Dynamic Time Scan Forecasting: A Benchmark With M4 Competition Data

Rodrigo Barbosa de Santis, Tiago Silveira Gontijo, *Reviewer, IEEE* and Marcelo Azevedo Costa, *Reviewer, IEEE*

***Abstract*—Univariate forecasting methods are fundamental for many different application areas. M-competitions provide important benchmarks for scientists, researchers, statisticians, and engineers in the field, for evaluating and guiding the development of new forecasting techniques. In this paper, the Dynamic Time Scan Forecasting (DTSF), a new univariate forecasting method based on scan statistics, is presented. DTSF scans an entire time series, identifies past patterns which are similar to the last available observations and forecasts based on the median of the subsequent observations of the most similar windows in past. In order to evaluate the performance of this method, a comparison with other statistical forecasting methods, applied in the M4 competition, is provided. In the hourly time domain, an average sMAPE of 12.9% was achieved using the method with the default parameters, while the baseline competition – the simple average of the forecasts of Holt, Damped, and Theta methods – was 22.1%. The method proved to be competitive in longer time series, with high repeatability.**

***Index Terms*—Univariate methods, M4 competition, benchmarking, dynamic time scan forecasting.**

I. Introduction

The development of predictive models is widely debated in the literature [1]–[4], since it assists the control of associated uncertainty intrinsic to random variables. Given the above, there are several categories of predictive models based on this physical knowledge (such as spectral analysis [5]) of intensive machine learning and statistical approaches [6]. Forecasting models associated with a single random variable as a function of time support univariate forecasting, which is a very important area given its application in various sectors such as [7]–[9], business [10]–[12], energy [13]–[15], among others. In this context, it is fundamentally valuable to develop meticulous criteria for selecting the models [16].

The M-competition [17]–[20] is the most important forecasting competition in academia, in which researchers from all around the world test their methods on real-life, anonymous time series from distinct areas of industry. The 4th edition took place in 2018 [17], and 17 methods based on combinations of statistical- and machine-learning or hybrids were tested on 100,000-time series. Outputs from these events are registered in review articles, pointing out the directions of development and refinement of the most promising forecasting techniques [21]. The 5th edition took place in 2020, and focused on a retail sales application with 42,850 unit sales hierarchical series, with the objective to produce the most accurate point forecast as well as the most accurate estimation of the uncertainty of these forecasts [22]. The 6th competition will take place this year and it will focus on predicting the overall market returns of individual stocks [23].

Whereas most well-known forecasting methods are based on identifying intrinsic components of the time series, such as level, trend, or seasonality, a particular group of methods based on similarity searches have been arousing interest in the areas of meteorology and renewable energy [24], [25]. These methods consist of identifying past weather patterns ("analogs") that closely resemble the current state. These methods are capable of handling lengthy historical time series in order to produce accurate and interpretive forecasts.

Among these methods, Dynamic Time Scan Forecasting (DTSF) consists of a new and simple analog-based forecasting technique [26]. It generates forecasts based on similar patterns, those with the highest R2 scores, calculated from the last available window.

The accuracy of analog-based methods is scarcely reported in areas other than energy prediction and is mostly limited to wind and solar energy forecasting applications [27], [28], which begs the question: "are analog-search-based models competitive compared to classical statistical prediction methods?". Additionally, no research was found that compared analog search methods and statistical methods.

To fill this gap, the current paper describes the DTSF forecasting method and discloses its performance on the M4 competition time series. We compare DTSF with eight classical statistical methods (Naive, Seasonal Naive, Simple Exponential Smoothing, Holt, Damped, Theta, AutoRegressive Integrated Moving Average (ARIMA), and ExponenTial Smoothing state space model (ETS)) and a combination of the outcomes of 3 individual methods (Holt, Damped, and Theta), which compose the baseline of the M4 competition. The M4 benchmark dataset was selected for this research because: (1) it consists of a reliable and curated benchmark base, adopted by other researchers and practitioners for developing and testing forecasting methods; (2) it has a significant number of series: 100,000 time series, with different frequencies (hourly, daily, monthly, weekly, quarterly, yearly); (3) it has been mostly predominated by statistical methods of forecasting; (4) and it is composed of univariate and independent series.

The major contributions of the present paper can be summarized as follows:

Rodrigo Barbosa de Santis is with LADEC Lab, Federal University of Minas Gerais (UFMG) e-mail:rsantis@ufmg.br

Tiago Silveira Gontijo is with LADEC Lab, Federal University of Minas Gerais (UFMG) e-mail:economista@ufmg.br

Marcelo Azevedo Costa is with LADEC Lab, Federal University of Minas Gerais (UFMG) e-mail:azevedo@est.ufmg.br

- the study applies a new method to M4 competition for benchmark purposes;
- the method is compared with nine classical statistical methods and a combination of the outcomes of three individual methods, which compose the baseline of the competition;
- in addition to applying the method, along with its default parameters, an exhaustive search with hold-out validation is adopted for model selection.

The major conclusions are:

- in the hourly time domain, an average error of 12.9% was obtained using the method with the default parameters, while the competition baseline was 22.1%;
- through the automatic selection of parameters, we boosted the accuracy of the method by 12.31% compared to the method application without parameters selection;
- the method proved to be competitive, both in terms of accuracy and computational cost, over long time series and with high repeatability.

