Automatic Load Model Selection Based on Machine Learning Algorithms

S. HERNÁNDEZ-PEÑA, S. PÉREZ-LONDOÑO, (Member, IEEE), AND J. MORA-FLÓREZ
Department of Electrical Engineering, Universidad Tecnológica de Pereira, Pereira 660003, Colombia
Corresponding author: S. Pérez-Londoño (saperez@utp.edu.co)

This work was supported by the Colombian Science Ministry (MinCiencias) under grant INTEGRA2023 (111085271060 CT-774-2020) and grant (111077657914 CT-031-2018) and by the VIE-Universidad Tecnológica de Pereira under research project (6-21-8).

ABSTRACT Technology development and decentralized operations create changes in conventional electric systems, where load modeling has been a challenge in dynamic analysis. Consequently, accurate dynamic load models are required to ensure the quality of the studies in current systems. This paper presents an automatic strategy based on clustering, classification, and optimization algorithms, to obtain the load models in the case of several system operating conditions. The obtained load models are helpful for the planning, operation, and protection of electric power systems. The proposed approach validation is performed using the IEEE 14-bus test system, where high performance is obtained. The average obtained cross-validation error for the load models assigned to the 13 clusters of disturbances is $5.36 \times 10^{-3}$. The cross-validation error is used as a tolerance value to determine when an online assigned load model is suitable to represent the measured disturbance. The proposed tests show the strategy’s capabilities of defining the load model online, making this approach suitable for field applications.

INDEX TERMS Dynamic load modeling, generalization capability, machine learning, parameter identification.

NOMENCLATURE

A. ACRONYMS

ADN Active Distribution Network.
ANN Artificial Neural Networks.
CL Composite Load model.
CMPLDW WECC Dynamic Composite Load Model.
EL + IM Exponential Load model and Induction Motor.
ERL Exponential Recovery Load model.
kNN k-Nearest Neighbors.
PSO Particle Swarm Optimization.
RBF Radial Basis Function.
RMS Root Mean Square.
SSE Sum of Squared Error.
SVM Support Vector Machine.
WECC Western Electricity Coordinating Council.
WSSD Within-Cluster-Sum of Squared Distances.
ZIP Polynomial load model.

B. VARIABLES

$e_{D_{k,1}}$ Error estimation for the first of the $V$ folds.
$P_d$ Measured signal of active power during disturbance $d$.
$P_e$ Estimated signal of active power.
$P_i$ Measured signal of active power.
$Q_d$ Measured signal of reactive power during disturbance $d$.
$Q_e$ Estimated signal of reactive power.
$Q_i$ Measured signal of reactive power.
$\text{par}^*$ Averaged optimal load model parameters.
$C$ Centroid of the cluster $k$.
$CrossE$ Cross-validation error.
$D$ Disturbances at the database.
$d$ Disturbance signal composed by sampled signals of voltage, active and reactive power.
$D_k$ Disturbances at the cluster $k$.
$E_m$ Error function to determine the differences between the measured and the estimated active and reactive power.

The associate editor coordinating the review of this manuscript and approving it for publication was Reinaldo Tonkoski.
Regarding the former question, several models (black-box modeling), among other aspects [1], [2]. Load modeling has been a research topic due to its particular aspects, such as the type, composition, variability, and diversity of loads. Many strategies have been published, considering different approaches (component-based or measurement-based modeling) and mathematical structures (static, dynamic, or composite) or their physical (induction motor equations) or non-physical meaning (data-based approach). These models are not fully automated for online model determination. As shown in Table 1, most proposed approaches focus on applying machine learning to identify load composition or specific load models that only perform correctly under specific operating conditions. Unfortunately, these strategies are not fully automated for online model determination.

### B. STATE OF THE ART - COMPARISON TO THE RELATED PROBLEM

Several approaches have been considered for load modeling; however, this paper focuses only on the measurement-based load modeling approach enhanced by machine learning and optimization algorithms.

Machine learning algorithms are widely used to address several problems in power systems, including protection and fault location [7], [8], voltage stability [9], [10], [11], [12], and operation and control, among others, [11], [13], [14], [15], [16].

Supervised learning algorithms have been applied in load forecasting in several studies [17], [18], [19], [20], [21]. However, according to the literature review summarized in Table 1, few papers consider a combination of data-based learning approaches to obtain an automatic adaptive load model, useful for system analysis under different disturbances. In [22] a k-means based clustering defines the consumption patterns used as a training set to obtain generalized load models constructed by RBF neural networks. Zhang et al. [23] used k-medoids clustering to analyze the parameter dependency for the complex WECC dynamic composite load model (CMPLDW). Kontis et al. [25] proposed a statistical analysis and an ANN-based approach to obtain generic dynamic load models for different operating conditions. Another approach based on ANN-based modeling is found in [26], where the data of an induction motor load considering several operating conditions were used to develop a black-box model. In [27], clustering and ANN-based strategies were proposed to derive parameters for a variable-order dynamic equivalent model. Wang et al. [28] used the support vector machine (SVM) approach to propose a parameter identification for composite ZIP and electronic loads. Liang et al. [29] proposed SVM to identify parameters for a differential equation-based dynamic load model structure.

