Article

Forecasting Irregular Seasonal Power Consumption. An Application to a Hot-Dip Galvanizing Process

Oscar Trull [1,*], Juan Carlos García-Díaz [1] and Angel Peiró-Signes [2]

1 Department of Applied Statistics, Operational Research and Quality, Universitat Politècnica de València, E-46022 Valencia, Spain; juagardi@eio.upv.es
2 Management Department, Universitat Politècnica de València, E-46022 Valencia, Spain; anpeisig@omp.upv.es
* Correspondence: otrull@eio.upv.es

Featured Application: The method described in this document makes it possible to use the techniques usually applied to load prediction efficiently in those situations in which the series clearly presents seasonality but does not maintain a regular pattern.

Abstract: Distribution companies use time series to predict electricity consumption. Forecasting techniques based on statistical models or artificial intelligence are used. Reliable forecasts are required for efficient grid management in terms of both supply and capacity. One common underlying feature of most demand–related time series is a strong seasonality component. However, in some cases, the electricity demanded by a process presents an irregular seasonal component, which prevents any type of forecast. In this article, we evaluated forecasting methods based on the use of multiple seasonal models: ARIMA, Holt-Winters models with discrete interval moving seasonality, and neural networks. The models are explained and applied to a real situation, for a node that feeds a galvanizing factory. The zinc hot-dip galvanizing process is widely used in the automotive sector for the protection of steel against corrosion. It requires enormous energy consumption, and this has a direct impact on companies' income statements. In addition, it significantly affects energy distribution companies, as these companies must provide for instant consumption in their supply lines to ensure sufficient energy is distributed both for the process and for all the other consumers. The results show a substantial increase in the accuracy of predictions, which contributes to a better management of the electrical distribution.

Keywords: time series; demand; load; forecast; DIMS; irregular; galvanizing

1. Introduction

Demand management is a primary process in the development of industrial activity. Distribution companies must ensure a supply is provided at a reasonable cost, and for this reason, they need to manage resources efficiently. The use of electrical prediction models contributes to their management of the distribution lines by offering tools to estimate future demand with great precision. The techniques allow for forecasting based on time series using statistical models or artificial intelligence (AI).

The most widely used univariate forecasting tools for electricity demand can be classified into three broad groups [1]: fundamental models, statistical models, and computational models. There is growing interest in the use of computational models, although the most widely used models are statistical models, both exponential smoothing models and autoregressive integrated moving average (ARIMA) models.

The fundamental models are made up of hybrid models that introduce all the possible physical variables, adopting a complex relationship between them and also using the techniques of statistical models.

Computational models are based on AI and emulate natural behaviors through the use of mathematical models. These are algorithms whose learning is automatic and are

part of the science of Machine Learning [2]. At present, deep learning techniques represent an evolution and have found applications in demand forecasting, especially in areas where prediction is difficult, such as renewable energies [3]. The most widely used techniques for electricity demand are artificial neural networks (ANN) [4], particularly non-linear autoregressive neural networks with exogenous variables (NARX) [5,6]. Support vector machines (SVM) [7] and bagged regression trees (BRT) [8] also stand out, and these occasionally apply fuzzy logic [9].

Electricity demand series show stochastic behavior, and they have traditionally been modeled using statistical methods. The ARIMA models are considered to be the econometric models par excellence. The Box–Jenkins methodology [10] is used to determine which ARIMA model to use, although some authors [11] state that simpler methods are better than this methodology at providing forecasts. The application of ARIMA models to demand is usually carried out in a general way in Seasonal Autoregressive Integrated Moving Average Exogenous (SARIMAX) models [12–14] in which exogenous variables are included to improve demand. The introduction of two seasonalities allows substantial improvement in the predictions of these models [15].

State-space models (SSM) are a form of exponential smoothing representation. They are commonly applied to demand [16], especially since the introduction of the Kalman filter (see [17]). They also allow the introduction of various seasonalities in a complex way [18] and with covariates [19]. De Livera and Hyndman include modifications that include adjustment of the error using autoregressive moving average (ARMA) models and with Box-Cox transformations (BATS [20]) and trigonometric seasonality (TBATS [18]).

Other very common smoothing techniques are the Holt-Winters models [20]. These models are excellent predictors for time series with marked seasonality [1,21]. The inclusion of more seasonality [22–24] improves their forecasts, leading to the development of multiple seasonal Holt-Winters models (nHWT). Trull et al. [25] introduce discrete seasonality that takes into account seasonalities whose occurrences are not regular (nHWT-DIMS models).

The current trend is to create hybrid models in which traditional techniques are combined with machine learning [26,27]. An example can be found in [28], which applies an exponential smoothing method and neural networks to divide the forecasting process between a linear part and a non-linear part.

The use of wavelets for irregular series has been combined with ARIMA models [29], Holt-Winters models [30], or ANN [31].

Regularization techniques have also been applied to prevent over- and under-fitting issues, based on a Least Absolute Shrinkage and Selection Operator (LASSO) [32], and have been applied to short-term load forecasting models based on multiple linear regression [33]. Banded regularization is also used to estimate parameters without overfitting in autoregressive models [34].

Newer methods use an anti-leakage least-squares spectral analysis (ALLSSA) to simultaneously estimate the trend and seasonal components before making a regularization and make forecasts [35]. The ALLSSA method determines the statistically significance of the components preventing under- and over-fitting issues. The least-squares wavelet analysis (LSWA) is a natural extension of the least-squares spectral analysis and allows the forecaster to obtain spectrograms for equally and unequally spaced time series and identify statistically significant peaks in the time series [36].

One common feature of most demand-related time series is their strong seasonality components [37]. In some cases, the electricity demanded by a process could present an irregular seasonal component that seriously distorts the behavior of the series in a way that the models cannot deal with.

The zinc hot-dip galvanizing process is a process that is widely used in the automotive sector to protect steel against corrosion [38,39]. It requires an enormous consumption of energy, and this has a direct impact on companies' income statements. However, the process also significantly affects energy distribution companies, since they must foresee

the instantaneous consumption in their lines in order to ensure the distribution of energy both for the process and for the other consumers. A characteristic of the demand in this process is the presence of seasonal patterns that resemble seasonality but, because of their irregular behavior, are difficult to assimilate to seasonality.

The structure shown by the series in this study means that it is more suitable to work with time series models rather than frequency or signal analysis. We have therefore considered it convenient to preferably use traditional time series models with seasonality.

In this article, we present several solutions to this problem based on the use of ARIMA models, multiple seasonal Holt-Winters models with and without discrete interval moving seasonalities (DIMS), state space models, and neural network models. To verify the effectiveness of the techniques described, they are applied to the industrial process of hot-dip galvanizing.

The article is organized as follows: Section 2 conducts a review of the forecasting methods as well as an explanation of the production process; Section 3 demonstrates the results and their analysis; Section 4 discusses the results; and finally, in Section 5, the conclusions are summarized.

2. Materials and Methods

2.1. Study Area

The study has been applied to the consumption node of a hot-dip galvanizing company. The process is carried out by coating extruded steel strips with a zinc foil that forms an alloy with the steel and gives the desired properties. This process is continuous and produces high-quality products [40].

Figure 1 shows a general scheme for the galvanizing process, where the greatest consumption is in the zinc bath. In the annealing furnace, the steel strip is preheated, and then it is immersed in a bath of molten zinc at 460 °C. Subsequently, the galvanized steel strip goes through the skin-pass process [41–43] after it has cooled down.
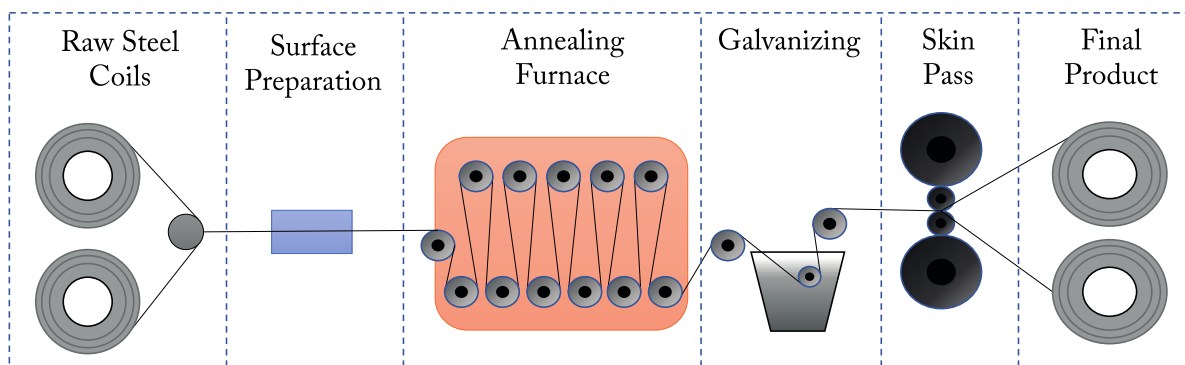
Figure 1. Representation of a hot dip galvanizing process.

The zinc bath consists of a molten alloy of Zn in a bath, which is kept at 460 °C by the action of two heating inductors located at the bottom of the bath. Figure 2 schematically shows the operation of the zinc bath. The bath temperature is measured as an average of the local temperatures provided by the thermocouples Ta, Tb and Tc. The inductors heat from the bottom of the bath and a natural flow inside the bath is produced so that the bath achieves the targeted temperature.
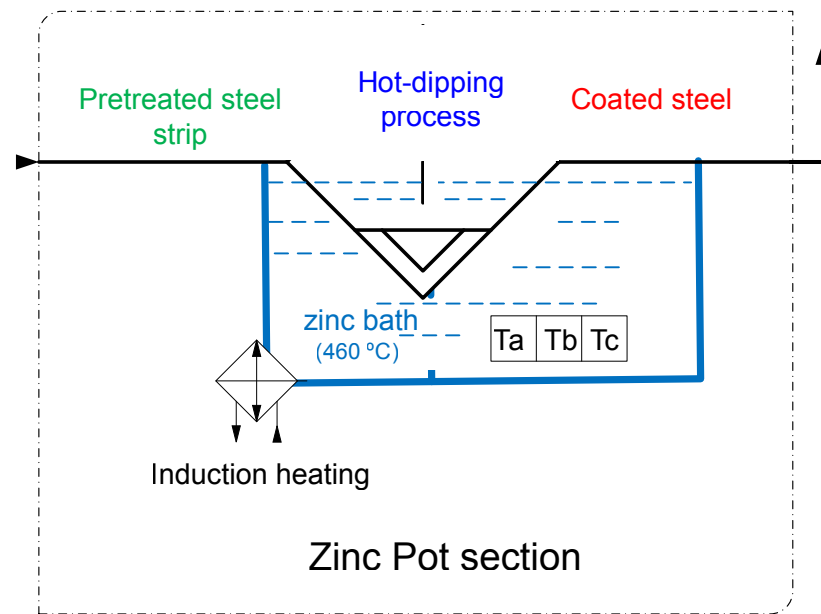
Figure 2. Galvanizing section (hot dip zinc bath). Pretreated steel goes into the zinc pot, which is filled with an Al–Zn solution at 460 °C. Thermocouples Ta, Tb and Tc measure the local temperatures in the bath. Induction heaters located at the base of the bath keep the temperature as targeted. After the bath, the steel is coated with Zn.

