Initialization Methods for Multiple Seasonal Holt–Winters Forecasting Models

Oscar Trull [1],* [ID]**, Juan Carlos García-Díaz [1]** [ID] **and Alicia Troncoso [2]** [ID]

[1]   Department of Applied Statistics and Operational Research and Quality, Universitat Politècnica de València, 46022 Valencia, Spain; juagardi@eio.upv.es
[2]   Department of Computer Science, Pablo de Olavide University, 41013 Sevilla, Spain; atrolor@upo.es
*   Correspondence: otrull@eio.upv.es

Abstract: The Holt–Winters models are one of the most popular forecasting algorithms. As well-known, these models are recursive and thus, an initialization value is needed to feed the model, being that a proper initialization of the Holt–Winters models is crucial for obtaining a good accuracy of the predictions. Moreover, the introduction of multiple seasonal Holt–Winters models requires a new development of methods for seed initialization and obtaining initial values. This work proposes new initialization methods based on the adaptation of the traditional methods developed for a single seasonality in order to include multiple seasonalities. Thus, new methods to initialize the level, trend, and seasonality in multiple seasonal Holt–Winters models are presented. These new methods are tested with an application for electricity demand in Spain and analyzed for their impact on the accuracy of forecasts. As a consequence of the analysis carried out, which initialization method to use for the level, trend, and seasonality in multiple seasonal Holt–Winters models with an additive and multiplicative trend is provided.

Keywords: forecasting; multiple seasonal periods; Holt–Winters, initialization

1. Introduction

Exponential smoothing methods are widely used worldwide in time series forecasting, especially in financial and energy forecasting [1,2]. These methods smooth the observed values of a time series, using an exponentially downward weighting strategy giving more weight to recent values compared to older values. Weights decrease geometrically over past observations. The most commonly used methods are: Simple exponential smoothing for the series that does not show any trend, or with a locally stable mean, having a single smoothing parameter α; double seasonal exponential smoothing (also known as Holt's method [3]) for trend-having series with two smoothing parameters: α and γ; finally, triple exponential smoothing (also known as Holt–Winters [4]) for the series with a trend and seasonal component, with α, γ, and δ as smoothing parameters. The smoothing parameters are bounded in the interval [0, 1], where the closer to 0 means the more weight of the old values against the newer ones.

The multiple seasonal Holt–Winters models (nHWT) outperform former Holt–Winters methods by including several seasonalities nested in the model. This development has been driven by the necessity of improving forecasts, especially in energy planning, where system operators need accurate forecasts. Hong [5] determines how a 1% improvement in the forecast accuracy can save up to $600,000 per year in a 1 GW power station.

However, these new forecasting methods are recursive and based on initial values, and the performance of the models highly depend on those initial values. Thus, new initialization methods to

include multiple seasonality must be developed. Another question is then raised: Is there any method that outperforms the rest?

This article intends to provide answers to these questions. Hence, new initialization methods are proposed for nHWT models and their performance is analyzed for forecast accuracy by using the Spanish hourly electricity demand.

Holt–Winters models were formerly introduced in the 1960s and have been very commonly used in forecasting techniques and signal processing. The Holt–Winters model (also known as triple exponential smoothing) introduced by Winters [4] consists of three smoothing equations and a forecasting equation as shown in Equations (1)–(4):

$$\text{Level}: \ S_t = \alpha\left(\frac{X_t}{I_{t-s}}\right) + (1-\alpha)(S_{t-1} + T_{t-1}) \tag{1}$$

$$\text{Trend}: \ T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1} \tag{2}$$

$$\text{Seasonality}: \ I_t = \delta\left(\frac{X_t}{S_t}\right) + (1-\delta)I_{t-1} \tag{3}$$

$$\text{Forecasting}: \ \hat{X}_t(k) = (S_t + k\,T_t)I_{t-s+k} \tag{4}$$

where α, γ, and δ are the smoothing parameters, and S_t, T_t, and I_t are the smoothing equations known as level, trend, and seasonality. The information from the observed values (X_t) is projected through a forecasting Equation (4) k steps ahead to obtain predictions ($\hat{X}_t\ (k)$).

Taylor [6,7] introduced the double and triple seasonal Holt–Winters models (HWT). These models are characterized by capturing the information contained in the seasonal component, split into several seasonalities of different lengths, as well as including an adjustment of the forecast including the first autocorrelation error. García-Díaz and Trull [8] generalized these models to adapt to an indeterminate number of seasonalities, proposing the multiple seasonal Holt–Winters models (nHWT).

The trend and seasonal equations can be expressed in an additive or multiplicative form. The way the equations are combined allow a total of 30 possible combinations. The nomenclature used to name the models is based on three letters: The first to indicate the trend method, the second to indicate the seasonality method, and a third to indicate if the model has an adjustment using the first order autocorrelation error, as shown in Table 1. Additionally, the seasonalities used are indicated by adding subscripts with period lengths.

Table 1. Multiple seasonal Holt–Winters models' nomenclature according to the trend and seasonal method.

Trend	Seasonal					
	AR (1) Adjusted			Non-Adjusted		
	None	Additive	Multiplicative	None	Additive	Multiplicative
None	NNC	NAC	NMC	NNL	NAL	NML
Additive	ANC	AAC	AMC	ANL	AAL	AML
Damped Additive	dNC	dAC	dMC	dNL	dAL	dML
Multiplicative	MNC	MAC	MMC	MNL	MAL	MML
Damped Multiplicative	DNC	DAC	DMC	DNL	DAL	DML

The notation used is based on the following criteria: The first letter stands for the trend method (N: None, A: Additive, M: Multiplicative, d: Damped Additive, and D: Damped Multiplicative). The second letter stands for the seasonality method (N: None, A: Additive, and M: Multiplicative). The third letter is used to designate if the model has an AR (1) adjustment (C: Adjusted, L: Not adjusted).

As an example, the $AMC_{24,168}$ model is a model with an additive trend and multiplicative seasonality; the model is adjusted by using the first-order autocorrelation error, and it has two

seasonalities: One daily (subscript 24) and one weekly (subscript 168). The equations defining the $AMC_{24,168}$ model are described in Equations (5)–(9). The general formulae of all possible models depending on the trend and seasonal type can be found in Appendix A:

$$S_t = \alpha \left(\frac{X_t}{I_{t-24}^{(24)} I_{t-168}^{(168)}} \right) + (1 - \alpha)(S_{t-1} + T_{t-1}) \tag{5}$$

$$T_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)T_{t-1} \tag{6}$$

$$I_t^{(24)} = \delta^{(24)} \left(\frac{X_t}{S_t I_{t-168}^{(168)}} \right) + \left(1 - \delta^{(24)}\right) I_{t-1}^{(24)} \tag{7}$$

$$I_t^{(168)} = \delta^{(168)} \left(\frac{X_t}{S_t I_{t-24}^{(24)}} \right) + \left(1 - \delta^{(168)}\right) I_{t-1}^{(168)} \tag{8}$$

$$\hat{X}_t(k) = (S_t + k\,T_t) I_{t-24+k}^{(24)} I_{t-168+k}^{(168)} + \varphi_{AR}^k \varepsilon_t \tag{9}$$

where the former seasonal equation I_t now splits into $I_t^{(s_i)}$ and δ splits into $\delta^{(s_i)}$, s_i being the seasonal periods considered, $s_1 = 24$ and $s_2 = 168$. Here, the value φ_{AR} is the adjustment factor to include the first-order autocorrelation error (ε_t).

