



Tiago Mendes Dantas

**Combining to succeed: A novel strategy to
improve forecasts from Exponential Smoothing
models**

Tese de Doutorado

Thesis presented to the Programa de Pós-graduação em Engenharia de Produção of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia de Produção.

Advisor: Prof. Fernando Luiz Cyrino Oliveira

Rio de Janeiro
September 2018



Tiago Mendes Dantas

**Combining to succeed: A novel strategy to
improve forecasts from Exponential Smoothing
models**

Thesis presented to the Programa de Pós-graduação em Engenharia de Produção of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Engenharia de Produção. Approved by the undersigned Examination Committee.

Prof. Fernando Luiz Cyrino Oliveira

Advisor

Departamento de Engenharia Industrial – PUC-Rio

Prof. Reinaldo Castro Souza

Departamento de Engenharia Industrial – PUC-Rio

Prof. José Francisco Moreira Pessanha

Centro de Pesquisas de Energia Elétrica – CEPEL

Prof. Eduardo Pestana de Aguiar

Universidade Federal de Juiz de Fora – UFJF

Prof. Daniel Takata Gomes

Fundação Instituto Brasileiro de Geografia e Estatística –
IBGE/ENCE

Prof. Márcio da Silveira Carvalho

Vice Dean of Graduate Studies
Centro Técnico Científico – PUC-Rio

Rio de Janeiro, September 28th, 2018

All rights reserved.

Tiago Mendes Dantas

Obtained his Bachelor's degree in Statistics from ENCE/IBGE in 2009 and his Master's degree in Electrical Engineering from PUC-Rio in 2011. Since 2010, he has been working at the Brazilian Institute of Geography and Statistics (IBGE) as a Statistician and currently holds the position of Research and Development Manager, working mostly with Statistical Modelling, Time Series and Big Data projects.

Bibliographic data

Dantas, Tiago Mendes

Combining to succeed: A novel strategy to improve forecasts from Exponential Smoothing models / Tiago Mendes Dantas; advisor: Fernando Luiz Cyrino Oliveira. – Rio de Janeiro: PUC-Rio, Departamento de Engenharia Industrial, 2018.

v., 80 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial.

Inclui bibliografia

1. Engenharia de Produção – Teses. 2. Bagging;. 3. Agrupamento de séries temporais;. 4. Amortecimento exponencial;. 5. Partitioning Around Medoids;. 6. Redução de variância.. I. Oliveira, Fernando Luiz Cyrino. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. III. Título.

CDD: 658.5

To Célia, Sidnei, Daiane and Das Dores for their love, support
and encouragement.

Acknowledgments

Esse é um momento de extrema alegria e que não seria possível sem fé em Deus e apoio incondicional dos que de alguma forma estiveram presentes. Portanto, gostaria de agradecer a todos que fizeram parte dessa incrível jornada, de perto ou de longe, nesse plano ou em qualquer outro. Em especial:

Célia e Sidnei, meus pais que além de me amar e apoiar incondicionalmente, me mostraram o valor da educação, fazendo o possível e o impossível para que eu sempre tivesse as melhores oportunidades.

Daiane, amor da minha vida e maior incentivadora da minha entrada no doutorado . Obrigado não só pelo alegria e amor diário, mas também pela paciência, compreensão e carinho nos momentos onde tudo que eu podia fazer era estar sentado escrevendo ou programando.

Das Dores e Teresa, minhas mães postizas que desde que me entendo por gente são só carinho e amor.

Nenzinha, minha avó mas, sobretudo, a definição de mulher guerreira, com um coração enorme e a fé capaz de mudar o mundo. Você será sempre o meu maior exemplo.

Berenice, Manoel e Tersilha, por me amarem sempre e serem os avós que todo mundo gostaria de ter.

Fernando Cyrino, meu orientador e amigo, pela confiança, por me ensinar o valor da pesquisa, pela amizade e boas risadas. Ter escolhido você foi essencial!

Joana, Renata e Henrique, por permitirem que eu fosse um filho único com 3 irmãos.

Alex e Lincoln por estarem sempre por perto (mesmo que nem sempre fisicamente) e sempre torcerem por mim.

Bruno, Maurício e Paulo por incentivarem minha entrada no doutorado.

Professores da PUC-Rio por todas as lições aprendidas, em especial ao professor Reinaldo Castro Souza por sempre abrir portas e ser a razão da minha entrada na PUC-Rio em 2009.

Aos funcionários do Departamento de Engenharia Industrial da PUC-Rio, em especial a Claudia Teti, sempre solícita e disposta a ajudar.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

Obrigado!

Abstract

Dantas, Tiago Mendes; Oliveira, Fernando Luiz Cyrino (Advisor). **Combining to succeed: A novel strategy to improve forecasts from Exponential Smoothing models**. Rio de Janeiro, 2018. 80p. Tese de doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

This thesis is inserted in the context of time series forecasting. In this sense, although many approaches have been developed, simple methods such as exponential smoothing usually produce extremely competitive results, often surpassing approaches with a higher level of complexity. Seminal papers in time series forecasting showed that the combination of forecasts has the potential to dramatically reduce the forecast error. Specifically, the combination of forecasts generated by Exponential Smoothing has been explored in recent papers. Although this can be done in many ways, a specific method called Bagged.BLD.MBB.ETS uses a technique called Bootstrap Aggregating (Bagging) in combination with Exponential Smoothing methods to generate forecasts, showing that the approach can generate more accurate monthly forecasts than all the analyzed benchmarks. The approach was considered the state of the art in the use of Bagging and Exponential Smoothing until the development of the results obtained in this thesis. This thesis initially deals with validating Bagged.BLD.MBB.ETS in a data set relevant from the point of view of a real application, thus expanding the fields of application of the methodology. Subsequently, relevant motifs for error reduction are identified and a new methodology using Bagging, Exponential Smoothing and Clusters is proposed to treat the covariance effect, not previously identified in the method's literature. The proposed approach was tested using data from three time series competitions (M3, CIF 2016 and M4), as well as using simulated data. The empirical results point to a substantial reduction in variance and forecast error.

Keywords

Bagging; Clustering Time Series; Exponential Smoothing; Partitioning Around Medoids; Variance Reduction.

Resumo

Dantas, Tiago Mendes; Oliveira, Fernando Luiz Cyrino. **Combinando para ter sucesso: uma nova estratégia para melhorar a previsões de modelos de amortecimento exponencial.** Rio de Janeiro, 2018. 80p. Tese de Doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

A presente tese se insere no contexto de previsão de séries temporais. Nesse sentido, embora muitas abordagens tenham sido desenvolvidas, métodos simples como o de amortecimento exponencial costumam gerar resultados extremamente competitivos muitas vezes superando abordagens com maior nível de complexidade. No contexto previsão, papers seminais na área mostraram que a combinação de previsões tem potencial para reduzir de maneira acentuada o erro de previsão. Especificamente, a combinação de previsões geradas por amortecimento exponencial tem sido explorada em papers recentes. Apesar da combinação de previsões utilizando Amortecimento Exponencial poder ser feita de diversas formas, um método proposto recentemente e chamado de Bagged.BLD.MBB.ETS utiliza uma técnica chamada Bootstrap Aggregating (Bagging) em combinação com métodos de amortecimento exponencial para gerar previsões mostrando que a abordagem é capaz de gerar previsões mensais mais precisas que todos os benchmarks analisados. A abordagem era considerada o estado da arte na utilização de Bagging e Amortecimento Exponencial até o desenvolvimento dos resultados obtidos nesta tese. A tese em questão se ocupa de, inicialmente, validar o método Bagged.BLD.MBB.ETS em um conjunto de dados relevante do ponto de vista de uma aplicação real, expandindo assim os campos de aplicação da metodologia. Posteriormente, são identificados motivos relevantes para redução do erro de e é proposta uma nova metodologia que utiliza Bagging, Amortecimento Exponencial e Clusters para tratar o efeito covariância, até então não identificado anteriormente na literatura do método. A abordagem proposta foi testada utilizando diferentes tipo de séries temporais da competição M3, CIF 2016 e M4, bem como utilizando dados simulados. Os resultados empíricos apontam para uma redução substancial na variância e no erro de previsão.

Palavras-chave

Bagging; Agrupamento de séries temporais; Amortecimento exponencial; Partitioning Around Medoids; Redução de variância.

Table of contents

1	Introduction	13
2	Literature Review and Contributions	15
2.1	Contributions	17
3	The state of the art of Bagging Exponential Smoothing - Bagged.MBB.BLD.ETS	19
3.1	Bagged.BLD.MBB.ETS	19
4	Bagged.BLD.MBB.ETS to an Air Transport problem	26
4.1	introduction	26
4.2	Literature Review on Air Transport Time Series Forecast	27
4.3	Data description	29
4.4	Methods	30
4.5	Forecasting Results	32
4.6	Summary and Concluding Remarks	34
5	Bagged.Cluster.ETS - A new approach to Bagging Exponential Smoothing	35
5.1	Introduction	36
5.2	Why Bagging tends to work	36
5.3	Proposed Approach - Bagged.Cluster.ETS	39
5.4	Computational Aspects of Bagged.Cluster.ETS	46
5.5	Bias and Variance analysis	50
5.6	Experiments on Forecasting Competitions Data	55
5.6.1	Effects on the number of clusters	55
5.6.2	Performance of the Proposed Approach on Competitions	57
5.6.2.1	M3 - Monthly Results	58
5.6.2.2	M3 - Quarterly and Yearly Results	59
5.6.3	Discussion	61
5.6.4	CIF 2016	62
5.6.5	Participation in the M4 Competition	66
5.7	Summary and Concluding Remark	69
6	Concluding Remarks e Future Directions	71
	Bibliography	74

List of figures

Figure 3.1	STL decomposition of the Logarithm of the CO2 series	20
Figure 3.2	MBB - Source: Petropoulos et al. in [47]	22
Figure 3.3	ETS - Source: Hyndman and Athanasopoulos in [75]	23
Figure 3.4	Bagged.BLD.MBB.ETS flowchart	25
Figure 4.1	Air Passenger demand by country	30
Figure 4.2	Bagging Holt Winters Flowchart	31
Figure 4.3	Forecasts Out of Sample by country	33
Figure 5.1	Interest Rates by Country	41
Figure 5.2	Silhouette method	42
Figure 5.3	Bagged.Cluster.ETS Flowchart	43
Figure 5.4	Bootstrapped versions from time series 1083	45
Figure 5.5	Forecasts for time series 1083	45
Figure 5.6	Map Reduce - Source: Salem in [85]	47
Figure 5.7	Map Reduce diagram - Source: Ramakrishnan et al. in [86]	48
Figure 5.8	Processing time	48
Figure 5.9	Bias and Variance from DGPs	53
Figure 5.10	sMAPE according to the number of clusters	56

List of tables

Table 4.1	sMAPE Out of Sample	33
Table 5.1	Squared Bias and Variance - AR	53
Table 5.2	Squared Bias and Variance - STAR	54
Table 5.3	Squared Bias and Variance - ETS	54
Table 5.4	Squared Bias and Variance - SARIMA	54
Table 5.5	Methodologies' comparison - M3 Monthly	58
Table 5.6	Friedman rank-sum test - M3 Monthly	59
Table 5.7	Methodologies' comparison - M3 Quarterly	60
Table 5.8	Methodologies' comparison - M3 Yearly	60
Table 5.9	Friedman rank-sum test - M3 Quarterly and Yearly	61
Table 5.10	Results with reduced variance - Bagged.Cluster.ETS	62
Table 5.11	CIF 2016 - Methodologies' comparison - Artificial Series	64
Table 5.12	CIF 2016 - Methodologies' comparison - Real Series	64
Table 5.13	CIF 2016 - Methodologies' comparison - All Series	65
Table 5.14	Friedman rank-sum test - CIF 2016	65
Table 5.15	M4 dataset	66
Table 5.16	M4 results	67

"Look up at the stars and not down at your feet. Try to make sense of what you see, and wonder about what makes the universe exist. Be curious"

Stephen Hawking, .

1 Introduction

With the acceleration of industrial development and the intensification of competitiveness, the lack of a well-designed planning stage that meets current and future needs has a direct impact on the entire production chain. Not planning properly, regardless of the size of the company, within the context of Production Systems, means incurring in several problems in the future. For example: lack of labour, raw material, among other limitations of productive processes that has as a consequence the inability to meet the demands of the consumer market.

The lack of knowledge about the future and the inherent randomness of short, medium and long term projections make forecasting methods extremely necessary in the planning and management stages of Production Systems. A forecasting method that is able to accurately predict the future has as its great asset the quantification of uncertainty so that decision making can be done more efficiently, see [1].

Having the importance of producing accurate forecasts in mind, time series forecasting researchers have been developing statistical models and methods for decades, see [38] for a comprehensive review.

Interestingly, a field called Machine Learning has become popular in recent years, driven by the evolution of computers, that allowed increasingly powerful machines to be available at low cost, and the creation of "Machine Learning as a Service" clouds which allowed powerful algorithms to be easily delivered to everybody, see [42].

The use of techniques that combine both classic statistical models and modern Machine Learning approaches are starting appear in time series forecasting literature and recent studies have shown promising results, for instance see [13] and [33].

Understanding that the results from the combination of techniques and approaches are the state of the art in time series forecasting, this thesis seeks to combine both classical time series forecasting methods and Machine Learning methods. Therefore, it starts by validating the methodology in [13] that combines Bootstrap Aggregating (Bagging) with Exponential Smoothing methods, on a real air transport demand forecast issue, expanding then the

fields of application of the method. Afterwards, this thesis seeks to identify and give intuition on why the method works. By doing so, possible aspects of improvement are identified and these improvements are packed into a new method called Bagged.Cluster.ETS, since it uses Bagging, Clusters and Exponential Smoothing.

The proposed approach is tested using simulated data to check and compare specific aspects, such as variance, and data from three forecasting competitions to assess forecast accuracy. It is shown that the proposed approach is able to not only address the drawbacks in [13], but also reducing forecast error in many cases when compared to all benchmarks.

The main motivation to create the proposed approach, hereby called Bagged.Cluster.ETS, which is also the main motivation for this thesis, is that despite of the fact that existing methods using Bagging and Exponential Smoothing to generate and aggregate a group of forecasts in order to reduce variance, none of them consider the effect of covariance among the group of forecasts, even though it could drastically impact the total variance of the group and, therefore, the forecast accuracy.

Next chapters are organized as follows: chapter 2 is dedicated to present the presents the literature review about Bagging in time series forecasting context to contextualize how this work is inserted in the state of the art. Chapter 3 explain Bagged.BLD.MBB.ETS in details. Chapter 4 presents the application of Bagged.BLD.MBB.ETS to a real problem on air transportation context, including a literature review on demand forecasting using time series forecasting methods. Chapter 5 is devoted to present the proposed approach called Bagged.Cluster.ETS. This is done by first explaining why Bagging works, the drawbacks in the methodology proposed by Bergmeir et al. in [13] and, then, proposing the methodology itself. The chapter is finalized validating the methodology using simulated data, and data from 3 competitions M3, CIF 2016 and M4. Finally, chapter 6 is dedicated to the final considerations and directions to future research.

2

Literature Review and Contributions

With the final goal of generating forecasts as accurate as possible, several time series forecasting methods have been developed in the literature, see [38]. However, while many complex methods have been proposed, it is yet remarkable that simple approaches usually obtain good forecasting performance, many times beating strong and more complex contestants, see [39] for a comparison between Dynamic Linear Models and simpler models such as the Exponential Smoothing and [33] for results on one of the latest time series forecasting competition up to this date.

Considering simplicity as a starting point, [40] presents how the Holt-Winters, a particular method from the class of Exponential Smoothing, see [37], have been evolving in order to incorporate new features such as multiple cycles, see [41]. Evolving the Exponential Smoothing is desirable, since the method is widely recognized as being both simple and highly accurate.

In [2] Breiman introduces the Bagging method as a way to reduce the prediction error using multiple versions of a predictor. Each of these versions are generated resampling the learning set using a technique called Bootstrap, see [3]. Breiman's paper is considered seminal and have been referenced thousands of times. However, despite the fact that many other papers using Bagging in the field of Machine Learning have been published, just a few of them use Bagging to improve time series forecasting performance. We next describe a chronological review of relevant works using Bagging in a time series forecasting context:

- The work of Inoue and Killian in [4] presents one of the first attempts to use Bagging in time series context. The authors concluded, using an econometric approach, that Bagging was responsible for reducing the forecast error.
- Lee and Yang in [5], used Bagging to model binary and quantile time series data (e.g. time series of sign of a financial return). A large reduction in the prediction error was found but no improvements with large samples were verified.
- Inoue and Killian in [6] proposed three variants of the Bagging algorithm in order to verify whether the inclusion of real economic activity

indicators to U.S consumer price inflation forecasting model was able to reduce the prediction Mean Squared Forecast Error (MSFE) estimates. The author's findings pointed that MSFE was indeed reduced.

- In [7], Cordeiro and Neves proposed a combination of Bagging and Exponential Smoothing in an approach called Boot.EXPOS. Although the idea from the authors was interesting, their results were not really consistent when tested using data from the M3 competition, see [22] on the original M3 competition.
- Hillebrand and Medeiros in [8] applies Bagging to a log-linear model and a non-linear specification with logistic transitions to increase forecast accuracy for realized volatility, showing that Bagging log-linear model provides larger improvements on forecast accuracy.
- Rapach and Strauss in [9] showed that a combination of Bagging and Dynamic Linear Regression models to forecast the U.S. employment growth often lead to MSFE reduction.
- Wang et al. proposed in [10] a multi-ensemble hybrid system using Bagging, Supporting Vector Machines (SVM) and Artificial Neural Networks (ANN) to make forecasts for chaotic time series. The authors demonstrated the approach using Bagging is capable of generating more accurate results in comparison with other ensemble methods and single-model SVM or ANN.
- Zontul et al. in [11] proposed a successful combination of Bagging with an algorithm called REPTree to produce wind speed forecasts in Kirklareli (Turkey). However, the lack of other regions poses as a drawback of the paper.
- The work of Jin et al. in [12] proposed a revised version of Bagging to consider the dependency in time series data. They also demonstrate that Bagging was able robustify financial time series forecasts even when models are poorly specified.
- Bergmeir et al. in [13] proposed an approach combining Bagging with exponential smoothing methods and perform an extensive evaluation by making forecasts for the M3 competition data set (645 yearly, 756 quarterly and 1428 monthly time series), demonstrating that their approach is extremely accurate, specially for monthly time series.

The results presented in [13] are quite important because they show an impressive reduction on the forecasting error for Exponential Smoothing methods, indicating an enormous potential to improve forecasts in many

fields. In this sense, in a result also verified in this thesis, Dantas and Cyrino Oliveira in [14] expanded the fields of application to the air transportation industry by showing the method proposed in [13] was able to outperform the benchmarks (SARIMA, Holt Winters, ETS, and Seasonal Naive). The paper poses itself as an important proof of the method's capacity due to its use on a real problem.

Despite the fact the work in [13] contributes significantly to an area yet poorly explored that is the combination of traditional times series and machine learning methods, the work leave some aspects unattended. First aspect is the lack of validation on other datasets than the M3 competition, since the competition, although the main reference, is a bit outdated by now. Second aspect refers to the lack of explanation or even intuition on why their method works. Finally, the third aspect derives straight from the second, since by not having a proper explanation on why their method works, the authors missed one important point of improvement that is to consider the covariance among the aggregated forecasts.

Filling the gaps left unattended in [13] is the main motivation to this thesis, where it is expected that making significant contributions to what is, up to this date, the documented state of the art in time series forecasting, would lead the whole forecasting field to better forecasts methods and practices.

2.1 Contributions

This thesis seeks to make contributions to the time series forecasting field. One important contribution lies on expanding the fields of application of the method proposed in [13]. This is done by applying the approach to real air transport demand data. This contribution has beneficial implications on how forecasts can be done in the air transportation field.

Another contribution is an analysis on why Bagging tends to improve forecast accuracy of the Exponential Smoothing method, filling a gap left unattended in [13] and showing that covariance plays a fundamental role in it.

Identifying the possible underlying reasons on why the method produces lower forecast errors is fundamental when one is trying to get even better results. This leads to the main contribution of the thesis that is to take into account aspects left unattended by the previous authors to propose a new method using Bagging, Exponential Smoothing and Clusters, called Bagged.Cluster.ETS. The method is extensively tested and compared with other methods, including the state of the art.

The first contribution has direct relation to the author's published paper

"Air transportation demand forecast through Bagging Holt Winters methods" in the *Journal of Air Transport Management*, Volume 59, March 2017, Pages 116-123 and can be found in chapter 3 and 4 of the thesis.

The core ideas and results developed in this thesis that constitute the other major contributions are found in chapter 5 and have a strong relation with the author's already published paper "**Improving Time Series Forecasting: an Approach Combining Bootstrap Aggregation, Clusters and Exponential Smoothing**" in the *International Journal of Forecasting*, Volume 34, October-December 2018, Pages 748-761.

