Hourly Electricity Consumption Forecasting for Antioquia-Colombia Using Statistical-Machine Learning Models

Adiel Ignacio Restrepo Gil

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2024

Hourly Electricity Consumption Forecasting for Antioquia-Colombia Using Statistical-Machine Learning Models

A Master's Degree Project Submitted by
Adiel Ignacio Restrepo Gil

In Partial Fulfillment of the Requirements for the Degree of:
Magister en Ciencias-Estadística

Director:
Norman Diego Giraldo Gómez
M.Sc., Associate Professor, with Tenure
Escuela de Estadística
Universidad Nacional de Colombia

Universidad Nacional de Colombia
Facultad de Ciencias, Escuela de Estadística
Medellín, Colombia
2024

You cannot create experience.
You must undergo it.

Albert Camus

Acknowledgments

To all who have supported me throughout this process.

Abstract

Energy sector plays a fundamental role in encouraging a country's economic growth and social progress due to its functionality as an input for productive processes and as a public service asset that provides greater welfare to the population. Electricity consumption forecasting is a valuable instrument for policy-makers to guide pricing, taxation and investment decisions, as well as energy and operational security planning, helping to ensure a continuous supply of electricity and reducing cost overruns associated with the provision of energy distribution services. The aim of this research is to forecast hourly electricity consumption in Antioquia-Colombia using Statiscal-Machine Learning models with exogenous variables such as day-type and maximum temperature. The results show that LSTM Neural Network can be an efficient model for the operational deployment of electricity distribution since its average electricity supply error for an operational week is estimated to be around 493 MWh, while XM[1] Market Operator's benchmark model obtained an error of 3420 MWh during the evaluated week.

Keywords: Forecast, Electricity Consumption, Machine Learning.

Resumen

El sector energético desempeña un papel fundamental en el fomento del crecimiento económico y el progreso social de un país debido a su funcionalidad como insumo de los procesos productivos y como activo de servicio público que proporciona mayor bienestar a la población. La previsión del consumo de energía eléctrica es un valioso instrumento para que los hacedores de política orienten las decisiones de tarifas, impuestos e inversión, así como la planificación de la seguridad energética y operativa, contribuyendo a garantizar un suministro continuo de electricidad y reduciendo los sobrecostos asociados a la prestación de los servicios de distribución de energía. El objetivo de esta investigación es pronosticar el consumo de electricidad horario en Antioquia-Colombia utilizando modelos de Statistical-Machine Learning con variables exógenas como el tipo de día y la temperatura máxima. Los resultados muestran que la Red Neuronal LSTM puede ser un modelo eficiente para el despliegue operativo de la distribución eléctrica debido a que su error promedio de suministro de electricidad para una semana operativa se estima en alrededor de 493 MWh, mientras que el modelo de referencia del Operador de Mercado XM obtuvo un error de 3420 MWh durante la semana evaluada.

Palabras Clave: Pronóstico, Consumo de Electricidad, Machine Learning. .

[1] www.xm.com.co

Content

1. Introduction

Energy sector has a key role to achieve a better development and economic growth due to its dual functionality as an input for productive processes and as a public service asset that provides greater welfare to the population.

One of the most important policy instruments used by decision makers worldwide is energy consumption forecasting (Meng et al., 2022) to provide information to support investment decisions on energy supply infrastructure and to facilitate consensus building on expansion and operation projects (UPME, 2021) with the aim to advance in the implementation of complementary mechanisms to ensure the effective provision of electricity service in the case of a lower availability of water resources or a lower supply of fossil fuels (CPC, 2020).

As reported by (EIA, 2023) electricity is the flow of electrical power which is both a basic part of nature and one of the most widely used forms of energy. The electricity that we use is a secondary energy source because it is produced by converting primary sources of energy such as coal, natural gas, nuclear energy, solar energy, and wind energy, into electrical power. According to (DNP, 2017) electricity is used for lighting, heating, cooling, refrigeration, operating appliances, computers, electronics, machinery, driving power and public transportation systems.

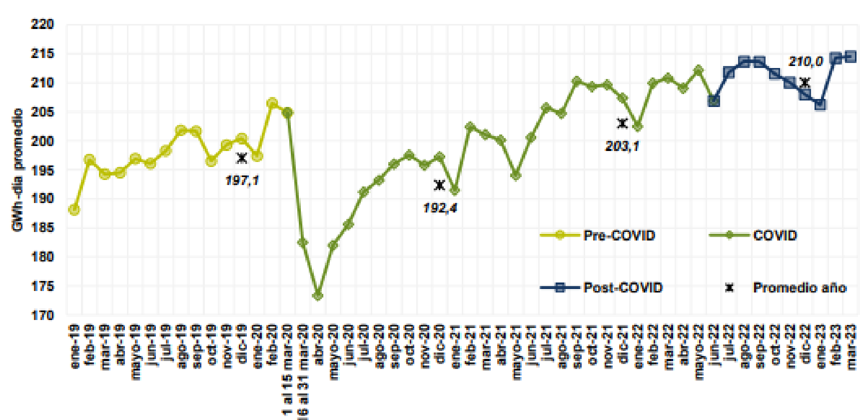


Figure 1-1.: Colombia's Average Daily Electricity Consumption (UPME, 2023).

The total influence of the COVID-19 health emergency on the worldwide energy sector has seen energy consumption drip dramatically (Garcia et al., 2023; Kumar et al., 2021). In Colombia, electricity demand in 2023 has been reporting strong signs of recovery from the reduction observed in 2020 as a result of the pandemic. The average monthly growth rates were -2.01 % for 2020, 5.4 % for 2021, and 3.48 % for 2022. Starting from June 2021, electricity consumption began to rise at higher levels compared to 2020 and 2019. (UPME, 2023).

Accurate electricity demand forecasting is of major relevance to any country's energy policy and planning due to inaccurate demand forecasting may raise the operation cost of electrical power systems (Li et al., 2021) which means considerable money cannot be saved (Jiang et al., 2020). In Colombia, the electric power market has experienced several drawbacks in the provision of energy distribution services (Jimenez Mares et al., 2019) which may be solved if a balance between energy generation and demand is maintained through the choice of a supply strategy based on forecasting information that allows for timely decisions necessary for energy transactions in the market between producers and consumers (Li et al., 2021).

The growing competition and deregulation within the electricity market have resulted in significant obstacles to satisfy energy demand with the least possible expense, while guaranteeing service quality, safety, and reliability. Reducing errors in electricity consumption forecasting streamlines the daily operation and scheduling of generation units, eliminating the need for new schedules or redispatches with more expensive plants or the cancellation of previously planned generations, thus avoiding unnecessary expenses (Valencia et al., 2007).

Traditional linear models may have several limitations in analyzing and predicting nonlinear, complex and irregular electricity consumption data (James et al., 2021; Wen et al., 2020; Zhou et al., 2020). Although artificial intelligence models, hybrid models and ensemble learning methods are an attractive alternative to overcome these problems. According to (James et al., 2021; Khalil et al., 2022), these methods may suffer from issues such as local optimization, overfitting, parameter sensitivity, complex structure, and usability challenges. Therefore, this research proposes different approaches to forecasting electricity consumption, aiming to compare and select the most appropriate methodology based on predictive performance measures.

2. Theoretical Framework

This section offers a concise overview of the statistical theory that forms the foundation of machine learning models, along with essential insights into the training aspects within the time series framework and modeling short-term electricity consumption.

2.1. Supervised Learning model approach

Probability theory and statistics as reported by (Wasserman, 2010) are based on the model of a random experiment or probability space $(\Omega, \mathcal{F}, P)$.Here, $\Omega = \{w_1, w_2, ..., w_n\}$ is a set containing all possible outcome, where the points w in Ω are called sample outcomes, realizations, or elements. $\mathcal{F}$ is a collection of events $A \subseteq \Omega$ and P is a measure assigning a probability to each event.

According to (Peters et al., 2017), given the above mathematical structure, probability theory allows us to reason about the results of random experiments, while statistical learning deals with the reverse problem: given the results of the random experiment, we want to infer from the properties of the underlying mathematical structure. The statistical learning is the process of using data to infer distribution that generated the data.

For instance, suppose we observe pairs of data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$ usually called *learning data set*, where $x_i \in \mathcal{X}$ are **inputs** (covariates or independent variables) and $y_i \in \mathcal{Y}$ are **outputs** (response variables or dependent variables). The statistical learning model assumes that the observations $\{(x_i, y_i)\}_{i=1}^{n}$ are realizations of the random variables $\{(X_i, Y_i)\}_{i=1}^{n}$ which are independent and indetically distributed with joint distribution function $P_{X,Y}$. However, such an assumption can be violated if there are changes in distributional processes or interventions in a system. These disturbances may be linked to causality.

We may now be interested in building a model to predict an output $Y \in \mathcal{Y}$ based on inputs $(X_1, ..., X_p) \in \mathcal{X} \subset \mathbb{R}^p$ using a learning data set $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n}$, $x_i \in \mathcal{X}$ such as: if $\mathcal{Y}$ is continuous, $\mathcal{Y} = \mathbb{R}$ is called *Regression Problem* and if $\mathcal{Y}$ is finite, $\mathcal{Y} = \{1, ..., C\}$ is called *Classification Problem*. The relationship between the response $Y \in \mathcal{Y}$ and $p \in \mathbb{N}$ predictors $\mathbf{X} = (X_1, ..., X_p) \in \mathcal{X}$ can be written in a very general form as:

$$Y = f(X) + \epsilon \tag{2-1}$$

As specified by (James et al., 2021) $f : \mathcal{X} \to \mathcal{Y}$ is a fixed unknown function that represents the systematic information that X provides about Y, and ϵ is a random error term independent of X with mean of zero. Since the error term averages to zero, we can predict Y using $\widehat{Y} = \widehat{f}(X)$. Statistical learning refers to a set of approaches for estimating f using learning data set. The accuracy of $\widehat{Y}$ as a prediction for $E(Y|X)$ depends on two quantities: *reducible error* and *irreducible error*.

$$\begin{aligned} E(Y - \widehat{Y})^2 &= E[f(X) + \epsilon - \widehat{f}(X)]^2 \\ &= \underbrace{[f(X) - \widehat{f}(X)]^2}_{Reducible} + \underbrace{Var(\epsilon)}_{Irreducible} \end{aligned} \tag{2-2}$$

The set of techniques to estimate f focuses on minimizing the reducible error since it is possible to significantly improve the accuracy of $\widehat{f}(X)$ by using the most appropriate statistical learning technique to learn about $f(X)$. The irreducible error is uncontrollable, no matter how well we estimate f.

2.2. Statistical-Machine Learning Models

2.2.1. Multiple Linear Regression (MLR)

Linear Regression Models are a simple approach to supervised learning. According to (James et al., 2021), linear regression is a useful tool for predicting a quantitative response variable and has been widely used for a long time. Supervised Learning Models seek to find a relationship between a response variable Y and a set of covariates X, which takes the form $Y = f(X) + \epsilon$. If f is to be approximated by a linear function, then we can write this relationship as:

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i\,, \;\; i = 1, 2, ..., n \tag{2-3}$$

Here, Y_i represents the i-th observation of the dependent variable, X_{ij} represents the j-th predictor, β_j is the average effect on Y of a one-unit increase in X_{ij}, holding all other predictors fixed, and ϵ_i is a random error term. The regression coefficients are unknown and can be estimated using the *Least Squares* (LS) approach. We choose $\beta_0, \beta_1, ..., \beta_p$ to minimize the sum of squared residuals:

$$RSS(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{2-4}$$

As specified by (Hastie et al., 2009), $RSS(\boldsymbol{\beta})$ is a quadratic function of the parameters such that its minimum always exists but may not be unique. There are several potential problems that can occur when fitting MLR to a particular dataset. For instance, as stated in (James et al., 2021):

1. **Non-linearity of the response-predictor relationships**: If the true relationship is far from linear, conclusions about the model's fit are suspect, and the model's predictive performance may be significantly diminished.

2. **Correlation of error terms**: If there is a correlation between the error terms, then the estimated standard errors will tend to underestimate the true standard errors, which may affect the predictive performance of the model. Such correlations frequently occur in the context of time series data.

3. **Non-constant variance of error terms**: An important assumption of MLR is that the error term has constant variance, $Var(\epsilon_i) = \sigma^2$ which is often difficult to achieve.

4. **Outliers**: If there are outliers, they can affect the estimation of the parameters through the LS approach and, therefore, reduce the fit of the MLR. Observations with high leverage can substantially impact the estimated regression line

5. **Collinearity**: This occurs when two or more variables are closely related to one another, meaning a variable can be expressed as a linear combination of the others. Usually, this problem is solved by removing some of the highly correlated variables using statistical techniques. However, with time series data, it is incorrect to eliminate variables since important information for predicting the target variable could be lost.

2.2.2. Least Absolute Shrinkage and Selection Operator (LASSO)

LASSO is a regularization or shrinkage method popular in high-dimensional estimation due to its statistical accuracy for prediction, variable selection, interpretability, and computational feasibility. As mentioned by (Hastie et al., 2015), there are two main reasons to consider alternatives to the least squares estimate: prediction accuracy and interpretability.