The present paper is organized into 5 sections. Following this Introduction, Section 2 provides a review of the proposed forecasting method. Section 3 provides a background of the datasets and methods applied in this study. Section 4 presents the results and discussions obtained from the application of the methods. Finally, Section 5 concludes the present paper and includes some recommendations for future studies.

II. Materials and Methods

A. M4 Competition Dataset

The data used in the current study comes from the M4 competition dataset [17]. It is composed of 100,000 time series, taken from different domains such as Economics, Finance, Demographics, and Industry, among others. The time series show different periods: yearly, quarterly, monthly, weekly, daily, or hourly.

Table I summarizes the information about the competition's dataset. Domain refers to the time period from which the data have been extracted, ranging from hourly to yearly. The number of Series shows how many time series are available, in total. The dataset is mostly composed of a collection of time series from yearly, quarterly or monthly domains - 95,000 time series. The minimum length is the shorter time series in the given domain: the more aggregated the domain, like yearly, the more difficult it is to retrieve data. For example, hourly time series are longer, having at least 700 available observation points. Horizon refers to how many steps are being predicted in the future and are being used for metric computation. Seasonality represents the expected recurrence of an event in a given time domain.

The dataset provides a public and reliable source for comparing statistical, machine learning, or hybrid methods on univariate time series forecasting [29]. It is internationally recognized by researchers and data scientists as the most important competition in this area [30].

TABLE I
SUMMARY OF M4 COMPETITION DATASET, INCLUDING TIME-FREQUENCY, MINIMUM LENGTH OF TIME SERIES, AND FORECAST HORIZON OF EACH TIME SERIES.

Domain	Number of series	Min. length	Horizon	Seasonality
Yearly	23,000	13	6	1
Quarterly	24,000	16	8	4
Monthly	48,000	42	18	12
Weekly	359	80	13	52
Daily	4,227	93	14	7
Hourly	414	700	48	24

B. Dynamic Time Scan Forecasting

DTSF is a forecasting method based on scan statistics [31] and was originally developed to address the problem of wind forecasting for Brazilian power generation plants. It consists of scanning a time series and identifying past patterns (called "analogs") similar to the last observations available of the time series (called "query") [26].

Let y_t be a time series of length N, $t = 1, ..., N$. Firstly, let vector $\mathbf{y}^{[w]}$ be defined as the last w observations of the series:

$$\mathbf{y}^{[w]} = [y_{N-w+1}, ..., y_N]. \tag{1}$$

The goal of DTSF is to identify analogs in the time series which are greatly correlated with vector $\mathbf{y}^{[w]}$. Hence, the set of candidate vectors can be defined by:

$$\mathbf{x}_t^{[w]} = [y_{t-w+1}, ..., y_{t-w}] \tag{2}$$

where $t = 1, ..., N-2\cdot w$. The upper limit of the time sequence $(N - 2 \cdot w)$ guarantees that vector $\mathbf{x}_t^{[w]}$ does not overlap with vector $\mathbf{y}^{[w]}$. Fig. 1 presents the DTSF procedure. Given the last w observed values, which comprises vector $\mathbf{y}^{[w]}$, a rolling window with the same size $(\mathbf{x}_t^{w})$ is used for scanning previous values of the series.

Lastly, DTSF provides a k steps ahead forecast of the time series, $y_{N+1}, ..., y_{N+k}$. To produce this outcome, the DTSF scans the series to find the closest analogs $\mathbf{x}_t^{[w]}$. The subsequent values of the time series are used as the forecast values:

$$y_{N+i} = f_{\mathbf{x}_t^{[w]}}(y_{t-w+i}) \tag{3}$$

where $f_{\mathbf{x}_t^{[w]}}$ is a function which correlates the elements of vector $\mathbf{x}_t^{[w]}$ and the elements of vector $\mathbf{y}^{[w]}$.

According to that, a first constraint can be set on $k : 1 \leq k \leq w$. This constraint guarantees that if the most correlated time series window comprises the most recent values, prior to vector $\mathbf{y}^{[w]}$, then the forecast values are a function of vector $\mathbf{y}^{[w]}$,

$$y_{N+i} = f_{\mathbf{x}_{N-2w}^{[w]}}(y_{N-w+i}). \tag{4}$$

As stated in Equations (3) and (4), forecast values depend on the window length w and the function $f_{\mathbf{x}_t^{[w]}}(.)$. A intuitive proposal for function $f_{\mathbf{x}_t^{[w]}}(.)$ is a linear scaling of the elements of vector $\mathbf{x}_t^{[w]}$, i.e., a linear model. This occurs due to the fact that previous values are likely similar to the last observations, except for a scale and/or offset shift. So, the method searches for values that may be similar to the last values, after applying a similarity function [26].