Next chapter is devoted to explaining in details the state of the art of Bagging and Exponential Smoothing, called Bagged.BLD.MBB.ETS.

3

The state of the art of Bagging Exponential Smoothing - Bagged.BLD.MBB.BLD.ETS

This chapter is devoted to explaining in details the main aspects of Bagged.BLD.MBB.ETS, the method proposed by Bergmeir et. al in [13]. Since the method was proposed a lot of attention was given to it and papers using the methodology have been written. Besides the paper on air transportation (product of this thesis), recently, Meira and Cyrino Oliveira in [84] successfully applied the approach to the energy sector and Petropoulos et al. in [47] have explored sources of uncertainty in the method, although they have not explored the one to be presented in this thesis that is the covariance.

3.1

Bagged.BLD.MBB.ETS

The approach can be divided into five major steps: transformation, decomposition, simulation, forecasting and aggregation. The steps and the specific choices made by Bergmeir and colleagues in [13] are described next:

Step 1 - Transformation

- Transform the time series;

The first step of the approach is to transform the time series. The transformation is conducted using the family of Box-Cox transformations, see [19]. This transformations are defined as follow:

$$w_t = \begin{cases} \log(y_t) & \text{if } \lambda=0; \\ (y_t^\lambda - 1)/\lambda & \text{otherwise.} \end{cases}$$

Note when $\lambda=0$ then transformation simply reduces to natural logarithm. However it is up to the user to define the λ that suits best to each series. In Bagged.BLD.MBB.ETS, the method proposed by Guerrero in [43] is used. The reason why choosing Guerrero's method is because it is know that it acts as a variance stabilizer. Stabilizing variance is a desired feature since high variances can severely impact the performance of the subsequent steps.

Step 2 - Decomposition

- Decompose the time series into three components: seasonal, trend and remainder;

Although there are several methods to decompose time series, Bagged.BLD.MBB.ETS uses the Seasonal-Trend decomposition using LOESS (STL decomposition), proposed by in [20]. STL is an additive method that applies a sequence of smoothing operations using Locally Weighted Regression (LOESS) to decompose the time series into three components: Seasonal (S_t), Trend (T_t), and Remainder (R_t):

$$y_t = S_t + T_t + R_t \quad (3-1)$$

The method has many interesting features, such as being flexible in a sense that the amount of variation over time in the seasonal and trend components can be specified by the user, robustness to outliers and missing values. As an example, see figure 3.1 for a STL decomposition of the logarithm of Mauna Loa Atmospheric CO2 Concentration time series, see [44] for reference on the time series.

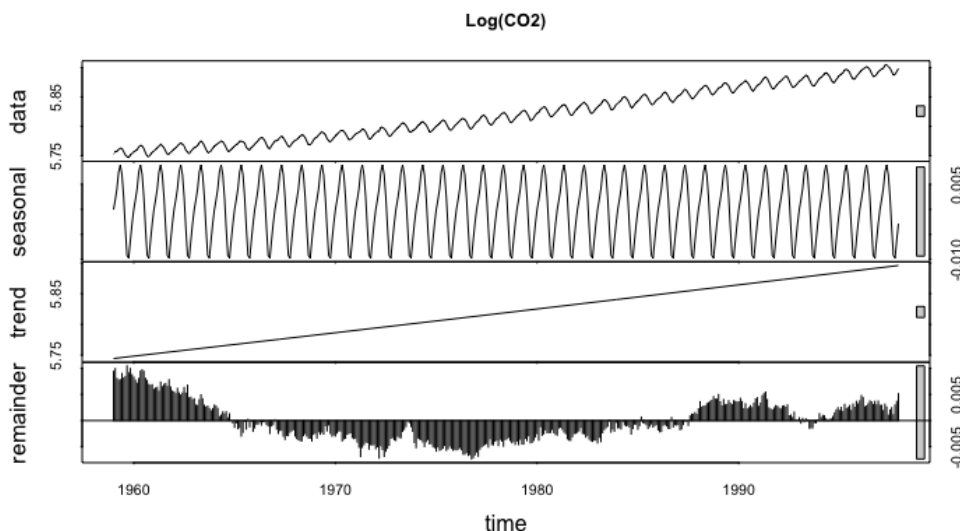


Figure 3.1: STL decomposition of the Logarithm of the CO2 series

When there is no seasonal pattern in the time series, the decomposition is made using a loess-based procedure, see [77]. In this case, the decomposition reduces to only Trend (T_t) and Remainder (R_t) components:

$$y_t = T_t + R_t \quad (3-2)$$

Step 3 - Simulation

- Generate new versions of the remainder;
- Add back the seasonal and trend components to each of the new versions of the remainders;
- Detransform the versions using the same λ used in step 1.

In order to generate new versions of the remainder the Bootstrap method, proposed by Efron in [3] is used. The method has been largely used in a wide variety of situations where there is uncertainty regarding an estimator or the quality of a method. Its great advantage lies in the fact that it can be easily implemented in a series of situations where the calculation of some parameter of interest cannot be found in a trivial way. Although extremely powerful, the method was initially developed for independent data. A theoretical proof is presented by Singh in [46], attesting the quality of the method for cases where the data is independent. However, the author indicates that for dependent data, the method elaborated by Efron in [3] is inadequate.

It is important to understand that the remainder component may still be serially correlated. This means that if one simply resamples remainder observations without taking serial correlation into account, then the dependency structure might be lost. To overcome this problem, Bagged.BLD.MBB.ETS uses Moving Block Bootstrap (MBB), a technique created by Kunsch in [45] to maintain the dependency structure almost intact by breaking the series into blocks and resampling only the blocks.

The MBB idea is to construct and randomly select blocks of consecutive observations. The method can be described as follows:

$$B_i = (y_i, \dots, y_{i+l-1}) \quad (3-3)$$

Is the block of size l , beginning at the observation y_i , from a stationary process, $1 \leq i \leq b$ where $b = n - l + 1$. Selecting by simple random sample with replacement from the set of blocks $\{B_1, B_2, \dots, B_b\}$ a new sample is formed by B_1^*, \dots, B_k^* where k is the number of selected blocks. The elements in B_i^* can be denoted by $(y_{(i-1)l+1}^*, \dots, y_{il}^*)$, $i = 1, \dots, k$. Thus, the MBB sample, of size $m = kl$ is y_1^*, \dots, y_m^* . A visual representation of the approach can be seen in the work of Fotios and colleagues in [47] reproduced in figure 3.2.

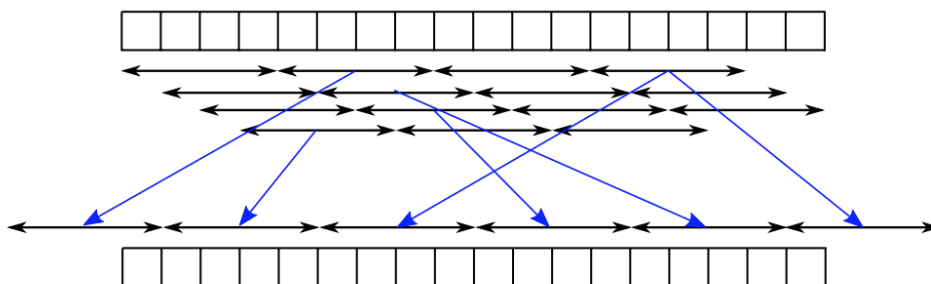


Figure 3.2: MBB - Source: Petropoulos et al. in [47]

One important aspect to consider is the size of the block. While one can determine it based on any criteria, Bergmeir and colleagues in [13] consider the size of the blocks as the double of the frequency of the time series (e.g. 24 for monthly time series and 8 for quarterly). This is justified by the fact that some seasonal effect may remain in the remainder. Thus, having blocks with this size would diminish the impact of seasonal effects not captured by the decomposition phase.

Step 4 - Forecasting

- Make forecasts for each Bootstrapped time series plus the original;

Although many forecasting methods exist (e.g. SARIMA, Neural Nets, THETA, among many others), the Exponential Smoothing is the forecasting method adopted by Bagged.BLD.MBB.ETS. The explanation given by Bergmeir and colleagues in [13] on the use of the method is due to its relevance even after 50 years of existence, see [40], simplicity and transparency and adaptiveness to changes. Also, the work of Hyndman and colleagues in [37] gave the method strong theoretical support in a form of a state space approach called ETS. The ETS models allow to produce not only point forecasts, as the Exponential Smoothing algorithm in the past did, but also an entire forecast distribution allowing to determine forecast intervals. Since ETS is an state space approach, the models have an observation equation and transition equations for each component (level, trend and seasonal) and each of the models have additive or multiplicative errors. Taking all this variations into account a total of 30 models are defined using the approach. The ETS stands for **E**rror, **T**rend, **S**easonal, and the models can be uniquely defined using the taxonomy below:

- Error={ A, M }
- Trend={ N, A, A_d, M, M_d }
- Seasonal={ N, A, M }

Where A is additive, M is multiplicative, A_d is additive damped, M_d is multiplicative damped and N is none. Figure 3.3 reproduces a table in [1] that presents the equation for each of the 30 models.

ADDITIVE ERROR MODELS			
Trend	N	Seasonal A	M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A _d	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$
M	$y_t = \ell_{t-1}b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1})$
M _d	$y_t = \ell_{t-1}b_{t-1}^\phi + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$	$y_t = \ell_{t-1}b_{t-1}^\phi + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1}^\phi + \beta\varepsilon_t/(s_{t-m}\ell_{t-1})$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1}b_{t-1}^\phi)$
MULTIPLICATIVE ERROR MODELS			
Trend	N	Seasonal A	M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A _d	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M	$y_t = \ell_{t-1}b_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1} + \alpha(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
M _d	$y_t = \ell_{t-1}b_{t-1}^\phi(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$	$y_t = (\ell_{t-1}b_{t-1}^\phi + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi + \alpha(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$ $b_t = b_{t-1}^\phi + \beta(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t/\ell_{t-1}$ $s_t = s_{t-m} + \gamma(\ell_{t-1}b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}b_{t-1}^\phi s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}b_{t-1}^\phi(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1}^\phi(1 + \beta\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Figure 3.3: ETS - Source: Hyndman and Athanasopoulos in [75]

The ETS models are implemented in R. The `ets` function from the `forecast` package implements this type of models, see [48]. If the model is not specified by the user, the `ets` function have a fully automatic selection of

model using Akaike's Information Criterion (AIC). In this thesis, version 8.0 of the package was used.

Step 5 - Aggregation

- Aggregate the forecasts to generate the final result.

The aggregation of the forecasts can be done using many measures, eg. simple mean, median, weighted mean, trimmed mean etc. However, it is important to have in mind that depending on the distribution of the forecasts, the results might change a lot. This is especially true in the presence of outliers. In this sense, Bagged.BLD.MBB.ETS uses the median as the aggregation method, since it is less sensitive to outliers. Figure 3.4 shows the flowchart of entire the approach.

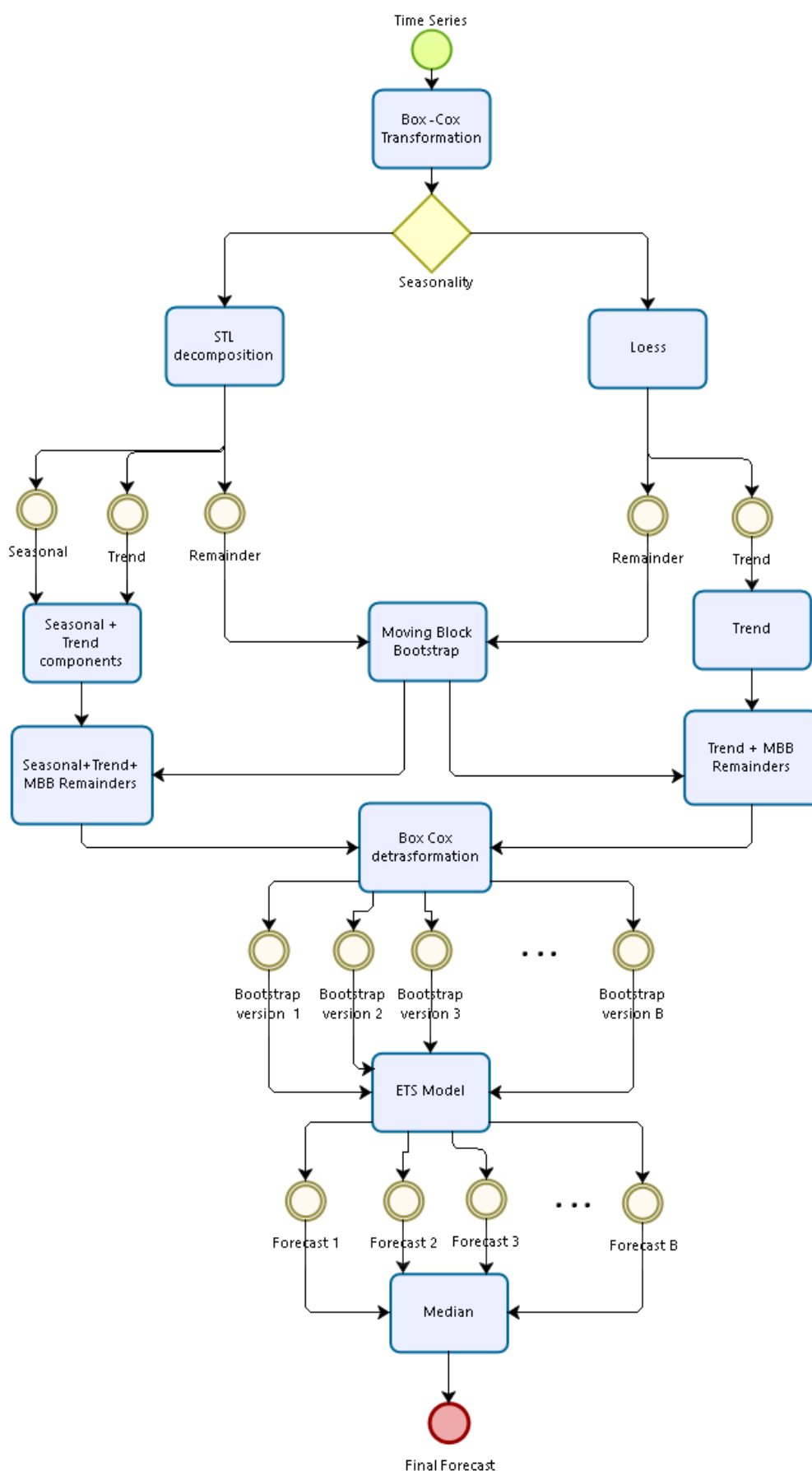


Figure 3.4: Bagged.MBB.BLD.ETS flowchart

4

Bagged.BLD.MBB.ETS to an Air Transport problem

The Bagged.BLD.MBB.ETS approach was proposed and successfully tested on data from the M3 competition. Inspired by the good results, in this chapter it is presented the first contribution of the thesis. This is an application of the method to a real air transport issue that is to forecast the air passenger demand.

In this sense, this chapter is a product of the feedbacks from the experts in forecasting and air transport about the following conference papers:

- Forecasting the Air Transportation Demand: An application at Amsterdam Schiphol Airport using Bagged Holt Winters assisted by Compression-Based Dissimilarity measure - *International Symposium on Forecasting - California - USA 2015* - Awarded with the travel grant award
- Bagging ARIMA para previsão de demanda de transporte aéreo - *Encontro Nacional de Engenharia de Produção - Fortaleza - Brazil 2015*

Furthermore, these feedbacks highly influenced not only the creation of this chapter but also this chapter's heavily based related publication in the *Journal of Air Transport Management*, one of the finest journals in the field of air transportation:

- Air transportation demand forecast through Bagging Holt Winters methods - *Journal of Air Transport Management*, Volume 59, March 2017, Pages 116-123.

4.1 introduction

The air transport sector has an enormous relevance in both social and economic aspects of country's developments, see [60]. Due to this, the problems related to the air transportation field usually are not simple and being subject to modern optimization techniques based on operations research (OR) theories, see [61] for an overview of OR techniques applied to the air transportation field.

These techniques apply to several areas such as aircraft and crew scheduling, revenue management, overbooking, leg-based seat inventory management and the planning and operations of aviation infrastructure.

Demand forecast plays an important role, since it affects the entire air transport operation. Inaccurate forecast can have disastrous consequences such as system congestion, excess infrastructure capacities, increased operator's costs, public and private drainage of funds, among others.

This chapter uses Bagged.BLD.MBB.ETS in order to generate short to medium term forecast (1 year ahead) for the air transport demand. This work was the first one to use the method in the air transport context and by the time the related paper was published, March 2017, was a novelty in the sector.

Next section briefly reviews the literature about air transportation forecasting. Section 3 explains the specific choices in the methodology. Section 4 is dedicated to present data and the evaluation metrics used in the study. Section 5 is dedicated to presenting the results. Finally, the last section presents the conclusions.

4.2

Literature Review on Air Transport Time Series Forecast

Several methods have been considered to forecasting the air transport demand. In general, two approaches are common to address this issue: causal econometric models and univariate time series models. The econometric approach seeks to identify relationships between, e.g. total air passengers and social, economic, and service-related factors. The univariate time series approach only depends on the past of the series to generate forecasts, making use of the correlation between present and past observations.

In order to check the state of the art of the air passenger demand forecast, we have conducted an online search for papers combining the words air transport and forecast. Considering only the most recent (since the year 2000), a total of 68 papers have been found. However, not all of them have a direct relationship with air transport demand forecast.

In [62] it is presented a modified version of the Holt Winters method called Holt Winters with damped trend. The authors consider monthly time series of total air Passengers at UK airports to demonstrate that their approach improves forecasting performance for long lead times. One interesting feature of their method is that it allows generating low, medium and high scenarios by just varying the future trend. Another application of forecasting was developed in [63] and predicts survival probabilities of Asian Airlines using an adapted Markov Model that takes into account the relationship between

airline growth and profitability. An application in [64] uses leading indicators such as macroeconomic variables as input for a Probit model for predicting short-term shifts in demand for business travels by air to and from the United Kingdom. The forecasts generated by the approach were shown to be more precise in comparison with the benchmark linear model.

Considering sensitive events, the work in [65] evaluate the impact of September 11 on the United States air travel demand. They consider an intervention model to show that there was a significant impact for 1 month, considering domestic demand, and 2 months, considering international demand. They also make forecasts using the approach and compare it with forecasts generated by an ARIMA model, demonstrating the superiority of their approach. Specifically to the United States, [67] analyses whether it is better to forecast the air travel demand using aggregated data at a national level or to aggregate the resulting forecasts of individual airports.

Considering Artificial Neural Networks, the work in [66] developed a multivariate hybrid approach to forecast the Brazilian air transportation demand, showing the approach was superior when compared with an econometric approach.

Regarding city-pairs, [68] not only considered an Artificial Neural Network Model but also logistic regression algorithms to forecast the probability of unconnected city-pairs being connected in the future by an air route. While, in [70] two gravity models are considered to forecast air passenger volume between city pairs.

Considering the performance of official forecasts, in [69] autoregressive and exponential smoothing methods were used to make independent forecasts for the total number of passengers at Lisbon Airport. The forecasts are compared to the one made by the government to demonstrate that the official forecasts were too optimistic.

A combination of methods is proposed in [71]. The authors put a high effort on developing a Neuro Fuzzy method based on Singular Spectral Analysis. The study uses data from the Hong Kong International Airport to show that the approach was capable of producing better forecasts when compared with several methods and models like ARIMA, Multi-Layer Perceptron, Wavelet Neural Network, Takagi-Sugeno-Kang Recurrent Fuzzy System, Classical Decomposition and Singular Spectral Analysis.

The literature review resulted in no evidence, as far as we are aware, of the application of a combined approach of Bootstrap aggregating (Bagging) and time series forecasting methods to predict air transport demand.

4.3

Data description

The data considered in the study are proxy time series for the air passenger demand (national level). A total of 14 countries was considered in the study. The majority of the countries are located in Europe (Belgium, Czech Republic, Denmark, Germany, Greece, Ireland, Italy, Netherlands, Portugal, Spain and the United Kingdom) but countries out of Europe were also considered (Australia, Brazil and the United States of America). For the European countries the monthly time series of total passengers carried was used and the data was obtained from EUROSTAT. The Australian data was obtained from the Bureau of Infrastructure, Transport and Regional Economics (BITRE). The Brazilian data provided by the National Civil Aviation Agency of Brazil (ANAC) and for the United States, data from the Bureau of Transportation Statistics (BTS) for the USA was used.

The period considered in the majority of the countries was from January 2007 to December 2014. The only exceptions are Ireland and the United Kingdom which data runs from January 2003 to December 2014. In all cases, data before 2014 was used as training set (in-sample) and from January 2014 to December 2014 was used as a test set (out-of-sample).