Prediction accuracy can often be enhanced by reducing regression coefficient values and even setting some to zero.When dealing with a large number of predictors, it is desirable to identify a subset that exhibits the strongest effects. According to (Hastie et al., 2009), LASSO incorporates regularization by penalizing the sum of the absolute values of the regression coefficients, known as l_1-*norm* given by $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^{p} |\beta_j|$. The LASSO estimate is defined by:

$$RSS(\boldsymbol{\beta};\lambda) = \frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 \tag{2-5}$$

Where $\lambda \geq 0$ is a complexity or tuning parameter that controls the amount of shrinkage and must be determined separately using, for instance, cross-validation. When $\lambda = 0$ the penalty term has no effect, and LASSO reduces to producing the least squares (LS) estimates. As $\lambda = 0$ increases, some coefficient estimates are forced to zero, effectively selecting the most influential predictors, which helps mitigate overfitting or multicollinearity issues.

Models suffering from multicollinearity or containing numerous explanatory variables often perform well in training samples but poorly in test samples, despite exhibiting high *R-Squared* estimates. As determined by (Hastie et al., 2015) there exists an optimal point on the path from underfitting to overfitting, where the test *Mean Square Error* (MSE) is minimized. This point corresponds to the ideal model estimation scenario, where there is sufficient information to achieve accurate forecasting

2.2.3. Generalized Additive Model (GAM)

GAM is a flexible statistical learning method which can be used to identify and characterize nonlinear regression effects. According to (James et al., 2021) to allow for non-linear relationships between each explanatory variable $(X_1, ..., X_p) \in \mathcal{X} \subset \mathbb{R}^p$ and the response variable $Y \in \mathcal{Y}$ is to replace each linear component $\beta_j X_{ij}$ with a non-linear function $f_j(X_{ij})$. Then we can write this relationship as:

$$Y_i = \beta_0 + \sum_{j=1}^{p} f_j(X_{ij}) + \epsilon_i\,, \;\; i = 1, 2, ..., n \tag{2-6}$$

The f_j's are unspecified smooth or nonparametric functions and may be expressed as $f_j(X_{ij}) = \sum_{k=1}^{K} \beta_k b_k(X_{ij})$ where $b_k(X_{ij})$ is the k-th basis function. As stated in (Wood, 2017), $f_j(X_{ij})$ can be represented, for instance, by a cubic spline which is a curve formed by sections of cubic polynomials connected in such a way that the first derivative exists to guarantee the continuity of the function at the point and the second derivative exists so that no concavity changes from one side of the point to the other, thus guaranteeing the smoothness of the curve around the point. The connection points are known as spline nodes, which can be equally spaced in the range of X_{ij} or be positioned in its quantiles.

GAM is more flexible than MLR, since it allows for nonlinear relationships between the response variable and the explanatory variables. However, this flexibility may come at a cost: if we want a very flexible relationship and, therefore, a more accurate fit to the data, the base function $f_j(X_{ij})$ will generally have a large dimension, which may lead to possible overfitting. As reported by (Wood, 2017) if it is assumed that Y_i is distributed according to an *Exponential Family*, i.e. $Y_i \sim EF(\mu_i, \phi)$ with mean μ_i and scale parameter ϕ, overfitting can be alleviated with the Penalized Likelihood Function given by:

$$l_p(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - \frac{1}{2\phi} \sum_j \lambda_j \boldsymbol{\beta}^T \boldsymbol{S}_j \boldsymbol{\beta} \tag{2-7}$$

$l(\boldsymbol{\beta})$ is the unrestricted log-likelihood, $\boldsymbol{S}_j$ denote the model and penalty matrix for f_j. The λ_j are a smoothing parameters and control trade-off between model goodness of fit and model smoothness. They are selected in such a way as to minimize the Double Cross Validation Score proposed by (Kim and Gu, 2004):

$$\mathcal{V}_d = \frac{n \left\| \boldsymbol{y} - \widehat{\boldsymbol{u}} \right\|}{\left[n - 1{,}5\text{tr}(\boldsymbol{A}) \right]^2} \tag{2-8}$$

Where $\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{S}_\lambda)\boldsymbol{X}^T$ is a influence or hat matrix which $\boldsymbol{S}_\lambda = \sum_j \lambda_j \boldsymbol{S}_j$. GAM's main limitation according to (James et al., 2021) is that the model is restricted to be additive, so with high-dimensional data, important interactions may be missed.

2.2.4. Multivariate Adaptive Regression Splines (MARS)

As explained by (Hastie et al., 2009) MARS is an adaptive regression procedure well-suited for high-dimensional problems. As stated by (Murat, 2022), it works effectively with a large number of predictor variables, automatically detects interactions between variables, and operates efficiently and swiftly. However, it has limitations such as susceptibility to overfitting and challenges in interpretation.

MARS is a nonparametric regression technique that captures the nonlinear relationship between a set of covariates and a response variable using basis functions, similar in structure to GAM. It can be expressed as:

$$Y_i = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{k=1}^{K_m} h_{km}(X_{i,v(k,m)}) + \epsilon_i \,, \quad i = 1, 2, ..., n \tag{2-9}$$

According to (Safer, 2004) basis functions h_{km} are first-order truncated power splines, $h_{km}(X) = \pm\,(X - t_{km})_+$ where t_{km} is the knot of the input variable and $v(k,m)$ is an index of the predictor for the m-th component of the k-th product. As defined in (Hastie et al., 2009) MARS uses expansions in piecewise linear basis functions of the form:

$$(x-t)_+ = \begin{cases} x - t, & \text{if } x > t \\ 0, & otherwise \end{cases} \quad \text{and} \quad (t-x)_+ = \begin{cases} t - x, & \text{if } x < t \\ 0, & otherwise \end{cases} \tag{2-10}$$

For $K_m = 1$, the model is additive. For $K_m = 2$, the model allows pairwise interactions. A maximum M is set on the number of base functions allowed in the model. The least important basis functions are individually taken out of the model based on the *Generalized Cross-Validation* (GCV) criterion, as specified by (Wood, 2017):

$$\mathcal{V}_d = \frac{n \left\| \boldsymbol{y} - \widehat{\boldsymbol{u}} \right\|}{\left[n - \text{tr}(\boldsymbol{A}) \right]^2} \tag{2-11}$$

2.2.5. Extreme Gradient Boosting (XGBoost)

XGBoost is an open-source machine learning project proposed by (Chen and Guestrin, 2016). It is a powerful ensemble learning algorithm that leverages the construction of multiple subtrees to make predictions. By efficiently combining the predictions from these subtrees, XGBoost generates precise final results.

The fundamental idea behind boosting algorithms is to sequentially adjust multiple weak learners, which are simple models that outperform random expectations. With each iteration, every new model learns from the mistakes of the previous model and thereby enhances its predictive capacity(James et al., 2021). For a given learning data set with n observations and p features, $\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$, $\boldsymbol{x}_i \in \mathbb{R}^p$ and $\mathcal{Y}$ is continuous therefore $y_i \in \mathbb{R}$, a tree ensemble model uses K additive functions to predict the target variable or the number of subtrees built by the model. The ensemble model of the XGBoost tree is written as follows:

$$\widehat{y}_i = \phi(\boldsymbol{x}_i) = \sum_{k=1}^{K} f_k(\boldsymbol{x}_i), \;\; f_k \in \mathcal{F} \tag{2-12}$$

Where $\mathcal{F} = \{f(\boldsymbol{x}) = w_{q(\boldsymbol{x})}\}(q : \mathbb{R}^p \to T,\ w \in \mathbb{R}^T)$ is the space of regression trees with q structures in each tree and T leaves in the tree. $\widehat{y}_i$ is the predicted value of the model for the i-th sample and each f_k represents to an independent tree structure q and leaf weights w or the score given by the k-th tree to the i-th observations in the data. To learn the set of functions used in the model, the XGBoost objective function is comprised of two parts: (i) the error between the predicted value of the model $\widehat{y}_i$ and the true value y_i and the regular term $\Omega(f_k)$ with γ, λ penalty coefficient that controls the complexity of the model which prevents the model from overfitting, as shown in the following regularized objective:

$$\begin{gathered} \mathcal{L}(\phi) = \sum_i l(\widehat{y}_i, y_i) + \sum_k \Omega(f_k) \\ \text{where } \Omega(f) = \gamma T + \frac{1}{2}\lambda \left\| w \right\|^2 \end{gathered} \tag{2-13}$$

In the above formula, γ is a parameter that controls the number of leaves T and λ is a parameter that controls the leaf weight w. However (2-13) cannot be optimized using traditional optimization methods in Euclidean space (Chen and Guestrin, 2016) and as a result the XGBoost algorithm minimizes the objective function through an iterative method with t-th iteration, as illustrated in the formula below:

$$\begin{gathered} \widetilde{\mathcal{L}}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(\boldsymbol{x}_i) + \frac{1}{2} h_i f_t^2(\boldsymbol{x}_i) \right] + \Omega(f_t) \\ \text{where } g_i = \partial_{\widehat{y}^{(t-1)}} l(y_i, \widehat{y}^{(t-1)}) \text{ and } h_i = \partial^2_{\widehat{y}^{(t-1)}} l(y_i, \widehat{y}^{(t-1)}) \end{gathered} \tag{2-14}$$

2.2.6. Long Short-Term Memory (LSTM) Networks

LSTM is a variation of Recurrent Neural Network (RNN) designed for processing long sequential data, proposed by (Hochreiter and Schmidhuber, 1997). LSTM introduces a memory cell and generates a computational unit to replace traditional artificial neurons in the hidden layer of a network. According to (Goodfellow et al., 2016), these memory cells address several challenges faced by traditional RNNs, such as the vanishing gradient problem that leads to difficulty in parameter estimation and limits the ability to find global optima.

As explained by (Ushiku, 2020), unlike standard RNNs, LSTM uses an additional memory cell $\boldsymbol{c}$ alongside the hidden state $\boldsymbol{h}$. This memory cell, known as the long-term memory, stores information over long periods, while $\boldsymbol{h}$ represents the short-term memory. The computations at the t-th step are as follows:

$$\boldsymbol{c}_{t-1/2} = \boldsymbol{c}_{t-1} \odot \boldsymbol{g}_f \tag{2-15}$$

$$\boldsymbol{g}_f = \sigma\left(\boldsymbol{W}_f \begin{bmatrix} \boldsymbol{g}_{t-1} \\ \boldsymbol{x}_t \end{bmatrix} + \boldsymbol{b}_f\right) \tag{2-16}$$

$$\boldsymbol{c}_t = \boldsymbol{c}_{t-1/2} + \boldsymbol{g}_i \odot \boldsymbol{z} \tag{2-17}$$

$$\boldsymbol{g}_i = \sigma\left(\boldsymbol{W}_i \begin{bmatrix} \boldsymbol{h}_{t-1} \\ \boldsymbol{x}_t \end{bmatrix} + \boldsymbol{b}_i\right) \tag{2-18}$$

$$\boldsymbol{z} = \tanh\left(\boldsymbol{W}_z \begin{bmatrix} \boldsymbol{h}_{t-1} \\ \boldsymbol{x}_t \end{bmatrix} + \boldsymbol{b}_z\right) \tag{2-19}$$

$$\boldsymbol{h}_t = \boldsymbol{g}_o \odot \tanh\left(\boldsymbol{c}_t\right) \tag{2-20}$$

$$\boldsymbol{g}_o = \sigma\left(\boldsymbol{W}_o \begin{bmatrix} \boldsymbol{h}_{t-1} \\ \boldsymbol{x}_t \end{bmatrix} + \boldsymbol{b}_o\right) \tag{2-21}$$

The vectors $\boldsymbol{g}_{(\cdot)}$ are gates to update memories in LSTM which are normalized between 0 and 1. Thus, each element of gates represents the strength for which elements of memories are filtered using the current input and the last memory. The notations $\boldsymbol{W}_{(\cdot)}$ and $\boldsymbol{b}_{(\cdot)}$ is a pair of a transformation matrix and a bias vector respectively. The Hadamard product is denoted by the symbol $\odot$, namely, element-wise product of two vectors. An illustration of LSTM's architecture, in which $\oplus$ denotes the summation of two vectors, is presented below:

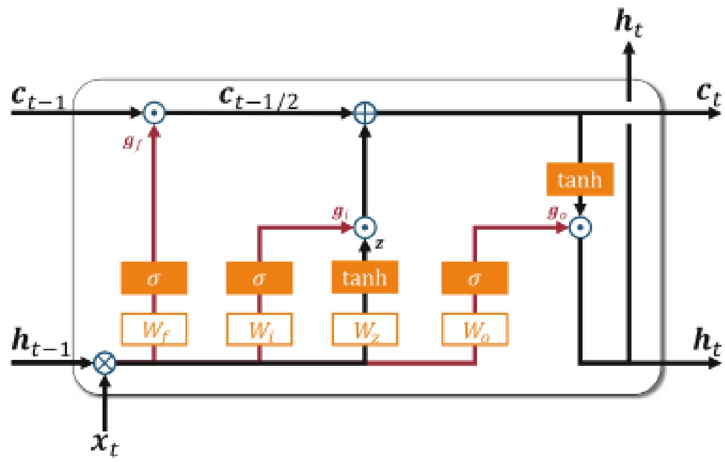


Figure 2-1.: Architecture of an LSTM Cell (Ushiku, 2020).

The paths for the gate vectors are shown in dark red, while the others are in black. At the t-th step, the memories $\boldsymbol{c}$ and $\boldsymbol{h}$ are updated and then the output $\boldsymbol{y}_t$ is computed as follows:

$$\boldsymbol{y}_t = \tanh\left(\boldsymbol{W}_y \boldsymbol{h}_t + \boldsymbol{b}_y\right) \tag{2-22}$$

2.3. Measure for Regression Problems

Evaluating the accuracy or efficiency of the estimated model in explaining the behavior of a target variable is an important aspect of the modeling process. Goodness-of-fit measures are a group of statistics that attempt to describe how well a calculated model fits to the observed data set. Typical regression measures are as follows:

1. **Mean Squared Error** (MSE) is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2 \tag{2-23}$$

   Since MSE is a squared measure, it penalizes extreme errors that occur during the estimation process. Furthermore, it is a measure that is very sensitive to scale changes and transformations.

