Electricity price forecasting on electricity spot market: a case study based on the Brazilian Difference Settlement Price

Tiago Silveira Gontijo[1], *Marcelo* Azevedo Costa[1,2], *Rodrigo* Barbosa de Santis[1,*]

[1] Graduate Program in Industrial Engineering - Universidade Federal de Minas Gerais - Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG, Brazil.
[2] Department of Industrial Engineering - Universidade Federal de Minas Gerais - Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG, Brazil.

Abstract. Developing predictive models is a complex task since it deals with the uncertainty and the stochastic behavior of variables. Specifically concerning commodities, accurately predicting their future prices allows us to minimize risks and establish more reliable decision support mechanisms. Although the discussion on this question is extensive, there is academic attention being paid to the construction of nonparametric models applied to energy markets, as they have presented promising predictive results, what justifies the present study. This paper applies classical statistical models and Dynamic Time Scan Forecasting (DTSF) to the short-term electricity market prices, in Brazil, from 2006 to 2019. DTSF consists of scanning a time series and then identifying past patterns (so-called "matches"), similar to the last available observations. We predict Brazilian electricity spot prices, according the most similar matches, using aggregation functions, such as median. Recent research on the electricity spot market is increasing, indicating research significance. Our predictive approach exhibited greater accuracy than seminal statistical models. Our approach was designed for a high frequency series. Its predictive performance remained robust when other models presented both high predictive errors (spring), as well as when those models are highly accurate (winter). For future research, we recommend a more finely-tune study on DTSF parameters.

1 Introduction

The construction of predictive models arouses interest in the literature [1-4], since economic agents deal with uncertainty in multiple spheres and aim to achieve the best level of results from the available resources [5]. Therefore, developing models with acceptable accuracy and adequate rigor presents a meaningful challenge to researchers. The Professor George Box synthesized this scheme: "All models are wrong, but some are useful" [6]. In other words, prediction is a technique that deals with risk, and there will always be a

---

* Corresponding author: rsantis@ufmg.br

fundamental error associated with it. The best model is the one that adequately represents the phenomenon of interest.

In relation to the object of our study, electricity prices, there are several forecasting applications: (i) classical time series models like the autoregressive moving average, autoregressive integrated moving average, generalized autoregressive conditional heteroscedastic, among others [7-8]; (ii) pre-processing techniques, e.g., spectrum analysis, wavelets and Fourier analysis [9-10]; and, (iii) machine learning approaches such as neural networks, fuzzy systems and support vector machine [11]. Additionally, an alternative class known as hybrid models aims to combine machine learning representations with deferent methods. Instances of these methods are focused time-delay neural networks [12], neural networks with fuzzy inputs [13], finite-impulse response neural networks [14], local feedback dynamic fuzzy neural networks [15], type recurrent fuzzy networks [16], neuro-fuzzy inference systems [17], among others.

Although the literature on this question is extensive, there is academic interest in the construction of nonparametric models applied to energy markets, as they have presented promising predictive results. In general, these models are designed to deal with long-time series and are chiefly based on analog search [18] and scan-clustering methodologies [19-20].

According to this framework, understanding price behavior takes singular importance [21] since it allows the minimization of risk and uncertainty [22-23] and thereby to provide reliable production plans [24] and to establish a fine electricity market design [25]. In this scenario, it is noteworthy that in Brazil, for instance, the National System Operator defines the spot price of electricity weekly [26]. Predicting future values can help to establish enterprise decisions and behaviors. For example, if a firm expects that the energy price will scarcely increase, it could provisionally suspend part of its production and sell the surplus electricity in the spot market later at a higher price [27-28]. Also, the value of mapping the main conceptual theories used in the literature regarding the electricity spot market is noteworthy, as those theories evolve and become more robust over time [29-30].

The present paper contributes to the debate about the electricity spot market in two ways: (i) it illustrates how big data tools can produce relevant information about the electricity market; and, (ii) it illustrates, as a case study, a new forecasting approach, proposed by Costa et al. [20] to the analyzed market. Accordingly, we aim to review the literature on the electricity spot market, through the employment of big data tools, presenting the main research trends in the electricity spot market; also, to introduce a new, predictive instrument based on dynamic time scan, using the Brazilian difference settlement price as a case study.

The paper is organized as follows. Section 2 presents the methodology, the data retrieval, and the dynamic time scan forecasting procedure. Section 3 illustrates an application of our forecasting approach, applied to the Brazilian electricity market, as a case study. Finally, section 5 highlights some patterns observed in the literature review and the potential use of dynamic time scan forecasting for future studies focusing on the energy spot market.

2 Materials and Methods

This section first presents the data and the procedures used to carry out a review of literature. Second, it introduces a promising statistical model based on dynamic time scan

forecasting. To illustrate the accuracy of the model, and to present its applicability, the Brazilian "difference settlement price" dataset was used as a case study.

2.1 A review of the literature on the Electricity Spot Market

We conducted an exhaustive survey of publications regarding papers on the electricity spot market. It is noteworthy that Web of Science (WoS), along with Scopus, are the most used academic citation databases for delineating fields of study. WoS, formerly known as the ISI, Web of Knowledge database, was used as the most complete and traditional bank of scientific publications in the world [31-32]. We limited our search to papers published in English. Moreover, we considered only the WoS database, since the WoS and Scopus databases may not differ significantly [30]. In addition, we chose to analyze only the core publication journals, which are usually indexed in both databases.

Data extraction from the WoS platform (2019-05-06) took into account publications that responded to the search expression "electricity spot market*", as shown in Figure 1. The use of the "*" character in the search expression helps to capture words in both the singular and in the plural, making search results more complete.