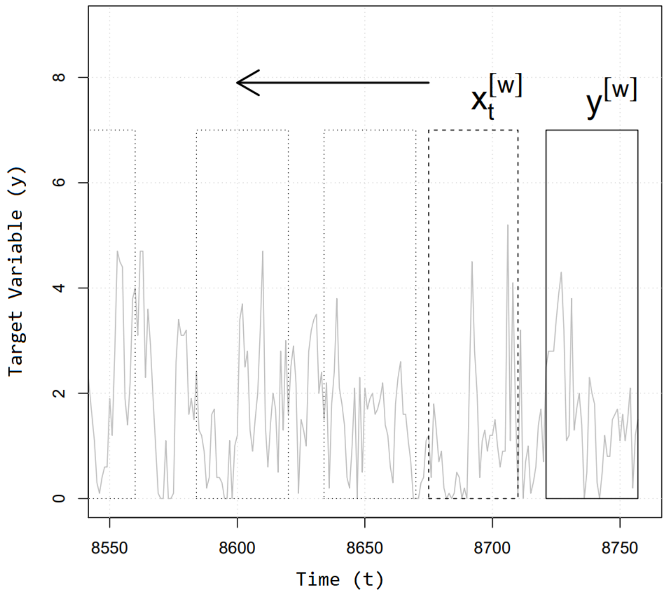

Fig. 1. Illustration of the DTSF time series scan procedure.

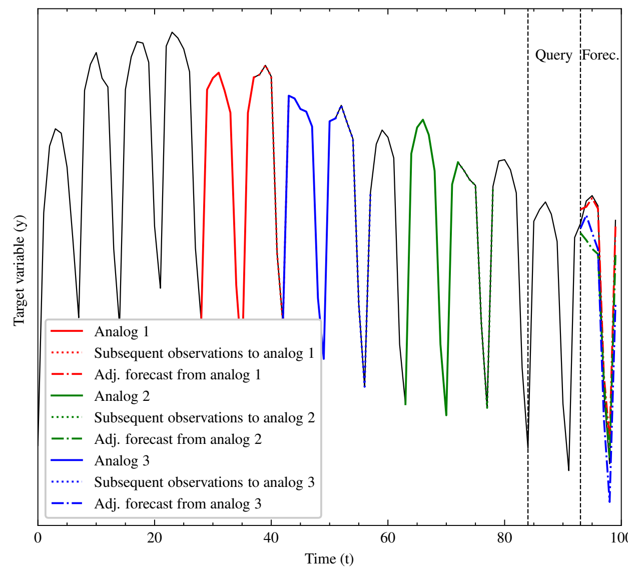

Fig. 2. Example of DTSF application to forecasting a time series. The three colored lines represent the top three analogs correlated to the queried period. The dashed lines are the subsequent observations of the analogs. The forecast is given by the median of the adjusted forecast from the subsequent observations of the top analogs.

By taking a linear function as the similarity function, the parameters of the model can be estimated to minimize the sum of squares between the elements of vector $\mathbf{y}^{[w]}$ and the linear equation: $\beta_0^{[t]} + \beta_1^{[t]} \times \mathbf{x}_t^{[w]}$. Moreover, the similarity statistic can be assumed as the linear regression coefficient of determination R^2 [26], [32]:

$$R^2 = 1 - \frac{\sum_j \left(\mathbf{y}_j^{[w]} - \hat{\mathbf{y}}_j^{[w]}\right)^2}{\sum_j \left(\mathbf{y}_j^{[w]} - \bar{y}_j^{[w]}\right)^2} \tag{5}$$

where $\mathbf{y}_j^{[w]}$ is the j-th value of vector $\mathbf{y}^{[w]}$ and $\hat{\mathbf{y}}_j^{[w]}$ is the j-th predicted value using the estimated linear function. Finally, the method calculates a similarity profile based on the R^2 score resulting from the comparison of the query with previous windows. The analogs with higher R^2 scores are considered closer analogs. Predictions of future steps are calculated from a predefined number of analogs using aggregation functions, such as median [26].

The DTSF model requires three parameters to be selected by the user: the length of the query window, the similarity function specification, and the number of analogs to be considered. The original implementation of DTSF is available on the R package, DTScanF. In the present study, the original implementation is the extent to which the aggregation function applied to analogs can be either the median or the mean, according to the user or the model selection procedure.

Fig. 2 illustrates the forecasting procedure, using time scanning in a given hourly time series, adopting a window with a length equal to 48 hours, a linear similarity function (degree equal to 1), and the three analogs. Windows 1, 2, and 3 are the ones most similar to the last window of available data. The forecast is given by the median (but other statistics can be used such as the mean) of the subsequential observations of the analogs.

As a data-driven method, DTSF usually performs better on time series with large numbers of observations and it can also be extended to search the patterns of secondary series related to the prediction. The main disadvantage of the method is the computational cost of scanning the entire time series and calculating the similarity profile. However, more efficient methods, such as the Maureen's Algorithm of Similarity Search (MASS) which applies convolution, have been applied for speeding up this task [27]. To keep it feasible, the linear similarity functions commonly adopted are from the first to the third-degree polynomials.

C. Statistical Forecasting Methods

A univariate forecasting method is a procedure for estimating a point. The forecast is based on past and present values of a given time series [33]. This method is generally applied when there is a large number of series to forecast, or when multivariate methods require forecasts for each explanatory variable. Given the advantage of simplicity and high usage, univariate forecasting methods are employed in most of the forecast applications in areas such as business, energy, and finance. The following methods are selected from the latest M4 competition benchmark [17], and a simple explanation is given for each one, as follows:

1) *Naive*: the simplest, yet still powerful forecasting method; assumes that the next steps to be predicted are equal to the last available observation [20].
2) *Seasonal Naive (sNaive)*: the same concept as Naive, with the adaptation that the time series is deseasonalized; method adjusted and forecast later, re-adjusted with the seasonal component [20].
3) *Naive2*: each time series uses the forecast of either Naive or sNaive, based on their score on the validation set.

