A Framework for Sensitivity Analysis of Parameters in a Dual-Diode Solar Cell Model

Yi-En Tang, I-Li Chiu, Yu-Tzu Chen, Wai-Chi Yang, Yan-Hao Huang*

Abstract

The performance of equivalent circuit models for solar cells is highly dependent on the interactions among multiple physical parameters. Accurately quantifying the extent to which each parameter influences the output current is critical for model calibration and performance optimization. However, the double-diode model exhibits highly nonlinear behavior and parameter coupling. The traditional One-Factor-At-A-Time (OFAT) method can only capture local effects, while global sensitivity analysis (SA) methods face the challenge of high computational costs.

This study proposes a systematic framework that integrates parameter identification, surrogate modeling, and interpretability analysis. First, particle swarm optimization (PSO) was employed for Double-Diode Models (DDM) parameter identification, yielding a model with excellent fitting performance (RMSE = 0.008508). A surrogate model was constructed using Random Forest (RF) to reduce the computational burden of the original physical model. Finally, SA and cross-validation were performed using two methods, Mean Decrease in Impurity (MDI) and OFAT.

The results show that the photocurrent *Iph* is the most significant influencing parameter, with a significantly higher importance than the other parameters. The first-order parameters n1 and Isd1 consistently rank as secondary factors, and their effects exhibit opposite trends,consistent with the underlying physical mechanisms. In contrast, the series resistance Rs and the shunt resistance Rsh have a relatively minor influence. Overall, the sensitivity results obtained from the various methods show a high degree of consistency.

This study validates the effectiveness of combining the surrogate model with MDI for SA of solar cell parameters. The proposed analytical framework combines computational efficiency with interpretability, providing physically meaningful quantitative insights for parameter optimization and fault diagnosis.

Keywords: Double-diode model, Sensitivity analysis, Surrogate model, Random forest

---

Yi-En Tang, Student, Department of Green Energy and Information Technology, National Taitung University, Email: 11222106@gm.nttu.edu.tw

I-Li Chiu, Student, Department of Green Energy and Information Technology, National Taitung University, Email: 11222129@gm.nttu.edu.tw

Yu-Tzu Chen, Student, Department of Green Energy and Information Technology, National Taitung University, Email: 11222128@gm.nttu.edu.tw

Wai-Qi Yang, Student, Department of Green Energy and Information Technology, National Taitung University, Email: 11222106@gm.nttu.edu.tw

Yan-Hao Huang (Corresponding Author), Assistant Professor, Department of Green Energy and Information Technology, National Taitung University, Email: yhhuang@nttu.edu.tw

太陽能電池雙二極體模型參數敏感度分析框架

唐以恩 、邱肖劣 、陳禹慈 、楊崴棋 、黃彥皓*

中文摘要

太陽能電池等效電路模型之性能高度依賴多項物理參數之交互作用，準確量化各參數對輸出電流之影響程度，對於模型校準與性能優化具有關鍵意義。然而，雙二極體模型具高度非線性與參數耦合特性，傳統單因子逐次變動法（One-Factor-At-A-Time, OFAT）僅能反映局部效應。而全域敏感度分析方法則面臨計算成本過高之問題。

本研究提出一套結合參數識別、替身模型與可解釋性分析之系統化框架。首先，採用粒子群最佳化進行 DDM 參數識別，所得模型具有良好擬合能力（RMSE = 0.008508）。使用隨機森林（Random Forest, RF）建立替身模型，以降低原始物理模型之計算負擔。最後使用平均不純度減少量（Mean Decrease in Impurity, MDI）與 OFAT 二種方法進行敏感度分析與交叉驗證。

結果顯示，光生電流 *Iph* 為最主要影響參數，其重要性顯著高於其他參數。第一支路參數 *n1* 與 *Isd1* 穩定列為次要因子，且其影響方向呈現相反趨勢，與物理機制一致。而串聯電阻 *Rs* 與並聯電阻 *Rsh* 則影響較小。整體而言，各方法所得敏感度結果具高度一致性。