The series in the study, exhibit visual trends. However, the perceived growth is not homogeneous in all 14 countries. For example, Europe and USA have slower growth rates in comparison to Australia and Brazil. Another example is the perceived effects of the world economic crisis in 2008 and 2010 that even persisted in the subsequent years in some countries, such as the case in Ireland, Spain, Greece, and U.K. Portugal. The non-European countries behaved in another way. The effects of the crisis lasted longer in the USA and a stationary period can be seen. The growth rates in Brazil were interestingly high and the country has only experienced the effects of the crisis and other internal problems after 2014. Finally, Australia experienced high growth rates. Despite the differences found in each country, it is clear that there is a strong seasonal pattern governing all time series, see 4.1.

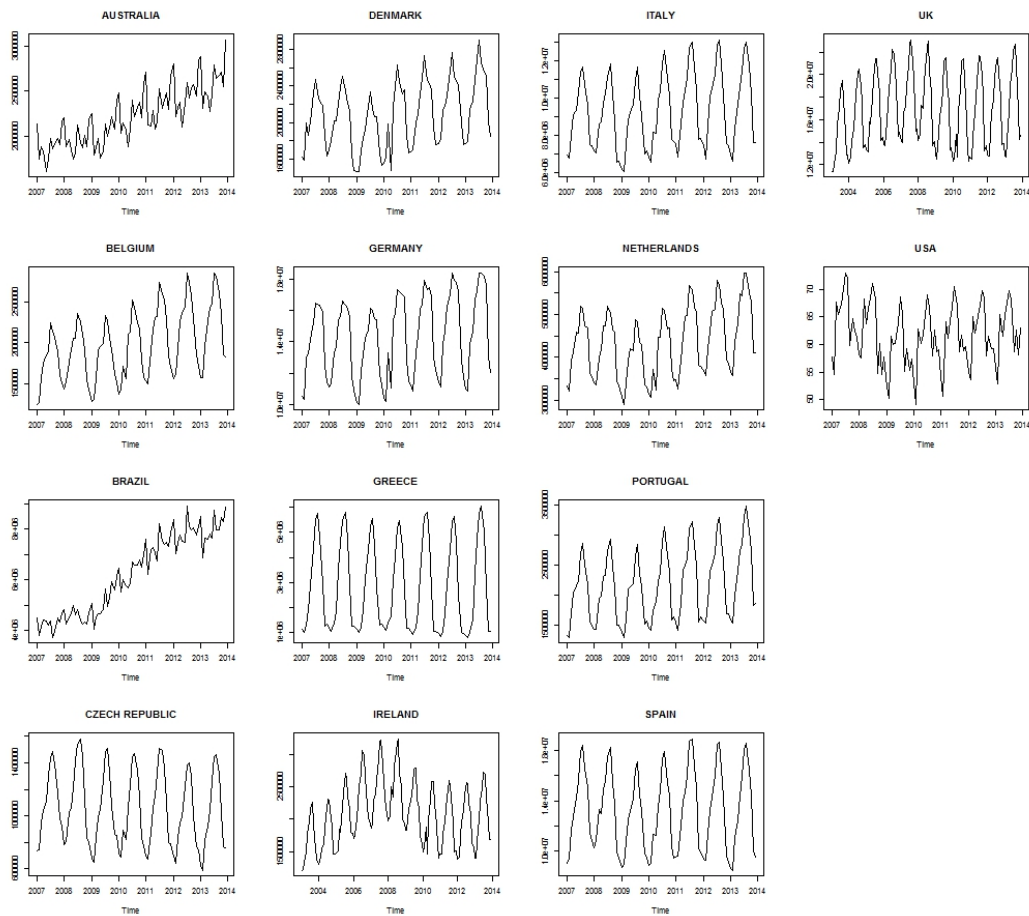


Figure 4.1: Air Passenger demand by country

4.4 Methods

The Bagged.BLD.MBB.ETS showed in chapter 3 is used, but it is also considered a particular case to specifically address the known properties of the time series. In this sense, knowing that all time series presents a strong seasonal pattern and in some cases, trend, the model in the Exponential Smoothing taxonomy are restricted do MAM and AAM (Holt Winters method). Also, the lack of evidence that the results were getting any better after the transformation-detransformation process led to the decision of not using the Box-Cox transformation. It its important to understand that the choices made are, in fact, particular cases of the original approach, since one of the possible Box Cox transformations is the identity and Holt Winters is part of the ETS family of models. Figure 4.2 shows a flowchart of the approach considering these choices.

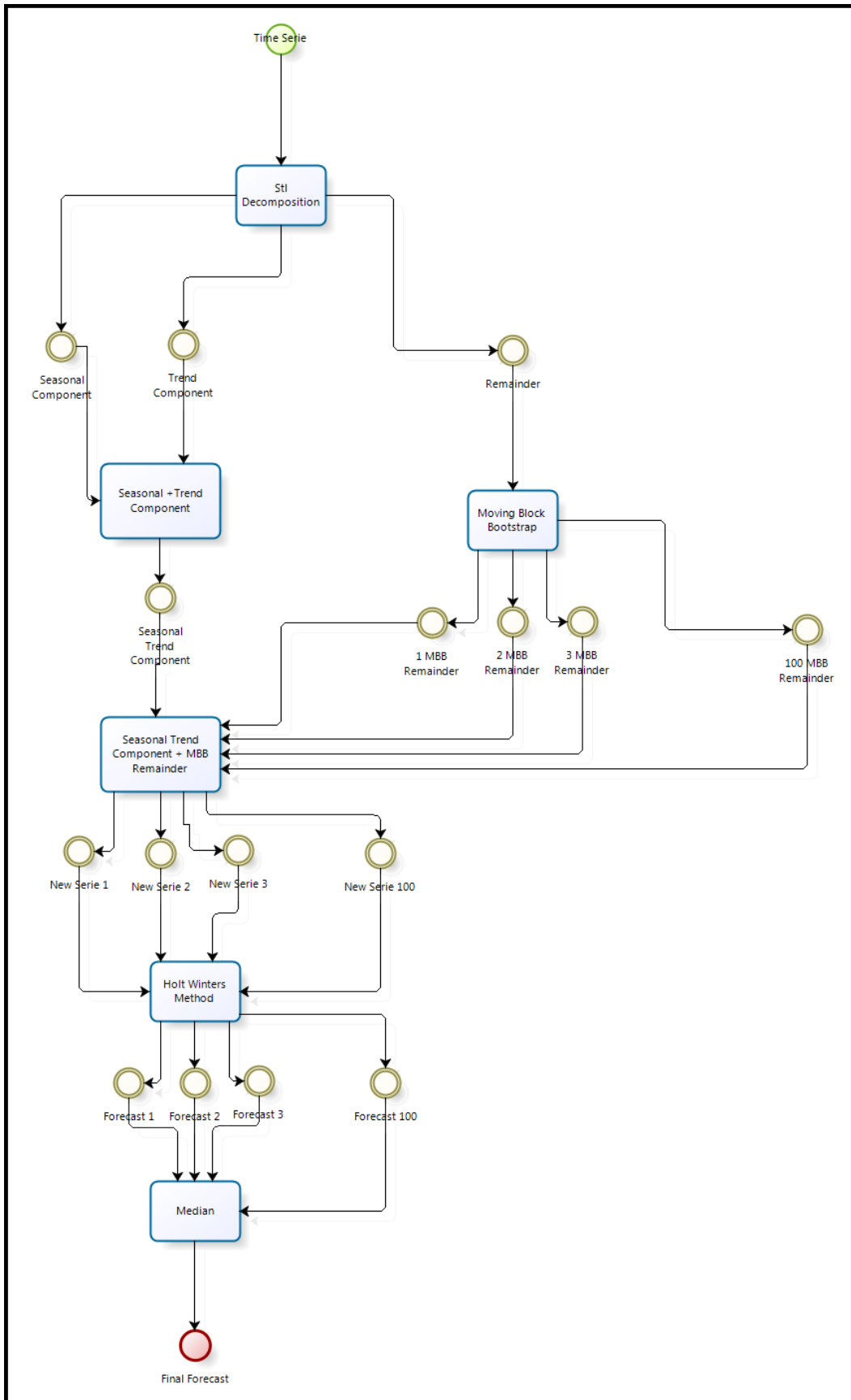


Figure 4.2: Bagging Holt Winters Flowchart

In order to access the quality of the approach the Symmetric Mean Absolute Percentage Error (sMAPE), defined by Makridakis in [72] was used. The metric gives a symmetric penalty to positive and negative value and is scale-free. This allows direct comparison between time series with different scales. See equation 4-1.

$$sMAPE = \frac{2}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{|y_t| + |\hat{y}_t|} \quad (4-1)$$

The method proposed by Bergmeir and colleagues in [13], Bagged.BLD.MBB.ETS, and its restricted version, Bagged Holt Winters, was compared to other four benchmarks: SARIMA, Exponential Smoothing (both the state-space approach called ETS and Holt Winters) and Seasonal Naive.

4.5 Forecasting Results

The results presented in table 4.1 indicated that Bagged.BLD.MBB.ETS was able to substantially reduce the forecast error in the majority of the countries in comparison to its single counterparts (ETS to Bagged.BLD.MBB.ETS and Holt Winters to Bagged Holt Winters). The sMAPE of the approaches using Bagging was on average 20% lower than their single counterparts additive Holt Winters, multiplicative Holt Winters and ETS. Specifically, Bagging with the multiplicative Holt Winters led to a reduction of, approximately, 33% in comparison to a single multiplicative Holt Winters. The case with Bagging with additive Holt Winters led to a reduction of, approximately, 17% when comparing to a single additive Holt Winters. Finally, Bagging with ETS models led to a reduction of, approximately 11%, in comparison to a single ETS model.

Comparing the best approaches using Bagging and the benchmarks (ETS, Holt Winters, SARIMA and Seasonal Naive) there were reductions in sMAPEs in almost all countries. In this sense, the sMAPE of the best Bagging approach was on average 32% lower in comparison to the best Holt Winters in each country, 37% in comparison to SARIMA, 38% in comparison to ETS and 43% in comparison to Seasonal Naive. The forecasts using the best model with Bagging along the actual values are shown in figure 4.3.

One of the possible explanations on why Bagging with ETS performed worse than with Holt Winters is due to the automatic selection of models. While using Bagging with additive or multiplicative Holt Winter only generate

an ensemble of forecasts created by the same model with different parameters, the ETS makes an ensemble of forecasts created by more than one model with different parameters.

Table 4.1: sMAPE Out of Sample

Countries	Bagging Holt Winters		Holt Winters		Bagged.BLD.MBB.ETS	ETS	SARIMA	Seasonal Naive
	Multiplicative	Additive	Multiplicative	Additive				
Australia	1.71	2.28	2.01	2.22	1.93	2.05	1.94	5.55
Belgium	3.73	6.47	4.52	7.81	7.16	7.47	5.82	8.74
Brazil	2.32	2.76	4.80	4.30	2.32	2.68	3.89	6.06
Czech Republic	2.88	1.74	3.39	1.85	1.95	2.45	3.57	2.39
Denmark	2.97	4.85	4.39	6.20	5.36	5.98	4.49	5.80
Germany	2.66	3.40	3.68	4.53	3.65	4.46	3.36	3.11
Greece	8.81	13.63	19.76	17.63	13.93	18.58	16.90	14.51
Ireland	5.73	4.98	7.48	6.45	5.72	7.37	6.68	7.32
Italy	2.54	3.86	4.70	5.71	3.76	5.17	3.17	2.77
Netherlands	1.63	2.71	2.36	4.04	2.09	3.38	3.88	4.87
Portugal	3.09	6.83	5.01	8.98	5.55	6.06	4.98	8.98
Spain	3.90	4.65	5.19	5.36	4.80	5.75	4.56	4.58
United Kingdom	2.55	3.22	3.51	4.31	4.20	4.93	4.09	4.49
United States	1.09	1.75	2.25	1.23	2.09	1.21	8.22	2.54

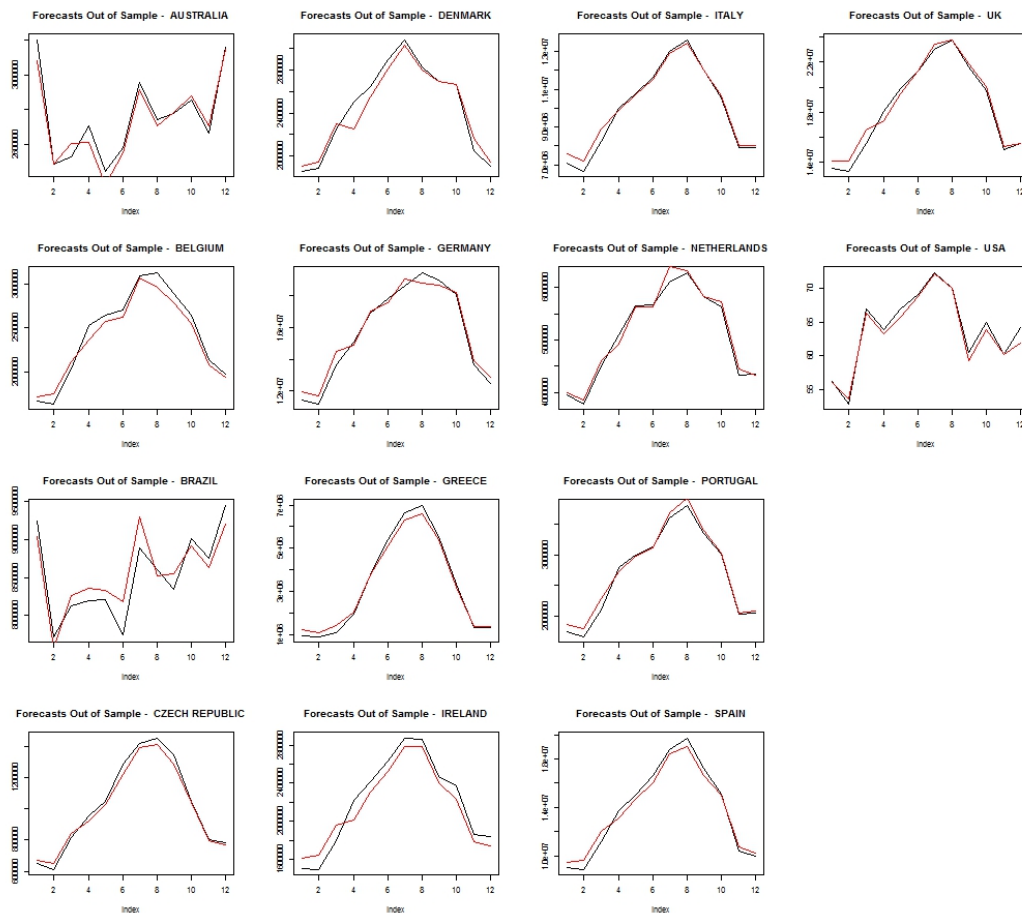


Figure 4.3: Forecasts Out of Sample by country

4.6

Summary and Concluding Remarks

The level of competitiveness in the air industry makes the relevance of reductions in the forecast error even more desirable, since the optimization of processes and the corresponding costs are usually discussed at decimal scale. In this sense, methods that can improve accuracy of demand forecast are very much appreciated.

In practice, reducing forecast error means that management decisions can be taken in a more precise way and the relation between profit and cost can be better estimated. In the air transport industry, it has a huge impact. Taking the case of airport management as an example, a forecast that overestimates the demand makes the operating unit costs higher (e.g., unnecessary staff, idle equipment) while the forecasts that underestimate the demand can lead to a reduction in the level of service (e.g., lack of staff, queues, higher waiting times).

In this chapter, for the first time in the literature, a combination of Bagging and Exponential Smoothing was applied to the air industry to predict future demand for air transportation. The method employed is Bagged.BLD.MBB.ETS proposed in [13].

The Bagged.BLD.MBB.ETS approach was able to consistently produce forecasts with reduced error when compared to other benchmarks: SARIMA, Holt Winters, ETS and Seasonal Naive. Among the 14 countries in the study, the method was able to reduce sMAPE in 13 of them. This result is an indication that the method should also produce satisfactory results if applied to other countries and can be seen as an important tool for the air industry.

5

Bagged.Cluster.ETS - A new approach to Bagging Exponential Smoothing

This chapter presents the main findings of the paper. After applying and stressing the method proposed by Bergmeir and colleagues, the obvious next step was to propose improvements. Although the solution seems pretty obvious now, it was a hard process that consumed months of thinking, writing and coding. During the process, some of the findings have been presented to other researchers in the following national and international events:

- Improving times series forecasting: an approach combining Bagging, Clusters and Exponential Smoothing - *International Symposium on Forecasting - Santander, Spain 2016*
- Improving times series forecasting: an approach combining Bagging, Clusters and Exponential Smoothing - *Encontro de Pesquisa e Pós-Graduação em Engenharia de Produção - Curitiba, Brazil 2017* *Paper awarded with Honourable Mention in statistical analysis.
- Bagging-Clustering Methods to Forecast Time Series - *International Symposium on Forecasting - Cairns, Australia 2017*
- Forecasting energy consumption combining Bagging and Clusters - *International Symposium on Forecasting - Boulder, USA 2018*

Every lesson learned during this period led to a novel approach called Bagged.Cluster.ETS. The proposed methods and the results were published on the première time series forecasting journal called *International Journal of Forecasting*. Therefore, this chapter is heavily based on the following work:

- Improving Time Series Forecasting: an Approach Combining Bootstrap Aggregation, Clusters and Exponential Smoothing - *International Journal of Forecasting*, Volume 34, October-December 2018, Pages 748-761.

5.1 Introduction

The proposed approach is indeed a powerful way to improve the forecasting performance of ETS models. However, there are unattended aspects in the approach that can increase the forecast error. The way the ensemble is constructed, via Bootstrap approach, might produce correlated forecasts (high covariance effect), affecting the Mean Squared Error of Forecast (MSFE). The MSFE is composed by a sum of variances and squared bias. Therefore, if one is interested in reducing the forecast error, the variance and the bias are the fundamental keys to better results.

Inspired by these ideas, this chapter proposes an approach, Bagged.Cluster.ETS, to reduce the forecast error by reducing the ensemble's variance. Along the explanations of the proposed approach, an extensive simulation is conducted, in order to access and compare bias and variance with the Bagged.BLD.MBB.ETS. The forecast itself is also verified using data from more than 103,000 time series from 3 different competitions M3, CIF 2016 and the recent M4.

5.2 Why Bagging tends to work

In order to propose an approach that generates lower forecast error it is vital to understand the reasons why Bagging works. In a recent work, Petropoulos and colleagues, see [47], describe some reasons why Bagging actually works. According the authors, Bagging is successful because it handles three sources of uncertainty: data, model and parameter uncertainty. The authors also argue that the main benefits of Bagging come from the model uncertainty that makes different models be selected as optimal due to the bootstrapped versions.

Despite the insights in [47], variance itself is one important missing point not addressed by the authors. In this sense, the work of Inoue and Kilian in [4] presents the properties of MSFE showing that it can be decomposed into three additive terms: variance of the real values, squared bias and variance of the forecasts:

$$\begin{aligned}
 MSFE &= E[(y_{t+1|t} - \hat{y}_{t+1|t})^2] = \\
 &= E[(y_{t+1|t} - E(y_{t+1|t}))^2] + [E(y_{t+1|t}) - E(\hat{y}_{t+1|t})]^2 + E[(\hat{y}_{t+1|t} - E[\hat{y}_{t+1|t}])^2] \\
 &= E[(y_{t+1|t} - E(y_{t+1|t}))^2] + [E(y_{t+1|t}) - E(\hat{y}_{t+1|t})]^2 + Var(\hat{y}_{t+1|t}) \\
 &= E[(y_{t+1|t} - E(y_{t+1|t}))^2] + bias(\hat{y}_{t+1|t})^2 + Var(\hat{y}_{t+1|t}) \\
 &= Var(y_{t+1|t}) + bias(\hat{y}_{t+1|t})^2 + Var(\hat{y}_{t+1|t})
 \end{aligned} \tag{5-1}$$

Note since $Var(y_{t+1|t})$ is the variance that comes from the data it cannot be controlled, and there is not much one can do, regardless the forecasting method, to make it smaller. However, the bias, $bias(\hat{y}_{t+1|t})^2$, and the variance, $Var(\hat{y}_{t+1|t})$, are strictly derived from forecasting method. This means that one can, potentially, affect them in order to obtain better results. In other words, if one is seeking to reduce MSFE, the last two terms are the key to do it.

The average forecast of the Bootstrapped forecasts can be written as:

$$\tilde{y}_{t+1|t} = \frac{1}{B} \sum_{i=1}^B \hat{y}_{(i)t+1|t}^* \tag{5-2}$$

where $\hat{y}_{(i)t+1|t}^*$ indicates the forecast for instant $t + 1$ at time t , using the Bootstrapped version i , and B is the total number of Bootstrap samples.