TABLE II
PARAMETERS RANGE ADOPTED FOR DTSF.

Parameters	**Range**
Polynomial degree	1
Analogs	10
Window size	48
Aggregation function	Median

4) *Simple Exponential Smoothing (SES)*: classic statistical method which applies an exponentially weighted average [34].
5) *Holt*: exponential smoothing with level and linear trend components [34].
6) *Damped*: exponential smoothing with dampened parameters for flattening trends, after a given period [35].
7) *Theta*: method based on a coefficient of curvature of the time-series, applied to the second difference of the data [36].
8) *Combined (Comb)*: the simple average of the forecasts of the previous three models: Holt, Damped and Theta.
9) *ARIMA*: general forecast method estimated from the autoregressive, moving average and integration components from the time series analysis [37].
10) *ETS*: automatic forecasting based on an extended range of exponential smoothing methods [38].
11) *DTSF*: the proposed method, adopting the defined default parameters, which are: (i) polynomial function degree equal to 1, (ii) analogs equal to 10, (iii) window size equal to length of forecast horizon, and (iv) median as aggregation function [26].

Table II presents the range adopted for the parameters of the proposed method. The polynomial degree is the degree of the function used for approximation, analogs are the number of analogs to be used to estimate the forecast, window size defines the length of the scan window, and aggregation function is the one that transforms the projection of the analogs into the final forecast.

D. Model Selection Procedure

The split of the data into training sets and test sets split is predefined and given by the competition organizers. The data come from different files for each of the time series domains. The test set has a fixed horizon for all the time series, and it is used only for computing the final scores. The evaluation metrics adopted are the same ones that are applied in the M4 Competition, and are those most used in literature [39], [40]: the Symmetric Mean Absolute Percentage Error (sMAPE), Mean Absolute Scaled Error (MASE) and Overall Weighted Average (OWA). The formula for calculating the metrics is given:

$$sMAPE = \frac{1}{h}\sum_{t=1}^{h}\frac{2|Y_t - \hat{Y}_t|}{|Y_t| + |\hat{Y}_t|} \tag{6}$$

$$MASE = \frac{1}{h}\frac{(n-m)\sum_{t=1}^{h}|Y_t - \hat{Y}_t|}{\sum_{t=m+1}^{n}|Y_t - Y_{t-m}|} \tag{7}$$

$$OWA = \frac{sMAPE_k/sMAPE_{base} + MASE_k/MASE_{base}}{2} \tag{8}$$

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

A hold-out cross-validation scheme is adopted to evaluate and select the best parameters for the methods, in which the last k observations are kept as the validation set, k being equal to the forecast horizon. All possible parameter combinations are enumerated within the defined ranges, and the methods are tuned using an exhaustive grid search procedure with sMAPE as the scorer.

E. Software and Hardware

Routines were implemented using the R 3.6.0 programming language with the official benchmarks and evaluation script of M4 Competition, available at the GitHub repository (https://github.com/M4Competition/M4-methods). The *Forecast 8.7* package is used for the SES, Holt, Damped, ARIMA, and ETS methods. DTSF comes from the official implementation of the method in R and C++, available from the public repository (https://rdrr.io/github/leandromineti/DTScanF/). All data and scripts are available from the authors upon request.

Computer specifications used to execute the algorithms and calculate the forecasts are as follows: CPU 8-core Intel Core i9 2.3 GHz, 16 GB of RAM, and macOS 12.5 operating system. Once the predictions are calculated, the error arrays are next calculated and saved as RDS files, allowing analysis of the results. Fitting time is computed from the time delta of the system, before and after each execution of the methods.

III. RESULTS AND DISCUSSION

Table III presents the average sMAPE achieved by each of the statistical methods and by the proposed method, computed for each of the time domains. The Theta method achieved the best scores for the yearly and monthly frequencies (14.603 and 13.003), which composed more than 70% of the total of the series, thus contributing to this particular method outperforming the other methods in the overall average (12.312). In the individual domains, Comb achieved the lowest error for both the daily (10.197) and the quarterly (10.197) domains, while the ARIMA method scored the lowest error on the weekly frequency (8.593).

The average error of all methods is the lowest for daily frequency (close to 3.00), and there seems to exist a trend toward increasing as the time domain becomes broader: the weekly average error is around 9, the monthly is around 13, and so on. The exception is for the hourly frequency, in which most of the statistical methods scored errors from 13.912 to 43.003.

DTSF exhibited errors considerably fewer errors methods considered for benchmarking in this particular kind of time series (12.927). This makes the DTSF method interesting for studying applications in which competitive estimators are sought.

Table IV presents the evaluation of the methods using OWA. This metric is understood as showing how one method is more accurate when compared to Naive2. If OWA is lower than 1 the method is more adequate than Naive2. Otherwise, Naive2 provides better forecasting performance. The DTSF scores for the hourly series imply a meaningful increase in accuracy over the Naive method (0.552). Moreover, when applying fine-tuning, the gain increases to nearly 50%. For all other domains, the only ones in which the method performed worse than Naive2 were the yearly and the daily, both of which have in common the longer term forecast period and the lowest seasonality traits in common.

