Applied Energy 348 (2023) 121439

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Short-term load forecasting using neural networks and global climate models: An application to a large-scale electrical power system

Lucas Barros Scianni Morais [a], Giancarlo Aquila [b], Victor Augusto Durães de Faria [c], Luana Medeiros Marangon Lima [d], José Wanderley Marangon Lima [e], Anderson Rodrigo de Queiroz [f,g,*]

[a] Institute of Electrical and Energy Systems, Federal University of Itajubá, Itajubá, MG, Brazil
[b] Institute of Production Engineering and Management, Federal University of Itajubá, Itajubá, MG, Brazil
[c] Operations Research Graduate Program at NC State University, Raleigh, NC, USA
[d] Nicholas School of Environment, Duke University, Durham, NC, USA
[e] Marangon Consulting and Engineering, Itajubá MG, Brazil
[f] Decision Sciences, Economics & Finance Department at NC Central University, Durham, NC, USA
[g] Civil, Construction and Environmental Eng Department at NC State University, Raleigh, NC, USA

HIGHLIGHTS

- We develop shallow and deep neural networks for the Short-Term Load Forecasting problem.
- Different neural networks architectures are tested, including uni- and bi-directional structures.
- Global climate models' information is used as input of the neural networks.
- We present a real study case of time series forecasting for the Brazilian power system.
- Relevant results are presented and systematically compared using Diebold-Mariano test.

ARTICLE INFO

Keywords:
Short-Term Load Forecasting
Artificial Neural Networks
Recurrent Neural Networks
Global Climate Models
Electricity Markets

ABSTRACT

This paper focuses on the development of shallow and deep neural networks in the form of multi-layer perceptron, long-short term memory, and gated recurrent unit to model the short-term load forecasting problem. Different model architectures are tested, and global climate model information is used as input to generate more accurate forecasts. A real study case is presented for the Brazilian interconnected power system and the results generated are compared with the forecasts from the Brazilian Independent System Operator model. In general terms, results show that the bidirectional versions of long-short term memory and gated recurrent unit produce better and more reliable predictions than the other models. From the obtained results, the recurrent neural networks reach Nash-Sutcliffe values up to 0.98, and mean absolute percentile error values of 1.18%, superior than the results obtained by the Independent System Operator models (0.94 and 2.01% respectively). The better performance of the neural network models is confirmed under the Diebold-Mariano pairwise comparison test.

1. Introduction

Short-Term Load Forecasting (STLF) plays an important role in supporting Independent System Operators (ISO) in many aspects of energy planning and operations, such as power generation reserve, system reliability, dispatch scheduling, demand management, and electricity pricing [1]. In the past decade, with the advance of smart grid technologies and the increasing penetration of wind and solar farms, the complexity associated with short-term operational planning has escalated in electricity power systems, posing considerable challenges for ISOs to operate power grids reliably [2]. The expected increase in the deployment of electric vehicles [3], distributed renewable generation at

* Corresponding author.
E-mail addresses: scianniufu@gmail.com (L.B.S. Morais), giancarlo.aquila@yahoo.com.br (G. Aquila), vaduraes@ncsu.edu (V.A.D. de Faria), luana.marangon.lima@duke.edu (L.M.M. Lima), jose.marangon@marangonenergia.com.br (J.W.M. Lima), adequeiroz@nccu.edu, ardequei@ncsu.edu (A.R. de Queiroz).

https://doi.org/10.1016/j.apenergy.2023.121439
Received 24 January 2023; Received in revised form 6 May 2023; Accepted 10 June 2023
Available online 7 July 2023

the regional level [4], and energy storage also exacerbate the uncertainties in supply and energy controllability in short-term operations [5]. In this context, accurate demand forecasts are crucial for systems operators and analysts to support planning and operational decisions in electricity power systems [6].

STLF models aim to improve electrical system planning from small-scale regional networks to large-scale interconnected grids. Accurate STLFs are essential to balance supply and demand for the hour- or the day-ahead operational planning [7]. Several methodologies have been applied over the years to model STLF problems; among them, non-linear models are frequently superior in representing the behavior of this type of time series [8]. In this context, non-linear machine learning methods have received considerable attention in the literature with applications of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) [9]. These models rely on large datasets with multiple variables (e.g., load, calendar variables, holidays, temperature, humidity, cloud cover, and others), aiming to improve the characterization of the problem features and better model hidden relationships between inputs and outputs to produce accurate and reliable forecasts.

Recent methodological developments in machine learning and significant improvements in computational hardware (e.g. usage of graphical processing units for training ANNs) and software (e.g. availability of open-source packages such as Keras and Google TensorFlow) helped to push the boundaries associated with time series forecasting using ANNs, SVMs, and other related methods. As a result of this significant development, machine learning models have been successfully applied in STLF studies. For example, Fan et al. [10] used SVMs to perform hourly and sub-hourly load forecasts for day ahead in New South Wales, and Barman et al. [11] used SVMs to perform STLFs for Assam (India). SVMs have been applied to regression and classification in many engineering problems for many years [12]. Other studies used recurrent ANNs to perform STLF in electricity markets like Belgium [13]; the work of Panapakidis [14] developed forecasting models for day-ahead and hour-ahead load predictions at the bus-level in Greece using ANNs; the authors in Jiao et al [15] developed ANN models in connection with k-means clustering to investigate non-residential load forecasting problem in China. Convolutional Neural Networks (CNNs), generally designed to exploit spatial correlation in data when dealing with images and speech recognition, have also been applied to STLF problems. For example, Voß et al. [16] used CNNs to generate load forecasts for individual households in Austin-TX and Boulder-CO, Tian and Hao [17] proposed combining unidirectional LSTMs and CNNs for STLF in Italy, and Sadaei et al. [18] proposed the use of fuzzy time series and CNNs connecting load information and temperature to generate forecasts for a company in Malaysia.

The existing STLF literature shows several applications of Recurrent Neural Networks (RNNs) with satisfactory results while using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, e.g. the work of Jin et al. [19] uses GRU models to improve predictor accuracy associated with temperature, wind speed and humidity data used in agriculture; the authors in Bouktif et al. [20] uses LSTMs in connection with genetic algorithms for feature selection and investigate electricity consumption in a metropolitan area in France; the work of Kong et al. [21] investigate the use of LSTMs for load forecasting using aggregated information from smart meters data from customers in New South Wales.

RNNs continue to be explored in recent STLF studies in different systems, such as the study by Veeramsetty et al. [22] that uses traditional RNNs for day ahead and hourly ahead with the support of the Principle Component Analysis (PCA) in the data processing stage. Haque and Rahman [23] present a heuristic analysis on the selection of relevant inputs and the train datset, adjustment of hyperparameters and fine-tuning for an ideal LSTM. In turn, Lin et al. [24] presents a two-stage LSTM for short-term probabilistic load forecasting, based on an encoder built to calculate the inputs correlation with load. Yang et al. [25] decompose the initial training dataset into a finite number of bivariate modal components, then reconstruct it using multivariate permutation entropy to identify internal factors, seeking better accuracy in LSTM results. Yue et al. [26] used LSTM combined with empirically decomposed techniques, in addition to permutation entropy and feature selection, and a Bayesian optimization algorithm in order to increase the accuracy of their models. Subbiah and Chinnappan [27] used LSTM with selection of inputs based on filtering features and data clustering to identify subsets of ideal data, aiming to reduce overfitting and improve the accuracy.

Thus, it is important to note that most of the previous studies that apply RNNs to solve STLF problems focus on: one (or a few) ANN architecture(s), the input variable selection, and hybrid models that combine ANNs with other techiniques such as clustering. Moreover, most of the previous literature apply ANN models only considering small scale, and/or low to medium complexity systems, seeking to validate the proposed methodology performance to those systems. In turn, the present study is focused on the analysis of several ANN architectures to the STLF problem in a large-scale electrical power systems; such system is composed by different geographic zones, climate conditions, population centers, which implies in complexities to the load profiles.

Despite recent advances using LSTMs and GRUs, the existent STLF literature and practical implementations still lack proper model/input definition and pre-processing procedures [28]. For example, bidirectional RNNs are not as popular as their unidirectional counterpart, despite the benefits of bidirectional information in the RNN learning process [29]. Only a few papers focused on developing bidirectional RNNs to STLFs. For example, the work of Cai et al. [30] shows a multi-layer stacked bidirectional LSTM to generate forecasts for a 35 kV substation in China using three years of data as training/test set and produced more accurate forecasts than a set of other three ANN architectures investigated. In Wahab et al. [31], the authors present the use of a bidirectional sequential RNN model with feature engineering applied to different STLF datasets of historical load and calendar data. Ullah et al. [32] show a hybrid model based on bidirectional LSTMs in combination with CNNs to forecast household electric power consumption, where authors point out that future studies should attempt to incorporate climate conditions in the ANN model inputs. Although the use of RNNs in STLF with observed temperature and calendar variables has been used in previous studies, to our knowledge there is lack of studies evaluating different RNNs architecture performance for the case of real large-scale systems considering information originated from Global Climate Models (GCMs).