2. **Root Mean Squared Error** ($RMSE$) is squared root of the average of the squared difference between actual and predicted values. It is defined as the square root of the MSE:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2} \tag{2-24}$$

This metric shares the same properties as MSE. However, it offers a more intuitive interpretation, as the unit of measurement used to calculate the data returns to its original scale.

3. **Mean Absolute Error** (MAE) is the average absolute deviation of the observed values from the predicted values. It is defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i| \tag{2-25}$$

This metric measures the total error magnitude, where positive and negative errors do not cancel each other out. Consequently, it becomes challenging to assess the bias of the estimates. As emphasized by (Hastie and Tibshirani, 2017), this metric tends to decrease with an increasing number of variables in the model training sample but does not exhibit the same behavior in the validation sample. Therefore, it is not a suitable metric for comparing models with varying numbers of explanatory variables

4. **Mean Absolute Percentage Error** ($MAPE$) shows the average deviation of the actual from prediction values percentage. It is given by:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100 \tag{2-26}$$

This metric is useful for model selection because, unlike the MAE, it is independent of the number of explanatory variables.

5. **Diebold-Mariano (DM) Test** is a statistical tool used to evaluate the significance of differences in forecasting accuracy proposed by (Diebold and Mariano, 1995). It operates as an asymptotic Z-test, specifically examining the hypothesis related to the mean of the loss differential series:

$$d_{ijk} = \mathcal{L}(\epsilon_{ik}) - \mathcal{L}(\epsilon_{jk}) \tag{2-27}$$

Where $\epsilon_{ik} = y_{ik} - \widehat{y}_{ik}$, $\epsilon_{jk} = y_{jk} - \widehat{y}_{jk}$ is the prediction error of model i and model j for time step k and $\mathcal{L}(\cdot)$ is a loss function. For example: $\mathcal{L}(\epsilon_k) = |\epsilon_k|^p$ with $p = 1$ or 2, which corresponds to the absolute and squared losses. The key hypothesis of equal predictive accuracy or equal expected loss corresponds to $E(d_{ijk}) = 0 \ \forall k$, in which case DM-Statistic is:

$$DM_{ijk} = \frac{\overline{d}_{ij}}{\widehat{\sigma}_{\overline{d}_{ij}}} \longrightarrow N(0,1) \tag{2-28}$$

Where $\overline{d}_{ij}$ is the sample mean loss differential and $\widehat{\sigma}_{\overline{d}_{ij}}$ is a consistent estimate of the standard deviation of $\overline{d}_{ij}$. However, (Harvey et al., 1997) suggests that improved small-sample properties can be obtained by (i) applying a bias correction to the DM test statistic and (ii) comparing the corrected statistic with a Student-t distribution with $(T-1)$ degrees of freedom, instead of the standard normal. The corrected DM-Statistic is obtained as:

$$DM^*_{ijk} = \sqrt{\frac{T+1-2h+h(h-1)}{T}} \cdot DM_{ijk} \longrightarrow t_{(T-1)} \tag{2-29}$$

The forecast horizon used in calculating ϵ_{ik} and ϵ_{jk} is determined by h, usually $h=1$. The null hypothesis is that the two methods have the same forecast accuracy:

$$H_o : E(d_{ijk}) = 0 \ \forall k \tag{2-30}$$

The alternative hypothesis can be:

- *Two-Sided*: the alternative hypothesis is that method i and method j have different levels of accuracy.
$$H_a : E(d_{ijk}) \neq 0 \tag{2-31}$$
- *Less-One-Sided*: the alternative hypothesis is that method j is less accurate than method i.
$$H_a : E(d_{ijk}) < 0 \tag{2-32}$$
- *Greater-One-Sided*: the alternative hypothesis is that method j is more accurate than method i.
$$H_a : E(d_{ijk}) > 0 \tag{2-33}$$

Some applications of the Diebold-Mariano test in modeling electricity consumption can be found in (Khan and Osińska, 2023; Ribeiro et al., 2020; Sadeghian Broujeny et al., 2023; Saranj and Zolfaghari, 2022; Zhou et al., 2023; Zolfaghari and Sahabi, 2019).

2.4. Literature Review

Accurate forecasting of electricity consumption is crucial for ensuring power system reliability, optimizing daily operations, efficiently managing energy resources, and developing effective demand response strategies. The scientific community has explored various modeling approaches, classified into three main categories according to (Khalil et al., 2022): Machine Learning approaches (such as Artificial Neural Network, Random Forest, Support Vector Machine, Boosting, Tree Regression), Deep Learning approaches (including Long-Short Term Memory, Convolutional Neural Network, Recurrent Neural Network) and Statistical Analysis approaches (such as Linear Regression, Multiple Linear Regression, Autoregressive Integrated Moving Average, Holt–Winters). The authors indicate that from 2015 to 2018, Machine Learning approaches dominated over Deep Learning and Statistical Analysis methods for electricity forecasting, but there has been significant growth in the use of Deep Learning approaches since 2019 and 2020. This shift is attributed to the ability of Machine Learning and Deep Learning models to handle large and complex datasets, capturing underlying relationships between input and output variables critical for electricity forecasting. In contrast, traditional approaches often struggle to capture non-linear variations in energy consumption data and typically rely on domain knowledge to select important features from the dataset.

Electricity consumption forecasting poses challenges due to its non-stationary, non-linear nature, and sensitivity to seasonal and holiday factors. Numerous methods have been proposed in the literature to address these challenges. Studies by (La Tona et al., 2023; Wen et al., 2020; Zhou et al., 2020; Zhu et al., 2021a) demonstrate that LSTM models outperform traditional statistical models in short-term electricity consumption prediction in terms of Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). These studies highlight the importance of considering external factors, such as holidays, weekends, special celebrations like New Year's, and atmospheric variables, which can significantly influence electricity consumption behavior.

Research by (Chapagain and Kittipiyakul, 2018) shows that including atmospheric variables like relative humidity, wind speed, precipitation, solar radiation, and cloudiness improves short-term demand forecasting accuracy compared to models that only consider temperature and calendar components. Additionally, (Arora and Taylor, 2018) demonstrates that incorporating working days, special days and days near holidays into models like SARMA enhances short-term electricity demand forecasting accuracy compared to using Artificial Neural Network (ANN) and Holt-Winters (HW) models. For this reason, it may be important to include external factors when modeling hourly electricity consumption.

Gradient boosting models have also proven effective in capturing non-linear patterns in electricity demand. Studies by (Maltais and Gosselin, 2022; Zhu et al., 2021b) using gradient boosting techniques report significant improvements in predictive performance compared to Linear Regression (LR), Support Vector Machine (SVM), ARMA, ARIMA, Neural Networks (NN), Backpropagation Neural Networks (BPNN) and Random Forests (RF). These researchers considered additional factors such as, extreme temperature, extreme temperature in the last 24 hours, extreme temperature in the last 168 hours, seasonality (spring, summer, autumn, winter), day-type, holiday position according to its duration (1-3 days or 1-7 days), legal holiday, month and hour when modeling short-term electricity demand. Hybrid approaches, such as those presented in (Divina et al., 2018; Sujan et al., 2022), combining Ensemble Gradient Boosting with Neural Networks, RF, and Evolutionary Tree (EVTree) models, aim to further enhance the accuracy of hourly electricity demand forecasts.

Models offering greater flexibility in capturing electricity demand behavior have shown to improve forecast efficiency and accuracy. For example, (Li et al., 2021) applies a variation of SVM known as Least Squared Support Vector Machine (LSSVM) to forecast short-term electricity consumption, outperforming models like BPNN, SVM, and ARIMA models. This fact can also be observed in (Young-Min et al., 2012), where the estimation of a Fuzzy Polynomial Regression has a higher predictive performance than the Fuzzy Linear Regression. The studies considered exogenous variables, such as business and non-business days, as well as meteorological variables, such as minimum temperature, maximum temperature, mean temperature, sunshine duration, relative humidity, wind speed, and discomfort index. These variables are typically used to model short-term electricity consumption, but (Son et al., 2022) also includes photovoltaic generation as an exogenous variable.

Studies in the Colombian literature have primarily focused on traditional models such as ARIMA, Linear Regression, Multiple Linear Regression, Moving Averages, and Spline Smoothing Models for hourly electricity demand estimation. Some studies like (Castaño, 2007; Murillo et al., 2003) have not incorporated external factors or compared different modeling approaches. Others studies like (Barrientos et al., 2007; Valencia et al., 2007) have integrated explanatory variables like day types, weekdays, and daily curves estimated from annual Gross Domestic Product (GDP), albeit infrequently in short-term applications. These studies underscore that short-term energy forecasting is heavily influenced by the immediate past consumption of the same day type and exhibits characteristic behaviors tied to weekdays.

Recent advancements include studies employing Artificial Neural Networks like Multilayer Perceptron, as seen in (Medina et al., 2012; Sarmiento and Villa, 2008) which incorporate exogenous variables such as day-type from preceding days and historical periods. These studies suggest that incorporating such variables into Artificial Neural Networks improves prediction accuracy compared to Autoregressive Exogenous (ARX) Models. Current research spans a range of methodologies including Gaussian Processes, SVM, RF, Gradient Boosting, Generalized Additive Models (GAM) (Berbesi and Pritchard, 2023; Rosero et al., 2023) and Functional Time Series (FTS) approaches, aiming to refine ARIMA and ARFIMA models (Barrientos et al., 2023). It is evident that the modeling of hourly electricity consumption has been studied from a traditional approaches to more advanced approaches, in which explanatory variables related to seasonality and calendar effects are used. While some studies have successfully integrated atmospheric variables, caution is advised due to the complexity and uncertainty these variables introduce.

Research	Country	Model	Time Horizon	Exogenous Variable
(Arora and Taylor, 2018)	France	ARMA, HWT, ANN	Hourly	Weekday, Weekend Bridging Proximity Days
(Barrientos et al., 2007)	Colombia	Spline Regression, ARIMAX	Hourly	Weekday, Typical Daily Curve
(Barrientos et al., 2023)	Colombia	FTS, ARIMA, ARFIMA	Hourly	
(Berbesi et al, 2023)	Colombia	GAM	Hourly	Weekday, Hour, Region
(Castaño, 2007)	Colombia	ARIMA	Hourly	
(Chapagain et al, 2018)	Japan	MLR, ARMAX	Hourly	Month, Day-Type, Temperature, Relative Humidity, Wind Speed, Precipitation, Solar Radiation, Cloud Cover
(Divina et al., 2018)	Spain	Gradient Boosting, ANN, RF, EVTree, LR, ARIMA, ARMA	Ten-Minute	
(La Tona et al., 2023)	France	LSTM, ANN, NARX, Naive Seasonal	Daily	Global Active and Reactive Power, Voltage, Global Intensity, Sub-Metering
(Li et al., 2021)	Australia	SVM, LSSVM, BPNN, ARIMAX	Half-Hour	Weekday, Weekend
(Maltais et al, 2022)	Canada	Gradient Boosting, ANN, LR	Ten-Minutes	Weekday, Weekend, Month, Day, Hour
(Medina et al., 2012)	Colombia	ANN, AR, ARX	Hourly	Day-Type: actual, before, after, three weeks ago and one year ago
(Murillo et al., 2003)	Colombia	ARIMA	Hourly	
(Rosero et al., 2023)	Colombia	Gradient Boosintg, RF, SVM, KNN, Gaussian Process	Hourly	
(Sarmiento et al, 2008)	Colombia	ANN	Hourly	Day-Type
(Son et al., 2022)	South Korea	LSTM, XGBoost, Fuzzy Linear Regression, WTSM Algorithm	Hourly	Day-Type, Weather, Photovoltaic Generation
(Sujan et al., 2022)	India	Gradient Boosting, LSTM, ANN, RF, EVTree, ARMA, ARIMA	Hourly	
(Valencia et al., 2007)	Colombia	Moving Averages	Hourly	Day-Type
(Wen et al., 2020)	EEUU	LSTM, RNN, ANN, ARIMA, SVM, MLR	Hourly	Weekday, Weekend, Temperature, Pressure, Wind Speed, Hour, Day
(Young-Min et al., 2012)	South Korea	Fuzzy Polynomial Regression, Fuzzy Linear Regression	Hourly	Temperature: minimum, maximum and mean. Sunshine Duration, Relative Humidity, Wind Speed, Discomfort Index
(Zhou et al., 2020)	China	LSTM, BPNN, ARIMA	Hourly Daily	
(Zhu et al., 2021b)	China	XGBoost, SVM, BPNN, RF	Hourly	Day-Type, Holiday-Type, Holiday Duration, Legal Holiday Extreme temperature: actual, in the last 24 and 168 hours. Seasonality: spring, summer, autumn, winter.
(Zhu et al., 2021a)	China	LSTM, ARIMA, Fbprophet Algorithm	Daily	

Table 2-1.: Short-Term Electricity Consumption Model Research

3. Modeling Hourly Electricity Consumption

Reducing errors in electricity consumption forecasting streamlines daily operations and the scheduling of generation units, thereby avoiding unnecessary expenses and ensuring service quality, safety, and reliability. If the forecast is lower than the actual consumption, a re-dispatch is performed using more expensive generation units to supply the demand. If the forecast is higher than the actual consumption, operating cost overruns due to the startup of generation units that were not required. Forecast errors can lead to increases in the price of electricity to end users (Valencia et al., 2007). This section explores different methodologies for modeling the hourly electricity consumption time series in Antioquia-Colombia. The models are evaluated over a seven-day time horizon, covering 24 hours per day, resulting in a total of 168 hours ahead.