本研究驗證了結合替身模型與 MDI 進行太陽能電池參數敏感度分析之有效性，所提出之分析框架兼具計算效率與解釋能力，可為參數優化與故障診斷提供具物理意義之量化依據。

關鍵字：雙二極體模型、敏感度分析、替身模型、隨機森林

唐以恩，國立臺東大學綠能與資訊科技學系學生，Email: 11222106@gm.nttu.edu.tw
邱肖劣，國立臺東大學綠能與資訊科技學系學生，Email: 11222129@gm.nttu.edu.tw
陳禹慈，國立臺東大學綠能與資訊科技學系學生，Email: 11222128@gm.nttu.edu.tw
楊崴棋，國立臺東大學綠能與資訊科技學系學生，Email: 11325127@gm.nttu.edu.tw
黃彥皓(通訊作者)，國立臺東大學綠能與資訊科技學系助理教授，Email: yhhuang@nttu.edu.tw

I. Background Information

As the world transitions to a net-zero carbon future, the ultimate performance of PV devices is significantly influenced by a combination of physical and process parameters. The DDM model includes parameters such as the photocurrent (*Iph*), reverse saturation current (*Isd*), diode ideality factor (*n*), series resistance (*Rs*), and shunt resistance (*Rsh*). Uncertainties in these parameters directly alter the shape of the I-V curve, thereby affecting efficiency and stability. Therefore, a quantitative understanding of the extent to which these parameters influence output performance is not only the foundation for model design optimization but also the basis for accurate diagnostics.

Sensitivity analysis (SA) is a tool used to explore the influence of PV model parameters. Early studies often employed the One-Factor-At-A-Time (OFAT) method, which involves fixing all other parameters and varying a single parameter to observe changes in the output. While OFAT has intuitive physical significance and is easy to implement, it is limited to capturing only local variations and cannot reflect the interactive relationships between input factors rendering it of only limited significance in highly nonlinear PV cell models.

With the continuous advancement of computational resources and machine learning (ML) technologies, using data surrogates has become an effective approach to alleviating computational burdens. Among these, the Random Forest (RF) algorithm is suitable for use as a surrogate model for PV cells due to its natural ability to perform nonlinear segmentation, capture strong parameter interactions, and demonstrate high stability. However, ML models are often viewed as “black boxes” because their internal workings cannot be understood through physical intuition. Therefore, in this study, the Mean Decrease in Impurity (MDI) metric built into the RF algorithm can be used to explain the global importance ranking.

As an advanced exploration, this study establishes a framework for interpreting parameters through cross-domain validation. By introducing MDI as the primary analytical tool, the study returns to the fundamental physical significance of the cell parameters. This approach, which prioritizes experimental data while incorporating physical significance, effectively demonstrates whether the parameter behaviors revealed by ML align with solar cell theory.

The contributions of this paper are as follows: (1) a systematic comparison of the performance of DDM models within a multi-indicator sensitivity framework; (2) verification of the dominance and stability of key parameters across different models; and (3) the provision of quantitative data for parameter optimization and calibration under resource-constrained conditions. This study demonstrates that even when structural changes occur, the sensitivity structures remain highly consistent, thereby providing a more interpretable pathway for the diagnosis and performance evaluation of PV systems.

The characteristics of solar cells are typically represented by equivalent circuit models to facilitate numerical simulation and performance analysis. The Double Diode Model (DDM), as shown in Figure 1, is currently the most widely used model. The DDM comprises seven main parameters: the photo-generated current *Iph*, the reverse saturation currents *Isd1*, *Isd2*, the ideal factors *n1* and *n2*, as well as the series resistance *Rs* and the shunt resistance *Rsh*, which capture the additional current loss caused by recombination effects in the space charge region under low-irradiance conditions. Compared to the single-diode model, the DDM offers higher modeling accuracy, but the difficulty of solving it also increases accordingly(Villalva et al., 2009).