The rest of the paper is organized as follows: In Section 2, the previous literature is analyzed; in Section 3 the initialization methods are described; in Section 4, an analysis is carried out and the results obtained are described; and in Section 5, these results are discussed. Section 6 lists the conclusions reached.

2. Related Work

The Holt–Winters models are recursive, and thus an initialization value is needed to feed the model. With the development of new models, it is mandatory to define new initialization methods for obtaining the initial values of the smoothing equations, considering that the older methods were defined only for single seasonal models. These initialization methods also depend on the method used for trend and seasonality.

The first methods of obtaining initial values date back to the proposal made by Winters [4] with the definition of triple exponential smoothing models. (Do not confuse triple exponential smoothing, with triple seasonal exponential smoothing, the former being a reference to the use of three equations, while the latter refers to the use of three seasonalities). In [4], the initial value of the trend (T_0) is calculated as the slope between the average of the last available full period q and the first, as indicated by Equation (10), where the only seasonality is of length s.

$$T_0 = \frac{\overline{D}_q - \overline{D}_1}{(q - 1)s} \tag{10}$$

Here, the value of $\overline{D}_1$ indicates the average value for the first period and $\overline{D}_q$ indicates the average value for the last period considered q. In many cases, using the first two periods (that is, $q = 2$) is sufficient [9]. Granger and Newbold [10] proposed leaving the trend in values of 0 for additive trend methods and 1 for multiplicative trend methods. In this way, the slope adjusts itself to the trend value.

The previous work of Holt [3,11] and Brown [12] in single and double exponential smoothing did not require calculations of initial values external to the model. Makridakis, Wheelright, and Hyndman [9] proposed a new method of calculating the trend and seasonality. Bowerman, O'Connell and Pack [13], and Bowerman, O'Connell, and Köhler [14] proposed a regression method where they consider that the intercept is the level and the slope is the trend. This method is criticized

by Hyndman [15] for the bias it introduces. Rasmussen [16] proposed obtaining initial values through minimization. With the emergence of multiple seasonal models, new initialization methods were introduced [6,7], although some authors prefer not to use them [17]. Taylor introduced a new initialization method of the level and trend applied to double seasonal Holt–Winters in [6]. In particular, the seasonality is initialized as the division of the simple values over the moving average of the period (called the "simple" method), and the level (S_0) is obtained once the series trend has been eliminated. It consists in first obtaining the trend of the series and then the level, as seen in Equation (11):

$$S_0 = \overline{D}_1 - \frac{s_m + 1}{2} T_0 \tag{11}$$

where s_m is the longest seasonality. This scheme to initialize the level was also adopted by Hyndman et al. [18]. Hyndman introduced new initialization methods incorporated in the exponential smoothing methodology based on state spaces and included them in the "forecast" package of the statistical software R. Trull, García-Díaz, and Troncoso [19] also used a series decomposition using STL methods of various seasonality, in order to obtain initial values in discrete seasonality.

Some papers [20–22] analyze the initial values but very succinctly. Segura and Vercher [23] collect the information about the initialization methods and perform an analysis similar to that carried out by Makridakis and Hibon [24]. They conclude that when the parameters are optimized, it is not possible to determine significant differences between the different initialization methods.

After an exhaustive review of these previously published works, it can be concluded that the analysis of the initialization methods is not very extensive in the literature. This fact justifies the need for research in the topic described in this paper.

3. Materials and Methods

The way to obtain the initial values must match the trend and seasonal method, as it would impact later on the accuracy of the model. The addition of more seasonal patterns in the model requires newer methods of obtaining the initial values or seeds. Therefore, several methods depending on the trend and seasonal method have been implemented. This paper proposes new initialization methods for multiple seasonalities based on the previous methods, which were defined for a single seasonal pattern.

The proposed methods for the initialization of the Holt–Winters models are separated into three groups. The first ones calculate the seeds for the level and trend separately. The last one implements each seasonality separately. The latter first obtains a seed, as the trend, and then obtains another one on a recurring basis, such as that of the level.

3.1. Level Methods

The level initialization basically consists in determining the longest period, or the longest seasonality (s_m) and determining the initial values around that point, by using the moving average (adapted from Holt [3]) or the average (adapted from Winters [4]). Williams and Miller [25] used the first value of the series for the level. This method is included although it remains unchanged for the further comparative analysis in Section 4. Table 2 shows the adapted methods to obtain the initial value of the level (S_0).

Table 2. Proposed methods and the method published in [25] for obtaining the initial value of the level in multiple-seasonal methods.

Denomination	Calculation
Moving average adapted from Holt [3]	$S_0 = \frac{1}{S_m}\left[\frac{X_{1+s_m}-X_1}{s_m} + \frac{X_{2+s_m}-X_2}{s_m} + \ldots + \frac{X_{2s_m}-X_{s_m}}{s_m}\right]$
First value [25]	$S_0 = X_1$
First period's average adapted from Winters [4]	$S_0 = \frac{\sum_1^{s_m} X_i}{s_m}$

3.2. Trend Methods

Based on Winters' method [4], a new initialization method has been developed to be able to include more than one seasonality in the model, according to the contribution to the trend of smaller seasonality, as indicated [24]. These proposals (methods overall and two periods) are shown in Table 3 along with the methods published in [6,26] for the further comparative analysis in Section 4. T'_{0i} is an intermediate step where individual contributions are calculated in each period of length s_i. The average of all of them will be taken into account, counting the maximum number of periods q for the overall case, and only two periods for that special case ($q = 2$). The seasonality of greater length is used as the basis of calculation. The value of s_m indicates the length of the longest seasonal period considered in the model. The rest is calculated in the form of differences. The value s_p indicates the length of the minor of the periods considered. ∇D_{12} indicates differences in values for the first two periods, while ∇D_i refers to differences in the period considered. Finally, the average of the contributions is calculated to obtain the value of T_0.

Table 3. Proposed methods and methods published in [6,25] for obtaining the initial value of the trend in multiple-seasonal methods.