The electrical consumption associated with the process can be seen in Figure 3. This graph shows the consumption for eight working days, measured every six minutes. It begins on 14th November, 2009 at 00:00 am and ends on 22nd November, 2009 at 08:00 am. There are in total 2000 measurements. The oscillations shown in the time series are produced by the action of induction heaters that keep the bath at the targeted temperature. The big peaks in consumption are produced when the bath needs to be recharged, and new Zn (dropped in ingots) is added into the bath. At this moment, the heaters must be put into full operation. From this dataset, the first 1800 observed values are used for training purposes, and the last 200 ones are used for validation.
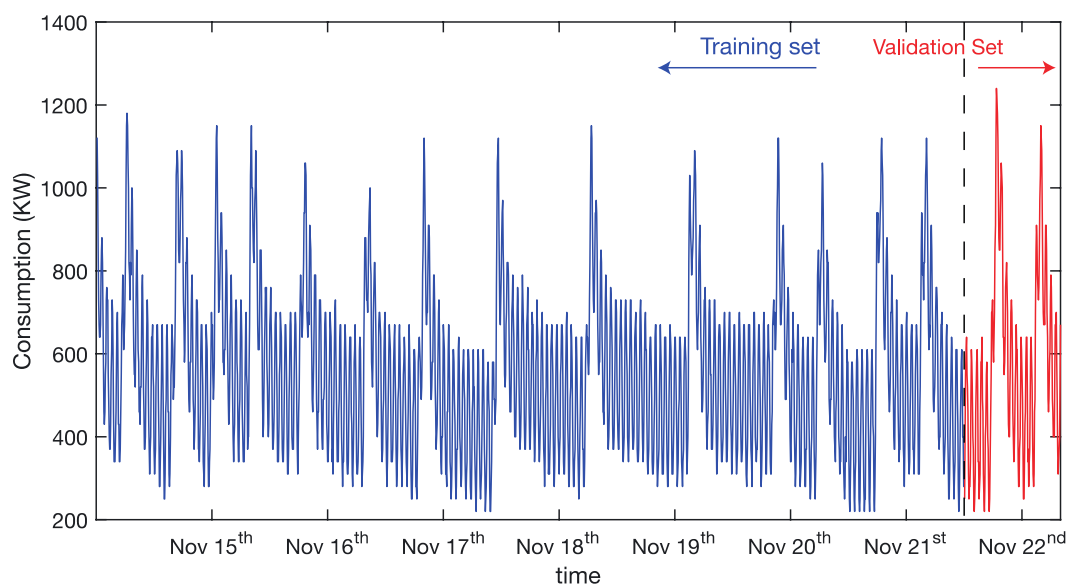
Figure 3. Electricity demand for the hot-dip galvanizing. The ticks represent the beginning of each day. The blue dataset designates the data used for training, whereas the red one represents the data used for testing and validation.

A series of cyclical patterns can be observed (oscillations) that are repeated throughout the series, with a short–length pattern that is repeated continuously throughout the series clearly standing out. There are other patterns that are repeated irregularly, with differentiated behaviors. A closer view of the series is shown in Figure 4. In graph (a) and graph (b), a common underlying pattern can be identified, with a length of around ten time units (which means an hour, as the temperature is measured every six minutes). This first pattern is repeated regularly over the whole time period, and it is considered as a seasonality.
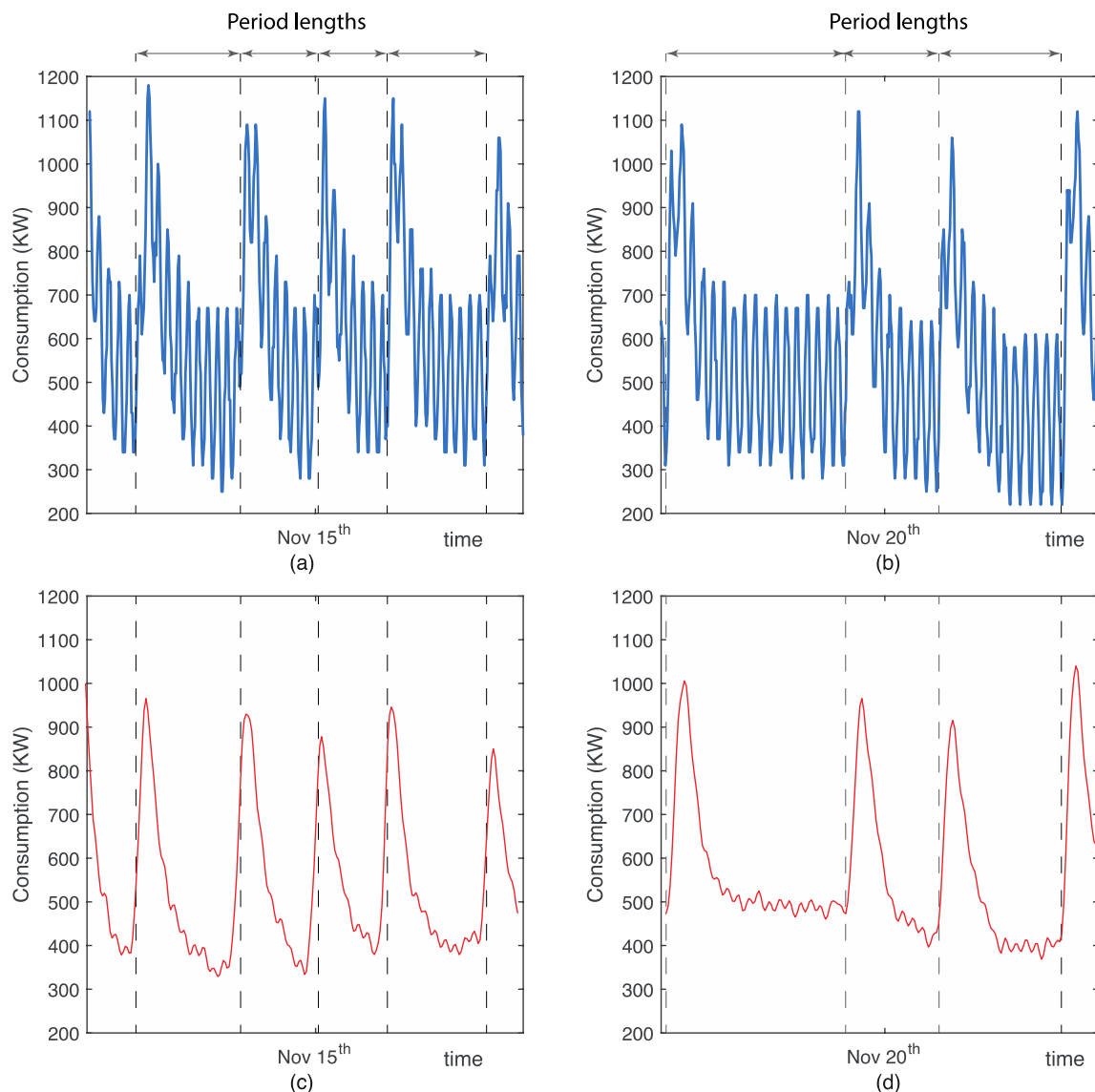
Figure 4. Close-up version of Figure 3, where different seasonal patterns can be located: a first pattern along the whole series, with sort oscillations as shown in (**a**,**b**); and a second pattern covering the consumption peaks, as shown in (**c**,**d**). This second pattern has a different length on every appearance.

Figure 4, graph (c) and (d) show the time series after removing the first seasonal pattern. It can also be seen that other patterns develop throughout the series in such a way that the time of their appearance or their length are not constant. Technically, this non-regular behavior cannot be considered as a seasonality, since it is not possible to predict the fluctuation pattern that will develop in the future. To make consumption

predictions, it is necessary to take into account this seasonal behavior, even though it is not regular.

2.2. Forecasting Methods

In this section, we describe the forecasting methods applied to the time series under study. The most common methods applied to short-term electricity demand forecasting, using both AI and statistical methods, have been chosen. First, the methods used with regular seasonal behavior are described, and then we describe the models with discrete seasonality.

2.2.1. Artificial Neural Networks

Neural networks are computational models structured in the form of layers with nodes interconnected as a network. They are named because of their resemblance to the human brain structure. The nodes, called neurons, perform simple operations in parallel and are located in each of the layers of the network. The layers are of three different types: the input layer, where neurons receive direct information from the inputs; hidden layer (s), whose neurons use the information from the neurons of previous layers and feed the next layers; and the output layer, where neurons use the information from the hidden layers to produce an output. Thus, there is an input layer, one or more hidden layers, and an output layer. The connections between the different layers are made through the connection of their neurons, which are called synapses. The strength of the connection between neurons is determined by a weighting established at the synapse.

The most suitable structure for forecasting time series is the NARX type structure [44,45]. It is a recurrent dynamic neural network, with feedback connections. Figure 5 shows a close-loop representation of the NARX structure [46]. Neurons receive information from exogenous input variables in addition to the target series itself and the feedbacks. In order to improve forecasts, it can be used the past predicted and observed values delayed through a tapped delay line (TDL) memory. The circles after the input layers denote the TPL delay (e.g., one to two delays in the figure).
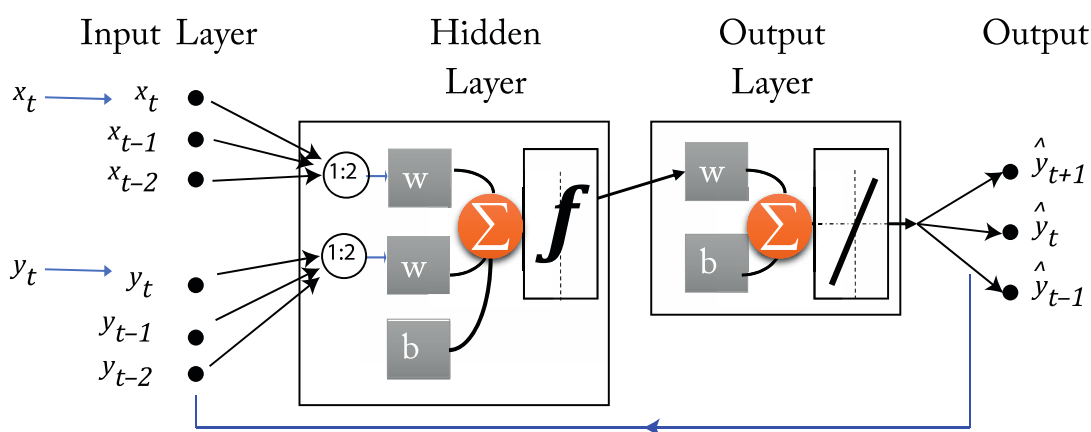
Figure 5. NARX neural network schema. There is an input layer with variables, one hidden layer and one output layers. Circles represent tapped delay line (TDL).

The input variables x_t are exogenous variables used in the model. Both x_i and y_t are connected by axioms to which weights w_i are assigned, and with an activation function f that is integrated with an aggregation function Σ. The output $\hat{y}_{t+1}$ provides future forecasts after the network has been trained. b stands for the bias whose presence increases or decreases the neuron's processing capacity.

The mathematical and non-linear representation that governs the network is shown in (1), where x_t represents the inputs and y_t the objective function, while $\hat{y}_{t+1}$ represents the prediction. D_x and D_y are the time delays applied in the network.

$$\hat{y}_{t+1} = f\left[x_t, x_{t-1}, \ldots, x_{t-D_x+1}, y_t, y_{t-1}, \ldots, y_{t-D_y+1}\right]. \tag{1}$$

The NARX neural network maps the function through the multilayer perceptron, using the time delays for both the input variables and the output feedback [47].