The outcome of the experiment can be explained by the intrinsic design of the DTSF method, which was originally conceived to deal with very long time series with recurrent patterns, such as its original application to 30-min frequency wind speed forecasting. Comparing results to Table I, which presents the seasonality, length, and forecast horizon of each time domain, it is shown that the DTSF accuracy is greater when the number of available data points is also greater.

Fig. 3 displays the average sMAPE for each one of the 414 hourly time series available in the competition database, listed in ascending order according to the calculated error of the DTSF method. The methods Naive, sNaive and SES methods were holdouts of the graphical representation. The y-axis is presented using the base-10 logarithmic scale in order to facilitate visual analysis.

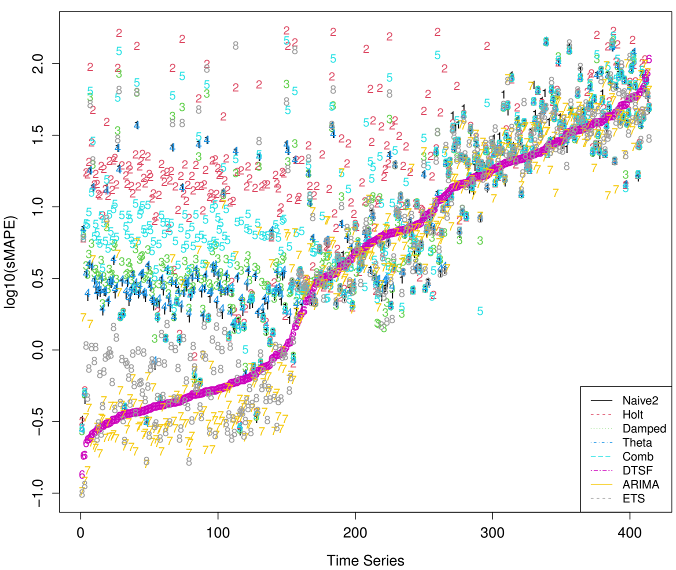

Fig. 3. Forecasting methods average sMAPE for each of the 414 hourly time series, ordered by the accuracy of the DTSF method. The proposed method obtained fewer errors for most of the time series in this particular domain of application.

In the first 170 time series with the lowest sMAPE – one-third of the total available – the method proposed in the present article achieved errors close to 10^{-2}, while most of the others obtained errors between $10^{0.5}$ and 10^{2}. This shows the enormous predictive power in this specific type of series, and the great gain in accuracy that explains the best performance of this method, on average. Analyzing the sets between the 170th and 300th time series with the smallest error, there is less distinction between all the methods which, in general, presented errors very close to each other. Other methods have shown a lower errors than DTSF along all time series, specially the methods ARIMA and ETS. In the set between 300th and 414th, DTSF again marginally outperformed the other benchmark methods in most of the series.

Table V presents the average sMAPE detailed by the forecast horizon, grouped by 6-hour periods. DTSF obtained lower errors, for all horizons than the other compared methods. Furthermore, the average error is 12.9%, and the highest errors were obtained during the periods between the hours from 19 to 30 and the hours from 43 to 48.

To provide better visualization of error evolution over time, Fig. 4 presents the mean errors per step of each method (excluding the three from the previous figure), for all hourly time series. An increase in error over time, according to the phenomenon of error propagation, is expected. This is better observed in the Holt method, in which error varied from 10% at the first step to 40% at the last step. Moreover, in such a visual representation, the Theta model is perceived to have been more accurate, on average, than the DTSF model for the 1st and 24th hours.

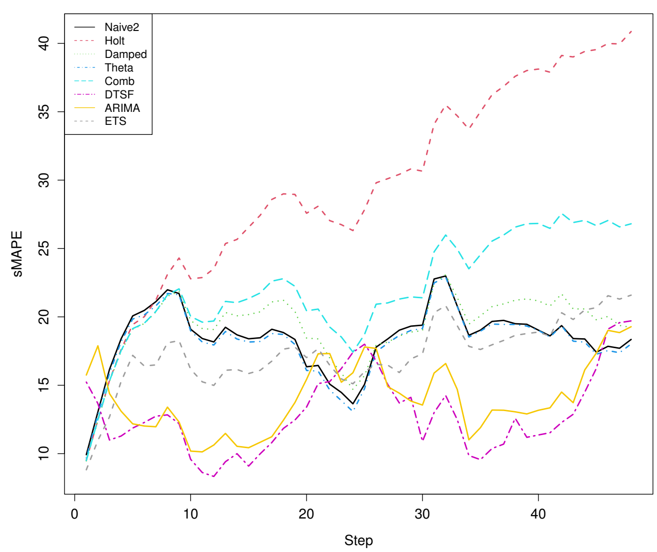

Fig. 4. Average sMAPE (obtained in the 414 hourly time series by all the methods for each step of the prediction, up to 48 hours – forecast horizon).

Most statistical methods presented a pattern of very similar curves, with the exception of the DTSF method. In DTSF, the errors presented a different pattern, alternating peaks, and valleys with the patterns of the other statistical methods. In general, DTSF appeared to remain more stable throughout the period, experiencing less of the error propagation effect and not exceeding the limit of 20%. These are more examples that explain the better performance of the DTSF method, compared to the benchmark, in the hourly domain.