Denomination	Calculation
Newbold [26]	$T_0 = 0$ (additive) $T_0 = 1$ (multiplicative)
Taylor [6]	$T_0 = \dfrac{\left(\dfrac{\overline{D}_1 - \overline{D}_2}{s_m} + \dfrac{\sum_1^{2s_m} \nabla D_{12}}{2s_p}\right)}{2}$
Overall and two periods method. Multiplicative.	$T_0 = \dfrac{\sum_{i=1,\dots,q} T'_{0i}}{q},$ with $T'_{0i} = \begin{cases} \dfrac{\overline{D}_q - \overline{D}_1}{q} & \textbf{if} \quad s_i = s_m \\ \dfrac{\sum_1^{s_i} \nabla D_i}{s_i} & \textbf{if} \quad rest \end{cases}$
Overall and two periods method. Multiplicative.	$T_0 = \prod_1^{n_s} T'_{0i},$ with $T'_{0i} = \begin{cases} \dfrac{\overline{D}_q - \overline{D}_1}{q} & \textbf{if} \quad s_i = s_m \\ \dfrac{\sum_1^{s_i} \nabla D_i}{s_i} & \textbf{if} \quad rest \end{cases}$

3.3. Seasonality Calculation Methods

The first methods determined for the calculation were developed by Winters [4]. The calculation of seasonality consists in calculating the series averages over the series without a trend. When it comes to more than one seasonality that are modeled annually, as in nHWT, it is necessary to avoid duplication of the effects of seasonality. Thus, based on Winters' method [4], we propose to obtain the seasonal indices $I_{t-s_i}^{(s_i)}$ for each seasonality (i) of length s_i as described in Equations (12)–(16):

$$I_{t-s_i}^{(i)} = \begin{cases} I_{t-s_1}^{*(i)}, \quad i = 1 \\ \dfrac{I_{t-s_i}^{*(i)}}{\prod_{j=1}^{i-1} I_{t-s_j}^{*(i)}}, \quad i > 1 \end{cases} \tag{12}$$

with the individual seasonal indices,

$$I_{t-s_i}^{*(i)} = \frac{1}{q_i} \sum {I'_{t-s_i}}^{(i)}, \, t = 1, \dots, s_i \tag{13}$$

$$I_{t-s_i}^{\prime(i)} = \frac{1}{s_i} \sum {I'_{t,j}}^{(j)}, \, j = 1, \dots, q_i \tag{14}$$

$$I'^{(i)}_{t,j} = \frac{X_t}{\overline{D}_j^{i)} - \left(\frac{s_i+1}{2} - i\right)T_0},$$
$$j = 1, \ldots, s_i, t = j, j + s_i, \ldots, (q_i - 1)s_i \tag{15}$$

$$\overline{D}_j^{(i)} = \frac{1}{q_i} \sum_{k=1}^{q_i} X_t,$$
$$j = 1, \ldots, s_i; \ t = j, j + s_i, \ldots, j + (q_i - 1)s_i \tag{16}$$

where q_i is the maximum number of complete cycles of length s_i that exist in the series. i refers to the seasonality considered, so that $i = 1, \ldots, n_s$.

Brockwell and Davis [27] proposed another method, in which the calculation of seasonality is done by obtaining the weights of the series against the average values of the cyclic pattern. This methodology is adopted by the National Institute of Standards and Technology (NIST) [28]. This method has also been adapted to multiple seasonality, and the proposed method is described in the following steps:

Step 1:

For each seasonality, it is necessary to reorganize the data and obtain the means in each internal position within the seasonality as the average of the values obtained. In this way, the values $A_m^{(i)}$ are obtained, for each seasonality of length s_i and that occurs n_q times in the data series, as shown in Equation (17):

$$A_m^{(i)} = \frac{\sum_{j=1}^{s_i} X_{(m-1)*s_i+j}}{s_i}, \ m = 1, \ldots, n_q \tag{17}$$

Step 2:

For each seasonality, the proportion of the series over the average is obtained. There are m_i occurrences of seasonal cycles in the range considered of length s_i.

	1	2	...	q
1	$X_1/A_1^{(i)}$	$X_{s_i+1}/A_2^{(i)}$		$X_{s_i+1}/A_2^{(i)}$
...				
s_i	$X_{s_i}/A_1^{(i)}$	$X_{2s_i}/A_2^{(i)}$		$X_{m_1 s_i+1}/A_{m_i}^{(i)}$

Step 3:

The provisional seasonal indices of each seasonality on the series are obtained as follows:

$$I_{1-s_i}^{*(i)} = \frac{\frac{X_1}{A_1^{(i)}} + \frac{X_{s_i+1}}{A_2^{(i)}} + \ldots + \frac{X_{(m_i-1)s_i+1}}{A_{m_i}^{(i)}}}{m_i}$$

$$I_0^{*(i)} = \frac{\frac{X_{s_i}}{A_1^{(i)}} + \frac{X_{2s_i}}{A_2^{(i)}} + \ldots + \frac{X_{m_i s_i+1}}{A_{m_i}^{(i)}}}{m_i}$$

Step 4:

The definitive seasonal indices, shown in Equation (18) are obtained taking into account that they are nested, and therefore the seasonal indices by removing the first one, need to compensate for the effect produced in the indices considered prior to the one calculated.

$$I_{t-s_i}^{(i)} = \begin{cases} I_{t-s_1}^{*(i)}, & i = 1, \ t = 1, \ldots s_i \\ \\ \dfrac{I_{t-s_i}^{*(i)}}{\prod_{j=1}^{i-1} I_{t-s_j}^{*(i)}}, & i > 1, \ t = 1, \ldots s_i \end{cases} \tag{18}$$

Finally, the initialization method proposed by Granger and Newbold [26], named the "normal" method, has been adapted, where the values of the first cycle are divided by their average. This proposal is described by Equations (19) and (20):

$$I^{*(i)}_{t-s_i} = \frac{D_t}{\overline{D_{s_i}}}, \; t = 1, \ldots, s_i \tag{19}$$

$$I^{(i)}_{t-s_i} = \begin{cases} I^{*(i)}_{t-s_1}, & i = 1 \\[2ex] \dfrac{I^{*(i)}_{t-s_i}}{\prod_{j=1}^{i-1} I^{*(j)}_{t-s_j}}, & i > 1 \end{cases} \tag{20}$$

