International Journal of Production Economics 212 (2019) 186–211

Contents lists available at ScienceDirect

International Journal of Production Economics

journal homepage: www.elsevier.com/locate/ijpe

Optimization of combined time series methods to forecast the demand for coffee in Brazil: A new approach using Normal Boundary Intersection coupled with mixture designs of experiments and rotated factor scores

Livio Agnew Bacci, Luiz Gustavo Mello, Taynara Incerti, Anderson Paulo de Paiva, Pedro Paulo Balestrassi*

Institute of Industrial Engineering and Management, Federal University of Itajuba, Itajuba, Minas Gerais, Brazil

ARTICLE INFO

Keywords:
Multi-objective optimization
Time series
Combining forecasts
Multivariate statistics
Decision support
Mixture design of experiments

ABSTRACT

This paper proposed a new multi-objective approach to find the optimal set of weight's combination of forecasts that were jointly efficient with respect to various performance and precision metrics. For this, the residues' series of each previously selected forecasts methods were calculated and, to combine them through of a weighted average, several sets of weights were obtained using Simplex - Lattice Design {m,q}. Then, several metrics were calculated for each combined residues' series. After, Principal Components Factor Analysis (PCFA) was used for extracting a small number series' factor scores to represent the metrics selected with minimal loss of information. The extracted series' factor scores were mathematically modeled with Mixture Design of Experiments (DOE-M). Normal Boundary Intersection method (NBI) was applied to perform joint optimization of these objective functions, allowing to obtain different optimal weights set and the Pareto frontier construction. As selection criteria of the best optimal weights' set were used the Shannon's Entropy Index and Global Percentage Error (GPE). Here, these steps were successfully applied to predict coffee demand in Brazil as a case study. In order to test the applicability and feasibility of the proposed method based on distinct time series, the coffee's Brazilian production and exportation were also foreseen by new method. Besides, the simulated series available in Montgomery et al. (2008) were also used to test the viability of the new method. The results showed that the proposed approach, named of FA-NBI combination method, can be successfully employed to find the optimal weights of a forecasts' combination.

1. Introduction

In 2017–2018 period, Brazil maintained the position of the world's largest coffee producer. According to the International Coffee Organization (ICO, 2018), for this period, Brazil produced around 51 million bags of 60 kg coffee, which represented approximately 31.94% of the global coffee production. Regarding the exportation, according to this same organization, Brazil – the world's largest exporter – exported 30.5 million bags of 60 kg coffee in 2017–2018, which represented approximately 26.40% of all coffee exported in the world. In addition to being the largest producer and exporter of green coffee beans, Brazil is also the second largest consumer behind only the United States. In the same period of 2017–2018, Brazil consumed around 20.5 million bags of 60 kg coffee according to ICO (2018).

Analyzing the numbers of the annual period of 2017–2018, one can note that almost all the coffee produced in Brazil is sold domestically and in foreign markets, which hardly leaves any security stocks. In relation to the coffee production chain, the lack of safety stocks and a likely coffee consumption growth in the coming years will require strategically constant adequacy of their physical and managerial structure. So that the domestic and external markets do not fail to be supplied. As shown above, the domestic consumption forecast of coffee is fundamental to the whole members of agribusiness coffee chain, with part of it represented by Fig. 1, in order to avoid new problems of shortages or overproduction. In relation to industries of soluble, roasted and ground coffee, the demand forecast is important so that they can plan their productive capacity for the future. For the grain production sector, which exports the product to the international market and still

* Corresponding author. Institute of Industrial Engineering and Management, Federal University of Itajubá, Brazil Avenida BPS 1303 Itajubá, MG 37500-903, Brazil.

E-mail addresses: livioab@yahoo.com.br (L.A. Bacci), contato@gustavomello.adm.br (L.G. Mello), taynaraincerti@gmail.com (T. Incerti), andersonppaiva@unifei.edu.br (A. Paulo de Paiva), ppbalestrassi@gmail.com, pedro@unifei.edu.br (P.P. Balestrassi).

https://doi.org/10.1016/j.ijpe.2019.03.001
Received 22 March 2018; Received in revised form 11 January 2019; Accepted 1 March 2019
Available online 06 March 2019
0925-5273/

Fig. 1. Part of the coffee agroindustrial chain.

provides feedstock for the processing of coffee industries in Brazil, the forecast is important so that they can plan the crop and the supply of green coffee grains.

In relation to the choice of strategy for forecasting, over the last 5 decades, many advantages have been pointed out in combining forecast methods, such as:

(I) it aggregates information about the form of the relationship between the variables, since one forecasting method is based on variables or information that the other forecasts has not considered (Bates and Granger, 1969; Bunn, 1975; Chan et al., 1999; Moreno and López, 2013; Graef et al., 2014);
(II) it allows identifying the underlying process, since those different forecasting models are able to capture different information's aspects available for prediction (Reeves and Lawrence, 1982; Clemen, 1989);
(III) it takes into account the relative accuracy of individual methods and forecast errors' covariance among the methods (Winkler and Makridakis, 1983);
(IV) it improves forecasting accuracy (Makridakis and Winkler, 1983; Winkler, 1989; Mahmoud, 1989; Hibon and Evgeniou, 2005; Bjornland et al., 2012; Bordignon et al., 2013; Cang and Yu, 2014; Graef et al., 2014; Barrow and Kourentzes, 2016);
(V) it decreases accuracy's variability for different variance's measures (Makridakis and Winkler, 1983; Mahmoud, 1989; Hibon and Evgeniou, 2005);
(VI) it allows the uncertainty reduction, being safer and less risky than relying on a single method (Makridakis and Winkler, 1983; Winkler, 1989; Hibon and Evgeniou, 2005; Bordignon et al., 2013);
(VII) it may result in more normally distributed errors (Barrow and Kourentzes, 2016).

There are basically two methodologies to perform the combination of forecasting methods: simple unweighted average (SA or AVG) and weighted average (WA). Regard to the WA, during the last 50 years many articles have been written on classical techniques to find the best weights for combining forecasting methods (weights assigned to each individual forecast). Most of these studies focused on minimizing only one goal or single criterion as the error variance's combination (Bates and Granger, 1969); the average square error (Newbold and Granger, 1974); mean absolute percentage error (MAPE) (Winker and Makridakis, 1983; Lam et al., 2001); out-of-sample sum of squared forecast errors (Granger and Ramanathan, 1984; Deutsch et al., 1994; Chan et al.,1999); mean squared error (MSE) (Diebold and Pauly, 1990; Lesage and Magura, 1992); mean absolute error (MAE) (Lesage and Magura, 1992), and maximum absolute percentage error (MAXAPE) (Lam et al., 2001).

However, few studies employed multi-objective approaches. Reeves and Lawrence (1982) developed a multiple objective linear programming (MOLP) framework for generating combined forecasts that are efficient in respect to multiple objectives, minimizing four objectives simultaneously (total forecast error, positive forecast error over all periods, total forecast error over the recent periods, and maximum forecast error). They justified the use of MOLP by arguing that selecting a single individual forecast based upon a single objective could not make the best use of available information as a combined forecast could, considering the minimization of multiple objectives and not only the minimization of forecast error variance. Likewise, Gulledge et al. (1986) used MOLP to determine weighted linear combinations of forecasts in which three forecast goals were minimized (sum of the absolute forecast errors over all 24 periods, sum of the absolute forecasts errors with double weighting in the last 8 periods, and maximum absolute forecast error). Their justification was the fact that multiple objectives allow the decision maker to have more flexibility in selecting forecasted variables as inputs in their policy analysis and often provide a better fit to the specific decision problem. Reeves et al. (1988) argued that multi-objective mathematical weighting scheme allows the direct incorporation of a varied set of management objectives to be brought directly into the forecasting process and created a multi-objective linear programming that minimized total forecast error, over forecast errors and recent forecast error. Once again, Reeves and Lawrence (1991) developed a mathematical programming framework for combining forecasts given multiple objectives and incorporated two types of objectives – accuracy and direction of change and concluded that combined forecasts generated in a multi-objective environment are at least as good as the best individual forecast with respect to one or more objectives. As an argument for using this approach, they claimed that it allows decision makers to trade-off weighting schemes for combining forecasts with respect to multiple conflicting measures of forecast accuracy. Ten years later, in a portfolio based-approach, Leung et al. (2001) calculated a set of weights of a combination using a goal programming (GP) multi-objective approach. The authors used a GP model to combine the forecasts, such that the expected return and skewness of return were maximized while the variance (risk level) of the return was minimized. More recently, Ustum and Kasimbeyli (2012) made use of a general mean-variance-skewness model, in which these metrics were coupled in the multi-objective portfolio optimization model to find the optimal weights of each action in a great portfolio.

Differently, this work proposes a new multi-objective framework - the FA-NBI combination method - to find the optimal weights' time series combination that are jointly efficient with respect to various performance and precision metrics. Specifically to forecast coffee consumption in Brazil as a case study to demonstrate the steps for FA-NBI's application, four different time series' methods were selected previously to participate in the combination - Double Exponential Smoothing (DSE), Holt-Winters Multiplicative (HW) and two specifications of Autoregressive Integrated Moving Average (ARIMA). After, properly starting the implementation of the new approach, the residues' series for each of these methods were calculated. In sequence, with the use of simplex-lattice design {4, 5}, a Mixture Design of Experiments (DOE-M) approach, 61 wt' sets to combine the residues series of the 4 selected methods were obtained. For each set of weights generated by DOE-M, combined residues' series of the selected time series methods were obtained through a weighted average. Then, for each of the 61 combined residues' series found, 14 forecasting metrics were calculated, generating 14 metrics' series. Seeking to reduce the dimensionality of the problem, Principal Components Factor Analysis (PCFA) multivariate technique was applied for extracting a small number of factors scores' series to represent the 14 selected performance and precision's metrics, with minimal loss of information. Thus, in the context of time series combination, PCFA was used to reduce the number of metrics to be jointly optimized from 14 to 2. As a result, allowed us to find the weights' combination not only minimizing one or two metrics, but more than 14 metrics together, which made different this work from those mentioned above. Other difference was that, in sequence, DOE-M was used to model two mathematical objective functions of factor scores that represented the 14 original metrics, which without joint optimization of the same would not be possible. After these two factor scores' functions were modeled, the Normal Boundary Intersection (NBI) method was applied for the multi-objective optimization, being found 21 Pareto-optimal weights. Then, to not sacrifice the information provided by each of the 14 selected metrics, it was proposed to optimize them together to find the best combination of time series methods. To select the best optimal set of weights was used Shannon's Entropy Index (S) by Shannon (1948) and Global Percentage Error (GPE). FA-NBI combination found was statistically compared in terms of performance and precision using DM test, introduced by Diebold and Mariano (1995), with the combined individual methods and with the other traditional weighting methods. In order to prove the competitiveness, applicability and feasibility of the proposed approach, coffee's Brazilian production and exportation were also foreseen using FA-NBI and analyzed in conjunction with the coffee demand. The proposed method was also applied and tested based on simulated series available in Montgomery et al. (2008).

This paper is organized as follows: the next section presents the methodology proposed in this study to get the optimal weights' combination together with a brief literature review focusing on the forecast metrics most used in the literature to make comparisons between forecasting methods, PCFA, DOE-M, NBI, S, GPE and some of the more traditional methods for finding weights' combination. Section 3 presents the results analysis and the conclusions are presented in Section 4.

2. Methodology and background literature review

Part of terminology used in this section is displayed in Table 1.

To obtain the optimal weights for a combination of time series, this study proposes a combination of the aforementioned techniques that

Table 1
Nomenclature.

A	Aggregation
API	Forecasts' comparison of accuracy made by institutions
AR	Average ranking
AvgRelMAE	Average Relative Mean Absolute Error
C	*c-th* set the weights, with c = 1, 2, …, *p*
C	Combination
CC	Combination of combinations
CIM	Computational intelligence method
FM	Forecast methods
FS_q	*q-th* factors scores series
FS_{qp}	*q-th* factor score calculated based on *p-th* combined residues series
G	*g-th* observation of a given residues series
GMRAE	Geometric Mean Relative Absolute Error
M	*m-th* forecasting method, with m = 1, 2, …, z
NN	Neural Network
M_j	*j-th* original performance metric
M_{jp}	*j-th* original performance metric calculated based on the *p-th* combined residues series
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MdAE	Median Absolute Error
MdAPE	Median Absolute Percentage Error
MdRAE	Median Relative Absolute Error
ME	Mean Error
MSE	Mean Squared Error
MSPE	Mean Squared Percentage Error
P	*p-th* set of weights/*p-th* combined residues series
PB	Percentage Better
R_z	Residues series of *z-th* forecasting method
R_{zg}	*g-th* observation of *z-th* residues series
R_t^m	*t-th* observed residue, belonging to a series of residues of the *m-th* forecasting method
RC_{pg}	*g-th* observation of *z-th* residues series, series originated from *z-th* forecasting method
RC_t^c	*t-th* combined residue, calculated based on the *c-th* set of weights
RMSE	Root Mean Squared Error
RMSPE	Root Mean Square Percentage Error
SD	Standard Deviation of Errors
sMAPE	Symmetric Mean Absolute Percentage Error
sMdAPE	Symmetric Median Absolute Percentage Error
U1	U1 Theil's statistical coefficient
U2	U2 Theil's statistical coefficient
VAR	Variance of errors
t	*t-th* observation of the residues series of each forecasting method, with t = 1, 2, …, g
W_z	Weight assigned to *z-th* forecasting method
W_{zp}	Weight attributed to the *z-th* forecasting method considering the *p-th* set of weights
$W_{zp}R_{zg}$	Weight pertaining to the *p-th* set of weights assigned to the *z-th* forecasting method, multiplied by the *g-th* observation of the *z-th* residues series originating from the *z-th* forecasting method
w_m^c	Weight of the *m-th* method in *c-th* combination

were applied according to the following step-by-step procedure:

Step 1: Data collection.

Step 2: Selection of the z individual time series methods to take part in the combination. After, generate corresponding residues series (R) for each of these z chosen methods (*R1*, *R2*, …, *Rz* series), each series with g observations. The Appendix A provides a brief explanation of the forecasting methods selected in this paper to take part in the combination in order to specifically forecast the demand for coffee in Brazil, i.e. Double Exponential Smoothing (DES), Holt-Winters Multiplicative (WM) and Autoregressive Integrated Moving Average (ARIMA). As will be observed, with different time series, other forecasting methods will be selected.

Step 3: Using DOE-M, definition of p sets of weights (proportions) through a mixture simplex-lattice design {q, m}, with each set of weights with z weights (w_1, w_2, w_3, …, w_z), to combine the z residues series generated in Step 2. The simplex-lattice design technique will be explained in step 6. In sequence, were produced p series of combined residues, each one with g observations, according to the procedure

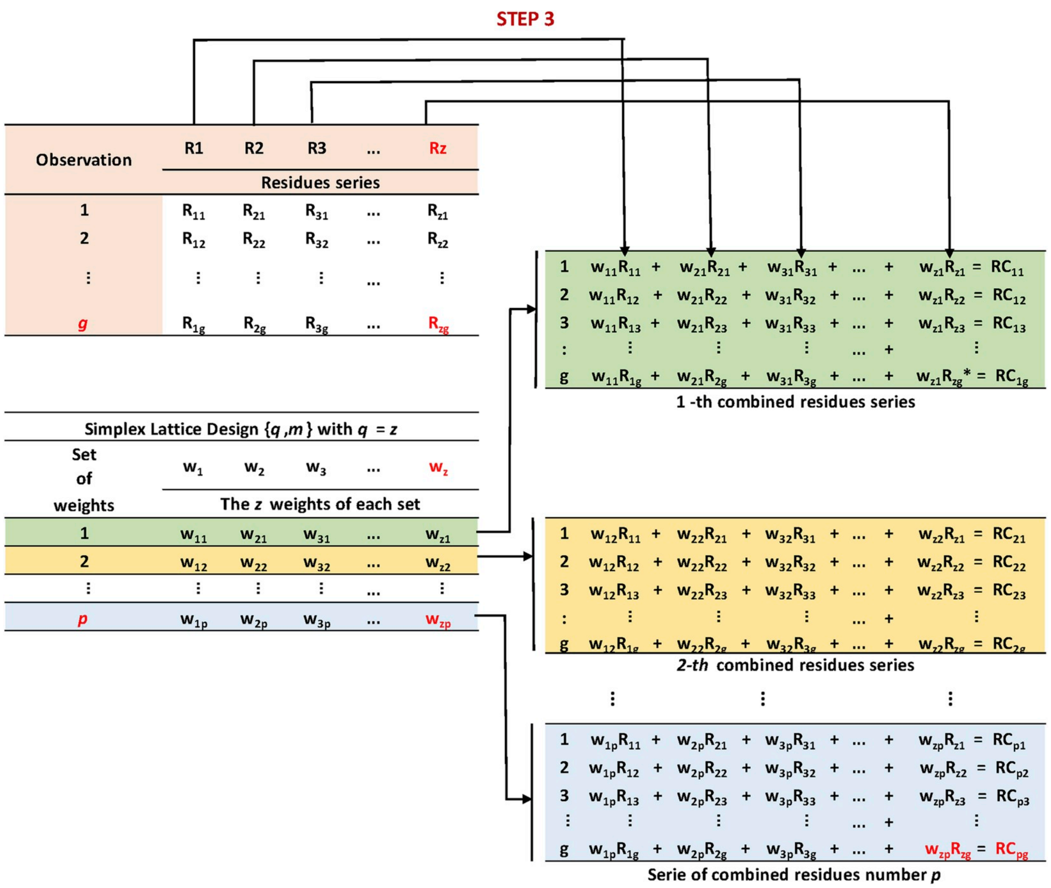

Fig. 2. Production of the p residues combined series.

described in Fig. 2.