Most of the previous STLF literature focuses on assessing performance accuracy of models based on classical metrics as the Mean Absolute Percent Error (MAPE), the Mean Absolute Error (MAE), the Mean Square Error (MSE), and the Root Mean Square Error (RMSE). For example, Li et al. [33] uses LSTM in connection with time series decomposition to provide STLF for a city in China and compares accuracy results using MAPE, MAE, and RSME. Similar comparison by these metrics is also presented in Bashir et al. [34] where a hybrid LSTM-Prophet is used to model load at Elia Grid in Belgium. Other metrics are also added to accuracy performance investigation in Javed et al. [35] that uses CNNs in connection with LSTMs for STLF in the city of Lahore in Pakistan (a related approach is presented in Rafi et al. [36] with application to Bangladesh). Other previous work [37–38] use MAPE, MSE, and other related metrics. While previous STLF literature has focused on classical metrics to assess model performance accuracy, it is still modest in numbers of applications the use of more robust pairwise comparisons for STLF models. Such pairwise-comparisons aim to provide a statistical hypothesis test to determine whether one forecasting model is significantly better than another and help with decisions regarding model selection.

This paper presents a novel approach to evaluating potential accuracy gains in the STLF problem by combining the use of a range of ANNs with temperature forecasts from GCMs. Our research breaks new ground by investigating the performance of unidirectional and bidirectional versions of RNNs based on LSTMs and GRUs, as well as the performance of Multi-layer Perceptrons (MLP) ANNs with shallow and deep structures. We also test our methodology using real data from the Brazilian Central-Southeast electricity market, which accounts for about 50% of the total energy consumption in Brazil [39]. To our knowledge, this is the first time bidirectional RNNs have been used in combination with climate variables to produce forecasts for a large-scale power system. Large-scale systems are subjected to unique STLF challenges due to diverse climates, consumer behaviors, limited information, increased coordination needs, and susceptibility to external factors. Additionally, this work fills the gap by performing systematic pairwise comparisons between forecasts generated by different models using the Diebold-Mariano significance test [40]. The overall contributions of the work can be synthetized as:

(i). Establishment of a novel time series forecasting framework for STLF based on ANNs that uses historical load, calendar, and population data in combination with temperature information from a GCM;
(ii). Investigation of bidirectional RNNs performance in comparison to other unidirectional RNNs and MLPs architectures, as well as existing model used by the Brazilian ISO;
(iii). Detailed analysis of STLF in a large-scale electrical power system, providing a realistic evaluation of different model performance in practical applications;
(iv). Application of the Diebold Mariano test to suggest a more reliable and accurate selection of different ANNs architectures for STLF applications.

The other sections of this paper are organized as follows: Section 2 presents the methodology explaining the application of ANNs to STLF and the analysis performed in this work. Section 3 presents the data and details the case study. Section 4 contains results and discussion. Section 5 presents the main conclusions of the work and points to future research directions.

2. Methodology

2.1. Time series Forecasting: Analysis framework

This study investigates the STLF problem using the framework described in Fig. 1. Different machine learning models and methods are employed to generate STLFs, and while our focus is on the use of ANNs (in the form of MLPs, LSTMs, and GRUs) SVMs are also considered.

Initially, the data inputs go through a pre-processing step that includes the treatment of missing data, one-hot encoding, and normalization. Subsequently, the machine learning models that produce the STLFs are applied, and the accuracy of each model is assessed through the use of the Mean Average Percentage Error (MAPE) and the Nash-Sutcliffe Error (NSE). Finally, the developed models are compared using the Diebold-Mariano pairwise test, which assesses whether there is a statistical difference between the accuracy of the different models investigated.

2.2. Data Pre-processing

In possession of the STLF dataset, a set of pre-processing procedures must be followed. In this work, missing data and outliers were identified and substituted by linear interpolations using the nearest neighbor values. After that, the resultant dataset was normalized by subtracting the hourly value of each variable (load, temperature, etc.) by its minimum value and dividing the result by the standard deviation of the corresponding variable.

As machine learning models cannot interpret categorical variables without a prior treatment, a process of encoding these variables was performed using the One Hot Encoding technique [41]. The One Hot Encoding is based on the transformation of a categorical variable into a binary form (dummy). For example, the calendar variables used as STLF input are: day of the week, day of the month, and month. Each variable is represented by a vector containing zeros and ones. The vector size is the number of possibilities that the variable can reach. For example, the day of the week variable has seven possibilities, so a vector with seven positions will represent it. If the day in question is a Monday, the first position of the vector assumes the value one, and the other positions will take the value zero.

As it is well known, ANNs have a large memorization capacity; however, it is not desirable that the model memorizes the training set because it affects its generalization capabilities [42]. Ideally, the model should be able to generalize the learning process and create reasonable extrapolations for events that have not been covered in the training set. Therefore, regularization techniques have been created to improve the ANN performance and reduce data overfitting. One of the main techniques that is used in this study is the Dropout, where part of the neurons in the ANN are randomly turned off during the training phase. More

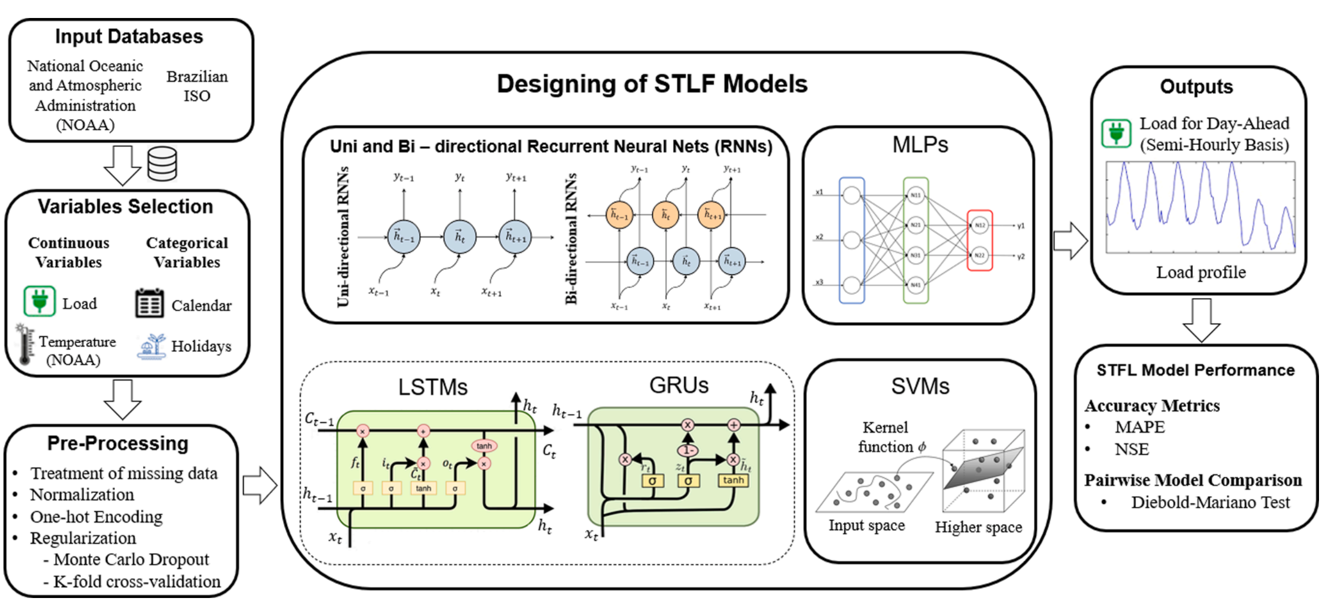


Fig. 1. Time Series Forecasting Analysis Framework for Short-term Electric Load.

specifically, this work uses the Monte Carlo Dropout technique proposed by Gal and Ghahramani [43], which provides a probabilistic approximation of model forecast, where part of the ANN neurons is randomly turned off not only during training but also during forecasting. In this framework, different ANN weight configurations can be evaluated in a single ANN model providing mathematical bases for uncertainty quantification and improving the model performance.

Another technique used here to improve model performance is the k-fold cross-validation which consists in dividing the training data into k mutually exclusive subsets of the same size [44]. From this, a subset is used for testing, and the remaining $k-1$ subsets are used for parameter estimation (model training). This process is performed k times, alternating the training and testing subsets. In the end, k different models are trained and tested with different batches of data [45]; similar procedure can also be observed in the work of [46] about global solar radiation in India.

2.3. Multi-Layer perceptron Neural Networks

Among the different ANN architectures, the MLP is one of the most popular and it is often used in several types of time series forecasting problems, including STLF. In MLPs, initially, the ANN signal moves forward from the input layer to the hidden layers until it reaches the output layer, where the cost function can be estimated. Then, posteriorly, a backward propagation step is performed where the partial derivatives of the cost function with respect to the weights and bias are used to update the model parameters [47].