4. Results

The proposed new methods and older methods published in [6,25,26] were applied to the hourly electricity demand series in Spain in the period 1999 to 2004. Figure 1 shows the time series worked. This period is characterized by strong continued growth and marked seasonality. This will allow working with models with clear seasonality and trend.
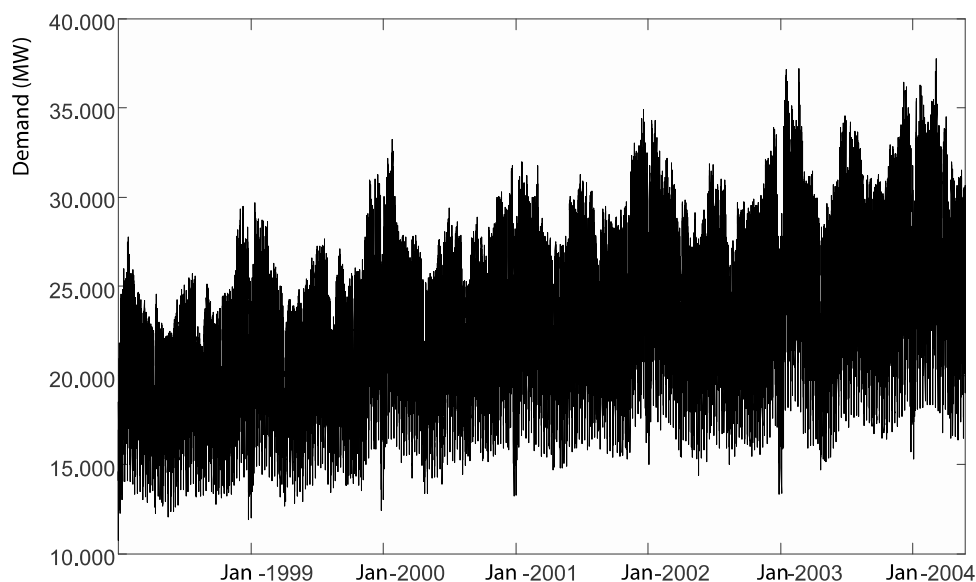

Figure 1. Spanish hourly electricity demand in the period 1998 to 2004.

The analysis has been raised in the form of an experimental design. The main factors to analyze are the different methods to estimate the initial values of level, trend, and seasonality. It has been used as two seasonal patterns. It is important to note that, when performing this analysis, the influence of the meteorological variables has been considered as a noise factor, so that windows have been created in the four seasons of the year where the analysis was performed. Within the same context, several replicas of the analysis have been included, using the different years.

The response variables used were the MAPE (see Equation (20)), an indicator of forecast accuracy widely used in the time series analysis [29]. To obtain the response variable, the seven-week series have been used randomly extracted from the series, although always taking into account that they belong to a specific season of the year—having at least six series in each seasonal period and each year.

$$MAPE(\%) = 100 \frac{1}{h} \sum_h \frac{\left|\hat{X}_t - X_t\right|}{|X_t|} \tag{21}$$

The process of comparing the initial values needs a specific methodology, since once the parameter optimization is performed, the differences between them are lost [9,24]. This situation can be avoided by performing the analysis based on the forecasts directly from the initial values and without adjusting. It is also necessary that the model does not cushion the results, which implies that the parameters used must be $\alpha = 0$, $\gamma = 0$, and $\delta^{(i)} = 0$ for all seasons. Table 4 shows all methods both proposed and published in the literature that will be used to initialize the level, trend, and seasonality factors of nHWT models.

Table 4. The factors and initialization methods analyzed.

Factor	Type	Initialization Methods	
		New Methods	**Older**
Level	Controllable	Moving average, average	First value [25], Taylor [6]
Trend	Controllable	Overall, two periods	Newbold [26], Taylor [6]
Seasonality	Controllable	Normal, Winters, NIST	Simple [6]

The names used are consistent with the methods presented in the previous section. In addition, they have been carried out in four different years, which is an additional factor to consider. The models studied have been $AMC_{24,168}$, $AAC_{24,168}$, $MAC_{24,168}$, and $MMC_{24,168}$.

The design of experiments includes 4096 tests, three controllable multilevel factors, and a noise factor that is the period of the year replicated for four different years. This study is shown in Table 5.

Table 5. Analysis of the variance for the factors studied using the models with multiplicative trend.

	Source	Sum of Squares	df	Mean Square	F-Ratio	*p*-Value
		Main effects				
A	Level	283.89	3	94.63	6.66	0.0002
B	Trend	2.96E6	3	984124.	69240.76	0.0000
C	Seasonality	1099.63	3	366.52	25.79	0.0000
D	Model	19.36	1	19.36	1.36	0.2432
E	Season	91.29	3	30.43	2.14	0.0931
		Interactions				
	AB	876.46	9	97.3843	6.85	0.0000
	BC	300.32	9	33.3691	2.35	0.0124
	BE	296.07	9	32.8968	2.31	0.0138
	CD	189.57	3	63.1903	4.45	0.0040
	Residuals	28483.0	2004	14.2131		
	Total	2.98E6	2047			

The first action taken, regardless of the design, was to be clear if there are significant differences between the models and from there determine which options to use. The result is shown in Figure 2, where you can see how the groups with additive trend differ significantly from the groups with multiplicative trend. Thus, the design will be separated in two to address each trend by its side.
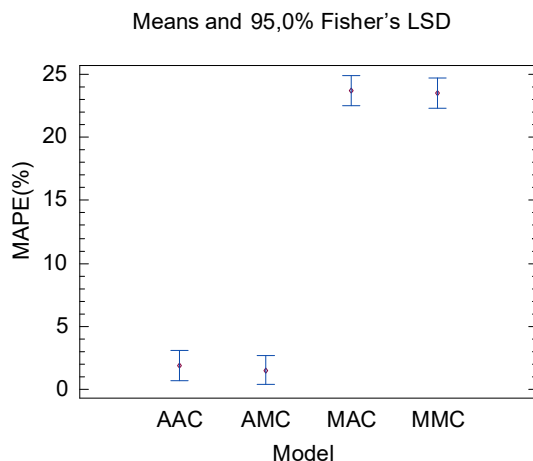
Figure 2. Means and Fisher's LSD (least significant differences) diagram of the MAPE obtained against the nHWT models considered.

Table 5 shows the results of the ANOVA (analysis of variance) obtained during the design. Only those significant factors and interactions are shown. This ANOVA has been carried out using the models with multiplicative trend and in response to the MAPE.

It is appreciated how the use of different seasonal methods is not significant, while the rest of the factors are. This model is also able to minimize the effect of the seasons of the year, becoming a nonsignificant factor, due to the influence of the trend being superior to that of the other parameters.

The same analysis performed on the models with additive trend offers slightly different results. Table 6 shows the ANOVA of this analysis. Only the significant factors and interactions are shown. The behavior is similar, although the additive trend methods are more sensitive to the annual period considered and the seasonality method considered.

Table 6. Analysis of the variance for the factors studied using the models with additive trend, without adjustment.