Because Bagging resamples the learning set, it is expected the average group of the Bootstrapped forecasts would have similar bias but reduced variance. This is due to the expressions below:

$$\begin{aligned}
 bias(\tilde{y}_{t+1|t}) &= \\
 &= E[\frac{1}{B} \sum_{i=1}^B \hat{y}_{(i)t+1|t}^*] - E[y_{t+1|t}] = \\
 &= \frac{1}{B} \sum_{i=1}^B bias(\hat{y}_{(i)t+1|t}^*)
 \end{aligned} \tag{5-3}$$

$$Var(\tilde{y}_{t+1|t}) = \frac{1}{B^2} \sum_{i=1}^B Var(\hat{y}_{(i)t+1|t}^*) + \frac{1}{B^2} \sum_{i \neq i'} Cov[\hat{y}_{(i)t+1|t}^*, \hat{y}_{(i')t+1|t}^*] \tag{5-4}$$

Note that according equation 5-3, if the Bootstrap versions are unbiased, then an unbiased ensemble is constructed. In this sense, if one already has an unbiased forecast then Bagging isn't very effective in bias reduction. However, equation 5-4 shows Bagging's real strengths since it shows that if the variances of the Bootstrapped forecasts are similar and there is no correlation among them, the variance of the group reduces to:

$$Var(\tilde{y}_{t+1|t}) \approx \frac{1}{B} Var(\hat{y}_{(1)t+1|t}^*) \tag{5-5}$$

One big drawback in the work of Bergmeir and colleagues, [13], and in all previous related works is that the authors do not give proper treatment to the covariance effect. This means that despite the fact that forecast error is reduced in previous works, the second part of the equation 5-4 is neglected. This has an important effect on the overall results, since increasing the covariance might increase the variance of the group and consequently increase the forecast error. This is the main gap that this thesis tries to bridge.

It is worth mentioning that Leo Breiman in a seminal paper, see [23], presented the famous Random Forest method as an extension of the Bagging Trees method that deals with the correlation among the ensemble. This is done by providing randomization of possible selected variables and is a clever solution for static data. The final result is a less correlated ensemble that usually results in lower prediction error.

5.3

Proposed Approach - Bagged.Cluster.ETS

In this section we explain the proposed approach called Bagged.Cluster.ETS. The approach combines Bagging, Cluster and Exponential smoothing models. The approach has similarities with Boot.Expos, see [7] and Bagged.BLD.MBB.ETS. The later share common steps with Bagged.Cluster.ETS since the process to generate the bootstrapped versions is the same. However, Boot.Expos and Bagged.BLD.MBB.ETS don't give any treatment to the covariance effect presented in equation 5-4 nor provide any means to prevent forecasts with large bias to be selected. The reduction of bias and variance is something desirable, but hardly achievable due to a well known trade-off between bias and variance. This means that usually a reduction in variance is followed by an increase in bias and vice versa, see [24] for more details on the bias variance trade-off.

Among the forecasts produced using the bootstrapped versions of the original time series, there are some that perform better in terms of accuracy. However, the forecasts that perform poorly have the potential to reduce the overall forecast performance of the group by affecting the bias of the group.

In order to prevent the inclusion of poorly accurate forecasts in the ensemble, the idea is to use a validation set as big as the amount of data that will be forecasted. In the cases where there is not enough data is available, it is recommended to use a validation set of equal size to the frequency of the time series. Thus, the proposal is to generate a considerable high number of Bootstrapped version (at least 1000) and aggregate only a fraction of the versions (e.g. 10%) that generate the best forecasts, considering some metric (e.g. sMAPE, MASE among others), in the validation set. In this thesis, the number of generated series is set to 1000 and of selected series is set to 100. The number of selected series follows the work of Bergmeir and colleagues, see [13], and is big enough in terms of convergence.

Regarding the effect covariance effect, the proposal to reduce the correlation among the ensemble is to create clusters of Bootstrapped series. Cluster methods tend to group together objects, in this case time series, that have lower distance. This means that series inside a cluster tend to have maximal similarity and series in different clusters tend to have minimum similarity. In this sense, the main idea is that selecting time series from different clusters would lead to an ensemble less correlated that would produce less correlated forecasts.

A great review of clustering techniques, showing not only methods but applications, can be found in [25]. In order to choose a clustering method, one

has to define what similarity is in their own context. Defining similarity means to choose a similarity/dissimilarity distance. Several distances are exemplified by Montero and Vilar, see [26], using their R package **Tsclust** (the same used in this thesis). The Euclidean distance is used in the proposed approach. Our interest here is to group profiles of series by pairing each pair of series and mapping them in a one-to-one scheme which makes the Euclidean distance a good match for it.

Moreover, the Euclidean distance produces good results with faster outputs when compared to other distances tested in this thesis such as Fréchet (see [51]), Dynamic Time Warping (see [49]), an adaptive distance/dissimilarity index covering both proximity on values and on behavior (see [51]), distances based on autocorrelations (see [52], [53] and [54]), distances based on periodograms (see [55] and [53]), a distance based on the discrete wavelet transform (see [56]), distances based on the symbolic representation SAX (see [57]) and distances based on compression (see [58] and [59]). In some cases, the calculation of the mentioned distances more than tripled the computational time, since some of them include also model estimation, with no clear gain in the final performance of the approach and, therefore, not justifying using a more complex distance/dissimilarity metric.

Regarding the clustering technique, the Partitioning Around Medoids (PAM) clustering algorithm was chosen due to its velocity and robustness to outliers, see [34]. PAM is a non-hierarchical clustering method that operates trying to discover iteratively objects that are defined as medoids, in this case time series, that are located in the center of a cluster. This is done by reducing the average dissimilarity of the objects to the object defined as the medoid.

The number of clusters, k , needs to be defined by the user. One way of doing it is using cross-validation. However, depending on the number of time series to be forecasted this might not be feasible since it is a computationally intensive process. As an alternative, one can use an automatic way to do it by using the Silhouette Information. The method measures similarity of an object to its cluster. It does so by calculating the following quantity:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (5-6)$$

where $a(i)$ is the average distance of time series i to all time series in its own cluster, and $b(i)$ is the lowest average distance between time series i to all other time series in any other clusters.

Note that $s(i)$ vary from -1 to 1, where higher values indicate that the time series is correctly allocated to its clusters. The ideal number of clusters is obtained using the average of the silhouette method, see [27]. To exemplify the approach, consider the example using data from OECD on monthly interest rates (10-year bonds) running January 1995 to November 2012 from: Ireland, Japan, Portugal, Switzerland, UK and USA. This data was subject analysis in [26]. See figure 5.1.

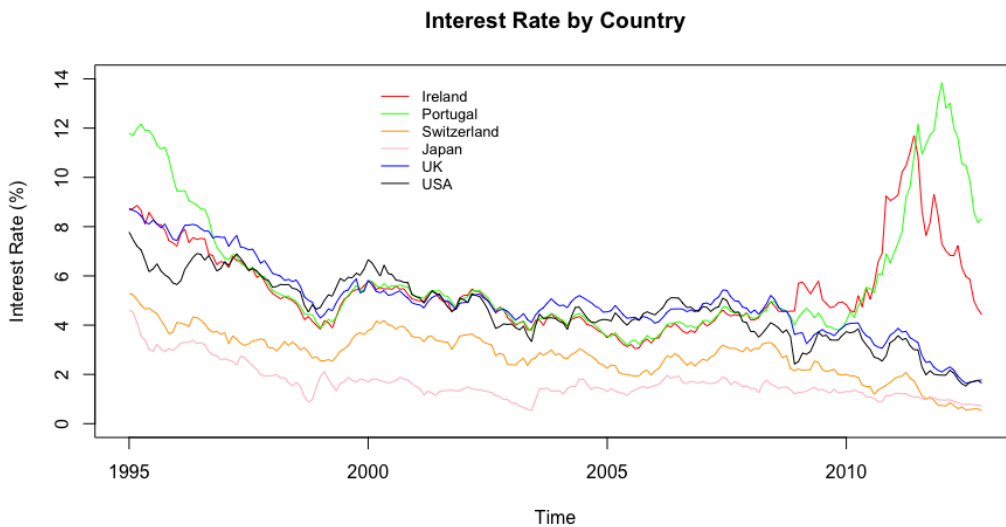


Figure 5.1: Interest Rates by Country

In order to generate the clusters the PAM algorithm with Euclidean distance was used and the Silhouette Information was calculated, leading to the ideal number of 3 clusters (Portugal and Ireland; Switzerland and Japan; UK and USA), since the average of the Silhouette Information is maximized when k is set to 3, see figure 5.2.

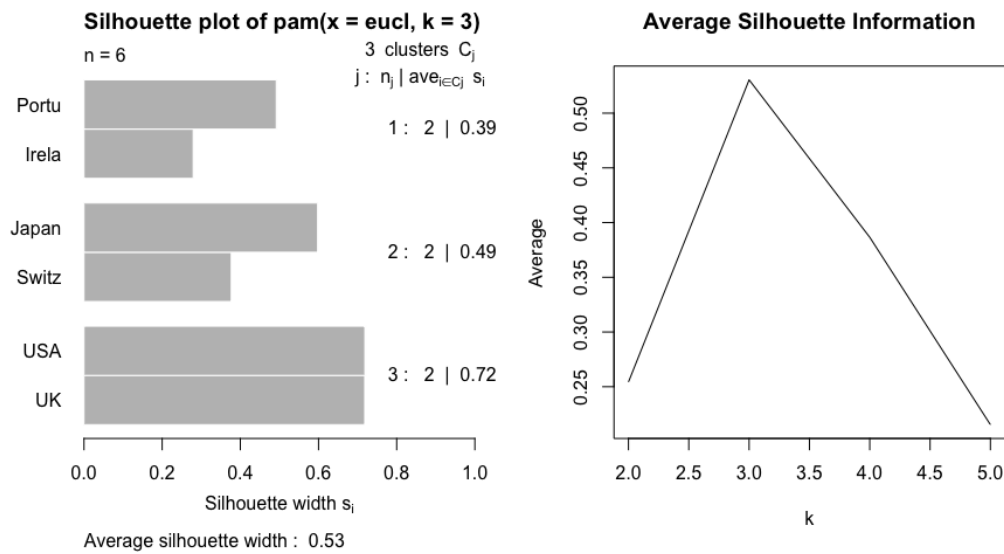


Figure 5.2: Silhouette method

An important aspect still not mentioned is the number of series selected in each cluster. To do so, we define the number of series selected in each cluster proportional to the size (total number of series) of each cluster. Thus, the number of series, n_h , picked in each cluster h is defined as follows:

$$n_h = \frac{N_h}{N} * n \quad (5-7)$$

Where n_h is the number of time series selected in each cluster h , n is the number of series to be aggregated (e.g. 100 time series), N_h is the total number of time series in cluster h and N is the total number of Bootstrapped versions.

To illustrate the approach a flowchart is presented in figure 5.3. The approach can also be divided into two parts: First part is devoted to generate the Bootstrap versions and the second to manipulating the versions in order to obtain the best results. The first part uses the same algorithm proposed by Bergmeir and colleagues in [13] and presented in algorithm 1. The second part implements the ideas exposed in this section, see algorithm 2.

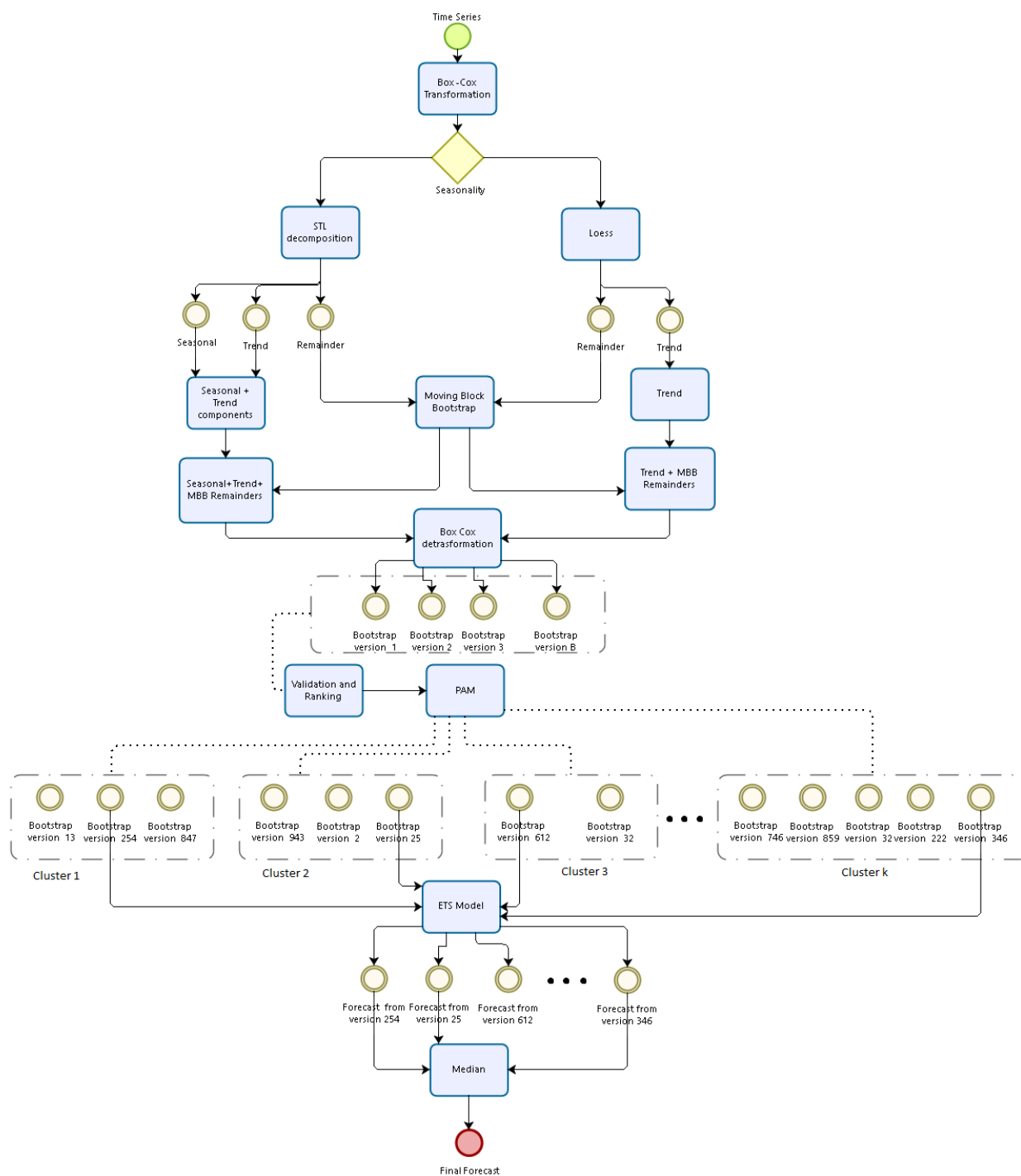


Figure 5.3: Bagged.Cluster.ETS Flowchart

Algorithm 1 Generating bootstrapped series

```

1: procedure BOOTSTRAP(ts,num.boot)
2:    $\lambda \leftarrow \mathbf{BoxCox.lambda}(ts,min=0,max=1)$ 
3:   ts.bc  $\leftarrow \mathbf{BoxCox}(ts,\lambda = 1)$ 
4:   if ts is seasonal then
5:     [trend seasonal remainder]  $\leftarrow \mathbf{stl}(ts.bc)$ 
6:   else
7:     seasonal  $\leftarrow 0$ 
8:     [trend,remainder]  $\leftarrow \mathbf{loess}(ts.bc)$ 
9:   end if
10:  recon.series[1]  $\leftarrow ts$ 
11:  for i in 2 to num.boot do
12:    boot.sample[i]  $\leftarrow \mathbf{MBB}(\text{remainder})$ 
13:    recon.series.bc[i]  $\leftarrow \text{trend} + \text{seasonal} + \text{boot.sample}[i]$ 
14:    recon.series[i]  $\leftarrow \mathbf{InvBoxCox}(\text{recon.series.bc}[i],\lambda)$ 
15:  end for
16:  return recon.series
17: end procedure

```

Algorithm 2 Proposed Approach

```

1: procedure PROPOSED_APPROACH(recon.series,k)
2:   if length(forecast.period) < prop*length(recon.series) then
3:     pseudo.recon.series  $\leftarrow$  recon.series without the last length(forecast.period) observations
4:   else
5:     pseudo.recon.series  $\leftarrow$  recon.series without the last frequency(recon.series) observations
6:   end if
7:   for i in 1 to num.boot do
8:     pseudo.model[i]  $\leftarrow \mathbf{ets}(\text{pseudo.recon.series}[i])$ 
9:     pseudo.forecast[i]  $\leftarrow \mathbf{forecast}(\text{pseudo.model}[i])$ 
10:    pseudo.forecast.sMAPE[i]  $\leftarrow \mathbf{sMAPE}(\text{pseudo.forecast}[i])$ 
11:   end for
12:   cluster.1, cluster.2, ..., cluster.k  $\leftarrow \mathbf{PAM}(\text{recon.series})$ 
13:   for h in 1 to k do
14:     ensemble  $\leftarrow$  select  $n_h$  recon.series from cluster.h with lowest rank(pseudo.forecast.sMAPE)
15:   end for
16:   for j in 1 to 100 do
17:     model[j]  $\leftarrow \mathbf{ets}(\text{ensemble}[j])$ 
18:     ensemble.forecast[j]  $\leftarrow \mathbf{forecast}(\text{model}[j])$ 
19:   end for
20:   final.forecast  $\leftarrow \mathbf{median}(\text{ensemble.forecast})$ 
21:   return final.forecast
22: end procedure

```

As an example, time series 1083 from the M3 competition dataset is used to demonstrate the two algorithms. The algorithm 1 can be visualized in figure 5.4 where the original time series is presented (series in black) on the left and the bootstrapped versions (series in blue) generated from it are presented on the right. Note the bootstrapped versions introduce the noise to the original time series that leads to different forecasts. Figure 5.5 presents algorithm 2 applied to the bootstrapped versions shown in figure 5.4. On the left side, it shows the forecasts from the series selected in each cluster (forecasts with

the same color belong to the same cluster) and on the right side it shows the aggregated forecast using the median.

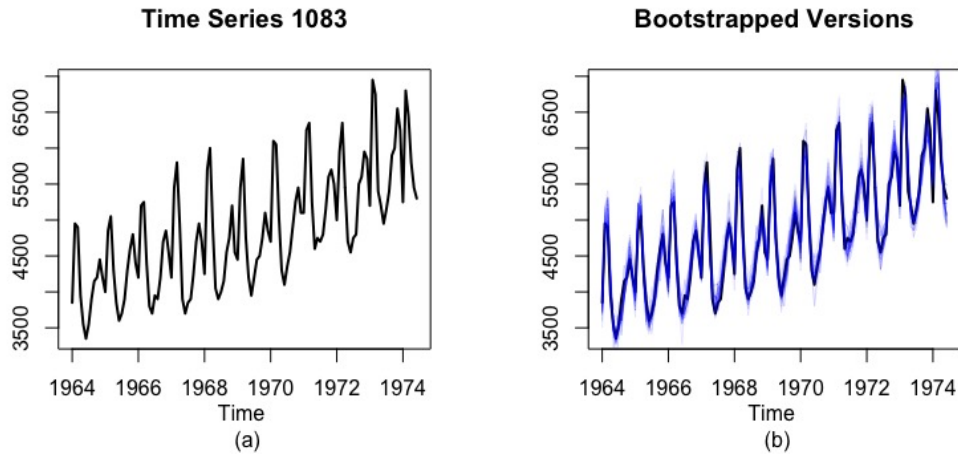


Figure 5.4: Bootstrapped versions from time series 1083

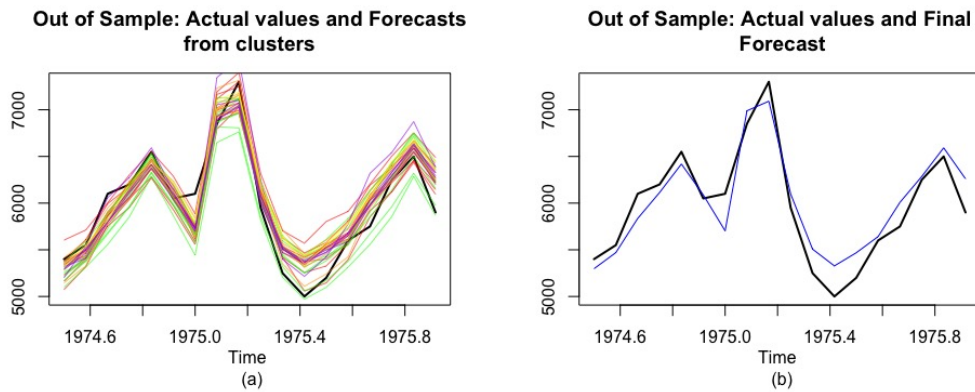


Figure 5.5: Forecasts for time series 1083

5.4 Computational Aspects of Bagged.Cluster.ETS

The proposed approach modifies the existing Bagged.BLD.MBB.ETS method in order to obtain the ultimate goal that is better forecasts. However, one important aspect of it is the feasibility of the approach in terms of computational time. In practice, this means that if one has to generate forecasts under limited time, the process has to be fast.