An alternative to this neural network is a function fitting neural network. This is a type of shallow neural network based on multilayer perceptron (MLP) with which we can make adjustments to non-linear functions (non-linear regression, NLR). The use and application of such a network for the prediction of electricity demand has been discussed previously [48]. The mathematical representation that governs this network is shown in (2).

$$\hat{y}_{t+1} = f[x_t, x_{t-1}, \ldots, x_{t-D_x+1}]. \tag{2}$$

Here x_t are the predictors, which are several variables (including the observed values of the time series) used to feed the model. A representative schema for this neural network is shown in Figure 6.
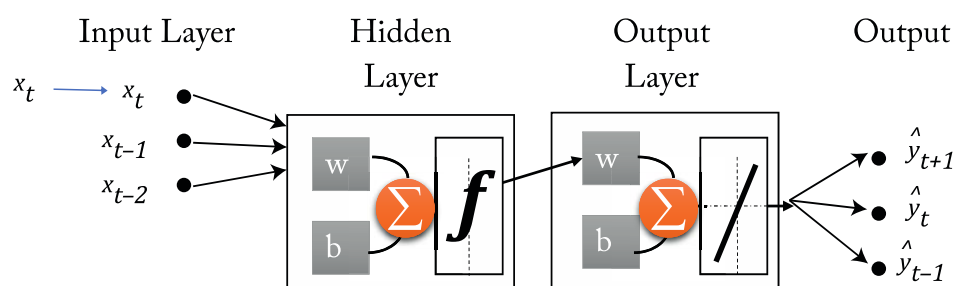
Figure 6. Function fitting neural network schema.

By training the network, weights are assigned to the synaptic connections, minimizing an error criterion. The ANNs used in this work are trained using the Levenberg-Marquardt algorithm [49], and minimizing the mean squared error (MSE). After the training process, to give the predictions, a closed loop network is performed, and forecasts are provided.

2.2.2. ARIMA Models

ARIMA models were introduced by Box and Jenkins [50] to model non–stationary series and allow predictions to be made. A description and in-depth analysis can be found in [51] and in the book by Brockwell and Davis [52]. Seasonal ARIMA models are usually denoted by $ARIMA(p, d, q) x (P, D, Q)_S$. S indicates the length of the seasonal pattern under consideration. The compact representation of the ARIMA model is usually, as shown in (3), a function of autoregressive polynomials and polynomials of moving means, and of the difference operators.

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D\left(y_t^{(\lambda)} - c\right) = \theta_q(B)\Theta_Q(B^S)\varepsilon_t,$$
$$\{\varepsilon_t\} \sim N(0, \sigma^2). \tag{3}$$

$\{y_t, \ t = 0, \pm1, \pm2, \ldots\}$ are the observed data of the univariate series. If the variability in the data grows with time, it is necessary to transform the data to stabilize the variance. The Box-Cox power transformation family is a general class of variance-stabilizing trans-

formations. The Box-Cox transformation of y_t with power parameter λ to the transformed data $y_t^{(\lambda)}$ is defined by (4).

$$y_t^{(\lambda)} = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}; & if\ \lambda \neq 0, \\ ln\ y_t; & if\ \lambda = 0\ . \end{cases} \tag{4}$$

The power parameter λ is estimated by the maximum–likelihood method. The polynomials $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ and $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ represent the regular or non–seasonal autoregressive and the moving averages components, respectively, and the polynomials $\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \cdots - \Phi_P B^{PS}$ and $\Theta_Q(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \cdots - \Theta_Q B^{QS}$ represent the seasonal autoregressive and the moving averages components, respectively, with B as the lag operator. ∇ is the is the backward difference operator, $[By_t = y_{t-1};\ B^S y_t = y_{t-S};\ \nabla = (1 - B);\ \nabla^d = (1 - B)^d;\ \nabla_S^D = (1 - B^S)^D]$. d and D are the number of differencings required to make the time series stationary ($d, D \leq 2$). $\{\varepsilon_t\}$ is a Gaussian white noise process, $[\{\varepsilon_t\} \sim N(0, \sigma^2)]$. c is the model constant.

The orders of the polynomials $\{p, d; P, Q\}$ are selected using the Akaike's Information Criterion (*AIC*, AICc) or Schwarz's or the Bayesian Information Criterion (*SIC* or *BIC*). The model coefficients $\{\phi_1, \phi_2, \ldots, \phi_p;\ \theta_1, \theta_2, \ldots, \theta_q;\ \Phi_1, \Phi_2, \ldots, \Phi_P; \Theta_1, \Theta_2, \ldots, \Theta_Q\}$ and σ^2 are estimated by the maximum likelihood method.

ARIMA models can present more than one seasonality, as indicated in (5). To do this, the models are expressed as ARIMA $(p, d, q) x (P_1, D_1, Q_1)_{S_1} x (P_2, D_2, Q_2)_{S_2}$ where $S_1\ and\ S_2$ indicate the two seasonalities to which they refer.

$$\phi_p(B)\Phi_{P_1}\left(B^{S_1}\right)\Omega_{P_2}\left(B^{S_2}\right)\nabla^d\nabla_{S_1}^{D_1}\nabla_{S_2}^{D_2}\left(y_t^{(\lambda)} - c\right) = \theta_q\ (B)\Theta_{Q_1}\left(B^{S_1}\right)\Psi_{Q_2}\left(B^{S_2}\right)\varepsilon_t. \tag{5}$$

The polynomials $\Omega_{P_2}(B^{S_2})$ and $\Psi_{Q_2}(B^{S_2})$ represent the second seasonal autoregressive and the moving averages components, respectively.

2.2.3. Multiple Seasonal Holt-Winters Models

Exponential smoothing uses information from the past through weighted averages to make predictions. The weight decreases as newer values are entered into the time series, giving more importance to newer data over older. A smoothing parameter determines this weight. The introduction of these models dates back to the 1960s with the work of Holt [53] and Brown [54]. Winters [20] presented the Holt-Winters models, in which exponential smoothing techniques are performed on the three components of the series: level (l_t), trend (b_t) and seasonality (s_t). The model includes a series of structured equations, called smoothing equations, the information from which is compiled by a forecast equation to provide forecasts. The equations can be combined with additive or multiplicative trends and seasonality.

Gardner and McKenzie [55] introduced a damping factor for the trend, and their model outperforms the previous models when the trend shows high variations [56]. Taylor broke down seasonality into two or three nested components so that the models can capture the series that present more than one seasonality, such as series for short-term demand [22,23]. Taylor also included in the model an adjustment using the one-step-ahead error as proposed by Chatfield [57]. This adjustment adds an AR(1) model for the residuals, obtaining the parameter at the same time as the smoothing parameters are obtained. In the same way, García-Díaz and Trull [24] generalized the model including the way the initial values are obtained, to n seasonalities. The nHWT models are shown in Equations (6)–(9).

$$l_t = \alpha\left(\frac{y_t}{\prod s_{t-s_i}^{(i)}}\right) + (1 - \alpha)(l_{t-1} + \varrho b_{t-1}), \tag{6}$$

$$b_t = \gamma(l_t - l_{t-1}) + (1 - \gamma)\varrho b_{t-1}, \tag{7}$$

$$s_t^{(i)} = \delta^{(i)} \left(\frac{y_t}{l_t \prod_{j \neq i} s_{t-s_j}^{(j)}} \right) + \left(1 - \delta^{(i)}\right) s_{t-s_i}^{(i)}, \tag{8}$$

$$\hat{y}_{t+k} = \left(l_t + \sum_{j=1}^{k} \varrho^j b_t \right) \prod_i s_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t. \tag{9}$$

The smoothing equations include the smoothing parameters α, γ and $\delta^{(i)}$ for smoothing the level, the trend and the different seasonal indices (i) of length s_i. The equation $\hat{y}_{t+k}$ provides the k–future prediction values from the observed values of the series y_t. Here, ε_t is the one–step–ahead error, and the parameter φ_{AR} is the parameter for the AR(1) adjustment. The damping parameter for the trend is denoted by ϱ [58].

The equations of the model are recursive, and therefore they need initial values so that they can fit the model. Several methodologies for initialization have been documented [56,57]. To be able to use the models, it is necessary to estimate the smoothing parameters by minimizing the error using non-linear algorithms [59,60]. The Nelder-Mead [61,62] simplex method has been used, which minimizes the root of mean squared error (RMSE).

2.2.4. State Space Models

The SSM refers to a form of graphical–probabilistic representation [63] to describe the dependence between an observed measurement and a series of latent state variables through equations called state equations that describe its evolution. Taking into account the fact that a time series can be decomposed into components of level, seasonality and trend, this terminology applied to time series would be understood as the model that interprets the evolution of the relationship of the observed variables (y_t) with the latent unobservable variables (level, trend, and seasonality).

SSMs have a great variety of formulations. In this paper, the formulation indicated by Durbin and Koopman [64] and Hyndman et al. [16] applied to univariate stochastic time series is used. These models are structured through a formulation of two matrix equations, as shown in (10)–(11):

$$y_t = \mu_t + \boldsymbol{r} x_{t-1} + \varepsilon_t, \tag{10}$$

$$x_t = \boldsymbol{f} x_{t-1} + \boldsymbol{g} \varepsilon_t. \tag{11}$$

Equation (11) is known as the state transition equation, and Equation (10) is known as the observation equation. Here x_{t-1} is known as the vector of states, y_t is the vector of observations, while ε_t is a vector of Gaussian white noise and is known as the innovation process. $\boldsymbol{r}$, $\boldsymbol{f}$ and $\boldsymbol{g}$ are matrices and vectors of coefficients with appropriate dimensions. $\boldsymbol{f}$ explains the evolution of x_t and $\boldsymbol{g}$ provides the innovation correction of ε_t. The term μ_t is the one step ahead forecast, and $\boldsymbol{r}$ is a term to include the error additively.

De Livera [18] introduced modified models, based on the exponential smoothing methods, in which a Box-Cox transformation is applied to the data, and the residuals are modeled using an ARMA process and include the damping factor for trend and multiple seasonalities. The acronym for this method is BATS (Box-Cox transform ARMA errors Trend and Seasonal Components). This model is described in (12)–(16).

$$y_t^{(\lambda)} = l_{t-1} + \varrho b_{t-1} + \sum_{i=1}^{n_s} s_{t-s_i}^{(i)} + d_t, \tag{12}$$

$$l_t = l_{t-1} + \varrho b_{t-1} + \alpha d_t, \tag{13}$$

$$b_t = (1 - \varrho)b + b_{t-1} + \gamma d_t, \tag{14}$$

$$s_t^{(i)} = s_{t-s_i}^{(i)} + \delta_i d_t, \tag{15}$$

$$d_t = \sum_{i=1}^{p} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t. \tag{16}$$

In these equations, $y_t^{(\lambda)}$ indicates the value of the observed data after the Box-Cox transformation with the value λ, described in (4). l_t, b_t and $s_t^{(i)}$ are the values of the level, trend and seasonalities with smoothing parameters α, γ and δ_i. The subscript i denotes the seasonality under consideration, of seasonal length s_i, and n_s is the number of seasonalities. d_t is an ARMA (p, q) process with residuals whose coefficients are determined by φ_i and θ_i. ϱ is the damping factor for trend. The term b_t stands for a long–run trend term. ε_t is a Gaussian white noise process $N(0, \sigma^2)$. The nomenclature for BATS includes the following arguments $(\lambda, \varrho, p, q, s_1, \ldots, s_{n_s})$.

Additionally, De Livera et al. [18] presented the same model but with seasonality based on trigonometric models. The seasonality Equation (16) is replaced by the set of Equations (17)–(20), a seasonal component based on Fourier series. These are known as TBATS (Trigonometric seasonal BATS).

$$s_t^{(i)} = \sum_{j=1}^{(k_i)} s_{j,t}^{(i)}, \tag{17}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos(\omega_{j,i}) + s_{j,t-1}^{*(i)} \sin(\omega_{j,i}) + \delta_1^{(i)} d_t, \tag{18}$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin(\omega_{j,i}) + s_{j,t-1}^{*(i)} \cos(\omega_{j,i}) + \delta_2^{(i)} d_t, \tag{19}$$

$$\omega_{j,i} = 2\pi j / s_i. \tag{20}$$

Every seasonal component of the model $s_t^{(i)}$ results from the sum of the k_i stochastic levels $s_{j,t}^{(i)}$ of period i. $s_{j,t}^{*(i)}$ is the stochastic growth for each period. $\delta_1^{(i)}$ and $\delta_2^{(i)}$ are the smoothing parameters. The nomenclature for TBATS includes the following arguments $(\lambda, \varrho, p, q, \{s_1, k_1\}, \ldots, \{s_i, k_i\})$.