The research universe encompasses publications from the years 1992 to 2018 (n = 173). 1992 is the year of the first scientific publication on the subject indexed in WoS. 2018 is the last year for which complete information is available. The sample for this project (1992 to 2018, n = 97) is henceforth called "primary publications". To filter these primary publications, we scanned for any of the descriptors shown in Figure 1 in the titles, abstracts, author's keywords, or in the Plus keywords.
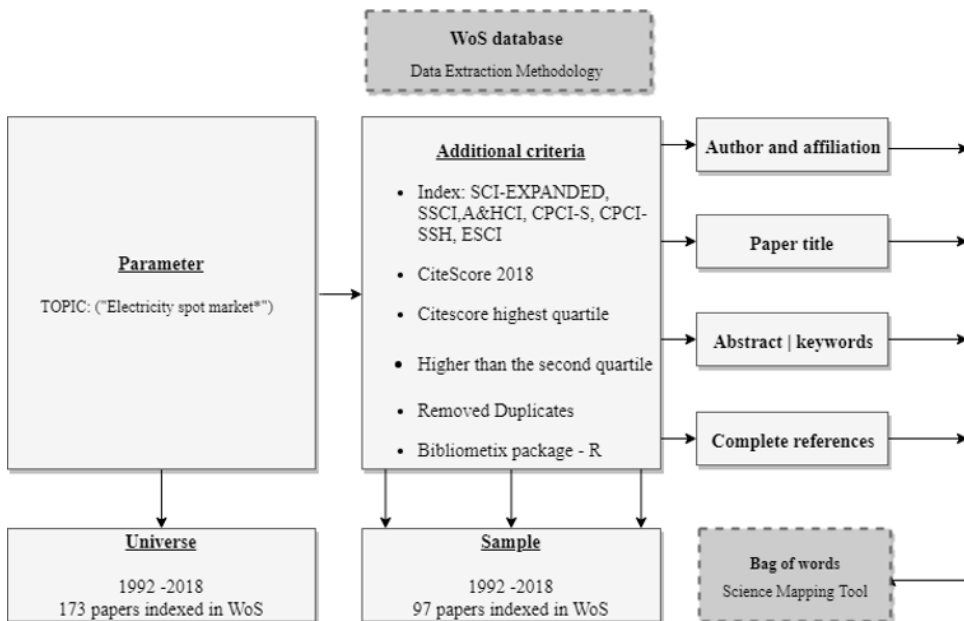

Fig. 1. Methodological procedure to obtain the analyzed sample
Note: CiteScore, SNIP and SJR metrics calculated using data from 30 April 2018.

Criteria for including a paper in the sample were based on these indicators: (i) CiteScore (measures average citations received per document published in the serial); (ii) SCImago

Journal Rank (SJR) (measures weighted citations received by the serial. Citation weighting depends on subject field and prestige, SJR of the citing serial); and, (iii) Source Normalized Impact per Paper (SNIP) (measures actual citations received relative to citations expected for the serial's subject field), all of them for the year of 2018. The present study considered only the periodicals rated above the second quartile for each of the three criteria.

After selecting journals, we conducted a descriptive analysis of our bibliographic data frame. To do this, we used the "Bibliometrix R-package" developed by Aria & Cuccurullo [33] to analyze the annual publication of electricity spot markets in the most relevant journals, as well as the most productive countries of corresponding authors and seminal works. Next, we developed a dictionary that aggregates similar words. For example, we considered "electricity prices" and "electricity pricing", or "electricity spot market" and "electricity trading", as equivalent expressions. Finally, we constructed a WordCloud using the VOSviewer software [34], that filters the hot and cold areas of interest for the theme according to the co-citation keywords, showing in detail how they interact with each other.

2.2 Data retrieval

The present study is based on the Brazilian difference settlement price (DSP). The DSP is determined weekly, considering three load levels for each sub-market. The submarkets are defined by the National Operator of the System, and consider the following geographical divisions: North, Northeast, Southeast/Center-West and South.

The DSP is determined ex-ante (considering expected availability and load information), based on weeks counted from Saturday to Friday. The prices must sell out all the energy, not just the contracted energy, among the agents [35]. The analyzed sample consists of weekly DSP data (R\$/MWh) collected by the Chamber of Electric Energy Commercialization, from January 2006 to May 2019 (n = 701 weeks). We utilized the twelve available series, divided into the four Brazilian sub-markets: North, Northeast, Southeast/Midwest and South (called, respectively, N, NE, SE and S); and three load levels of energy: Heavy, Average, and Light (called, respectively, P, M and L)[†].

2.3 Dynamic time scan forecasting methodology and benchmark comparison

Dynamic time scan forecasting (DTSF) is an R-package based on scan statistics [36]. It was originally formulated to deal with wind forecasting and power generation by industrial plants [20]. It consists of scanning a time series and then identifying past patterns (so called "matches") similar to the last available observations. Future values are predicted from the most similar matches using aggregation functions, such as median. Mathematical formulation and the results of original applications can be found in Costa et al. [20].

Here, we present an innovative use of DTSF based on the electricity spot market. To present the method, we utilize a predictive window equal to four weeks, the number of matches as the ten best-correlated ones and, finally, a polynomial function of order one (Figure 2). In this way, it is possible to find patterns between last available data and the old information, without any statistical test (green highlights). This procedure is based only on

---

[†]According to this, we have: L_N; L_{NE}; L_S; L_{SE}; M_N; M_{NE}; M_S; M_{SE}; P_N; P_{NE}; P_S and P_{SE}.

high similarity statistics (R²). The most similar patterns, called "median value of matches," allow us to predict the DSP with the highest level of accuracy.
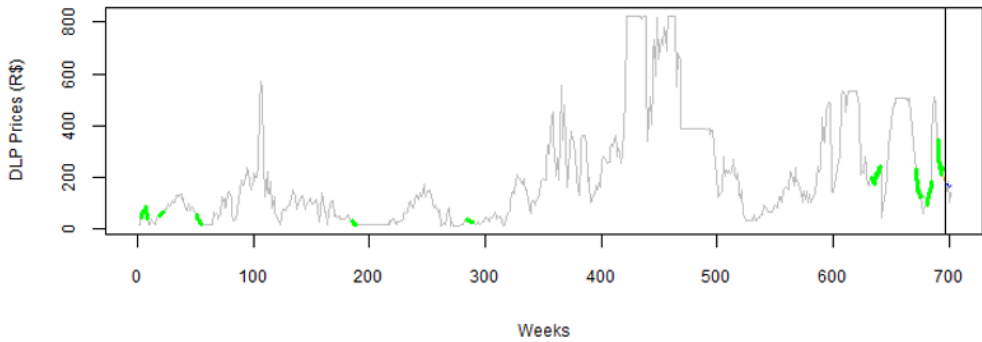

Fig. 2. Time series of the DSP (P-SE) and ten best matches found using DTSF.

