of the observed data.

Only one specific day of the total validation set is shown. This day includes data on the power generated during half of the day when the sky is clear (cloudless) (08:00 am - 14:00), while the second half of the day (from 14:00 onwards), is cloudy, as it can be seen in Fig. 4. On the other hand, most of the days, the real data are within the 95% confidence interval (2σ) of both models. It points out that both models are mostly fitted to real data and are considered acceptably accurate in that region. However, during peak power generation between 13:00 and 15:00 hours, the real data is outside the GMM/GMR confidence interval, indicating that the model is not adequately capturing the variability of the actual data in that region.

The validation indexes for the entire validation set can be found in Table 3. In addition, the table includes the average run time for each sample ($t_v/sample$), per day (t_v/day) and the total ($t_v/total$) evaluation time. It can be seen that NF and GMM/GMR can estimate the real value of active power with a degree of error, as shown in the Table 3.

Validation metrics indicate that the NF model have good accuracy and capture the nonlinear dynamics over the entire operating range (day and night). Based on the mean error, it can be deduced that in both cases the models subestimate the active power, as $\bar{E}$ is negative. The $RMSE$ for the NF is 4.725 kW, whereas for the GMM/GMR is 5.406 kW, represent approximately 2.5% of the power, considering that the photovoltaic system has a nominal operating power of 210 kW.

Furthermore, the linear regression between the estimated and real active power of the output gives the coefficients of determination $R^2 = 0.996$ for the NF model and $R^2 = 0.995$ for the GMM/GMR model. This indicates that the models explain more than 99% of the variation in the output through the inputs.

[4] R^2 is a number between 0 and 1, that measures how well a statistical model predicts an outcome. If $R^2 = 0$, the model does not describe the outputs, if $0 < R^2 < 1$, the model partially predicts the outputs, and if $R^2 = 1$ the the model perfectly predicts the outputs.

Table 3. Overall performance indices

Validation index		
Model Output	NF	GMR
$\bar{E}$ [kW]	-0.924	-1.312
σ_E [kW]	4.634	5.244
RMSE [kW]	4.725	5.406
R^2	0.996	0.995
$t_v/sample\ (us)$	1.956	0.876
$t_v/days\ (ms)$	2.817	1.261
$t_v/total\ (s)$	0.0507	0.0227

Therefore, it can be said that the obtained models show a good fit with the data, a high ability to explain the variability in the output variable through the input variable, and that both are good at making predictions.

According to Table 3, the time to estimate a day is under 3 millisecond, being faster GMM/GMR than NF; nevertheless, both models are sufficiently fast to estimate data in a real-time situation as samples are taken every minute.

5 Conclusion

This work has developed two prediction models for a photovoltaic field. The first, an ANFIS-based predictive models, which at the end of the learning result in a FIS with two rules. The second one, based on statistical and probabilistic methods GMM/GMR. GMM is used to model the distribution of the data, while GMR is used to perform regressions on data that are generated by a mixture of Gaussian distributions. The neurofuzzy model obtained showed good performance in non-linear dynamics predictions throughout the operating range in the systems. It has been shown that both the neurofuzzy model and the GMM/GMR perform well and capture the dynamics of the systems. In addition, the time required to update the models is relatively low. Furthermore, the execution time per sample is less than 1 millisecond in both cases.

In both cases the GMM/GMR has shorter update and run times. But when it comes to predicting the output, during peak power generation, the actual data is outside the GMR confidence interval, indicating that the model is not adequately capturing the variability of the actual data in that region. In either case, these models can be used in control and optimization problems that require a solution in a limited time. Future work may explore improving the modeling methodology presented in a combination of genetic algorithms and also using the models to perform predictive control strategies.

Acknowledgment

The authors thank to the European Commission for funding this work under project DENiM. This project has received funding from the European Union's

Horizon 2020 research and innovation programme under grant agreement No. 958339. Also, this work has been funded by PID2022-142069OB-I00/AEI/10.13039/501100011033/ FEDER, UE and by 0091_AGERAR_PLUS_6_E, under VI-A Spain-Portugal Programme (POCTEP) 2021-2027.

References

1. Beckman, W., Blair, N., Duffie, J.: Solar engineering of thermal processes, photovoltaics and wind, fifth edition. John Wiley & Sons, Ltd pp. 1–905 (1 2021). https://doi.org/10.1002/9781119540328
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer New York, NY (2006)
3. Chicaiza, W.D., Machado, D.O., Len, A.J.G., Gonzalez, J.M.E., Alba, C.B., de Andrade, G.A., Normey-Rico, J.E.: Neuro-fuzzy digital twin of a high temperature generator. IFAC-PapersOnLine **55**(9), 466–471 (2022). https://doi.org/https://doi.org/10.1016/j.ifacol.2022.07.081, 11th IFAC Symposium on Control of Power and Energy Systems CPES 2022
4. Chicaiza, W.D., Gómez, J., Sánchez, A.J., Escaño, J.M.: El gemelo digital y su aplicación en la automática. Revista Iberoamericana de Automática e Informática industrial (2024). https://doi.org/10.4995/riai.2024.20175, https://polipapers.upv.es/index.php/RIAI/article/view/20175
5. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society **39**(1), 1–38 (1977)
6. Gómez, J., Chicaiza, W.D., Escaño, J.M., Bordons, C.: A renewable energy optimisation approach with production planning for a real industrial process: An application of genetic algorithms. Renewable Energy **215**, 118933 (2023). https://doi.org/https://doi.org/10.1016/j.renene.2023.118933
7. Jang, J.S.: Anfis: adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernetics **23**(3), 665–685 (1993). https://doi.org/10.1109/21.256541
8. Lindskog, P.: Fuzzy Identification from a Grey Box Modeling Point of View, pp. 3–50. Springer Berlin Heidelberg, Berlin, Heidelberg (1997). https://doi.org/10.1007/978-3-642-60767-7-1
9. Loutzenhiser, P., Manz, H., Felsmann, C., Strachan, P., Frank, T., Maxwell, G.: Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation. Solar Energy **81**(2), 254–267 (2007). https://doi.org/https://doi.org/10.1016/j.solener.2006.03.009
10. Machado, D.O., Chicaiza, W.D., Escaño, J.M., Gallego, A.J., de Andrade, G.A., Normey-Rico, J.E., Bordons, C., Camacho, E.F.: Digital twin of a fresnel solar collector for solar cooling. Applied Energy **339**, 120944 (2023). https://doi.org/https://doi.org/10.1016/j.apenergy.2023.120944
11. Suganthi, L., Iniyan, S., Samuel, A.A.: Applications of fuzzy logic in renewable energy systems – a review. Renewable and Sustainable Energy Reviews **48**, 585–607 (2015). https://doi.org/https://doi.org/10.1016/j.rser.2015.04.037
12. Zhong, Z.: Modeling, Control, Estimation, and Optimization for Microgrids: A Fuzzy-Model-Based Method. CRC Press, New York, NY (2020)