As shown in Table 1, most proposed approaches focus on applying machine learning to identify load composition or specific load models that only perform correctly under specific operating conditions. Unfortunately, these strategies are not fully automated for online model determination.

### C. CONTRIBUTIONS

The measurement-based load modeling approach requires a time-consuming parameter estimation process; thus, a strategy based on clustering and classification algorithms is proposed to reduce the computation effort in determining suitable load models for online applications. This allows the automatic definition of a parameterized load model that best represents the online measured power disturbance. The

---

**I. INTRODUCTION**

**A. MOTIVATION**

Load modeling has been a research topic due to its particular aspects, such as the type, composition, variability, and diversity of loads. Many strategies have been published, considering different approaches (component-based or measurement-based modeling) and mathematical structures (static, dynamic, or composite) or their physical (induction motor equations) or non-physical meaning (data-based black-box modeling), among other aspects [1], [2].

Two questions frequently appear when load modeling is analyzed: *Which model best represents the load behavior?* and *How can the model’s generalization capability be improved?* Regarding the former question, several models can represent some of the required characteristics in specific applications, e.g., static load models for the power flow calculation [3], dynamic load models for transient, and voltage stability [1], and black-box or grey-box models for active distribution networks (ADN) [4], [5].

On the other hand, due to the load variability both in composition and power demand, it is difficult to guarantee that the same load model adequately represents operational conditions that are different from those used in the parameter estimation when the measurement-based load approach is applied. This characteristic is known as generalization capability and must be evaluated for each load model. Usually, the generalization capability is tested for disturbances similar to those used in the model parameter estimation, which is not a recommended test of the load models [6]. Otherwise, in conventional distribution systems, a load model type is generally defined for a specific bus, and its parameters are not changed regardless of the studies performed. However, considering the stochastic nature of distributed energy sources, the everyday use of power converters, and the characteristics of modern loads, new modeling approaches must be developed that allow one to shift between the model types according to the load time-varying composition. This paper proposes an automatic load model selection strategy based on machine learning and optimization algorithms to solve these problems. The proposed strategy obtains the best load model to analyze the electric system, considering different disturbances and operating conditions.
proposed approach offers a quick and reliable load model to perform forensic or planning studies.

II. PROPOSED STRATEGY
Supervised learning is related to using labeled datasets to train algorithms that later classify data or predict outcomes accurately [30]. This feature is used in the proposed approach to select the best suitable parameterized load model online from a set of predefined models. As load model determination through an estimation process can be time-consuming, this strategy reduces the effort using a two-stage methodology. The offline stage obtains several load models for a power system disturbance database through the measurement-based approach. Then, using clustering and cross-validation techniques, the load model that best fits is selected for each disturbance at the database. For a new disturbance, the online stage quickly determines the load model, using the information obtained in the offline stage and a classification technique that reduces the computation time required to obtain new load models.

Figure 1 shows the general framework of the proposed strategy. The stages and respective steps involved are described below.

A. OFFLINE STAGE: ESTIMATION, CLUSTERING AND SELECTION
The main objective in this stage is to determine several load models for a disturbance, applying a measurement-based approach, optimization, and clustering algorithms. Finally, using cross-validation, the best load model is selected. This section contains five steps, including the disturbance database determination, normalization of the database, clustering of the normalized disturbance database, parameterization of the selected load models, and the selection of the best representative model to describe the load behavior for each cluster of disturbances. All of the steps are explained below.

1) STEP 1. DETERMINATION OF THE DISTURBANCE DATABASE
The database contains RMS voltage and power (active and reactive) measurements at the load bus, obtained during disturbances at the analyzed system or from simulation responses (synthetic data). The database must contain a wide variety of large and small disturbances to achieve diversity and successfully apply machine learning tools. This strategy considers $D$ disturbances caused by shunt faults, lines/loads switching, and variations of the transformer tap changer, for example.

2) STEP 2. NORMALIZATION OF THE DISTURBANCE DATABASE
Once the database has been obtained, a normalization process of $3 \times D$ signals (voltage and active/reactive power) is performed before using the machine learning tools. The objective is to change the data set values to a standard scale without distorting the differences in the ranges of values. In general, for machine learning, the data set does not require normalization; however, it is required in the case of three measurements in different scale ranges, such as the voltages and active and reactive power signals used in this proposed approach.

There are several normalization strategies as min-max, means, median, and stdv [31]. With this strategy, the min-max normalization algorithm is applied, considering adequate results obtained in similar engineering applications. The normalized voltage and active and reactive power signals are merged for each disturbance at the database, into one single vector as presented in (1), to generate $D$ normalized sampled signals. This set is used as input for the clustering algorithm.

$$S_d = [V_d, P_d, Q_d] \ \forall \ d \in D$$ (1)

3) STEP 3. CLUSTERIZATION OF THE DATABASE DISTURBANCE
A clustering step is applied to group the available $D$ normalized sampled signals based on their similar behavior. The $k$-means algorithm is used in this approach, considering its simplicity and adequate performance obtained in similar clustering approaches, although the strategy can use any clustering approach [32]. This unsupervised clustering method is used to differentiate samples based on their similarity [33].