To illustrate the DTSF, we compare it to the classical statistical methods utilized in the "Makridakis Competitions" (also known as the M Competitions or M-Competitions). These competitions are a series of open disputes organized by the Professor Spyros Makridakis to evaluate and compare the accuracy of different forecasting methods [37-38]. Table 1 provides our benchmark comparison.

Methods	Initials	Description
Statistical benchmarks	Naïve 1	A random walk model; future values will be the same as that of the last known observation.
	Naïve S	Forecasts are equal to the last known observation of the same period.
	Naïve 2	Like Naïve 1 but the data are seasonally adjusted, if needed, by applying a classical multiplicative decomposition. A 90% autocorrelation test is performed to decide whether the data are seasonal.
	SES	Exponentially smoothing the data and extrapolating assuming no trend. Seasonal adjustments are considered as per Naïve 2.
	Holt	Exponentially smoothing the data and extrapolating assuming a linear trend. Seasonal adjustments are considered as per Naïve 2.
	Damped	Exponentially smoothing the data and extrapolating assuming a damped trend. Seasonal adjustments are considered as per Naïve 2.
	Theta	As applied to the M3 Competition using two Theta lines, $\Theta 1 = 0$ and $\Theta 2 = 2$, with the first one being extrapolated using linear regression and the second one using SES. The forecasts are then combined using equal weights. Seasonal adjustments are considered as per Naïve 2.
	Comb	The simple arithmetic average of SES, Holt and Damped exponential smoothing (used as the single benchmark for evaluating all other methods).
Scan-clustering analysis	DTSF	Based on scan-clustering, parametrized with a predictive window equals to four week, assuming the number of matches as the ten best correlated ones and a polynomial function of order one.

Table 1- DTSF, benchmarks, and standards for comparison of the M4 Competition
Source: Makridakis, Spiliotis & Assimakopoulos [39], adapted by the authors.

The test set has a fixed horizon for all of the time series, and it is used only to compute the final scores. We utilized M4 competition accuracy metrics, based on the Overall Weighted Average (OWA) of two accuracy measures: The Mean Absolute Scaled Error (MASE, Eq.

01) and the Symmetric Mean Absolute Percentage Error (sMAPE, Eq. 02), since they are among those most found in the literature [40-41].

$$sMAPE = \frac{1}{h}\sum_{t=1}^{h} \frac{2|Y_t - \hat{Y}_t|}{(|Y_t| + |\hat{Y}_t|)} \qquad\qquad (Eq.\ 01)$$

$$MASE = \frac{1}{h} \frac{\sum_{t=1}^{h}|Y_t - \hat{Y}_t|}{\frac{1}{n-m}\sum_{t=m+1}^{n}|Y_t - Y_{t-m}|} \qquad\qquad (Eq.\ 02)$$

where Y_t is the post sample value of the time series at point t, $\hat{Y}_t$ the estimated forecast, h the forecasting horizon and m the frequency of the data.

3 Results and Discussion

3.1 Academic research on electricity spot market

Research on electricity spot markets began with Green & Newbery [42], who analyzed the competition in the British market and demonstrated the Nash equilibrium in supply schedules, implying a high markup on marginal costs and substantial deadweight losses. However, there was a notable gap of seven years until the next three papers were published [43-45]. The first research [43] studied transmission efficiency in the western US. The second and the third studies [44-45] focused on measuring duopoly power in the British electricity spot market and the reform of electricity trading in England and Wales, respectively.

In the period between 1992 and 2018, 215 authors published in leading selected journals (n = 25) indexed in WoS, totaling (n = 97) publications. Here, we have additional information for our sample: (i) Author's Keywords (included in records of papers by the authors) (n = 299); (ii) Keywords Plus (index terms automatically generated by WoS, considering the titles of cited papers) (n = 253); (iii) all keywords (n = 432) (the sum of Author's Keywords and Keywords Plus, excluding duplicates); and, (iv) Average citations per documents (n = 45.54). During the survey period, the mean annual growth of publications was 8.33% (from 1 paper in 1992 to 8 papers in 2018). Additionally, an average of 4.62±2.78 papers was published per year, demonstrating academic relevance.

Seminal works on the energy spot market are concentrated, according to Bradford's Law [46], in a few journals, namely: Energy Economics (n = 20), Energy Policy (n = 19) and IEEE Transactions on Power Systems (n = 14). This growth has aroused the interest of several researchers and favors the emergence of some questions, namely: (i) which theoretical approaches have shown the best results in describing electricity spot market? (ii) what are the potential gaps in the literature that future works should address?

The leading countries in publications are the USA (n = 31), followed by Germany (n = 23) and Spain (n = 21). Together, they account for 41.44% of the world's research on the electricity spot markets. On the other hand, the countries with the highest total number of citations are, respectively, the USA (n = 1127), Spain (1042) and the United Kingdom (n = 993). Last, but not least, Ireland (n = 101), the United Kingdom (n = 90.27) and Spain (n = 86.83) are the three countries with the highest average number of citations per paper, which is a good indicator of the degree of academic relevance.