Recently, more powerful hardware has enabled the implementation of Deep Neural Networks (DNN). DNNs have a large number of hidden layers when compared to shallow ANNs. MLPs can be trained as deep ANNs and may benefit from the enhanced representation. Given an appropriate amount of data and computational time for training, DNN contributes to a better abstraction of the system parameters and, therefore, better representation of the non-linear relationships between inputs and outputs in problems such as the STLF [48]. For a discussion about the characteristics of DNNs in comparison with shallow ANNs for streamflow forecasting, see [49].

2.4. Recurrent Neural Networks (RNN)

Another ANN model that is constantly employed for STLF is the RNN. These models are based on architectures that segment the ANN prediction by time steps, allowing optimal applications in time series problems [50].

LSTM is a special RNN model capable of learning long-term dependencies [51]. In an LSTM, an internal memory cell gate (C_t) is defined to store long-term information. This gate interacts with previous output and subsequent input to select which elements of the internal vector will be updated, kept, or deleted. In this architecture, the memory cell state is defined as in (1–4), where (1) and (2) describe the cell gate vector update, (3) represents the input gate (i_t), and (4) describes the computations for the forget gate (f_t). The input gate is combined with the cell update vector as a mechanism to integrate new information into the cell state; similarly, the forget gate is combined with the previous cell state in order to delete information from the RNN memory. After computing C_t the final output of the RNN cell (h_t) can be determined by (5–6), where (5) is called the output gate. In equations (1–6), w_c, w_i, w_f, w_0, and are the weight matrices, b_c, b_i, b_f, and b_0 are bias vectors, σ is the logistic sigmoidal function, x_t is the input vector, h_t is the output vector of the current cell, "*" represents the element-wise multiplication of matrices (Hadmard Product), and "•" represents a normal matrix multiplication.

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{1}$$

$$\widetilde{C}_t = \tanh(w_c \bullet [h_{t-1}, x_t] + b_c) \tag{2}$$

$$i_t = \sigma(w_i \bullet [h_{t-1}, x_t] + b_i) \tag{3}$$

$$f_t = \sigma(w_f \bullet [h_{t-1}, x_t] + b_f) \tag{4}$$

The output gate (o_t) and the final output (h_t) are described by (5) and (6), respectively. Here, w_0 is the weight matrix associated with o_t and b_0 is the bias vector.

$$o_t = \sigma(w_0 \bullet [h_{t-1}, x_t] + b_0) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

Another important RNN model is the GRU. The GRU architecture does not include a cell state and uses its hidden state (h_t) to transfer information from previous steps of the time series [52]. The GRU architecture has two gates the update gate (z_t) and reset gate (r_t), and can be mathematically described by (7–10). Where w_h, and w_r are weight matrices, and b_h, b_z and b_r are bias vectors.

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \tag{7}$$

$$\widetilde{h}_t = \tanh(w_h \bullet [x_t, r_t * h_{t-1}] + b_h) \tag{8}$$

$$z_t = \sigma(w_z \bullet [x_t, h_{t-1}] + b_z) \tag{9}$$

$$r_t = \sigma(w_r \bullet [x_t, h_{t-1}] + b_r) \tag{10}$$

2.5. Bidirectional RNN models

In addition to RNN structures that have a unidirectional flow of information (forward in time), there are also models characterized by a bidirectional data flow, known bi-RNNs (Fig. 2). Unlike RNNs with unidirectional flow (Fig. 2a.), architectures with bidirectional flow process data in two directions (Fig. 2b.). For time series problems, bi-RNNs process data forwards and backward in time through different layers [53]. The combination of the bidirectional flow structure with LSTMs and GRUs generates what is known as bi-LSTMs and bi-GRUs.

Consider a time sequence $\overrightarrow{t} = \{1, \cdots, T\}$, for the forward layer of the bi-RNN, and and $\overleftarrow{t} = \{T, \cdots, 1\}$ for the backward layer. Following the notation of Fig. 2, a forward hidden sequence ($\overrightarrow{h}$) would, in general, be computed as (11), a backward hidden sequence ($\overleftarrow{h}$) would be computed as (12), and the output (y_t) would be computed as (13) [54]. In these equations $w_{x\overrightarrow{h}}$, $w_{y\overrightarrow{h}}$, $b_{\overrightarrow{h}}$ are the weight and biases of forward layer, $w_{x\overleftarrow{h}}$, $w_{y\overleftarrow{h}}$, $b_{\overleftarrow{h}}$ are the weight and biases of backward layer, and b_y is a bias parameter for the output.

$$\overrightarrow{h}_t = \tanh\left(w_{x\overrightarrow{h}} \bullet \left[x_t, \overrightarrow{h}_{t-1}\right] + b_{\overrightarrow{h}}\right) \tag{11}$$

$$\overleftarrow{h}_t = \tanh\left(w_{x\overleftarrow{h}} \bullet \left[x_t, \overleftarrow{h}_{t+1}\right] + b_{\overleftarrow{h}}\right) \tag{12}$$

$$y_t = w_{y\overrightarrow{h}} \overrightarrow{h}_t + w_{y\overleftarrow{h}} \overleftarrow{h}_t + b_y \tag{13}$$

2.6. Support vector Machines for regression

In this work we use SVMs with the purpose of time series forecasting, therefore we consider only SVMs based on Support Vector Regression (SVR). Given a dataset $(\boldsymbol{X}, \boldsymbol{Y})$ with $\boldsymbol{X} \in \mathbb{R}^{N \times M}$, and $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ where N is the number of samples, M is the number of input elements, and T is the number of output elements, the parameters $\boldsymbol{W}$ and $\boldsymbol{b}$ of a SVR model can be determined by solving (14–17), where $\phi(\boldsymbol{X}_i)$ maps $\boldsymbol{X}_i$ to a higher-

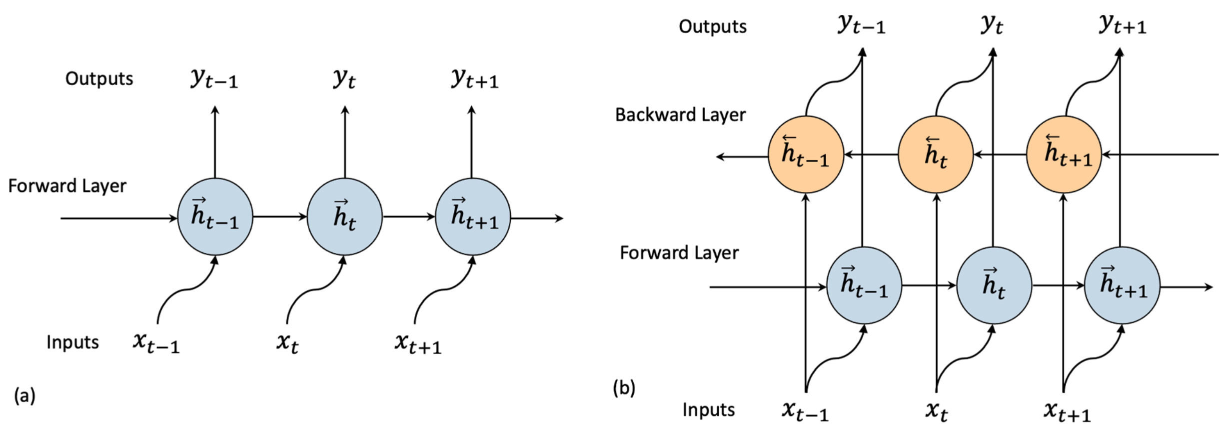


Fig 2. (a)Uni-RNN; (b)Bi-RNN architecture.

dimensional space by the use of a kernel function [55]. In this framework, an estimate $\boldsymbol{Y}$ for an arbitrary input vector $\boldsymbol{X}$ has the form of Equation (18).

$$\min_{\boldsymbol{W},b,\xi,\xi^*} \frac{1}{2}||\boldsymbol{W}||^2 + C\sum_{k=1}^{N}(\xi_k + \xi_k^*) \tag{14}$$

$$\text{s.t.} \quad y_k - \boldsymbol{W}^T\phi(\boldsymbol{X}_k) - \boldsymbol{b} \leq \varepsilon + \xi_k, \quad \forall k = 1, \cdots, N \tag{15}$$

$$-y_k + \boldsymbol{W}^T\phi(\boldsymbol{X}_k) + \boldsymbol{b} \leq \varepsilon + \xi_k^*, \quad \forall k = 1, \cdots, N \tag{16}$$

$$\xi_k, \xi_k^* \geq 0, \quad \forall k = 1, \cdots, N \tag{17}$$

$$y_i = \boldsymbol{W}^T\phi(\boldsymbol{X}_i) + \boldsymbol{b} \tag{18}$$

In (14–17), the parameter ε is a margin of tolerance in the approximation (18), ξ_k and ξ_k^* are slack variables of the model constraints, and $C \geq 0$ is a regularization hyperparameter that controls de degree of the deviations above ε.

By using kernel functions to implicit transform the input data into a higher dimensional space, we can solve the dual version of the model (14–17), find the dual variables ($\boldsymbol{\alpha}$) associated with the primal model constraints, and easily determine the forecast $\boldsymbol{Y}$ of any input variable $\boldsymbol{X}$ using the well-known kernel trick [55]. In this work, the radial basis kernel function [55] was used in the implementation of the model (14–18) as this is the kernel used by the Brazilian ISO in its load forecasting model [56].