By the time this thesis was written there was only one implementation of Bagged.BLD.MBB.ETS. This is a function called **baggedETS** in the **forecast** package from R. This function was used to generate the comparisons in this thesis study and used as the base to the construction of the **baggedClusterETS** function that implements the approach described in this chapter.

Although it was important to use the same function as the previous author's to be sure the results were not being influenced by the way the methodology was coded, **baggedETS** is a slow function that suffer mainly because of the number of computations required. One point not considered by the authors was that, despite the fact that Bagging can be slow depending on the number of replicas generated by Bootstrap, the method could be easily parallelized, something not implemented in their code.

In this sense, parallelization can be done using Google's general Map Reduce idea, see figure 5.6, proposed by Dean and Ghemawat in [80] and customized to machine learning problems by Chu and colleagues in [79]. The later demonstrated that algorithms fitting the Statistical Query Model can be written in a summation form. This summation can be expressed in a Map Reduce framework, since the summation over the data points can be batched.

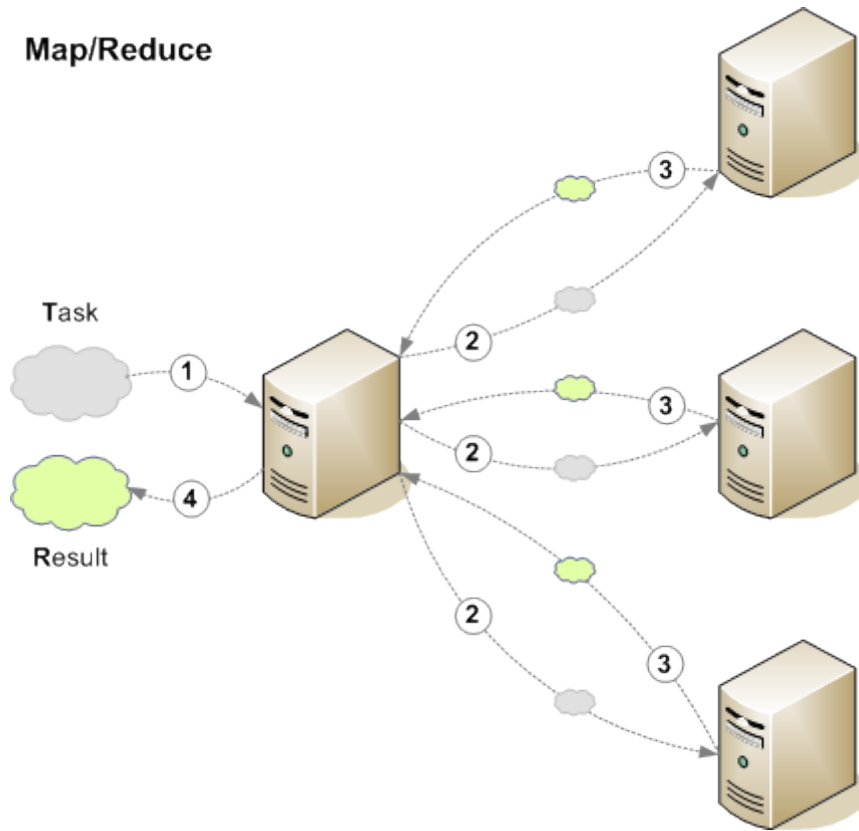


Figure 5.6: Map Reduce - Source: Salem in [85]

Occurs that Bagging can be thought this way, where the summation in equation 5-2 corresponds to a reduce operation and the $\hat{y}_{(i)t+1|t}^*$, is generated by the map function $\hat{f}(y_{(i)t}^*)$, an ETS model, see figure 5.7. This allows to distribute the calculations over the machines (or cores in the case of a single machine with multiple cores). Taking this into account, the implementation of Bagged.Cluster.ETS, **baggedClusterETS** function, parallelizes parts of the calculations, making the process much faster and scalable, which is not the case for the implementation of Bagged.BLD.MBB.ETS, **baggedETS**. This means that if one get access to more cores/machines, something easily achievable using cloud services, the code run faster. In fact, in the development of this thesis, in several occasions, machines in the cloud (Google and AWS Amazon) with a large number of cores (16, 32 and 64) were used, making the process feasible.

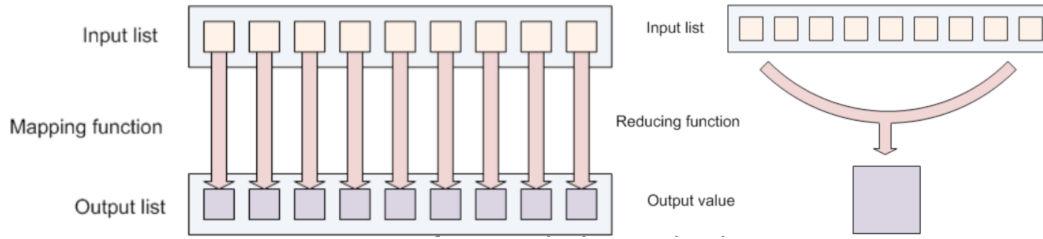


Figure 5.7: Map Reduce diagram - Source: Ramakrishnan et al. in [86]

In order to exemplify how parallel computing can speed up the computing process, a m4.16xlarge machine (188 ECUs, 64 vCPUs, 2.3 GHz, Intel Xeon E5-2686v4, 256 GiB memory) on the AWS Amazon cloud was used to generate 12 months ahead forecasts for the monthly totals of accidental deaths in the USA from 1973 to 1978, see [74] for reference on the time series. Figure 5.8 illustrates how the computational time decreases for **baggedClusterETS** (the implementation of the proposed approach using R) as the number of CPUs increases.

PUC-Rio - Certificação Digital N° 1421392/CA

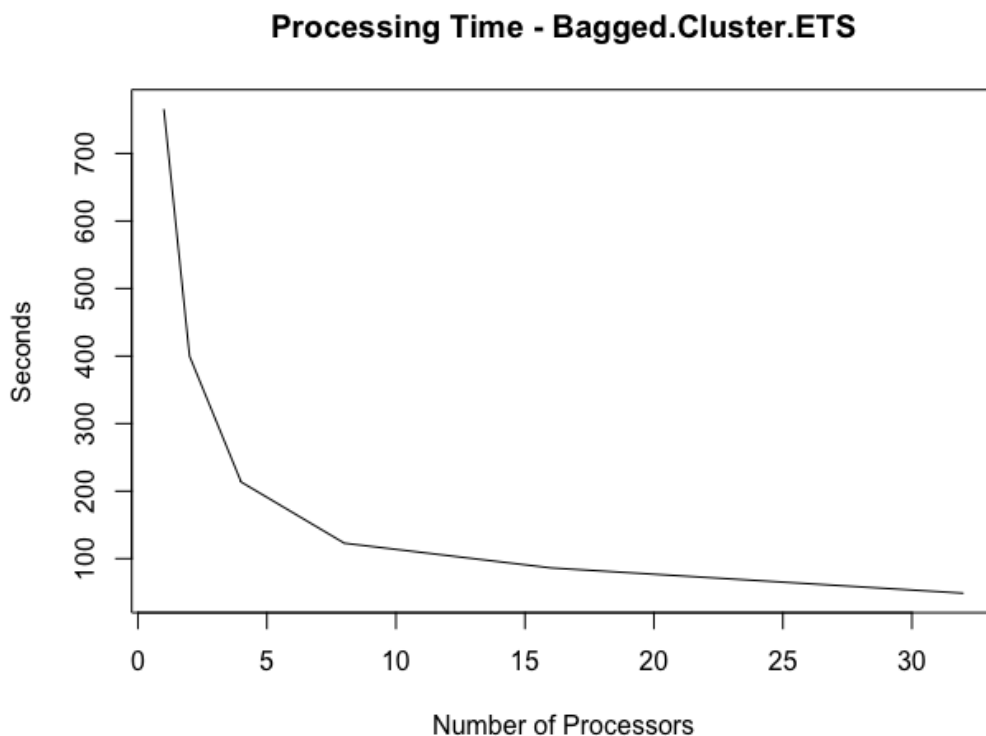


Figure 5.8: Processing time

Finally, the full code that implements `Bagged.Cluster.ETS` is already

available for free on github (<https://github.com/tiagomendesdantas/Bagged.Cluster.ETS>).

5.5

Bias and Variance analysis

The proposed approach, Bagged.Cluster.ETS, has as the main contribution the application of time series clustering for reducing correlated forecasts that affects the overall variance and, consequently, the forecast error. Therefore, it is important to verify whether the method, in fact, is able to reduce variance without compromising bias in comparison to Bagged.BLD.MBB.ETS. To do so, we conduct a Monte Carlo simulation study considering four Data Generation Processes (DGPs) listed in details below:

Autoregressive - AR(6)

We first consider an Autoregressive model of order 6. This model fits the famous Sunspot dataset, see [73], and the parameters obtained from adjusting the AR(6) to the data were the ones considered in the simulation. This follows the same approach used by Taieb and Atiya in [28]. The model can be seen below:

$$y_t = 100 + 1.2401y_{t-1} - 0.419y_{t-2} - 0.1797y_{t-3} + 0.1267y_{t-4} - 0.2259y_{t-5} + 0.1697y_{t-6} + \epsilon_t \quad (5-8)$$

where ϵ_t is independently and identically distributed (i.i.d.), $N(0, 17.28)$

Smooth Transition Autoregressive - STAR

The second model is an Smooth Transition Autoregressive. The model have been subject to many simulation studies for purposes of model selection, evaluation and comparisons. The parameters are the same ones used in [28]. The DGP is shown below:

$$y_t = 500 + 0.3y_{t-1} + 0.6y_{t-2} + (0.1 - 0.9y_{t-1} + 0.8y_{t-2})[1 + e^{-10y_{t-1}}]^{-1} + \epsilon_t \quad (5-9)$$

where ϵ_t is independently and identically distributed (i.i.d.), $N(0, 1)$

Exponential smoothing state space model - ETS (A,N,A)

The third model is an Exponential Smoothing using the state space approach. The parameters were obtained applying the model to the Accidental Deaths in the US from 1973 to 1978 time series. The time series is widely known due to its use in as example in [74] and in the forecast package from R to demonstrate ETS models. The DGP can be seen below:

$$\begin{aligned}
y_t &= l_{t-1} + s_{t-m} + \epsilon_t \\
l_t &= l_{t-1} + 0.5891\epsilon_t \\
s_t &= s_{t-m} + 0.001\epsilon_t
\end{aligned}
\tag{5-10}$$

where ϵ_t is independently and identically distributed (i.i.d.), $N(0, 264.75)$

Seasonal Autoregressive Integrated Moving Average - SARIMA

Since the SARIMA is one of the most used models when it comes to seasonal time series, it was considered in the simulations. The specific model considered, $SARIMA(3, 0, 1)(0, 1, 2)_{12}$, was the best to fit the monthly Corticosteroid Drug Sales in Australia from 1992 to 2008 time series. The steps to obtain the best model (identification and also parameter estimates) are the same reported in [75]. The model is shown bellow:

$$\begin{aligned}
y_t &= y_{t-12} + 0.1603(y_{t-1} - y_{t-13}) - 0.5481(y_{t-2} - y_{t-14}) - 0.5678(y_{t-3} - y_{t-15}) + \\
&\epsilon_t - 0.5222\epsilon_{t-12} - 0.1768\epsilon_{t-24} + 0.3827\epsilon_{t-1} - 0.1998459\epsilon_{t-13} - 0.06766136\epsilon_{t-25}
\end{aligned}
\tag{5-11}$$

where ϵ_t is independently and identically distributed (i.i.d.), $N(0, 0.06)$.

The experiment was conducted considering 1000 time series of length 100 for each DGP. In this sense, first 88 observations from each series were used as training set and the last 12 as test set. The bias and variance were estimated considering the forecasts for the test set. For each time series, the number of clusters for the proposed approach, Bagged.Cluster.ETS, was set to 5, 10, 20, 40, 80. The automatic selection using the Silhouette Information was also considered. For the sake of fair comparisons, in both cases, Bagged.BLD.MBB.ETS and Bagged.Cluster.ETS, the number of time series considered in the aggregation phase were set to 100. In both methods the size of the block was set to 24 to ensure that even if some seasonality was missed in the decomposition phase would still be captured inside the blocks.

In order to compare the results between Bagged.BLD.MBB.ETS and Bagged.Cluster.ETS, ratios of bias and variance among the methods were calculated. This is shown in figure 5.9, where lines in black represent the Bagged.BLD.MBB.ETS, which is considered the benchmark. The other lines were generated using the Bagged.Cluster.ETS varying the number of clusters where lines in blue, green, orange, yellow, red and pink are, respectively, the approach with 5, 10, 20, 40, 80 clusters and the automatic selection. Figure 5.9 shows the results, where values smaller or greater than 1 indicates the proposed approach have, respectively, higher or lower values for bias or variance.

Regarding the squared bias, no consensus between the results were found for all DGPs, since the values oscillated above and below 1, (see figure 5.9). Generally speaking, the proposed approach increased bias for almost all DGPs, the only exceptions are the AR with 5 and 80 clusters and SARIMA. However, as intended, variance was reduced in all cases, see tables 5.1, 5.2, 5.3 and 5.4.

The simulations indicate the proposed approach sometimes increases bias, but in general reduces variance. Moreover, the reduction in variance was higher than the increase in bias in all the cases, meaning the trade-off weights more on the variance side due to its magnitude. This is precisely the expected effect that leads towards better forecasts.

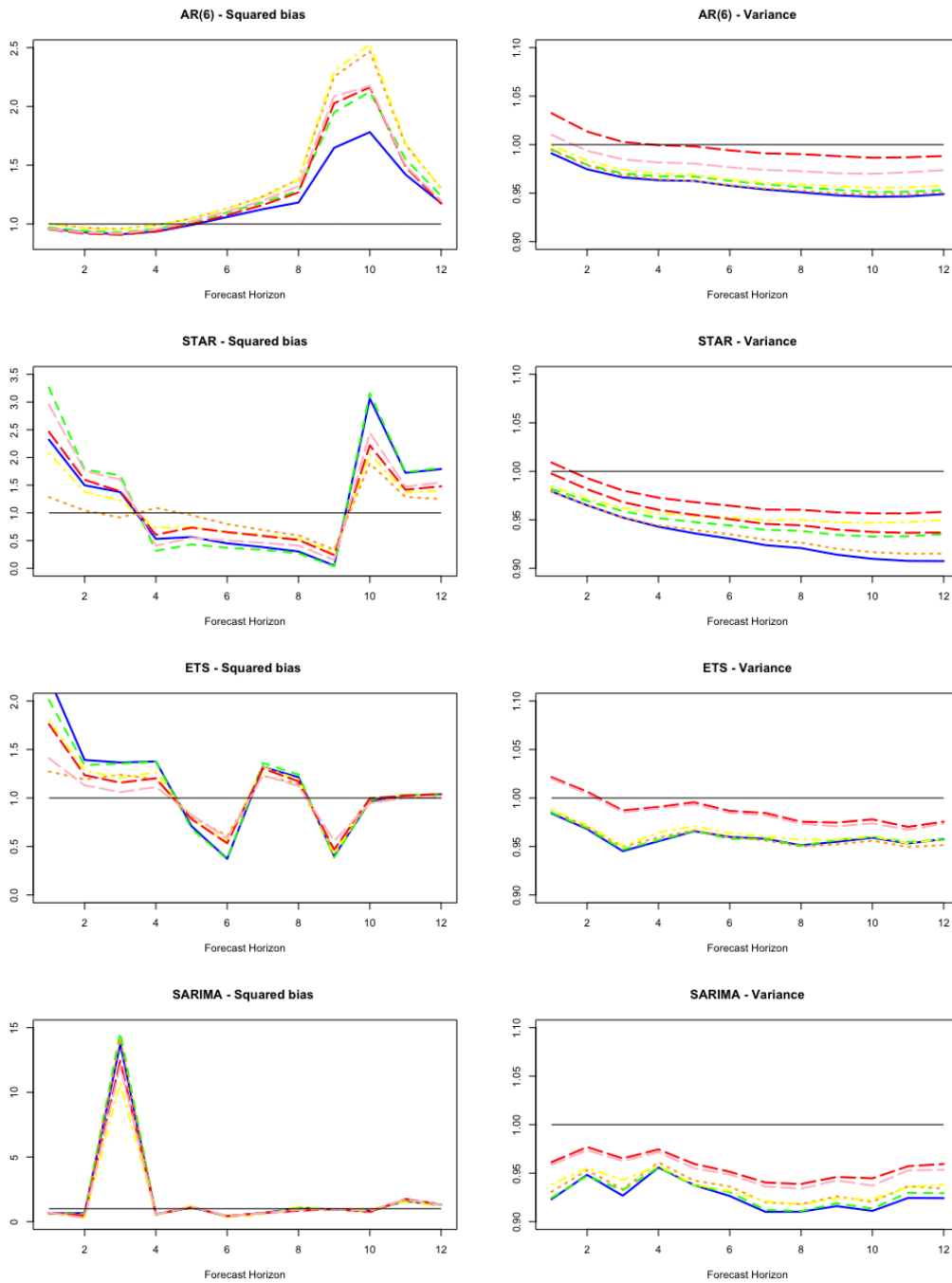


Figure 5.9: Bias and Variance from DGPs

Table 5.1: Squared Bias and Variance - AR

Measures	Bagged.BLD.MBB.ETS	Proposed Approach					Automatic
		k=5	k=10	k=20	k=40	k=80	
Squared Bias	24.24961	23.95768	24.60746	25.50372	25.42617	24.08474	24.48792
Variance	1878.05877	1796.92845	1805.87991	1800.33918	1812.08169	1869.32690	1836.71178

Table 5.2: Squared Bias and Variance - STAR

Measures	Bagged.BLD.MBB.ETS	Proposed Approach					
		k=5	k=10	k=20	k=40	k=80	Automatic
Squared Bias	0.00188	0.00214	0.00215	0.00190	0.00194	0.00201	0.00198
Variance	6.00350	5.55466	5.65972	5.58463	5.72152	5.79445	5.69631

Table 5.3: Squared Bias and Variance - ETS

Measures	Bagged.BLD.MBB.ETS	Proposed Approach					
		k=5	k=10	k=20	k=40	k=80	Automatic
Squared Bias	172.29797	183.12050	183.40112	176.13825	181.46536	179.87989	173.58747
Variance	41154.93728	39434.44026	39456.12883	39401.36690	39572.46175	40531.28877	40429.45484

Table 5.4: Squared Bias and Variance - SARIMA

Measures	Bagged.BLD.MBB.ETS	Proposed Approach					
		k=5	k=10	k=20	k=40	k=80	Automatic
Squared Bias	3.66877e-05	3.49816e-05	3.51077e-05	3.57547e-05	3.4729e-05	3.58039e-05	3.5621e-05
Variance	0.00387	0.00357	0.00358	0.00361	0.00361	0.00369	0.00368

In the simulation the number of clusters that lead to better results seemed to vary from DGP but, in general, the best results were the ones up to 40 clusters. However to specifically address the effects of the number of clusters in forecast accuracy, next section uses data from forecasting competitions and afterwards makes forecasts using the proposed approach comparing the results with the benchmark and Bagged.BLD.MBB.ETS in an ex-ante analysis.

5.6 Experiments on Forecasting Competitions Data

In order to understand the impacts of the selection of the number of clusters and also to verify the performance of the proposed approach in terms of forecast accuracy, data from the M3, CIF 2016 and M4 competitions are used. The M3 competition was a widely known time series forecasting competition organized in 2000 by Spyros Makridakis with more than 3000 series, being 1428 monthly, 756 quarterly and 645 yearly time series from several fields (e.g. macroeconomics, demographics, industry, among others). The dataset from the competition is still the main reference to evaluate and compare time series forecasting methods, see [22]. In [76], Rob Hyndman, Editor-in-Chief of International Journal of Forecasting (IJF), the première journal on forecasting, stated:

"The M3 data have continued to be used since 2000 for testing new time series forecasting methods. In fact, unless a proposed forecasting method is competitive against the original M3 participating methods, it is difficult to get published in the IJF".