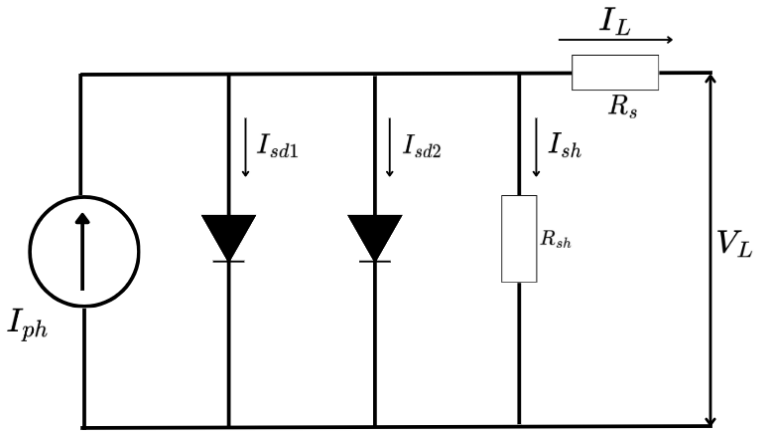

Figure 1 *DDM Model Diagram*

SA is used to quantify the extent to which each parameter affects the output current, and it is a key tool for model diagnosis and parameter optimization. OFAT is the earliest widely adopted SA method. It fixes all other parameters at their baseline values, varies a single parameter at a time, and observes the resulting changes in the output to determine the local linear sensitivity and direction of effect for each parameter. OFAT is intuitive to perform and computationally efficient. It clearly reveals the polarity of individual parameters' effects near the baseline and is suitable as a preliminary tool for directional validation(Hoshyarzadeh et al., 2026) .

However, OFAT essentially reflects only local gradient effects near the reference point; for highly nonlinear models such as the DDM, where strong interactions exist among parameters, OFAT results cannot represent the overall distribution of influence across the entire parameter space(Li et al., 2019) .

As ML technologies have matured, surrogate models have been widely adopted to address the high computational costs associated with high-dimensional nonlinear models in global SA. By learning the input-output mapping of the underlying physical model through a data-driven approach, surrogate models enable the evaluation of a large number of parameter combinations within a limited computational budget, thereby supporting subsequent SA and uncertainty quantification(Razavi & Gupta, 2016).

RF is an ensemble learning method that trains multiple decision trees using bootstrap resampling and aggregates their outputs. It effectively captures nonlinear interactions among input parameters and mitigates overfitting, demonstrating robust generalization capabilities in tasks such as solar radiation prediction, I-V curve modeling, and fault diagnosis for solar power systems(Breiman, 2001).

The key advantage of using RF s as a surrogate model for DDM lies in the fact that, once training is complete, SA can be performed efficiently on the surrogate model without the need to repeatedly solve the original system of nonlinear equations. More importantly, the structure of decision tree ensembles naturally supports the extraction of feature importance, and the built-in MDI metric provides a fast global ranking, overcoming the lack of physical interpretability found in traditional black-box models(Louppe et al., 2013; Lundberg & Lee, 2017) .

II. Method

2.1 Overview of the Research Process

This study establishes a systematic framework that integrates parameter identification and global SA, balancing calibration accuracy with parameter interpretability. First, parameter identification is performed on the double-diode model (DDM) using the Particle Swarm Optimization (PSO) algorithm to search for the optimal parameter combination within a predefined search space that minimizes the root mean square error (RMSE) between the measured current and the model-predicted current. Next, centered on the optimal PSO calibration solution, Latin Hypercube Sampling (LHS) is performed to uniformly sample the neighborhood, and a surrogate model is constructed using RF to approximate the input–output mapping relationship of the PV physical model. Finally, the MDI is adopted as the core quantitative metric, and OFAT cross-validation is performed to quantify the relative contribution and direction of effect of each parameter on the output current.

The overall workflow of the proposed framework is illustrated in Figure 2, including parameter optimization, data sampling, surrogate modeling, and sensitivity analysis.