In mathematical terms, the *t-th* observation from the *c-th* series of combined residues (RC_t^c) can be calculated by Eq. (1).

$$RC_t^c = \sum_{m=1}^{z} w_m^c R_t^m \tag{1}$$

The subsequent steps 4, 5 and 6 are indicated in Fig. 3.

Step 4: Calculation of the performance and precision (variability) forecast measures or metrics (M). In this step 4, j metrics must be calculated for each p combined series of weighted residues obtained in Step 3. With this, j metrics based on each series of combined residues will be produced, with each series of metrics with p observations,

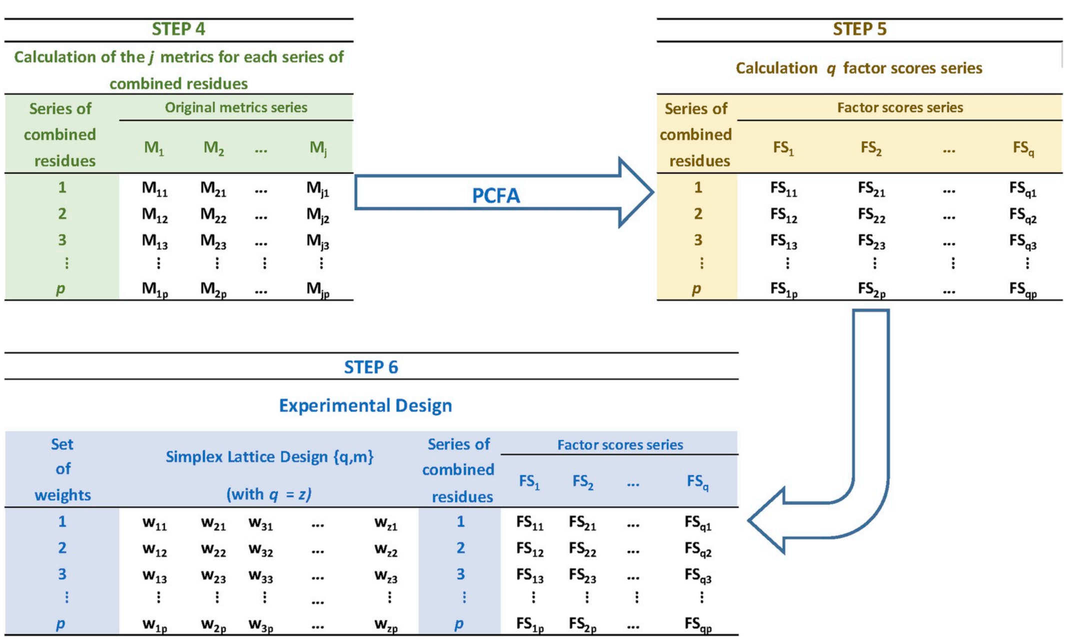

Fig. 3. Steps to obtain the experimental design.

according to Fig. 3.

The 14 metrics (j = 14) used in this work were MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1, U2, VAR, SD, MdAE, MdAPE and sMdAPE. The Equations to calculate these metrics are described in Appendix B at the end of this article. The variance (VAR) and standard deviation (SD) of errors or residues are metrics of risk associated to the choice of method in respect to the precision (variability), that is, of occurrence of high values thereof of above average and below average (Hibon and Evgeniou, 2005). The twelve others are metrics of accuracy (performance), that demonstrate the "ability of a forecasting method to forecast actual data, either when a model is fitted to such data, or for future periods (post-sample) whose values have not been used to develop the forecasting model" (Makridakis, 1993). MAE measures how much the estimated or forecasted value differ from actual values (Xu et al., 2017). MSE is the average of the prediction error squares, which can be applied for estimating the change of forecasting models (Du et al., 2018). RMSE summarizes the difference between the actual and forecasted values (Xu et al., 2017). MAPE is a prediction accuracy's measure of a forecasting method in statistics (Du et al., 2018). Like MAPE, sMAPE is an average of the absolute percent errors but these errors are computed using a denominator representing the average of the forecast and observed values. MASE is a measure of forecast accuracy proposed by http://www.sciencedirect.com/science/article/pii/S0169207006000239 Hyndman and Koehler (2006). It is the absolute error (MAE) of a given method divided (i.e. scaled) by the Mean Absolute Error (MAE) of the naïve benchmark model. RMSPE is the square root of the average sum of squared errors expressed as a percentage. U1 Theil's statistic coefficient is conceived as a measure of accuracy, while U2 Theil's is a statistical coefficient of forecasting quality to make comparisons between the chosen forecast and the naïve prediction. Both measures are bounded by zero (the case of perfect forecasting) and one (allegedly, the worst case) (Martínez-Rivera et al., 2012). MdAE is a robust measure of the variability of a univariate sample of quantitative data. MdAPE is calculated by ordering the absolute percentage error (APE) from the smallest to the largest and using its middle value as median. It is more resilient to outliers than MAPE and sMAPE and is recommended as a model performance evaluation criterion when forecasting models need to be compared across different series (Xu et al., 2017). The sMdAPE is obtained by ordering the symmetric absolute percentage errors and using the middle value.

The choice of the error measure to be used in each case depended on the characteristics presented by the data series, such as differences in scale across the series, the amount of change that occurs over the horizon forecast, the presence of extreme forecast error outliers, and the purpose of research (Armstrong and Collopy, 1992; Ahlburg, 1992). Due to the fact that each measure of error has its weak points, several empirical studies found in the literature used more than one performance measure to select the best forecasting method and/or the best combination of forecasting methods, as can be observed in the Table 2. In these studies, the best method or the best combination of methods was chosen based on those who had the lowest errors measures independently. In a different way, because each metric has its weak and strengths points, in addition to each providing a different information, in the present work the choice of combination weights will be the one that optimizes at the same time all the 14 metrics (the objectives).

Step 5: Application of PCFA to reduce the dimensionality of the problem, obtaining q series of factor scores, with each one of these q series representing one or more of one of the j metrics obtained in Step 4, such that $j < q$. Considering the linear factor analysis model extracted by principal components and applying the Varimax rotation, the number of q retained factors was determined by Kaiser's rule (1960). Each of the observations of the q factor scores series represented a response related to a given set of combined residues and, consequently, to a given set of weights selected in Step 3 (Fig. 3).

PCFA is multivariate analysis technique that allows reducing the dimensionality of a data set, separating each group of highly correlated variables, forming, for each of these groups of correlated variables, a construct or factor responsible for the observed correlations. In PCFA, using covariance matrix **S** or correlation matrix **R** for factor extraction, the p original variables, which form a random vector of observed variables $\mathbf{X} = [X_1, X_2, \ldots, X_p]$, with p components, mean μ and covariance matrix **S** or correlation matrix **R,** are represented by p linear functions. As can be seen in Eq. (2), the set of these linear functions form the linear factor analysis model with m common factors. Each one of these functions is linearly dependent of m random, hypothetical, latent and unobservable variables ($F_1, F_2, \ldots, F_m$), known as common factors or constructors (F), plus a random error component (e_j), with $m < p$ (Johnson and Wichern, 2007).

$$\begin{aligned} X_1 - \mu_1 &= \boldsymbol{l}_{11}F_1 + \boldsymbol{l}_{12}F_2 + \ldots + \boldsymbol{l}_{1m}F_m + e_1 \\ X_2 - \mu_2 &= \boldsymbol{l}_{21}F_1 + \boldsymbol{l}_{22}F_2 + \ldots + \boldsymbol{l}_{2m}F_m + e_2 \\ &\vdots \\ X_p - \mu_p &= \boldsymbol{l}_{p1}F_1 + \boldsymbol{l}_{p2}F_2 + \ldots + \boldsymbol{l}_{pm}F_m + e_m \end{aligned} \tag{2}$$

It may be noted in Eq. (2) that PCFA reduces the problem of dimensionality from p to m dimensions, modeling representative functions of the original variables, with each of these functions being formed by m factors or hypothetical variables $F_1, F_2, \ldots, F_m$. Thus, a small number of m factors can be used to explain many p original variables.

The common factors $F_1, F_2, \ldots, F_m$, extracted from the covariance matrix **S** (or correlation matrix **R**), are calculated on the basis of pairs of eigenvalues-eigenvectors estimated $(\hat{\lambda}_1, \hat{e}_1), (\hat{\lambda}_2, \hat{e}_2), \ldots, (\hat{\lambda}_p, \hat{e}_p)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$. Each one of the functions is also formed by l_{jk} terms known as factor loadings, with $j = 1, 2, \ldots, p$ and $k = 1, 2 \ldots, m,$ such that the l_{jk} term is the loading of the *j-th* variable on the *k-th* factor. Loadings close to 1 or 1 indicate that the factor strongly represents the variable, while loadings close to zero indicate that the factor has a weak influence on the variable (Johnson and Wichern, 2007). A good rule of thumb is that standardized loading estimates should be 0.5 or higher, and ideally 0.7 (Hair et al., 2010). The error terms e_j are also known as specific factors, because each e_j is specific to each function (X_j - μ_j),

Table 2
Performance measures used individually to select and/or combine forecasting methods.

Authors	Choose between	Performance metrics used
Winkler and Makridakis (1983)	FM, C, CC	MAPE, MSE
Kang (1986)	FM, C	MAE, MSE
Sankaran (1989)	C, C	MAPE, MSE
Diebold and Pauly (1990)	FM, C	RMSE, MAE
Makridakis (1990)	FM, C	MSE, MAPE, MdAPE
Tseng et al. (2002)	FM, NN, C	MSE, MAE, MAPE
Meade (2002)	FM	MdAE, RMSE, PB
Weatherford and Kimes (2003)	FM, C	MAE, MAPE
Fang (2003)	FM, C	RMSPE, MAE, RMSE
Dekker et al. (2004)	FM, A, C	sMAPE, MAE, MSE, AR
Faria and Mubwandarikwa (2008)	C, CC	MAE, RMSE, GMRAE
Jose and Winkler (2008)	FM, C	sMAPE, sMdAPE
Wallstrom and Segerstedt (2010)	FM	MSE, MAE, sMAPE
Crone et al. (2011)	FM, NN, CIM, C	sMAPE, MdRAE, MASE, AR
Andrawis et al. (2011)	C, CC	sMAPE, MASE
Martins and Werner (2012)	FM, C, CC	MAPE, MSE, MAE
Bordignon et al. (2013)	FM, C	MSE, MSPE, MAE, MAPE
Simionescu (2013)	API	ME, MAE, RMSE, U1, U2
Adhikari and Agrawal (2014)	FM, NN, C	MSE, sMAPE
Petropoulos et al. (2014)	FM, C, CC	sMAPE, PB, MASE
Cang and Yu (2014)	FM, NN, C, CC	MAPE, MASE
Zhao et al. (2014)	FM, C	RMSE, MAE, MAPE
Fildes and Fotios (2015)	FM, C	MdAPE, MAPE, AvgRelMAE
Tselentis et al. (2015)	FM, C	RMSE, MAPE
Barrow and Crone (2016)	NN, C	sMAPE, MASE

which does not happen with the common factors $F_1, F_2, \ldots, F_m$, since it integrates all linear functions (X_j-μ_j) representing each variable (Jolliffe, 2002; Johnson and Wichern, 2007). In each *j-th* function (X_j - μ_j), the part of variance explained by m common factors ($F_1, F_2, \ldots, F_m$) is known as communality (h_j^2). In addition, the part of variance explained by a specific factor (e_j) is known as uniqueness or specific variance (ψ) Thus, $\sigma_{jj} = l_{j1}^2 + l_{j2}^2 + \ldots + l_{jm}^2 + \psi_j$ or $\sigma_{jj} = h_j^2 + \psi_j$ explain the total variance of the variable *j* in the generated linear factor analysis model. The *j-th* estimated communality ($\tilde{h}_j^2$) is calculated by the sum of the squares of the estimated loadings ($\tilde{l}_{jk}$) of the *j-th* variable in m common factors ($\tilde{h}_j^2 = \tilde{l}_{j1}^2 + \tilde{l}_{j2}^2 + \ldots + \tilde{l}_{jm}^2$) , with $k = 1, 2, \ldots, m$. The estimated $\tilde{l}_{ji}^2$ is the contribution of *i-th* common factor to the variance of *j-th* variable (s_{ji}), and $\hat{\lambda}_1 = \tilde{l}_{11}^2 + \tilde{l}_{21}^2 + \ldots + \tilde{l}_{p1}^2$ is the estimated contribution of the first common factor for the total sample variance. (Johnson and Wichern, 2007).

In case where there are many high loadings in a given factor complicating the selection of a single variable to represent the set of variables series, factor scores series can be calculated for creating smaller set of variables series to replace the original set. Factor scores series representative of the original variable series with high loadings in a given factor are computed based on the factor loadings of all variables on the factor (Hair et al., 2010). The factor scores $\hat{f}_j$ are estimated values for the unobserved random factor vectors F_j, with $j = 1, 2, \ldots, n$, that is, $\hat{f}_j$ is the estimate of the values f_j attained by F_j. The factor scores can be calculated by Weighted Least Squares or Regression methods (Johnson and Wichern, 2007).

The key issue in PCFA is to know the number of factors (m) to be retained because, according to Haylton et al. (2004), specifying too few factors or too many factors are substantial errors that affect the results, although specifying too few is considered more severe. When extracting factors employing the correlation matrix (***R***), the Kaiser's rule (1960) can be used, where the factors with eigenvalues greater than 1 will be retained. Still, the initial factor loadings can be rotated through many rotation methods, each one using a different algorithm in the search for the factor structure simplification. Orthogonal rotation produces factors that are uncorrelated (i.e., maintain a 90° angle between axes), being more easily interpretable. There is a wide variety of orthogonal rotation types - equimax, varimax, quatmax orthomax, although varimax is the most used because it was developed as an improvement in the quartimax and equamax algorithms (Osborn, 2015).

Step 6: Mathematical modeling of the q factor scores series representative of *j* original metrics with the use of DOE-M. Each one of the p observations of the q factor scores series represented a response related to a given set of combined residues and, consequently, to a given set of weights in simplex lattice design $\{q, m\}$ selected in Step 3. The sets of weights defined in Step 3 combined with the responses (components) produced in Step 5 determined the experimental design of mixtures and, therefore, the designs of the response surfaces, one for each series of factors scores series, from which were extracted the mathematical models (objective functions). In design of experiments, the exploration of the response surfaces $\boldsymbol{\phi}FS_1, \boldsymbol{\phi}FS_2, \ldots, \boldsymbol{\phi}FS_q$ over the simplex region design will provide the respective polynomial equations $\widehat{FS_1}(w), \widehat{FS_2}(w), \widehat{FS_3}(w), \ldots, \widehat{FS_q}(w)$ which represents the surface over the region of interest. Which w represents a given set of weights formed by the weights of each of the z forecast methods selected. For each selected w, the objective functions will give as responses, respectively, the predicted values of $\widehat{FS_1}(w), \widehat{FS_2}(w), \widehat{FS_3}(w), \ldots, \widehat{FS_q}(w)$.

The main goal of DOE-M is to try to model the dependence of the response variable on the relative proportions of the components with some form of mathematical equation (Cornell, 2002; Coronado et al., 2014). As a special type of response surface, in DOE-M the factors are the mixture components, and the response n is a function of the proportions w_i of each component in the mixture, with $w_i \geq 0$ and $\sum_{i=1}^{q} w_i = w_1 + w_2 + \cdots + w_q = 1$, where $i = 1, 2, \ldots, q$, with q being the number of components in the mixture (Cornell, 2002; Myers et al., 2009). These constraints for w_i provide the geometric description of the factor space S^{q-1} (experimental region) containing q components, with each component representing a space vertex. The factor space consists of all points on or inside the boundaries (vertices, edges, faces etc.) of a regular (q - 1) dimensional simplex. Thereby, the set of points defined in an experimental region describes the possible proportions for the mixture components. The uniformly spaced distribution of points on a simplex is known as a lattice. In a simplex-lattice design $\{q, m\}$, the number of points in the design depends not only on the number of components (q) in the mixture, but also on the degree (m) of the polynomial model. Considering that the proportions assumed by each component take the $m + 1$ equally space values form 0 to 1, that is $w_i = 0, \frac{1}{m}, \frac{2}{m}, \ldots, 1$, plus the simplex-lattice$\{q, m\}$ consists of all possible combinations (mixtures) of components, the number of design points in the simplex-lattice $\{q, m\}$will be $N = \frac{(q+m-1)!}{m!(q-1)!}$ (Cornell, 2002; Myers et al., 2009).