This algorithm aims to minimize an objective function, in this case, a squared error function, as shown in (2):

$$\min_{c_1, \ldots, c_k} \frac{1}{K} \sum_{k=1}^{K} \sum_{d=1}^{D} \|S_d - c_k^2\|^2, \ \forall \ d \in D$$ (2)

where $c_k$ is the centroid of the cluster $k$.

A fair value of the number of clusters ($K$) is defined using several strategies, such as the elbow and silhouette methods [34]. The elbow method, used in this paper, is based

| Reference | Machine learning technique | Load Model | Online |
|-----------|---------------------------|------------|--------|
| [22]      | ✓                        | Generalized L. | - |
| [23]      | ✓                        | CMPLDW     | - |
| [24]      | ✓                        | CMPLDW     | - |
| [25]      | ✓                        | ERL        | - |
| [26]      | ✓                        | Black box  | - |
| [27]      | ✓                        | Generalized L. | - |
| [28]      | ✓                        | ZIP, Electronic m. | - |
| [29]      | ✓                        | Z-Transfer function | - |
| Prop. Strategy | ✓          | CL, EL, IM, ZIP, ERL | ✓ |
on calculating the Within-Cluster-Sum of Squared Distances (WSSD) to the cluster centroid, for different values of \( K \), and choosing the value for which WSSD becomes first starts to diminish. In the plot of WSSD-versus-\( K \), this is visible as an elbow, it defines the fine value of \( K \) [35].

4) **STEP 4. PARAMETER ESTIMATION FOR LOAD MODELS**

Simultaneous with the normalization and clustering steps, the database performs an estimation process for each \( d \) disturbance. The measurement-based load modeling requires an initial selection of the model structure for the identification and parameter fitting; then, several load models are considered because each disturbance is better represented by a specific load model [1]. The best model for each disturbance is automatically selected to avoid delegating the model selection to the power system analyst. \( M \) models are considered in this strategy to represent the static and the dynamic load behavior.

The load model parameter estimation is performed in this step by using the particle swarm optimization technique (PSO), considering the adequate performance reported in other applications [36], [37]. Despite the previous, the approach presented in this paper is not constrained by any specific optimization algorithm; thus, other options can be considered.

The accuracy of the estimated and parameterized load model is assessed based on the error function given in (3). This considers the difference between the power measured at the load bus \((P_s, Q_s)\), and the power estimated using the parameterized model \((P_e, Q_e)\) for each model \( m \), set of model parameters \( \text{par} \), and a specific disturbance \( d \).

\[
E_{\text{in,par,d}} = \frac{1}{J} \sum_{j=1}^{J} \left[ (P_{sj} - P_{ej})^2 + (Q_{sj} - Q_{ej})^2 \right]^{\frac{1}{2}}
\]  

(3)
\[ J \] is the total number of samples at the measured signals. The load model parameters \( \text{par} \) are updated through the optimization algorithm to minimize the error. Finally, the optimal parameters for each disturbance are defined as \( \text{par}^* \).

At this step and for each disturbance \( d \), the optimal parameters for the \( M \) load models are obtained.

5) STEP 5. SELECTION OF THE MOST REPRESENTATIVE MODEL FOR EACH CLUSTER

As previously explained, the set of disturbances \( D \) has been grouped into \( K \) clusters. Each cluster contains several disturbances \( (D_k) \) with similar behavior as presented in Step 3, and each disturbance \( d \) has been represented by \( M \) load models as described in Step 4. Therefore, each cluster has \( D_k \times M \) load models.

Considering the previous explanation, Step 5 is oriented to determine the most representative load model \( m_k^* \) for each cluster \( k \). The cross-validation technique is selected in this step because it allows an adequate estimation of the error defined in (3). The selection of the best load model using the cross-validation is depicted in Figure 2.

For cross-validation, the disturbance set in cluster \( k \) \( (D_k) \) is divided in \( F \)-folds. The method is repeated \( F \) times, such that each time, one of the \( F \) folds is used as the validation set, and the other \( F - 1 \) subsets are grouped to form a fitting set to define models with optimal average parameters. The error estimation for the first fold is presented in (4), where \( d \) represents the disturbances in each fold \( D_{k,f} \) of cluster \( k \) and \( \text{par}^* \) corresponds to the average parameters of each model for each one of the disturbances at the \( F - 1 \) folds.

\[
e_{D_{k,1}} = \frac{1}{D_{k,1}} \sum_{d=1}^{D_{k,1}} E_{m, \text{par}^* , d} \tag{4}
\]

The error estimation is averaged over all \( F \) trials to obtain the total effectiveness of the model \( m \) for representing the disturbances in cluster \( k \), as presented in (5). This reduces bias, as the data are used for fitting, and it simultaneously reduces variance, as the data are also used in the validation set.

\[
\text{CrossError}_{m,k} = \frac{1}{F} \sum_{f=1}^{F} e_{D_{k,f}} \tag{5}
\]

The criterion used to define the representative model of each cluster \( k \) is that the model must have the minimum cross-validation error, according to (5). The parameters of this model are obtained using the average of the model parameters for the \( D_k \) disturbances.