	Source	Sum of Squares	df	Mean Square	F-ratio	*p*-Value
	Main effects					
A	Level	0.8485	3	0.282849	5.59	0.0008
B	Trend	1.3228	3	0.440938	8.71	0.0000
C	Seasonality	1864.54	3	621.514	12276.07	0.0000
D	Model	73.2721	1	73.2721	1447.26	0.0000
E	Season	1.1902	3	0.396725	7.84	0.0000
	Interactions					
	AE	1.19074	9	0.132305	2.61	0.0053
	BD	1.1674	3	0.389135	7.59	0.0000
	CD	299.458	3	99.8192	1971.62	0.0000
	Residuals	102.218	2019	0.0506281		
	Total	2345.21	2047			

The analysis for each factor is shown in the graphs of Figure 3. This figure shows the average graphs of the different factors separated by the trend method. MAPE (%) is displayed before optimizing. You can see how the behaviors are different depending on the trend. For those with multiplicative trend only the method proposed by Taylor is significantly different, both for the level and for the trend, but not for the seasonality. However, for additive trend methods, there are differences in all cases. Seasonal values in additive trend methods show differences between them, with the simple method and NIST being the best results.
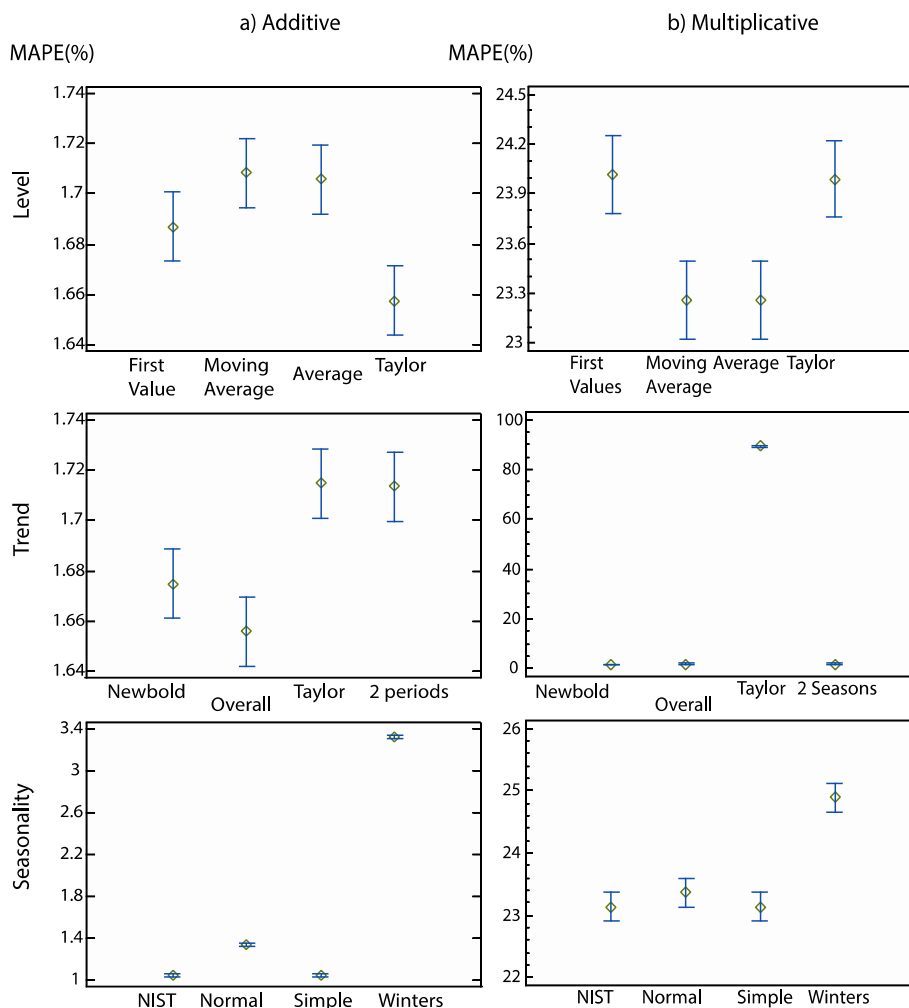
Figure 3. Mean graphs for the MAPE according to the initialization methods analyzed. Left column: Additive trend methods, right column: Multiplicative. From top to bottom: Level, trend, and seasonality. The levels used are those described in Table 4.

Following these results, the values obtained during the experiment were analyzed, and each of the analyzed factors was studied separately.

4.1. Trend Analysis

The trend determines the behavior of the series in the long term. Using the seven-week series, this influence is remarkable. The models that include multiplicative trend obtain results for the trend close to 1. It is difficult to assess this situation. However, with the additive trend models, more information can be obtained. The distribution of the initial values for the additive trend is shown in Figure 4. It should be noted that the initial value of the Newbold method is always 0, and that the overall method has little variability. This is not the case with the two-period method. This tells us that the larger the set of data we take for the calculation of initial values, the more the variability is reduced.

Taylor's method has a different behavior. Its variability is much greater than the rest. This is in contrast to using only the first two periods, in addition to introducing the contribution of the minor seasonality to obtain the overall trend.
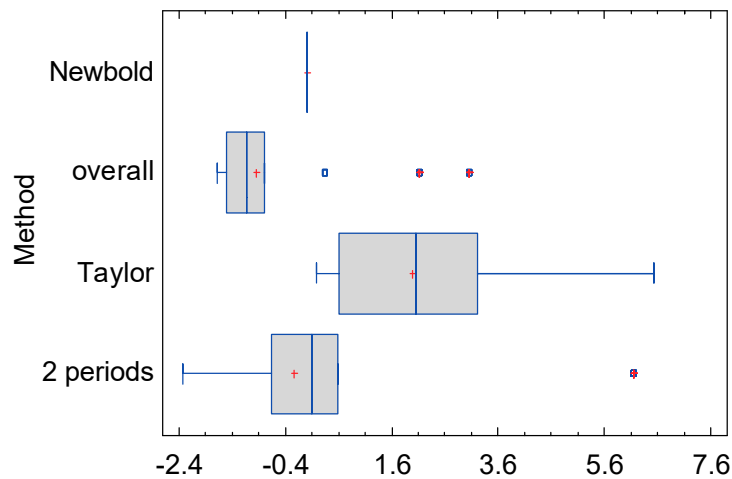
Figure 4. Distribution of initial values obtained for the trend in additive trend models $AMC_{24,168}$ and $AAC_{24,168}$.