The set of n responses give rise to a response surface $\boldsymbol{\phi}$. Therefore, there is a functional relationship $n = \boldsymbol{\phi}(w_1, w_2, \ldots, w_q)$, which exactly describes the surface $\boldsymbol{\phi}$, with n dependent on the proportions $w_1, w_2, \ldots, w_q$ of the q components. Generally, polynomial functions are used to represent the response surface $\phi(w_1, w_2, \ldots, w_q)$. The justification being that one can expand $\boldsymbol{\phi}$ $(w_1, w_2, \ldots, w_q)$ using a Taylor series, and thus a polynomial can be used also as an approximation (Cornell, 2002, 2011). The parameters in the $\{q, m\}$ polynomials are estimated as observed values' functions of responses at the points over $\{q, m\}$ simplex-lattice designs (Cornell, 2011). So, there is a correlation between the number of points in the simplex and the number of terms in the polynomial (Oliveira et al., 2011). Therefore, the properties of the polynomials used to estimate the response function depend to a substantial extent on the specific experimental design. Normally, the canonical form of the polynomial (or Scheffé form) that can be used to fit the data and build a model (objective function) from a mixture experiment can be, for example, quadratic, as in Eq. (3) (Myers et al., 2009).

$$n = E(y) = \beta_0 + \sum_{i=1}^{q} \beta_i w_i + \sum \sum_{i<j=2}^{q} \beta_{ij} w_i w_j + \varepsilon \tag{3}$$

Step 7: Once the mathematical objective functions of the q factor scores series have been modeled leading to the creation of multiple mixture response surfaces, one for each factor scores function, these multiple surfaces will need to be combined into a multi-objective optimization problem. After, $\widehat{FS_1}$ (w), $\widehat{FS_2}$ (w), …, $\widehat{FS_q}$ (w) functions were simultaneously optimized using the NBI, being that the joint optimization of $\widehat{FS_1}(w), \widehat{FS_2}(w), \ldots, \widehat{FS_q}(w)$ in reality represented the joint optimization of the *j* original metrics represented by them. Thus, each of the v Pareto-optimal solutions found using NBI and, consequently, each of the v sets of optimal weights found represented a point at the Pareto frontier in which it would not be possible to reduce the value of one of those metrics without causing the simultaneous increase of, at least, another metric.

In contrast to single-objective optimization (SOP), which only has one global optimum response, a multi-objective optimization problem (MOP) involves simultaneous multiple objectives' optimization and generates, rather than a single optimal solution, several Pareto-optimal solutions, where no solution can be considered better or worse than the others. This way, multi-objective optimization algorithms can search for and gather a series of optimal solutions for a problem with more than one objective at the same time, without either being considered better or worse than the other (Du et al., 2017). Therefore, a vector of viable solutions (***x***) is Pareto-optimal if there is no other feasible point y to reduce any of the objective functions without causing the simultaneous increase of, at least, another objective function (Vahidinasab et al., 2010). This set of optimal non-dominated solutions (***x***) form the Pareto frontier or surface. That is, the set which includes values of the Pareto-optimal solutions is called the Pareto-optimal front. (Wang et al., 2018). The Eq. (4) describes a minimization of regular multi-objective optimization problem (Wang et al., 2017; Du et al., 2017).

$$Minimize: F(X) = \{f_1(x), f_2(x), ..., f_0(x)\}$$

$$Subject\ to: \quad g_j(x) \geq 0, j = 1,2, ..., m$$

$$h_j(x) = 0, j = 1,2, ..., p$$

$$L_j \leq x_j \leq U_j, j = 1,2, ..., n \tag{4}$$

where n, o, m and p represent the number of the variables, objective functions, inequality constraints and equality constraints, respectively; g_j and h_j are the *j-th* inequality and equality constraints; and $[L_j, U_j]$ stand for the boundaries of *j-th* variable.

The definitions of Pareto dominance, Pareto optimality, Pareto-optimal set and Pareto-optimal are respectively given by Equations (5)–(7) e 8 (Wang et al., 2017; Du et al., 2017).

Taking two vectors, $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ and $\boldsymbol{y} = (y_1, y_2, ..., y_n)$, $\boldsymbol{x}$ dominates $\boldsymbol{y}$ (Pareto dominance), that is, $x > y$ if:

$$\forall\ i \in [1, n], [f(x_i) \geq f(y_i)] \ \wedge [\exists\ i \in [1, n]: f(x_i)] \tag{5}$$

The Pareto-optimal x belongs to $\mathbf{X}$ if:

$$\nexists y \in XF(y) > F(x) \tag{6}$$

A Pareto-optimal set is defined as:

$$P_s: \{x, y \in X | \ \exists\ F(y) > F(x)\} \tag{7}$$

A Pareto-optimal front is a set including the value of objectives functions for Pareto solutions set:

$$P_f = \{F(x) | \ x \in P_s\} \tag{8}$$

NBI, introduced by Das and Dennis (1998), is an optimization routine traditionally used to generate a near-uniform spread Pareto frontier for MOPs – regardless to the distribution of weights – on which the multiple solutions with gradual trade-offs in the objectives are obtained (Ganesan et al., 2013; Lopes et al., 2016). Another advantage of the NBI is its scales' independence of different objectives functions (Jia et al., 2007; Shukla, 2007). The first step in the NBI method is the calculation of the Payoff matrix $\mathbf{\Phi}$ elements for m objectives functions, as in Eq. (9). The $\mathbf{\Phi}$ will be a $(m\ x\ m)$ matrix calculated by obtaining the individual minimum value of each objective function $f_i(x)$. The solution that minimizes the *i-th* objective function $f_i(x)$ is $f_i^*(x_i^*)$, which indicates the minimum of $f_i(x)$ obtained in the point x_i^*. Thus, x_i^* is the solution that minimizes $f_i(x)$, with $(i = 1, ..., m)$. The remaining elements of each row of the $\mathbf{\Phi}$, $f_i(x_i^*)$, are calculated by replacing each optimal point x_i^* obtained, in all other objective functions. In the Payoff matrix, these values are represented by $f_1(x_i^*)$, $f_2(x_i^*)$, ..., $f_{i-1}(x_i^*)$, ..., $f_{i+1}(x_i^*)$, ..., $f_m(x_i^*)$. Therefore, each row of the matrix $\mathbf{\Phi}$ consists of optimal values $f_i^*(x_i^*)$ and non-optimal values $f_i(x_i^*)$ of the *i-th* objective function, indicating the upper and lower limits of these objective functions (Vahidinasab et al., 2010; Brito et al., 2014).

$$\Phi = \begin{bmatrix} f_1^*(x_1^*) & \cdots & f_1(x_i^*) & \ldots & f_1(x_m^*) \\ \vdots & \ddots & & & \vdots \\ f_i(x_1^*) & & f_i^*(x_i^*) & & f_i(x_m^*) \\ \vdots & & & \ddots & \vdots \\ f_m(x_1^*) & \ldots & f_m(x_i^*) & \ldots & f_m^*(x_m^*) \end{bmatrix} \tag{9}$$

The Utopia points are generally points outside the feasible region and they can be represented by vector $f_i^U = [f_1^*(x_1^*), f_2^*(x_2^*), ..., f_i^*(x_i^*), ..., f_m^*(x_m^*)]^T$. They are formed by the global minimum value of each *i-th* optimized objective function (the best values in each row of the Φ). The components of the vector $\boldsymbol{f_i^U}$ will form the main diagonal of the $\mathbf{\Phi}$. In addition, the Nadir points, represented by vector $\boldsymbol{f_i^N} = [f_1^N, f_2^N, ..., f_i^N, ..., f_m^N]^T$, are the worst values found for each *i-th* objective function (worse values of each row of the $\mathbf{\Phi}$) (Utyuzhnikov et al., 2009; Vahidinasab et al., 2010). After defining the Utopia and Nadir points, the objective functions must be normalized, especially if they are in different scales or represent different meanings. The normalization allows getting the Pareto-optimal solutions that represent the Pareto frontier (Vahidinasab et al., 2010). These normalized values can be calculated as in Eq. (10), and they are subsequently used to define the elements of a normalized payoff matrix $\bar{\mathbf{\Phi}}$ in Eq. (11).

$$\bar{f}(\mathrm{x}) = \frac{f_i(x) - f_i^U}{f_i^N - f_i^U}, \quad \mathrm{i}= 1, ..., \mathrm{m} \tag{10}$$

$$= \begin{bmatrix} \overline{f_1} & \cdots & \overline{f_1} & \ldots & \overline{f_1}(x_m^*) \\ \vdots & \ddots & & & \vdots \\ \overline{f_i} & & \overline{f_i} & & \overline{f_i}(x_m^*) \\ \vdots & & & \ddots & \vdots \\ \overline{f_m}(x_1^*) & \ldots & \overline{f_m}(x_i^*) & \ldots & \overline{f_m}(x_m^*) \end{bmatrix} \tag{11}$$

The convex combinations of each row of the normalized payoff matrix $\bar{\mathbf{\Phi}}$ will form the set of points in "*The Convex Hull of Individual Minima*" (CHIM), also called the Utopia line for bi-objective problems. In multi-objective problems with more than two objective functions, the anchor points will form a Utopia hyperplane. Considering the bi-objective problem, the Utopia line is drawn by connecting the two anchor points, that is, points obtained when the *i-th* objective function is minimized independently. Points as a, b and e divide the straight Utopia line (the CHIM) into equal and proportional segments within the normalized space (Jia et al., 2007; Shukla et al., 2007; Vahidinasab et al., 2010; Brito et al., 2014). The anchor points, in addition to defining the ends of the Utopia line, also define the ends of the Pareto Frontier (Fig. 4).

The next step is to find equidistant points in the Utopia line. They can be calculated by using $\bar{\Phi}$ (Eq. (11)), making W_i, with W_i being a convex weight vector W_i, used to obtain a uniformly distributed set of efficient points (Shukla et al., 2007; Jia et al., 2007). Finally, considering $\hat{n}$ the normal direction unit, with a distance D to be maximized, starting from a point $\boldsymbol{W_i}$ on the Utopia line, toward the origin, then $\bar{\mathbf{\Phi}}$ $\boldsymbol{W_i} + D\hat{n}$ is the vector equation of an orthogonal straight line to a surface or plane formed, with $D\ \varepsilon\ R$; $\boldsymbol{W_i}$ being the position vector; $\hat{n}$, a direction vector; and D, a scalar. The point where the line formed intersects the boundary of feasible region closest to the origin will correspond to the maximization of the distance (D) between the Utopia line and the Pareto frontier, and, therefore, will mark a point of optimal solution on Pareto frontier. (Shukla et al., 2007; Utyuzhnikov et al., 2009). Eq. (12), adapted from Jia et al. (2007), shows the maximization of distance (D):

$$\underset{x,t}{\mathrm{Max}}\, t = \mathrm{D} \ \ s.\ t.\ \ W_i + D\hat{n} = \bar{F}(x) \quad h(x) = 0 \quad b \geq x \geq a \tag{12}$$

The constraint $\boldsymbol{W_i} + D\hat{n}$ensures that the point x is mapped toward a point on the normal, while the other constraints guarantee the feasibility of x considering the original problem multi-objective (MOP) (Jia et al., 2007; Ahmadi et al., 2015). Varying $\boldsymbol{W_i}$ will lead to equidistant points on the Pareto frontier, as can be seen in Fig. 4. Hence, the optimization problem described by Eq. (12) can be solved repeatedly for different values of W_i, creating a Pareto frontier evenly distributed with equidistant points (Vahidinasab et al., 2010; Brito et al., 2014). For a bi-objective problem, D can be eliminated of the Eq. (12), originating the simplified Eq. (13), where $\overline{f_1}(x)$ e $\overline{f_2}(x)$ are the normalized $f_1(x)$ e $f_2(x)$ functions (Brito et al., 2014).

$$Min\ \bar{f}(x)\ \ subject\ to: \ \overline{f_1}(x) - \overline{f_2}(x) + 2w - 1 = 0 \quad g_j(x) \geq 0 \quad 0 \leq w \leq 1 \tag{13}$$

The NBI generates several Pareto-optimal solutions, with each solution being originated from a different set of weights $(\boldsymbol{W_i})$, without that a Pareto-optimal solution could be considered better or worse than

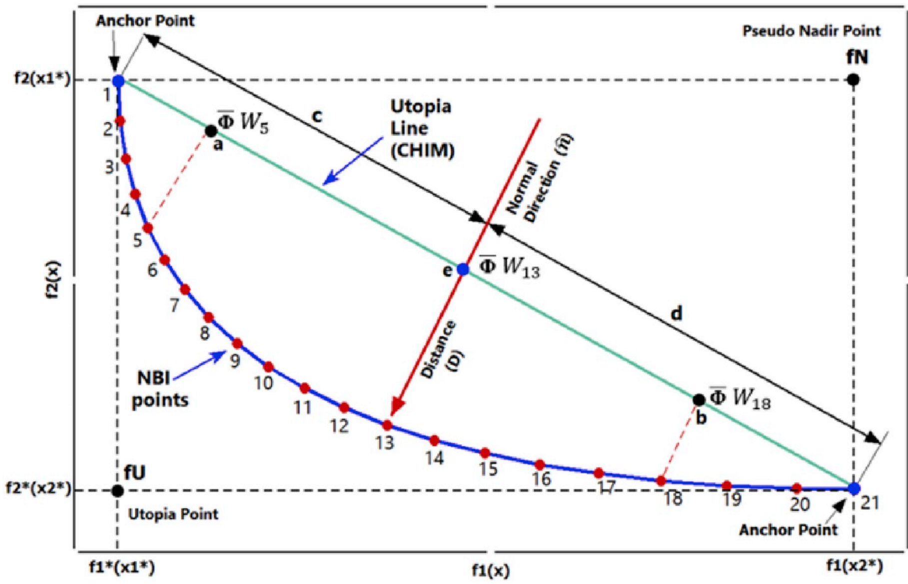

Fig. 4. Graphical description of NBI method.

the other. In order to find the best solution among the various Pareto-optimal solutions it was used Shannon's Entropy Index (S) in combination with Global Percentage Error (GPE). For each one of the v Pareto-optimal solutions obtained in the previous step, it was calculated the respective GPE_{Total}, using Eq. (14), and S_{Total}, using Eq. (15). After, the best set of weights and, consequently, the best solution was defined as the one that presented the highest relation (S_{Total}/GPE_{Total}). Each GPE's Pareto-optimal solutions are calculated defining how far the analyzed point (y_i^*) is from the objective function's ideal value, namely the target (T_i) (Rocha et al., 2015):

$$GPE_{Total} = \sum_{i=1}^{m} \left| \frac{y_i^*}{T_i} - 1 \right| \quad (14)$$

where:

y_i^*– Value of Pareto optimal responses;
T_i – targets defined;
m– Number of objectives.

(S) is used to diversify the weights in the context of multi-objective optimization. In order to diversify the weights belonging to each $\boldsymbol{W_i}$, used to attribute weights to the objective functions in the NBI, (S_{Total}) is calculated according to Eq. (15):

$$S_{Total}(x) = -\sum_{i=1}^{m} w_i \ln(w_i) \quad (15)$$

One can choose as the best point for a given problem to the one in which the relation (S/GPE) has the highest value (Rocha et al., 2015).

Step 8: After the weights of the FA-NBI combination method were obtained using the above strategic, the set of weights were also calculated using some of the more traditional weighting schemes found in the literature. The weights were estimated using regression-based method (RB) by Granger and Ramanathan (1984) and variance-covariance weighting methods by Bates and Granger (1969), extended by Dickson (1973) for more than two methods in combination, considering two procedures: one with covariance (BG/D with COV); other without considering the covariance (BG/D without COV). The results presented by these methods were compared with the FA-NBI in terms of performance and precision, as well as with those presented by these other forecast combination methods, in addition to comparing with those presented by simple average (AVG), in which equal weights are assigned for all methods, and with individual methods used to combine. Bates and Granger (1969) used two procedures - one considering covariance (ρ) between the errors of two individual methods in Eq. (16), and another without considering it (ρ = 0) in Eq. (17) - with the objective of estimating w_1 and, by difference, w_2 weights ($w_2 = 1 - w_1$) to combine two forecasting methods. The method developed by them was intended to give greater weight to the set of forecasts that seemed to contain the lower mean-square errors, minimizing the overall variance.

$$w_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2 - \rho\sigma_1\sigma_2} \quad (16)$$

$$w_1 = \frac{\sigma_2^2}{\sigma_1^2 - \sigma_2^2} \quad (17)$$

In Equations (16) and (17), σ_1^2 and σ_2^2 are the variance of errors for two forecasts, σ_1 and σ_2 are the respective standard deviations and ρ is covariance between the errors of method 1 and 2.

Dickson (1973) extended the results of Bates and Granger (1969) to combinations of n forecasts using the matrix form in Eq. (18), where ($\mathbf{w} = w_1, w_2, \ldots, w_n$) is the vector of n weights, $\boldsymbol{\Sigma}$ is the variance-covariance matrix of forecast errors (n x n), and ($\mathbf{I_n}$) is the ($n \times 1$) matrix with all elements equal to 1.

$$w = \Sigma^{-1} I_n (I_n' \Sigma^{-1} I_n)^{-1} \quad (18)$$

Granger and Ramanathan (1984) proposed three different combining procedures to find the weights minimizing the sum of squared forecast errors using linear regression. The first without restrictions with respect to the weights (Eq. (19)), the second considering the case in which the weights are constrained to worth one (Eq. (20)), and the third has no restrictions on the weights, but a constant term (w_0) is added (Eq. (21)).

$$\hat{y}_t = w_1\hat{y}_t^1 + w_1\hat{y}_t^2 + \varepsilon_t \quad (19)$$

$$\hat{y}_t = w_1\hat{y}_t^1 + w_1\hat{y}_t^2 + \varepsilon_t \quad , \; s.\ t.\ \ w_1 + w_2 = 1 \quad (20)$$

$$\hat{y}_t = w_0 + w_1\hat{y}_t^1 + w_1\hat{y}_t^2 + \varepsilon_t \quad (21)$$

where $\hat{y}_t$ is the combined forecast in period t, $\hat{y}_t^1$ and $\hat{y}_t^2$ are the 1 and 2 methods forecasts in period t, $\mathcal{E}_t$ is the error term, w_1 and w_2 are the weights (coefficients) assigned to method 1 and 2 in the regression model, and wo is the constant term.