2.7. Systematic comparison of forecasts

Traditional accuracy measures such as MAPE, and NSE can lead to wrong conclusions while comparing the performance of different models if no statistical significance test is performed in the analysis. To face this problem, Diebold and Mariano [40] presented an accuracy test capable of statistically validating the performance of a given model in relation to a benchmark. The Diebold-Mariano (DM) test is presented below.

Here, we define y_i, $i \in \{1, \cdots, N\}$ as the ground truth observations of the forecasts made by models M_1 ($\hat{y}_i^{(M_1)}$), and M_2 ($\hat{y}_i^{(M_2)}$), and $e_i^{(M_1)}$, $e_i^{(M_2)}$ the forecast errors of each model (19). By applying a loss function to the errors of each model as in (20), the statistical value of the Diebold-Mariano test can be computed as in (21), where S^2 is a consistent estimator of the asymptotic variance of (22) [57].

$$e_i^{(M_\alpha)} = y_i - \hat{y}_i^{(M_\alpha)}, \forall i \in \{1, \cdots, N\}, and M_\alpha \in \{M_1, M_2\} \tag{19}$$

$$F\left(e_i^{(M_\alpha)}\right) = \left(e_i^{(M_\alpha)}\right)^2 \tag{20}$$

$$DM = \frac{\sum_{i=1}^{N}\left(F\left(e_i^{(M_1)}\right) - F\left(e_i^{(M_2)}\right)\right)\Big/N}{\sqrt{S^2/N}} \tag{21}$$

$$\mathbb{E}(d) = F\left(e_i^{(M_1)}\right) - F\left(e_i^{(M_2)}\right) \tag{22}$$

Finally, the Diebold-Mariano hypothesis test can be described as:

- H_0: loss function generates predictions that are not statistically different ($\mathbb{E}(d) = 0$);
- $H_\alpha = \mathbb{E}(d) > 0$, where model M_1 has better prediction performance than M_2;
- $H_\alpha = \mathbb{E}(d) < 0$, where model M_2 has better prediction performance than M_1.

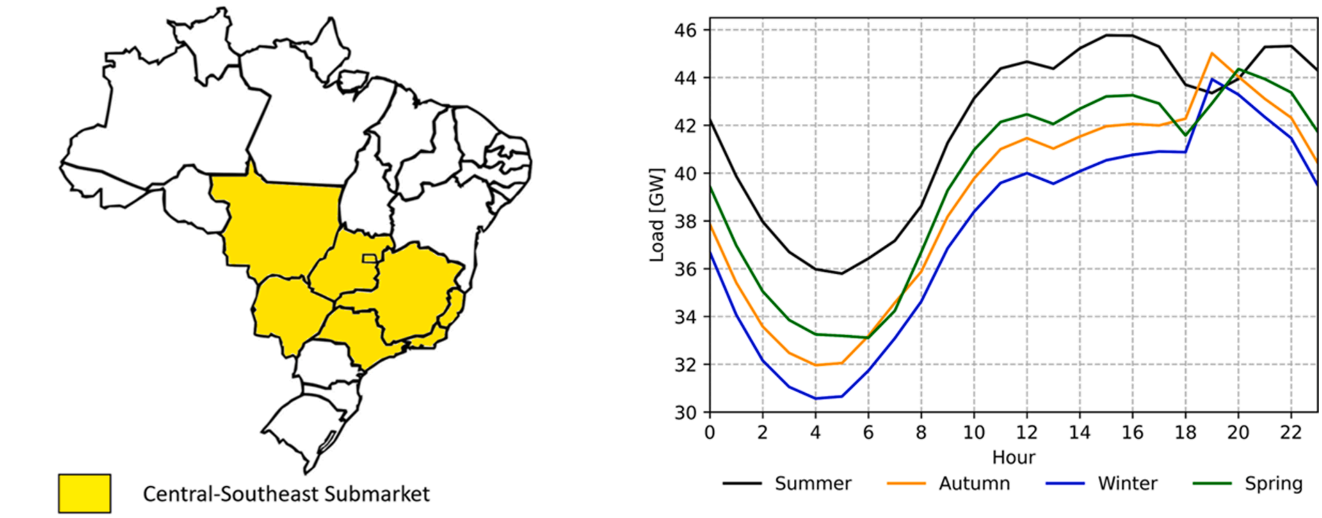
3. Short-Term load forecasting case study

3.1. Case study description

The framework presented in Section 2 is applied and tested using a real case study representing a portion of Brazilian electricity market. The Brazilian interconnected power system has a load of approximately 73GW divided in four subregions (South, Central-Southeast, North and Northeast) [39]. Our focus in this work is directed to the Central-Southeast portion of the system which corresponds to approximately 60% of the system total load and is the region where the largest load centers in the country (São Paulo, Rio de Janeiro and Belo Horizonte) are located (Fig. 3a). Fig. 3b shows the load profile of the Central-Southeast for typical days representing the four seasons of the year in 2019.

STLF has been receiving increasing attention from the Brazilian ISO, as the daily operational planning of the Brazilian interconnected power system expanded to hourly and sub-hourly time discretization, relying on a unit commitment model named DESSEM [28]. DESSEM uses as input demand for the day-ahead on a sub-hourly basis. The day-ahead forecasts are made for 48 half-hour intervals, which are provided by STLF models, and determine the scheduling of the power generation plants and the hourly locational marginal electricity prices (LMPs). The LMPs are the basis to represent the electricity prices in the Brazilian short-term market, and therefore, short-term demand forecasts play a significant role in the pricing formation.

In the daily operation schedule, the most recent STLF model developed by the Brazilian ISO is named PrevCargaDESSEM, which considers as input variables: load historical time series data; calendar variables (date, time, month, year), including holidays and days; and minimum, average, and maximum observed temperatures for the Brazilian Central-Southeast submarket [58].



a. Brazilian Central-Southeast Market

b. Typical day Load Profile in 2019

Fig 3. Typical day Load Profile 2019 - Central-Southeast Submarket.

The PrevCargaDESSEM works with an ensemble of 14 models, with different inputs for temperature data, including: four SVMs with linear kernel (one with maximum temperatures; one with average and maximum temperatures; one with average temperatures; and one without temperature data); four radial-kernel SVMs (one with maximum temperatures; one with average and maximum temperatures; one with average temperatures; and one with no temperature data); 4 MLPs 3-layers (one with maximum temperatures; one with average and maximum temperatures; one with average temperatures; and one without temperature data); and two dynamic regressions based on the ARIMA model (p,q,d) (one with maximum temperature; and one with mean temperature). The weights of each model are obtained by an optimization algorithm that aims at the weighting that results in the smallest forecast error.

3.2. Load and calendar data

Load data is based on historical information from the Brazilian ISO available in hourly discretization in the SINtegre platform [59] for all four Brazilian submarkets (Central-Southeast, South, Northeast, and North). Furthermore, the load profile is segmented into light, medium, and heavy, with the light level being concentrated in the early hours of the morning and the heavy level mainly between 6:00 pm and 11:00 pm. Initially, forecasts are made on an hourly basis, which are later discretized on a semi-hourly basis by a cubic monotonic spline approximation made by ISO.

Besides that, calendar variables, holiday information, and daylight savings time (DST) are available daily in SINtegre [59]. In Brazil, DST is defined between the months of October to February; Among the main holidays are national holidays, such as Independence Day, Carnival, Christmas, New Year's Eve, and days with special events such as World Cup soccer matches. For the models develop in this work, when the STLF is performed for a holiday, the calendar variables are reported as Sunday since there is a small sample of holiday data available on the input dataset.

3.3. Observed temperature and ISO temperature forecasts

The verified temperature for each submarket is gathered from civil and military aviation databases [60], being available on an hourly basis. Also, the temperature forecasts used by the ISO are gathered from the Center for Weather Prediction and Climate Studies - CPTEC/INPE [61] and the National Center for Environmental Prediction - NCEP [62], being available for seven days ahead at hourly time discretization [59]. In order to determine an aggregate temperate value for a specific submarket, the ISO assigns weights for the temperatures of the main cities inside each submarket based on historical data and take the mean of the resultant sum. As this is the official procedure adopted by the Brazilian regulatory agencies, it was also followed in this work.

3.4. Temperature forecasts from the Global climate model

Global climate models focus on short-term atmospheric changes, in addition to spatial and temporal accuracy through high-precision algorithms [63]. These models also accurately describe the evolution of the present climate and adapt to external changes forced within the climate system. Therefore, the greater accuracy that GMCs may have, the better will be the meteorological data for inputs for STLF models. In this study, temperature data from the Global Ensemble Forecast System (GEFS) [64], a GCM provided by the National Oceanic and Atmospheric Administration (NOAA), are tested in the STLF models evaluated. This data is acquired for the twenty largest cities of the Brazilian Central-Southeast submarket, and subsequently weighted by population in order to obtain the equivalent temperatures for the system.