B. ONLINE STAGE- AUTOMATIC LOAD SELECTION

In the online stage, the information obtained in the offline stage (clustering, parameter estimation, and model selection) is used to derive a load model automatically when a new disturbance is presented. This strategy reduces the computation time required to obtain load models when a measurement-based approach is used. The proposed approach automates the load model selection, giving a significant advantage over other proposals, which generally require the continuous execution of the parameter estimation process for each new disturbance.

1) STEP 6. NORMALIZATION OF THE ONLINE RECORDED DISTURBANCE DATA

In this step, a signal processing phase is applied to each new disturbance, similar to that performed in the offline stage. The voltage and active and reactive power measurements are normalized using the strategy in Section II-A2 and considering the same maximum and minimum values used at the offline stage; this strategy avoids errors caused by comparing signals normalized using different scales.

This step allows the new disturbance signal to be prepared, leading to the classification step.

2) STEP 7. CLASSIFICATION OF THE DISTURBANCE

A distance-based classification algorithm known as the k-nearest neighbors (kNN) is applied to automatically associate the normalized measurements representing the new disturbance to a specific cluster previously defined in the offline stage. This cluster’s best load model (representative model) is selected to represent the load behavior in the new disturbance.

The classification algorithm is used due to its simplicity, and adequate results for a wide range of applications [38]. As in the case of the clustering algorithm, the proposed methodology is not constrained to using a specific classification method.

3) STEP 8. LOAD MODEL VALIDATION

Once a parameterized load model (the representative model) is associated with the online acquired disturbance, the model performance is evaluated through the error presented in (3). The results of this process are compared to the defined reference error \( \text{tol} \). This paper uses the error for each cluster obtained from the cross-validation as \( \text{tol} \). In the case of an error higher than the reference, the offline obtained load models do not adequately represent the actual disturbance behavior; then, this disturbance serves to update the database as described in the next step. Otherwise, the proposed approach quickly offers a reliable load model for forensic or planning studies.

4) STEP 9. ORIGINAL DISTURBANCE DATABASE UPDATE

When the behavior of the assigned load model has a low similitude with the corresponding disturbance, this measurement is included in the disturbance database, and the complete offline stage is performed. Otherwise, the assigned model adequately represents the new disturbance, and the database need not be updated.
III. RESULTS AND ANALYSIS

A. TEST SYSTEM

Methodology validation is performed using several operating conditions at the IEEE 14-bus system. This system has five synchronous machines with IEEE type-1 exciters and three synchronous compensators used for reactive power support; it additionally contains 11 loads (259 [MW] and 81.3 [MV AR]). The dynamic data for the generator exciters are obtained from [39], and the static data are obtained from [40].

B. OFFLINE STAGE: ESTIMATION, CLUSTERING AND SELECTION

1) STEP 1 - DETERMINATION OF THE DISTURBANCE DATABASE

The dynamic load behavior is captured through voltage and active and reactive power signals. A total of 319 disturbances are included in the database \( D = 319 \). The disturbances consider shunt faults, load size variation, and the connection or disconnection of generators and lines.

2) STEP 2 - NORMALIZATION OF THE DISTURBANCE DATABASE

Sampled voltage and active and reactive power signals are normalized using the min-max strategy. The maximum and minimum values for each signal are stored, as these are required in the online stage.

Next, the normalized signals of voltage and active and reactive power are merged for each disturbance at database \( S_d = \{ V_d, P_d, Q_d \} \) to generate \( D \) normalized sampled signals. Figure 3 depicts some of the merged normalized sampled signals; it is worth noting that diversity of disturbances is needed (from small to high voltage disturbances) to represent the expected system behavior adequately.

3) STEP 3 - CLUSTERIZATION OF THE DISTURBANCE DATABASE

At this stage, clustering is applied to assemble the available normalized signals into \( k \) groups. Based on the elbow criterion presented in section II-A3, an optimal value of \( k = 13 \) is obtained. As presented in Figure 4, the clustering algorithm adequately differentiates the load response’s behavior in the case of each disturbance. Some clusters contain small perturbations, such as cluster 12; others, such as clusters 4 and 10, contain oscillatory signals or high variations of the normalized signal.

4) STEP 4 - PARAMETERIZATION OF THE LOAD MODELS

As described in Section II-A4, for each disturbance \( d \) represented by the sampled signals of \( V, P \) and \( Q \), the selected \( M \) load models are parameterized.

Four load models \( (M = 4) \), which represent static and dynamic behavior, are considered in this paper: the polynomial load model (ZIP), the exponential recovery load (ERL), the classical composite model (CL), and the induction motor with an exponential model (EL+IM). These models are adequately described in [41], [42], [43] and [44] and summarized in the appendix.

In the proposed tests, as \( D = 319 \) and \( M = 4 \), 1276 load models are estimated. This process involves a high computation time; however, it is performed offline.

5) STEP 5 - SELECTION OF THE MOST REPRESENTATIVE MODEL OF EACH CLUSTER

Only a load model is selected to represent the disturbances in a cluster at this stage. This is the most representative model of the cluster, and it is selected considering the cross-validation strategy presented in Section II-A5. As a consequence, the error estimated according to (5) for each load model in each cluster is presented in Figure 5. In this figure, the blue-colored
FIGURE 3. Examples of the normalized signals from the database for the IEEE 14-bus system.