The principal references, cited in the 97 publications, are, respectively: (i) Green & Newbery [42]. Richard Green is Professor of Sustainable Energy Business at the Imperial College London; David Newbery is a professor at the University of Cambridge; (ii) Klemperer [47]. Paul Klemperer is an economist and Professor of Economics at Oxford University; and, (iii) Arroyo & Conejo [48]. José M. Arroyo is Professor at the Universidad de Castilla-La Mancha; Antonio J. Conejo is Professor at Ohio State University.

It is possible to create maps based on bibliographic information and then explore, for instance, the degree of appearance of a particular keyword and its temporal evolution. This type of analysis allows us to verify the hot and cold areas of interest in the scientific debate on the electricity spot market (Figure 3). Our purpose is to take out the frequency of papers with certain keywords. We set our "breakpoint" (n = 2) as the minimum number of occurrences considering all the keywords (n = 432). As it is possible to see, below, 37 keywords (nodes) meet that threshold. The diameter of the circle is proportional to the number of occurrences of, and the lines show the number of links between, these keywords. In recent research, some topics have gained prominence: (i) non-parametric models; (ii) impact analysis; (iii) merit order effect; (iv) virtual power plant; and, (v) electric vehicles.

Figure 3 highlights a defined trajectory of research over time. The first published research (in purple) emphasized economic aspects such as the supply of energy, as well as technical elements (capacity and transmission, e.g.) concerning spot energy markets. Following this, there was a growing importance of studies on electricity generation and pricing (in blue), studies on contract theory, energy storage, and technologies (in green). Finally, the areas of the most current interest are those linked to non-parametric models, impact analysis, merit order effect, virtual power plant, and electric vehicles.
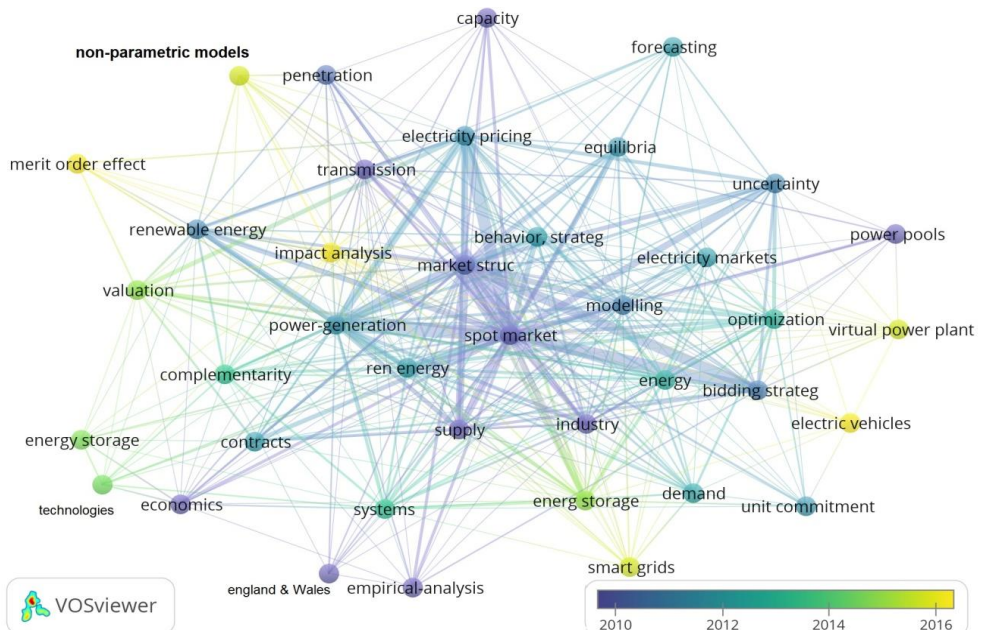

Fig. 3. Temporal WordCloud for the 50 most repeated keywords (all) in the analyzed sample utilizing VOSviewer software [34].

Note: This WordCloud considers only articles with a maximum number of authors per document (n=25). The inclusion criterion for a single word is that it appear at least (n=2) times in different papers' keywords.

Let us look at the non-parametric node. This node has been used in papers after 2016, which indicates a possible trend for further research. Studies that have used the keyword "non-parametric models" have also used other keywords, namely: (i) behavior strategies, (ii) complementarity, (iii) electricity pricing, (iv) impact analysis, (v) market structure, (vi) penetration, (vii) power generation, (viii) renewable energy, (ix) spot market, and (x) transmission. Thus, our paper relates to current research trends on electricity and innovates by proposing the first application of the nonparametric model (DTSF) for forecasting energy prices in the spot market.

The main articles that addressed nonparametric models applied to energy focused on: (i) assessing the influence of high penetration of wind power on the market-splitting behavior between West and East Denmark, using logit and non-parametric models [49]; and, (ii) expressing the probability response for market-splitting of day-ahead spot electricity prices as a function of the explanatory variables representing the main technologies in the generation mix, including wind, hydro, thermal and nuclear power, together with the available transfer capacity and electricity demand [50].

3.2 Statistical analysis of difference settlement price time series (DSP)

In order to test the hypothesis of equality of DSP prices, among groups 1 to 12 (PSE, MSE, LSE, PS, MS, LS, PNE, MNE, LNE, PN, MN and LN) ($H_0 = \mu_1 = \mu_2 = \cdots = \mu_{12}$), we performed an Analysis of Variance (ANOVA) (Table 2). The initial objective is to identify the existence of at least one difference among the prices in the DSP groups. As the P-value for the F statistic is significant, there is evidence that at least one difference among the average prices would be significant.

Table 2 - Analysis of Variance (ANOVA): response of DSP prices for the analyzed groups.