Obtaining the values of the previous matrices and vectors requires the application of an algorithm based on the Kalman filter and the maximum likelihood function using the sum of squared errors (SSE) as the minimization criterion. This algorithm carries a high computational load and manages to obtain these parameters iteratively. The reference [18] explains in detail the process to be carried out in order to use the BATS and TBATS methods.

2.2.5. Multiple Seasonal Holt-Winters Models with Discrete Interval Moving Seasonalities

nHWT models are robust to variations in the series, but sometimes special situations occur in which it is interesting to take these anomalies into account. One of the clearest examples is the influence of the calendar effect on electricity demand [65]. These anomalous and specific situations can sometimes be modeled as a discrete seasonality, if they follow a repetitive pattern. Despite being seasonal, since they are discrete, they have the particular quality that they are not located at fixed moments in the time series; therefore, they are not linked to a deterministic appearance, as would be the case for regular seasonality. These seasonalities are called discrete interval moving seasonality (DIMS).

Trull et al. [25] include the use of discrete seasonality in their model, so that the model seen in (6)–(9) now results in (21)–(25), which is named nHWT–DIMS:

$$l_t = \alpha \left(\frac{y_t}{\prod s_{t-s_i}^{(i)} \prod D_{t_h^* - s_m^*}^{(m)}} \right) + (1 - \alpha)\,(l_{t-1} + \varrho b_{t-1}), \tag{21}$$

$$b_t = \gamma(l_t - l_{t-1}) + (1 - \gamma)\varrho b_{t-1}, \tag{22}$$

$$s_t^{(i)} = \delta^{(i)} \left(\frac{y_t}{l_t \prod_{j \neq i} s_{t-s_j}^{(j)} \prod_m D_{t_h^* - s_m^*}^{(m)}} \right) + \left(1 - \delta^{(i)}\right) s_{t-s_i}^{(i)}, \tag{23}$$

$$D_{t_h^*}^{(h)} = \delta_D^{(h)} \left(\frac{y_t}{l_t \prod_j s_{t-s_j}^{(j)} \prod_{m \neq h} D_{t_h^* - s_m^*}^{(m)}} \right) + \left(1 - \delta_D^{(h)}\right) D_{t_h^* - s_h^*}^{(h)}, \tag{24}$$

$$\hat{y}_{t+k} = \left(l_t + \sum_{v=1}^{k} \varrho^v b_t \right) \prod_i s_{t-s_i+k}^{(i)} \prod_j D_{t_h^* - s_h^* + k}^{(h)} + \varphi_{AR}^k \varepsilon_t. \tag{25}$$

Here the term $D_{t_h^*}^{(h)}$ is included, which represents the discrete seasonal indices, for each DIMS (h) considered up to n_{DIMS}. DIMS are only defined in the time intervals in which the special event takes place. These time intervals are designated using t_h^* for each DIMS (h). This nomenclature is chosen in order to distinguish this from the continuous time interval t.

The great difference between this model and other methods of modeling special situations is that the effect produced by the anomaly in the series is modeled as an internal part of the model, as one more seasonality, and is smoothed with each new appearance, unlike the use of models with dummy variables and/or modifications of the original series.

In the nHWT models, the seasonality equation shows a fixed recurrence for each seasonal pattern (s_i) being considered. With DIMS, this is not possible, since the occurrences of special events are not subjected to a deterministic pattern in the series. Therefore, the use of the variable s_h^* indicates, for each DIMS and each occurrence, which is the recurrence to consider.

One possible situation with special events is the simultaneous occurrence of two events. In such a case, the forecaster should consider the option of using only one of the DIMS that occur at that time, or using both, if the effects produced by the two special events add up.

An important aspect to consider is the initialization of the DIMS. A DIMS may have few occurrences in the time series and, therefore, its seasonal indexes must be calculated in such a way that it converges rapidly to the desired effect.

The initialization method consists in first obtaining the initial values of the level, the trend, and the seasonal indices for the regular seasonality. Subsequently, a decomposition of the series is carried out using trend and multiple seasonality. It is common to use the multiple STL method (Seasonal–Trend decomposition procedure using Loess [66], where Loess is a method to estimate linear relationships).

From the decomposition, the series can be reconstructed without including the irregular part, which is where the information necessary to obtain the desired indices is found. The initial values are obtained by weighting the time series against the reconstructed series.

The adjustment of the parameters is carried out following the same procedure as for the nHWT, with the exception that, if necessary, this adjustment can be carried out in two steps—first adjusting the parameters of the regular model and then adjusting the parameters associated with the DIMS. Adjusting all the parameters simultaneously obtains models with more reliable predictions, while the second option is faster. Thus, the first option is chosen for this work.

3. Results

The approach proposed for the work described below has the following scheme. First, a study of the series is carried out to determine the seasonal periods. The study is carried out using continuous seasonality models and discrete seasonality models, all as described in the previous section. Although it is preferable to use an error minimization criterion for

each technique when fitting the models, the RMSE—defined in (26)—is used to standardize and compare the fitted results.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2} \, . \tag{26}$$

Here N stands for length of the dataset used for the training. The final comparison will be made according to the forecasts made in the validation set.

3.1. Analysis of the Seasonality of the Series

The series shown in Figure 3 clearly presents a seasonality with periodicity of one hour. However, to study the following seasonal patterns it is necessary to perform an analysis on the frequency domain. To investigate the appreciable frequencies in the time series, a spectral density analysis is carried out, the result of which is shown in Figure 7 in the form of a smoothed periodogram. A smoothed periodogram is the preferred tool here as the periodic cycles do not show a regular periodicity [67].
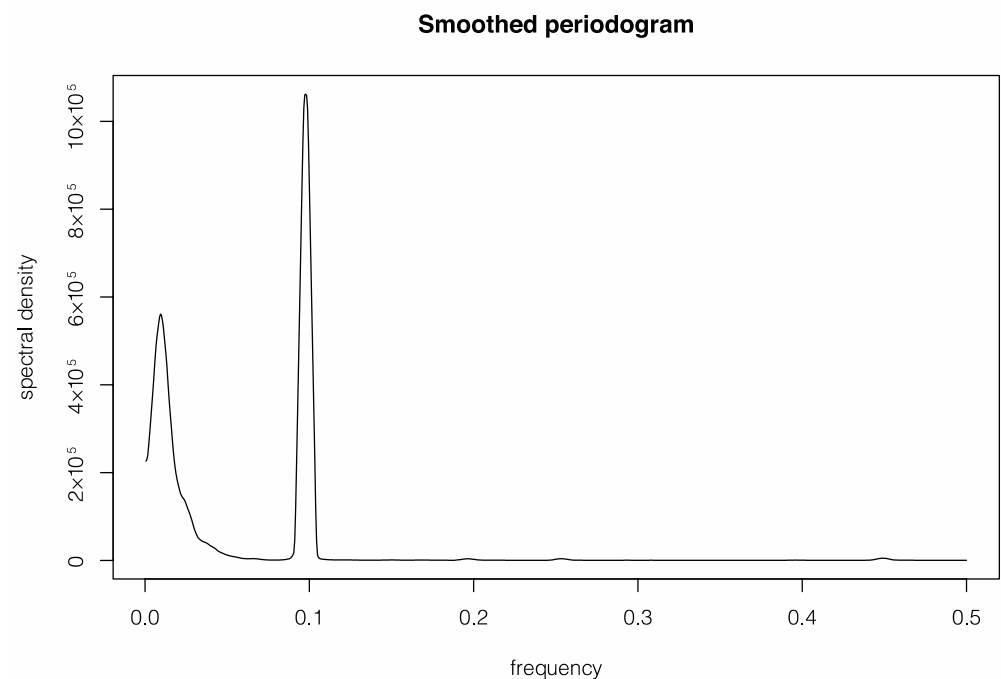
**Figure 7.** Smoothed periodogram obtained from the time series shown in Figure 3.

Analyzing the figure, the presence of a clearly dominant frequency is observed, which corresponds to the periodicity of ten units of time (one hour). Also, the presence of another dominant frequency can be observed. This corresponds to a second seasonality with a period of 106 time-units. However, this is the second seasonality and is associated with a greater variability around its value, which confirms what is seen in Figure 3.

To confirm these hypotheses, an ALLSSA analysis is performed. This method is robust against unequally spaced time series, estimating trend and seasonal components simultaneously, and providing statistically significant components in the time series [35]. The analysis shows three main and significant frequencies at periodicities of 10, 118, and 203 time-units. This disagreement between the two methods suggests that, despite various seasonalities clearly coexisting, non-dominant seasonalities do not occur continuously and may influence the analysis.

In contrast to this result, an analysis based on the use of wavelets is also carried out. The advantage of using wavelets to analyze the spectral content of the series is that we

obtain a map in the time-scale plane. The concept of frequency in spectral analysis is now replaced by the scale factor, and therefore, instead of using a periodogram, we use a scalogram. The scale measures the stretching of the wavelets, being directly related to frequency, as the greater the scale is, the higher the frequency of the series, which is related to the inverse of a frequency, that is, to a period [68]. This map allows the non–stationary characteristics of the signal, including changes in periodicity, to be studied, which is the objective.

Figure 8 shows the average wavelet power graph. This graph shows the means of the powers developed over time for each period or frequency. Although the results are similar to those shown in the previous graph, a greater variability is observed in the longer periods. Three main periods are located at 10, 94, and 196 time units. The results are very close to the previous one.
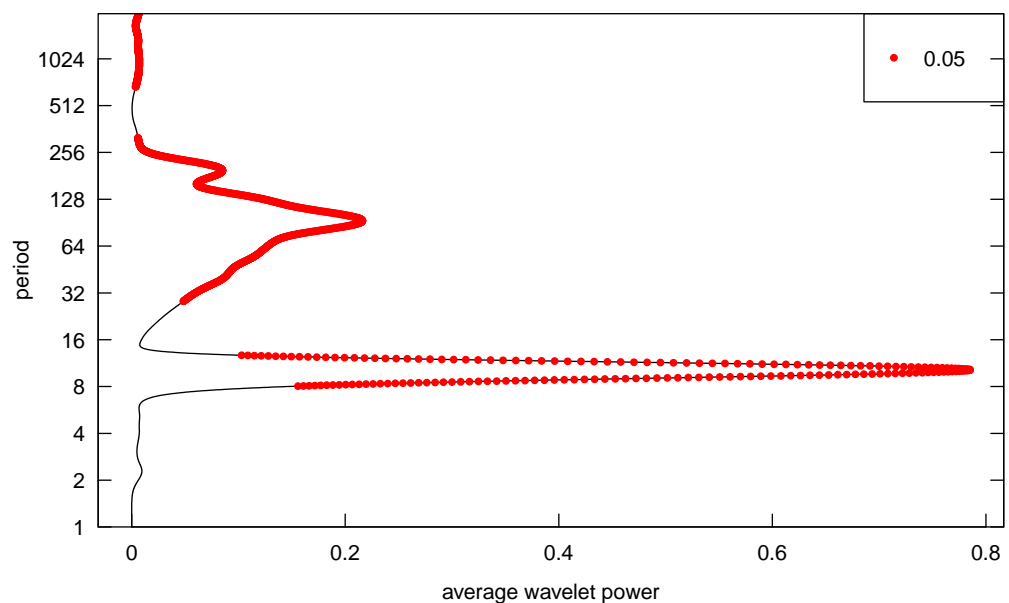

Figure 8. Plot of wavelet power averages across time. The red bullets show the significance level (0.05).

The need for a robust analysis using the time and frequency domain motivates the use of LSWA [35,69]. The software LSWAVE [70] in MATLAB® is an easy and intuitive tool for performing this analysis. This software computes the least square wavelet spectrum (LSWS) for the series, with no need for preprocessing, transforming, or detrending. LSWA considers the correlations of the sinusoidal functions and constituents and the noise at the same time. We apply LSWA to the training set, with the results shown in Figure 9.