In sequence, to verify if there were statistically significant

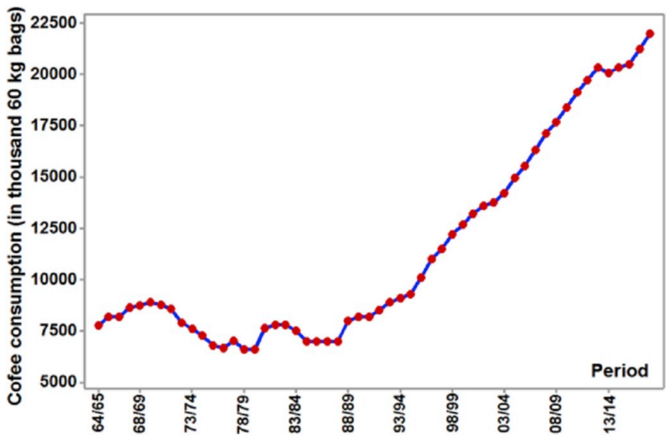

Fig. 5. Y_t series for the evolution of domestic consumption of coffee in Brazil (ICO, 2018).

differences of predictive accuracy between two different competing forecasts methods we used DM test, proposed by Diebold and Mariano (1995). In the present paper, the statistic differences of the predictive accuracy were measured between the combination obtained (FA-NBI) and the individual methods that were combined, as well as between the FA-NBI and the four other weighting methods - AVG, LSE, BG/D with COV and BG/D without COV. For one-step ahead forecasts, the DM test is computed considering the errors series of the two comparison models – (e_1) and (e_2) series. The Appendix D brings more details about DM test.

3. Results and discussion

Thus, in order to find the optimal weights of a combination of time series methods (the FA-NBI combination) to be used to predict coffee demand in the Brazilian domestic market in the periods from 2006/2007 to 2017/2018, the methodology described in section 3 was applied as it follows:

Step 1: The Time Series Plot of the original series of coffee domestic consumption data in Brazil from 1964/1965 to 2017/2018 annual period (Y_t series) is shown in Fig. 5. These data were made available by the ICO (2018). In order to validate the methods of time series to be tested, the consumption of coffee data for the annual periods from 2006/2007 to 2017/2018 were left out from Y_t series' sample to perform the out-of-sample analysis. Thus, the tested methods were fit considering only the fitting data segment between 1964/1965 and 2005/2006, with the Y_t series having 42 observations to perform the in-sample analysis. The residues of each forecast method were calculated from this last period and the prediction errors were calculated from the period left out from Y_t series' sample.

The Y_t series' observations did not follow a normal distribution, as demonstrated by the Anderson-Darling normality test in Fig. 6. Even so, it was not decided to transform it mathematically because, according to Makridakis et al. (1998), such procedure would only be justified if the data showed strong tendency at the same time with a strong seasonal pattern whose variation increases with time. As can be seen in Fig. 5, there is a strong upward trend, but there is no seasonal pattern whose variability grows over time in conjunction with this trend. The Y_t series did not contain missing values or redundant observations, no observation needed to be removed or appended. Additionally, the Dixon's r22 ratio outlier test (Dixon, 1950) was performed, and on the significance level of 0.05 did not detect the presence of suspected outliers in Y_t series (P-Value as 1 and r_{22} statistic as 0.15), taking into account the smallest data value or the largest date, with p-value being 1.

Step 2: Four methods ($z = 4$) were chosen to take part in the weighted combination: Double Exponential Smoothing method (DES), Holt-Winter's multiplicative (WM), and Autoregressive Integrated Moving Average models $ARIMA_{(1,1,1)}$ and $ARIMA_{(2,2,3)}$. The choice of these methods occurred due to the good adjustment to the Y_t series. A brief review of these time series methods is given in Appendix A. For the DES method, the optimal values for the smoothing parameters α and β (computed by fitting an $ARIMA_{(0,2,2)}$ model) were 1.31541 and 0.15515, respectively. Considering the WM method, the optimal values found for α, β and γ (chosen by minimizing the MAE) were 0.75, 0.60 and 0.30, respectively, with length of seasonality 2 ($s = 2$). Regarding the Box-Jenkins methodology (ARIMA), in the data preparation phase, the original Y_t series proved be nonstationary. For this specific application, it was necessary to differentiate Y_t in data preparation and model selection phases. However, the differentiated Y_{t-1} series seemed to follow growing trend from the twenty observations. For this reason, the Y_{t-1} series was differentiated for a second time, resulting in the Y_{t-2} series. As a result, two ARIMA models were selected and specified: the first model, with the use of Y_{t-1} series differentiated once ($d = 1$), and the second model with the Y_{t-2} series differentiated a second time ($d = 2$). The p and q orders of the two selected ARIMA models ($p = 1/q = 1$ for $ARIMA_{(1,1,1)}$, and $p = 2/q = 3$ for $ARIMA_{(2,2,3)}$) were obtained through the Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions and simulations in Minitab®, to obtain models with smaller MAE and VAR of the residuals. Tables 3 and 4 show the calculated coefficients (*Coef*) and the significance test of each parameter.

No correlations were detected in the residues generated for each of the 4 forecast methods, being the correlations considered statistically equal to zero. The P-values of Run-Chart test done over the $ARIMA_{(1,1,1)}$, $ARIMA_{(2,2,3)}$, DES and WM's residues series did not detect trends (0.500; 0.920; 0.809 and 0.159), oscillation (0.500; 0.080; 0.191 and 0.841), mixtures (0.561; 0.168; 0.106 and 0.826) and clustering (0.439; 0.832; 0.894 and 0.826) in their data, since all P-values were greater than 0.050. The Augmented Dickey-Fuller statistic test, with P-Values < 0.05 (0.000; 0.0000; 0.0000 and 0.008) proved that the residues series had no unit root and the process could be considered statistically stationary. Besides that, P-Values for the Anderson-Darling normality test were greater than the chosen significance level of 0.05 for $ARIMA_{(1,1,1)}$ (0.073), $ARIMA_{(2,2,3)}$ (0.344) and DES (0.132) methods, accepting the hypothesis that the residues following a normal distribution. With tests carried out so far, the residues series of theses three methods proved to be Gaussian white noise, with uncorrelated observations, constant variance and normally distributed, which can be verified in the time series plot of residues in Fig. 7. Although WM did not present residues normally distributed, it has been decided to keep it because of the good results found in the statistics tests (see Fig. 8).

Based on the in-sample performance analysis, considering the period from 1964/1965 to 2005/2006 the $ARIMA_{(2,2,3)}$ presented, on average, the best performance and precision (less residues' dispersion) from the $ARIMA_{(1,1,1)}$, DES and WM, as can been observed in Table 5. Moreover, in the in-sample period, the VAR and SD, calculated based on the residues of $ARIMA_{(2,2,3)}$, were lower than in the other methods, which indicated that the selection of this method to forecast future periods was less risky. This means that the decision maker, initially not taking into account the strategy of combining the individual methods, would choose the $ARIMA_{(2,2,3)}$ to forecast coffee demand in the Brazilian market for the subsequent periods (2006/2007 to 2017/2018).

Step 3: A Simplex-Lattice Design {4,5} with four components and lattice degree 7 was created, composed of 4 vertices in 3 (4-1) dimensions, augmented with one center point, four axial points and with one replicate, which resulted in 61 sets of weights ($p = 61$). These 61 sets of weights can be observed in Table C1 in Appendix C. The 61 wt sets were used to combine 42 observations ($g = 42$) of the 4 series of residues (components) for the four selected methods. In this way, these 61 wt sets were used to produce 61 weighted sets of combined residues of the 4 forecast methods chosen, with each observation of each series being calculated using Eq. (1). The geometric description of the experimental region containing the 4 components is shown in Fig. 9a, with each

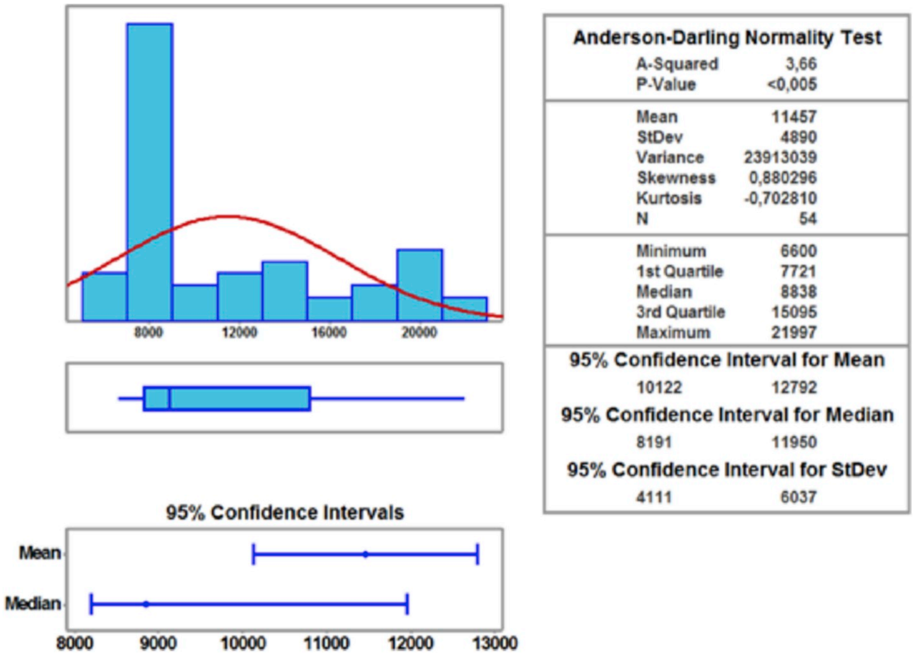

Fig. 6. Summary report for Y_t series.

Table 3
Coefficients and significance test of the estimated parameters of $\text{ARIMA}_{(1,1,1)}$.

Parameter	Coef	SE Coef	T	P
AR 1	0.8999	0.1218	7.3900	0.0000
MA 1	0.5075	0.2089	2.4300	0.0020

Table 4
Coefficients and significance test of the estimated parameters of $\text{ARIMA}_{(2,2,3)}$.

Parameter	Coef	SE Coef	T	P
AR 1	−1.2374	0.2241	−5.5200	0.0000
AR 2	−0.4898	0.2104	−2.3300	0.0260
MA 1	−0.9104	0.2110	−4.3100	0.0000
MA 2	−0.5008	0.2061	−2.4300	0.0200
MA 3	0.7409	0.1461	5.0700	0.0000

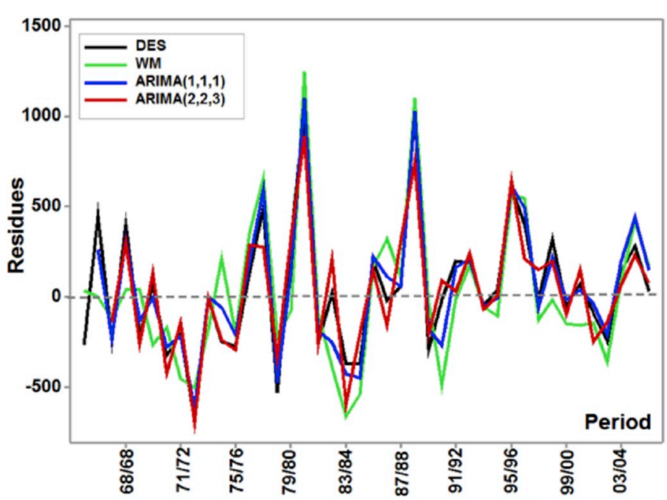

Fig. 7. Residues' time series plot of forecast methods (in thousands 60 kg bags).

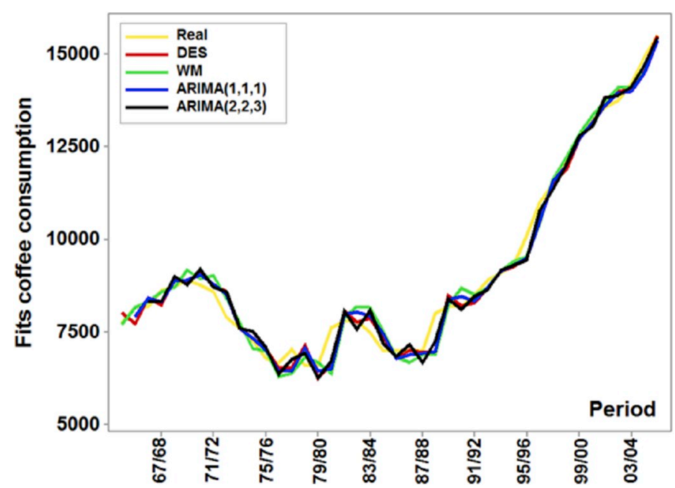

Fig. 8. Real coffee consumption and forecasting methods' fits to historical data (in thousands 60 kg bags).

Table 5
Comparison between individual methods on the in-sample analysis.

Metrics	DES	WM	$\text{ARIMA}_{(1,1,1)}$	$\text{ARIMA}_{(2,2,3)}$
MAE	267.9937	280.3940	260.5738	243.8930
MSE	128,080.9566	153,362.5660	128,743.9337	100,477.9804
RMSE	357.8840	391.6153	358.8090	316.9826
MASE	1.0988	1.1497	1.0684	1.0000
RMSPE	4.4852	4.8891	4.4591	3.9684
MAPE	3.2405	3.3531	3.1188	2.9461
sMAPE	0.0328	0.0340	0.0316	0.0296
U1	0.0188	0.0206	0.0189	0.0167
U2	0.0376	0.0411	0.0377	0.0333
VAR	129,539.4380	156,803.8746	129,075.6035	102,036.0894
SD	359.9159	395.9847	359.2709	319.4309
MdAE	243.6130	173.5543	211.0836	207.2514
MdAPE	2.5630	2.1440	2.3914	2.2063
sMdAPE	0.0256	0.0214	0.0239	0.0221

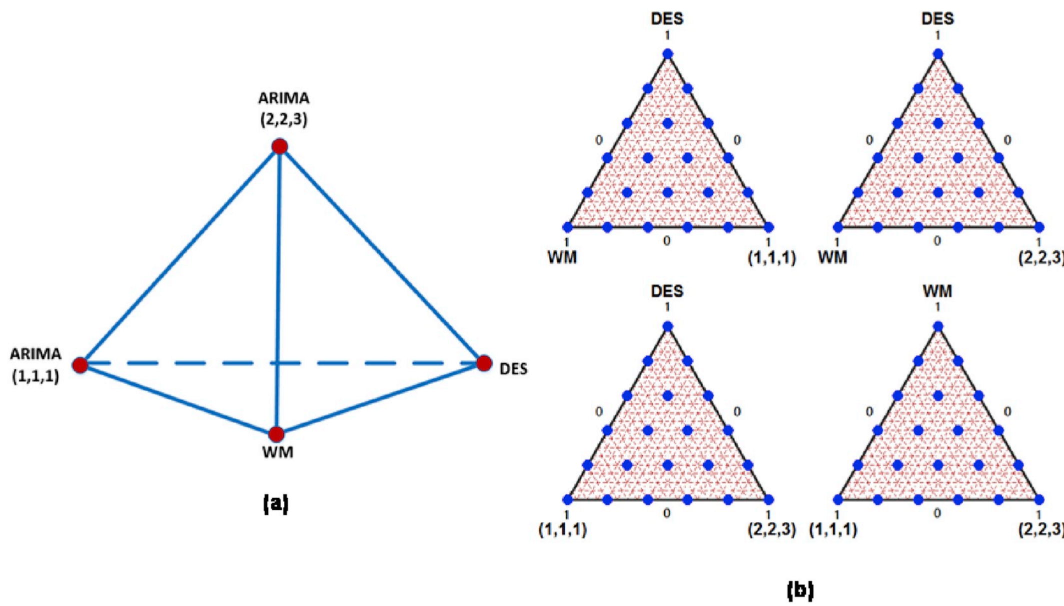

Fig. 9. (a). Experimental region represented by four-component tetrahedron. **(b)**. Simplex-Lattice design {4, 5} on the four plane triangular faces of a tetrahedron.

component representing the vertex of a tetrahedron with four plane triangular faces. The set of the points defined in the simplex lattice region describes the possible proportions of the mixtures to combine the components. The simplex lattice in the four plane triangular faces of the tetrahedron is shown in Fig. 9b.

Step 4: Fourteen ($j = 14$) (MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1, U2, VAR, SD, MdAE, MdAPE and sMdAPE) were calculated for each of the 61 combined residues series. The benchmark method (random walk) chosen was the $ARIMA_{(2,2,3)}$, since it was the individual method that obtained better accuracy and lower variability of the residuals in-sample analysis (Table 5).

Step 5: The PCFA was used to further dimensionality reduction, using principal components. As can be seen in the Table 6, only PC_1 and PC_2 presented eigenvalues greater than one, determining the extraction of only 2 factors ($q = 2$) based on Kaiser's rule (1960).

Table 7 presents the sorted rotated factor analyze model generated with the use of principal components, Varimax rotation as extraction method and based on correlation matrix of the 14 metrics.