We also analyze the use of temperature forecasts from the GEFS database and evaluate the performance of the developed STLF models against using temperature forecasts from the Brazilian ISO database. GEFS is a weather/climate model developed by NCEP that generates 21 separate forecasts to overcome uncertainties in the inputs and model limitations. GEFS quantifies these uncertainties and produces a range of potential outcomes based on data perturbations [64]. This model's data are available from 2000 to 2019 in a three-hour discretization.

We weight the GEFS temperature by the populations of the twenty most populous cities in each state of Central-Southeast region based on a geometric mean as previously done in Cawthorne et al [65] for two states in the southern US and in a large scale analysis of the 48 states representing the contiguous US territorial area in Esraghi et al. [66]. In addition, to ensure a bias correction of the GEFS data, a regression analysis is performed between estimated population-weighted GEFS temperatures with the temperature available from the Brazilian ISO dataset. The regression is described by (23) and shows an R^2 value equal to 0.874.

$$t^{*} = 3.5416 + 0.9717 t_{GEFS} \tag{23}$$

In (23), t^* is the estimated temperature based on the regression equation that uses the GEFS temperature (t_{GEFS}) as the predictor.

3.5. Time series forecasting considerations

All forecasts are performed for the day-ahead at half-hour time discretization. Our models use as input data the verified load and temperature values in the past two days, the predicted temperatures, a calendar indicator for the day of the forecast, and additional calendar information representing special days (e.g., holidays). After testing several combinations of parameters and input variable selection, the best results were obtained using five timesteps for RNNs. Each timestep is composed of the set of features mentioned earlier (calendar variables, historical data and temperature predictions), such that in its totality, an input sample of the RNNs contains the information from seven sequential days prior to the forecast day.

The forecasting model used by the Brazilian ISO is based on a combination of a radial SVM, a linear SVM, and a single-layer MLP, combined in an ensemble of twelve different model configurations [56]. This ensemble considers different configurations of temperature inputs, such as: neglecting predicted temperature, using mean temperatures, using maximum temperatures, and using maximum and minimum temperatures. Therefore, when comparing different model performances in Section 4, we use the ISO STLF forecast results of the ensemble model available [59].

4. Results and discussions

4.1. Experimental design for time series forecasting Analysis

As mentioned earlier, we focus on evaluating STLF models performance for the Brazilian Central-Southeast electricity submarket. The STLFs are developed using the electricity market data and consider different MLP architectures, as well as unidirectional and bidirectional LSTMs and GRUs. Forecasts are obtained on a semi-hourly basis, and in addition to performance evaluations through MAPE and NSE the DM test is employed for systematic pairwise comparisons. Three different comparison analyses among models are carried out using the Brazilian ISO and GEFS temperature data.

For the experiments of Analysis 1 and 2, data is available for the period of 01/01/2016 to 09/20/2021. From April to September 2020, the load shape changed considerably due to COVID-19 initial quarantine effects, and therefore, this time segment was discarded from the input data. Also, the last 90 days of the dataset were excluded from the training batch to be used as an out-of-sample test for the models. As the GEFS temperature data is stored in the NOAA servers from 2000 to the end of 2019, for the Analysis 3, the period between 01/01/2016 and 12/31/2019 was selected to match other ISO available data and the entire year of 2019 was left to test the model.

- **Analysis 1: Shallow vs. Deep MLPs.** Initially, the performances of MLPs with different layers are compared to assess whether there is an accuracy gain with deeper ANNs. Brazilian ISO historical load data, verified temperature, predicted temperature (up to 7 days ahead), holidays and calendar variables are considered as input.
- **Analysis 2: Models use datasets with temperature information from the Brazilian ISO database for a three-month period.** Brazilian ISO historical load data, verified temperature, predicted temperature (up to 7 days ahead), holidays and calendar variables are considered as input. Results from MLP, Uni-LSTM, Uni-GRU, Bi-LSTM, Bi-GRU, and the Brazilian ISO model are analyzed.
- **Analysis 3: Models use predicted temperatures from the GEFS database and the Brazilian ISO database for one-year period.** Brazilian ISO historical load data, existent and predicted GEFS temperature (up to 10 days ahead), holidays and calendar variables are considered as input. Unlike the temperature data from Brazilian ISO database, where the predicted temperature values are available hourly, the GEFS temperature is available every 3 h, a total of 8 predicted temperature values per day. The performance of the models with the GEFS data is evaluated considering a test set with one year of the existent data. Results from MLP, Uni-LSTM, Uni-GRU, Bi-LSTM, Bi-GRU are analyzed, and the Brazilian ISO model is not compared as no ISO information is available for the test period.

4.2. Analysis 1: Deep Neural Networks vs shallow Neural Networks

MLPs with one to five hidden layers are designed and compared, considering 10 k to 60 k training epochs and ten folds (k-fold cross validation). Table 1 summarizes the MAPE results for each configuration of hidden layers, and Table 2 shows the results of DM test for MLPs considering 10 k epochs. In Table 2 the cells marked in grey indicate that the models with the best performance in the pairwise comparison are in the columns, cells in blue indicate the MAPE and NSE values of each model individually. We note that MLPs with only one hidden layer had the lowest MAPE, the highest Nash-Sutcliffe Error (NSE) and, the best DM test results were obtained with MLPs with 2, 3 and 5 hidden layers but close to the DM test limits (p-value < 0.05 and DM > 1.96). Therefore, given the proximity of results only MLPs with one hidden layer are further considered in the following comparisons with the Brazilian ISO model and RNNs models.

To avoid model overfitting, a dropout of 10% is used in the MLP training process. The dropout procedure is also extended to the prediction of the test set (Monte Carlo Dropout), which allows an assessment of the model uncertainly. Furthermore, a k-fold cross-validation with ten splits is also considered to improve the model performance.

4.3. Analysis 2: Pairwise comparison based on Brazilian ISO temperature data

Next, we compare results produced by the MLP, unidirectional and bidirectional versions of LSTMs and GRUs with the Brazilian ISO STLF model. Historical data are separated into two sets: training and testing sets, according to the description in Section 4.1. Table 3 summarizes the comparisons between the results of the models, the cells marked in grey indicate that the models with the best performance in the pairwise comparison are in the columns.

Initially, the results of the Diebold-Mariano test with a significance degree of 0.05 indicate that the recurrent models produced statistically significant better results (p-value < 0.05 and DM > 1.96). More specifically, the results indicate that the best model is the Bi-LSTM, followed by the Bi-GRU.

Although most models presented MAPEs between 1.2% and 2.11% and NSE values close to 1.0, it is observed that the accuracy of the models naturally alternate over time. From Fig. 4, it is possible to notice that the ISO model and the MLP had their worst performance on holidays. This is similar to what has been observed by different authors in other STLF studies [67–68]. In this aspect, RNNs performed better in some of the holidays, such as 9th of July (a holiday in Sao Paulo state), and on 7th of September (Brazilian Independence Day).

Table 1
MAPE Results for Deep and Shallow MLPs and Different Training Epochs.

Epochs / Number of Folds			Number of Hidden Layers
10 k | 10	**30 k | 10**	**60 k | 10**	
2.22%	2.19%	2.36%	1
2.36%	2.16%	2.55%	2
2.35%	2.21%	2.45%	3
2.25%	2.13%	2.46%	4
2.35%	2.26%	2.48%	5

Table 2
Results for Pairwise Comparisons of MLPs – Analysis 1.

	1 Hidden Layer	2 Hidden Layers	3 Hidden Layers	4 Hidden Layers	5 Hidden Layers
1 Hidden Layer	MAPE: 2.22% NSE: 0.897	DM: 2.08 p-value: 0.041	DM: 2.75 p-value: 0.007	DM: 1.28 p-value: 0.204	DM: 2.16 p-value: 0.034
2 Hidden Layers	–	MAPE: 2.36%NSE: 0.857	DM: 0.93 p-value: 0.350	DM: 0.56 p-value: 0.573	DM: 0.02 p-value: 0.986
3 Hidden Layers	–	–	MAPE: 2.35%NSE: 0.864	DM: 1.80 p-value: 0.074	DM: 0.96 p-value: 0.338
4 Hidden Layers	–	–	–	MAPE: 2.25% NSE: 0.859	DM: 0.707 p-value: 0.481
5 Hidden Layers	–	–	–	–	MAPE: 2.35%NSE: 0.859

Table 3
Results for Pairwise Comparisons – Analysis 2.