The abscissa axis indicates the time scale used, while the ordinate axis shows the cyclical frequencies (as 1/period). The level of variance explained is reflected by colors, according to the scale to the right of the graph.

The first conclusion is clear from the graph: the one–hour seasonality remains practically throughout the series as the predominant seasonality (with a percentage of variance greater than 90%), but discontinuously. In the sections where this does not occur, a series of sawtooth-shaped formations stand out from the rest, although the percentage of variance that it reflects does not exceed 30%. Some areas are shaded with more than 40% of the variation within high frequencies areas. This graph is shown in closer detail of in Figure 10.
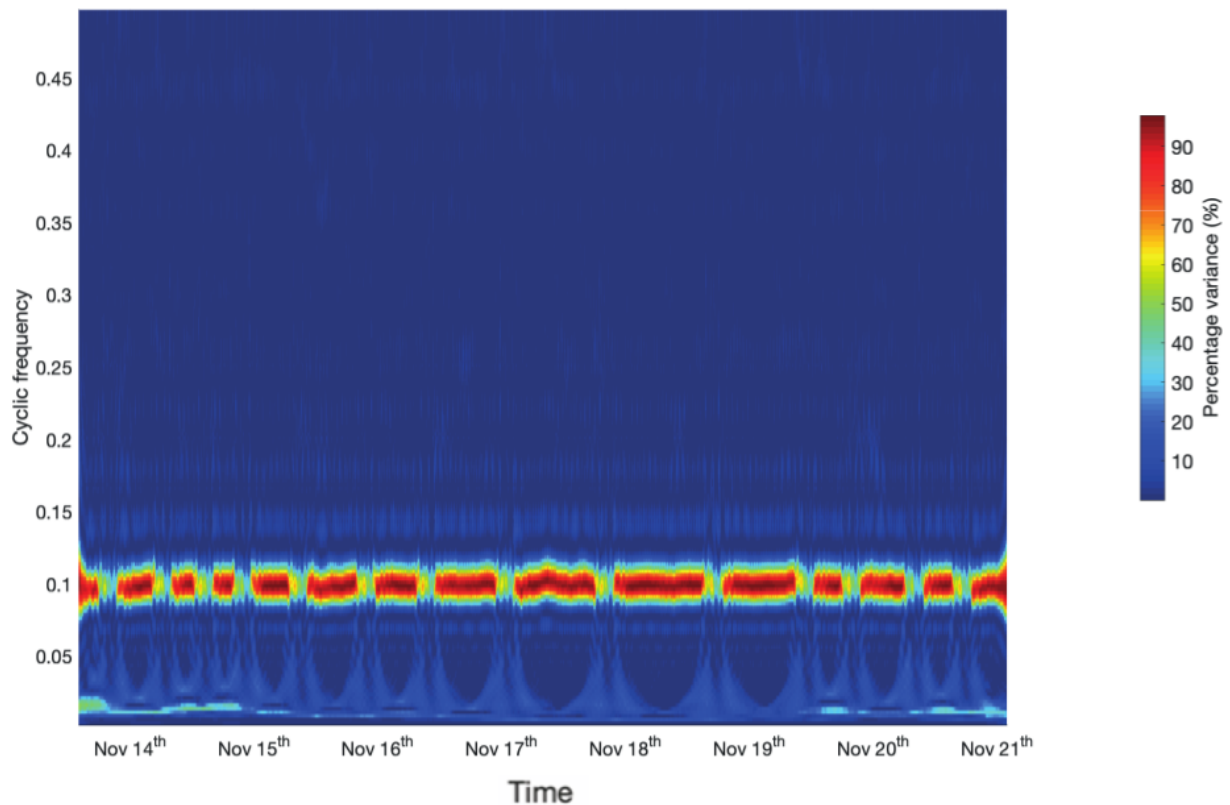
Figure 9. Least-squares wavelet analysis (LSWA) applied to the training set of electricity consumption.
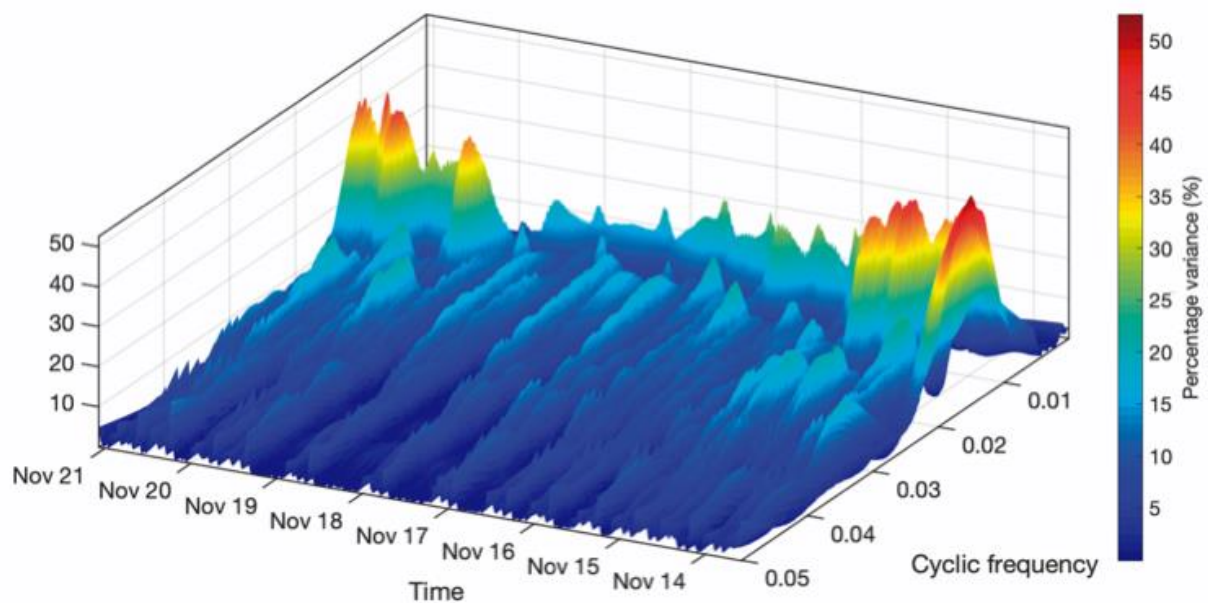
Figure 10. Detail in 3D for the lowest cyclic frequencies in the LSWA analysis.

We decided to use a 3D representation because it is then easier to appreciate the lowest cyclic frequencies. Between 14 November and 15 November and later between 19 November and 21 November, two frequencies with a percentage of variance of over 40% appear. This corresponds to a period of 100 time-units. In the middle, between the two intervals, some peaks with 30% of the variance are also located. This corresponds to a periodicity of 200-time units.

The conclusion from this analysis is that there is clearly one seasonality that occurs every hour (ten-time units), and a second pattern with unregular behavior over time and a periodic length that has been established at between 94 and 118 units. Although it is not strictly a seasonality, it can be modeled as a second seasonality. A marginal seasonality can be obtained for long cycles but will not be taken into account as it seems that its influence on the time series is very small compared to the previous one.

3.2. Application of Models with Regular Seasonality

Given this disparity of values for the determination of the length of the seasonal periods, we choose to carry out one analysis with the models using a single seasonality (ten-time units) and another using two seasonalities.

For the models with regular seasonality, the second seasonality to be tested will be for a range of periods of between 90 and 110 time-units.

3.2.1. Application of ANN

One of the most powerful tools for working with neural networks is the MATLAB® Deep Machine Learning toolbox. This toolbox includes a great variety of possibilities and different neural networks. From this range of possibilities, the NARX network and the NLR network are used. These two networks have been proven to be efficient in predicting future electricity demand values. The method of working with them is described in Figure 11. Here, it can be seen that it is first necessary to use the historical information about demand.
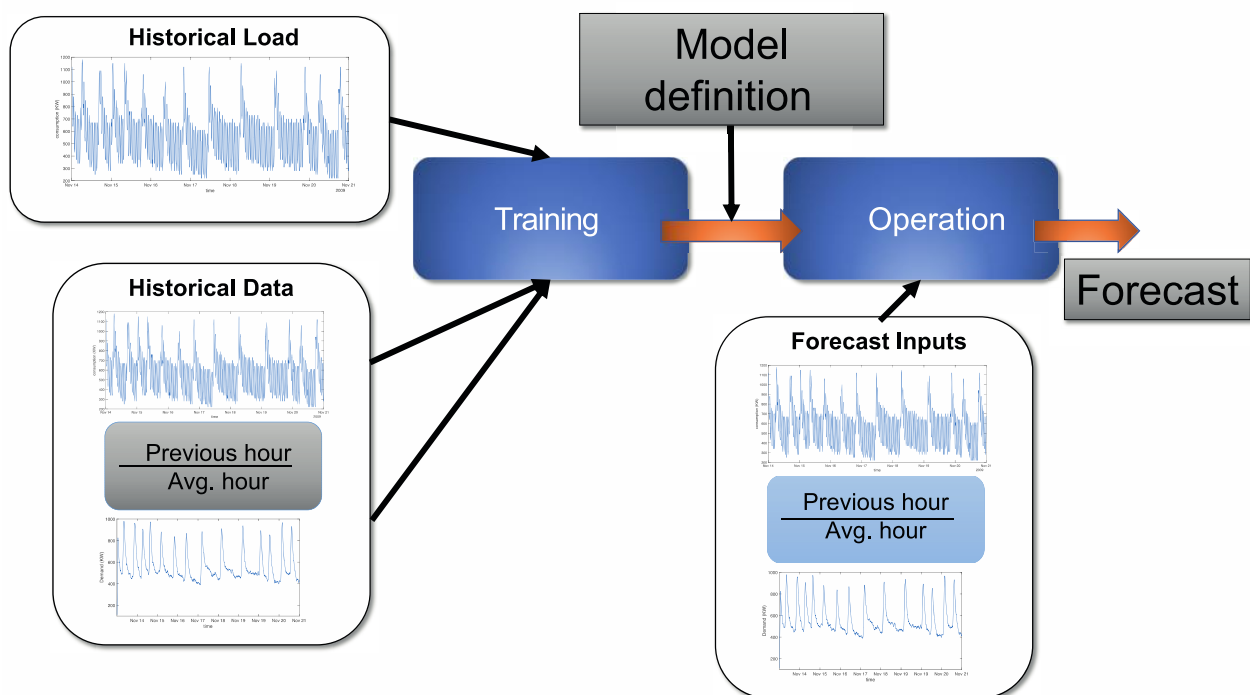
Figure 11. Working scheme with neural networks using the Deep learning machine tool from MATLAB.

To address the observed seasonality, the series is additionally given a delay, according to the seasonality. In this case, the seasonality of one hour corresponds to ten units of time of six minutes, so a delay of ten units is introduced to the series. Additional variables are added to the information provided by the time series. The exogenous information supplied to the model is:

- Previous hour's average electricity consumption.
- Consumption of electricity from the previous hour.
- Timestamp (only in NARX model).

The networks used are described in Table 1. The NARX model includes a single hidden layer and a TDL of three time units. It was checked that greater TDL did not improve the results. NLR model also include a single hidden layer.

Table 1. Neural network parameters and RMSE for fitted results.

Neural Network	Parameters	Fit RMSE
NARX	1 input layer 1 hidden layer with 20 neurons 1 output layer TDL = 3	33.95
NLR	1 input layer 1 hidden layer with 20 neurons 1 output layer	33.21

The training process is performed by minimizing the MSE and using the Levenberg-Marquardt algorithm. The training result is displayed as the RMSE in Table 1.

In the same way as the first seasonality was introduced, the second seasonality is added, following what we have seen in Section 3.2. The result of adding a new seasonality does not improve the result. The chosen model has only one seasonality.