Despite all the relevance of M3, by this time it is a bit outdated, therefore we also considered data from two more recent competitions, CIF 2016 and the recently finished but in preliminary results phase, M4. On CIF 2016, the competitors were supposed to forecast 72 monthly time series. Among the time series, 24 were real time series from banking domain and 48 were artificially generated, see [33]. When this study was conducted the competition was already over. This was not the case for the M4 that we actually participated and were supposed to make forecasts for 100,000 time series from many fields and frequencies.

5.6.1 Effects on the number of clusters

To verify the behave of the forecast error according to the number of clusters in Bagged.Cluster.ETS, monthly data from M3 and CIF 2016 was used (data from the test set on the M4 competition was not available to conduct the same analysis on it). To do so 1000 bootstrapped versions using algorithm 1 was generated and only 100 were actually aggregated using algorithm 2. The number of clusters tested varied from 3 to 90. It is important to note that this simulation process is computationally intensive, since it means that for each variation of cluster (1100 ETS models are estimated and used to forecast).

The sMAPE was the metric used to access the forecast performance. In figure 5.10 the behavior of sMAPE using the proposed approach when the number of clusters varies from 3 to 90 (lines in black) is shown. The graph also shows the sMAPE for Bagged.BLD.MBB.ETS (lines in red) and the proposed approach using the Silhouette information (lines in blue). Note that there is a relatively fast decay in forecast error when the number of clusters increase until between 40 and 50 clusters and afterwards the sMAPE starts to increase.

The results are in fact coherent with the design of the approach. The intuition about it comes from the fact that the approach forces the aggregated forecast to be composed by at least one forecast from each cluster. In this sense, when the number of clusters is too small, the majority of the forecasts comes from the same cluster. This means that the selected forecast may be too similar. On the other hand, a large number of clusters also forces the procedure to select forecasts that are not that distant, since each cluster is not very different from each other. In this sense, having an equilibrium between the number of clusters and the number of time series (in this case, the number of clusters between 40% to 50% of the number of series) is preferable and led to smaller forecast errors.

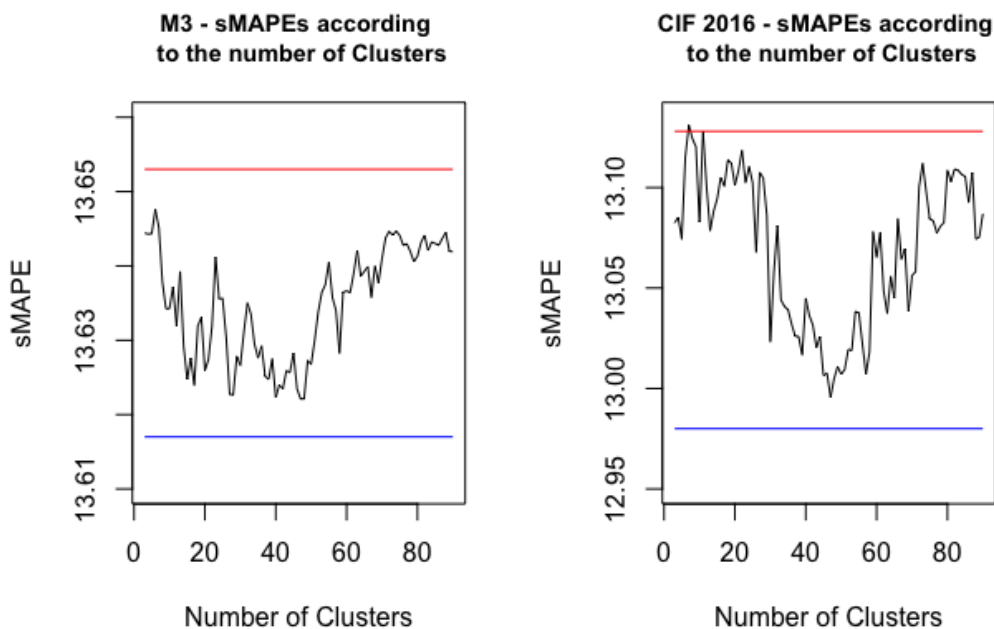


Figure 5.10: sMAPE according to the number of clusters

5.6.2

Performance of the Proposed Approach on Competitions

The proposed approach, Bagged.Cluster.ETS was evaluated using M3 and CIF 2016 data. In order to do so, time series were divided into training and test set (out of sample), where the forecast performance metrics were calculated. The proposed approach was compared with the original M3 and CIF 2016 competition methods (in this case, the performance metrics was calculated using the original submissions by each participant in the competition) and also with Bagged.BLD.MBB.ETS, the method proposed by [13].

The performance metrics considered were the Symmetric Mean Absolute Percentage Error (sMAPE), previously defined in equation 4-1, and the Mean Absolute Scaled Error (MASE). The first metric was used to classify the competing methods in both competitions. The second metric, has desirable statistical properties, as showed by Hyndman in [29]. Also, the metric has better interpretation when compared to sMAPE, since it allows a direct comparison between the considered approach and the naive forecast. See equation 5-12:

$$MASE = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|} \quad (5-12)$$

In order to present the results, six columns were constructed: mean of the sMAPEs ranks (Rank sMAPE) in each series, mean of the sMAPEs (Mean sMAPE), median of sMAPEs (Median sMAPE), mean of the MASEs ranks (Rank MASE) in each series, mean of the MASEs (Mean MASE) and median of MASEs (Median MASE). The inclusion of the median in the analysis is due to the possible asymmetries in sMAPE and MASE distribution of the results. The results were sorted using the first column (Rank sMAPE), this follows the results in [13].

To search for statistically significant differences among the methods, the Friedman rank-sum test with the post-hoc procedure from Hochberg and Rom was employed, see [30] for details on the procedure and implementations. Specifically, the method with lower Rank sMAPE was used as control method and the others were tested for statistical significance.

The entire approach was conducted using R. The results for Bagged.BLD.MBB.ETS were obtained using baggedETS function from the forecast package which improved the results for most of the cases in comparison to the ones presented in [13].

5.6.2.1 M3 - Monthly Results

The results indicate the superiority of the proposed approach, Bagged.Cluster.ETS, in comparison to all other benchmarks for monthly time series, considering all metrics, see table 5.5. The overall Friedman rank-sum test had a p-value of 3.15×10^{-10} , indicating the rejection of null hypothesis of no difference among the results. When considering the proposed approach as control method, the adjusted p-values from the post-hoc procedure indicated statistically significant differences at $\alpha = 5\%$ between the approach and all other methods but Bagged.BLD.MBB.ETS, THETA and Forecast Pro, see table 5.6.

Table 5.5: Methodologies' comparison - M3 Monthly

Methods	Rank sMAPE	Mean sMAPE	Median sMAPE	Rank MASE	Mean MASE	Median MASE
Proposed Approach	11.553	13.617	8.738	11.558	0.835	0.685
Bagged.BLD.MBB.ETS	11.709	13.653	8.848	11.737	0.837	0.691
THETA	11.983	13.892	8.925	11.930	0.858	0.706
ForecastPro	12.000	13.898	8.809	12.023	0.848	0.702
COMB S-H-D	13.028	14.466	9.374	13.095	0.896	0.736
ETS	13.056	14.135	9.073	13.074	0.865	0.716
ForcX	13.260	14.466	9.212	13.314	0.894	0.741
HOLT	13.288	15.795	9.281	13.243	0.909	0.730
WINTER	13.582	15.926	9.305	13.575	1.165	0.735
RBF	13.808	14.76	9.209	13.840	0.910	0.762
DAMPEN	14.006	14.576	9.441	14.088	0.908	0.75
AAM1	14.009	15.670	9.675	13.841	0.905	0.769
AutoBox2	14.151	15.731	9.282	14.209	1.082	0.758
B-J auto	14.220	14.796	9.320	14.239	0.914	0.749
AutoBox1	14.250	15.811	9.268	14.268	0.924	0.748
SMARTFCS	14.374	15.007	9.517	14.283	0.919	0.749
AAM2	14.388	15.938	9.621	14.184	0.923	0.779
Flors-Pearc2	14.414	15.186	9.614	14.474	0.950	0.790
Auto-ANN	14.483	15.031	9.616	14.505	0.928	0.778
PP-Autocast	14.699	15.328	9.897	14.783	0.994	0.759
ARARMA	14.743	15.826	9.800	14.774	0.907	0.777
AutoBox3	14.800	16.590	9.397	14.697	0.962	0.775
Flors-Pearc1	15.126	15.986	9.959	15.159	1.008	0.797
THETA _{sm}	15.177	15.380	9.650	15.176	0.950	0.771
ROBUST-Trend	15.372	18.931	9.733	15.293	1.039	0.830
SINGLE	15.834	15.300	10.028	15.919	0.974	0.810
NAIVE2	16.687	16.891	10.115	16.721	1.037	0.838

Table 5.6: Friedman rank-sum test - M3 Monthly

hypothesis	Adjusted p-value
Proposed Approach	-
Bagged.BLD.MBB.ETS	0.599
THETA	0.147
ForecastPro	0.132
COMB S-H-D	6.791E-7
ETS	4.159E-7
ForcX	8.991E-9
HOLT	5.125E-9
WINTER	8.304E-12
RBF	3.114E-14
DAMPEN	1.449E-16
AAM1	1.325E-16
AutoBox2	2.205E-18
B-J auto	2.744E-19
AutoBox1	1.064E-19
SMARTFCS	2.108E-21
AAM2	1.339E-21
Flors-Pearc2	5.817E-22
Auto-ANN	5.838E-23
PP-Autocast	3.198E-26
ARARMA	6.636E-27
AutoBox3	8.026E-28
THETAsm	3.062E-34
Flors-Pearc1	2.498E-33
ROBUST-Trend	7.661E-38
SINGLE	4.242E-47
NAIVE2	6.212E-67

5.6.2.2

M3 - Quarterly and Yearly Results

The forecasts for quarterly and yearly time series indicate a heavy decline in performance for the proposed method. This results is not unexpected since in the work of Bergmeir and colleagues Bagged.BLD.MBB.ETS had poor results for this frequencies, see [13]. However, the results for the proposed approach performed were worse than Bagged.BLD.MBB.ETS in both frequencies. Considering the quarterly data, the THETA method obtained the best results for the considered metrics but the Median MASE, see table 5.7. The results on yearly data varied a lot depending on the accuracy metric. In this sense, the ForcX method obtained the best results considering Rank sMAPE, Rank MASE and Median MASE. The RBF method obtained the best results considering Mean sMAPE and Flors-Pearc1 was the best according to Median sMAPE and, finally, considering the Mean MASE, the ROBUST-Trend obtained the best results, see table 5.8.

The overall p-value for the Friedman rank-sum test, considering quarterly data was 9.54×10^{-11} and for yearly data meaning that that are statistically significant differences among the methods in both frequencies. Considering

quarterly data and using the THETA method as control, there are statistically significant differences between the method and all other methods. On yearly data and selecting the ForcX method as the control method, the adjusted p-value indicated significant differences between ForcX and all methods but RBF, AutoBox2, Flors-Pearc1, THETA, ForecastPro, Robust-Trend and PP-Autocast see table 5.9.

Table 5.7: Methodologies' comparison - M3 Quarterly

Methods	Rank sMAPE	Mean sMAPE	Median sMAPE	Rank MASE	Mean MASE	Median MASE
THETA	11.817	8.956	5.369	11.821	1.087	0.774
COMB S-H-D	12.620	9.216	5.315	12.614	1.105	0.817
ROBUST-Trend	12.915	9.789	5.000	12.944	1.152	0.823
DAMPEN	13.109	9.361	5.586	13.101	1.126	0.839
PP-Autocast	13.272	9.395	5.256	13.278	1.128	0.825
ForcX	13.349	9.537	5.620	13.339	1.155	0.810
Bagged.BLD.MBB.ETS	13.464	9.803	5.810	13.464	1.163	0.855
B-J auto	13.655	10.260	5.685	13.655	1.188	0.880
ETS	13.717	9.605	5.761	13.687	1.186	0.872
ForecastPro	13.729	9.815	5.837	13.763	1.204	0.853
Proposed Approach	13.742	9.891	5.817	13.684	1.171	0.862
HOLT	13.771	10.938	5.711	13.731	1.225	0.861
RBF	13.796	9.565	5.665	13.757	1.173	0.847
AutoBox2	13.871	10.004	5.595	13.906	1.185	0.85
WINTER	13.895	10.840	5.710	13.871	1.217	0.874
Flors-Pearc1	13.988	9.954	5.612	14.007	1.184	0.844
ARARMA	14.005	10.186	6.108	13.975	1.185	0.860
Auto-ANN	14.416	10.199	6.282	14.444	1.241	0.923
THETAsm	14.705	9.821	5.647	14.683	1.211	0.942
AAM1	14.798	10.165	6.365	14.852	1.24	0.944
SMARTFCS	14.813	10.153	5.708	14.855	1.226	0.858
Flors-Pearc2	14.832	10.431	6.220	14.913	1.255	0.925
AutoBox3	14.931	11.192	6.150	14.882	1.272	0.921
AAM2	14.966	10.260	6.443	15.017	1.256	0.956
SINGLE	15.203	9.717	6.184	15.151	1.229	0.980
AutoBox1	15.257	10.961	6.145	15.278	1.331	0.957
NAIVE2	15.362	9.951	6.184	15.328	1.238	0.985

Table 5.8: Methodologies' comparison - M3 Yearly

Methods	Rank sMAPE	Mean sMAPE	Median sMAPE	Rank MASE	Mean MASE	Median MASE
ForcX	11.596	16.480	11.337	11.567	2.769	1.809
RBF	11.929	16.424	10.738	11.947	2.720	1.902
AutoBox2	11.953	16.593	11.309	11.970	2.754	1.835
Flors-Pearc1	12.044	17.205	10.724	12.054	2.938	1.914
THETA	12.068	16.974	11.252	12.112	2.806	1.971
ForecastPro	12.238	17.271	11.049	12.253	3.026	1.886
ROBUST-Trend	12.302	17.033	11.298	12.347	2.625	1.887
PP-Autocast	12.366	17.128	10.825	12.36	3.016	1.919
Bagged.BLD.MBB.ETS	12.402	17.397	11.200	12.422	2.891	2.000
DAMPEN	12.426	17.36	10.948	12.416	3.032	1.911
COMB S-H-D	12.499	17.072	11.682	12.454	2.876	1.950
ETS	12.535	17.114	11.535	12.576	2.893	2.011
Proposed Approach	12.727	17.560	11.417	12.715	2.931	1.978
SMARTFCS	12.901	17.706	11.834	12.922	2.996	2.095
HOLT	13.160	20.021	11.766	13.174	3.182	2.079
WINTER	13.160	20.021	11.766	13.174	3.182	2.079
Flors-Pearc2	13.556	17.843	12.548	13.584	3.016	2.189
B-J auto	13.572	17.726	11.699	13.578	3.165	1.918
ARARMA	13.595	18.356	11.353	13.688	3.481	1.933
Auto-ANN	13.891	18.565	13.079	13.865	3.058	2.112
AutoBox3	14.091	20.877	12.891	14.078	3.177	2.232
THETAsm	14.116	17.922	12.215	14.036	3.006	2.179
AutoBox1	14.395	21.588	12.747	14.401	3.679	2.256
NAIVE2	14.712	17.880	12.369	14.629	3.172	2.267
SINGLE	14.766	17.817	12.445	14.674	3.171	2.262

Table 5.9: Friedman rank-sum test - M3 Quarterly and Yearly

Quarterly		Yearly	
hypothesis	Adjusted p-value	hypothesis	Adjusted p-value
THETA	-	ForcX	-
COMB S-H-D	0.049	RBF	0.417
ROBUST-Trend	0.007	AutoBox2	0.384
DAMPEN	0.002	Flors-Pearc1	0.274
PP-Autocast	3.674E-4	THETA	0.249
ForcX	1.754E-4	ForecastPro	0.117
Bagged.BLD.MBB.ETS	5.486E-5	ROBUST-Trend	0.085
B-J auto	6.729E-6	PP-Autocast	0.060
ETS	3.276E-6	Bagged.BLD.MBB.ETS	0.049
ForecastPro	2.842E-6	DAMPEN	0.043
Proposed Approach	2.425E-6	COMB S-H-D	0.027
HOLT	1.719E-6	ETS	0.022
RBF	1.252E-6	Proposed Approach	0.006
AutoBox2	4.900E-7	SMARTFCS	0.001
WINTER	3.609E-7	WINTER	1.350E-4
Flors-Pearc1	1.055E-7	HOLT	1.350E-4
ARARMA	8.365E-8	Flors-Pearc2	1.738E-6
Auto-ANN	1.951E-10	B-J auto	1.425E-6
THETA _{sm}	1.515E-12	ARARMA	1.070E-6
AAM1	2.880E-13	Auto-ANN	2.134E-8
SMARTFCS	2.156E-13	AutoBox3	1.151E-9
Flors-Pearc2	1.535E-13	THETA _{sm}	7.782E-10
AutoBox3	2.401E-14	AutoBox1	8.589E-12
AAM2	1.245E-14	NAIVE2	2.916E-14
SINGLE	1.105E-16	SINGLE	1.039E-14
AutoBox1	3.583E-17		
NAIVE2	3.897E-18		

5.6.3 Discussion

In order to better understand the behavior of Bagged.Cluster.ETS in comparison to Bagged.BLD.MBB.ETS, the variance of the group of Bootstrapped forecasts was calculated for both methods for the entire forecasting horizon (18 months for monthly data, 8 quarters for quarterly data and 6 years for yearly data). Considering monthly data, the results indicate that in comparison to Bagged.BLD.MBB.ETS, Bagged.Cluster.ETS was able to reduce variance for the entire forecasting horizon and for the majority of time series. This is precisely the expected behavior of the proposed method by the introduction of the changes that distinguish both methods. However, this result does not hold for the quarterly and yearly cases, this helps to understand why the method didn't produce good forecasting results for these frequencies. See table 5.10 for the percentage of the results where the proposed method produced results with smaller variance in comparison to Bagged.BLD.MBB.ETS.

Table 5.10: Results with reduced variance - Bagged.Cluster.ETS

h	Monthly (%)	Quarterly (%)	Yearly (%)
1	50.98	49.47	45.43
2	53.15	50.4	48.06
3	53.15	49.47	47.6
4	53.99	47.22	47.75
5	53.78	46.96	48.68
6	55.25	48.81	48.53
7	56.93	48.54	–
8	58.26	48.15	–
9	55.39	–	–
10	53.99	–	–
11	53.01	–	–
12	51.54	–	–
13	52.59	–	–
14	52.94	–	–
15	53.99	–	–
16	53.29	–	–
17	53.99	–	–
18	54.97	–	–

One possible explanation on why the divergent results for monthly and quarterly/yearly is the size of the considered time series. While the median length for monthly time series is 115, for quarterly and yearly these numbers drop, respectively, to 44 and 19. Having a small number of data points may hurt the entire procedure.

Intuitively, Bagged.BLD.MBB.ETS is affected by the short length of the time series. Specifically, the MBB phase of the procedure would have less blocks to sample, leading to more similar time series and, consequently, similar forecasts. This would increase the covariance effect. The case for the proposed approach is even worse since besides the MBB problem, the clustering phase would also hurt the procedure. This is due to design of the method itself that forces at least one time series from each created cluster to be picked and when the time series generated by MBB are too similar, the PAM would create clusters with unsatisfactory dissimilarity.

5.6.4 CIF 2016

The Bagged.BLD.MBB.ETS method has its merits on monthly data by outperforming all other methods from the original M3 competition. The proposed approach improved Bagged.BLD.MBB.ETS at monthly frequency by reducing the forecasting error even more. However, the results for Bagged.BLD.MBB.ETS are substantially worse for quarterly and yearly data. The performance for the proposed method adds an additional clustering step that can be a potential source of error, therefore the results are even worse

than Bagged.BLD.MBB.ETS. This is an indication that having enough data points is vital to the process.

The data from the M3 competition is the benchmark dataset but a bit dated by now. In this sense, a dataset from the recent CIF 2016 competition is used to validate the proposed method. The data consists of 72 time series divided between real and artificial ones, where the last accounts for 2/3 of the series. The data is exclusively monthly, which taking into account the results from the M3 competition makes it a perfect match to find more evidence in favour of the quality of the approach.

Considering all metrics, but median sMAPE and median MASE, the proposed approach generated the best forecasts for the entire 48 artificial time series, see table 5.11. The results on table 5.12, indicate that considering the 24 real time series, the Ensemble of LSTMs and ETS was the winner for all metrics. It is yet important to note that considering all metrics but median sMAPE and median MASE, Bagged.Cluster.ETS generated better results than its direct contestant Bagged.BLD.MBB.ETS.

The table 5.13 shows the results considering all time series in the competition. The results indicated the Ensemble of LSTMs and ETS as the top performer in the competition. The proposed approach obtained the second best results according to all metrics, but Median sMAPE and Median MASE, confirming its superiority against Bagged.BLD.MBB.ETS.