Response: Prices					
Degrees of Freedom	Sum of Squares	Mean of Squares	F	P-value	Pr(>F)
11	976023	88729	2.299	0.008	**
8400	324076244	38581			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

To check whether the results obtained in the ANOVA are satisfactory, it is recommended that the following conditions be tested: (i) are the standard deviations of groups constant over time? (ii) is the response distribution normal for each group? To test for homoscedasticity, we performed the Bartlett Test [51], and obtained a P-value equal to 0.448. To detect normality, we performed the Shapiro test [52], which gave a P-value of 0.501. The non-normality of these results corroborates findings previously reported in the literature [53]. Furthermore, since the DTSF is based on a non-parametric model, our goal was not to perform statistical inference analysis.

From this evidence, it is interesting to identify where the differences among the means of the analyzed groups lie. In particular, the process of comparing means examines them two by two, using the Multiple Comparisons method proposed by Tukey [54] (Figure 4). In

general, most of the series have equal mean values. This reinforces the thesis that the DSP behavior is stable concerning the charge amount and the associated geographic region. Additionally, at the 0.1 level of significance of, the means of the DSP prices differ statistically from the others only in the following groups: M_{SE}-L_N, P_{NE}-L_N and P_{SE}-L_N.

In order to test the hypothesis of equality of DSP prices, among groups 1 to 12 (P_{SE}, M_{SE}, L_{SE}, P_S, M_S, L_S, P_{NE}, M_{NE}, L_{NE}, P_N, M_N and L_N) ($H_0 = \mu_1 = \mu_2 = \cdots = \mu_{12}$), we performed an Analysis of Variance (ANOVA) (Table 2). The initial objective is to identify the existence of at least one difference among the prices in the DSP groups. As the P-value for the F statistic is significant, there is evidence that at least one difference among the average prices would be significant.
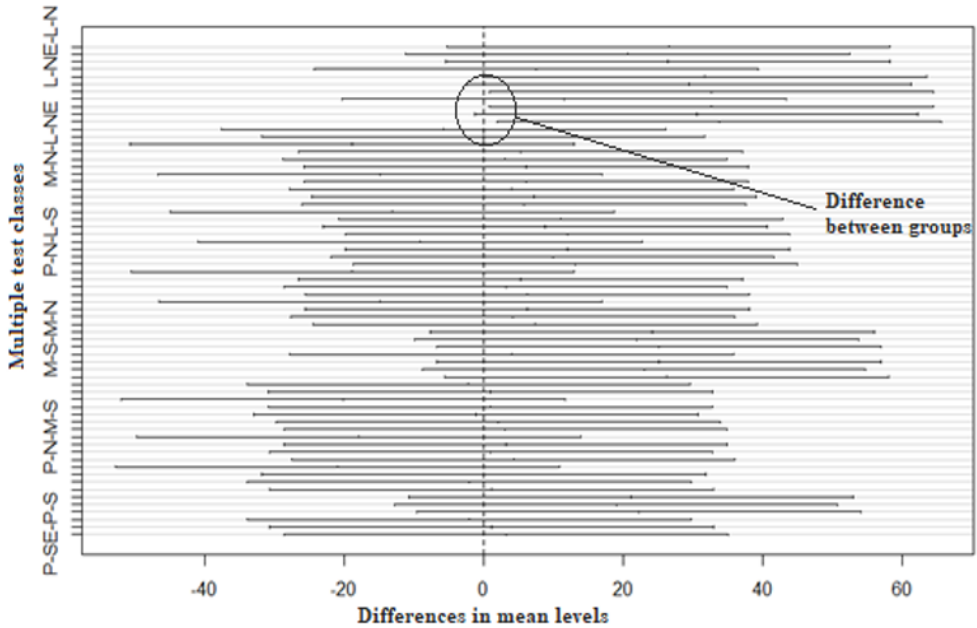

Fig. 4. Tukey's Honestly Significant Difference (HSD) test.

In general, the main differences are centered in the north of Brazil. This is reasonable, since this region is not fully connected to the national electricity system and, thus, has specific characteristics [55].

3.3 Case study: applying DTSF to the heavy southeast energy submarket

According to the statistical evidence presented in section 3.2, there is robust indication of a similar pattern among the mean values for the twelve analyzed time series. Thus, we assume that the DTSF approach will generate similar forecastings for each group. Given that, our case study is based only on the DSP in the southeast of Brazil, and considers the heavy load level of energy. This choice is justified because the southeast is the wealthiest region of the country and concentrates nearly 55% of the country's gross development product [56]. Lastly, we kept the heavy load level as it relates to industrial energy demand [57].

As expressed in the Material and Methods (section 2.3), the new predictive DTSF model is faced to the eight "seminal statistical models" (classified here as the benchmark models), frequently used in the technical literature on forecasting competitions, such as the classical one "M4 Competition" [37-38].

There is no evidence of seasonal influence in the time series analyzed, since the variations of the Naïve method resulted in equivalent predictive metrics (Table 3). We also observed that the benchmark models presented similar results. On the other hand, according to the OWA, the DTSF method presented better performance than the Naïve (baseline) model. DTSF presents an efficient resolution for the Brazilian DSP prediction in the southeast region, considering the heavy load level of energy. Figure 5 graphically illustrates the fit of the models according to the sMAPE and MASE statistics.

Table 3 - Comparison between the DTSF and the analyzed forecasting models (benchmark).