TABLE III
THE PERFORMANCE OF DTSF COMPARED TO M4 BENCHMARK STATISTICAL METHODS – SMAPE METRIC.

Method	**sMAPE** Yearly (23k)	Quarterly (24k)	Monthly (48k)	Weekly (359)	Daily (4,227)	Hourly (414)	Average (100k)
Naive	16.342	11.610	15.255	9.161	3.405	43.003	14.207
sNaive	16.342	12.521	15.994	9.161	3.405	13.912	14.660
Naive2	16.342	11.012	14.429	9.161	3.405	18.383	13.565
SES	16.398	10.600	13.620	9.012	3.405	18.094	13.089
Holt	16.535	10.955	14.833	9.706	3.070	29.474	13.839
Damped	15.162	10.243	13.475	8.867	3.063	19.277	12.655
Theta	**14.603**	10.312	**13.003**	9.094	3.053	18.138	**12.312**
Comb	14.874	**10.197**	13.436	8.947	**2.985**	22.114	12.567
ARIMA	15.150	10.408	13.486	**8.593**	3.185	14.081	12.679
ETS	15.356	10.291	13.525	8.727	3.046	17.307	12.725
DTSF	16.816	11.006	13.823	8.983	3.313	**12.927**	13.370

TABLE IV
THE PERFORMANCE OF DTSF COMPARED TO M4 BENCHMARK STATISTICAL METHODS – OWA METRIC.

Method	**OWA** Yearly (23k)	Quarterly (24k)	Monthly (48k)	Weekly (359)	Daily (4,227)	Hourly (414)	Average (100k)
Naive	1.000	1.066	1.095	1.000	1.000	3.593	1.072
sNaive	1.000	1.153	1.147	1.000	1.000	0.628	1.106
Naive2	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SES	1.003	0.970	0.951	0.975	1.000	0.990	0.970
Holt	0.956	0.935	0.989	0.964	0.997	2.760	0.976
Damped	0.888	0.893	0.924	0.916	0.996	1.140	0.912
Theta	0.872	0.917	0.907	0.971	0.999	1.006	**0.906**
Comb	**0.868**	0.891	0.920	0.926	**0.979**	1.559	**0.906**
ARIMA	0.891	0.898	**0.904**	0.927	1.041	0.950	**0.906**
ETS	0.903	**0.890**	0.914	0.931	0.996	1.824	0.913
DTSF	1.002	0.961	0.950	**0.914**	1.092	**0.552**	0.969

TABLE V
AVERAGE SMAPE OBTAINED IN THE 414 HOURLY TIME SERIES BY THE PREDICTED STEPS, GROUPED IN 6-HOUR PERIODS.

Methods	**Steps** 1-6	7-12	13-18	19-24	25-30	33-36	37-42	43-48	1-48
Naive2	16.3	20.1	18.8	15.7	18.2	20.7	19.3	18.0	18.4
Naive2	16.3	20.1	18.8	15.7	18.2	20.7	19.3	18.0	18.1
Holt	15.7	23.0	27.1	27.5	29.9	34.9	37.9	39.8	29.5
Damped	15.5	20.3	20.5	17.5	18.1	21.2	21.2	19.9	19.3
Theta	16.1	19.9	18.5	15.3	17.8	20.5	19.2	17.8	18.1
Comb	15.6	20.6	21.8	19.7	20.8	24.9	26.7	26.8	22.1
ARIMA	14.2	11.4	11.2	15.8	15.4	13.9	13.4	17.0	14.1
ETS	13.6	16.5	16.4	16.6	16.5	19.0	18.9	17.4	17.3
DTSF	**12.6**	**10.7**	**10.2**	**15.0**	**14.8**	**11.6**	**11.6**	**11.6**	**12.9**

Table VI shows the time necessary to fit the methods for all of the 100,000 time series. The methods Naive2 and Comb have been omitted as these two are a combination/selection of individual methods. Total fitting time is given in seconds, while the average time per series is given in microseconds. The Ratio Naive column compares the average time of a particular method compared to the execution time of the Naive method.

DTSF was the method that consumed the most computational time, almost 9 times more than Naive. It is worth mentioning that the default parameters for DTSF adopt 10 analogs to estimate the forecast. Also, part of the method is executed in the C compiled language, and part of it is executed in R.

TABLE VI
TOTAL AND AVERAGE TIMES NECESSARY FOR FITTING THE METHODS.

Methods	**Total fitting time (s)**	**Average time per series (ms)**	**Ratio to naive**
Naive	0.458	1.106	1.00
sNaive	0.656	1.584	1.43
SES	2.219	5.360	4.85
Holt	5.947	14.365	12.99
Damped	12.789	30.892	27.94
Theta	2.964	7.159	6.47
ARIMA	18437.598	44535.261	40278.22
ETS	1838.638	4441.155	4016.63
DTSF	6.241	15.074	13.63

IV. Conclusions

The current paper presents the results of applying the dynamic time scan forecasting method with the M4 competition data and compares it with statistical methods used as baselines in the same competition. The results point to a significant gain in accuracy in hourly time domain problems, compared to the reference, which justifies adopting this method for problems of this particular nature.

Since the method was developed for problems with long time series and high repeatability, DTSF has been proved competitive. In the present experiment, the DTSF method reduced the sMAPE by 12.13%.

Furthermore, the dissemination of this method may be interesting for other researchers who wish to extend it to existing methods, either by combining it with other techniques or by adapting its operation to other applications.