	Brazilian ISO	MLP	Uni-LSTM	Bi-LSTM	Uni-GRU	Bi-GRU
Brazilian ISO	MAPE: 2.006% NSE: 0.945	DM: 0.90 p-value: 0.37	DM: 5.46 p-value: 0.00	DM: 6.71 p-value: 0.00	DM: 5.45 p-value: 0.00	DM: 6.29 p-value: 0.00
MLP	–	MAPE: 2.114% NSE: 0.936	DM: 6.87 p-value: 0.00	DM: 8.55 p-value: 0.00	DM: 7.25 p-value: 0.00	DM: 8.18 p-value: 0.00
Uni-LSTM	–	–	MAPE: 1.306% NSE: 0.976	DM: 4.67 p-value: 0.00	DM: 0.57 p-value: 0.56	DM: 3.87 p-value: 0.0
Bi-LSTM	–	–	–	MAPE: 1.179% NSE: 0.979	DM: 3.24 p-value: 0.002	DM: 0.97 p-value: 0.33
Uni-GRU	–	–	–	–	MAPE: 1.329% NSE: 0.976	DM: 3.21 p-value: 0.001
Bi-GRU	–	–	–	–	–	MAPE: 1.205% NSE: 0.979

4.4. Analysis 3: Pairwise comparison based on GEFS temperature data

Tables 4 and 5 summarize the comparisons between model performances, the cells marked in grey indicate that the model listed in the column present the best performance in the pairwise comparison. In the Table 4 the results are about the dataset considering the Brazilian ISO temperature information as input for ANNs, and in the Table 5 is about the results considering the GEFS temperature forecasts data (Table 5). The test set for this comparison was defined as the year 2019, with the other years of data being used in the training set.

Tables 4 and 5 show that unidirectional and bidirectional LSTMs and GRUs present MAPEs smaller than the MLP, with a NSE metrics closer to 1. Furthermore, the Diebold-Mariano test indicates that the accuracy performance is statistically better for all RNN models. In this comparison, it is also possible to notice that most models obtained lower MAPE and higher NSE while considering the GEFS temperature forecasts. Finally, although in the analysis considering the ISO temperature information with a test set of only three months (Subsection 4.3), the bidirectional models performed significantly better than the unidirectional ones, the same result was not observed with the GEFS information and with a one-year test set.

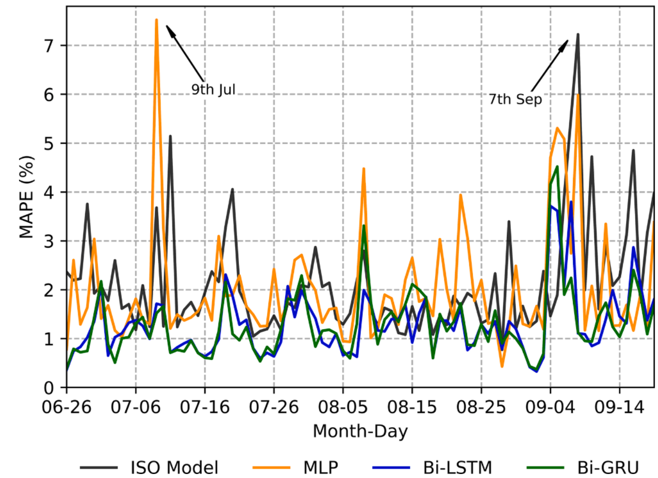


Fig 4. Accuracy performance of ANNs and ISO model: Holidays Indicated by Arrows.

Overall, the results presented in this section show the superiority of the RNN models for STLF. As observed in the accuracy metrics and the pairwise comparisons, the analyzed RNN models significantly outperformed the best MLP model. Lastly, Figs. 5 and 6 show the accumulated verified daily load and the percentage error from the hourly forecasts of all models during the 2019 test year. From these Figures, it is possible to observe that the models were able to accurately capture the seasonality present in the load during the year (small percentage error), and that the results of the bidirectional RNNs are generally more accurate than their unidirectional counterparts. Therefore, this class of RNNs should be further explored in time series forecasting tasks. Finally, our results also show that temperature forecasts from GCMs can be successfully used as input for ANN models and provide independence from ISO temperature information, which is very useful for electricity market agents who follow the ISO procedures.

5. Conclusions

The short-term load forecasting is an important and complex problem in electrical power systems literature. In real and large-scale systems, the problem is even more challenging, as there are different geographic areas with very specific climate conditions, population centers, and load profiles. This paper presented an assessment of different neural networks architectures to model the STLF problem in a large-scale system representing the Brazilian Central-Southeast market.

Long-Short Term Memory and Gated Recurrent Unit have shown superior performance than MLPs and the forecasting model currently used by the Brazilian Independent System Operator. Results of the present study validate the potential of the RNN models to STLF problems, especially bidirectional LSTMs and GRUs, which obtained reliable results (with Nash-Sutcliffe reaching values up to 0.98 and Mean Absolute Percentile Error values of 1.18%) while using different test sets, and temperature information weighted by population. Based on the Diebold-Mariano test comparisons, the Long-Short Term Memory and Gated Recurrent Unit have shown superior performance than MLPs and the Brazilian ISO forecasting model.

This paper has also shown the use of a Global Climate Models (GCM) in STLF in combination with ANN models. ANNs have shown reliable performance when using GCM data (with Nash-Sutcliffe reaching values

Table 4
Comparison results using Brazilian ISO Data – Analysis 3.

	MLP	Uni-LSTM	Bi- LSTM	Uni-GRU	Bi-GRU
MLP	MAPE: 2.196% NSE: 0.895	DM: 6.58 p-value: 0.000	DM: 6.98 p-value: 0.000	DM: 6.59 p-value: 0.000	DM: 7.30 p-value: 0.000
Uni-LSTM	–	MAPE: 1.668% NSE: 0.938	DM: 0.55 p-value: 0.585	DM: 1.19 p-value: 0.242	DM: 1.24 p-value: 0.231
Bi-LSTM	–	–	MAPE: 1.652% NSE: 0.943	DM: 1.62 p-value: 0.116	DM: 0.91 p-value:0.362
Uni-GRU	–	–	–	MAPE: 1.70% NSE: 0.944	DM: 2.76 p-value: 0.006
Bi-GRU	–	–	–	–	MAPE: 1.630% NSE: 0.945

Table 5
Comparison results using GEFS Data – Analysis 3.

	MLP	Uni-LSTM	Bi- LSTM	Uni-GRU	Bi-GRU
MLP	MAPE: 2.179% NSE: 0.886	DM: 6.12 p-value: 0.000	DM: 6.16 p-value: 0.000	DM: 6.81 p-value: 0.000	DM: 6.88 p-value: 0.000
Uni-LSTM	–	MAPE: 1.649% NSE: 0.936	DM: 0.31 p-value: 0.756	DM: 0.88 p-value: 0.377	DM: 1.24 p-value: 0.215
Bi-LSTM	–	–	MAPE: 1.659% NSE: 0.940	DM: 1.12 p-value: 0.261	DM: 1.85 p-value: 0.063
Uni-GRU	–	–	–	MAPE: 1.618% NSE: 0.941	DM: 0.34 p-value: 0.730
Bi-GRU	–	–	–	–	MAPE: 1.608% NSE: 0.945

up to 0.95 and Mean Absolute Percentile Error values of 1.6%). From these results, the GEFS GCM can potentially serve as a reliable alternative of input information to the ANN forecasting models used to model the STFL in the region. Also, other information from this GCM such as precipitation, air humidity, wind speed, and solar radiation could be investigated in the future as other predictors to the models. For a real large-scale system this is extremely significant because of the distinct characteristics of each region.

The results of bidirectional RNNs obtained in this work are promising, future studies should explore this architecture with modifications in the input data selection, activation functions and training process. Future studies should also aim to evaluate the performance of other ANN architectures in STLF, such as Bayesian Neural Networks and Convolutional Neural Networks. New studies evaluating input from other models, such as the European Center for Medium-Range Weather Forecasts are necessary.