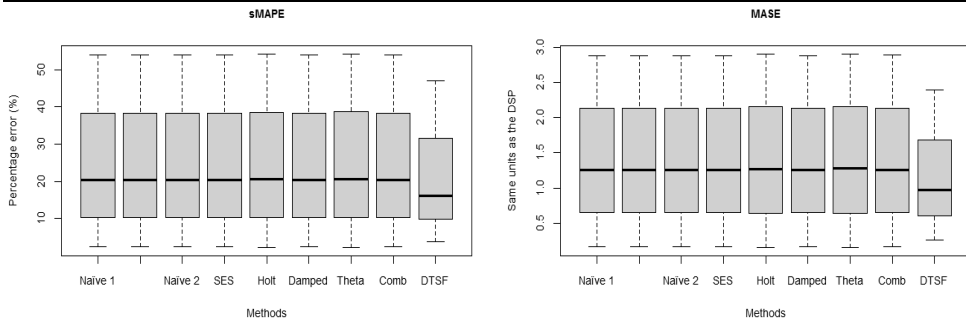| Method	Indicators		
	sMAPE	MASE	OWA
Naïve 1	24.266	1.389	1.000
Naïve S	24.266	1.389	1.000
Naïve 2	24.266	1.389	1.000
SES	24.267	1.389	1.000
Holt	24.415	1.399	1.007
Damped	24.261	1.388	1.000
Theta	24.453	1.402	1.009
Comb	24.314	1.392	1.002
DTSF	**20.729**	**1.148**	**0.840**

Fig. 5. Comparison between the DTSF and the analyzed models, considering sMAPE and MASE.

Nevertheless, DTSF should be examined in more detail. In Figure 2 (section 2.3) we presented the time series utilized and, according to that time series, the discontinuous nature of the observations over the weeks is evident. This occurs for different reasons, ranging from political influence on the energy pricing process [58] to theoretical limitations in defining the difference settlement price in Brazil [59].

To strengthen our predictive study, we divided the time series according to the seasons (spring, summer, autumn, and winter). Following this division, the predictive potential of the models was re-examined (Table 4). In general, the best performances were, respectively: (i) spring (DTSF); (ii) summer (Naïve and Damped); (iii) autumn (Naïve) and (iv) winter (Damped).

Table 4 - Comparisons between the DTSF and the analyzed forecasting models.

Seasons	Method	Indicators		
		sMAPE	MASE	OWA
Spring	Naïve 1	45.898	1.149	1.000
	Naïve S	45.898	1.149	1.000
	Naïve 2	45.898	1.149	1.000
	SES	45.949	1.150	1.001
	Holt	45.756	1.144	0.996
	Damped	45.935	1.150	1.001
	Theta	47.993	1.221	1.054
	Comb	45.880	1.148	1.000
	DTSF	**41.727**	**1.013**	**0.896**
Summer	Naïve 1	**37.220**	**2.645**	**1.000**
	Naïve S	**37.220**	**2.645**	**1.000**
	Naïve 2	**37.220**	**2.645**	**1.000**
	SES	37.224	**2.645**	**1.000**
	Holt	38.329	2.743	1.034
	Damped	37.224	**2.645**	**1.000**
	Theta	37.617	2.680	1.012
	Comb	37.594	2.678	1.011
	DTSF	63.340	5.435	1.878
Autumn	Naïve 1	**24.266**	**1.060**	**1.000**
	Naïve S	**24.266**	**1.060**	**1.000**
	Naïve 2	**24.266**	**1.060**	**1.000**
	SES	24.443	1.070	1.008
	Holt	25.248	1.114	1.046
	Damped	24.437	1.070	1.008
	Theta	25.380	1.122	1.052
	Comb	24.712	1.085	1.021
	DTSF	28.745	1.327	1.218
Winter	Naïve 1	0.901	0.149	1.000
	Naïve S	0.901	0.149	1.000
	Naïve 2	0.901	0.149	1.000
	SES	0.901	0.149	1.000
	Holt	3.081	0.517	3.447
	Damped	**0.862**	**0.142**	**0.957**
	Theta	1.560	0.259	1.737
	Comb	1.614	0.268	1.797
	DTSF	0.901	0.149	1.000

It is noteworthy that our results are generally consistent. DTSF, based on the scanning method, exhibited predictive worsening due to the time series cutoff. Nevertheless, even with the cutoff, it was more statistically accurate in the spring and winter forecasts. Although all models have high predictive errors (spring), DTSF produces better results. The same goes for the inverse situation: whenever models are highly accurate (winter), DTSF is among the best predictors. Finally, due to the high discontinuity pattern of the analyzed series, the Naïve model presented good overall performance.

Our results, although early, shed light on some future research possibilities such as detailed examination of DTSF performance in other global markets, such as in the Nordic countries. These countries have a sturdy energy pricing system, "Nord Pool" [60], that use hourly pricing information for electricity. Thus, the available information exceeds 50,000 observations for each market, (country) which allows it to produce a bright body of investigation. Last, but not least, we suggest the examination of different DTSF parameters through a grid search inquiry. Also, we recommend testing the precision of models by cutting time series into hundreds of subsamples and then verifying the predictive power of DTSF under different circumstances.

4 Conclusions

Analysis of the electricity spot market is complicated, involving the relationship between forecasting models and uncertainty, distinctly regarding the stochastic behavior of variables. The National System Operator regulates the Brazilian spot price of electric energy, which weekly discloses a new electric energy price to minimize the operating cost of the whole system. The present paper is aimed at the policymakers, offering a big data analysis of the scientific research of electricity. It also proposes a new forecasting approach, based on Scan-clustering modelling of the buying and selling of energy in future markets.

Although still nascent, research on the electricity spot market is increasing. As proof of this, the annual growth of publications during the survey period (1992 to 2018) corresponds to 8.33%, with an average of 4.62±2.78 papers published per year. This indicates the growing significance of this research. There may be several reasons for this, including the development of technologies or the growing interest in intelligent and automated networks by agents. Given the above, we noted that recent research shown a interest in some specific study domains, such as non-parametric models.

We present an application of DTSF based on the electricity spot market and apply our results to the eight "seminal statistical models" regularly used in forecasting competitions. In general, the eight benchmark models presented similar results. However, DTSF exhibited the best performance, as measured by all accuracy metrics. Subsequently, we examined the accuracy of the DTSF against timeframes in the observations, considering the seasons. DTSF was designed for series with high frequency and its predictive performance remained high when all other models had high predictive errors (spring). Even when other models were highly accurate (winter), DTSF remained among the best ones. There were other cases, however, in which we observed the DTSF performance to falter.