Future research should extend the method to multivariate forecasting problems and hierarchical time series and should assess its performance in other applications with this characteristic (the M5 competition, for instance). Also, some extensions of the method itself are foreseen, in order to improve its accuracy on time series for which its performance was less satisfactory than the performance of other statistical methods, for example, adopting k-fold instead of hold-out cross-validation for model selection [41].

Acknowledgments

The authors would like to thank the National Council for Scientific and Technological Development (CNPq), Companhia Energética Integrada (CEI), and Brasil Energia Inteligente (BEI) for its financial support of the project (Grant No. 141740/2019-1 and Grant No. 141177/2019-2, respectively).

References

[1] T. Hill, L. Marquez, M. O'Connor, and W. Remus, "Artificial neural network models for forecasting and decision making," *International journal of forecasting*, vol. 10, no. 1, pp. 5–15, 1994.

[2] P.-F. Pai and C.-S. Lin, "A hybrid arima and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, 2005.

[3] G. Dudek, "Pattern-based local linear regression models for short-term load forecasting," *Electric Power Systems Research*, vol. 130, pp. 139–147, 2016.

[4] R. Shanmugam, "Book review," *Journal of Statistical Computation and Simulation*, vol. 76, no. 10, pp. 935–940, 2006.

[5] T. T. Tchrakian, B. Basu, and M. O'Mahony, "Real-time traffic flow forecasting using spectral analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 519–526, 2011.

[6] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[7] H. Hassani and E. S. Silva, "Forecasting uk consumer price inflation using inflation forecasts," *Research in Economics*, vol. 72, no. 3, pp. 367–378, 2018.

[8] W. Cai, J. Chen, J. Hong, and F. Jiang, "Forecasting chinese stock market volatility with economic variables," *Emerging Markets Finance and Trade*, vol. 53, no. 3, pp. 521–533, 2017.

[9] E. Bernardini and G. Cubadda, "Macroeconomic forecasting and structural analysis through regularized reduced-rank regression," *International Journal of Forecasting*, vol. 31, no. 3, pp. 682–691, 2015.

[10] Z. R. Khan Jaffur, N.-U.-H. Sookia, P. Nunkoo Gonpot, and B. Seetanah, "Out-of-sample forecasting of the canadian unemployment rates using univariate models," *Applied Economics Letters*, vol. 24, no. 15, pp. 1097–1101, 2017.

[11] Y. Zhang, M. Zhong, N. Geng, and Y. Jiang, "Forecasting electric vehicles sales with univariate and multivariate time series models: The case of china," *PloS one*, vol. 12, no. 5, p. e0176729, 2017.

[12] G. A. Tularam and T. Saeed, "Oil-price forecasting based on various univariate time-series models," *American Journal of Operations Research*, vol. 6, no. 03, p. 226, 2016.

[13] G. Girish, A. K. Tiwari, *et al.*, "A comparison of different univariate forecasting models forspot electricity price in india," *Economics Bulletin*, vol. 36, no. 2, pp. 1039–1057, 2016.

[14] M. Rana, I. Koprinska, and V. G. Agelidis, "Univariate and multivariate methods for very short-term solar photovoltaic power forecasting," *Energy Conversion and Management*, vol. 121, pp. 380–390, 2016.

[15] E. Raviv, K. E. Bouwman, and D. Van Dijk, "Forecasting day-ahead electricity prices: Utilizing hourly prices," *Energy Economics*, vol. 50, pp. 227–239, 2015.

[16] B. Billah, R. J. Hyndman, and A. B. Koehler, "Empirical information criteria for time series forecasting model selection," *Journal of Statistical Computation and Simulation*, vol. 75, no. 10, pp. 831–840, 2005.

[17] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The m4 competition: Results, findings, conclusion and way forward," *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, 2018.

[18] S. Makridakis and M. Hibon, "The m3-competition: results, conclusions and implications," *International journal of forecasting*, vol. 16, no. 4, pp. 451–476, 2000.

[19] S. Makridakis, C. Chatfield, M. Hibon, M. Lawrence, T. Mills, K. Ord, and L. F. Simmons, "The m2-competition: A real-time judgmentally based forecasting study," *International Journal of Forecasting*, vol. 9, no. 1, pp. 5–22, 1993.

[20] S. Makridakis and M. Hibon, "Accuracy of forecasting: An empirical investigation," *Journal of the Royal Statistical Society: Series A (General)*, vol. 142, no. 2, pp. 97–125, 1979.

[21] M. P. P. Flores, J. F. Solís, G. C. Valdez, J. J. G. Barbosa, J. P. Ortega, and J. D. T. Villanueva, "Hurst exponent with arima and simple exponential smoothing for measuring persistency of m3-competition series," *IEEE Latin America Transactions*, vol. 17, no. 05, pp. 815–822, 2019.

[22] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The m5 competition: Background, organization, and implementation," *International Journal of Forecasting*, 2021.

[23] "The m6 financial forecasting competition."

[24] D. Yang and S. Alessandrini, "An ultra-fast way of searching weather analogs for renewable energy forecasting," *Solar Energy*, vol. 185, pp. 255–261, 2019.

[25] L. E. B. Hoeltgebaum, N. L. Dias, and M. A. Costa, "An analog period method for gap-filling of latent heat flux measurements," *Hydrological Processes*, vol. 35, no. 4, p. e14105, 2021.