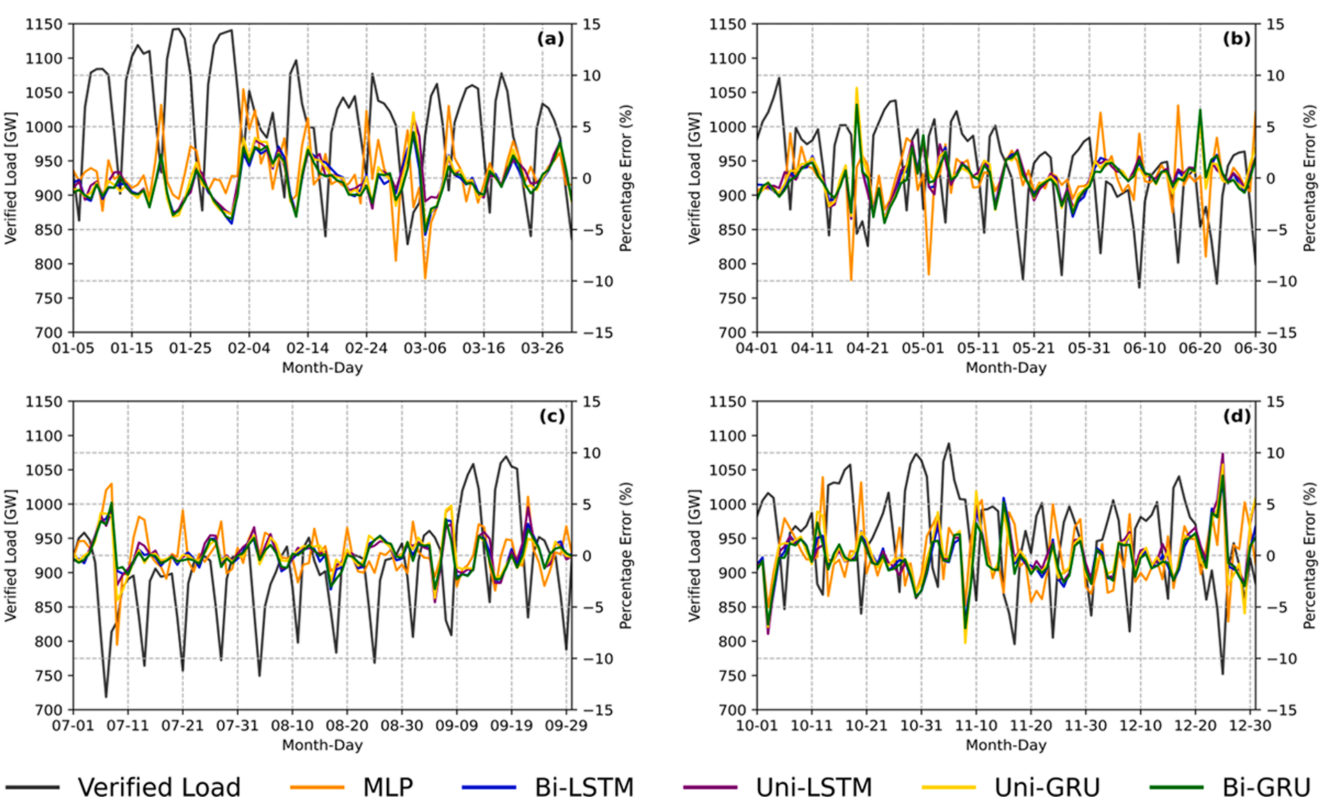


Fig. 5. Verified Electricity Demand (first y-axis) and Percentage Forecasts Error (second y-axis) for MLP, LSTM, and GRU Models Using the Brazilian ISO Temperature Data. (a) Summer (b) Fall (c) Winter (d) Spring.

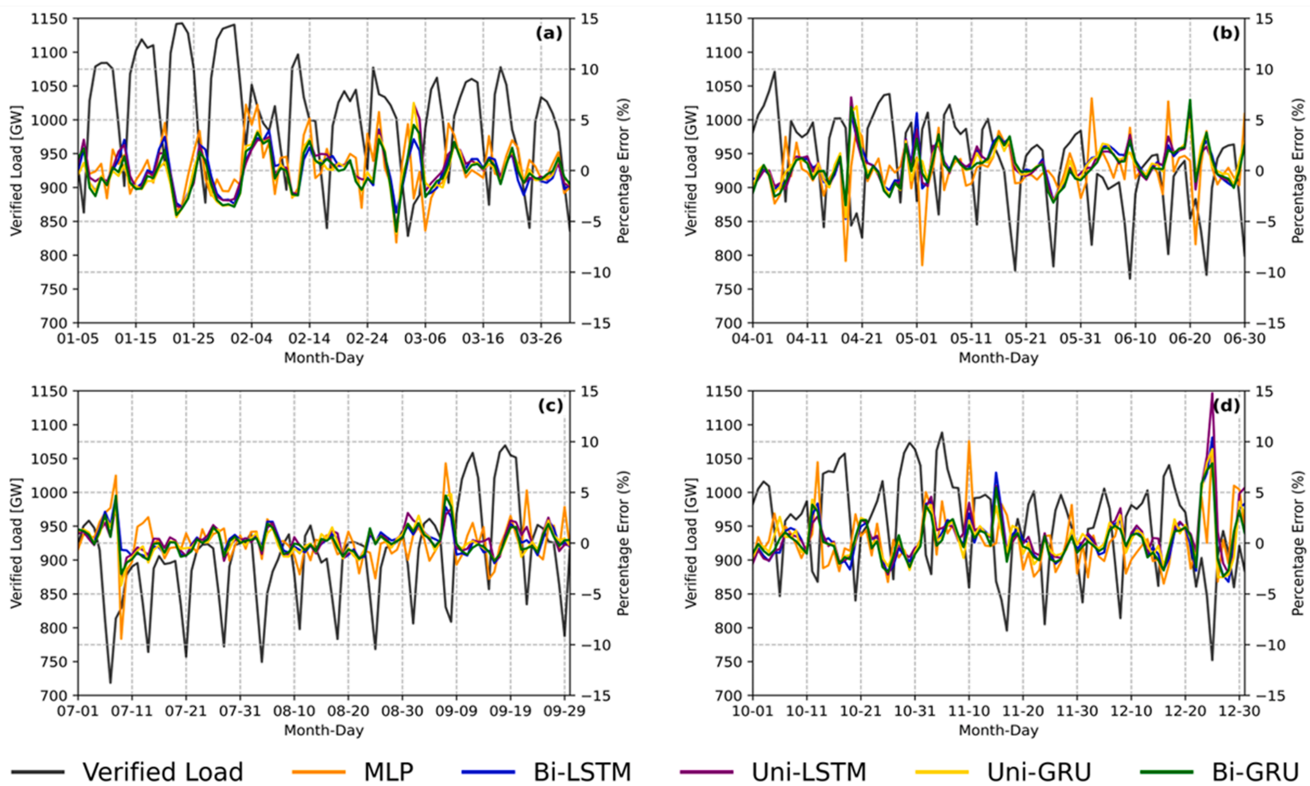


Fig. 6. Verified Electricity Demand (first y-axis) and Percentage Forecasts Error (second y-axis) for MLP, LSTM, and GRU Models Using the GEFS temperature data. (a) Summer (b) Fall (c) Winter (d) Spring.

CRediT authorship contribution statement

Lucas Barros Scianni Morais: Data curation, Methodology, Writing – original draft, Software. **Giancarlo Aquila:** Data curation, Writing – review & editing, Formal analysis. **Victor Augusto Durães de Faria:** Methodology, Writing – review & editing, Software. **Luana Medeiros Marangon Lima:** Methodology, Writing – review & editing. **José Wanderley Marangon Lima:** Data curation, Writing – review & editing. **Anderson Rodrigo de Queiroz:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The authors thank Energisa for the financial support on the ANEEL R&D project PD-06585-2003/2020. We also thank CAPES in Brazil Finance Code 001 for the financial support of the first author, CNPq in Brazil for the scholarship support number 318139/2021-9, and SemeAD (FEA-USP) of Foundation Institute of Administration and Cactus Payment Institution (SemeAD Scholarship - PQjr) of the second author.

References

[1] Shahidehpour M, Yamin H, Li Z. Market operations in electric power systems: forecasting, scheduling, and risk management. John New York: Wiley & Sons; 2003. p. 552.
[2] Mocanu E, Nguye PH, Gibescu M, Kling WL. Deep learning for estimating building energy consumption. Sustainable Energy Grids Networks 2016;6:91–9.
[3] Hafiz F, de Queiroz AR, Husain I. Solar generation, storage, and electric vehicles in power grids: challenges and solutions with coordinated control at the residential level. IEEE Electrif Mag 2018;6(4):83–90.
[4] Mehigan L, Deane JP, Gallachóir BÓ, Bertsch V. A review of the role of distributed generation (DG) in future electricity systems. Energy 2018;163:822–36.
[5] Sioshansi R, Denholm P, Arteaga J, Awara S, Bhattacharjee S, Botterud A, et al. Energy-storage modeling: state-of-the-art and future research directions. IEEE Transactions on Power Systems; 2021.
[6] Niu D, Wei Y. Short-term power load combinatorial forecast adaptively weighted by FHNN similar-day clustering. Automation of Electr Power Syst 2013;37(3):54–7.
[7] Hong T, Fan S. Probabilistic electric load forecasting: a tutorial review. Int J of Forecasting 2016;32(3):914–38.
[8] Kwon BS, Park RJ, Song KB. Short–term load forecasting based on deep neural networks using LSTM layer. J Electr Eng Technol 2020;15:1501–9.
[9] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell 2013;35:1798–828.
[10] Fan GF, Peng LL, Hong WC. Short term load forecasting based on phase space reconstruction algorithm and bi-square kernel regression model. Appl Energy 2018;224:13–33.
[11] Barman M, Choudhury ND, Sutradhar S. A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. Energy 2018;145:710–20.
[12] S.R. Gunn, 1998. "Support Vector Machine for Classification and Regression". Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science, Southampton, Tech Rep.
[13] C. Liu, Z. Jin, J. Gu, C. Qiu, 2017. Short-term load forecasting using a long short-term memory network. In Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Torino, Italy, 1–6.
[14] Panapakidis IP. Clustering based day-ahead and hour-ahead bus load forecasting models. I J Electr Power Energy Syst 2016;80:171–8.
[15] Jiao R, Zhang T, Jiang Y, He H. Short-Term non-residential load forecasting based on multiple sequences LSTM recurrent neural network. IEEE Access 2018;6:59438–48.
[16] Voß M, Bender-Saebelkampf C, Albayrak S. Residential short-term load forecasting using convolutional neural networks. In: In 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). IEEE; 2018. p. 1–6.