For future research, we recommend fine-tuning the study of DTSF parameters, as well as testing the accuracy of the models by randomly splitting time series into subsamples. Further, we recommend the development of case studies that will take the pricing of the electricity spot market in different regions, such as the Nordic and Iberian countries that dispose of massively available, high frequency data, into consideration.

Finally, the present research contributes to the energy planning processes of different players, given that understanding of the price patterns has singular importance in minimizing risks and supporting reliable production planning. Good forecasts for future energy pricing can support operational arrangements, e.g., when the energy price is high, it may be more valuable for an industry temporarily to delay part of its production, trade the surplus electricity, and carry out preventive maintenance on machines and accessories.

Acknowledgments

Nomenclature

Variables

L_N	light load level of energy to the nort market
L_{NE}	light load level of energy to the northeast market
L_S	light load level of energy to the south market
L_{SE}	light load level of energy to the southeast market
P_N	heavy load level of energy to the nort market
P_{NE}	heavy load level of energy to the northeast market
P_S	heavy load level of energy to the south market
P_{SE}	heavy load level of energy to the southeast market
M_N	medium load level of energy to the nort market
M_{NE}	medium load level of energy to the northeast market
M_S	medium load level of energy to the south market
M_{SE}	medium load level of energy to the southeast market

Acronyms

DSP	difference settlement price
ONS	national system operator
SJR	Scientific Journal Rankings
WoS	web of science

Accuracy

MASE	mean absolute scaled error
sMAPE	symmetric mean absolute percentage error
OWA	overall weighted average

Algorithms

DTSF	dynamic time scan forecasting

References

1. Hamm, G., & Borison, A. Forecasting long-run electricity prices. The electricity journal, **19(7)**, 47-57 (2006)
2. Kuhn, M. Building predictive models in R using the caret package. Journal of statistical software, **28(5)**, 1-26 (2008)
3. Bui, D. T., Tuan, T. A., Klempe, H., Pradhan, B., & Revhaug, I. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides, **13(2)**, 361-378 (2016)
4. Geisser, S. Predictive inference. Routledge (2017)

5. Choi, Y. B. Paradigms and conventions: Uncertainty, decision making, and entrepreneurship. University of Michigan Press (1993)

6. Box, G. E. Science and statistics. Journal of the American Statistical Association, **71(356)**, 791-799 (1976)

7. Pappas, S. S., Ekonomou, L., Karamousantas, D. C., Chatzarakis, G. E., Katsikas, S. K., & Liatsis, P. Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models. Energy, **33(9)**, 1353-1360 (2008)

8. Liu, H., & Shi, J. Applying ARMA–GARCH approaches to forecasting short-term electricity prices. Energy Economics, **37**, 152-166 (2013)

9. Simonsen, I. Measuring anti-correlations in the Nordic electricity spot market by wavelets. Physica A: Statistical Mechanics and its applications, **322**, 597-606 (2003)

10. Miranian, A., Abdollahzade, M., & Hassani, H. Day-ahead electricity price analysis and forecasting by singular spectrum analysis. IET Generation, Transmission & Distribution, **7(4)**, 337-346 (2013)

11. Chen, X., Dong, Z. Y., Meng, K., Xu, Y., Wong, K. P., & Ngan, H. W. Electricity price forecasting with extreme learning machine and bootstrapping. IEEE Transactions on Power Systems, **27(4)**, 2055-2062 (2012)

12. Chen, Y., He, Z., Shang, Z., Li, C., Li, L., & Xu, M. A novel combined model based on echo state network for multi-step ahead wind speed forecasting: A case study of NREL. Energy conversion and management, **179**, 13-29 (2019)

13. Liu, H., Tian, H. Q., Liang, X. F., & Li, Y. F. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Applied Energy, **157**, 183-194 (2015).

14. Pir, M. Y., Shah, F. A., & Asger, M. Comparative study of different wavelet based neural network models for IIP growth forecasting using different yield spreads. International Journal of Electrical Electronics & Computer Science Engineering, **4(6)**, 5-13 (2017)

15. Nagaraja, Y., Devaraju, T., Kumar, M. V., & Madichetty, S. A survey on wind energy, load and price forecasting: (Forecasting methods). In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 783-788). IEEE (2016, March)

16. Jain, L. C., Seera, M., Lim, C. P., & Balasubramaniam, P. A review of online learning in supervised neural networks. Neural computing and applications, **25(3-4)**, 491-509 (2014)

17. Moreno, S. R., & dos Santos Coelho, L. Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro fuzzy inference system. Renewable energy, **126**, 736-754 (2018)

18. Yang, D., & Alessandrini, S. An ultra-fast way of searching weather analogs for renewable energy forecasting. Solar Energy, **185**, 255-261 (2019)

19. Simmhan, Y., & Noor, M. U. Scalable prediction of energy consumption using incremental time series clustering. In 2013 IEEE International Conference on Big Data (pp. 29-36). IEEE (2013, October)

20. Costa M. A., Mineti L. B., Prates M. O., Cardenas, R. R. Dynamic Time ScanForecasting. arXiv e-prints. Jun;:arXiv:1906.05399 (2019)

21. Rostamnia, N., & Rashid, T. A. Investigating the effect of competitiveness power in estimating the average weighted price in electricity market. The Electricity Journal, **32(8)**, 106628 (2019)

22. Khosravi, A., Nahavandi, S., Creighton, D., & Naghavizadeh, R. Uncertainty quantification for wind farm power generation. In The 2012 International Joint Conference on Neural Networks (IJCNN) (pp. 1-6). IEEE (2012, June)

23. Heck, N., Smith, C., & Hittinger, E. A Monte Carlo approach to integrating uncertainty into the levelized cost of electricity. The Electricity Journal, **29(3)**, 21-30 (2016)

24. Schuh, G., Prote, J. P., Sauermann, F., & Franzkoch, B. Databased prediction of order-specific transition times. CIRP Annals (2019).