[26] M. A. Costa, R. Ruiz-Cárdenas, L. B. Mineti, and M. O. Prates, "Dynamic time scan forecasting for multi-step wind speed prediction," *Renewable Energy*, vol. 177, pp. 584–595, 2021.

[27] T. S. Gontijo, M. A. Costa, and R. B. de Santis, "Similarity search in electricity prices: An ultra-fast method for finding analogs," *Journal of Renewable and Sustainable Energy*, vol. 12, no. 5, p. 056103, 2020.

[28] T. S. Gontijo, M. A. Costa, and R. B. de Santis, "Electricity price forecasting on electricity spot market: a case study based on the brazilian difference settlement price," in *E3S Web of Conferences*, vol. 239, EDP Sciences, 2021.

[29] G. Bontempi, "Comments on m4 competition," *International Journal of Forecasting*, vol. 36, no. 1, pp. 201–202, 2020.

[30] R. Fildes and S. Makridakis, "The impact of empirical accuracy studies on time series analysis and forecasting," *International Statistical Review/Revue Internationale de Statistique*, pp. 289–308, 1995.

[31] J. Glaz and N. Balakrishnan, *Scan statistics and applications*. Springer Science & Business Media, 2012.

[32] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[33] C. Chatfield, *Time-series forecasting*. Chapman and Hall/CRC, 2000.

[34] R. Hyndman, A. B. Koehler, J. K. Ord, and R. D. Snyder, *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.

[35] E. S. Gardner and E. McKenzie, "Why the damped trend works," *Journal of the Operational Research Society*, vol. 62, no. 6, pp. 1177–1180, 2011.

[36] V. Assimakopoulos and K. Nikolopoulos, "The theta model: a decomposition approach to forecasting," *International journal of forecasting*, vol. 16, no. 4, pp. 521–530, 2000.

[37] G. E. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *Journal*

of the American statistical Association, vol. 65, no. 332, pp. 1509–1526, 1970.

[38] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," *International Journal of forecasting*, vol. 18, no. 3, pp. 439–454, 2002.

[39] S. M. Al-Alawi and S. M. Islam, "Principles of electricity demand forecasting. i. methodologies," *Power Engineering Journal*, vol. 10, no. 3, pp. 139–143, 1996.

[40] A. Azadeh, S. Ghaderi, and S. Sohrabkhani, "Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors," *Energy Conversion and management*, vol. 49, no. 8, pp. 2272–2278, 2008.

[41] C. Bergmeir, R. J. Hyndman, and B. Koo, "A note on the validity of cross-validation for evaluating autoregressive time series prediction," *Computational Statistics & Data Analysis*, vol. 120, pp. 70–83, 2018.

Rodrigo Barbosa de Santis graduated in Production Engineering from the Federal University of Juiz de Fora (2015) and Master in Computer Modeling from the Federal University of Juiz de Fora (2018). He has experience in operations management, having worked as a consultant at Natura, BASF, Johnson & Johnson, and B2W. His skills include developing and implementing predictive and decision support systems. His main interests are Machine Learning, Artificial Intelligence, Operations Management, and Supply Chain.

Tiago Silveira Gontijo is PhD student and Master in Production Engineering. Bachelor of Business Administration and Economics. He has published books and papers and works as a reviewer for international journals. He has experience in the administration of production systems and in the sanitation production chain. Worked at the Regulatory Agency for Water Supply and Sanitary Sewage Services of the State of MG. Has experience in applied statistics and quantitative methods, with emphasis on renewable energies and time series. He works as an independent consultant in economic regulation.

Marcelo Azevedo Costa is a full professor of Applied Statistical Methods in the Department of Production Engineering at the Federal University of Minas Gerais (UFMG) and faculty member of the Graduate Program in Production Engineering (Stochastic Modeling and Simulation) at the same institution. He holds a degree in Electrical Engineering from the Federal University of Minas Gerais (1999), a Ph.D. in Electrical Engineering from the Federal University of Minas Gerais (2002) in Computational Intelligence. He has done post-doctorate from Harvard Medical School & Harvard Pilgrim Health Care (2007) in Spatial Statistics and Epidemiological Surveillance, and a post-doctorate from Linköping University (2018/Sweden) in Statistical Analysis, Diagnosis and Fault Detection in Industrial Environments. He is currently a collaborating professor in the Laboratory of Intelligence and Computational Technology (LITC/UFMG), developing research projects in artificial neural networks. He is also a collaborating professor at the Center of Research in Efficiency, Sustainability and Productivity (NESP/UFMG), developing research projects in economic regulation in the electricity sector. He has published several papers in international journals such as Socio-Economic Planning Sciences, IEEE Transactions on Power Delivery, Statistical Methods in Medical Research, PLOS One, Measurement, among others. He is a reviewer of international and national journals, as well as the author of book chapters published in English. He is a CEMIG/FAPEMIG R&D project coordinator and researcher in R&D projects. He supervises undergraduate, specialization, master and doctoral students in the following areas: statistical models applied to the electrical sector, applied statistics, network analysis, spatial statistics, time series analysis, artificial neural network theory, and applications. He is enthusiastic about using R language programming, which he has applied to solve many practical statistical problems including Big Data analysis.