3.2.2. Application of ARIMA Models

To apply the ARIMA models, MATLAB® is the chosen platform, using Econometrics Toolbox and SSMMATLAB [71] for double seasonal ARIMA. R is also tested with the 'forecast' package, but the MATLAB® results outperformed the R results. Like the previous models, models with one and two seasonalities according to Section 3.2 are tested. The best results are found using a single seasonality of one hour (ten-time units). The best model is ARIMA $(4,2,4) \times (4,1,4)_{10}$, for which the parameters are shown in Table 2.

Table 2. ARIMA parameters and RMSE for fitted results.

Parameters		Fit RMSE
AR	$\phi_1 = -0.145$; $\phi_2 = -0.472$; $\phi_3 = -0.190$; $\phi_p = 0.170$	50.42
MA	$\theta_1 = 0.861$; $\theta_2 = -0.601$; $\theta_3 = 0.443$; $\theta_4 = 0.174$	
SAR	$\Phi_1 = -0.487$; $\Phi_2 = 0.067$; $\Phi_3 = -0.260$; $\Phi_4 = -0.067$	
SMA	$\Theta_1 = 0.142$; $\Theta_2 = 0.069$; $\Theta_3 = -0.292$; $\Theta_4 = -0.065$	

3.2.3. Application of nHWT Models

To perform the analysis using the nHWT models, a proprietary tool developed in MATLAB® is used. This tool comes in the form of a toolbox, but it has not been published yet. Models with one and two seasonalities are tested.

The results show that the model that best adapts to this series presents a single seasonality, with the parameters $\alpha = 0.0001$, $\gamma = 0.0001$, $\delta_{10} = 0.4856$ and $\varphi_{AR} = 0.9286$. The damping factor ϱ is set to 0. Models including this parameter were tested, but results were not improved, thus it was removed from the model. The RMSE of the fit process is 54.93.

The result is not surprising since the nHWT models are structural and do not allow for a relaxation of seasonality. Once seasonality is established, the model will continue to develop the same pattern, even though this is not really reflected in the series. When using a single seasonality, the information provided by the second seasonal pattern is lost, but it

has less influence than the error caused by a second seasonality that does not coincide with the real seasonality of the series.

3.2.4. Application of SSM Models

To work with the state spaces, the 'forecast' library is used in R [72]. Models with a single seasonality are tested, as well as models that include several seasonalities as indicated in Section 3.1. Here again, the use of the trend damping parameter did not provide better results and was removed from the model.

As in the previous cases, the models with several seasonalities do not show better results than the models with a single seasonality. Table 3 shows the models used—including their arguments—in the first column, the parameters obtained after the adjustment process in the second column, and the RMSE value of the adjustment in the third column.

Table 3. SSM models, their parameters, and fit RMSE for fitted results.

Model Arguments	Parameters	Fit RMSE
BATS (0.717,0, 5,2, 10)	$\lambda = 0.717, \alpha = 0.120, \gamma = 0.206,$ $\delta = -0.004,$ $\phi_1 = 0.945, \phi_2 = -0.625, \phi_3 = 0.022,$ $\phi_4 = 0.104, \phi_5 = -0.500,$ $\theta_1 = -0.153, \theta_2 = 0.515.$	46.77
TBATS (0.756,0, 5,2, {10,1})	$\lambda = 0.756, \alpha = 0.862, \gamma = 0.105, \delta_1 = -0.0004,$ $\delta_2 = 0.0001,$ $\phi_1 = 1.293, \phi_2 = -0.552, \phi_3 = -0.359, \phi_4 = 0.227,$ $\phi_5 = -0.1968,$ $\theta_1 = -1.358, \theta_2 = 0.783.$	45.44

3.3. Application of Discrete Seasonality Models (nHWT-DIMS)

The application of discrete seasonality carries with it a differentiated strategy. The use of nHWT-DIMS models makes it possible to model localized seasonality at specific instants of time, independently of other seasonality. In Figure 12, we show two different periods for the series. In addition to the seasonal pattern described at the beginning (of one hour), a new pattern can also be observed in Figure 12a, whose length is established at 27 time units (2 h and 42 min). This pattern is framed between two dashed lines including the demand peaks. Figure 12b shows another seasonal pattern that has the same length, but a different behavior. These two patterns will be called DIMS a and DIMS b.

The appearance of each discrete seasonality does not occur on a regular basis. This situation causes the recursion required in the Holt-Winters models to be variable. This is indicated in Figure 12 by the lines with arrows. The solid lines indicate the recursion for the DIMS a, and the dashed lines indicate it for the DIMS b. The information regarding the DIMS is organized in Table 4. This table includes the locations of the discrete seasonalities on every appearance (starting and ending time when the DIMS is defined, used in the variable t_h^*) and the associated recursivity in minutes, which corresponds to s_h^*. As an example, the time interval when the second appearance of DIMS a is defined starts at 04:00 pm on the 14th and ends at 06:42 pm on the 14th. The recursivity s_h^* during this interval is 618 min.
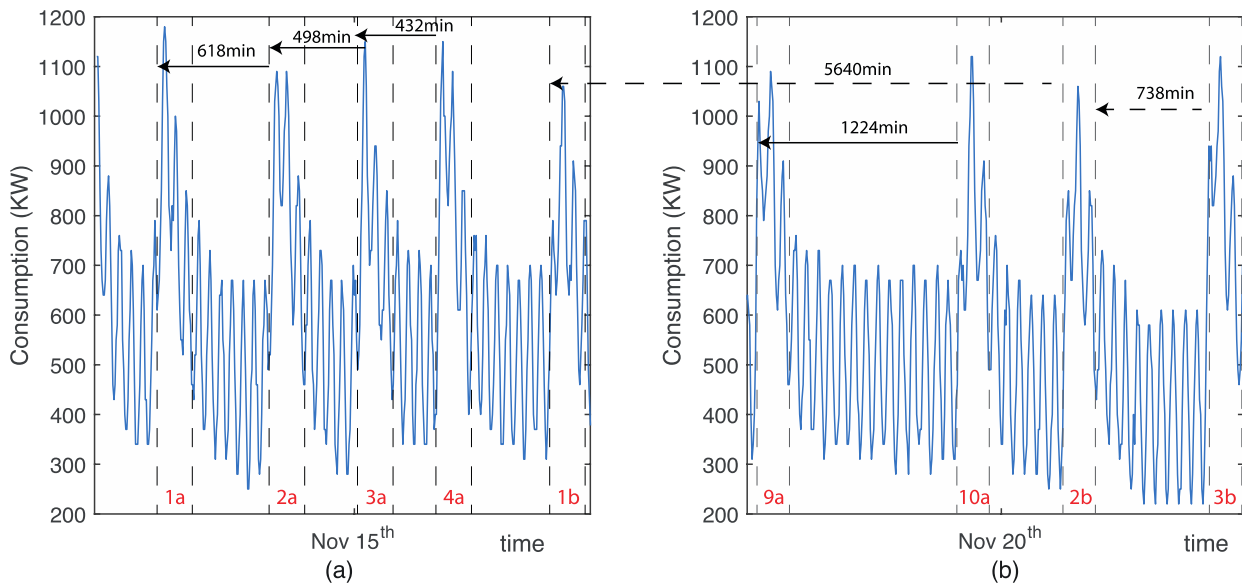
Figure 12. Discrete interval moving seasonality (DIMS) locations and recursion design. Two additional seasonal patterns are located, the first mostly appears in (**a**) while the second pattern appears in (**b**). Vertical dashed lines delimitate the DIMS period length. The appearance number of each DIMS is numbered at the bottom of the figure in red. Lines with arrows represent the recursivity of the DIMS, with full lines for the DIMS in (**a**) and dashed lines for the DIMS in (**b**). The time span of the previous appearance is shown in minutes over the line.

Table 4. Location in the time series of the discrete seasonalities (DIMS) and their recursivity. The column Nr. indicates the order of appearance of the corresponding DIMS. 'Time starts' and 'Time ends' reflect the moving interval in which DIMS is defined, and 'Recursivity' shows the length of time since the previous appearance.

DIMS	Nr.	Time Starts	Time Ends	Recursivity
DIMS a	1	14th November at 05:42 am	14th November at 08:24 am	————
	2	14th at 04:00 pm	14th at 06:42 pm	618 min.
	3	15th at 00:18 am	15th at 03:00 am	498 min
	4	15th at 07:30 am	15th at 10:12 am	432 min
	5	15th at 06:42 pm	15th at 09:24 pm	672 min
	6	16th at 07:24 pm	16th at 10:06 pm	1482 min
	7	17th at 10:42 am	17th at 01:24 pm	918 min
	8	18th at 06:06 am	18th at 08:48 am	1164 min
	9	19th at 02:30 am	19th at 05:12 am	1224 min
	10	19th at 08:48 pm	19th at 23:30 pm	1098 min
	11	21th at 06:06 pm	21th at 20:48 pm	2718 min
DIMS b	1	16th November at 07:06 am	16th November at 09:48 am	————
	2	20th at 05:06 am	20th at 07:48 am	5640 min
	3	20th at 05:24 pm	20th at 08:06 pm	738 min
	4	21th at 02:36 am	21th at 05:18 am	552 min
	5	22th at 02:24 am	22th at 05:06 am	1428 min

The general model described by Equations (21)–(25) now results in the Equations shown in (27)–(32), with one seasonality of length ten time units and two DIMS as described in Table 4.

$$l_t = \alpha \left(\frac{y_t}{I_{t-s_{10}}^{(10)} D_{t_a^*-s_a^*}^{(a)} D_{t_b^*-s_b^*}^{(b)}} \right) + (1-\alpha)\,(l_{t-1} + b_{t-1}), \tag{27}$$

$$b_t = \gamma(l_t - l_{t-1}) + (1-\gamma)b_{t-1}, \tag{28}$$

$$s_t^{(10)} = \delta^{(10)} \left(\frac{y_t}{l_t D_{t_a^*-s_a^*}^{(a)} D_{t_b^*-s_b^*}^{(b)}} \right) + \left(1 - \delta^{(10)}\right) s_{t-10}^{(10)}, \tag{29}$$

$$D_{t_a^*}^{(a)} = \delta_D^{(a)} \left(\frac{y_t}{l_t s_{t-s_{10}}^{(10)} D_{t_b^*-s_b^*}^{(b)}} \right) + \left(1 - \delta_D^{(a)}\right) D_{t_a^*-s_a^*}^{(a)}, \tag{30}$$

$$D_{t_b^*}^{(b)} = \delta_D^{(b)} \left(\frac{y_t}{l_t s_{t-s_{10}}^{(10)} D_{t_a^*-s_a^*}^{(a)}} \right) + \left(1 - \delta_D^{(b)}\right) D_{t_b^*-s_b^*}^{(b)}, \tag{31}$$

$$\hat{y}_t(k) = (l_t + kb_t)s_{t-s_{10}+k}^{(10)} D_{t_a^*-s_a^*+k}^{(a)} D_{t_b^*-s_b^*+k}^{(b)} + \varphi_{AR}^k \varepsilon_t. \tag{32}$$

Here, $s_t^{(10)}$ is the seasonal equation for the regular seasonality of ten-time units with smoothing parameter $\delta^{(10)}$. $D_{t_a^*}^{(a)}$ and $D_{t_b^*}^{(b)}$ are the DIMS as described in Table 4 with smoothing parameters $\delta_D^{(a)}$ and $\delta_D^{(b)}$ defined only in time t_a^* and t_b^*. The recursivity s_a^* and s_b^* is defined in Table 4.

To use the model, the procedure described in [25] is carried out. Initially, the initial values for the level are obtained as the moving average of the first period of one hour; for the trend as the slope between the first and second cycle of one hour; and for the seasonal indices, the weighting of the series in the first cycle on the moving average. Subsequently, the seasonal indices of the DIMS are obtained. The time series is decomposed into its trend, seasonality, and irregular components using STL decomposition with the period length of one hour. From these components, the series is rebuilt, but without the irregular component being included. The seasonal indices are obtained by weighting the original series over the reconstructed one.