[17] Tian C, Hao Y. A novel nonlinear combined forecasting system for short-term load forecasting. Energies 2018;11(4):712.
[18] Sadaei HJ, Silva PCDL, Guimaraes FG, Lee MH. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. Energy 2019;175:365–77.
[19] Jin XB, Yang NX, Wang XY, Bai YT, Su TL, Kong JL. Hybrid deep learning predictor for smart agriculture sensing based on empirical mode decomposition and gated recurrent unit group model. Sensors 2020;20(5):1334.
[20] Bouktif S, Fiaz A, Ouni A, Serhani MA. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches. Energies 2018;11:1636.
[21] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans on Smart Grid 2017;10(1):841–51.
[22] Veeramsetty V, Chandra DR, Grimaccia F, Mussetta M. Short term electric power load forecasting using principal component analysis and recurrent neural networks. Forecasting 2022;4(1):149–64.
[23] Haque A, Rahman S. Short-term electrical load forecasting through heuristic configuration of regularized deep neural network. Appl Soft Comput 2022;122:108877.
[24] Lin J, Ma J, Zhu J, Cui Y. Short-term load forecasting based on LSTM networks considering attention mechanism. Int J Electr Power Energy Syst 2022;137:107818.
[25] Yang D, Guo JE, Sun S, Han J, Wang S. An interval decomposition-ensemble approach with data-characteristic-driven reconstruction for short-term load forecasting. Appl Energy 2022;306:117992.
[26] Yue W, Liu Q, Ruan Y, Qian F, Meng H. A prediction approach with mode decomposition-recombination technique for short-term load forecasting. Sustain Cities Soc 2022;85:104034.
[27] Subbiah SS, Chinnappan J. Deep learning based short term load forecasting with hybrid feature selection. Electr Pow Syst Res 2022;210:108065.
[28] Santos TN, Diniz AL, Saboia CH, Cabral RN, Cerqueira LF. Hourly pricing and day-ahead dispatch setting in Brazil: the dessem model. Electr Pow Syst Res 2020;189:106709.
[29] Wang Y, Gan D, Sun M, Zhang N, Lu Z, Kang C. Probabilistic individual load forecasting using pinball loss guided LSTM. Appl Energy 2019;235:10–20.
[30] Cai C, Tao Y, Zhu T, Deng Z. Short-Term load forecasting based on deep learning bidirectional LSTM neural network. Appl Sci 2021;11(17):8129.
[31] A. Wahab, M. A. Tahir, N. Iqbal, F. Shafait, S. M. R. Kazmi, 2020. Short-Term Load Forecasting using Bi-directional Sequential Models and Feature Engineering for Small Datasets. arXiv preprint arXiv:2011.14137.
[32] Ullah FUM, Ullah A, Haq IU, Rho S, Baik SW. Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks. IEEE Access 2019;8:123369–80.
[33] Li K, Huang W, Hu G, Li J. Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network. Energ Build 2023;279:112666.
[34] Bashir T, Haoyong C, Tahir MF, Liqiang Z. Short-term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. Energy Rep 2022;8:1678–86.
[35] Javed U, Ijaz K, Jawad M, Khosa I, Ansari EA, Zaidi KS, et al. A novel short receptive field based dilated causal convolutional network integrated with Bidirectional LSTM for short-term load forecasting. Expert Syst Appl 2022;205:117689.
[36] Rafi SH, Deeba SR, Hossain E. A short-term load forecasting method using integrated CNN and LSTM network. IEEE Access 2021;9:32436–48.
[37] Guo X, Zhao Q, Zheng D, Ning Y, Gao Y. A short-term load forecasting model of multi-scale CNN-LSTM hybrid neural network considering the real-time electricity price. Energy Rep 2020;6:1046–53.
[38] Liang Y, Niu D, Hong WC. Short term load forecasting based on feature extraction and improved general regression neural network model. Energy 2019;166:653–63.
[39] ONS – Brazilian Independent System Operator (2022). Electricity Load. Available at:< http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/carga_energia.aspx>.
[40] Diebold FX, Mariano RS. Comparing predictive accuracy. J Bus Econ Stat 1995;13:253–63.
[41] Potdar K, Pardawala TS, Pai CD. A comparative study of categorical variable encoding techniques for neural network classifiers. Int j comput applications 2017;175(4):7–9.
[42] I. Goodfellow, Y. Bengio, A. Courville, 2016 Deep Learning, MIT Press. Available at: <http://www. deeplearningbook.org>.
[43] Y. Gal, Z. Ghahramani, 2016. A Theoretically Grounded Application Of Dropout In Recurrent Neural Networks. In: Lee, D.D.; Sugiyama, M.; Luxburg, U.V.; Guyon, I.; Garnett, R. Advances in Neural Information Processing Systems 29 (NIPS 2016), pp. 1019–1027. Curran Associates Inc., Barcelona.
[44] D. Anguita, L. Ghelardoni, A. Ghio, L. Oneto, S. Ridella, 2012. The 'K'in K-fold cross validation. In 20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), pp. 441-446.
[45] Wong T, Yeh P. Reliable accuracy estimates from k-fold cross validation. IEEE Trans Knowl Data Eng 2020;32(8):1586–94.
[46] Saud S, Jamil B, Upadhyay Y, Irshad K. Performance improvement of empirical models for estimation of global solar radiation in India: a k-fold cross-validation approach. Sustainable Energy Technol Assess 2020;40:100768.
[47] Hippert HS, Pedreira CE, Souza RC. Neural networks for short-term load forecasting: a review and evaluation. IEEE Trans On Power Systems 2001;16(1):44–55.
[48] Graves A. Generating sequences with recurrent neural networks. Computer Science; 2013.
[49] Faria VAD, de Queiroz AR, Lima LM, Lima JWM, Silva BC. An assessment of multi-layer perceptron networks for streamflow forecasting in large-scale interconnected hydro systems. Science and Technology: Int. J. Environ; 2021.
[50] Hochreiter S, Schmidhuber J. Long short-term memory. J Neural Comput 1997;9(8):1735–80.
[51] Zheng J, Xu C, Zhang Z, Li X. Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network. IEEE: Information Sciences and Systems; 2017.
[52] Veeramsetty V, Reddy KR, Santhosh M, Mohnot A, Singal G. Short-term electric power load forecasting using random forest and gated recurrent unit. Electr Eng 2021:1–23.
[53] F.M. Bianchi, E. Maiorino, M.C. Kampffmeyer, A. Rizzi, R. Jenssen, 2017. An overview and comparative analysis of recurrent neural networks for short-term load forecasting.
[54] Yu Z, Ramanarayanan V, Suendermann-Oeft D, Wang X, Zechner K, Chen L, et al. Using bidirectional lstm recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech. IEEE Workshop on Automatic Speech Recognition and 2015. Understanding.
[55] Haykin S. Neural Networks: a comprehensive foundation, vol 2. Porto Alegre: Bookman; 2003.
[56] Centro de Pesquisas de Energia Elétrica – CEPEL (2019). Modelo de Previsão de Carga para o DESSEM – PrevCargaDESSEM. p.37 (in Portuguese).
[57] Chen H, Wan Q, Wand Y. Refined diebold-mariano test methods for the evaluation of wind power forecasting models. Energies 2014;7.
[58] ONS – Brazilian Independent System Operator (2022). PrevCargaDESSEM 2.0 Methodology and Use Manual. *Centro de Pesquisas de Energia Elétrica* (CEPEL).
[59] ONS – Brazilian Independent System Operator (2022). SINtegre Platform. Available at:< http://www.ons.org.br/topo/acesso-restrito>.
[60] REDEMET - Air Force Command Meteorology Network. REDEMET 5.0. Available at:< https://www.redemet.aer.mil.br/>.
[61] CPTEC/INPE - Center for Weather Prediction and Climate Studies. Weather Prevision Center and Climate Studies. Available at:< https://www.cptec.inpe.br/>.
[62] NCEP - National Centers for Environmental Prediction. National Weather Service. Available at: <https://www.ncep.noaa.gov/>.
[63] Yin R, Li D, Wang Y, Chen W. Forecasting method of monthly wind power generation based on climate model and long short-term memory neural network. Global Energy Interconnection 2020;3(6):571–6.
[64] Global Ensemble Forecast System – GEFS. Products: Weather Climate Models – GEFS, 2022.
[65] Cawthorne D, De Queiroz AR, Eshraghi H, Arumugam S, DeCarolis JF. The role of temperature variability on seasonal electricity demand in the southern US. Frontiers in Sustainable Cities 2021;3:43.
[66] Eshraghi H, de Queiroz AR, Sankarasubramanian A, DeCarolis JF. Quantification of climate-induced interannual variability in residential US electricity demand. Energy 2021;236:121273.
[67] Dahl M, Brun A, Kirsebom OS, Andresen GB. Improving short-term load forecasts with calendar and holiday data. Energies 2018;11(7):1678.
[68] Song K, Baek Y, Hong DU, Jang G. Short-term load forecasting for the holidays using fuzzy-linear method. IEEE Trans on Power Syst 2005;20(1):96–101.