FIGURE 4. 13 clusters obtained for the disturbance database.

The following concluding remarks are derived from this figure:

- The ZIP model is the best representative model for six of the 13 clusters (1, 4, 8, 10, 11, and 12), according to the calculated error (minimum cross-validation error). Although there are no significant differences for some models, it is highlighted how the ZIP model does respond with good precision to severe disturbances as these are included in clusters 1, 4, 8, 10, and 11. However, for some critical disturbances, such as those included in cluster 5, dynamic load models such as EL+IM show better response because the disturbances present high signal variations from the pre-disturbance value. It is essential to note that these conclusions should not be generalized because they are related to the system behavior under specific disturbances. In power systems with different characteristics (topology, compensation systems, and size, among others), the response will be different, and therefore other models can better represent the load behavior. This argues that a load model used correctly for one specific condition will not necessarily suit another.

- Although three dynamic load models are considered, each one has unique characteristics that allow a specific load model to be more appropriate under specific operating conditions (more or minor oscillations in the load response, a fast load response, and low drop voltage, among others). For instance, cluster 5 can...
be better represented for an EL+IM model than an ERL model.

• When there is a similar error among load models, the computation time is used as a selection criterion. This time is usually associated with the number of parameters. For instance, cluster 6 can be represented by ERL or EL+IM models (if the CL load model is not available) due to the similarity between errors; however, ERL has fewer parameters (6 parameters) than EL+IM (11 parameters), representing a reduction in computation time, especially in huge power systems.

• The selected load model must suit the proposed application. For example, a dynamic load model is preferred over a static one for stability analysis, and a static load model is preferred when a power load flow is executed.

According to these observations, it is worth developing adaptive approaches to derive measurement-based equivalent models to manage the changing conditions of the electric systems.

Finally, the minimum cross-validation error explained in Section II-A5 is applied to define the load model’s parameter. The parameters obtained for the best load model for each cluster are presented in Tables 2 to 5.

C. ONLINE APPLICATION OF THE PROPOSED STRATEGY

The proposed strategy is validated using ten new disturbances obtained from the IEEE 14-bus system, not previously used in the offline stage. Variations in load size and voltage disturbances are considered to generate additional operating conditions not considered in the offline stage.

1) STEP 6 - PRE-PROCESSING OF THE ONLINE RECORDED DISTURBANCE DATA

Once a new disturbance is presented, signal normalization is performed using the same minimum and maximum values obtained in the offline stage, as mentioned in Section III-B2.

As a result of the data processing, the normalized signals representing each disturbance are shown in Figure 6. This figure includes the normalized time-domain signals of the RMS values of voltage and active and reactive power.

2) STEP 7 - CLASSIFICATION OF THE ONLINE RECORDED DISTURBANCES

The assignment of a load model to a new disturbance in an agile and fast way, avoiding the parameter estimation process (a time-consuming process), is performed by applying the classification task explained in Section II-B2. The results for the online disturbance classification are presented in Table 6. According to the results, each disturbance is assigned to one cluster, and consequently, the respective average model defined in the offline stage for that cluster is selected. As shown in Table 6, although the ZIP or ERL models are assigned to several of the new disturbances, each one has different parameters (e.g., the ZIP model parameters for disturbance A are different from the ZIP model parameters for disturbances F, G, or J) because each model corresponds to a different cluster that groups similar load responses.
to different operating system conditions. This variability of the load behavior under several disturbances reinforces the applicability of the adaptive approach presented in this paper.

3) STEP 8 - LOAD MODEL VALIDATION

Once a load model is assigned to the new disturbances, the model’s adaptability to this signal is evaluated. Table 6 shows the results for model assignment to each disturbance and the error obtained when using the model assigned (column four) to represent the new disturbance. The error calculated according to (3) shows that it is possible to obtain load models with an appropriate fit. This table shows that, although the ZIP load model can represent four of ten disturbances, sometimes dynamic load models must be incorporated, such as the EL+IM (disturbance B) or ERL model (disturbances C, D, E, and I).

In all of the previous tests, the load is strictly constant (magnitude and composition) at the bus where the measurements are obtained and is composed of induction motors. It appears evident that if a bus has a fixed load, only one model can represent this load; however, as demonstrated, the load response varies according to the system operating conditions. The previous situation proves that considering a fixed load model for a bus would not be appropriate for several operating conditions.

Table 6 also includes the minimum error model obtained with other average load models. Although the selected model’s error is higher than the minimum error given by other models in three disturbances (A, F, and I), these errors are very close (percentage error of approximately 3.7% for A and I disturbances, 2% for disturbance F).

4) STEP 9 - UPDATING THE ORIGINAL DISTURBANCE DATABASE

Finally, to demonstrate the applicability of the proposed strategy, Figure 7 depicts the disturbance H classified in cluster 9 and the estimated load model response obtained with the assigned load model (in this case CL); the other two load models’ responses are used for comparison. As expected, cluster 9 assigns the CL load model, which allows a better adjustment to the disturbance than the ZIP or ERL load models, which are unable to follow the system’s power response.
in the offline stage. In this case, the analyzed disturbance is included in the disturbance database, and the entire offline stage is repeated, updating the clustering and the corresponding load models.