Once the initial values of the model have been determined, the model is fitted by minimizing the RMSE, and the smoothing parameters are obtained. The tool for this analysis is software developed in MATLAB (R) for this purpose. The obtained RMSE is 58.65. The smoothing parameters of the model obtained are $\alpha = 0.0001$, $\gamma = 0.0001$, $\delta^{(10)} = 0.4853$ for the first regular seasonality, $\delta^{(a)} = 0.0005$ for DIMS type a (see Figure 12a), $\delta^{(b)} = 0.0652$ for DIMS type b (see Figure 12b) and $\varphi_{AR} = 0.9056$. Here again, it has is decided not to use the damping parameter for trend. The RMSE of the fitted model is 58.65.

3.4. Model Fit Comparison

A benchmark summary is reported in Table 5, where the RMSE in the fit process is summarized. The RMSE used to compare the models while fitting shows that the ANN fits better than the other models to the time series. The worst case seems to be nHWT-DIMS. The comparison shows that the state space models and the ARIMA models fit the observed data better than the nHWT and nHWT–DIMS models. Similar behavior is expected in the forecasts.

Table 5. Main benchmarking results. RMSE used to compare fitted results. MAPE used to compare forecasts accuracy.

	RMSE on Fit	Average MAPE for Forecasts
NARX	33.95	26.63%
NN-NLR	33.21	13.94%
ARIMA	50.42	24.03%
nHWT	54.93	18.55%
TBATS	46.77	37.60%
BATS	45.44	37.61%
nHWT-DIMS	58.65	16.00%

3.5. Forecasts Comparison

The greatest interest is in forecast reliability. To compare the results of the forecasts given by the different methods, the mean absolute percentage error (MAPE) as a percentage is used, as indicated in (33). This is a common indicator used to compare forecasts of demand [73].

$$\text{MAPE}(h) = \frac{1}{h} \sum_{t=1}^{h} \left| \frac{y_t - \hat{y}_t}{y_t} \right|. \tag{33}$$

Here, h is the forecast horizon to evaluate. As the forecasts are made one hour ahead (ten units) throughout the validation subset, h can take values from one to ten time units. From the forecasts of one hour ahead, the MAPE is obtained by comparing these with the real values of the validation set, using the procedure described in [74]. The benchmark summary in Table 5 includes the average of the MAPE. The average is obtained as the mean of the MAPE(h) with h = 1,2, . . . ,10. The best forecasts, on average, are produced by the NLR. The nHWT-DIMS models are revealed as a competitive method against the regular seasonal models, outperforming the other models.

Figure 13 shows the MAPE of these forecasts as a function of the forecasting horizon. It is clear from the results obtained that traditional models with one seasonality are not capable of working with this type of series. The BATS and TBATS models of state spaces do not drop below 30% MAPE. The ARIMA model starts by making very short-term forecasts that have MAPE of below 15%, but beyond three time units it is not capable of making good predictions. The nHWT models improve the forecasts with respect to the previous ones, although the use of the DIMS allows the level of the predictions to be always kept below 20%. However, the method that produces the best results is NN–NLR. These models give forecasts that remain almost constant with an accuracy of about 14% of MAPE.
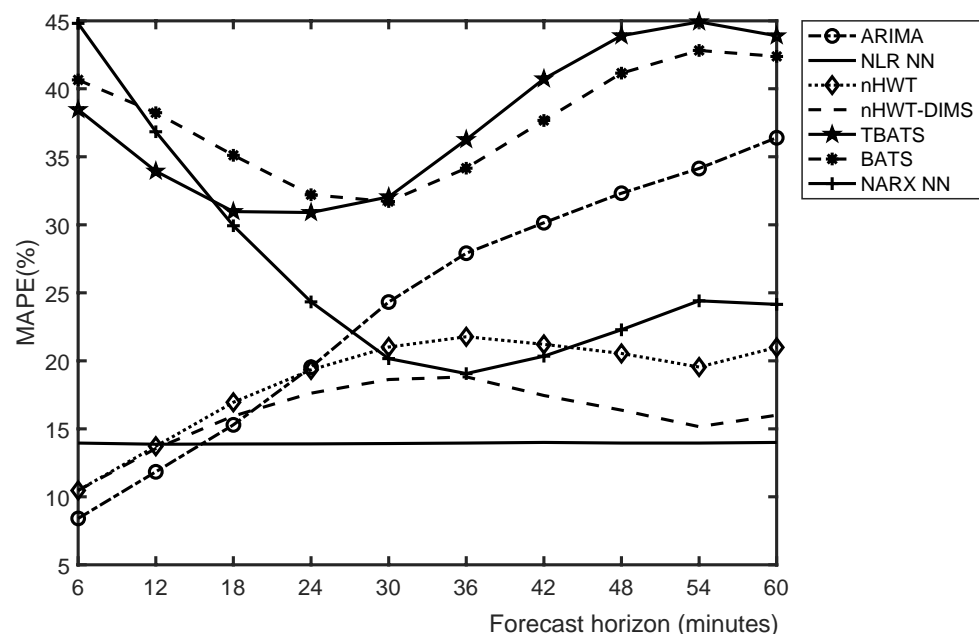

Figure 13. Mean absolute percentage error (MAPE) comparison of one hour-ahead forecasts.

4. Discussion

The results obtained in the previous exercise show that the fact that having irregularities in the series has an enormous influence on the result in statistical models used in this article. The models that use regular seasonalities require that they appear with a regular length, regardless of whether the pattern varies. When dealing with series whose seasonality is not completely defined, the models cannot overcome these variations.

The use of models with discrete seasonality allows for progress in this problem, since it is capable of introducing seasonality only where it occurs.

Though the periodicity of 200-time units did not show a consistent pattern over time in this data set, having a longer time series more than seven days (e.g., two-month record or more) may reveal discontinuous patterns repeating themselves at that low frequency, which may help to better train the model for forecasting such signals. This requires further investigation and research.

However, the best tool for this series is the use of AI to address these irregularities. Curiously, the NARX neural network does not offer good results, but the NLR neural network manages to improve the results. This situation responds to the fact that the previously described models require seasonal stability if they are to make predictions, since they are based on linear or structural models. The neural network model is not subject to these restrictions and uses these irregularities to make forecasts.

Future studies in this area should aim to ensure that the structural models are capable of introducing an ambiguity between their seasonal processes produced by the inconsistency of the series in terms of seasonality.

5. Conclusions

In this article, we have analyzed time series forecasting methods applied to a pattern of electricity demand that has an irregular periodicity, so that the seasonality is not well defined. We have analyzed models of neural networks, ARIMA, multiple seasonal Holt-Winters models and state spaces using regular seasonalities, and multiple seasonal Holt-Winters models with discrete interval moving seasonalities.

To compare the behavior of all the models discussed, they were applied to the situation of a connection node with a hot-dip galvanizing company, where the time series of electricity consumption due to the heating of the bath causes seasonalities. A frequency analysis using spectral density and least square wavelets with the series showed that a first seasonality of one hour could be easily located; some other seasonalities could be considered, but their period was not clear. The problem with irregular seasonality is that the models need to use patterns that constantly repeat themselves, so the pattern must be defined for the entire time series. Nevertheless, the use of Holt-Winters models with discrete seasonality (nHWT-DIMS) allows these seasonalities to be separated efficiently and reliable predictions to be made.

The results showed that the use of nHWT–DIMS models improves the results compared to the rest of the models. This is an interesting proposal for companies because of the simplicity of its application and good results—the MAPE obtained is around 16%. However, NLR (ANN) showed better predictions, with a MAPE of 14%.

Our study contributes to the improvement of forecasting systems with time series by including discrete seasonality in the model. This allows for an efficient method of prediction to be applied in situations of electrical demand with marked seasonality but non–regular periodic appearances.

Author Contributions: Conceptualization, O.T. and J.C.G.-D.; methodology, O.T. and J.C.G.-D.; software, O.T.; validation, J.C.G.-D. and A.P.-S.; formal analysis, A.P.-S.; investigation, O.T.; data curation, J.C.G.-D.; writing—original draft preparation, O.T.; writing—review and editing, J.C.G.-D. and A.P.-S.; supervision, J.C.G.-D. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

AI	Artificial intelligence
Al	Aluminum
AIC, AICc	Akaike's information criterion
ALLSSA	Anti-leakage least-squares spectral analysis
ANN	Artificial neural networks
AR(1)	Auto regressive model of order 1
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
BATS	Exponential smoothing state space model with Box-Cox transformation, ARMA errors, trend and seasonal components
BIC	Bayesian information criterion
BRT	Bagged regression trees
DIMS	Discrete interval moving seasonalities
LASSO	Least absolute shrinkage and selection operator
LSWA	Least-squares wavelet analysis
LSWS	Least square wavelet spectrum
MSE	Mean squared error
MAPE	Mean absolut percentage error
MLP	Multilayer perceptron
NARX	Non-linear autoregressive neural networks with exogenous variables
nHWT	Multiple seasonal Holt-Winters
nHWT-DIMS	Multiple seasonal Holt-Winters with discrete interval moving seasonalities
NLR	Non-linear regression
RMSE	Root of mean squared error
SARIMAX	Seasonal autoregressive integrated moving average exogenous model
SIC	Schwarz's information criterion
SSM	State-space models
STL	Seasonal–trend decomposition procedure using Loess
SVM	Support vector machines
TBATS	Exponential smoothing state space model with Box-Cox transformation, ARMA errors, trend and trigonometric seasonal components
TDL	Tapped delay line
Zn	Zinc

References

1. Weron, R. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*; John Wiley & Sons, Ltd.: Chichester, UK, 2006; ISBN 978-0-470-05753-7.
2. Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M. (Eds.) *Machine Learning. An Artificial Intelligence Approach*; Morgan Kaufmann: San Francisco, CA, USA, 1983; ISBN 978-0-08-051054-5.
3. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799. [CrossRef]
4. Fallah, S.; Ganjkhani, M.; Shamshirband, S.; Chau, K.; Fallah, S.N.; Ganjkhani, M.; Shamshirband, S.; Chau, K. Computational Intelligence on Short–Term Load Forecasting: A Methodological Overview. *Energies* **2019**, *12*, 393. [CrossRef]
5. Buitrago, J.; Asfour, S. Short–term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs. *Energies* **2017**, *10*, 40. [CrossRef]
6. López, M.; Valero, S.; Senabre, C. Short–term load forecasting of multiregion systems using mixed effects models. In Proceedings of the 2017 14th International Conference on the European Energy Market (EEM), Dresden, Germany, 6–9 June 2017; pp. 1–5.
7. Ahmad, W.; Ayub, N.; Ali, T.; Irfan, M.; Awais, M.; Shiraz, M.; Glowacz, A. Towards short term electricity load forecasting using improved support vector machine and extreme learning machine. *Energies* **2020**, *13*, 2907. [CrossRef]
8. Khan, A.R.; Razzaq, S.; Alquthami, T.; Moghal, M.R.; Amin, A.; Mahmood, A. Day ahead load forecasting for IESCO using Artificial Neural Network and Bagged Regression Tree. In Proceedings of the 2018 1st International Conference on Power, Energy and Smart Grid (ICPESG), Mirpur, Pakistan, 12–13 April 2018; pp. 1–6.
9. Zahedi, G.; Azizi, S.; Bahadori, A.; Elkamel, A.; Wan Alwi, S.R. Electricity demand estimation using an adaptive neuro–fuzzy network: A case study from the Ontario province—Canada. *Energy* **2013**, *49*, 323–328. [CrossRef]
10. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2015; ISBN 978-1-118-67502-01.
11. Makridakis, S.; Hibon, M. ARMA models and the Box–Jenkins methodology. *J. Forecast.* **1997**, *16*, 147–163. [CrossRef]

12. Cancelo, J.R.; Espasa, A.; Grafe, R. Forecasting the electricity load from one day to one week ahead for the Spanish system operator. *Int. J. Forecast.* **2008**, *24*, 588–602. [CrossRef]

13. Elamin, N.; Fukushige, M. Modeling and forecasting hourly electricity demand by SARIMAX with interactions. *Energy* **2018**, *165*, 257–268. [CrossRef]

14. Bercu, S.; Proïa, F. A SARIMAX coupled modelling applied to individual load curves intraday forecasting. *J. Appl. Stat.* **2013**, *40*, 1333–1348. [CrossRef]

15. Taylor, J.W.; de Menezes, L.M.; McSharry, P.E. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *Int. J. Forecast.* **2006**, *22*, 1–16. [CrossRef]

16. Hyndman, R.J.; Koehler, A.B.; Ord, J.K.; Snyder, R.D. *Forecasting with Exponential Smoothing: The State Space Approach*; Springer: Berlin/Heidelberg, Germany, 2008; ISBN 978-3-540-71916-8.