4.2. Level Analysis

The analysis on the level shown in Figure 5 reveals that the level values in the additive trend models are close. However, following what has been verified in the ANOVA, they are clearly significant when determining the accuracy of the model. The fact that Taylor's method for the level is mixed, and that the trend has been previously calculated, causes this method to decrease variability, transferring it to the trend. For cases of multiplicative trend, the results are very similar to those obtained in this analysis.
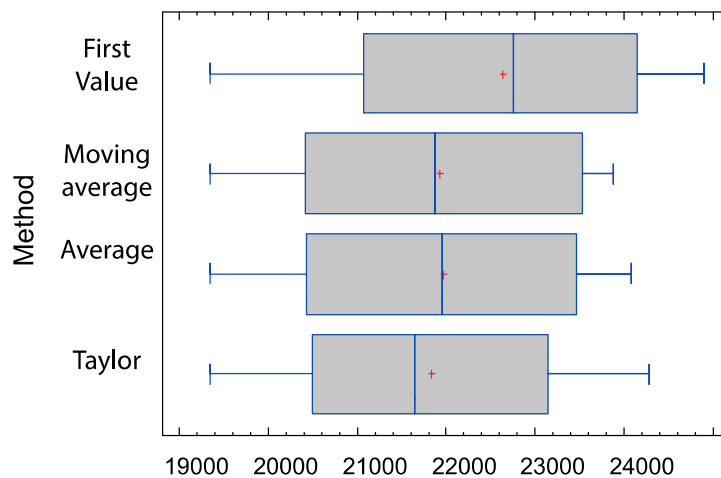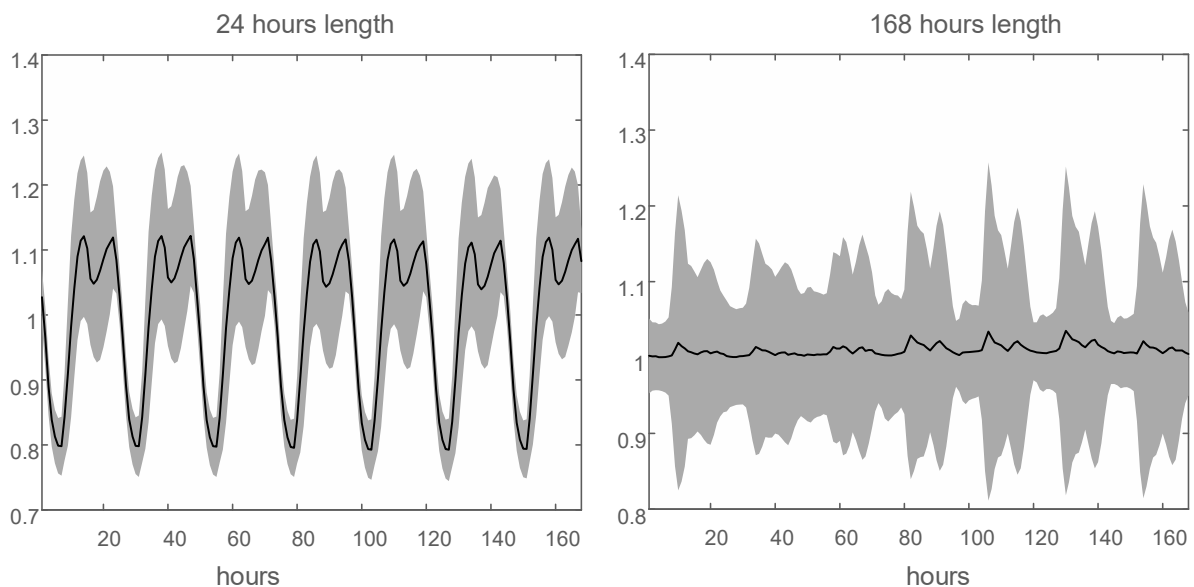

Figure 5. Distribution of the initial values obtained for the level in the additive trend models $AMC_{24,168}$ y $AAC_{24,168}$.

4.3. Seasonality Analysis

The ANOVA indicates that there are significant differences only in the additive trend models. Therefore, this analysis focuses on these models. There are two large groups between which you can include the original adapted Winters method: The NIST method and the simple adapted one. The analysis focuses on these two methods—simple and Winters—and the results obtained with the $AMC_{24,168}$ model have been used. Therefore, the initial values for the 24- and 168-hour length seasonality are shown. Figure 6 shows the distribution of the initial values for these seasonalities using the simple method. Similarly, Figure 7 reflects this same distribution but using the Winters method.
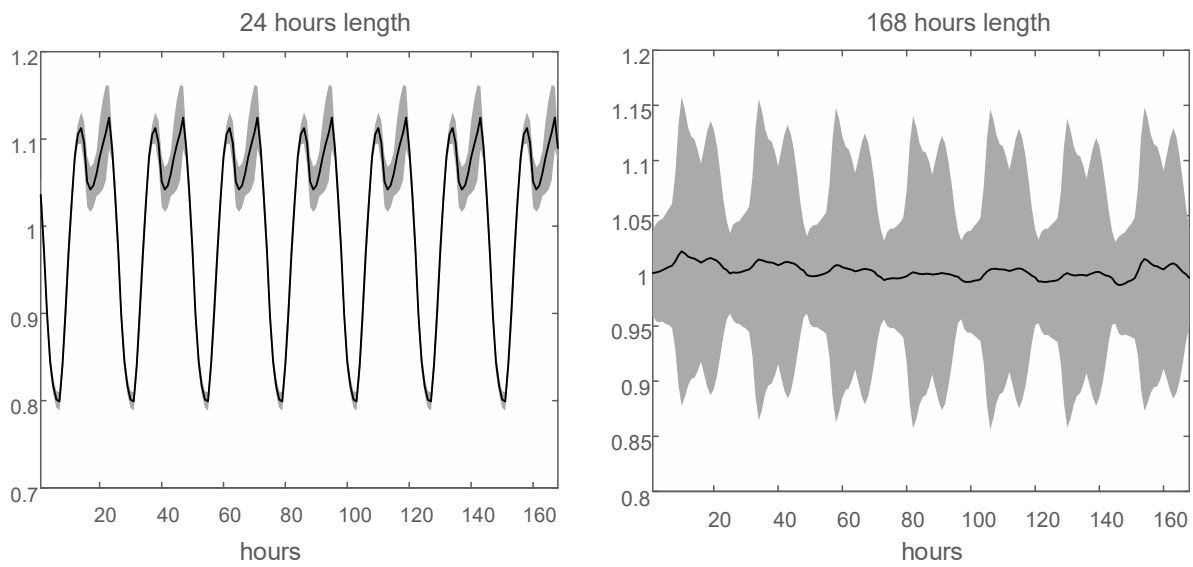
Figure 6. Distribution of the initial seasonal indices for the $AMC_{24,168}$ model using the simple method. In the left part, the seasonality of 24 hours is used, and in the right part the seasonality of 168 hours is used. The central line represents the average of the values, regardless of the time of the year, while the shaded area represents the variability with ± standard deviation.

Figure 7. Distribution of the initial seasonal indices for the $AMC_{24,168}$ model using the Winters method. In the left part, the seasonality of 24 hours is used, and in the right part the seasonality of 168 hours is used. The central line represents the average of the values, regardless of the time of the year, while the shaded area represents the variability with ± standard deviation.