The factor analyze model in Table 7 and the loading plot in Fig. 10 showed that MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1, U2, VAR and SD metrics defined Factor 1 because they had high loadings on it and small or negligible loadings on Factor 2. While MdAE, MdAPE and sMdAPE metrics defined Factor 2 because it had high loading on it and small or negligible loadings on Factor 1. The high values of communalities in Table 7 displayed that the two factors together explain a high percentage of variability from each metric. In addition, total factor model explained 97.6% of the total variance of these 14 metrics, which demonstrated that it properly represented them. Because there were many high loadings in Factors 1 and 2, two factor scores series were produced – FS_1 and FS_2. FS_1 series represented MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1, U2, VAR and SD metrics, and FS_2 series represented MdAE, MdAPE and sMdAPE. The dendrogram in Fig. 11 and the loading plot in Fig. 10 confirmed these representations.

Table 6
Correlation matrix's eigenanalysis.

Principal Components	PC_1	PC_2	PC_3	PC_4	PC_5
Eigenvalues	10.438	3.227	0.226	0.102	0.003
Proportion	0.746	0.230	0.016	0.007	0.000
Cumulative	0.746	0.976	0.922	0.999	1.000

Table 7
Factor analyze model.

Function	Loadings		Communalities
Variable	Factor 1	Factor 2	
RMSPE	0.995	−0.069	0.994
MSE	0.993	−0.079	0.993
U1	0.993	−0.064	0.990
U2	0.993	−0.072	0.991
RMSE	0.993	−0.072	0.991
SD	0.992	−0.105	0.994
VAR	0.991	−0.112	0.995
MASE	0.936	0.318	0.976
MAE	0.936	0.318	0.976
sMAPE	0.927	0.341	0.975
MAPE	0.911	0.370	0.967
sMdAPE	0.088	0.977	0.962
MdAPE	0.069	0.971	0.948
MdAE	−0.116	0.948	0.911
Variance	10.366	3.299	13.665
% Var	0.740	0.236	0.976

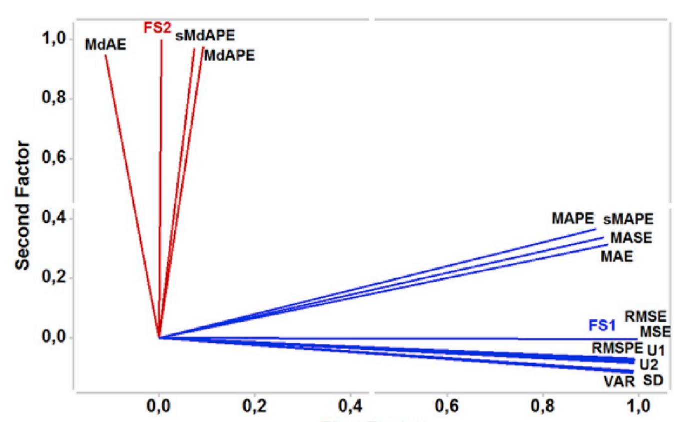

Fig. 10. PCFA loading plot of 14 metrics, FS_1 and FS_2.

The dendrogram, in Fig. 11 was generated through the use of Ward's linkage method with absolute correlation distance measure in cluster analysis to discover natural groupings of the items (or variables) based

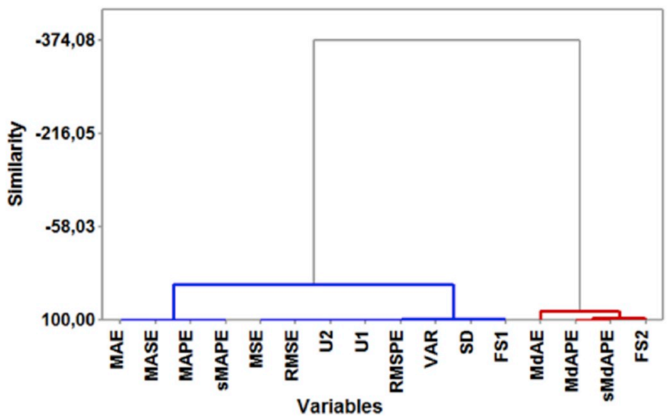

Fig. 11. Dendrogram for similarities of 14 metrics, FS_1 and FS_2.

on minimizing the 'loss of information' from joining two groups taken to be an increase in an error sum of squares criterion (Johnson and Wichern, 2007). The FS_1 and FS_2 series can be viewed into the experimental design in Table C1 and C2 in Appendix C.

As can be observed in Tables 8 and 9, FS_1 and FS_2 showed positive correlation with the metrics that each represented. This fact determined that the objective function to be modeled of FS_1 and FS_2 with the use of DOE-M should be minimized together when applying the NBI routine in Step 7 to calculate Nadir and Utopia points in Payoff matrix (**Φ**).

Step 6: The 61 sets of weights combined with FS_1 and FS_2 series formed the experimental design (shown part in Table C-1 in Appendix C at the end of this article), and, consequently, the designs of the response surfaces ϕFS_1, for FS_1 series, and ϕFS_2, for FS_2 series. Figs. 12 and 14 demonstrate a three-dimensional view that may provide a clearer picture of the response surfaces (3D Surfaces Plots) generated through of experimental design. The exploration of the response surfaces ϕFS_1 and ϕFS_2 over the simplex region design provided the proper polynomial quadratic mixture (functions objectives) $\widehat{FS_1}$ $(\boldsymbol{W})$ (Eq. (22)) and $\widehat{FS_2}$ $(\boldsymbol{W})$ (Eq. (23)) to approximate the surface over the region of interest. In both cases, the model fitting method was a mixture regression with backward eliminator; analyze components in pseudocomponents and quadratic terms. $\widehat{FS_1}$ $(\boldsymbol{W})$ and $\widehat{FS_2}$ $(\boldsymbol{W})$ responses values denote, respectively, the predicted or estimated value of FS_1 and FS_2 responses for a given set of weights $\boldsymbol{W}$ ($w_{(1,1,1)}$, $w_{(2,2,3)}$, $w_{(DES)}$ and $w_{(WM)}$).

$$\begin{aligned}\widehat{FS_1}(\boldsymbol{W}) = {} & 0.984w_{DES} + 3.025w_{WM} + 0.764w_{(1,1,1)} - 1.450\, w_{(2,2,3)}\\ & - 5.288(w_{DES} \times w_{WM}) - 0.889(w_{DES} \times w_{(1,1,1)})\\ & - 1.582(w_{DES} \times w_{(2,2,3)}) - 2.048(w_{WM} \times w_{(1,1,1)})\\ & - 7.403(w_{WM} \times w_{(2,2,3)}) - 3.141(w_{(1,1,1)} \times w_{(2,2,3)})\end{aligned} \tag{22}$$

$$\begin{aligned}\widehat{FS_2}(\boldsymbol{W}) = {} & 1.950\, w_{DES} - 0.864\, w_{WM} + 1.237\, w_{(1,1,1)} + 0.638\, w_{(2,2,3)}\\ & - 3.311\left(w_{DES} \times w_{WM}\right) - 1.564\left(w_{DES} \times w_{(1,1,1)}\right)\\ & - 2.935\left(w_{WM} \times w_{(1,1,1)}\right) - 7.343\left(w_{WM} \times w_{(2,2,3)}\right)\\ & - 2.980\left(w_{(1,1,1)} \times w_{(2,2,3)}\right)\end{aligned} \tag{23}$$

Table 9
Positive correlations between MdAE, MdAPE and sMdAPE with FS_2.

	MdAE	MdAPE	sMdAPE	FS_1
FS_1	−0.116	0.069	0.088	1.000
FS_2	0.948	0.971	0.977	0.000

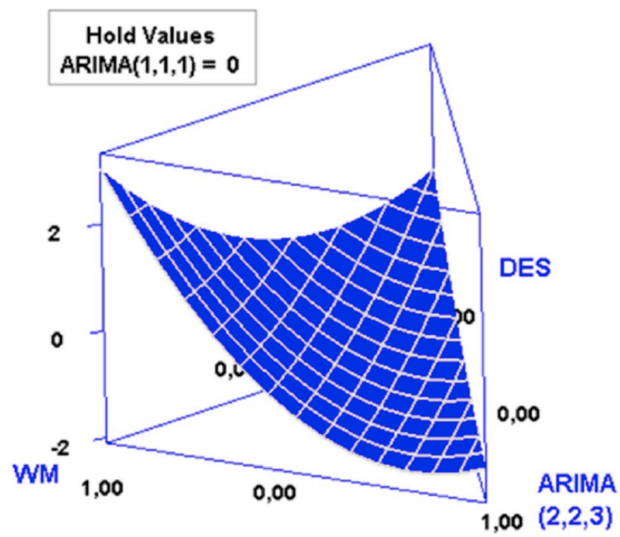

Fig. 12. Mixture surface plot of $\widehat{FS_1}(\boldsymbol{W})$.

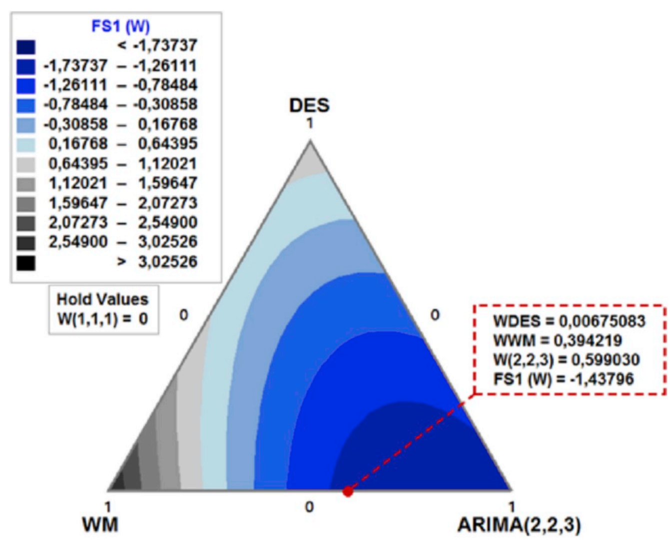

Fig. 13. Mixture contour plot of $\widehat{FS_1}(\boldsymbol{W})$.

Mixture Contour Plots in Figs. 13 and 15 show ranges of intervals containing the values assumed respectively by $\widehat{FS_1}$ $(\boldsymbol{W})$ and $\widehat{FS_2}$ $(\boldsymbol{W})$, as well as the corresponding set of weights at each point of the Figure.

Regarding the adjustment of the models, for $\widehat{FS_1}$ $(\boldsymbol{W})$ function the R^2Adj (adjusted coefficient of multiple determination R^2) was 99.79%, which means that it explained about 99.79% of the variability observed

Table 8
Positive correlations between MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1, U2, VAR and SD with FS_1.

	MAE	MSE	RMSE	MASE	RMSPE	MAPE	sMAPE	U1	U2	VAR	SD
FS_1	0.936	0.993	0.993	0.936	0.995	0.911	0.927	0.993	0.993	0.991	0.992
FS_2	0.318	−0.079	−0.072	0.318	−0.069	0.370	0.341	−0.064	−0.072	−0.112	−0.105

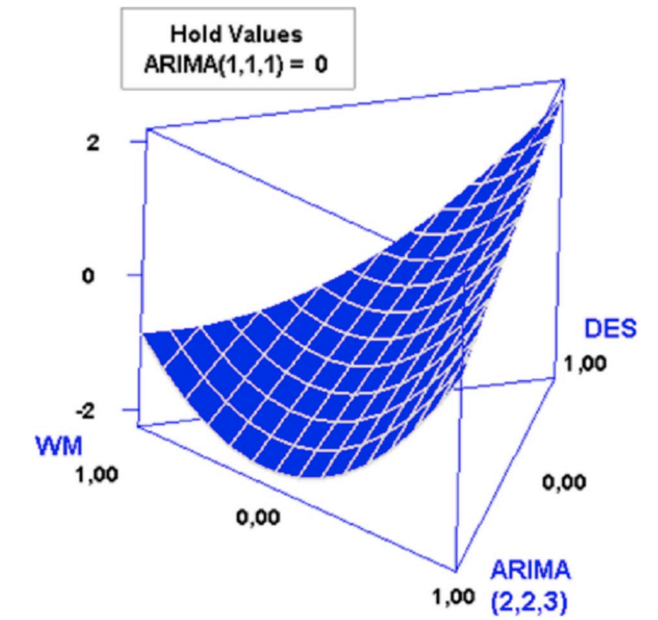

Fig. 14. Mixture surface plot of $\widehat{FS}_2(\boldsymbol{W})$.

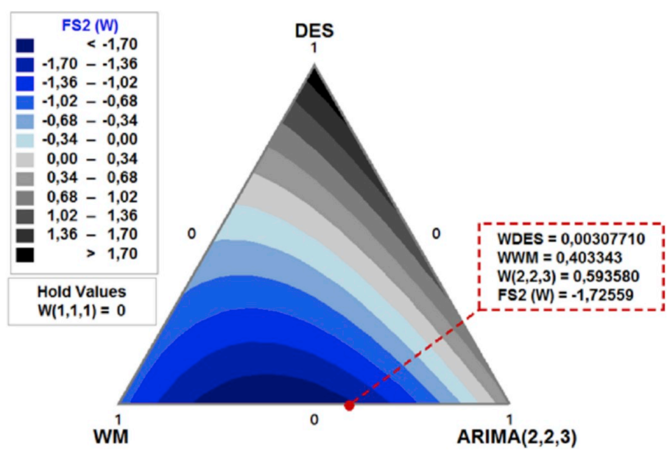

Fig. 15. Mixture contour plot of $\widehat{FS}_2(\boldsymbol{W})$.

in the responses. The $\widehat{FS}_2$ *(W)* function explained about 88.65% of the variability in the responses. In relation to the R-Sq (PRED), statistic that gives some predictive indication capability of the regression model, the $\widehat{FS}_1$ *(W)* explained about 99.72% of the variability in predicting new observations, while $\widehat{FS}_2$ *(W)* explained 85.41%. Therefore, all the models have a satisfactory overall predictive capability to explain a high percentage (99.72% and 85.41%) of the variability in new data. The Run-Chart residues' series did not detect trends (0.055 and 0.050), oscillation (0.975 and 0.950), mixtures (0.740 and 0.740), and clustering (0.260 and 0.260) in their data, since all P-values were greater than 0.050. The Augmented Dickey-Fuller statistic test, with P-values < 0.05 (0.000 and 0.000), proved that the residues series has not unit root, and the process can be considered statistically stationary. Besides that, for $\widehat{FS}_1$ *(W)* and $\widehat{FS}_2$ *(W)* the P-values for the Anderson-Darling test were greater than the chosen significance level of 0.05 (0.898 and 0.829), accepting the hypothesis that the residues following a normal distribution. With tests carried out so far, the residuals series of $\widehat{FS}_1$ *(W)* and $\widehat{FS}_2$ *(W)* proved be Gaussian white noise, with uncorrelated observations, constant variance and normally distributed.

Table 10 shows the ANOVA for the $\widehat{FS}_1(\boldsymbol{W})$ and $\widehat{FS}_2(\boldsymbol{W})$ quadratic models. The P-Values revealed that all terms of $\widehat{FS}_1(\boldsymbol{W})$ were statistically significant at a 5% level. $\widehat{FS}_2(\boldsymbol{W})$ presented one term with P-Value slightly higher than 0.05 (0.096), but it was decided to use $\widehat{FS}_2(\boldsymbol{W})$ because the tested model with the term's exclusion did not show better statistical results.

The trace curves of each component in Cox trace plots in Figs. 16 and 17 presented how the changes' proportions from each of the four components (w_{DES}, w_{WM}, $w_{(1,1,1)}$ and $w_{(2,2,3)}$) in the mixture designs affected, respectively, $\widehat{FS}_1$ *(W)* and $\widehat{FS}_2$ *(W)* estimated responses. In each response trace plot, the intersection of four lines represented the reference blends of ¼. Moving from zero to the right shows the relative proportion of a component increasing and moving from the left to zero shows the relative proportion of a component decreasing, as the other components are held in equal proportions. In Fig. 16 it can be seen that when the proportion of $\text{ARIMA}_{(2,2,3)}$ increases in the mixture (in the combination), the value of $\widehat{FS}_1$ *(W)* decreases, and when the ratio of $\text{ARIMA}_{(1,1,1)}$, DES or WM increases the value of $\widehat{FS}_1$ *(W)* increases. Considering Fig. 17, when the proportion of DES and $\text{ARIMA}_{(1,1,1)}$, increases in the mixture, $\widehat{FS}_2$ *(W)* predicted response also increases, and when the proportion of WM increases, $\widehat{FS}_2$ *(W)* decreases. Initially additions of $\text{ARIMA}_{(2,2,3)}$ in the combination decrease the $\widehat{FS}_2$ *(W)* responses, but from a certain proportion (about 0.2) increase the values thereof.