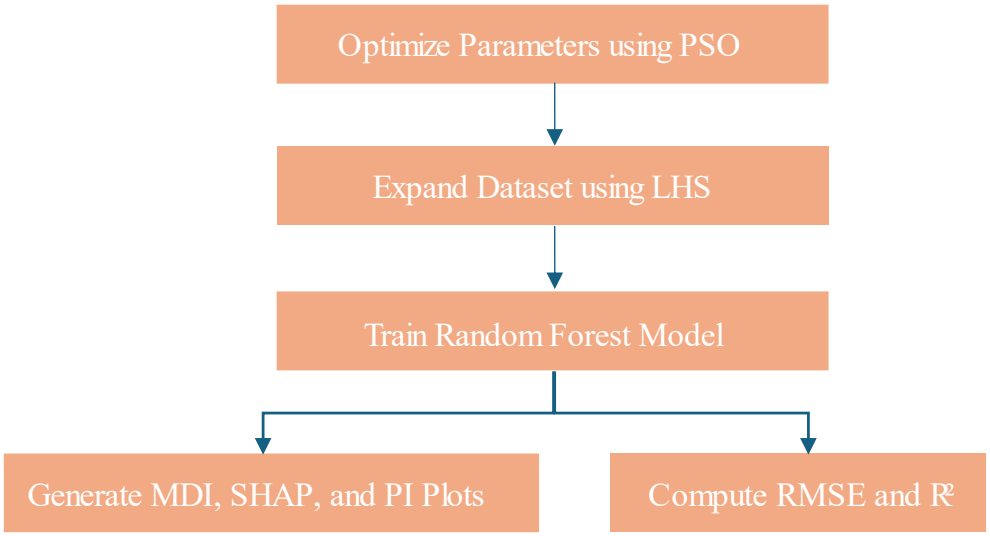

Figure 2 *Flowchart of the proposed methodology*

2.2 Particle Swarm Optimization Algorithm (PSO)

To identify parameter sensitivities, this study employs the PSO algorithm. PSO is a meta-heuristic optimization method that performs global search through the search behavior of a virtual swarm without requiring gradient information, making it a superior approach for high-dimensional and nonlinear parameter problems. PSO is a swarm-based optimization search technique(Ali et al., 2021) .Each particle represents a set of candidate parameter solutions, and their positions are continuously updated based on both local and global optimal solutions to gradually converge toward the optimal solution. PSO offers the advantages of fast convergence, simple parameter settings, and high efficiency.

The core of PSO lies in the iterative updating of particle velocity and position, where velocity is determined by a combination of the inertia term, the individual best, and the global best, and particle positions are updated accordingly to search for the optimal solution.

In the velocity update formula, the particle's velocity is composed of three components: an inertial term, guidance toward its own optimal position, and guidance toward the group's optimal position, with a random factor added to increase search diversity(Kennedy & Eberhart, 1995) . The mathematical representation is as follows:

$$V_i^{l+1} = c_0 V_i^l + c_1 r_{1,i}^l \left(X_i^{best,l} - X_i^l\right) + c_2 r_{2,i}^l (X_{swarm}^{best,l} - X_i^l) \quad (1)$$

Here V_i^l and X_i^l represent the velocity and position of a particle at the lth iteration, respectively ; $X_i^{best,l}$ and $X_{swarm}^{best,l}$ correspond to the individual particle's best position and the group's best position, respectively. c_0, c_1, c_2 These are the inertia weight, individual learning factor, and collective learning factor, which are used to adjust the movement tendencies and learning capabilities of particles during the search process ; $r_{1,i}^l$、$r_{2,i}^l$. A random number in the range [0,1] used to increase search diversity and prevent premature convergence.

The particle's position is adjusted based on the updated velocity:

$$X_i^{l+1} = X_i^l + V_i^{l+1} \quad (2)$$

This formula illustrates how a particle's position is updated to a new location by summing its velocity, thereby continuously exploring the search space for an optimal solution. Combined with the velocity formula, this approach balances global search capabilities with local convergence(Bechouat et al., 2017).