As shown in Table 6, the disturbances A, F, and G meet the previously exposed criterion. The error analysis determines a value of $2.55 \times 10^{-3} > 1.2 \times 10^{-3}$ for disturbance A, $5.32 \times 10^{-4} > 1.82 \times 10^{-4}$ for disturbance F, and finally, $1.08 \times 10^{-3} > 4.3 \times 10^{-4}$ for disturbance G. According to the obtained results, disturbances A, F, and G have to be included in the offline stage to update the disturbance database.

### IV. CONCLUSION

The characteristics of the current power systems, where new elements and technologies appear, increase the importance of achieving adequate load models according to the study type. Although the measurement-based load approach is extensively used, considering the availability of measurements in an electric system, these models’ adaptability is limited. A load model obtained from specific data measurements may not be adequate to represent other operating scenarios since it may lack the ability to fit new data measurements.

This paper presents a strategy that uses optimization, classification, and prediction algorithms to determine the suitable load model. The two-stage strategy allows the offline definition of several parameterized load models for each system bus. In the case of an online measured disturbance, the strategy speedily selects the best suitable load model in a low computational burden process. This is a remarkable advantage because the proposed strategy can be extended to online monitoring schemes.

According to the obtained results in the proposed tests, applying the proposed approach improves the online estimated load models’ generalization capability. These models usually are required for adequate control, operation, protection, and planning studies in electric systems.

### APPENDIX. LOAD MODELS

#### A. ZIP MODEL

The ZIP model is commonly used as a static load model and can be described by (A.1):

$$P_{ZIP} = P_0 \left[ a_0 + a_1 \left( \frac{U}{U_0} \right) + a_2 \left( \frac{U}{U_0} \right)^2 \right]$$

#### B. COMPOSITE LOAD MODEL (CL)

The composite load model (CL) combines a static polynomial load model ZIP and a third-order induction motor model (IM) as presented in (B.1) and (B.2) [41]:

$$\frac{dE_d'}{dt} = -\frac{1}{T_s} \left[ E_d' + (X - X') I_q \right] - (w - 1) E_q'$$

$$\frac{dE_q'}{dt} = -\frac{1}{T_s} \left[ E_q' + (X - X') I_d \right] + (w - 1) E_d'$$

$$\frac{dI_d}{dt} = -\frac{1}{2H} \left[ T_0 - X' \left( E_d' I_d + E_q' I_q \right) \right]$$

$$I_d = \frac{1}{R_s^2 + X_s^2} \left[ R_s \left( U_d - E_d \right) + X' \left( U_q - E_q \right) \right]$$

$$I_q = \frac{1}{R_s^2 + X_s^2} \left[ R_s \left( U_q - E_q \right) - X' \left( U_d - E_d \right) \right]$$

where

$$T' = \frac{X_r + X_m}{R_r} \cdot \dot{X} = X_0 + X_m \cdot \dot{X}' = X_0 + \frac{X_m X_r}{X_m + X_r}$$

and

$$a_0 + a_1 + a_2 = 1; \quad b_0 + b_1 + b_2 = 1$$

where $a_0$, $a_1$, $a_2$, $b_0$, $b_1$, and $b_2$ are proportional coefficients of the constant impedance, constant current and constant power, respectively, for the static active and reactive load. $P_0$ and $Q_0$ are the rated powers, and $U_0$ is the pre-disturbance voltage.

### TABLE 6. Load models selected for the ten online acquired new disturbances.

| Disturbance | Classification | Model assigned | Error | Minimum error model | Error |
|-------------|----------------|----------------|-------|---------------------|-------|
| A           | Cluster 8      | ZIP            | $2.35 \times 10^{-3}$ | ERL    | $2.46 \times 10^{-3}$ |
| B           | Cluster 5      | EL+IM          | $1.96 \times 10^{-3}$ | EL+IM  | $1.96 \times 10^{-3}$ |
| C           | Cluster 2      | ERL            | $9.06 \times 10^{-3}$ | ERL    | $9.06 \times 10^{-3}$ |
| D           | Cluster 13     | ERL            | $3.92 \times 10^{-3}$ | ERL    | $3.92 \times 10^{-3}$ |
| E           | Cluster 7      | ERL            | $4.19 \times 10^{-3}$ | ERL    | $4.19 \times 10^{-3}$ |
| F           | Cluster 11     | ZIP            | $5.32 \times 10^{-4}$ | ERL    | $5.21 \times 10^{-4}$ |
| G           | Cluster 12     | ZIP            | $1.08 \times 10^{-3}$ | ZIP    | $1.08 \times 10^{-3}$ |
| H           | Cluster 9      | CL             | $9.72 \times 10^{-4}$ | CL     | $9.72 \times 10^{-4}$ |
| I           | Cluster 13     | ERL            | $3.43 \times 10^{-4}$ | ZIP    | $3.31 \times 10^{-4}$ |
| J           | Cluster 10     | ZIP            | $5.66 \times 10^{-4}$ | ZIP    | $5.66 \times 10^{-4}$ |
C. EXPONENTIAL LOAD MODEL AND INDUCTION MOTOR (EL+IM)

The exponential load model (EL) and the induction motor (IM), defined as EL+IM, use the third-order induction model shown in (A3) and (A4) and the static part defined in (C.1) [36]:

$$P_{EL} = P_{EL0} \left( \frac{U}{U_0} \right)^{\alpha}; \quad Q_{EL} = Q_{EL0} \left( \frac{U}{U_0} \right)^{\beta} \quad (C.1)$$

where $U$ is the measured voltage at the load bus, $P_{EL0}$ and $Q_{EL0}$ are the pre-disturbance active and reactive power consumed by the static load, and $\alpha$ and $\beta$ are the static exponent's parameters.