Step 7: NBI was applied for problem optimization. The Payoff matrix (**Φ**) was calculated, according to Eq. (9), resulting in the matrix shown in Eq. (24), where the main diagonal was formed by the optimal values of the $\widehat{FS}_1$ *(W)* and $\widehat{FS}_2$ *(W)* individual minimization, that is, by the Utopia points. The Nadir points formed the secondary diagonal. $\overline{FS}_1(W)$ and $\overline{FS}_2(W)$ normalized objective functions were then performed as in Eq. (25), using Eq. (10).

$$\Phi = \begin{bmatrix} -1.739 & -0.528 \\ -0.824 & -2.025 \end{bmatrix} \tag{24}$$

$$\bar{\mathbf{F}}(\mathbf{x}) = \begin{bmatrix} \overline{FS}_1(w) = \dfrac{\widehat{FS}_1(w) - (-1.739)}{-0.528 - (-1.739)} \\ \overline{FS}_2(w) = \dfrac{\widehat{FS}_2(w) - (2.025)}{-0.824 - (-2.025)} \end{bmatrix} \tag{25}$$

The NBI minimization procedure was applied for the system of equations as described in Eq. (13), employing the GRG algorithm and using increments of 5% for the weight distribution (w). As though DOE-M was used, the sum of the ($w_{(1,1,1)}$, $w_{(2,2,3)}$, w_{DES}, w_{WM}) variables is a constraint and must be equal to 1. Table 11 presents the results found for the NBI routine, with the normalized values $\overline{FS}_1(W)$ and $\overline{FS}_2(\boldsymbol{W})$, the predicted values $\widehat{FS}_1$ *(W)* and $\widehat{FS}_2$ *(W)*, and the weights assigned for each forecast methods. Fig. 18 display the Pareto frontier obtained with these results.

In the 21 Pareto optimum solutions found, DES and $\text{ARIMA}_{(1,1,1)}$ did not participate in the combination, since they added to the combination (mixture) contributed to increase both the values of $\widehat{FS}_1$ *(W)* and $\widehat{FS}_2$ *(W)* responses, as can be seen in Figs. 16 and 17. It is worth remembering that respectively increasing or decreasing the value of $\widehat{FS}_1$ *(W)* responses also means respectively increasing or decreasing MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1, U2, VAR and SD metrics. Likewise, increasing or decreasing respectively the value of $\widehat{FS}_2$ *(W)* responses also means respectively increasing or decreasing MdAE, MdAPE and sMdAPE median performance metrics. In the Pareto optimal solution of ($v = 1$), in the first line of Table 11, it was given zero as weight to $\widehat{FS}_1$ *(W)* objective function ($w_i = 0$) and one to $\widehat{FS}_2$ *(W)* ($1 - w_i = 1$), i.e. only $\widehat{FS}_2$ *(W)* took part in the joint minimization

Table 10
ANOVAS for $\widehat{FS}_1(W)$ and $\widehat{FS}_2(W)$.

Source	ANOVA for quadratic model of $\widehat{FS}_1(\boldsymbol{W})$					ANOVA for quadratic model of $\widehat{FS}_2(\boldsymbol{W})$				
	DF	Adj SS	Adj MS	F	P-Value	DF	Adj SS	Adj MS	F	P-Value
Regression	9	59.891	6.546	3123.65	0.000	8	54.098	6.762	58.580	0.000
Linear	3	21.598	7.199	3379.34	0.000	3	10.361	3.454	30.430	0.000
Quadratic	6	10.031	1.672	784.72	0.000	5	8.840	1.768	15.580	0.000
$w_{(DES)}$ x $w_{(WM)}$	1	3.629	3.629	1703.35	0.000	1	1.458	1.458	12.840	0.001
$w_{(DES)}$ x $w_{(1,1,1)}$	1	0.103	0.103	48.15	0.000					
$w_{(DES)}$ x $w_{(2,2,3)}$	1	0.325	0.325	152.40	0.000	1	0.325	0.325	2.870	0.096
$w_{(WM)}$ x $w_{(1,1,1)}$	1	0.544	0.544	255.42	0.000	1	1.118	1.118	9.850	0.003
$w_{(WM)}$ x $w_{(2,2,3)}$	1	7.112	7.112	3338.33	0.000	1	7.169	7.169	63.170	0.000
$w_{(1,1,1)}$ x $w_{(2,2,3)}$	1	1.280	1.280	600.84		1	1.181	1.181	10.400	0.002
Residual Error	51	0.109	0.002			52	5.902	0.114		
Error	60	60.000				60	60.000			

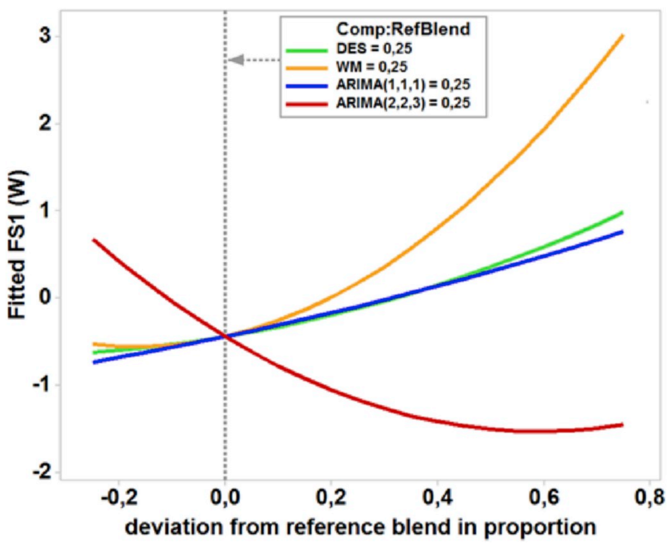

Fig. 16. Cox trace plot of FS_1 ($\boldsymbol{W}$).

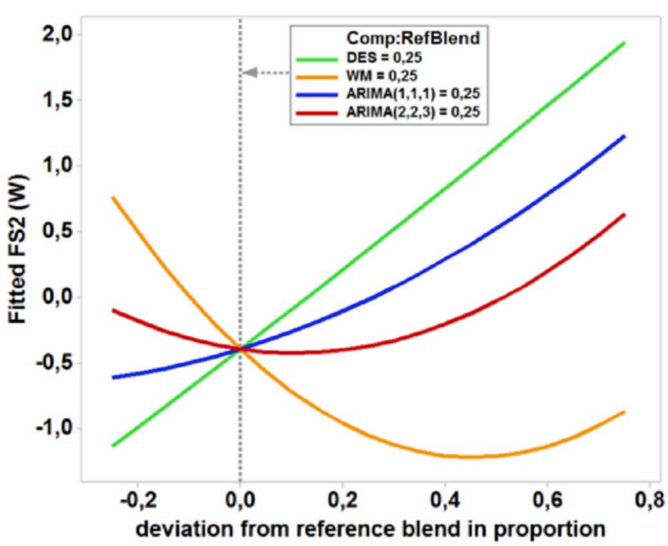

Fig. 17. Cox trace plot of FS_2 ($\boldsymbol{W}$).

process of the NBI. Consequently, to WM was given more weight in the combination because it is the forecasting method that most contributes to the minimization of $\widehat{FS}_2$ ($\boldsymbol{W}$) as it is added to the combination (the mix), which can be observed in Fig. 17. From the optimal solution ($v = 1$) to the other solutions, it can be seen that the more weight (w_i) is attributed to $\widehat{FS}_1$ ($\boldsymbol{W}$) and less weight (1-w_i) to $\widehat{FS}_2$ ($\boldsymbol{W}$) in the NBI process, the weight given to the $ARIMA_{(2,2,3)}$ increases and the WM decreases in the other Pareto optimal solutions found (Table 11 and Fig. 16). This is because it is $ARIMA_{(2,2,3)}$ which further decreases the value of $\widehat{FS}_1$ ($\boldsymbol{W}$) as more as it is being added in the combination, as in Fig. 16.

Fig. 19-a shows that the larger the weight provided to the function $\widehat{FS}_1$ ($\boldsymbol{W}$) in the NBI process, the smaller the $\widehat{FS}_1$ ($\boldsymbol{W}$) responses. And so that $\widehat{FS}_1$ ($\boldsymbol{W}$) responses to be smaller, higher should be the $ARIMA_{(2,2,3)}$ weight in the combination (Fig. 19-b). Taking SD for example, one of the metrics that objective function $\widehat{FS}_1$ ($\boldsymbol{W}$) represents, the smaller the value of $\widehat{FS}_1$ ($\boldsymbol{W}$) due to the increase of $w_{(2,2,3)}$ in the combination, the lower is the SD value (Fig. 19-c). On the other hand, the higher w_i, the higher the values of $\widehat{FS}_2$ ($\boldsymbol{W}$) responses (Fig. 19-d), since the weights of WM in the combination will be smaller (Fig. 19-e). In addition, the smaller the WM weights (w_{WM}), the larger will be, for example, MdMAE, which objective function $\widehat{FS}_2$ ($\boldsymbol{W}$) represents (Fig. 19-f).

To find the best solution among the 21 Pareto-optimal solutions in Table 11, for each one of them was calculated the respective S_{Total}. In this case, each set of $\boldsymbol{W}_i$ weights, with two weights each (w_i and 1-w_i) as in Table 8, were diversified using Eq. (15). In order to calculate the GPE_{total} for each of the 21 Pareto-optimal responses, the $\widehat{FS}_1$ ($\boldsymbol{W}$) was minimized, obtaining the minimum or target value (T_{f1}) of (−1.739). In addition, $\widehat{FS}_2$ ($\boldsymbol{W}$) was minimized, obtaining the minimum or target value (T_{f2}) of (−2.025). Then, taking the values of the $\widehat{FS}_1$ ($\boldsymbol{W}$) and $\widehat{FS}_2$ ($\boldsymbol{W}$) Pareto-optimal responses in Table 11, GPE_{total} was calculated according to Eq. (14). Table 11 shows that the best combination (FA-NBI combination), among the 21 Pareto-optimal solutions, was of $v = 11$, with weights of 0.400 assigned to WM and 0.60 to $ARIMA_{(2,2,3)}$, since that this solution presented the highest relation (S_{Total}/GPE_{Total}) (0.934). The FA-NBI's errors were normally distributed, with P-Value of 0.104 and (K-S) statistic of 0.124 in Kolmogorov-Smirnov normality test. There was also no autocorrelation or partial autocorrelation between them. The Augmented Dickey-Fuller statistic test, with P-Values < 0.05 (0.003) and t-statistic of 5.457 proved that the residues series had no unit root and the process could be considered statistically stationary. The P-Values of Run-Chart test did not detect trends (0.691), oscillation (0.309), mixtures (0.941) and clustering (0.059) in error's series.

The forecast equation by the FA-NBI combination is described in Eq. (26).

$$\hat{y}_{t(NBI-FA)} = 0.40\hat{y}_{tWM} + 0.60\ \hat{y}_{tARIMA_{(2,2,3)}} \tag{26}$$

Table 11
Optimization results and (S/EPG) relation for each Pareto-optimum response.

v	w_i	$1\text{-}w_i$	$\overline{FS}_1(\boldsymbol{W})$	$\overline{FS}_2(\boldsymbol{W})$	$\widehat{FS}_1(\boldsymbol{W})$	$\widehat{FS}_2(\boldsymbol{W})$	$w_{(DES)}$	$w_{(WM)}$	$w_{(1,1,1)}$	$w_{(2,2,3)}$	S_{Total}	EPG $_{(Total)}$	S/EPG
1	0.00	1.00	1.000	0.000	−0.528	−2.025	0.000	0.602	0.000	0.398	0.696	1.728	0.001
2	0.05	0.95	0.903	0.003	−0.646	−2.022	0.000	0.582	0.000	0.418	0.630	1.621	0.137
3	0.10	0.90	0.810	0.010	−0.758	−2.013	0.000	0.562	0.000	0.438	0.570	1.518	0.248
4	0.15	0.85	0.723	0.023	−0.864	−1.998	0.000	0.542	0.000	0.458	0.517	1.419	0.355
5	0.20	0.80	0.640	0.040	−0.964	−1.977	0.000	0.521	0.000	0.479	0.469	1.324	0.463
6	0.25	0.75	0.563	0.063	−1.058	−1.950	0.000	0.501	0.000	0.499	0.429	1.234	0.570
7	0.30	0.70	0.490	0.090	−1.146	−1.917	0.000	0.481	0.000	0.519	0.395	1.151	0.672
8	0.35	0.65	0.423	0.123	−1.228	−1.878	0.000	0.461	0.000	0.539	0.367	1.074	0.766
9	0.40	0.60	0.360	0.160	−1.303	−1.833	0.000	0.440	0.000	0.560	0.346	1.005	0.846
10	0.45	0.55	0.303	0.203	−1.373	−1.782	0.000	0.420	0.000	0.580	0.331	0.946	0.903
11	0.50	0.50	0.250	0.250	−1.437	−1.725	0.000	0.400	0.000	0.600	0.322	0.899	0.934
12	0.55	0.45	0.203	0.303	−1.494	−1.662	0.000	0.380	0.000	0.620	0.320	0.866	0.932
13	0.60	0.40	0.160	0.360	−1.546	−1.593	0.000	0.360	0.000	0.640	0.325	0.838	0.899
14	0.65	0.35	0.123	0.423	−1.591	−1.518	0.000	0.339	0.000	0.661	0.336	0.832	0.837
15	0.70	0.30	0.090	0.490	−1.630	−1.436	0.000	0.319	0.000	0.681	0.353	0.864	0.751
16	0.75	0.25	0.063	0.563	−1.664	−1.349	0.000	0.299	0.000	0.701	0.377	0.899	0.647
17	0.80	0.20	0.040	0.640	−1.691	−1.256	0.000	0.279	0.000	0.721	0.408	0.938	0.533
18	0.85	0.15	0.023	0.723	−1.712	−1.157	0.000	0.258	0.000	0.742	0.444	0.982	0.413
19	0.90	0.10	0.010	0.810	−1.727	−1.052	0.000	0.238	0.000	0.762	0.488	1.035	0.290
20	0.95	0.05	0.003	0.903	−1.736	−0.941	0.000	0.218	0.000	0.782	0.537	1.108	0.160
21	1.00	0.00	0.000	1.000	−1.739	−0.824	0.000	0.198	0.000	0.802	0.593	1.210	0.001

The overlaid contour plots (Fig. 20) displays that the optimal solution found, represented in the line 11 of the Table 11 (v = 11), lies within the feasible region, since the restrictions referring to $\widehat{FS}_1$ *(w)* and $\widehat{FS}_2$ *(w)* were all satisfied.

It should be noted that in addition to the solution of v = 11 (the best FA-NBI solution found) all the other 20 points of the Pareto frontier also provided optimal solutions. The simultaneous multiple objectives' optimization point of v = 11 was the best solution considering the highest relation (S_{Total}/GPE_{Total}) criterion. This is what made the optimal solution of v = 11 the best solution to the problem, that is, to forecast coffee demand in Brazil using FA-NBI combination method.

To illustrate this statement, the Fig. 21 shows the values of MAE, RMSE, U2, MdMAE, sMdAPE and SD in each of the 21 Pareto-optimal solutions, where each of these cannot be considered better or worse than the others. In addition, it also shows the increase or decreases in the values assumed by these metrics when moving from one Pareto-optimal solution towards another. MASE, MAPE and sMAPE metrics do not appear in the Fig. 21 because they presented the same evolution of MAE. For the same reason, MSE and RMSPE were also omitted because they were represented by RMSE. Similarly, U1 was represented by U2, VAR was by SD, and MdAPE by sMdAPE. It is clearly seen that the vector of feasible solutions (v = 11) is Pareto-optimal, since there is no feasible point which can reduce any of the objective functions (metrics) without causing the simultaneous increase of, at least, another objective function. As an example, it is noticed that it is not possible to reach a solution in which SD or VAR is smaller (Fig. 21-f) without increasing the RMSE value (Fig. 21-b) and, consequently, those of MSE and RMSPE. Likewise, it was not possible to reduce SD or VAR without increasing the value of U2 (Fig. 21-c) and therefore U1, which is represented by U2 in Fig. 21.

Step 8: In out-of-sample forecasting analysis (cross-validation) was done a comparative analysis between the made forecasts and the actual coffee's consumption that took place for the periods 2006/2007 to 2017/2018. It was FA-NBI combination that got, on average, better performance (accuracy) than DES, WM, $ARIMA_{(1,1,1)}$ and $ARIMA_{(2,2,3)}$,

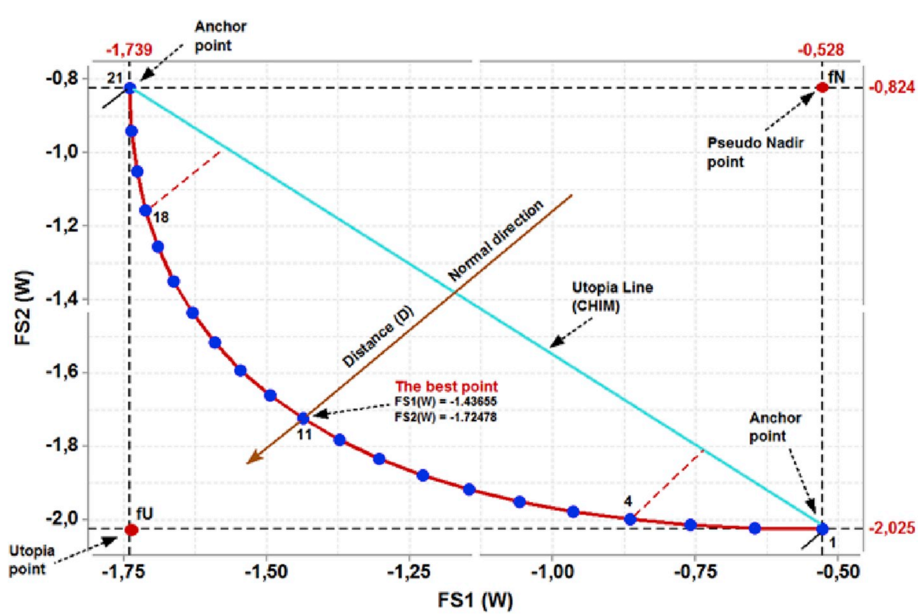

Fig. 18. Pareto frontier.