25. Woo, C. K., & Zarnikau, J. A nice electricity market design. The Electricity Journal, **32(9)**, 106638 (2019)

26. Resende, L., Soares, M., & Ferreira, P. Electric power load in Brazil: view on the long-term forecasting models. Production, **28** (2018)

27. Ioakimidis, C. S., Oliveira, L. J., & Genikomsakis, K. N. Wind power forecasting in a residential location as part of the energy box management decision tool. IEEE Transactions on Industrial Informatics, **10(4)**, 2103-2111 (2014).

28. Tian, P., Xiao, X., Wang, K., & Ding, R. A hierarchical energy management system based on hierarchical optimization for microgrid community economic operation. IEEE Transactions on Smart Grid, **7(5)**, 2230-2241 (2015)

29. Weron, R. Modeling and forecasting electricity loads and prices: A statistical approach (Vol. 403). John Wiley & Sons (2007)

30. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. International journal of forecasting, **30(4)**, 1030-1081 (2014)

31. Strozzi, F., Colicchia, C., Creazza, A., & Noè, C. Literature review on the 'Smart Factory' concept using bibliometric tools. International Journal of Production Research, **55(22)**, 6572-6591 (2017)

32. Analytics, C. Web of science core collection. Citation database. Web of Science. Retrieved from https://www.webofknowledge.com/WOS (2017)

33. Aria, M. & Cuccurullo, C. Bibliometrix: An R-tool for comprehensive science mapping analysis, Journal of Informetrics, **11(4)**, pp 959-975, Elsevier (2017)

34. Van Eck, N., & Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, **84(2)**, 523-538 (2009)

35. Chamber of Electric Energy Commercialization. DSP Historical Prices. Retrieved from https://www.ccee.org.br (2019)

36. Glaz, J., & Balakrishnan, N. (Eds.). Scan statistics and applications. Springer Science & Business Media (2012)

37. Makridakis, S., & Hibon, M. The M3-Competition: results, conclusions and implications. International journal of forecasting, **16(4)**, 451-476 (2000)

38. Hyndman, R. J. A brief history of forecasting competitions (No. 3/19). Monash University, Department of Econometrics and Business Statistics (2019)

39. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, **36(1)**, 54-74 (2020)

40. Al-Alawi, S. M., & Islam, S. M. Principles of electricity demand forecasting. I. Methodologies. Power Engineering Journal, **10(3)**, 139-143 (1996)

41. Azadeh, A., Ghaderi, S. F., & Sohrabkhani, S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. Energy Conversion and management, **49(8)**, 2272-2278 (2008)

42. Green, R. J., & Newbery, D. M. Competition in the British electricity spot market. Journal of political economy, **100(5)**, 929-953 (1992)

43. De Vany, A. S., & Walls, W. D. Cointegration analysis of spot electricity prices: insights on transmission efficiency in the western US. Energy Economics, **21(5)**, 435-448 (1999)

44. Wolfram, C. D. Measuring duopoly power in the British electricity spot market. American Economic Review, **89(4)**, 805-826 (1999)

45. Green, R. Draining the Pool: the reform of electricity trading in England and Wales. Energy Policy, **27(9)**, 515-525 (1999)

46. Brookes, B. C. Bradford's law and the bibliography of science. Nature, **224(5223)**, 953-956 (1969).

47. Klemperer, P. What really matters in auction design. Journal of economic perspectives, **16(1)**, 169-189 (2002)

48. Arroyo, J. M., & Conejo, A. J. Optimal response of a thermal unit to an electricity spot market. IEEE Transactions on power systems, **15(3)**, 1098-1104 (2000)

49. Figueiredo, N. C., da Silva, P. P., & Cerqueira, P. A. It is windy in Denmark: Does market integration suffer?. Energy, **115**, 1385-1399 (2016).

50. Figueiredo, N. C., da Silva, P. P., & Cerqueira, P. A. Evaluating the market splitting determinants: evidence from the Iberian spot electricity prices. Energy Policy, **85**, 218-234 (2015)

51. Bartlett, M. S. Properties of sufficiency and statistical tests. Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences, **160(901)**, 268-282 (1937)

52. Shapiro, S. S., & Wilk, M. B. An analysis of variance test for normality (complete samples). Biometrika, **52(3/4)**, 591-611 (1965)

53. Veloso, R. G. Real Options Model to Evaluate the Production Strategy in a Electricity Industry vs. Electricity Price (Doctoral dissertation). Federal University of Minas Gerais, Belo Horizonte, Brazil (2015)

54. Tukey, J. W. Comparing individual means in the analysis of variance. Biometrics, **5(2)**, 99-114 (1949)

55. Böckler, L., & Pereira, M. G. Consumer (Co-) Ownership in Renewables in Brazil. In Energy Transition (pp. 535-557). Palgrave Macmillan, Cham (2019)

56. Santos, E. S. D., Juchem, L., & Maduro, L. A. R. Performance sport, tax waiver and sports incentive law. Journal of Physical Education, **28** (2017)

57. Munhoz, F. C. The necessity of more temporal granularity in the Brazilian short-term electricity market. Sustainable Energy, Grids and Networks, **11**, 26-33 (2017)

58. Guimarães, C., & Piefer, N. Brazil: (Future) green energy power and strategic partner for the EU?. Comparative European Politics, **15(1)**, 23-44 (2017)

59. Gomes, R., & Poltronieri, R. The Electricity Sector and the Structure of the Short-Term Market in Brazil. In Energy Law and Regulation in Brazil (pp. 113-135). Springer, Cham (2018)

60. Botterud, A., Kristiansen, T., & Ilic, M. D. The relationship between spot and futures prices in the Nord Pool electricity market. Energy Economics, **32(5)**, 967-978 (2010)