3.1. Electricity Market Overview

Colombia's electricity market operates primarily under Laws 142 and 143 of 1994, enacted by the Congress of the Republic of Colombia (Santa María et al., 2009). These laws govern activities such as generation, interconnection, transmission, distribution, and commercialization of electric energy (Garcia et al., 2023). The functions assigned to the main entities in the sector are as follows: **(i) Direction**: Ministerio de Minas y Energía, **(ii) Planning**: Unidad de Planeación Minero Energética (UPME), **(iii) Regulation**: Comisión de Regulación de Energía y Gas (CREG), **(iv) Market Operation and Administration**: XM Compañía de Expertos en Mercados S.A., **(v) Council and Committee**: Consejo Nacional de Operación (CNO), Comité Asesor de Comercialización (CAC) and **(vi) Control and Surveillance**: Superintendencia de Servicios Públicos. The market is composed of users classified as regulated, non-regulated and agents (XM, 2023):

1. **Regulated**: Natural or legal person whose electricity purchases are subject to rates established by the Energy and Gas Regulation Commission. This includes most commercial, official, and residential users classified by socioeconomic strata, as well as some industrial users.

2. **Non-Regulated**: Natural or legal person making an energy demand exceeding 2 megawatts (2 MW). They can freely negotiate the costs of activities related to energy

generation and commercialization. At this consumption level, there are industrial and commercial entities that are large consumers.

3. **Agents**: Electricity market agents perform specific roles in the production, transportation, and sale of energy to the end-user. They are divided into generators, transmitters, distributors, and marketers based on their responsibilities. Generators produce energy through thermal, hydraulic, and wind power plants. Transmitters transport it over long distances from power plants to transformation substations. Distributors deliver energy to the end consumer, and marketers buy energy in the wholesale market and sell it to end-users.

According to (CNO, 2021) the Market Operator, *XM Compañía de Expertos en Mercados S.A.*, is responsible for creating and distributing the official forecast for hourly electricity demand in Colombia. The official forecasts are evaluated based on the number of daily deviations that have occurred above 5 % compared to the actual demand for each Electricity Marketing Market. This situation may pose a risk to ensuring a secure and reliable supply of demand. If the daily deviations exceed 5 % for two consecutive days in an Electricity Marketing Market or a load connected to the National Transmission System, the responsible forecasting entity must:

1. Perform a comprehensive analysis of the deviations and propose improvement actions, reporting to the Centro Nacional de Despacho (CND) within the next two business days.

2. Adjust the daily energy demand forecast for the next 7 days, starting from the day following the CND's deviation report.

3. Evaluate the effectiveness of the improvement actions and report to the Distribution Committee and the CND using a format defined by the Committee within the first 10 days of the month following the occurrence of the deviations. This process is crucial for ensuring the reliability and accuracy of energy demand forecasts, contributing to the overall efficiency and stability of the electricity market.

3.2. Description Data and Time Series Characteristics

The dataset corresponds to the hourly electricity consumption of Antioquia-Colombia for the period from 2017-12-01 00:00 to 2023-10-31 23:00 measured in megawatt-hours (MWh). The main source of the information is provided by *XM Compañía de Expertos en Mercados S.A.*, the operator of the National Interconnected System and administrator of the Colombian Wholesale Energy Market.

A preliminary attempt to identify the model that best represents the temporal evolution of electricity demand in Antioquia-Colombia is presented in the following graphical analysis:

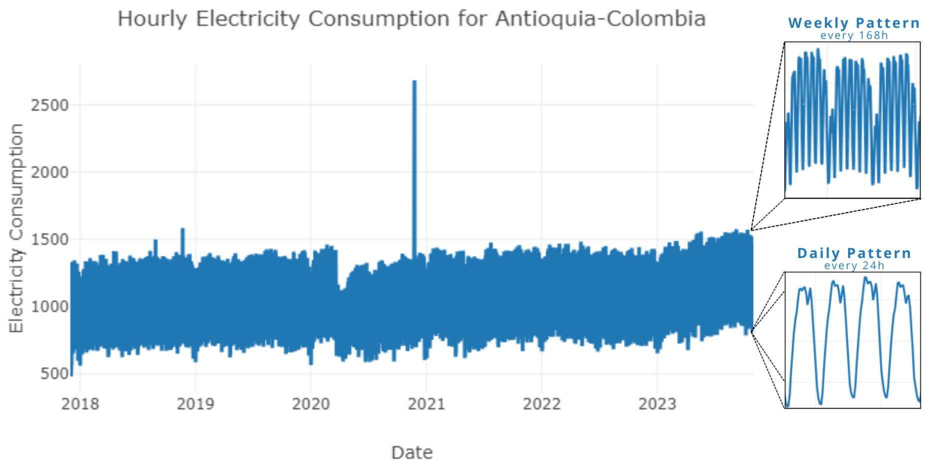


Figure 3-1.: Hourly Electricity Consumption (MWh)

The **Figure 3-1** illustrates the evolution of hourly electricity demand in Antioquia-Colombia, which oscillated around 1000 MWh until 2020. The COVID-19 health emergency induced a structural change in the time series trend and it was not until mid-2021 that it returned to these levels. It should be noted that electricity consumption in Antioquia-Colombia follows recurring patterns with similar intensity and frequency. These patterns exhibit a seasonal period observed both weekly (every 168 hours) and daily (every 24 hours), along with an atypical value recorded in November 2020.

Considering the above, when modelling hourly electricity consumption in Antioquia-Colombia, it is crucial to take into account three key aspects: (i) incorporating a dummy variable to identify the pandemic period, (ii) including up to 168 lags of electricity demand to capture the seasonal behaviour of the time series, and (iii) adding a dummy variable to identify the outlier that occurred in November 2021.

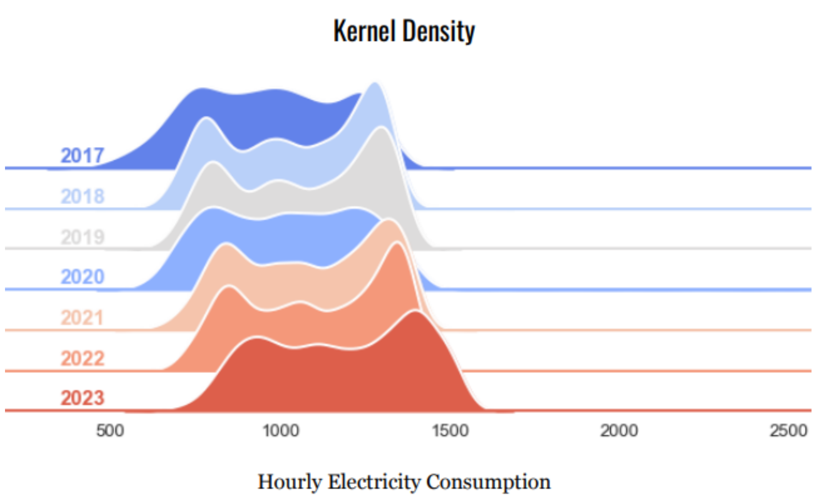


Figure 3-2.: Kernel Density Estimation by Year

The **Figure 3-2** shows the estimated density by kernel method for hourly electricity consumption in Antioquia-Colombia. The distribution process of electricity demand is non-normal with two peaks or modes around 900 MWh and 1400 MWh. Moreover, it does not change drastically, except for 2020, but there are slight changes around its mean level. This is reflected by the slight rightward shifts in the density functions per year. It is worth noting that the year 2017 was not included in this analysis of distributive processes due to its lack of representation, as it only has records corresponding to a single month.

Although year could be considered as a potential explanatory variable, the similarity in the density functions of electricity consumption by year suggests that it may not provide useful information for modeling the evolution of the time series. The most significant change in the density functions of hourly electricity consumption was observed in 2020, which would be captured by the dichotomous variable considered for the pandemic period.

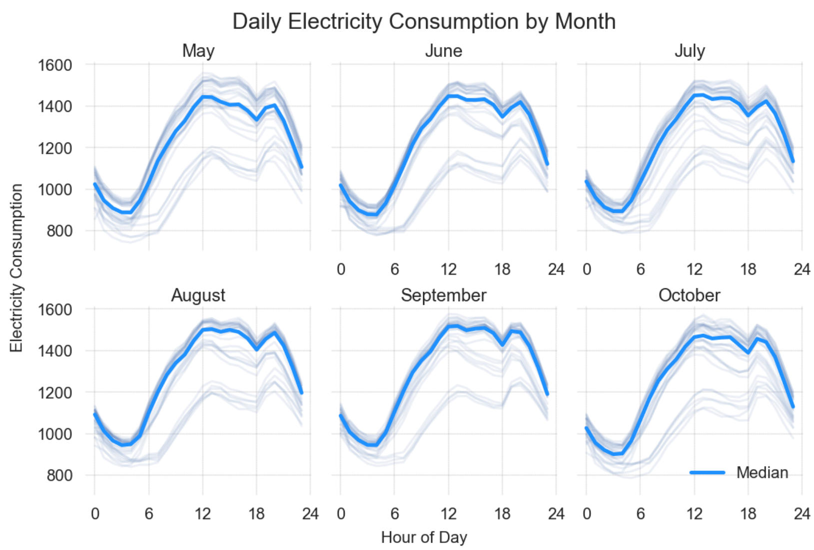


Figure 3-3.: Electricity Consumption for six months in 2023 (MWh)

The **Figure 3-3** represents the daily evolution of hourly electricity consumption in Antioquia-Colombia over the last six months for which data is available. During these months, the daily electricity demand fluctuated between 800 MWh and 1500 MWh. The lowest consumption occurs between 00:00 and 06:00, while the highest demand is observed between 12:00 and 19:00. Two peaks in electricity consumption occur around 12:00 and 19:00 which are likely due to lunch and dinner hours, as well as people finishing their workday and returning home for leisure activities, including watching television.

Based on these characteristics of the time series, it can be assumed that the hour and month variables are important factors to include in the modeling of electricity consumption in Antioquia-Colombia. They can be useful to improve the identification of seasonal effects within the time series and thus contribute to the predictive performance of the model.

The **Figure 3-4** shows the behavior of hourly electricity consumption along a typical seven-day week in 2023 for Antioquia-Colombia. It is observed that the highest electricity demand is experienced during the week while electricity consumption trends to be lower on weekends. Trend changes in electricity consumption can also be observed on holidays, days before and after holidays, Easter week, vacation periods, Colombian Independence Day, New Year's Day, among others. Hence, adding a day-type explanatory variable that captures the variability of electricity consumption presented on weekdays, weekends, and holidays could help build an appropriate model for forecasting hourly electricity demand.

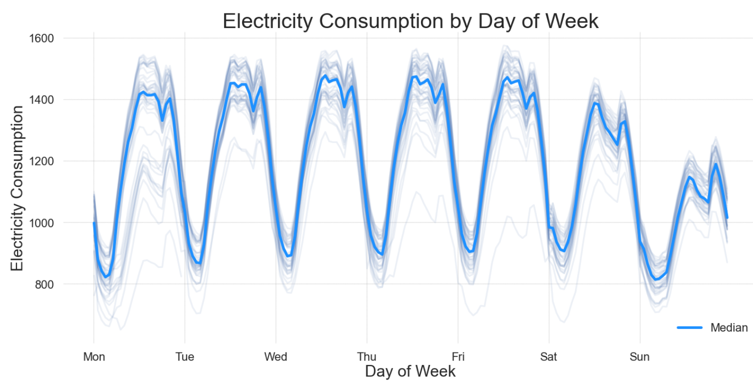


Figure 3-4.: Electricity Consumption by Day of Week in 2023 (MWh)

The Market Operator, XM, has identified 42 different categories of days in which electricity consumption could experience significant trend changes which are detailed below:

Day-Type	Description	Day-Type	Description
1–ENE	January 1st.	MSS	Wednesday of Easter Week
1–MAY	May 1st	JSS	Wednesday of Easter Week
20–JUL	July 20th	JUEVES	Thursday of Easter Week
24–DIC	December 24th	JUVDIC	Thursday of December Vacation
25–DIC	December 25th	JUVENE	Thursday of January Vacation
2–ENE	January 2nd	VIERNES	Friday
31–DIC	December 31st	VIVDIC	Friday of December Vacation
7–AGO	August 7th	VIVENE	Friday of January Vacation
8–DIC	December 8th	VSS	Friday of Easter Week
LF	Holiday Monday	SAALF	Saturday before a Holiday Monday
LFENE	Holiday Monday in January	SAALFENE	Saturday before a Holiday Monday in January
LUNES	Monday	SABADO	Saturday
LUVDIC	Monday of December Vacation	SAVDIC	Saturday of December Vacation
LUVENE	Tuesday of January Vacation	SAVENE	Saturday of January Vacation
MADLF	Tuesday following a Holiday Monday	SSS	Saturday of Easter Week
MARTES	Tuesday	DOALF	Sunday before a Holiday Monday
MAVDIC	Tuesday of December Vacation	DOALFENE	Sunday before a Holiday Monday in January
MAVENE	Tuesday of January Vacation	DOMINGO	Sunday
MIERCOLES	Wednesday	DOVDIC	Sunday of December Vacation
MIVDIC	Wednesday of December Vacation	DOVENE	Sunday of January Vacation
MIVENE	Wednesday of January Vacation	DSS	Sunday of Easter Week