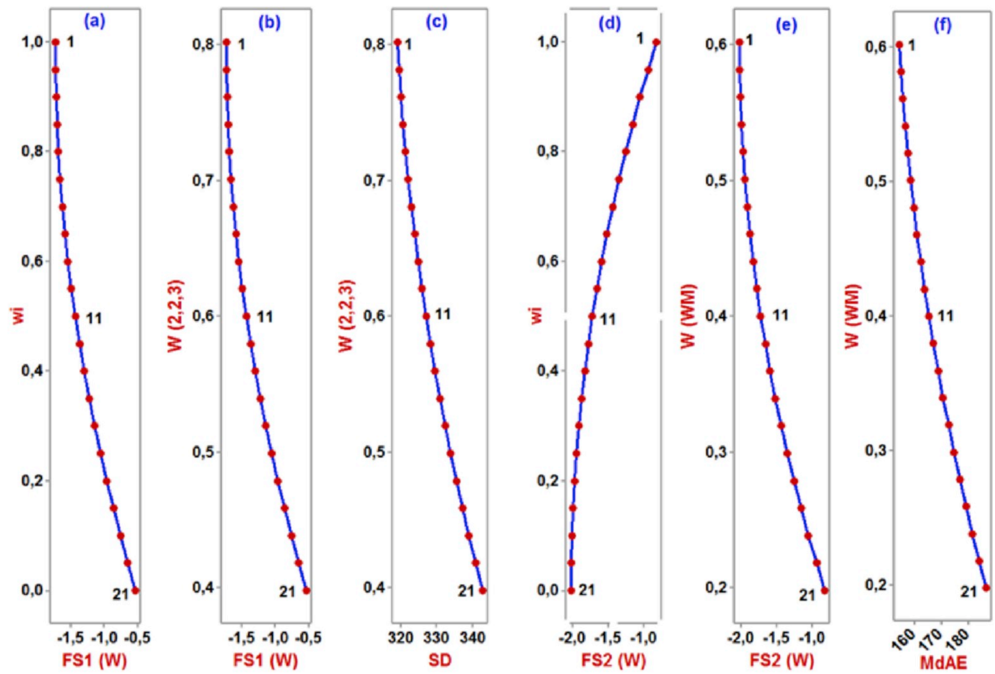

Fig. 19. Relationship between the weights of $\widehat{FS_1}$ *(W)* and $\widehat{FS_2}$ *(W) in NBI process* and the weights' combination methods.

considering MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1, U2, MdAE, MdAPE and sMdAPE performance metrics (Table 12). In addition, FA-NBI also presented the lowest VAR and SD of the forecast errors than WM and $ARIMA_{(1,1,1)}$, but larger than DES and $ARIMA_{(2,2,3)}$. As can be seen in Fig. 20, the decrease of SD and VAR precision metrics would only be possible with the increase of MSE, RMSE, RMSPE, U1 and U2 performance metrics. That is, to improve the precision one has to sacrifice performance and vice versa.

The performance and precision of FA-NBI combination method were compared with those of other weighting methods - AVG, RB, BG/D with COV and BG/D without COV. Considering RB, the weights were estimated by linear regression and constrained to sum 1, without taking into account the constant term (Eq. (20)). In order to calculate the weights using BG/D with COV in Eq. (18) was used a complete matrix **Σ** containing the error variances of each of the four methods on the main diagonal and the covariance between the errors of these methods in the other terms. Differently, to calculate the weights using BG/D without COV, only the variances were considered in main diagonal matrix **Σ**, being the covariance between the errors series equal to zero. The weights for AVG, RB, BG/D with COV and BG/D without COV are in the Table 13. Table 12 indicates that FA-NBI presented better performance than all other weighting methods with respect to MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1 and U2 mean metrics, but larger MdAE, MdAPE and sMdAPE median metrics than BG/D without COV.

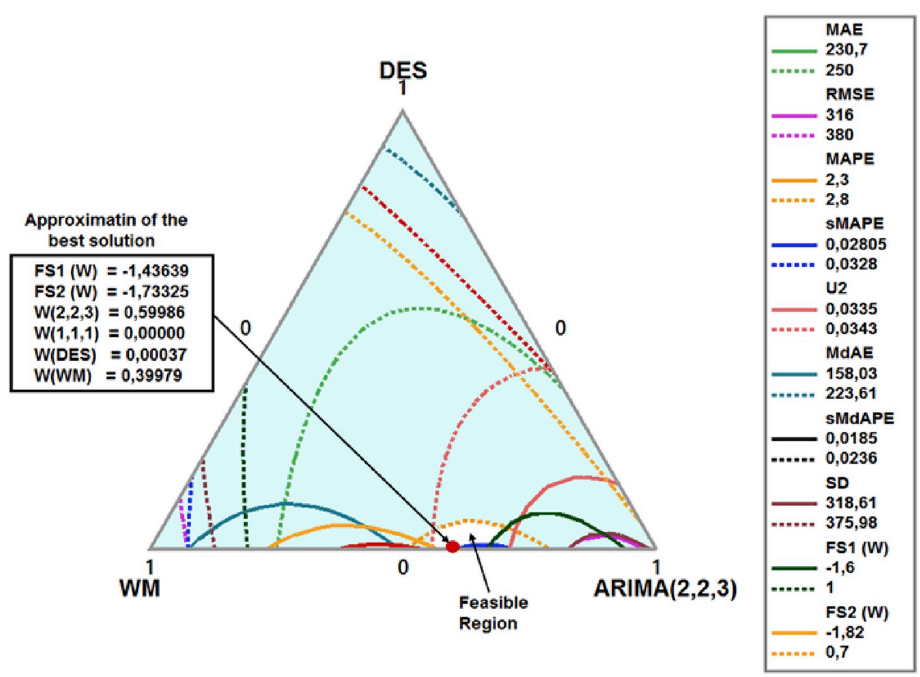

Fig. 20. Overlaid contour plot of FS_1 and FS_2.

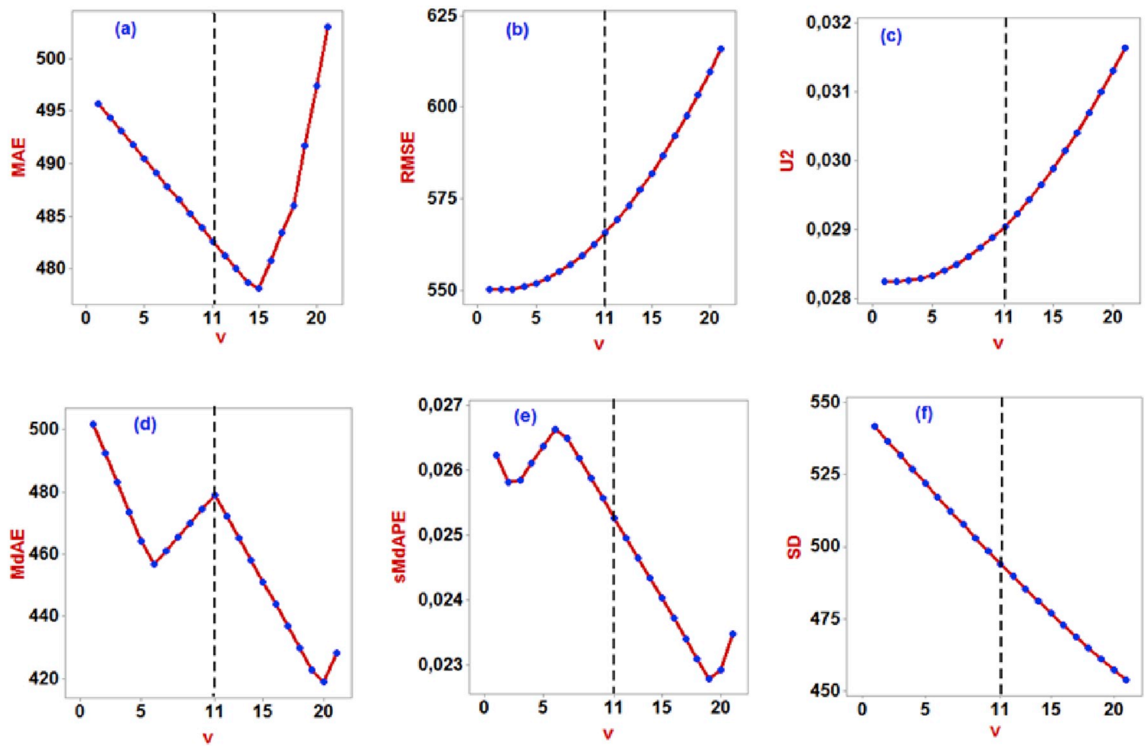

Fig. 21. Variations between metrics values in each optimal solution of the Pareto frontier.

In terms of precision, FA-NBI presented higher SD and VAR than AVG, RB and BG/D without COV.

The DM tests results in Table 14, based on the absolute and squared-error loss, indicated that there is significant statistic difference in relation to the prediction performance (accuracy) between the proposed method FA-NBI and DES, WM, $ARIMA_{(1,1,1)}$, ARIMA $_{(2,2,3)}$, AVG, RB and BG/D with and without COV compared methods. Since DM statistics values for the comparisons between FA-NBI with individual and weighting methods were greater than Z-value, the DM tests' H_0 hypothesis (both forecasts have the same accuracy) were rejected at the 5% level of significance (presenting all P-values < 0.05). Thus, the alternative hypotheses were accepted, proving that the forecasts do not have the same accuracy.

In order to test the FA-NBI's applicability and feasibility in relation to other series of different coffee consumption series' characteristics (Fig. 5), which has a positive upward trend but without seasonality, two tests were performed, one making use of real series with trend and variability, and another with simulated series. In consequence it was examined the capability of FA-NBI for other linear time series using both simulated and real data, proving that FA-NBI is quite competent in modeling and forecasting linear time series in a variety of situations. Each combination using FA-NBI was composed of three time-series methods chosen using the MINITAB software package. The 25 sets of weights to make the residues' combinations came out of a simplex-

Table 13
Weights' set for weighting methods for consumption.

Weighted Methods	DES	WM	$ARIMA_{(1,1,1)}$	$ARIMA_{(2,2,3)}$
FA-NBI	0.0000	0.4000	0.0000	0.6000
AVG	0.2500	0.2500	0.2500	0.2500
RB	0.0000	0.0000	0.0000	1.0026
BG/D with COV	0.5832	0.6808	−1.1493	0.8853
BG/D without COV	0.2491	0.2058	0.2441	0.3010

Table 12
Comparison between individual and other weighting methods with FA-NBI in coffee's consumption out-of-sample analysis.

	FA-NBI	DES	WM	ARIMA (1,1,1)	ARIMA (2,2,3)	AVG	RB	BG/D with COV	BG/D without COV
MAE	482.60	731.10	568.33	1737.98	577.32	741.08	537.63	1237.15	761.60
MSE	319,959	680,278	392,122	3,631,493	478,496	674,916	427,733	3,270,028	706,442
RMSE	565.65	824.79	626.20	1905.65	691.73	821.53	654.01	1808.32	840.50
MASE	1.000	1.515	1.178	3.602	1.196	1.536	1.114	2.564	1.578
RMSPE	2.934	4.239	3.089	9.324	3.583	4.188	3.385	8.592	4.280
MAPE	2.516	3.775	2.861	8.675	3.008	3.802	2.801	5.973	3.904
sMAPE	0.026	0.039	0.028	0.091	0.031	0.039	0.029	0.056	0.040
U1	0.015	0.022	0.016	0.051	0.018	0.022	0.017	0.045	0.022
U2	0.029	0.042	0.032	0.098	0.036	0.042	0.034	0.093	0.043
VAR	244,064	159,027	423,103	666,473	179,119	137,152	180,540	2,200,563	137,904
SD	494.0	398.8	650.5	816.4	423.2	370.3	424.9	1483.4	371.4
MdMAE	479.0	677.7	555.8	1949.4	521.0	656.8	474.5	377.7	674.4
MdAPE	2.522	3.482	2.785	9.605	2.835	3.478	2.585	1.999	3.569
sMdAPE	0.025	0.035	0.028	0.101	0.029	0.035	0.026	0.020	0.036

Table 14
DM tests between FA-NBI and individual and weighting methods.

Methods	Absolute error		Squared error	
	Statistic	P-Value	Statistic	P-Value
DES	−6.1980	0.0001	−3.8738	0.0026
WM	−5.8796	0.0001	−3.7761	0.0031
$ARIMA_{(1,1,1)}$	−4.6606	0.0007	−3.0154	0.0118
$ARIMA_{(2,2,3)}$	−5.8796	0.0001	−3.7761	0.0031
AVG	−4.8867	0.0005	−3.1261	0.0096
RB	−5.1569	0.0003	−3.4380	0.0055
BG/D with COV	−4.4876	0.0009	−2.9317	0.0136
BG/D without COV	−4.9470	0.0004	−3.1615	0.0091

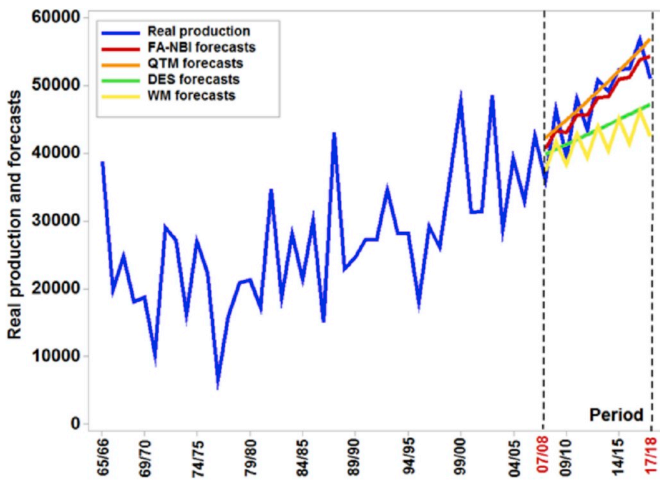

Fig. 22. Real production series and forecasts.

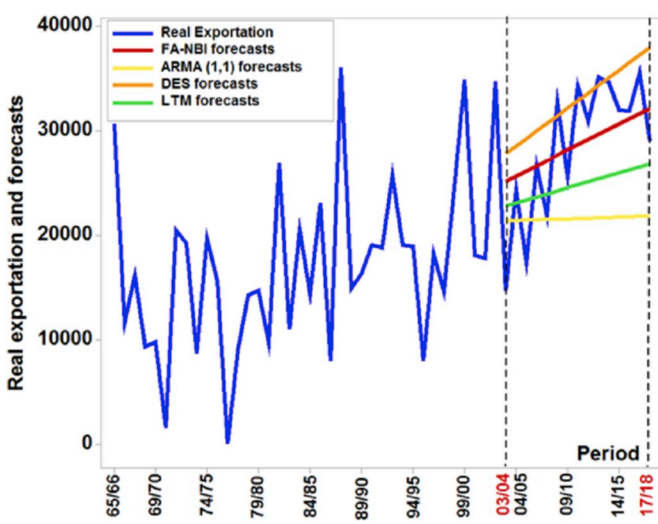

Fig. 23. Real exportation series and forecasts.

Table 16
Weights' set for weighting methods for coffee's production and exportation series.

Methods	Production			Exportation		
Weighted/ Individual	QTM	DES	WM	ARMA (1,1)	DES	LTM
FA-NBI	0.6966	0.0733	0.2301	0.0000	0.4721	0.5279
AVG	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
RB	1.4451	−0.7261	0.2700	−0.1741	0.0557	1.1084
BG/D with COV	1.4494	−0.7170	0.2675	−0.1526	0.0528	1.0998
BG/D without COV	0.4225	0.3260	0.2515	0.3369	0.2829	0.3801

lattice {3.5}. The rest of the procedures were identical to find the combination to predict the demand for coffee.

In the first test based on real series, the actual values of the coffee's Brazilian production from 1965/66 to 2006/07 (Real production in Fig. 22), made available by ICO (2018), were used to forecast production for 2007/08 to 2017/18 periods using FA-NBI. The coffee's production series data present trend and variability with one year increasing and other decreasing. The other real series used was the coffee's Brazilian exportation from 1965/66 to 2002/03 (Real exportation in Fig. 23), also available from ICO (2018), to forecast from 2003/04 to 2017/18 periods. The exportation series follows the production series in positive trend and variability. In this cases, the three individual methods selected in the MINITAB package to be combined are in Table 15.

Table 16 brings the weights' set for FA-NBI combination and for other weighting methods for production and exportation series.

Figs. 22 and 23 display the forecasts of these three methods and FA-NBI compared to the actual values. Table 17 compares the FA-NBI's performance and precision with the individual forecasting methods and with weighting methods in out-of-sample analysis, concerning production and exportation series. All the complete numerical comparison's Tables are in supplementary materials.

In Tables 12 and 17, it was observed that FA-NBI presented acceptable results in terms of performance and precision when applied in series with trend and variability (non-stationary time series). It should be noticed that Table 2 shows the main metrics used in the literature to make comparisons between forecasting methods. It can be observed that the most used metrics are MAE, MSE, RMSE, MAPE, sMAPE and MASE. Considering these six metrics, FA-NBI, in Table 16, overcomes all the individual methods used to combine and all the methods to find weights' combinations. Fig. 24 demonstrates a tendency for Brazilian coffee's production to increase more than domestic and external coffee's demand based on FA-NBI forecasts.