D. EXPONENTIAL RECOVERY LOAD (ERL)

The exponential recovery load model (ERL) is based on the power exponential response after a step disturbance at the bus voltage [44]. This is modeled through first-order nonlinear differential equations shown in (D.1):

$$
T_p \frac{dP_r}{dt} = -P_r + P_0 \left( \frac{U}{U_0} \right)^{N_{ps}} + P_0 \left( \frac{U}{U_0} \right)^{N_{pt}} \quad \text{(D.1)}
$$

$P_t$ is the measured voltage at the load bus; $U_0$ is the rated voltage; $x_p$ is the state variable related to active power dynamics; $T_p$ represents the time constants of the exponential recovery response; $N_{ps}$ and $N_{pt}$ are related to the steady-state and transient load response, respectively; $P_r$ is the active power recovery; $P_t$ is the total active power response; and $P_0$, $Q_0$ are the rated active and reactive power at the load bus. A similar notation applies to reactive power.

ACKNOWLEDGMENT
This research was developed by the ICE³ research group at the Universidad Tecnológica de Pereira, Colombia.

REFERENCES
[1] A. Arif, Z. Wang, J. Wang, B. Mather, H. Hashbualdo, and D. Zhao, “Load modeling—A review,” IEEE Trans. Smart Grid, vol. 9, no. 6, pp. 5986–5999, Nov. 2018, doi: 10.1109/TSG.2017.2700436.
[2] W. W. Price, H. D. Chiang, H. K. Clark, C. Cordincia, D. C. Lee, J. C. Hu, S. Ibara, C. A. King, C. J. Lin, Y. Mansour, and K. Srinivasan, “Load representation for dynamic performance analysis of power systems,” IEEE Trans. Power Syst., vol. 8, no. 2, pp. 472–482, May 1993, doi: 10.1109/59.260837.
[3] W. W. Price, C. W. Taylor, and G. J. Rogers, “Standard load models for power flow and dynamic performance simulation,” IEEE Trans. Power Syst., vol. 10, no. 3, pp. 1302–1313, Aug. 1995, doi: 10.1109/59.466523.
[4] P. N. Papadopoulos, T. A. Papadopoulos, P. Crolla, A. J. Roscoe, G. K. Papagiannis, and G. M. Burt, “Black-box dynamic equivalent model for microgrids using measurement data,” IET Gener, Transmiss. Distrib., vol. 8, no. 5, pp. 851–861, May 2014, doi: 10.1049/iet-gtd.2013.0524.
[5] S. M. Zali and J. V. Milanic, “Generic model of active distribution network for large power system stability studies,” IEEE Trans. Power Syst., vol. 28, no. 3, pp. 3126–3133, Aug. 2013, doi: 10.1109/TPWRS.2012.2333223.
[6] M. E.-N. Jahromi and M. T. Ameli, “Measurement-based modelling of composite load using genetic algorithm,” Electr. Power Syst. Res., vol. 158, pp. 82–91, May 2018, doi: 10.1016/j.epsr.2017.12.023.
[7] H. Livan and C. Y. Evrenosoglu, “A machine learning and wavelet-based fault location method for hybrid transmission lines,” IEEE Trans. Smart Grid, vol. 5, no. 1, pp. 51–59, Jan. 2014, doi: 10.1109/TSG.2013.2260421.
[26] M. Bostanci, J. Koplowitz, and C. W. Taylor, “Identification of power system load dynamics using artificial neural networks,” *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1468–1473, Nov. 1997, doi: 10.1109/59.627843.

[27] E. O. Kontis, T. A. Papadopoulos, M. H. Syed, E. Guillo-Sansano, G. M. Burt, and G. K. Papagiannis, “Artificial-intelligence method for the derivation of generic aggregated dynamic equivalent models,” *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2947–2956, Jul. 2019, doi: 10.1109/TPWRS.2019.2894185.

[28] C. Wang, Z. Wang, J. Wang, and D. Zhao, “SVM-based parameter identification for composite ZIP and electronic load modeling,” *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 182–193, Jan. 2019, doi: 10.1109/TPWRS.2018.2865966.

[29] X. Liang, Y. He, M. Mitolo, and W. Li, “Support vector machine based dynamic load model using synchrophasor data,” in *Proc. IEEE/IAS 54th Ind. Commercial Power Syst. Tech. Conf. (ICPS)*, May 2018, pp. 1–11, doi: 10.1109/ICPS.2018.8369961.

[30] M. Darnstädt, H. U. Simon, and B. Szörényi, “Supervised learning and co-training,” *Theor. Comput. Sci.*, vol. 519, pp. 68–87, Jan. 2014, doi: 10.1016/j.tcs.2013.09.020.