2.3 Surrogate Model Construction

To perform large-scale parameter sampling within a controlled computational budget, this study focuses on the calibrated optimal solution and performs LHS sampling within a ±5% neighborhood of each parameter (McKay et al., 2000). For the DDM model, 100 corresponding parameter vectors $\theta^{(k)}$ are generated, where k = 1, …, 100. The entire training set comprises 2,600 instances (26×100). Additionally, 650 independent test samples (26×25) were generated to evaluate the generalization capability of the surrogate model, with this ratio approximating an 80/20 training-to-test split.

RF is an ensemble learning method based on decision trees. In this study, a seven-dimensional vector of I-V and DDM model parameters is used as the input features for the RF, with the theoretical current obtained from the physical model serving as the output target. The surrogate model is trained on LHS samples within a ±5% neighborhood of the calibration optimal solution. RF hyperparameters are determined via cross-validation, and the DDM surrogate model employs 750 decision trees.

2.4 Multi-indicator sensitivity assessment framework

To comprehensively investigate the effect of PV model parameters on the output current, this study combines two different sensitivity metrics for cross-validation.

First is the MDI, a measure of feature importance built into the RF algorithm(Breiman, 2001). It quantifies the contribution of a feature during model training by measuring the reduction in node impurity achieved when splitting nodes based on that feature(Hoshyarzadeh et al., 2026).

MDI's strength lies in its high efficiency and compatibility with the learning mechanism of RF, which allows for rapid parameter selection. However, it tends to overemphasize features with high cardinality, as these are more likely to be selected as split nodes, leading MDI to overestimate the actual evaluation value(Li et al., 2019).

Next is the OFAT method, a local SA technique in which all other parameters are held at their baseline values while only the i-th parameter is varied to observe changes in the model output, thereby assessing the local sensitivity of that parameter and the direction of its influence(Saltelli et al., 2008).

However, OFAT can only reflect local linear effects near the reference point and cannot capture nonlinear interactions between parameters; therefore, this study defines it as a validation tool rather than a primary quantification method(Iooss & Lemaître, 2015).

III. Results and Discussion

3.1 Parameter Sensitivity Assessment Framework

Before conducting SA, this study first performed a per-unit value standardization procedure. Since this study uses different numerical definitions for MDI and OFAT, a direct comparison of their absolute values would lead to distorted results. The analysis using per-unit values is based on the following formula:$pu = \frac{|Value|}{\max |Value|}$.Using the maximum value of each parameter as the reference (1 pu) effectively eliminates dimensional differences, thereby verifying the consistency among different SA methods.

After completing the normalization of the target values, this study employs Pareto analysis to systematically classify the DDM model. The core logic of Pareto analysis lies in identifying the small number of key parameters that contribute the vast majority of the system's performance. This principle is described by the 80/20 rule, which states that 80% of the system's output is generated by a small number of factors, while the remaining 20% consists of a multitude of factors with negligible influence(Majka, 2024) .In this study, the model scores were quantified and ranked using the MDI and OFAT. After training the RF model, this study utilized the root mean square error (RMSE) from the independent test set to validate the approximation quality of the surrogate model relative to the physical model. The experimental results show that the DDM model achieved an optimal RMSE of 0.008508.

The minimal prediction error clearly demonstrate that the RF model accurately captures the output characteristics of the original physical model within the parameter range under consideration. This high-precision approximation not only ensures the physical credibility of subsequent sensitivity analyses but also provides a robust and generalizable data foundation for quantifying parameter influences, allowing the conclusions to be regarded as reliable global sensitivity approximations within the specified parameter range.