It can be seen how the simple method has a lower variability in obtaining the seasonal indices of 24 hours, which is accompanied by a lower variability in the seasonal index of 168 hours. This indicates that the simple method ensures that the initial values are better suited to the seasonal patterns shown in the series. That is why it shows a smaller MAPE series. Moreover, you can see how the daily seasonal index is the primary index, and the weekly index has to be adapted according to the series. You might think that this situation occurs by the method of obtaining the initial values, where the first seasonality is considered the daily seasonality. However, we have performed tests where the first seasonality considered is the weekly one, and the results are the same.

5. Discussion

As a result of the analyses carried out in the study, it can be seen how the use of one or another method of obtaining initial values can greatly affect the accuracy of the forecasts. Although once the values of the parameters are optimized, these differences can be minimal (as other authors already describe), their influence on accuracy is clear. As a consequence, the parameters may not be optimal starting from unsuitable initial values, which complicates the optimization of the model in the adjustment, and later, in the forecasts.

The first impression is to see how multiplicative trend models start with a much higher MAPE than additive trend models do. In principle, it could be thought that it is due to the trend method used, but if the MAPE values are observed in Figure 3b, it is observed how this is not the case, but the accuracy is worsened by the rest of the equations with the trend. Clearly, the work of the parameters in this case is to mitigate the effect of the initial seasonal and level values. It has also been observed that the result of the MAPE depends a lot on the trend and seasonality methods used. This does not mean that one method is better than another method, but there are differences between them at the initial state that the model must deal with. For each group of methods, we tried to locate differences within the trend, level, and seasonal methods, looking for the MAPE of the model without optimizing. The cases of multiplicative trend, for example, do not show large differences between initialization methods. On the contrary, the methods that use the additive trend do show a great dependence on how to obtain the initial values. Another factor to highlight is the behavior of seasonal indices. Depending on the method, a greater or lesser variability can be observed.

Thus, for the multiple seasonal methods of additive trend it is proposed to use the Taylor method for the level, the overall method for the trend, and the simple method or that of the NIST for seasonality. For models with a multiplicative trend, the use of the average method for the level, the Granger and Newbold method for the trend, and any of the methods proposed for seasonality are proposed.

6. Conclusions

In this article, we have presented new methods of obtaining initial values for multiple seasonal Holt–Winters models. These methods are based on the adaptation of existing methods for models with a single seasonality. In addition, a comparative analysis between our proposals and well-known methods published in the literature has been carried out in order to determine their influence on the accuracy of the models. All these initialization methods have been applied to the prediction of the electricity demand time series in Spain, in a period between 1999 and 2004. The characteristics of the series allow all methods to be analyzed effectively. An experimental design has been performed in which MAPE has been analyzed as a response variable, using the different initialization methods of the smoothing equations as factors to influence the response variable. The results show that the influence of the methods of obtaining initial values is greater on multiple seasonal Holt–Winters models with an additive trend. Finally, and based on the results obtained, a proposal on which method of initialization to use has been made.

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

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

Appendix A

The multiple seasonal Holt–Winters formulae are shown in Table A1. They are organized according to the trend (rows) and seasonal method (columns), as described in [8].

Table A1. Multiple seasonal methods equations.

		Trend			Trend	
		None	Additive	Damped Additive	Multiplicative	Damped Multiplicative
Seasonality	None	$S_t = \alpha X_t + (1-\alpha)S_{t-1}$ $\hat{X}_t(k) = S_t + \varphi_{AR}^k \varepsilon_t$	$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $\hat{X}_t(k) = S_t + kT_t + \varphi_{AR}^k \varepsilon_t$	$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\phi T_{t-1}$ $\hat{X}_t(k) = S_t + T_t \sum_{n=1}^{k} \phi^n + \varphi_{AR}^k \varepsilon_t$	$S_t = \alpha X_t + (1-\alpha)(S_{t-1}R_{t-1})$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}$ $\hat{X}_t(k) = S_t R_t^k + \varphi_{AR}^k \varepsilon_t$	$S_t = \alpha X_t + (1-\alpha)\left(S_{t-1}R_{t-1}^{\phi}\right)$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}^{\phi}$ $\hat{X}_t(k) = S_t R_t^{\sum_{j=1}^{k}\phi^j} + \varphi_{AR}^k \varepsilon_t$
	Additive	$S_t = \alpha(X_t - \sum_i I_{t-s_i}^{(i)}) + (1-\alpha)S_{t-1}$ $I_t^{(i)} = \delta^{(i)}\left(X_t - S_t - \sum_{j\neq i} I_{t-s_j}^{(j)}\right) + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = S_t + \sum_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$	$S_t = \alpha(X_t - \sum_i I_{t-s_i}^{(i)}) + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $I_t^{(i)} = \delta^{(i)}\left(X_t - S_t - \sum_{j\neq i} I_{t-s_j}^{(j)}\right) + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = S_t + kT_t + \sum_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$	$S_t = \alpha(X_t - \sum_i I_{t-s_i}^{(i)}) + (1-\alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\phi T_{t-1}$ $I_t^{(i)} = \delta^{(i)}\left(X_t - S_t - \sum_{j\neq i} I_{t-s_j}^{(j)}\right) + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = S_t + T_t \sum_{n=1}^{k}\phi^n + \sum_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$	$S_t = \alpha(X_t - \sum_i I_{t-s_i}^{(i)}) + (1-\alpha)(S_{t-1}R_{t-1})$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}$ $I_t^{(i)} = \delta^{(i)}\left(X_t - S_t - \sum_{j\neq i} I_{t-s_j}^{(j)}\right) + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = S_t R_t^k + \sum_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$	$S_t = \alpha(X_t - \sum_i I_{t-s_i}^{(i)}) + (1-\alpha)\left(S_{t-1}R_{t-1}^{\phi}\right)$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}^{\phi}$ $I_t^{(i)} = \delta^{(i)}\left(X_t - S_t - \sum_{j\neq i} I_{t-s_j}^{(j)}\right) + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = S_t R_t^{\sum_{j=1}^{k}\phi^j} + \sum_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$
	Damped Multiplicative	$S_t = \frac{\alpha X_t}{\prod_i I_{t-s_i}^{(i)}} + (1-\alpha)S_{t-1}$ $I_t^{(i)} = \delta^{(i)} \frac{X_t}{S_t \prod_{j\neq i} I_{t-s_j}^{(j)}} + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = S_t \prod_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$	$S_t = \frac{\alpha X_t}{\prod_i I_{t-s_i}^{(i)}} + (1-\alpha)(S_{t-1} + T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$ $I_t^{(i)} = \delta^{(i)} \frac{X_t}{S_t \prod_{j\neq i} I_{t-s_j}^{(j)}} + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = (S_t + kT_t) \prod_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$	$S_t = \frac{\alpha X_t}{\prod_i I_{t-s_i}^{(i)}} + (1-\alpha)(S_{t-1} + \phi T_{t-1})$ $T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)\phi T_{t-1}$ $I_t^{(i)} = \delta^{(i)} \frac{X_t}{S_t \prod_{j\neq i} I_{t-s_j}^{(j)}} + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = \left(S_t + T_t \sum_{n=1}^{k}\phi^n\right) \prod_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$	$S_t = \frac{\alpha X_t}{\prod_i I_{t-s_i}^{(i)}} + (1-\alpha)(S_{t-1}R_{t-1})$ $T_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}$ $I_t^{(i)} = \delta^{(i)} \frac{X_t}{S_t \prod_{j\neq i} I_{t-s_j}^{(j)}} + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = \left(S_t R_t^k\right) \prod_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$	$S_t = \frac{\alpha X_t}{\prod_i I_{t-s_i}^{(i)}} + (1-\alpha)\left(S_{t-1}R_{t-1}^{\phi}\right)$ $R_t = \gamma(S_t/S_{t-1}) + (1-\gamma)R_{t-1}^{\phi}$ $I_t^{(i)} = \delta^{(i)} \frac{X_t}{S_t \prod_{j\neq i} I_{t-s_j}^{(j)}} + (1-\delta^{(i)})I_{t-s_i}^{(i)}$ $\hat{X}_t(k) = \left(S_t R_t^{\sum_{j=1}^{k}\phi^j}\right) \prod_i I_{t-s_i+k}^{(i)} + \varphi_{AR}^k \varepsilon_t$