[31] M. R. Asif, T. S. Bording, A. S. Barfod, D. J. Grombacher, P. K. Maurya, A. V. Christiansen, E. Auken, and J. J. Larsen, “Effect of data preprocessing on the performance of neural networks for 1-T transient electromagnetic forward modeling,” *IEEE Access*, vol. 9, pp. 34635–34646, 2021, doi: 10.1109/ACCESS.2021.3061761.

[32] M. S. Ibrahim, W. Dong, and Q. Yang, “Machine learning driven smart electric power systems: Current trends and new perspectives,” *Appl. Energy*, vol. 272, Aug. 2020, Art. no. 115237, doi: 10.1016/j.apenergy.2020.115237.

[33] L. Kaufman and P. J. Rousseeuw, *Finding Groups Data: Introduction to Cluster Analysis*. Hoboken, NJ, USA: Wiley, 1990.

[34] F. Liu and Y. Deng, “Determine the number of unknown targets in open world based on elbow method,” *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 5, pp. 986–995, May 2021, doi: 10.1109/TFUZZ.2020.2966182.

[35] M. Gao, S. Pan, S. Chen, Y. Li, N. Pan, D. Pan, and X. Shen, “Identification method of electrical load for electrical appliances based on K-means ++ and GCN,” *IEEE Access*, vol. 9, pp. 27026–27037, 2021, doi: 10.1109/ACCESS.2021.3057722.

[36] P. Regulski, D. S. Vilchis-Rodriguez, S. Djurovic, and V. Terzija, “Estimation of composite load model parameters using an improved particle swarm optimization method,” *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 553–560, Apr. 2015, doi: 10.1109/TPWRD.2014.2301219.

[37] M. Papadimitrakis, N. Giamarelos, M. Stogiannos, E. N. Zois, N. A.-I. Livanos, and A. Alexandridis, “Metaheuristic search in smart grid: A review with emphasis on planning, scheduling and power flow optimization applications,” *Renew Sustain. Energy Rev.*, vol. 145, Jul. 2021, Art. no. 111072, doi: 10.1016/j.rser.2021.111072.

[38] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, “Applications of machine learning to machine fault diagnosis: A review and roadmap,” *Mech. Syst. Signal Process.*, vol. 138, Apr. 2020, Art. no. 106587, doi: 10.1016/j.ymssp.2019.106587.

[39] P. Anderson and A. Fouad, *Power System Control Stability*, 2nd ed. Hoboken, NJ, USA: Wiley, 2002.

[40] T. Yong and R. Lasseter, “Optimal power flow formulation in market of retail wheeling,” in *Proc. IEEE Power Eng. Soc. Winter Meeting*, vol. 1, Jan./Feb. 1999, pp. 394–398, doi: 10.1109/PESW.1999.747487.

[41] H. Renmu, M. Jin, and D. J. Hill, “Composite load modeling via measurement approach,” *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 663–672, May 2006, doi: 10.1109/TPWRS.2006.873130.

[42] J. V. Milanovic, K. Yamashita, S. M. Villanueva, S. Ž. Djokic, and L. M. Korunović, “International industry practice on power system load modeling,” *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3038–3046, Aug. 2013, doi: 10.1109/TPWRS.2012.2231069.

[43] C.-J. Lin, A.-T. Chen, C.-Y. Huang, H.-D. Chiang, J.-C. Wang, and L. Fekih-Ahmed, “Dynamic load models in power systems using the measurement approach,” *IEEE Trans. Power Syst.*, vol. 8, no. 1, pp. 309–315, Feb. 1993, doi: 10.1109/59.221226.

[44] D. Karlsson and D. J. Hill, “Modelling and identification of nonlinear dynamic loads in power systems,” *IEEE Trans. Power Syst.*, vol. 9, no. 1, pp. 157–166, Feb. 1994, doi: 10.1109/59.317546.

S. HERNÁNDEZ-PEÑA received the B.Sc. degree in electrical engineering from the Universidad Tecnológica de Pereira (UTP), Colombia, in 2018, where he is currently pursuing the M.Sc. degree. His research interests include microgrid modeling and machine learning. He is a member of the ICE3-UTP Research Group.

S. PÉREZ-LONDOÑO (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Universidad Tecnológica de Pereira (UTP), in 2000 and 2005, respectively, and the Ph.D. degree from the Universidad Nacional de Colombia, in 2013. She is currently an Associate Professor with the Electrical Engineering School, UTP. Her research interests include electric machines, load modeling, power stability, fault analysis, and soft computing. She is a member of the ICE3-UTP Research Group.

J. MORA-FLÓREZ received the B.Sc. and M.Sc. degrees in electrical engineering from the Universidad Industrial de Santander (UIS), Colombia, in 1996 and 2001, respectively, and the M.Sc. and Ph.D. degrees in information technologies from the University of Girona (UdG), Spain, in 2003 and 2006, respectively. He is currently an Associate Professor with the Electrical Engineering School, Universidad Tecnológica de Pereira (UTP). His research interests include power systems, protective relaying, and soft computing techniques. He is a member of the ICE3.