The results from the Friedman rank-sum test indicated an overall p-value of 1.62×10^{-10} pointing to the rejection of the null hypothesis of no difference between the results at $\alpha = 5\%$. Having the Ensemble of LSTMs and ETS as the control method no statistically significant differences at $\alpha = 5\%$ were found between the control and the following methods: Bagged.Cluster.ETS, Bagged.BLD.MBB.ETS, LSTM deseasonalized, ETS, MLP and FRBE.

It is interesting to check that the top 3 performers (Ensemble of LSTMs and ETS, Bagged.Cluster.ETS and the Bagged.BLD.MBB.ETS) were the ones that did some form of combination of forecasts. These results are related to equation 5-4 and corroborate with section 5.2.

Table 5.11: CIF 2016 - Methodologies' comparison - Artificial Series

Methods	Rank sMAPE	Mean sMAPE	Median sMAPE	Rank MASE	Mean MASE	Median MASE
Proposed Approach	6.896	6.308	5.078	6.792	0.964	0.559
Bagged.BLD.MBB.ETS	7.146	6.319	5.000	6.938	0.968	0.552
Ensemble of LSTMs and ETS	8.604	6.706	5.373	8.604	0.979	0.640
ETS	8.771	6.615	5.340	8.646	1.003	0.573
FRBE	9.208	7.024	5.375	9.312	1.064	0.630
LSTM deseasonalized	9.312	6.710	5.235	9.250	0.986	0.662
Boot.EXPOS	9.729	6.904	5.496	9.688	1.054	0.682
MLP	9.979	6.761	5.368	10.104	1.021	0.657
ARIMA	10.958	7.349	5.492	10.875	1.088	0.677
HEM	10.979	7.322	5.129	11.062	1.085	0.668
REST	11.688	7.342	6.259	11.792	1.076	0.717
PB-GRNN	11.917	7.754	5.591	11.917	1.140	0.711
PB-RF	11.917	7.754	5.591	11.917	1.140	0.711
PB-MLP	12.229	7.718	5.649	12.333	1.133	0.692
AVG	13.042	8.208	6.642	13.000	1.236	0.85
LSTM	13.479	7.795	6.619	13.458	1.124	0.835
MTSFA	14.292	9.464	6.414	14.312	1.333	0.775
Fuzzy c-regression m	14.625	9.588	7.274	14.667	1.430	1.219
FCDNN	15.646	8.587	7.475	15.625	1.259	0.833
Random Walk	17.062	10.69	8.855	17.125	1.621	1.195
THETA	17.750	10.834	8.790	17.917	1.604	1.396
TSFIS	17.833	10.697	9.489	17.792	1.625	1.279
HFM	19.021	14.564	9.607	18.979	3.675	1.293
MSAKAF	19.229	14.634	12.840	19.208	2.003	1.568
CORN	23.688	19.327	18.867	23.688	2.758	2.349

Table 5.12: CIF 2016 - Methodologies' comparison - Real Series

Methods	Rank sMAPE	Mean sMAPE	Median sMAPE	Rank MASE	Mean MASE	Median MASE
Ensemble of LSTMs and ETS	8.292	19.090	14.649	8.167	0.424	0.321
LSTM deseasonalized	9.125	18.178	15.735	8.958	0.494	0.327
MLP	10.583	22.882	23.514	10.583	0.491	0.343
Fuzzy c-regression m	10.750	22.010	20.041	10.750	0.521	0.353
REST	10.917	22.654	19.696	11.208	0.541	0.409
TSFIS	11.062	23.928	20.880	11.021	0.555	0.449
AVG	11.458	22.746	19.343	11.542	0.520	0.377
Random Walk	11.792	22.379	17.451	11.917	0.526	0.419
ETS	11.833	22.409	18.732	11.708	0.513	0.391
Proposed Approach	12.000	26.325	20.141	11.938	0.550	0.376
HEM	12.042	24.466	20.777	12.083	0.532	0.371
FRBE	12.250	24.667	18.616	12.167	0.531	0.369
LSTM	12.958	24.414	16.683	12.708	0.593	0.344
THETA	13.125	22.599	20.776	13.208	0.546	0.320
Bagged.BLD.MBB.ETS	13.250	26.748	20.117	13.146	0.552	0.388
MSAKAF	13.625	31.891	22.721	13.708	0.709	0.569
ARIMA	13.771	28.988	21.679	13.729	0.584	0.441
PB-GRNN	13.875	27.992	24.830	14.042	0.744	0.527
PB-RF	13.875	27.992	24.830	14.042	0.744	0.527
MTSFA	14.167	30.608	27.100	13.958	0.715	0.472
PB-MLP	14.458	29.392	25.906	14.042	0.711	0.469
Boot.EXPOS	15.125	31.942	21.633	15.042	0.674	0.458
HFM	16.500	38.055	24.736	16.917	2.463	0.543
FCDNN	17.125	32.698	26.650	17.417	0.908	0.620
CORN	21.042	47.624	34.024	21.000	1.207	1.100

Table 5.13: CIF 2016 - Methodologies' comparison - All Series

Methods	Rank sMAPE	Mean sMAPE	Median sMAPE	Rank MASE	Mean MASE	Median MASE
Ensemble of LSTMs an ETS	8.500	10.834	6.598	8.458	0.794	0.559
Proposed Approach	8.597	12.980	6.048	8.507	0.826	0.545
Bagged.BLD.MBB.ETS	9.181	13.128	5.978	9.007	0.829	0.537
LSTM deseasonalized	9.250	10.532	7.017	9.153	0.822	0.597
ETS	9.792	11.880	6.666	9.667	0.840	0.532
MLP	10.181	12.135	6.923	10.264	0.845	0.545
FRBE	10.222	12.905	6.769	10.264	0.886	0.566
HEM	11.333	13.037	7.317	11.403	0.900	0.590
REST	11.431	12.446	7.574	11.597	0.898	0.591
Boot.EXPOS	11.528	15.250	6.923	11.472	0.928	0.614
ARIMA	11.896	14.562	7.027	11.826	0.920	0.562
AVG	12.514	13.054	8.020	12.514	0.997	0.676
PB-GRNN	12.569	14.500	7.856	12.625	1.008	0.653
PB-RF	12.569	14.500	7.856	12.625	1.008	0.653
PB-MLP	12.972	14.943	8.052	12.903	0.992	0.681
LSTM	13.306	13.334	8.202	13.208	0.947	0.684
Fuzzy c-regression m	13.333	13.729	10.036	13.361	1.127	0.722
MTSFA	14.250	16.512	9.692	14.194	1.127	0.707
Random Walk	15.306	14.586	9.141	15.389	1.256	0.833
TSFIS	15.576	15.107	10.183	15.535	1.269	0.911
FCDNN	16.139	16.624	8.713	16.222	1.142	0.818
THETA	16.208	14.756	11.012	16.347	1.251	0.744
MSAKAF	17.361	20.386	14.239	17.375	1.572	1.314
HFM	18.181	22.394	11.890	18.292	3.271	1.144
CORN	22.806	28.760	19.858	22.792	2.241	1.826

Table 5.14: Friedman rank-sum test - CIF 2016

hypothesis	Adjusted p-value
Ensemble of LSTMs and ETS	-
Proposed Approach	0.937
Bagged.BLD.MBB.ETS	0.579
LSTM deseasonalized	0.541
ETS	0.292
MLP	0.171
FRBE	0.160
HEM	0.021
REST	0.017
Boot.EXPOS	0.014
ARIMA	0.006
AVG	0.001
PB-RF	9.080E-4
PB-GRNN	9.080E-4
PB-MLP	2.664E-4
LSTM	8.941E-5
Fuzzy c-regression m	8.137E-5
MTSFA	2.764E-6
Random Walk	2.887E-8
TSFIS	7.977E-9
FCDNN	4.739E-10
THETA	3.297E-10
MSAKAF	5.051E-13
HFM	2.974E-15
CORN	1.982E-31

5.6.5 Participation in the M4 Competition

The M4 competition was announced in November 2017 and officially started in the first day of January 2018. The deadline was 31 May 2018 with final results expected to 28th of September 2018. The competition was organized by Spyros Makridakis and is a continuation to the previous M1, M2 and M3 competitions. On the official M4 competition website, see [81] it says that:

"M-Competitions more generally seek to identify the most accurate forecasting method(s) for different types of predictions. To get precise and compelling answers, the M4 is utilizing 100,000 real-life series, and incorporates all major forecasting methods, including those based on Artificial Intelligence (AI), as well as the more traditional statistical ones".

On total, more than 248 teams and individuals from 48 countries participated in the competition. However, probably by the influence of the large number of series, only 50 individuals and teams completed the submission process and produced forecasts for the 100,000 time series from many fields (Demographic, finance, industry, macro, micro and other) and frequencies (Yearly, quarterly, monthly, weekly, daily and hourly), see table 5.15 and [82] for the paper on the preliminary results.

Table 5.15: M4 dataset

Frequency	Micro	Industry	Macro	Finance	Demographic	Other	Total
Yearly	6538	3716	3903	6519	1088	1236	23000
Quarterly	6020	4637	5315	5305	1858	865	24000
Monthly	10975	10016	10016	10987	5728	277	48000
Weekly	112	6	41	164	24	12	359
Daily	1476	422	127	1559	10	633	4227
Hourly	-	-	-	-	-	414	414
Total	5121	18798	19402	24534	8708	3437	100000

The team from PUC-Rio was composed by Tiago Mendes Dantas and Fernando Luiz Cyrino Oliveira and participated with Bagged.Cluster.ETS. In order to make the participation viable in terms of time and costs, Bagged.Cluster.ETS ran on the AWS cloud using a c4.8xlarge machine (132 ECUs, 36 vCPUs, 2.9 GHz, Intel Xeon E5-2666v3, 60 GiB memory).

The preliminary results, released on June 18 2018, evinced the level of the competition by the presence of not only high profile researchers, such as Sir David Hendry and Rob Hyndman, but also companies such as UBER and Wells Fargo. On total, there were more than 30 submissions from academia/universities and more than 10 from companies-organizations.

The evaluation of the methods was done using a combination of sMAPE and MASE called Overall Weighted Average (OWA), see equation 5-13, and the results showed that Bagged.Cluster.ETS was one among the only 17 out of 50 methods (on total there were 60 contestants: 50 methods from team and individuals and 10 benchmarks) that outperformed the benchmarks. The approach ranked 17th considering the OWA, 23th considering sMAPE and 15th considering MASE. Note the method, as expected, outperformed all the single Exponential Smoothing benchmarks. See table 5.16 for the full the list of submitted methods.

$$OWA = \frac{MASE}{NAIVE2} + \frac{sMAPE}{NAIVE2} \quad (5-13)$$

Table 5.16: M4 results

Team Members	Affiliation	Type of Method	sMAPE	MASE	OWA	General Rank (OWA)	General Rank (sMAPE)	General Rank (MASE)
Savil, S.	Uber Technologies	Hybrid	11.374	1.538	0.821	1	1	1
Montero-Manso, P., Talaga, T., Hyndman, R. J. & Athanasopoulos, G.	University of A Coruña & Monash University	Combination	11.720	1.551	0.838	2	3	3
Pawlikowski, M., Chorowska, A. & Yanchuk, O.	ProLogistica Soft	Combination	11.845	1.547	0.841	3	5	2
Jaganathan, S. & Prakash, P.	Individual	Combination	11.095	1.571	0.842	4	2	6
Fiorucci, J. A. & Louzada, F.	University of Brasília & University of São Paulo	Combination	11.836	1.554	0.843	5	4	4
Petroponis, F. & Svetunkov, I.	University of Bath & Lancaster University	Combination	11.887	1.565	0.848	6	6	5
Skaub, D.	Harvard Extension School	Combination	12.020	1.595	0.86	7	9	7
Legaki N. Z. & Koutsoyi K.	National Technical University of Athens	Statistical	11.986	1.601	0.861	8	8	8
Dvorak, J., Hendry, D. & Cserlo, J.	University of Oxford	Combination	11.924	1.627	0.865	9	7	10
Pedregal, D.J., Trapero, J. R., Villegas, M. A. & Madruga, J. J.	University of Castilla-La Mancha	Combination	12.114	1.614	0.869	10	10	9
Spliotis, E. & Assimakopoulos, V.	National Technical University of Athens	Statistical	12.148	1.628	0.874	11	12	11
Roubinichien, A.	Washington State Employment Security Department	Combination	12.183	1.631	0.876	12	13	13
Boulinas, M.	Georgia Institute of Technology	Statistical	12.198	1.634	0.88	13	14	14
Tarfa M seminar	University of Terzu	Combination	12.496	1.633	0.888	14	19	12
Wahneh, W.	Universiti Tun Hussein Onn Malaysia	Combination	12.146	1.706	0.894	15	11	22
Darin, S. & Stellwagen, E.	Business Forecast Systems (Forecast Pro)	Statistical	12.279	1.693	0.895	16	15	20
Dantas, T. & Cyrino Oliveira, F.	Pontifical Catholic University of Rio de Janeiro	Combination	12.553	1.657	0.896	17	21	15
THEIA	Benchmark	Statistical	12.369	1.696	0.897	18	16	21
Combination of Exponential Smoothing Models	Benchmark	Combination	12.555	1.663	0.898	19	23	16
Nkazal, A.	Individual	Combination	12.370	1.724	0.907	20	17	25
Dampas Exponential Smoothing	Benchmark	Statistical	12.661	1.683	0.907	21	25	18
Segura-Heras, J. V., Vercher-González, E., Bernádez-Eido, J. D., & Corberán-Vallat, A.	Universidad Miguel Hernández & Universitat de València	Combination	12.507	1.717	0.91	22	20	24
Trosta, B.	Individual	Machine Learning	12.894	1.692	0.915	23	27	17
Chen, W. & Francis, J.	Fortham University	Combination	12.554	1.739	0.915	24	22	26
Svetunkov, I., Abolghasemi, M. & Kourratzou, N.	Lancaster University & University of Newcastle	Combination	12.464	1.745	0.916	25	18	28
Talaga, T., Hyndman, R. J. & Athanasopoulos, G.	Monash University	Statistical	12.902	1.687	0.917	26	28	19
Su, M. & Rengifo, E.	Fortham University	Combination	12.855	1.743	0.93	27	26	27
Kharagiani, S.	Individual	Combination	13.065	1.716	0.93	28	29	23
Smart Forecast	Smart Cube (Smart Forecast)	Combination	13.214	1.788	0.955	29	31	30
Wainwright, E., Butz, E. & Raychaudhuri, S.	Oracle Corporation (Crystal Ball)	Statistical	13.336	1.798	0.962	30	32	32
Holt's Exponential Smoothing	Benchmark	Statistical	13.775	1.772	0.971	31	40	29
Simple Exponential Smoothing	Benchmark	Statistical	13.087	1.885	0.975	32	30	34
Vello de Santos, R., Araújo, C. & Accioly, R.	Individual	Combination	13.820	1.780	0.977	33	41	31
Reilly, T.	Automatic Forecasting Systems, Inc. (AutoBox)	Statistical	13.756	1.873	0.997	34	39	33
Naive 1	Benchmark	Statistical	13.564	1.912	1.00	35	34	36
Iqbal, A., Seery, S. & Sílvia, J.	Wells Fargo Securities	Statistical	14.312	1.892	1.022	36	44	35
Fritsch, R.	Individual	Statistical	13.530	2.069	1.040	37	33	44
Baeremga, G.	Université Libre de Bruxelles	Combination	13.990	2.023	1.045	38	42	39
Naive 2	Benchmark	Statistical	14.208	2.044	1.058	39	43	40
Bandara, K., Bergmeir, C. & Hewamalage, C.	Monash University	Combination	12.653	2.334	1.077	40	24	47
Seasonal Naive	Benchmark	Statistical	14.657	2.057	1.078	41	47	42
Patel, A.	Individual	Combination	14.430	2.098	1.081	42	45	45
Kyriakidis, I. & Artzi, A.	University of Nicosia	Combination	14.880	2.068	1.090	43	48	43
Viole, F. & Vinod, H.	Fortham University	Combination	15.392	2.015	1.094	44	49	38
Chirikhin, K. & Ryabko, B.	Novosibirsk State University	Combination	16.468	1.957	1.119	45	52	37
Alves Santos Junior, J. G.	Individual	Machine Learning	16.638	2.056	1.151	46	53	41
Mohamed, A.	SerjocKow	Combination	15.901	2.316	1.190	47	51	46
Clark, C.	Individual	Statistical	15.881	2.583	1.261	48	50	48
Schmaler, H. T.	Dokuz Eylul University	Statistical	13.587	3.305	1.365	49	37	52
Taylan, A. S.	Yapı Kredi Invest	Statistical	13.577	3.322	1.369	50	36	53
Yapar, G., Çapar, S. & Yavuz, I.	Dokuz Eylul University	Statistical	13.573	3.371	1.382	51	35	55
Yilmaz, T. E.	Dokuz Eylul University	Statistical	13.027	3.676	1.464	52	38	57
Çetin, B.	Dokuz Eylul University	Statistical	14.480	3.576	1.469	53	46	56
Mukhopadhyay, S.	University of Texas	Machine Learning	18.469	3.059	1.481	54	54	50
Recurrent neural network (RNN)	Benchmark	Machine Learning	21.152	2.685	1.482	55	57	49
Pitka, P.	Crestochoma University of Technology	Machine Learning	19.573	3.341	1.595	56	56	54
Multi-layer Perceptron (MLP)	Benchmark	Machine Learning	21.853	3.225	1.642	57	58	51
Dudek, G.	Crestochoma University of Technology	Other	26.137	6.119	2.561	58	59	58
Sinotin, R.	Siberian State University of Telecommunications and Information Sciences	Statistical	19.136	14.081	4.387	59	55	59

Unfortunately, on the preliminary results there is no full description of the methods in the competition. These are expected in the final version that will be available on September 28 2018. However, the competition organizers detailed the top 6 in the end of June at the International Symposium on Forecasting held in Boulder, Colorado - USA. According to the authors the methods from top 6 were:

- **Smyl,S (UBER)**: Hybrid model of Exponential Smoothing with LSTM and Hierarchical modeling where the parameters were estimated using data from the entire dataset and individual series. In the end a combination was also considered.
- **Montero-Manso, P., Talagala,T., Hyndman, R.J. & Athanassopoulos, G. (University of la Coruña and Monash University)**: Combination of ARIMA, ETS, tbats, THETA, Naïve, Seasonal Naïve, Neural Network and LSTM where the aggregation was made by a weighted average. The weights were calculated using the extreme gradient boosting algorithm (Xgboost) considering holdout sets.
- **Pawlikoski,M, Chorowska,A. & Yanchuk,O. (Prologistica Soft)** - Combination of several statistical methods via weighted average. The pool of methods was defined manually based on the characteristics of the time series.
- **Jaganathan, S. & Prakash, P. (individual)** - Combination of statistical methods using the approach defined in [83].
- **Fiorucci, J.A. & Louzada,F. (University of Brasilia and University of São Paulo)** - Combination of ARIMA, ETS and THETA via weighted average where the weights were estimated using cross-validation.
- **Petropoulos, F & Sventunkov, I. (University of Bath and Lancaster University)** - Combination of ETS, CES, ARIMA and THETA via the median of the forecasts.

An analysis on the type of methods that outperformed the benchmarks evince that combination was the right direction to follow in this competition. Moreover, a closer look at the top 6 approaches indicated also that the weighting scheme do matters. In this sense, while Bagged.Cluster.ETS is indeed a combination approach, there is no weighting in the proposed methodology. This is certainly a point of improvement that might have played an important role in the final result. Findings like these and probably many others from other researchers are precisely what makes competitions useful.

5.7

Summary and Concluding Remark

This chapter introduced an innovative way of improving forecasting performance of Bagging Exponential Smoothing methods by combining Bagging, Exponential Smoothing and Clusters methods. The proposed approach, Bagged.Cluster.ETS, takes into account the covariance effect, an important aspect left unattended by previous authors, and seeks to minimize it by generating clusters and, subsequently, selecting series from them. By doing so, the proposed approach is expected to reduce the forecast error.

The proposed approach was submitted to a series of validations. In this sense, a simulation exercise using Monte Carlo and 4 DGPs was conducted to understand the behavior of the Bagged.Cluster.ETS's bias and variance in comparison to Bagged.BLD.MBB.ETS and also the impact of the number of clusters in bias and variance. The overall result was that, as intended, Bagged.Cluster.ETS was able to reduce variance without sacrificing bias a lot. The number of clusters also played an important role in the magnitude of this reduction.