Table 3-1.: Day-Types

It has been determined that the hourly electricity consumption in Antioquia-Colombia has very strong seasonal patterns in which the month, hour and day-type are fundamental for modeling. To determine unambiguously the seasonal component of the time series, the Multiple Seasonal-Trend Decomposition using Loess (MSTL) technique was employed, which is a useful method for analyzing time series with multiple seasonal patterns proposed by (Bandara et al., 2021):

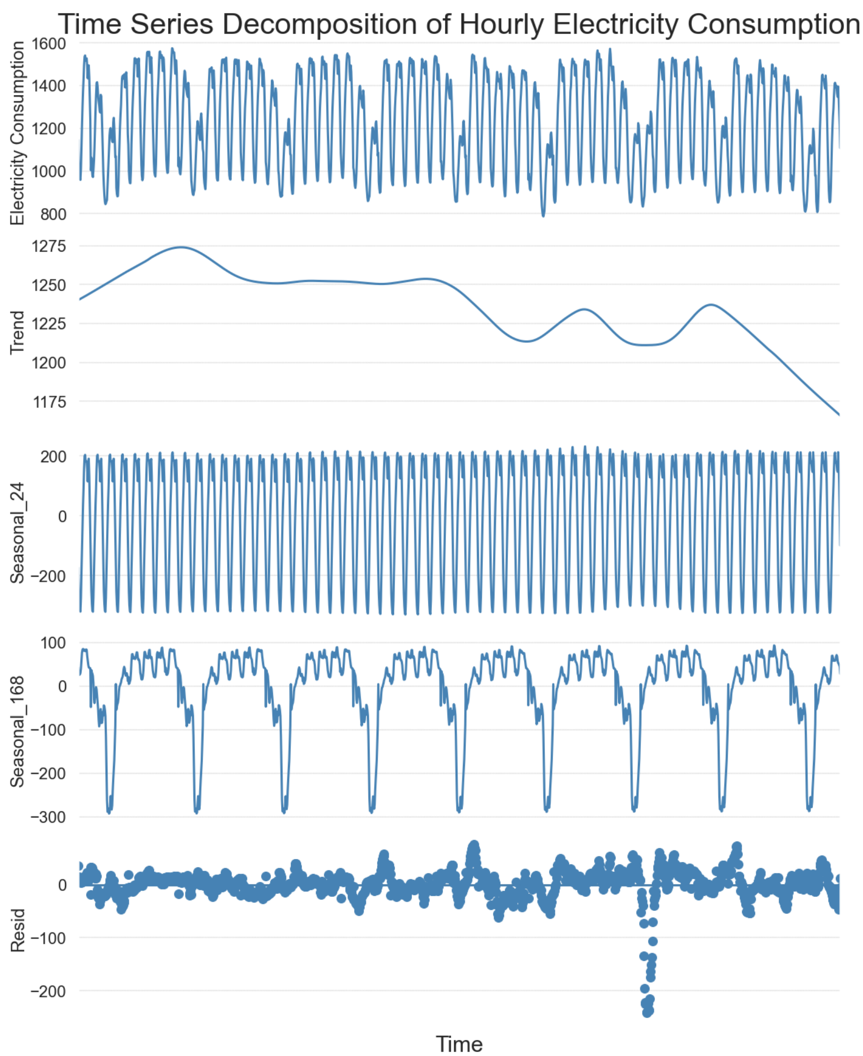


Figure 3-5.: MSTL Decomposition of Electricity Consumption (MWh)

The **Figure 3-5** displays the MSTL decomposition of electricity demand for September and October 2023 to provide a detailed visualization of seasonal patterns. The graph illustrates that hourly electricity consumption has at least two seasonal cycles: a daily cycle (each 24h) and a weekly cycle (each 168h). Note that the weekly cycle captures the behavior of the daily cycle, providing additional information that can be useful when modeling electricity consumption. Therefore, it is recommended to include up to 168 lags.

3.3. Modeling

3.3.1. Dataset

Based on the characteristics of the time series and literature recommendations, the selected dataset for constructing a model that adequately reproduces the evolution of hourly electricity consumption in Antioquia-Colombia is as follows:

Variable	**Description**	**Variable Type**	**Source**
Y-lags	Historical data, including 168 lags.	Float	XM
Month	Month of year	Category	
Hour	Time of Day	Integer	
Day-Type	Weekdays, Weekends, Holidays, EasterWeek,...	Category	
Pandemic-Period	Covid-19 period: between March 2020 and June 2022	Integer	Own work
Outlier-Value	Outlier on November 23rd, 2020	Integer	
Hour-Sin	Sine transformation of the hour variable: $sin\left(\frac{Hour}{23 \cdot 2\pi}\right)$	Float	
Hour-Cos	Cosine transformation of the hour variable: $cos\left(\frac{Hour}{23 \cdot 2\pi}\right)$	Float	
Temperature	Maximum Temperature of Antioquia-Colombia	Float	IDEAM

Table 3-2.: Explanatory Variables Dataset

The inclusion of atmospheric variables is a common suggestion in the literature to improve the predictive performance of the model. However, only the maximum temperature variable is considered in this study. In addition, the hour variable is used in sine and cosine transformations. This implementation of trigonometric functions such as sine and cosine enables the representation of cyclic patterns, helps avoid inconsistencies in data representation, and can enhance the predictive capability of models (Gomez, 2001).

One reason why socioeconomic variables are not considered is that the movements in these time series primarily generate long-term effects on electricity consumption, but they do not necessarily influence short-term fluctuations predictably. The concept of cointegration plays a crucial role in this scenario, according to (Engle and Granger, 1987; Johansen, 1991); two or more series may be cointegrated, indicating a long-term equilibrium relationship, while their short-term dynamics may be independent and not directly influenced by the long-term relationship (Diebold and Rudebusch, 1999; Hamilton, 1994). Another reason is that electricity producers and traders adjust their operating schedules and choose the best short-term energy supply strategy based on the information available at the time of electricity consumption forecasting (Li et al., 2021). Consequently, it is not feasible to incorporate socioeconomic variables in hourly electricity supply decisions.

3.3.2. Data Matrix Preprocessing

The structure dataset used to train Machine Learning models for forecasting a time series requires an adjustment where each value is associated with the preceding time window.

Therefore, it depends on the number of lags to be included in the models. Given a univariate time series $\{y_i\}_{i=1}^{n}$ and exogenous variable $\{x_i\}_{i=1}^{n}$ comprising n-observations, we can adjust the dataset for model training as follows:

Y: Time Series

y_1	y_2	y_3	y_4	y_5	...	y_n

X: Exogenous Variable

x_1	x_2	x_3	x_4	x_5	...	x_n

$\Rightarrow$

Training data matrix

Y	**Y-lags**						**X**
y_{p+1}	y_p	y_{p-1}	y_{p-2}	y_{p-3}	...	y_1	x_{p+1}
y_{p+2}	y_{p+1}	y_p	y_{p-1}	y_{p-2}	...	y_2	x_{p+2}
y_{p+3}	y_{p+2}	y_{p+1}	y_p	y_{p-1}	...	y_3	x_{p+3}
...	...	...	...	...	...	...	...
y_n	y_{n-1}	y_{n-2}	y_{n-3}	y_{n-4}	...	y_{n-p}	x_n

Table 3-3.: Data Structure for Training Machine Learning Models

Where $p \in \mathbb{N}$ denotes the number of lags to be included in the model which implies that the first p records in the database are lost. Once the data has been restructured in this way, it is possible to estimate machine learning models for time series.

3.3.3. Recursive Multi-Step Forecasting

In the time series framework, there is typically a requirement to predict more than one future value. Therefore, forecasting in a Machine Learning model with target variable lags must follow a process where each prediction uses the previous one. As stated in (Bontempi et al., 2013) a multi-step forecasting strategy involves predicting the values for the next h-steps $[y_{n+1}, ..., y_{n+h}]$ of a historical time series $[y_1, ..., y_n]$ composed of n-observations, where $h > 1$ denotes the forecasting horizon and p denotes the number of lags to be included in the model. We want to forecast the next h-values of the time series:

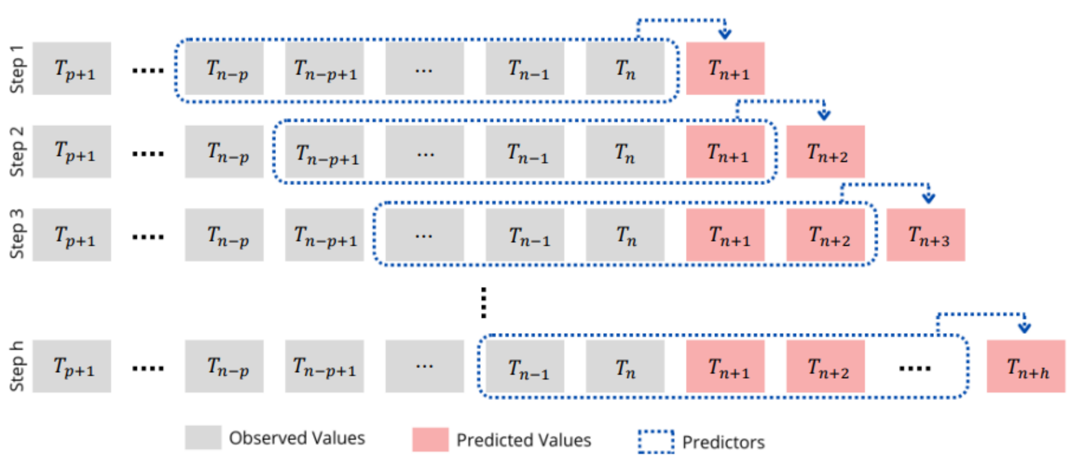


Figure 3-6.: Recursive Multi-Step Forecasting

The recursive method has demonstrated success in forecasting real-world time series using diverse machine learning models. However, a notable drawback lies in its susceptibility to estimation errors, as the reliance on estimated values grows with the forecasting horizon, diverging from actual values. Nevertheless, it is the most appropriate way to implement models in a real-world production environment. Applications often implement direct forecasting, but it is important to consider that this approach lacks realism when predicting multiple values of a time series and including lags as explanatory variables (Bontempi et al., 2013).

3.3.4. Hyperparameter Tuning and Cross-Validation

Hyperparameters are settings of the learning algorithm which are freely chosen within a certain range and influence model performance. While model parameters are chosen by the model itself during the learning process. It is important to understand that hyperparameters are different from these parameters. As explained in (Bartz et al., 2023) hyperparameter tuning is the determination of the best possible hyperparameters using tools to explore the space of possible hyperparameter settings in a systematic and structured approach.

Hyperparameter Tuning can be formulated as an optimization problem since, as mentioned by (Hastie et al., 2009), the objective of a learning algorithm $\mathcal{A}$ is to find a function f that minimizes some expected loss $\mathcal{L}(y, f(x))$ over samples $(x, y) \in (X, Y)$. This requires split dataset into three parts: (i) a training set $(X, Y)^{(train)}$ used to fit the models, (ii) a validation set $(X, Y)^{(val)}$ to estimate prediction error for model selection and (iii) a test set $(X, Y)^{(test)}$ used to assess generalization error:

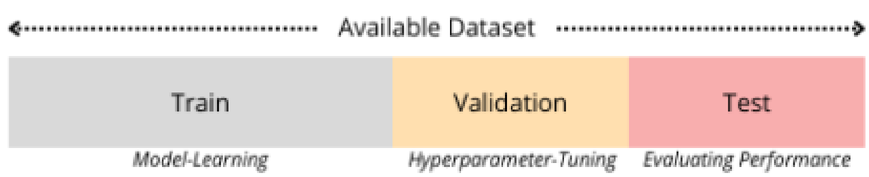


Figure 3-7.: Split Available Dataset for Hyperparameter-Tuning

In this case study, the training set collects information records from 2017-12-08 00:00 to 2023-04-23 23:00, the validation set from 2023-04-24 00:00 to 2023-10-24 23:00 and the test set from 2023-10-25 00:00 to 2023-10-31 23:00. The learning algorithm can compute an estimate of $f : \mathcal{X} \to \mathcal{Y}$ through optimization of a training criterion with respect to a set of parameters $\lambda \in \Lambda$.The hyperparameter optimization can be formulated as follows:

$$\lambda^{(*)} = \arg\min_{\lambda \in \Lambda} E_{(x,y)\in(\mathcal{X},\mathcal{Y})} \left[\mathcal{L}\left(y, \mathcal{A}_\lambda((X,Y)^{(train)})\right)\right] \tag{3-1}$$

However, the underlying space $(\mathcal{X}, \mathcal{Y})$ can be too large or the true relationship between $\mathcal{X}$ and $\mathcal{Y}$ is unknown. Therefore, it is useful to use the following equation:

$$\lambda^{(*)} \approx \arg\min_{\lambda \in \Lambda} \frac{1}{|(X,Y)^{(val)}|} \sum_{x \in (X,Y)^{(val)}} \mathcal{L}\left(y, \mathcal{A}_\lambda((X,Y)^{(train)})\right) \tag{3-2}$$

It is worth noting that the test error estimate obtained from a single hold-out test set usually has a high variance. For this reason (Hastie and Tibshirani, 2017) concluded that Cross-Validation (CV) and related methods may provide reasonable estimates of the expected error.

According to (Hastie and Tibshirani, 2017) is one of the most widely used methods to assess the generalizability of learning algorithms. However, when it comes to time series forecasting, the standard application of the Cross-Validation method can be problematic when using future data to predict the past (Bergmeir et al., 2018). A widely used adaptation involves conducting Cross-Validation on a rolling basis. In this method, the learning model undergoes training prior to each prediction using all data available up to the given time. This approach deviates from the conventional Cross-Validation technique, which arbitrarily splits the data into training and validation sets.