Although they share a common underlaying structure, differences are notable in the equations. S_t stands for the level equation. T_t stands for an additive trend smoothing equation, whereas R_t stands for a multiplicative one. This name has been chosen according to Pegels' classification. $I_t^{(i)}$ are the seasonal indices; there are as many indices as seasonalities considered. The superscript (i) stands for the seasonality of length s_i considered. α, β, and $\delta^{(i)}$ are the smoothing parameters. X_t are the observed values and $\hat{X}_t(k)$ are the k-ahead forecasted values. The factor ϕ stands for the damping factor when using a damping trend, regardless of it being additive or multiplicative. φ_{AR} is the factor of the adjustment with the first autocorrelation error. When the factor is null, the model does not include this adjustment. ε_t is the one-step-ahead forecasting error.

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. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. [CrossRef]
3. Holt, C.C. Forecasting seasonals and trends by exponentially weighted averages. In Proceedings of the O.N.R. Memorandum, Carnegie Institute of Technology, Pittsburgh 1957; Volume 52.
4. Winters, P.R. Forecasting sales by exponentially weighted moving averages. *Management* **1960**, *6*, 324–342.
5. Hong, T. Crystal Ball Lessons in Predictive Analytics. *Energybiz* **2015**, *12*, 35–37.
6. Taylor, J.W. Short-term electricity demand forecasting using double seasonal exponential smoothing. *J. Oper. Res. Soc.* **2003**, *54*, 799–805. [CrossRef]
7. Taylor, J.W. Triple seasonal methods for short-term electricity demand forecasting. *Eur. J. Oper. Res.* **2010**, *204*, 139–152. [CrossRef]
8. 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.
9. Makridakis, S.; Wheelwright, S.C.; Hyndman, R.J. *Forecasting: Methods and Applications*; John Willey and Sons, Inc.: Chichester, UK, 1998.
10. Granger, C.W.J.; Newbold, P. *Forecasting Economic Time Series*, 1st ed.; Academic Press Inc.: New York, NY, USA, 1977; ISBN 0-12-295150-6.
11. Holt, C.C. Forecasting seasonals and trends by exponentially weighted moving averages. *Int. J. Forecast.* **2004**, *20*, 5–10. [CrossRef]
12. Brown, R.G. *Statistical Forecasting for Inventory Control*; McGraw-Hill: New York, NY, USA, 1959.
13. Bowerman, B.L.; Koehler, A.B.; Pack, D.J. Forecasting time series with increasing seasonal variation. *J. Forecast.* **1990**, *9*, 419–436. [CrossRef]
14. Bowerman, B.L.; O'Connell, R.T.; Koehler, A.B. *Forecasting, Time Series, and Regression: An Applied Approach*; Thomson Brooks/Cole: Belmot, CA, USA, 2005; ISBN 0534409776.
15. Hyndman, R.J. Initializing the Holt–Winters Method. Available online: https://robjhyndman.com/hyndsight/hw-initialization/ (accessed on 10 August 2019).
16. Rasmussen, R. On time series data and optimal parameters. *Omega* **2004**, *32*, 111–120. [CrossRef]
17. Souza, R.C.; Barros, M.; Miranda, C.V.C. Short term load forecasting using double seasonal exponential smoothing and interventions to account for holidays and temperature effects. *Tlaio Ii-2 Do Taller Lat. Iberoam. Investig. Oper. Acapulco México* **2007**, 1–8.
18. Hyndman, R.J.; Koehler, A.B.; Ord, J.K.; Snyder, R.D. *Forecasting with Exponential Smoothing: The State Space Approach*; Springer, Ed.; Springer: Berlin, Germany, 2008; ISBN 978-3-540-71916-8.
19. 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]
20. Kalekar, P. Time series forecasting using Holt–Winters exponential smoothing. *Kanwal Rekhi Sch. Inf. Technol.* **2004**, 1–13.
21. Montgomery, D.C.; Jennings, C.; Kulahci, M. *Introduction to Time Series Analysis and Forecasting*; John Wiley & Sons, Inc.: New Jersey, USA, 2008; ISBN 978-0471653974.

22. Suppalakpanya, K.; Nikhom, R.; Booranawong, A.; Booranawong, T. An evaluation of Holt–Winters methods with different initial trend values for forecasting crude palm oil production and prices in Thailand. *Suranaree J. Sci. Technol.* **2019**, *26*, 13–22.

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

24. Makridakis, S.; Hibon, M. Exponential smoothing: The effect of initial values and loss functions on post-sample forecasting accuracy. *Int. J. Forecast.* **1991**, *7*, 317–330. [CrossRef]

25. Williams, D.W.; Miller, D. Level-adjusted exponential smoothing for modeling planned discontinuities. *Int. J. Forecast.* **1999**, *15*, 273–289. [CrossRef]

26. Granger, C.W.J.; Newbold, P. *Forecasting Economic Time Series*, 2nd ed.; Academic Press, Inc.: New York, NY, USA, 1986; ISBN 9780122951831.

27. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer: New York, NY, USA, 2002.

28. NIST/SEMATECH e-Handbook of Statistical Methods. Available online: http://www.itl.nist.gov/div898/handbook/ (accessed on 1 April 2012).

29. Torres, J.; Fernández, A.M.; Troncoso, A.; Martínez-Álvarez, F. Deep Learning-Based Approach for Time Series Forecasting with Application to Electricity Load. In *Biomedical Applications Based on Natural and Artificial Computing*; Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Toledo Moreo, J., Adeli, H., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 203–212.