This study defines three order-of-magnitude categories. By combining per-unit values with Pareto analysis, the DDM model parameters are classified into three levels based on their impact intensity: primary parameters, secondary parameters, and minor parameters. Major parameters are defined as those with pu ≥ 0.5, primarily to identify a clear distinction in magnitude from secondary parameters. Based on data observations, the influence of major parameters is typically several times greater than that of secondary parameters; thus, in terms of contribution, they adhere to the Pareto principle, where a critical few account for over 80%

of the total parameter influence. Secondary parameters are defined as those with pu between 0.05 and 0.5. From an analytical perspective, while the influence of secondary parameters is less than that of primary parameters, they are crucial for capturing nonlinear inflection points. Minor parameters are those with a standard score below 0.05. These model parameters contribute less than 5% to the overall contribution and are considered to be of extremely low importance.

3.2 Sensitivity Analysis of the DDM Model

Table 1
Sensitivity analysis results of DDM parameters *(MDI and OFAT)*

	Iph	*n1*	*Isd1*	*n2*	*Isd2*	*Rsh*	*Rs*
MDI	0.0376	0.0115	0.0111	0.0108	0.0103	0.0095	0.0093
OFAT	0.0116	0.0025	0.0024	0.0003	0.0002	0.0001	0.0000

Table 1 presents the results of a SA of the DDM model's global importance for MDI and OFAT. The values of the two metrics are consolidated and presented, and the parameters are ranked using per-unit values and Pareto analysis.

Regarding the MDI results, there are significant differences in the distribution of feature importance among the seven DDM parameters. *Iph* has the highest value, reaching 0.0376, far exceeding the other parameters, followed in order by *n1* (0.0115), *Isd*1 (0.0111), *n2* (0.0108), *Isd*2 (0.0103), *Rs* (0.0095), and *Rsh* (0.0093). Using *Iph* as the reference (1 pu) for standardization, the results are: *n1* = 0.306 pu, *Isd*1 = 0.295 pu, *n2* = 0.287 pu, *Isd*2 = 0.274 pu, *Rs* = 0.253 pu, and *Rsh* = 0.247 pu.

By combining per-unit values and Pareto analysis, this study defines a classification criterion: primary parameters are those with pu ≥ 0.5; secondary parameters are those with 0.05 ≤ pu < 0.5; and minor parameters are those with pu <0.05. It can be seen that *Iph* is the sole primary parameter, with an order-of-magnitude difference of approximately 3.2 times compared to the secondary parameters; the pu values of the remaining six parameters (*n1*, *Isd*1, *n2*, *Isd*2, *Rsh*, and *Rs*) all have pu values falling between 0.05 and 0.5, and are all classified as secondary parameters. Furthermore, the differences among them are similar, indicating that the DDM model structure tends to equalize the relative influence of each parameter.

The OFAT analysis results also show that the *Iph* value is the most prominent, reaching 0.0116, which reflects its dominant influence on the output current near the local reference point. The next most significant values are *n1* and *Isd*1, at 0.0026 and 0.0024, respectively, while the values for *n2* and *Isd*2 are relatively lower, at 0.0003 and 0.0002, respectively. Finally, the values of *Rs* and *Rsh* are -0.0001 and 0.0000, respectively, indicating that they make almost no significant contribution within the considered neighborhood. After performing the per-unit conversion with *Iph* as the reference, *n1* and *Isd*1 are 0.216 pu and 0.207 pu, respectively, *n2* and *Isd*2 are 0.026 pu and 0.017 pu, respectively, both below 0.03 pu, while *Rsh* and *Rs* are less than 0.01 pu. According to the classification criteria, *Iph* is the primary parameter, *n1* and *Isd*1 are classified as secondary parameters, and the rest are categorized as minor parameters.

3.3 Discussion

By synthesizing the results of the two SA methods, MDI and OFAT, this study arrived at highly consistent conclusions regarding parameter ranking within the DDM model: *Iph* was identified as the leading parameter across all method and model combinations by a substantial margin, with its influence far exceeding that of secondary and minor parameters. This aligns with the core characteristic of Pareto Analysis, wherein "a critical few dominate system output." In the DDM model, the seven parameters of the DDM show a more evenly distributed importance compared to a simpler single-branch structure. *Iph* retains its status as the primary parameter. The two methods maintain consistency in the relative ranking of parameters, indicating that the SA framework possesses robustness across model structures.