Rather than employing randomization, this method systematically expands the training set, preserving the temporal sequence of the data. This sequential expansion allows the learning model to be evaluated on incrementally larger segments of historical data, resulting in a more precise evaluation of its predictive performance. The standard practice for evaluating a model found by Cross-Validation is to report the hyperparameter configuration that minimizes the loss on the validation data as presented in (3-2).

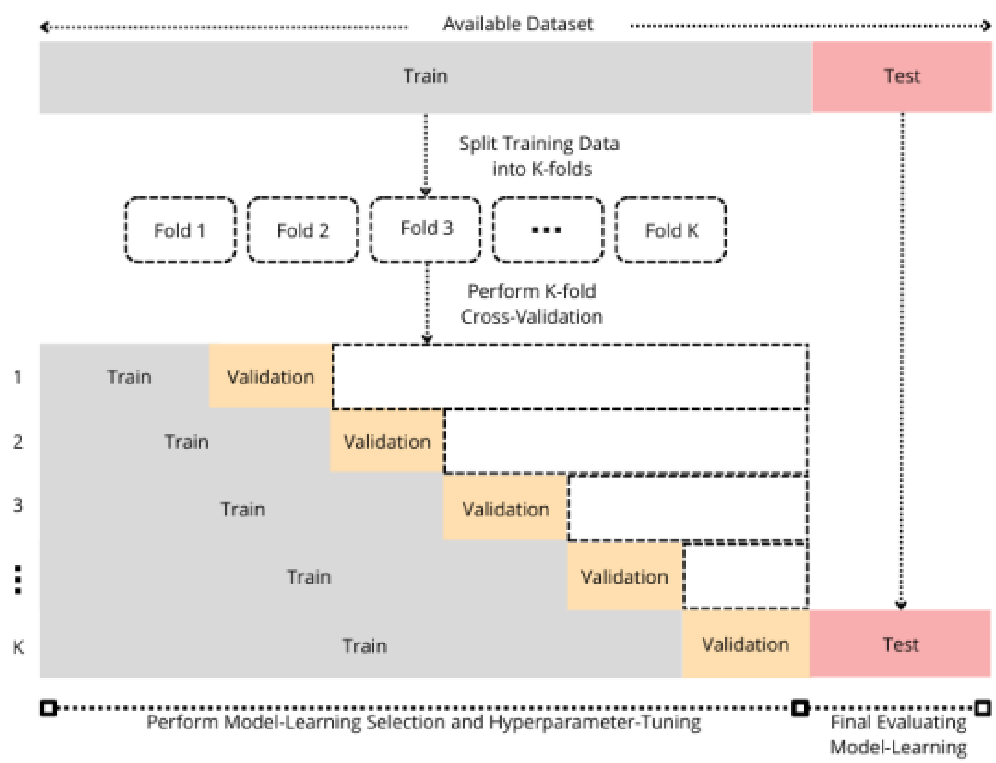


Figure 3-8.: Cross-Validation on Time Series

This learning process suggests that the optimal hyperparameters for model estimation are as follows: (i) LASSO with a penalty parameter of 0.01. (ii) GAM with a penalty parameter of 0.3, five partitions and fourth-degree polynomials. (iii) MARS with a penalty parameter of 0.07, four partitions, second-degree polynomials and interactions between the month, hour and day-type variables. (iv) XGBoost with 500 boosting rounds, a maximum tree depth of 10 and a learning rate of 0.1. (v) LSTM with a ReLu activation function and three neurons with 45, 25 and 10 hidden layers, respectively.

3.3.5. Predictive Performance of the Models

The assessment of a model's predictive performance is crucial to measure its ability to anticipate future outcomes. In the framework of a country's energy policy, an accurate forecast allows for improved provision of energy distribution services and optimization operating costs in the electrical power system. As forecast error are reduced, the supply strategy achieves greater efficiency, fostering economic sustainability and stability of energy resources, which is fundamental to business continuity.

In Colombia, the electric power market has faced several challenges in providing energy distribution services (Jimenez Mares et al., 2019), which could be mitigated by maintaining a balance between energy generation and demand through the selection of a supply strategy based on forecasting information. This enables timely decisions necessary for energy transactions in the market between producers and consumers (Li et al., 2021). Furthermore, an accurate demand forecast significantly influences the price of electricity in the power exchange, serving as a benchmark for all other transactions and contracts, thus playing a crucial role in the competition of the deregulated energy market(Valencia et al., 2007).

To reduce hourly electricity consumption forecasting errors in Antioquia-Colombia, this study employs six models that use parametric, semi-parametric, and non-parametric approaches. This diversified approach aims to enhance prediction accuracy by comprehensively addressing the inherent complexity of electricity consumption data variability in the region. The models will be evaluated in two scenarios: one without maximum temperature as an explanatory variable and another including maximum temperature. The second scenario aims to assess whether adding maximum temperature as an explanatory variable improves the predictive performance of Statistical-Machine Learning models. It is important to note that this assumes the reliability of maximum temperature forecasts provided by the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM[1], Spanish acronym).

In the search for effective models for electricity demand forecasting, the results presented in the **Table 3-4** provide insight into the predictive performance of in-sample models:
In the first scenario (1), where the explanatory variable of maximum temperature is not

[1] `www.ideam.gov.co`

	Metric	LSTM	XGBoost	GAM	MLR	LASSO	MARS
1	MAE	17.64	**1.65**	13.58	14.14	14.14	46.36
	RMSE	548.77	**5.02**	458.95	505.92	506.04	3372.81
	MAPE	1.649 %	**0.157 %**	1.360 %	1.420 %	1.419 %	4.479 %
2	MAE	15.72	**1.58**	13.47	14.03	14.03	46.36
	RMSE	458.64	**4.56**	453.13	500.22	500.34	3372.83
	MAPE	1.500 %	**0.150 %**	1.351 %	1.411 %	1.410 %	4.479 %

Table 3-4.: In-Sample Predictive Performance

included, XGBoost stands out notably, displaying lower values in MAE, RMSE, and MAPE compared to other models. Similarly, upon incorporating the temperature variable in the second scenario (2), all models experience slight improvements in the performance metrics. The **Table 3-5** reports the predictive performance of the out-of-sample models. The models are evaluated over a seven-day time horizon, covering 24 hours per day, resulting in a total of 168 hours ahead. The testing horizon for the models aligns with the forecast horizon of the XM Market Operator. The operator is required to provide hourly forecasts of electricity consumption for a week, as regulated by (CNO, 2021).

	Metric	LSTM	XGBoost	GAM	MLR	LASSO	MARS	XM
1	MAE	**17.4**	28.09	34.33	38.76	38.75	88.94	50.59
	RMSE	**493.43**	1211.19	1763.34	2134.45	2143.41	12692.63	3420.27
	MAPE	**1.445 %**	2.363 %	3.088 %	3.469 %	3.468 %	7.960 %	4.358 %
2	MAE	**18.62**	24.63	31.39	35.59	35.59	88.78	50.59
	RMSE	**507.81**	930.59	1545.51	1873.84	1881.72	12658.47	3420.27
	MAPE	**1.552 %**	2.089 %	2.821 %	3.193 %	3.195 %	7.946 %	4.358 %

Table 3-5.: Out-of-Sample Predictive Performance

In the first scenario, the estimated models, except for MARS, demonstrate higher predictive power than the XM Market Operator model. LSTM achieves the most favorable result in terms of predictive performance, significantly outperforming the forecast accuracy provided by the XM Market Operator's benchmark model. Specifically, LSTM achieves a MAPE of 1.445 % and an RMSE of 493.43, while XM's reference model achieves a MAPE of 4.358 % and an RMSE of 3420.27. The XGBoost, GAM, MLR, and LASSO models produced satisfactory results, with a MAPE of 2.363 %, 3.088 %, 3.468 %, and 3.468 %, respectively. However, the MARS model showed inferior predictive performance compared to the reference model, achieving only a MAPE of 7.96 %.

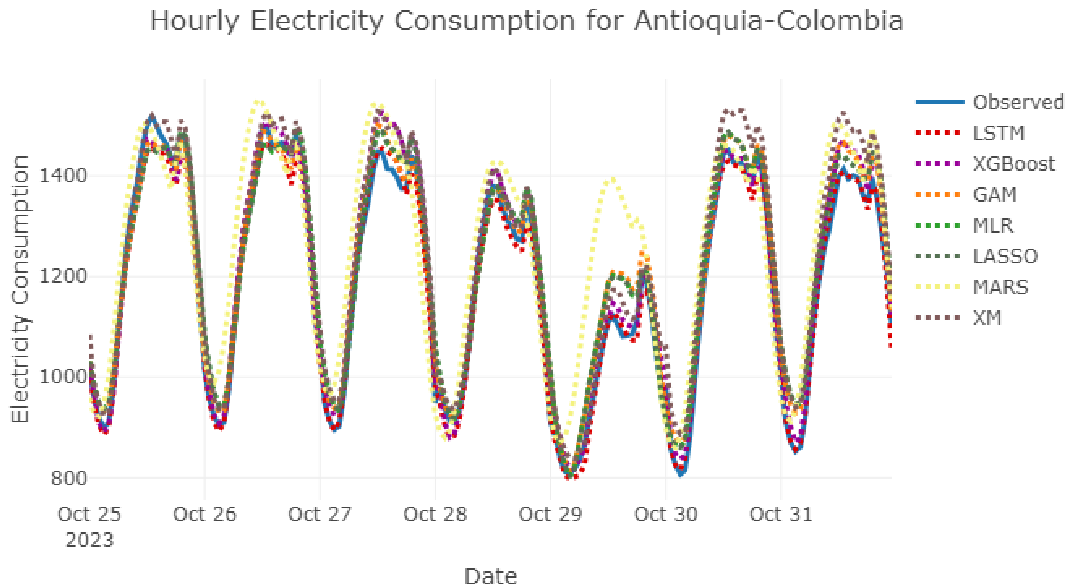
Interpreting these results within the framework of generation unit scheduling has crucial implications for ensuring adequate supply of demand. The average electricity supply error would decrease from 3420 MWh to 493 MWh, representing a significant optimization of energy resources of up to 2927 MWh, which would help reduce the cost overruns associated with the provision of electricity distribution services.

In the second scenario, considering the maximum temperature as an explanatory variable improves the predictive performance of the XGBoost, GAM, MLR, LASSO and MARS models, which achieved MAPE values of 2.089 %, 2.821 %, 3.193 %, 3.195 % and 7.946 %, respectively. However, there was a slight decline in the predictive performance of the LSTM model, which achieved a MAPE of 1.552 %.

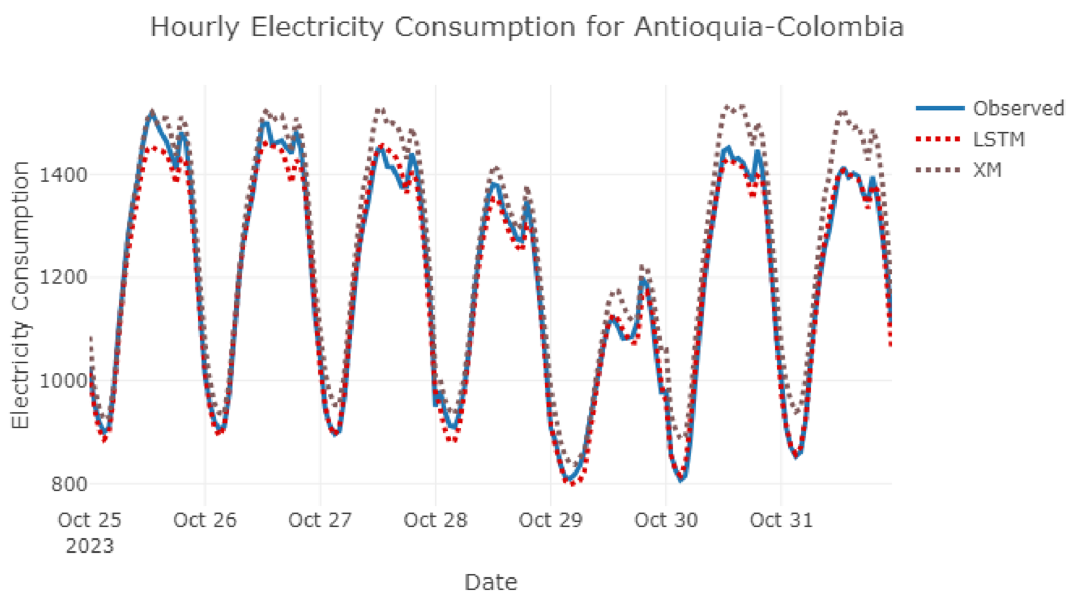
These findings suggest that incorporating maximum temperature as a variable can enhance the modeling of hourly electricity consumption using parametric, semi-parametric, and non-parametric approaches. However, relying on this variable could increase the complexity of the modeling process due to the direct transfer of uncertainty associated with temperature variability to electricity demand forecasts. The LSTM model's predictive capability is advantageous as it does not necessarily depend on the maximum temperature explanatory variable to achieve accurate forecasts. This indicates that LSTM has a remarkable ability to capture complex patterns in the data and effectively adapt to the behavior of electricity consumption, providing hourly forecasts with greater reliability for the daily operation of electricity distribution services.