17. Welch, G.; Bishop, G. An Introduction to the Kalman Filter. *Proc. Siggraph Course* **2006**, *7*, 1–16.

18. De Livera, A.M.; Hyndman, R.J.; Snyder, R.D. Forecasting time series with complex seasonal patterns using exponential smoothing. *J. Am. Stat. Assoc.* **2011**, *106*, 1513–1527. [CrossRef]

19. Bermúdez, J.D. Exponential smoothing with covariates applied to electricity demand forecast. *Eur. J. Ind. Eng.* **2013**, *7*, 333–349. [CrossRef]

20. Winters, P.R. Forecasting sales by exponentially weighted moving averages. *Management* **1960**, *6*, 324–342.

21. Taylor, J.W.; McSharry, P.E. Short–term load forecasting methods: An evaluation based on European data. *Power Syst. IEEE Trans.* **2007**, *22*, 2213–2219. [CrossRef]

22. Taylor, J.W. Short–term electricity demand forecasting using double seasonal exponential smoothing. *J. Oper. Res. Soc.* **2003**, *54*, 799–805. [CrossRef]

23. Taylor, J.W. Triple seasonal methods for short–term electricity demand forecasting. *Eur. J. Oper. Res.* **2010**, *204*, 139–152. [CrossRef]

24. García–Díaz, J.C.; Trull, O. Competitive Models for the Spanish Short–Term Electricity Demand Forecasting. In *Time Series Analysis and Forecasting: Selected Contributions from the ITISE Conference*; Rojas, I., Pomares, H., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 217–231. ISBN 978-3-319-28725-6.

25. Trull, O.; García–Díaz, J.C.; Troncoso, A. Application of Discrete–Interval Moving Seasonalities to Spanish Electricity Demand Forecasting during Easter. *Energies* **2019**, *12*, 1083. [CrossRef]

26. Hong, T.; Xie, J.; Black, J. Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting. *Int. J. Forecast.* **2019**, *35*, 1389–1399. [CrossRef]

27. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: Results, findings, conclusion and way forward. *Int. J. Forecast.* **2018**, *34*, 802–808. [CrossRef]

28. Smyl, S. A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. *Int. J. Forecast.* **2020**, *36*, 75–85. [CrossRef]

29. Choi, T.-M.; Yu, Y.; Au, K.-F. A hybrid SARIMA wavelet transform method for sales forecasting. *Decis. Support Syst.* **2011**, *51*, 130–140. [CrossRef]

30. Sudheer, G.; Suseelatha, A. Short term load forecasting using wavelet transform combined with Holt–Winters and weighted nearest neighbor models. *Int. J. Electr. Power Energy Syst.* **2015**, *64*, 340–346. [CrossRef]

31. Fard, A.K.; Akbari-Zadeh, M.-R. A hybrid method based on wavelet, ANN and ARIMA model for short-term load forecasting. *J. Exp. Theor. Artif. Intell.* **2014**, *26*, 167–182. [CrossRef]

32. Tibshirani, R. Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2011**, *73*, 273–282. [CrossRef]

33. Dudek, G. Pattern–based local linear regression models for short-term load forecasting. *Electr. Power Syst. Res.* **2016**, *130*. [CrossRef]

34. Bickel, P.J.; Gel, Y.R. Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2011**, *73*, 711–728. [CrossRef]

35. Ghaderpour, E.; Vujadinovic, T. The potential of the least-squares spectral and cross-wavelet analyses for near-real-time disturbance detection within unequally spaced satellite image time series. *Remote Sens.* **2020**, *12*, 2446. [CrossRef]

36. Ghaderpour, E.; Pagiatakis, S.D. Least-Squares Wavelet Analysis of Unequally Spaced and Non-stationary Time Series and Its Applications. *Math. Geosci.* **2017**, *49*, 819–844. [CrossRef]

37. Taylor, J.W.; Buizza, R. Using weather ensemble predictions in electricity demand forecasting. *Int. J. Forecast.* **2003**, *19*, 57–70. [CrossRef]

38. Shibli, S.M.A.; Meena, B.N.; Remya, R. A review on recent approaches in the field of hot dip zinc galvanizing process. *Surf. Coat. Technol.* **2015**, *262*, 210–215. [CrossRef]

39. Bush, G.W. Developments in the continuous galvanizing of steel. *JOM* **1989**, *41*, 34–36. [CrossRef]

40. Debón, A.; García-Díaz, J.C. Fault diagnosis and comparing risk for the steel coil manufacturing process using statistical models for binary data. *Reliab. Eng. Syst. Saf.* **2012**, *100*, 102–114. [CrossRef]

41. García-Díaz, J.C. Fault detection and diagnosis in monitoring a hot dip galvanizing line using multivariate statistical process control. *Saf. Reliab. Risk Anal. Theory Methods Appl.* **2009**, *1*, 201–204.

42. Ajersch, F.; Ilinca, F.; Hétu, J.F. Simulation of flow in a continuous galvanizing bath: Part II. Transient aluminum distribution resulting from ingot addition. *Metall. Mater. Trans. B* **2004**, *35*, 171–178. [CrossRef]

43. Tang, N.Y. Characteristics of continuous-galvanizing baths. *Metall. Mater. Trans. B* **1999**, *30*, 144–148. [CrossRef]

44. Hippert, H.S.; Pedreira, C.E.; Souza, R.C. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Trans. Power Syst.* **2001**, *16*, 44–55. [CrossRef]

45. Baliyan, A.; Gaurav, K.; Kumar Mishra, S. A review of short term load forecasting using artificial neural network models. *Procedia Comput. Sci.* **2015**, *48*, 121–125. [CrossRef]

46. Xie, H.; Tang, H.; Liao, Y.H. Time series prediction based on narx neural networks: An advanced approach. *Proc. Int. Conf. Mach. Learn. Cybern.* **2009**, *3*, 1275–1279.

47. Liu, Q.; Chen, W.; Hu, H.; Zhu, Q.; Xie, Z. An Optimal NARX Neural Network Identification Model for a Magnetorheological Damper with Force–Distortion Behavior. *Front. Mater.* **2020**, *7*, 1–12. [CrossRef]

48. Deoras, A. Electricity Load and Price Forecasting Webinar Case Study. Available online: https://es.mathworks.com/matlabcentral/fileexchange/28684--electricity--load--and--price--forecasting--webinar--case--study (accessed on 6 July 2020).

49. Moré, J.J. The Levenberg–Marquardt algorithm: Implementation and theory. In *Numerical Analysis. Notes in Mathematics*; Watson, G.A., Ed.; Springer: Berlin/Heidelberg, Germany, 1978; Volume 630, pp. 105–116. ISBN 978-3-540-08538-6.

50. Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1970.

51. Lu, Y.; AbouRizk, S.M. Automated Box–Jenkins forecasting modelling. *Autom. Constr.* **2009**, *18*, 547–558. [CrossRef]

52. Brockwell, P.J.; Davis, R.A.; Fienberg, S.E. *Time Series: Theory and Methods: Theory and Methods*, 2nd ed.; Springer Science & Business Media: New York, NY, USA, 1991; ISBN 0387974296.

53. Holt, C.C. *Forecasting Seasonals and Trends by Exponentially Weighted Averages*; Carnegie Institute of Technology, Graduate school of Industrial Administration.: Pittsburgh, PA, USA, 1957.

54. Brown, R.G. *Statistical Forecasting for Inventory Control*; McGraw-Hill: New York, NY, USA, 1959.

55. Gardner, E.S., Jr.; McKenzie, E. Forecasting Trends in Time Series. *Manag. Sci.* **1985**, *31*, 1237–1246. [CrossRef]

56. Gardner, E.S., Jr.; McKenzie, E. Why the damped trend works. *J. Oper. Res. Soc.* **2011**, *62*, 1177–1180. [CrossRef]

57. Chatfield, C. The Holt–Winters forecasting procedure. *Appl. Stat.* **1978**, *27*, 264–279. [CrossRef]

58. Gardner, E.S., Jr. Exponential smoothing: The state of the art—Part II. *Int. J. Forecast.* **2006**, *22*, 637–666. [CrossRef]

59. Segura, J.V.; Vercher, E. A spreadsheet modeling approach to the Holt–Winters optimal forecasting. *Eur. J. Oper. Res.* **2001**, *131*, 375–388. [CrossRef]

60. Bermúdez, J.D.; Segura, J.V.; Vercher, E. Improving demand forecasting accuracy using nonlinear programming software. *J. Oper. Res. Soc.* **2006**, *57*, 94–100. [CrossRef]

61. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [CrossRef]

62. Lagarias, J.C.; Reeds, J.A.; Wright, M.H.; Wright, P.E. Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM J. Optim.* **1998**, *9*, 112–147. [CrossRef]

63. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*; MIT Press: Cambridge, MA, USA, 2009; ISBN 978-0262013192.

64. Durbin, J.; Koopman, S.J. A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **2002**, *89*, 603–615. [CrossRef]

65. Trull, O.; García–Díaz, J.C.; Troncoso, A. Stability of multiple seasonal holt–winters models applied to hourly electricity demand in Spain. *Appl. Sci.* **2020**, *10*, 2630. [CrossRef]

66. Cleveland, R.B.; Cleveland, W.S.; McRae, J.E.; Terpenning, I. STL: A seasonal–trend decomposition procedure based on loess. *J. Off. Stat.* **1990**, *6*, 3–73.

67. Fan, J.; Yao, Q. Characteristics of Time Series. In *Nonlinear Time Series. Nonparametric and Parametric Methods*; Springer: New York, NY, USA, 2003; ISBN 978-0-387-26142-3.

68. Alessio, S.M. *Digital Signal Processing and Spectral Analysis for Scientists*; Springer: Cham, Switzerland, 2016; ISBN 978-3-319-25466-1.

69. Ghaderpour, E.; Ince, E.S.; Pagiatakis, S.D. Least-squares cross-wavelet analysis and its applications in geophysical time series. *J. Geod.* **2018**, *92*, 1223–1236. [CrossRef]

70. Ghaderpour, E.; Pagiatakis, S.D. LSWAVE: A MATLAB software for the least-squares wavelet and cross-wavelet analyses. *GPS Solut.* **2019**, *23*, 1–8. [CrossRef]

71. Gómez, V. SSMMATLAB: A Set of MATLAB Programs for the Statistical Analysis of State Space Models. *J. Stat. Softw.* **2015**, *66*, 1–37. [CrossRef]

72. Hyndman, R.J.; Khandakar, Y. Automatic time series forecasting: The forecast package for R. *J. Stat. Softw.* **2008**, *27*. [CrossRef]

73. Kim, S.; Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* **2016**, *32*, 669–679. [CrossRef]

74. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 2nd ed.; OTexts: Melbourne, Australia, 2018; Chapter 3.4; ISBN 978-0-9875071-1-2.