The main objective of a forecasting method is to be as accurate as possible. Therefore, data from M3 competition and CIF 2016 was used in two ways. First to understand the impact of the number of cluster in forecasting performance and second to conduct an ex-ante forecasting analysis of the proposed approach with the original methods from the M3 competition, plus Bagged.BLD.MBB.ETS and ETS and the methods in the CIF 2016 competition (this already includes Bagged.BLD.MBB.ETS).

The results on the number of clusters indicated that the ideal number was between 40 and 50 clusters, considering 100 time series in the ensemble. Also, the results showed that the automatic selection of clusters using the Silhouette information performed very well.

The ex-ante analysis indicated that Bagged.Cluster.ETS is a top competitor, since for the 1428 monthly time series from the M3 competition and 72 from CIF 2016 the approach consistently outperformed Bagged.BLD.MBB.ETS and all other benchmarks in the first competition and all but one, an Ensemble of LSTM and ETS, in the second competition.

A controversial point in the result, but not uncommon in the forecasting time series area, was the fact that the methods did not show statistically significant differences between the proposed approach and some of the methods. Nevertheless, in the case of forecasting the results are still valid and important findings as [31] and [32] clarify.

The preliminary results of the fresh M4 competition showed the approach

was able to outperform the competition benchmarks and although there were better ranked methods. Bagged.Cluster.ETS produced good results and is aligned with the top 6 performers as it also uses combination in order to generate the final forecasts. The competition also demonstrated that an introduction of a weighting scheme is a possible direction for future steps on the Bagged.Cluster.ETS.

It is important to note that for CIF 2016, M3 and M4, the Silhouette information was the method used to select the ideal number of clusters for each series. However, if one has the necessary computational power, the cross-validation approach is indeed a valid option that should be considered in determining the optimal number of clusters.

6

Concluding Remarks e Future Directions

Acknowledging that precise forecasts are a fundamental part of planning processes and the basis for every efficient system, this thesis finds its motivation on improving time series forecasting performance by proposing the use of the state of the art method, Bagged.BLD.MBB.ETS, on a very real problem and going further by identifying points of improvement in the state of the art and finalizing proposing a new approach to improving existing methods.

The minor contribution of the thesis lies on explaining the state of the art, Bagged.BLD.MBB.ETS, in detail and applying it to the problem of forecasting demand for air transport. On total, data from 14 countries was used and it was shown that Bagging was able to reduce the forecast error in almost all the cases when compared to single versions of the approach and the other benchmarks. At this point, Bagged.BLD.MBB.ETS has never been used in this context, making it a valid direction to practitioners in the industry.

It is important to understand that precise forecasts are of fundamental importance in the air transport industry where even small errors can lead to disastrous effects such as system congestion, excess infrastructure capacities, increased operators' costs. Thinking about companies, having more accurate forecast than its rivals can translate into competitive advantages since the data inputs on planning are more reliable.

The major contribution of the thesis rests on understanding why Bagging actually improve performance and how it works on Bagged.BLD.MBB.ETS, leading to the proposition of a new approach called Bagged.Cluster.ETS. The method uses the methodology proposed by Bergmeir and colleagues in [13] as a starting point, but goes further by addressing a key point of improvement that was the covariance among the ensemble of forecasts that are aggregated, impacting the variance and, consequently the forecast error.

To understand how the variance and bias behave in both Bagged.BLD.MBB.ETS and Bagged.Cluster.ETS a simulation study is conducted using the Monte Carlo simulation method. In order to access how the variance and bias behave in both methods in the presence of diverse characteristics (e.g. non-linearity and seasonality) 4 DGPs are used. It is shown that, in comparison to Bagged.BLD.MBB.ETS, Bagged.Cluster.ETS,

in general was able to reduce variance without hurting bias a lot, therefore, making a good balance in the bias variance trade-off.

The quality of the forecasts generated by Bagged.Cluster.ETS was accessed using data from time series forecasting competition. In this sense data from M3, CIF 2016 and the yet not fully finished M4 were used. The first 2 were initially used to verify how the number of clusters affected the quality of the forecasts, where the results indicated a number between 40 and 50 in the case, considering an aggregation of 100 time series.

Afterwards, the proposed approach was used to generate forecasts in an ex-ante fashion for M3 and CIF 2016. The results were compared to benchmarks in the original competitions (plus Bagged.BLD.MBB.ETS that was not available in the case of the M3). Bagged.Cluster.ETS was able to outperform Bagged.BLD.MBB.ETS in both competitions (considering monthly data), representing an advance in the state of the art and validated using the main reference. Furthermore, considering monthly data, the proposed approach outperformed all other methods in the M3 and ranked second in CIF 2016.

A drawback of the method, evinced in the quarterly and yearly data from M3 competition, was its inability to generate good forecasts in the presence of time series with small length. As it happened to Bagged.BLD.MBB.ETS, Bagged.Cluster.ETS experienced a deterioration in its forecasts, suffering even more than the first, probably, due to the extra phase (clustering) in the proposed methodology.

The evaluation of the M4 competition showed the approach was one among the 17 to actually beat the benchmarks. Although it did not rank in the top 3 methods, as we would like, the performance of Bagged.Cluster.ETS was aligned to what is the future in time series forecasting, since, the method is a combination of forecasts like the top methods.

It is important to say that the high number of time series to forecasts, 100,000, and the length of the competition, approximately 4 months, posed as a great difficulty. Since Bagged.Cluster.ETS is a computationally intensive method, a lot of validation and evaluation of the performance of the method that would lead to improvements, and that are common among participants in this type of competitions, were not implemented. This means that the results can probably be improved.

The competition itself, although in their preliminary results, is already providing guidance to further research, as immediately after the results it became clear that a weighting scheme can lead Bagged.Cluster.ETS to even better forecasts, something that before the competition was not crystal clear, see [17]. Therefore, the weighting scheme constitutes as a clear extension of

this work. As possible extensions of this work, the following stand out:

- The study of weighting schemes;
- The investigation of other alternative decomposition methods than STL;
- A detailed investigation on the gains of including other distances and clustering algorithms;
- The analytical calculations of the ensemble's variance;
- The inclusion of probabilistic forecasts;
- The investigation of computational cost/complexity reduction techniques;
- Other forms of variance reduction less computationally intensive.

Bibliography

- [1] HYNDMAN, R. J.; ATHANASOPOULOS, G.. **Forecasting: principles and practice**. OTexts, Melbourne, Australia, 2014.
- [2] BREIMAN, L.. **Bagging predictors**. *Machine learning*, 24(2):123–140, 1996.
- [3] EFRON, B.. **Bootstrap methods: Another look at the jackknife**. *Ann. Statist.*, 7(1):1–26, 01 1979.
- [4] INOUE, A.; KILIAN, L.. **Bagging time series models**. Discussion Paper, Centre for Economic Policy Research, London, 2004.
- [5] LEE, T.-H.; YANG, Y.. **Bagging binary and quantile predictors for time series**. *Journal of econometrics*, 135(1):465–497, 2006.
- [6] INOUE, A.; KILIAN, L.. **How useful is bagging in forecasting economic time series? a case study of us consumer price inflation**. *Journal of the American Statistical Association*, 103(482):511–522, 2008.
- [7] CORDEIRO, C.; NEVES, M. M.. **Forecasting time series with boot. expos procedure**. *Revstat*, 7(2):135–149, 2009.
- [8] HILLEBRAND, E.; MEDEIROS, M. C.. **The benefits of bagging for forecast models of realized volatility**. *Econometric Reviews*, 29(5-6):571–593, 2010.
- [9] RAPACH, D. E.; STRAUSS, J. K.. **Bagging or combining (or both)? an analysis based on forecasting us employment growth**. *Econometric Reviews*, 29(5-6):511–533, 2010.
- [10] WANG, Y.; XIAO, M. ; ZHOU, Y.. **A hybrid ensemble approximation method for chaotic time series forecast**. *Journal of Information & Computational Science*, 9(18):5849–5856, 2012.
- [11] ZONTUL, M.; AYDIN, F.; DOAN, G.; SENER, S. ; KAYNAR, O.. **Wind speed forecasting using reptree and bagging methods in kirkclareli-turkey**. *Journal of Theoretical and Applied Information Technology*, 56:17–29, 10 2013.

- [12] JIN, S.; SU, L. ; ULLAH, A.. **Robustify financial time series forecasting with bagging**. *Econometric Reviews*, 33(5-6):575–605, 2014.
- [13] BERGMEIR, C.; HYNDMAN, R. J. ; BENITEZ, J. M.. **Bagging exponential smoothing methods using stl decomposition and box–cox transformation**. *International Journal of Forecasting*, 32(2):303–312, 2016.
- [14] DANTAS, T. M.; OLIVEIRA, F. L. C. ; REPOLHO, H. M. V.. **Air transportation demand forecast through bagging holt winters methods**. *Journal of Air Transport Management*, 59:116–123, 2017.
- [17] TIMMERMANN, A.. **Forecast combinations**. *Handbook of economic forecasting*, 1:135–196, 2006.
- [19] BOX, G. E.; COX, D. R.. **An analysis of transformations**. *Journal of the Royal Statistical Society. Series B (Methodological)*, p. 211–252, 1964.
- [20] CLEVELAND, R. B.; CLEVELAND, W. S. ; TERPENNING, I.. **Stl: A seasonal-trend decomposition procedure based on loess**. *Journal of Official Statistics*, 6(1):3, 1990.
- [22] MAKRIDAKIS, S.; HIBON, M.. **The m3-competition: results, conclusions and implications**. *International journal of forecasting*, 16(4):451–476, 2000.
- [23] BREIMAN, L.. **Random forests**. *Machine learning*, 45(1):5–32, 2001.
- [24] GEMAN, S.; BIENENSTOCK, E. ; DOURSAT, R.. **Neural networks and the bias/variance dilemma**. *Neural Computation*, 4(1):1–58, 1992.
- [25] LIAO, T. W.. **Clustering of time series data—a survey**. *Pattern recognition*, 38(11):1857–1874, 2005.
- [26] MONTERO, P.; VILAR, J.. **Tsclust: An r package for time series clustering**. *Journal of Statistical Software, Articles*, 62(1):1–43, 2014.
- [27] ROUSSEEUW, P. J.. **Silhouettes: a graphical aid to the interpretation and validation of cluster analysis**. *Journal of computational and applied mathematics*, 20:53–65, 1987.
- [28] TAIEB, S. B.; ATIYA, A. F.. **A bias and variance analysis for multistep-ahead time series forecasting**. *IEEE transactions on neural networks and learning systems*, 27(1):62–76, 2016.

- [29] HYNDMAN, R. J.; KOEHLER, A. B.. **Another look at measures of forecast accuracy**. *International journal of forecasting*, 22(4):679–688, 2006.
- [30] GARCÍA, S.; FERNÁNDEZ, A.; LUENGO, J. ; HERRERA, F.. **Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power**. *Information Sciences*, 180(10):2044–2064, 2010.
- [31] ARMSTRONG, J. S.. **Significance tests harm progress in forecasting**. *International Journal of Forecasting*, 23(2):321–327, 2007.
- [32] KOSTENKO, A. V.; HYNDMAN, R. J.. **Forecasting without significance test**. manuscript, Monash University, Australia, 2008.
- [33] ŠTĚPNIČKA, M.; BURDA, M.. **On the results and observations of the time series forecasting competition cif 2016**. In: *FUZZY SYSTEMS (FUZZ-IEEE), 2017 IEEE INTERNATIONAL CONFERENCE ON*, p. 1–6. IEEE, 2017.
- [34] KAUFMAN, L.; ROUSSEEUW, P. J.. **Finding groups in data: an introduction to cluster analysis**, volumen 344. John Wiley & Sons, 2009.
- [37] HYNDMAN, R. J.; KOEHLER, A. B.; SNYDER, R. D. ; GROSE, S.. **A state space framework for automatic forecasting using exponential smoothing methods**. *International Journal of Forecasting*, 18(3):439–454, 2002.
- [38] DE GOOIJER, J. G.; HYNDMAN, R. J.. **25 years of time series forecasting**. *International journal of forecasting*, 22(3):443–473, 2006.
- [39] FILDES, R.. **An evaluation of bayesian forecasting**. *Journal of Forecasting*, 2(2):137–150, 1983.
- [40] GOODWIN, P.; OTHERS. **The holt-winters approach to exponential smoothing: 50 years old and going strong**. *Foresight*, 19:30–33, 2010.
- [41] TAYLOR, J. W.. **Short-term electricity demand forecasting using double seasonal exponential smoothing**. *Journal of the Operational Research Society*, 54(8):799–805, 2003.

- [42] RIBEIRO, M.; GROLINGER, K. ; CAPRETZ, M. A. M.. **Mlaas: Machine learning as a service**. In: 2015 IEEE 14TH INTERNATIONAL CONFERENCE ON MACHINE LEARNING AND APPLICATIONS (ICMLA), p. 896–902, Dec 2015.
- [43] GUERRERO, V. M.. **Time-series analysis supported by power transformations**. Journal of Forecasting, 12(1):37–48, 1993.
- [44] CLEVELAND, W. S.. **Visualizing data**. Hobart Press, 1993.
- [45] KUNSCH, H. R.. **The jackknife and the bootstrap for general stationary observations**. The annals of Statistics, p. 1217–1241, 1989.
- [46] SINGH, K.. **On the asymptotic accuracy of efron’s bootstrap**. The Annals of Statistics, p. 1187–1195, 1981.
- [47] PETROPOULOS, F.; HYNDMAN, R. J. ; BERGMEIR, C.. **Exploring the sources of uncertainty: Why does bagging for time series forecasting work?** European Journal of Operational Research, 268(2):545–554, 2018.
- [48] HYNDMAN, R.; KHANDAKAR, Y.. **Automatic time series forecasting: The forecast package for r**. Journal of Statistical Software, 27(3), 2008.
- [49] BERNDT, D. J.; CLIFFORD, J.. **Using dynamic time warping to find patterns in time series**. In: KDD WORKSHOP, número 16, p. 359–370. Seattle, WA, 1994.
- [51] EITER, T.; MANNILA, H.. **Computing discrete fréchet distance**. Technical report, Citeseer, 1994.
- [51] CHOUAKRIA, A. D.; NAGABHUSHAN, P. N.. **Adaptive dissimilarity index for measuring time series proximity**. Advances in Data Analysis and Classification, 1(1):5–21, 2007.
- [52] GALEANO, P.; PEÑA, D. P.. **Multivariate analysis in vector time series**. Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo, 4(4):383–403, 2000.
- [53] CAIADO, J.; CRATO, N. ; PEÑA, D.. **A periodogram-based metric for time series classification**. Computational Statistics & Data Analysis, 50(10):2668–2684, 2006.
- [54] MAHARAJ, E. A.; D’URSO, P.. **Fuzzy clustering of time series in the frequency domain**. Information Sciences, 181(7):1187–1211, 2011.

- [55] CASADO DE LUCAS, D.. **Classification techniques for time series and functional data**. PhD thesis, Universidad Carlos III de Madrid, 2010.
- [56] ZHANG, H.; HO, T. B.; ZHANG, Y. ; LIN, M.-S.. **Unsupervised feature extraction for time series clustering using orthogonal wavelet transform**. *Informatica*, 30(3), 2006.
- [57] LIN, J.; KEOGH, E.; WEI, L. ; LONARDI, S.. **Experiencing sax: a novel symbolic representation of time series**. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.
- [58] KEOGH, E.; LONARDI, S. ; RATANAMAHATANA, C. A.. **Towards parameter-free data mining**. In: *PROCEEDINGS OF THE TENTH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, p. 206–215. ACM, 2004.
- [59] KEOGH, E.; LONARDI, S.; RATANAMAHATANA, C. A.; WEI, L.; LEE, S.-H. ; HANDLEY, J.. **Compression-based data mining of sequential data**. *Data Mining and Knowledge Discovery*, 14(1):99–129, 2007.
- [60] SINGH, D.; DALEI, N. N. ; RAJU, T. B.. **Forecasting investment and capacity addition in indian airport infrastructure: Analysis from post-privatization and post-economic regulation era**. *Journal of Air Transport Management*, 53:218–225, 2016.
- [61] BARNHART, C.; BELOBABA, P. ; ODONI, A. R.. **Applications of operations research in the air transport industry**. *Transportation science*, 37(4):368–391, 2003.
- [62] GRUBB, H.; MASON, A.. **Long lead-time forecasting of uk air passengers by holt–winters methods with damped trend**. *International Journal of Forecasting*, 17(1):71–82, 2001.
- [63] CHIN, A. T.; TAY, J. H.. **Developments in air transport: implications on investment decisions, profitability and survival of asian airlines**. *Journal of Air Transport Management*, 7(5):319–330, 2001.
- [64] NJEGOVAN, N.. **A leading indicator approach to predicting short-term shifts in demand for business travel by air to and from the uk**. *Journal of Forecasting*, 24(6):421–432, 2005.
- [65] LAI, S. L.; LU, W.-L.. **Impact analysis of september 11 on air travel demand in the usa**. *Journal of Air Transport Management*, 11(6):455–458, 2005.

- [66] ALEKSEEV, K.; SEIXAS, J.. **A multivariate neural forecasting modeling for air transport—preprocessed by decomposition: A brazilian application.** *Journal of Air Transport Management*, 15(5):212–216, 2009.
- [67] CARSON, R. T.; CENESIZOGLU, T. ; PARKER, R.. **Forecasting (aggregate) demand for us commercial air travel.** *International journal of Forecasting*, 27(3):923–941, 2011.
- [68] KOTEGAWA, T.; DELAURENTIS, D. A. ; SENGSTACKEN, A.. **Development of network restructuring models for improved air traffic forecasts.** *Transportation Research Part C: Emerging Technologies*, 18(6):937–949, 2010.
- [69] SAMAGAIO, A.; WOLTERS, M.. **Comparative analysis of government forecasts for the lisbon airport.** *Journal of Air Transport Management*, 16(4):213–217, 2010.
- [70] GROSCHE, T.; ROTHLAUF, F. ; HEINZL, A.. **Gravity models for airline passenger volume estimation.** *Journal of Air Transport Management*, 13(4):175–183, 2007.
- [71] XIAO, Y.; LIU, J. J.; HU, Y.; WANG, Y.; LAI, K. K. ; WANG, S.. **A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting.** *Journal of Air Transport Management*, 39:1–11, 2014.
- [72] MAKRIDAKIS, S.. **Accuracy measures: theoretical and practical concerns.** *International Journal of Forecasting*, 9(4):527–529, 1993.
- [73] TONG, H.. **Non-linear time series: a dynamical system approach.** Oxford University Press, 1990.
- [74] BROCKWELL, P. J.; DAVIS, R. A.. **Time series: theory and methods.** Springer Science & Business Media, 2013.
- [75] HYNDMAN, R. J.; ATHANASOPOULOS, G.. **Forecasting: principles and practice.** OTexts, 2014.
- [76] **Rob hyndmans’blog post on the m4 competition.** <https://robjhyndman.com/hyndsight/m4comp/>. Accessed: 2017-11-17.
- [77] SHYU, W. M.; GROSSE, E. ; CLEVELAND, W. S.. **Local regression models.** In: *STATISTICAL MODELS IN S*, p. 309–376. Routledge, 2017.

- [79] CHU, C.-T.; KIM, S. K.; LIN, Y.-A.; YU, Y.; BRADSKI, G.; OLUKOTUN, K. ; NG, A. Y.. **Map-reduce for machine learning on multicore**. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, p. 281–288, 2007.
- [80] DEAN, J.; GHEMAWAT, S.. **Mapreduce: simplified data processing on large clusters**. Communications of the ACM, 51(1):107–113, 2008.
- [81] **Official m4 competition webpage**. <https://www.m4.unic.ac.cy/>. Accessed: 2018-07-25.
- [82] MAKRIDAKIS, S.; SPILIOTIS, E. ; ASSIMAKOPOULOS, V.. **The m4 competition: Results, findings, conclusion and way forward**. International Journal of Forecasting, 2018.
- [83] ARMSTRONG, J. S.. **Principles of forecasting: a handbook for researchers and practitioners**, volumen 30. Springer Science & Business Media, 2001.
- [84] DE OLIVEIRA, E. M.; OLIVEIRA, F. L. C.. **Forecasting mid-long term electric energy consumption through bagging arima and exponential smoothing methods**. Energy, 144:776–788, 2018.
- [85] SALEM, A. M.. **Facebook distributed system case study for distributed system inside facebook datacenters**. International Journal of Technology Enhancements and Emerging Engineering Research, 2, 2014.
- [86] RAMAKRISHNAN, L.; ZBIEGEL, P. T.; CAMPBELL, S.; BRADSHAW, R.; CANON, R. S.; COGLAN, S.; SAKREJDA, I.; DESAI, N.; DECLERCK, T. ; LIU, A.. **Magellan: experiences from a science cloud**. In: PROCEEDINGS OF THE 2ND INTERNATIONAL WORKSHOP ON SCIENTIFIC CLOUD COMPUTING, p. 49–58. ACM, 2011.