The **Figures 3-9** and **3-10** illustrate the comparison between actual and predicted demand by the different models over the test horizon. One noteworthy observation from the figures above is that the XM Market Operator's model exhibits higher forecast errors on October 30th and 31st with daily deviations exceeding 5 % consecutively. Typically, electricity consumption is lower on holidays than on weekdays. However, XM's calculation of the day-type variable does not take into account the Halloween holiday. This could be a plausible explanation for the Market Operator's overestimation of electricity demand during these days. Therefore, it is recommended to closely monitor electricity demand during holidays. Currently, there are no official available documents regarding the methodological processes used by the Market Operator to calculate hourly electricity consumption forecasts in Colombia. Thus, the employed model remains a mystery, making it difficult to provide more detailed suggestions and to discuss the best methodological approaches.

The following Diebold-Mariano test aims to evaluate the significance of differences in forecasting accuracy between the models over the test horizon. The null hypothesis is that the two methods have the same forecast accuracy and the alternative hypothesis is that method j is less accurate than method i, as stated in equation (2-29) and (2-32). The $P-Values$ correspond to the entries of the matrix and the $DM-Statistics$ is associated with the bar located on the right side of the **Figure 3-11**. The main diagonal is not considered in the analysis.

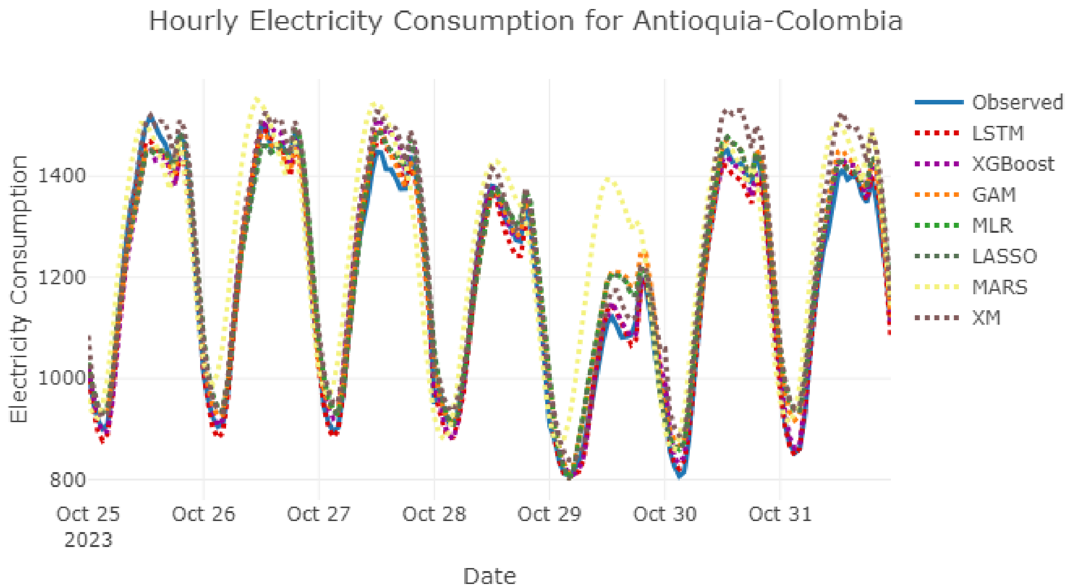


((a)) All Models

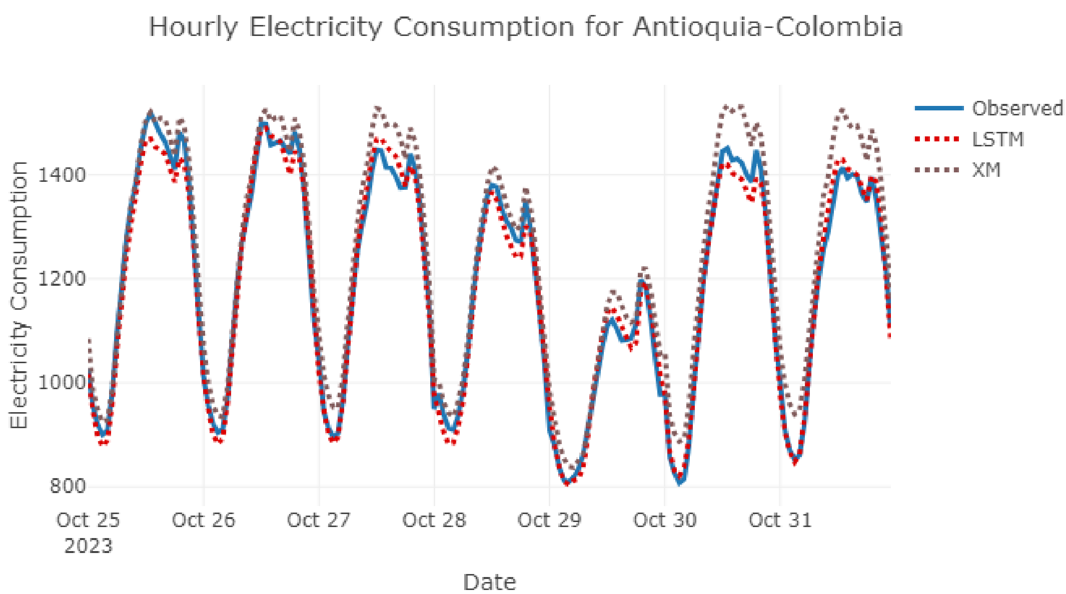


((b)) LSTM Vs XM

Figure 3-9.: Forecasting without maximum temperature as an exogenous variable



((a)) All Models



((b)) LSTM Vs XM

Figure 3-10.: Forecasting with maximum temperature as an exogenous variable

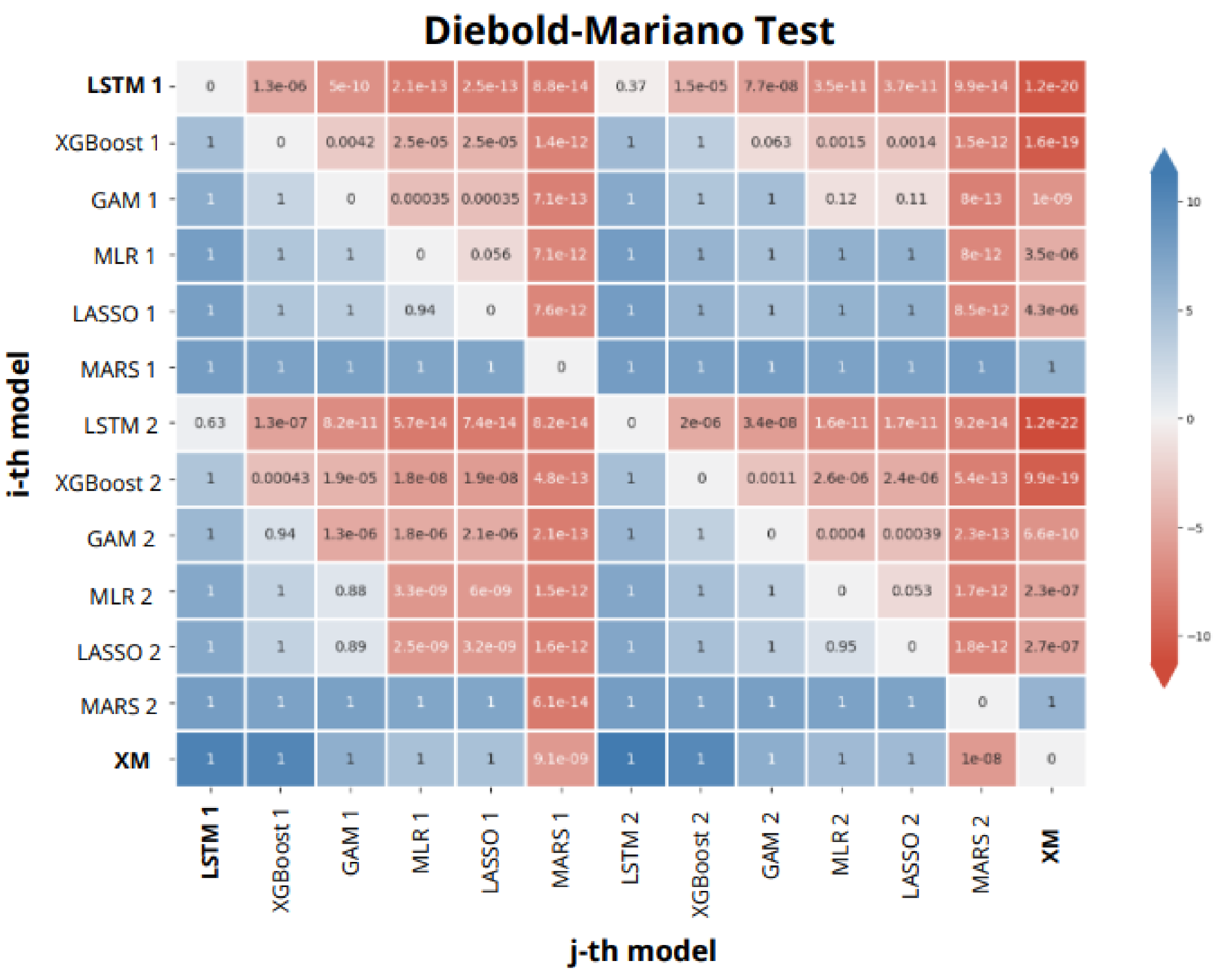


Figure 3-11.: Diebold-Mariano Test for Comparing Prediction Accuracy

When the $P-Value$ is less than $\alpha = 0{,}05$, there is enough evidence to reject the null hypothesis and therefore to conclude that (i) the LSTM model, without the maximum temperature variable, produces more accurate forecasts than the XM Market Operator, XGBoost, GAM, MLR, LASSO and MARS models. (ii) The models that incorporate the maximum temperature variable provide more accurate forecasts than those that do not include it, except for the LSTM model. In this case, the null hypothesis is not rejected, and both models are considered to have the same forecasting power. (iii) The predictive ability order for scenarios without the maximum temperature variable is LSTM 1 $>$ XGBoost 1 $>$ GAM 1 $>$ MLR 1 $>$ LASSO 1 $>$ XM $>$ MARS 1. For the scenario where the maximum temperature variable is included, the order is LSTM 2 $>$ XGBoost 2 $>$ GAM 2 $>$ LASSO 2 $>$ MLR 2 $>$ XM $>$ MARS 2. Therefore, all estimated models, except for MARS, achieve better predictive performance than the XM Market Operator's reference model.

4. Conclusions

4.1. Conclusions

Electricity consumption forecasting is a valuable instrument for policymakers to guide pricing, taxation, investment decisions, as well as energy and operational security planning, helping to ensure a continuous supply of electricity and reducing cost overruns associated with the provision of energy distribution services. Minimizing forecasting errors is fundamental to support the decision-making process of market agents. This research proposes novel methodological approaches that advance this strategic process, particularly for short-term electricity consumption forecasting in Antioquia-Colombia. The hourly forecast errors are lower than those obtained by the XM Market Operator's benchmark model.

The LSTM Neural Network proves to be an efficient model for operational deployment in electricity distribution in Antioquia-Colombia. Its average electricity supply error for an operational week is estimated to be around 493MWh. This represents an optimization of energy resources of up to 2927MWh, as the average supply error of the Market Operator for the evaluated week was 3420MWh. The XGBoost, GAM, and MLR models are attractive alternatives for efficiently allocating energy resources. In contrast, the MARS model could lead to inefficient electricity distribution planning in the market.

The incorporation of external factors, such as maximum temperature, results in statistically significant improvements in the forecasts generated by XGboost, GAM, MLR, LASSO, and MARS models. However, this is not the case for the LSTM model. This suggests that the LSTM model can achieve outstanding predictive capability without relying on the maximum temperature as an exogenous variable. This simplifies the modeling process and avoids potential uncertainty associated with external variables.

4.2. Recommendations

The identification of day-types is helpful to capture trends in hourly electricity demand. However, the information provided by the Market Operator does not identify the Halloween holiday, which could explain the forecast error of its model during this date and the neighboring days. Therefore, it is suggested to evaluate the relevance of including this holiday in the predictive performance of the models.