IV. Conclusion

This study establishes a parameter sensitivity framework centered on PSO-based parameter identification, with RF serving as both a surrogate model and an interpretive tool, and supplemented by MDI and OFAT for cross-validation. The framework systematically quantifies the magnitude and direction of the effects of each parameter in the DDM model on the output current.

In terms of parameter identification, the PSO algorithm demonstrated stable convergence in the nonlinear search space of the seven DDM parameters, achieving a minimum RMSE of 0.008508. The identified parameter combination was able to accurately reproduce the experimentally measured I-V curves, thereby establishing a physically reasonable baseline for subsequent SA.

In terms of SA, the ranking of per-unit values obtained using the MDI and OFAT methods showed a high degree of consistency, both pointing to the following conclusions. First, *Iph* is the overwhelmingly dominant parameter in the DDM model. Second, the parameters *n1* and *Isd*1 consistently rank as secondary parameters. Furthermore, the DDM model parameters *n2* and *Isd*2 also fall within the range of secondary parameters, reflecting that the double-diode structure introduces quantifiable multi-parameter coupling effects. Finally, the series and parallel resistances *Rs* and *Rsh* are minor parameters within the considered domain, and their global impact on the output current can be neglected.

Overall, this study demonstrates the effectiveness of using ML surrogate models combined with interpretability tools to perform SA of PV parameters under limited computational resources. The analytical framework established in this study exhibits physical consistency and cross-metric robustness, providing a quantifiable and easily interpretable analytical foundation for the parameter optimization and calibration, performance diagnosis, and fault identification of solar cells.

References

Ali, I. B., Naouar, M. W., & Monmasson, E. (2021). Parameters identification for a photovoltaic module: comparison between PSO, GA and CS metaheuristic optimisation algorithms. *International Journal of Modelling, Identification and Control*, *38*(3-4), 211–230.

Bechouat, M., Younsi, A., Sedraoui, M., Soufi, Y., Yousfi, L., Tabet, I., & Touafek, K. (2017). Parameters identification of a photovoltaic module in a thermal system using meta-heuristic optimization methods. *International Journal of Energy and Environmental Engineering*, *8*(4), 331–341.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Hamby, D. M. (1994). A review of techniques for parameter sensitivity analysis of environmental models. *Environmental monitoring and assessment*, *32*(2), 135–154.

Hoshyarzadeh, R., Hafshejani, L. D., Tishehzan, P., & Pa*Rs*aie, A. (2026). A hybrid framework of feature selection and interpretability for dissolved oxygen prediction in drinking water treatment plants. *Scientific Reports*.

Iooss, B., & Lemaître, P. (2015). A review on global sensitivity analysis methods. *Uncertainty management in simulation-optimization of complex systems: algorithms and applications*, 101–122.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization (PSO). Proc. IEEE international conference on neural networks, Perth, Australia,

Li, X., Wang, Y., Basu, S., Kumbier, K., & Yu, B. (2019). A debiased MDI feature importance measure for random forests. *Advances in Neural Information Processing Systems*, *32*.

Louppe, G., Wehenkel, L., Sutera, A., & Geurts, P. (2013). Unde*Rs*tanding variable importances in forests of randomized trees. *Advances in Neural Information Processing Systems*, *26*.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Majka, M. (2024). Pareto analysis: The 80/20 rule for effective decision-making. In: Researchgate.

McKay, M. D., Beckman, R. J., & Conover, W. J. (2000). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *42*(1), 55–61.

Razavi, S., & Gupta, H. V. (2016). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research*, *52*(1), 440–455.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: the primer*. John Wiley & Sons.

Villalva, M. G., Gazoli, J. R., & Ruppert Filho, E. (2009). Comprehensive approach to modeling and simulation of photovoltaic arrays. *IEEE Transactions on power electronics*, *24*(5), 1198–1208.