A. Appendix: software, libraries and examples

The models in this research were computed using Python programming language and supported by the libraries *sklearn*, *xgboost*, *pyGAM*, *pyearth* and *keras*. Below is a programming guide for computing the basic functions of each model:

```
#-------------------------------------------------
# Libraries
#-------------------------------------------------

from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from xgboost import XGBRegressor
from pygam import LinearGAM
from keras.models import Sequential
from keras.layers import Dense, LSTM
import keras.regularizers

#-------------------------------------------------
# Models
#-------------------------------------------------

MLR = LinearRegression(fit_intercept=True).fit(X,y)
LASSO = Lasso(alpha=penalty_parameter, max_iter=1000).fit(X,y)
MARS = Earth(max_degree=max_degree_polynomials, penalty=penalty_parameter,
             smooth=True, allow_missing=True).fit(X,y)
XGBoost = XGBRegressor(tree_method = 'hist', n_estimators=number_trees,
                       max_depth=max_depth_tree, learning_rate=learning_rate,
                       random_state=123).fit(X,y)
GAM = LinearGAM(n_splines = number_splines,
                spline_order = polynomial_order, lam=penalty_parameter
                constraints='none', basis='ps').fit(X,y)
LSTM = Sequential()
LSTM.add(LSTM(hidden_layers_1, input_shape=(1, X.shape[1]),
              activation=activation_function, return_sequences=True,
              dropout=fraction_units_drop))
LSTM.add(Dense(1))
LSTM.compile(loss='mean_squared_error', optimizer='adam')
LSTM.fit(X_reshaped, y_reshaped, epochs=number_epochs,
         batch_size=batch_size, verbose=0)
```

To obtain a structured time series database from the XM data source[1], processing is required. The data were downloaded into separate folders based on time period. Therefore, the following processing was carried out using R programming language:

```
#------------------------------------------------
# Libraries
#------------------------------------------------
# Packages to install
list_packages <- c('pacman','dplyr', 'readxl', 'openxlsx',
                   'lubridate', 'stringr', 'tidyverse')

# Identify which packages are new
new_packages <- list_packages[!(list_packages %in% installed.packages()[,"Package"])]

# Conditional: if the package is new then install it
if(length(new_packages)) install.packages(new_packages)

# Load Packages
pacman::p_load(dplyr, readxl,  openxlsx, lubridate, stringr, tidyverse)

#------------------------------------------------
# Dataset
#------------------------------------------------
# Folder Path: Data storage path
folder_path <- 'Indicadores_de_Pronosticos_Oficiales_de_Demanda'

# Folders
folders <- list.files(path=folder_path, full.names=TRUE)

# Data storage list
data <- list()

#--------------------
# Load databases
#--------------------
# Department name to extract data
departament <- 'antioquia|Antioquia'

# Iterator for loading databases
for (folder in folders){

  # Get the names of all 'xlsx' archives from the folder.
  files <- list.files(path=folder, pattern=".*\\.xlsx$|.*\\.xls$", full.names = TRUE)
  # Retrieve only the files that match: 'antoquia' o 'Antioquia'.
  files <- files[grepl(departament, files)]
  # Iterate through each file in the list of files.
  for (file in files){
```

[1] Available at: https://bitly.ws/3bgGj

```
    # Read each file and store the data in the list 'data'.
    data[[length(data)+1]] <- read_excel(file, sheet = 'real') %>%
                                # Convert column names to uppercase.
                                rename_all(toupper) %>%
                                # Transform column to Datetime with different
                                # formats and create column 'MES'.
                                mutate(FECHA = as.Date(parse_date_time(FECHA,
                                                        orders = c("ymd", "dmy"),
                                                       select_formats='%Y-%m-%d')),
                                MES = factor(month(FECHA))) %>%
                                # Add an underscore character to any column names
                                # that contain spaces.
                                rename_with(~str_replace_all(.x, "\\s", "_"),
                                            .cols = everything()) %>%
                                # Modify column name "VARIABLE"
                                rename_with(
                                           ~str_replace(.x, ".*VAR.*", "VARIABLE"),
                                            .cols = contains("VAR")
                                            )
  }
}

# Concatenate the dataframes in the list by rows.
data  <- bind_rows(data) %>%
         # Select rows from 'FECHA' to 'P24' and 'MES'
         select(FECHA:P24, MES) %>%
         # Create variable 'HORA' and 'Demanda'
         pivot_longer(cols = -c(FECHA, TIPO_DIA, MES),
         names_to = "HORA", values_to = "Demanda") %>%
         # Extract letter 'P' of the variable 'Hora'
         # and replace 24h by 0h
         mutate(HORA = as.numeric(str_replace(HORA, "P", "")),
                HORA = ifelse(HORA==24, 0, HORA)
         # Sort by 'FECHA' and 'HORA' in ascending order
         arrange(FECHA, HORA)
```

References

Arora, S. and Taylor, J. W. (2018). Rule-based autoregressive moving average models for forecasting load on special days: A case study for france. *European Journal of Operational Research*, 266(1):259–268.

Bandara, K., Hyndman, R. J., and Bergmeir, C. (2021). Mstl: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns.

Barrientos, A. F., Olaya, J., and González, V. M. (2007). A spline model for electricity demand forescasting. *Revista Colombiana de Estadística*, 30(2):187.

Barrientos, J., Marquez Marulanda, L., and Villada Duque, F. (2023). Analyzing electricity demand in colombia: A functional time series approach. *International Journal of Energy Economics and Policy*, 13(1):75–84.

Bartz, E., Bartz-Beielstein, T., Zaefferer, M., and Mersmann, O. (2023). *Hyperparameter Tuning for Machine and Deep Learning with R - A Practical Guide*. Springer.

Berbesi, L. and Pritchard, G. (2023). Modelling energy data in a generalized additive modelmdash: A case study of colombia. *Energies*, 16(4).

Bergmeir, C., Hyndman, R. J., and Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics Data Analysis*, 120:70–83.

Bontempi, G., Ben Taieb, S., and Le Borgne, Y.-A. (2013). *Machine Learning Strategies for Time Series Forecasting*, volume 138.

Castaño, E. (2007). Reconstrucción de datos de series de tiempo: una aplicación a la demanda horaria de la electricidad. *Revista Colombiana de Estadística*, 30:247 – 263.

Chapagain, K. and Kittipiyakul, S. (2018). Performance analysis of short-term electricity demand with atmospheric variables. *Energies*, 11.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA. Association for Computing Machinery.

CNO (2021). Consejo nacional de operación: Acuerdo 1303 por el cual se actualizan los procedimientos para la gestión integral de la demanda. Retrieved Dec 10, 2023 from https://bit.ly/3KSCHoI.

CPC (2020). *Informe Nacional de Competitividad 2020-2021.* Bogotá: Consejo Privado de Competitividad.

Diebold, F. and Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business*

Economic Statistics, 13(3):253–63.

Diebold, F. X. and Rudebusch, G. D. (1999). Business cycles: Durations, dynamics, and forecasting. (6636).

Divina, F., Gilson, A., Goméz-Vela, F., García Torres, M., and Torres, J. F. (2018). Stacking ensemble learning for short-term electricity consumption forecasting. *Energies*, 11(4).

DNP (2017). *Energy Demand Situation in Colombia*. Enersinc.

EIA (2023). Energy Information Administration: Electricity explained. Retrieved Dec 1, 2023 from www.eia.gov/energyexplained/electricity.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276.

Garcia, J., Rey Londoño, F., Arango Restrepo, L. J., and Bohorquez Correa, S. (2023). Sectoral analysis of electricity consumption during the covid-19 pandemic: Evidence for unregulated and regulated markets in colombia. *Energy*, 268:126614.

Gomez, D. (2001). Análisis espectral: consideraciones teóricas y aplicabilidad. *Economía y Sociedad*, 6.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Hamilton, J. D. (1994). Time series analysis. 10.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291.

Hastie, T. and Tibshirani, R. (2017). Generalized Additive Models. *Taylor Francis Group.*, pages 249–307.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Series in Statistics. Springer, 2nd ed. 2009. corr. 3rd printing 5th printing. edition.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman Hall/CRC Monographs on Statistics Applied Probability. Chapman and Hall/CRC.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9:1735–80.

James, G., Tibshirani, R., Witten, D., and Hastie, T. (2021). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics. Springer, 2 edition.

Jiang, P., Li, R., Lu, H., and Zhang, X. (2020). Modeling of electricity demand forecast for power system. *Neural Computing and Applications*, 32:1–19.

Jimenez Mares, J., Pertuz, A., Quintero M., C., and Montana, J. (2019). Multivariate statistical analysis based methodology for long-term demand forecasting. *IEEE Latin America Transactions*, 17:93–101.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580.

Khalil, M., McGough, S. A., Pourmirza, Z., Pazhoohesh, M., and Walker, S. (2022). Machine learning, deep learning and statistical analysis for forecasting building energy consumption — a systematic review. *Engineering Applications of Artificial Intelligence*, 115:105287.

Khan, A. M. and Osińska, M. (2023). Comparing forecasting accuracy of selected grey and time series models based on energy consumption in brazil and india. *Expert Systems with Applications*, 212:118840.

Kim, Y.-J. and Gu, C. (2004). Smoothing spline gaussian regression: More scalable computation via efficient approximation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(2):337–356.

Kumar, S., Viral, R., Deep, V., Sharma, P., Kumar, M., Mahmud, M., and Stephan, T. (2021). Forecasting major impacts of covid-19 pandemic on country-driven sectors: challenges, lessons, and future roadmap. *Personal and Ubiquitous Computing*.

La Tona, G., Luna, M., and Di Piazza, M. (2023). Day-ahead forecasting of residential electric power consumption for energy management using long short-term memory encoder–decoder model. *Mathematics and Computers in Simulation*.

Li, R., Chen, X., Balezentis, T., Streimikiene, D., and Niu, Z. (2021). Multi-step least squares support vector machine modeling approach for forecasting short-term electricity demand with application. *Neural Comput. Appl.*, 33(1):301–320.

Maltais, L.-G. and Gosselin, L. (2022). Forecasting of short-term lighting and plug load electricity consumption in single residential units: Development and assessment of data-driven models for different horizons. *Applied Energy*, 307:118229.

Medina, S., Moreno, J., and Gallego, J. P. (2012). Pronóstico de la demanda de energía eléctrica horaria en colombia mediante redes neuronales artificiales. *Revista Facultad de Ingeniería Universidad de Antioquia*, (59):98–107.

Meng, Z., Sun, H., and Wang, X. (2022). Forecasting energy consumption based on svr and markov model: A case study of china. *Frontiers in Environmental Science*, page 363.

Murat, N. (2022). Outlier detection in statistical modeling via multivariate adaptive regression splines. *Communications in Statistics - Simulation and Computation*, 0(0):1–12.

Murillo, J., Trejos, A., and Carvajal, P. (2003). Estudio del pronostico de la demanda de energia electrica, utilizando modelos de series de tiempo. *Scientia Et Technica, 3(23)*.

Peters, J., Janzing, D., and Schlkopf, B. (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press.

Ribeiro, A., Carmo, P., Silva, I., Sadok, D., Lynn, T., and Endo, P. (2020). Short-term firm-level energy-consumption forecasting for energy-intensive manufacturing: A comparison of machine learning and deep learning models. *Algorithms*, 13:274.

Rosero, J., Gonzalez, F. A., Zambrano, A., and Toledo-Cortesa, S. (2023). Short-term forecasting of power demand: A case study with data from colombia. *SSNR*.

Sadeghian Broujeny, R., Ben Ayed, S., and Matalah, M. (2023). Energy consumption forecasting in a university office by artificial intelligence techniques: An analysis of the exogenous data effect on the modeling. *Energies*, 16(10).

Safer, A. M. (2004). Multivariate adaptive regression splines and insider trading data for stock prediction. *Journal of Interdisciplinary Mathematics*, 7(1):79–93.

Santa María, M., Von Der Fehr, N.-H., Millán, J., Benavides, J., and Gracia, O. (2009). El mercado de la energía eléctrica en colombia: características, evolución e impacto sobre otros sectores. *Bogotá: Fedesarrollo*, 304 p.

Saranj, A. and Zolfaghari, M. (2022). The electricity consumption forecast: Adopting a hybrid approach by deep learning and arimax-garch models. *Energy Reports*, 8:7657–7679.

Sarmiento, H. O. and Villa, W. M. (2008). Inteligencia artificial en pronóstico de demanda de energía eléctrica: una aplicación en optimización de recursos energéticos. *Revista Colombiana de Tecnologías de Avanzada*.

Son, J., Cha, J., Kim, H., and Wi, Y.-M. (2022). Day-ahead short-term load forecasting for holidays based on modification of similar days' load profiles. *IEEE Access*, 10:17864–17880.

Sujan, R. A., Akashdeep, S., Harshvardhan, R., and Sowmya, K. S. (2022). Stacking deep learning and machine learning models for short-term energy consumption forecasting. *Advanced Engineering Informatics*, 52:101542.

UPME (2021). *Proyección de Demanda de Energía Eléctrica, Gas Natural y Combustibles Líquidos 2022-2036*.

UPME (2023). *Proyección de la Demanda de Energía Eléctrica y Potencia Máxima 2023-2037*.

Ushiku, Y. (2020). *Long Short-Term Memory*, pages 1–7. Springer International Publishing, Cham.

Valencia, A., Moreno, C. A., and Lozano, C. A. (2007). Modelo de promedios móviles para el pronóstico horario de potencia y energía eléctrica. *El Hombre y la Máquina*.

Wasserman, L. (2010). *All of statistics : a concise course in statistical inference*. Springer, New York.

Wen, L., Zhou, K., and Yang, S. (2020). Load demand forecasting of residential buildings using a deep learning model. *Electric Power Systems Research*, 179:106073.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second Edition*.

XM (2023). Mercados. Retrieved Dec 10, 2023 from www.xm.com.co/consumo/mercados.

Young-Min, W., Sung-Kwan, J., and Kyung-Bin, S. (2012). Holiday load forecasting using fuzzy polynomial regression with weather feature selection and adjustment. *IEEE Transactions on Power Systems*, 27(2):596–603.

Zhou, C., Fang, Z., Xu, X., Zhang, X., Ding, Y., Jiang, X., and ji, Y. (2020). Using long short-term memory networks to predict energy consumption of air-conditioning systems. *Sustainable Cities and Society*, 55:102000.

Zhou, W., Tao, H., Ding, S., and Li, Y. (2023). Electricity consumption and production forecasting considering seasonal patterns: An investigation based on a novel seasonal discrete grey model. *Journal of the Operational Research Society*, 74(5):1346–1361.

Zhu, G., Peng, S., Lao, Y., Su, Q., and Sun, Q. (2021a). Short-term electricity consumption forecasting based on the emd-fbprophet-lstm method. *Mathematical Problems in Engineering*, 2021:1–9.

Zhu, K., Geng, J., and Wang, K. (2021b). A hybrid prediction model based on pattern sequence-based matching method and extreme gradient boosting for holiday load forecasting. *Electric Power Systems Research*, 190:106841.

Zolfaghari, M. and Sahabi, B. (2019). A hybrid approach to model and forecast the electricity consumption by neurowavelet and arimax-garch models. *Energy Efficiency*, 12.