In the second applicability and viability test, Figs. 25–30 show a collection of linear time series implemented and simulated to test FA-NBI method in this present study. In each case, the errors ε_t: N (0,1) are assumed to be random variables independent and identically distributed (IID). These six time series were chosen to represent a variety of problems that have time series with different characteristics. The time series in Figs. 25–28, taken from Montgomery et al. (2008), have

Table 15
The three individual selected forecasting methods for production and exportation series.

Series	Selected Individual methods		
Production	QTM* with fitted trend equation of ($Y_t = 23885 - 411t + 19.51t^2$)	DES (α and β of 0.291522 and 0.106042 respectively).	WM (α, β and γ of 0.31, 0.31 and 0.23 respectively, with length of seasonality 2).
Exportation	$ARMA_{(1,1)}$	DES (α and β of 0.432264 and 0.092629 respectively).	LTM** with fitted linear trend equation of ($Y_t = 11616 + 287t$)

QTM*: Quadratic Trend Model.
LTM**: Linear Trend Model.

Table 17
Comparison between individual and other weighting methods with FA-NBI in coffee's production and exportation out-of-sample analysis.

Series	Means' perforfance metrics	Median's performance metrics	Precisions' metrics
Production	Better than all individual and weighting methods.	Better than DES and WM.	Better than all individual methods, AVG and BG/D without COV.
		Better than all weighting methods. Worse than QTM.	Worse than RB and BG/D with COV.
Exportation	Better than all individual and weighting methods, except LTM, RB and BG/D with COV in RMSPE metric.	Better than all individual methods, except DES.	Better than all individual methods, except DES.
		Better than all weighted methods.	Better than all weighted methods.

Mean's performance metrics: MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1 and U2.
Median's performance metrics: MdAE, MdAPE and sMdAPE.
Precision's metrics: SD and VAR.
Weighting's methods: AVG, RB, BD/D with COV and BG/D without COV.

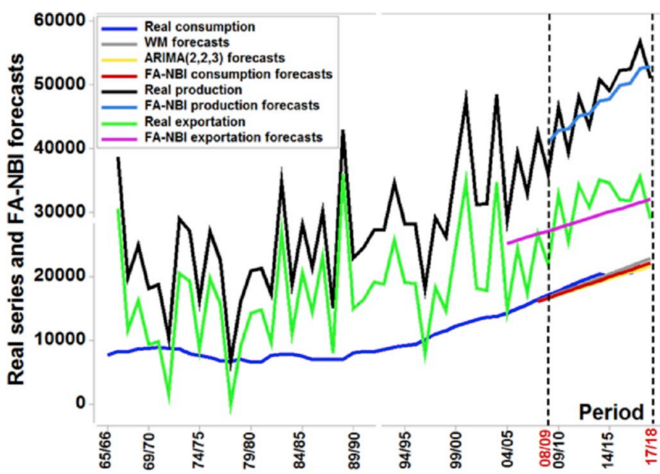

Fig. 24. Real series of coffee's consumption, production and exportation with respective FA-NBI forecasts.

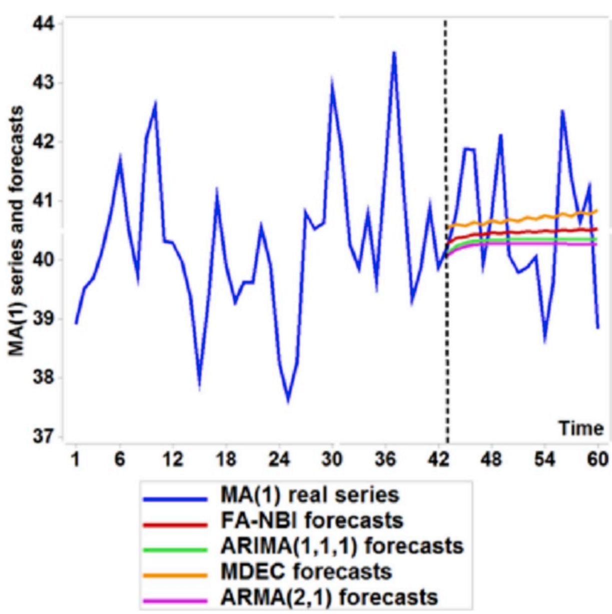

Fig. 25. Simulated time series describing the first order moving average MA (1), with equation of ($Y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$).

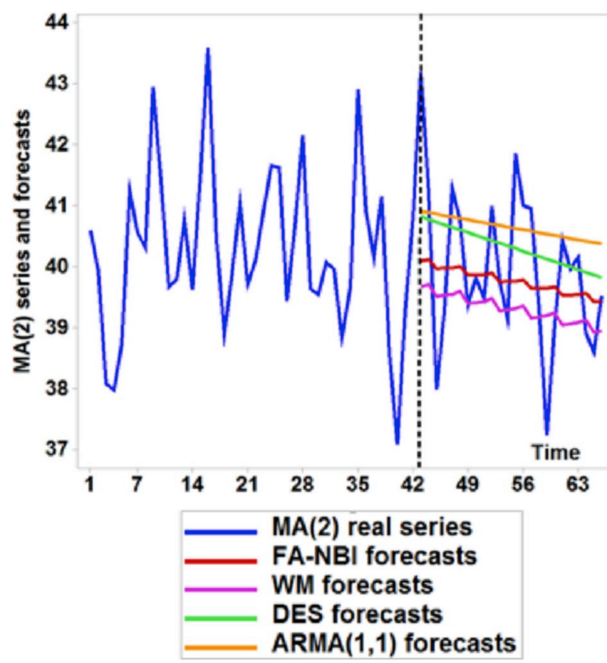

Fig. 26. Simulated time series describing the second-order moving average MA (2), with equation of ($Y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$).

pure autoregressive (AR) or pure moving average (MA) correlation structures, representing stationary process. The MA series reveal that the mean and variance remain stable while there are some short runs where successive observations tend to follow each other for very brief durations, suggesting that there is indeed some positive autocorrelation in the data. It can be observed, therefore, some short runs during which observations tend to move upward or downward. As opposed to the MA series, however, the duration of these runs tends to be longer and the trend tends to linger in AR series (Montgomery et al., 2008). FA-NBI was also tested in time series with mixed AR and MA components - ARMA (1,1) process in Fig. 29, and with seasonal time series – SARIMA (1,1,1) $(1,1,1)_{12}$ process with length of seasonality 12 (Fig. 30), both equally available in Montgomery et al. (2008).

The three individual selected forecasting methods for each of the six simulated series are in Table 18.

The FA-NBI combination weights for MA (1), MA (2) and AR (1) series are in Table 19, and for AR (2), ARMA (1,1) and SARIMA (1,1,1) $(1,1,1)_{12}$ in Table 20.

As noticed in Table 21 that FA-NBI also obtained, on average, good performance and precision in the comparisons. In results' view presented by FA-NBI in Tables 12, 16 and 21 it has been demonstrated that it can be applied successfully to find the time series combination's weights in series with different characteristics.

The complete tables bringing the comparisons of FA-NBI with the other methods, in view of their application in simulated series, are in supplementary materials.

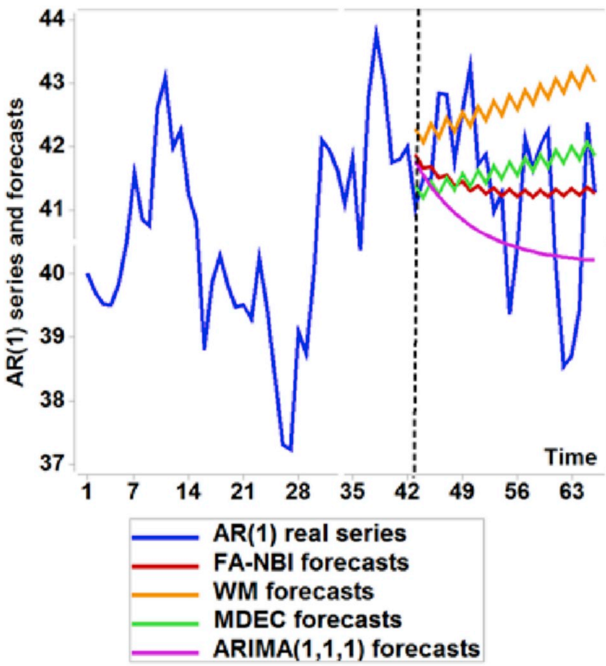

Fig. 27. Simulated time series describing the first-order autoregressive process AR (1), with equation of ($Y_t = 8 + 0.8Y_{t-1} + \varepsilon_t$).

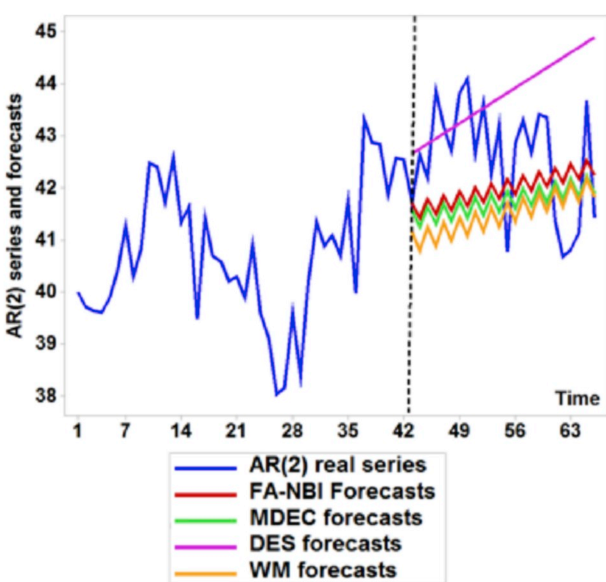

Fig. 28. Simulated time series describing the second-order autoregressive process AR (2) with equation ($Y_t = 8 + 0.4Y_{t-1} + 0.5Y_{t-2} + \varepsilon_t$).

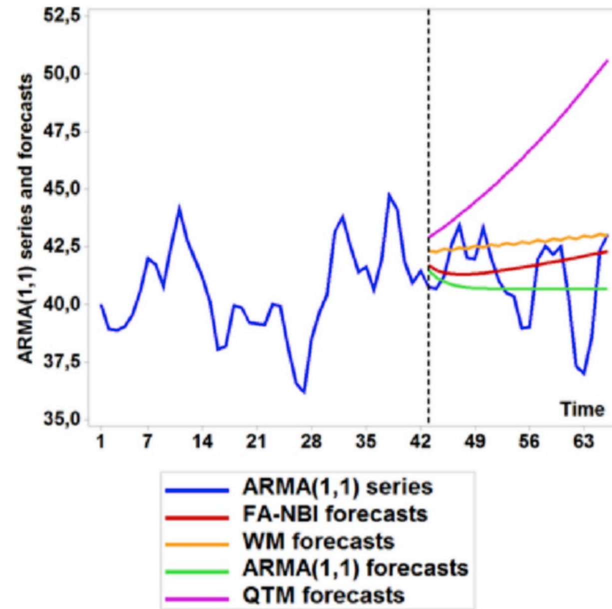

Fig. 29. Simulated time series describing mixed autoregressive – moving average process, ARMA (1,1), with equation ($Y_t = 16 + 0.6Y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1})$).

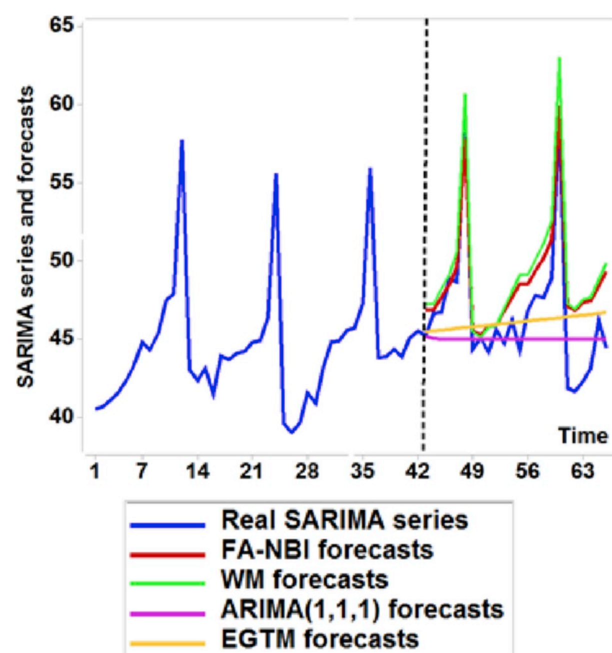

Fig. 30. Simulated time series describing SARIMA (1,1,1) $(1,1,1)_{12}$ process.

4. Conclusions

All the steps for the application of FA-NBI combination method were developed within a case study whose objective was, besides demonstrating the new method's applicability, to forecast the domestic Brazilian coffee's consumption. Thus, in the context of time series methods' combination, PCFA was used to reduce the number of metrics to be optimized at the same time from 14 to 2. These two metrics - the FS_1 and FS_2 factors scores metrics - represented, with minimal loss of information, the 14 original metrics of performance and precision, which allowed us to find the weights' combination not only minimizing one or two metrics, but the 14 metrics at once. Another important point was to use DOE-M to model the FS_1 and FS_2 objective functions, generating $\widehat{FS_1}$ *(w)* and $\widehat{FS_2}$ *(w)* mathematical objective functions representative of the 14 original metrics, without the same joint optimization would not be possible. After, $\widehat{FS_1}$ *(w)* and $\widehat{FS_2}$ *(w)* were simultaneously optimized using the NBI multi-objective optimization routine, being the $\widehat{FS_1}$ *(w)* and $\widehat{FS_2}$ *(w)* joint optimization in reality represented the joint optimization of the 14 original metrics represented by them. Thus, each of the 21 Pareto-optimal solutions found using NBI and, consequently, each of the 21 sets of optimal weights found represented a point at the Pareto frontier in which it would not be possible to reduce the value of one of the 14 metrics without causing the simultaneous increase of, at least, another metric. To find the best among the 21 sets of optimal weights – the FA-NBI combination, a highlight of the present study was to use the relation (S_{Total}/GPE_{Total}) in

Table 18
The three individual selected forecasting methods for each of the six simulated series.

Series	Selected Individual methods		
MA (1)	$ARIMA_{(1,1,1)}$	MDEC* with ($Y_t = 39.947 + 0.0145t$) equation and seasonal length of 2.	$ARMA_{(2,1)}$
MA (2)	WM (α, β and γ of 0.47, 0.16 and 0.14 respectively, with length of seasonality 2).	DES (α and β of 1.02818 and 0.03947 respectively).	$ARMA_{(1,1)}$
AR (1)	WM (α, β and γ of 0.50, 0.05 and 0.18 respectively, with length of seasonality 2).	MDEC with ($Y_t = 39.947 + 0.0145t$) equation and seasonal length of 2.	$ARIMA_{(1,1,1)}$
AR (2)	MDEC* with ($Y_t = 40.126 + 0.0296t$) equation and seasonal length of 2.	DES (α and β of 0.689070 and 0.060339 respectively).	WM (α, β and γ of 0.05, 0.008 and 0.04 respectively, with length of seasonality 2).
ARMA (1,1)	WM (α, β and γ of 0.27, 0.20 and 0.20 respectively, with length of seasonality 2).	$ARMA_{(1,1)}$	QTM with ($Y_t = 41.145 - 0.1523t + 0.00448t^2$) equation.
SARIMA (1,1,1) $(1,1,1)_{12}$	WM (α, β and γ of 0.35, 0.35 and 0.10 respectively, with length of seasonality 12)	$ARIMA_{(1,1,1)}$	EGTM** with $Y_t = 43.2652 \times (1.00116^t)$ equation

MDEC*: Multiplicative Decomposition Model.
EGTM**: Exponential Growth Trend Model.

Table 19
Weights' set for weighting methods in MA (1), MA (2) and AR (1) series.

Weighted Methods	MA (1) series		MA (2) series			AR (1) series			
	ARIMA (1,1,1)	MDEC	ARMA (2,1)	WM	DES	ARMA (1,1)	WM	MDEC	ARIMA (1,1,1)
FA-NBI	0.634	0.366	0.000	0.655	0.021	0.325	0.303	0.121	0.576
AVG	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333
RB	−15.064	1.103	14.965	0.492	−0.020	0.532	−0.110	0.051	1.059
BG/D with COV	−13.634	1.032	13.602	0.492	−0.023	0.531	−0.119	0.051	1.068
BG/D without COV	0.357	0.285	0.358	0.314	0.348	0.338	0.339	0.184	0.476

Table 20
Weights' set for weighting methods in AR (2), ARMA (1,1) and SARIMA (1,1,1) (1,1,1) series.

Weighted Methods	AR (2) series			ARMA (1,1) series			SARIMA (1,1,1) (1,1,1) series		
	MDEC	DES	WM	WM	ARMA (1,1)	QTM	WM	ARIMA (1,1,1)	EGTM
FA-NBI	0.882	0.118	0.000	0.000	0.836	0.164	0.300	0.700	0.000
AVG	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.333
RB	0.833	0.746	−0.592	−0.027	0.969	0.058	0.899	−0.120	0.217
BG/D with COV	0.843	0.746	−0.589	−0.027	0.969	0.058	0.085	0.454	0.461
BG/D without COV	0.281	0.446	0.273	0.165	0.656	0.179	0.075	0.449	0.477

making decision about the choice of the best solution to forecast coffee's demand. Based on the out-of-sample analysis in Table 12, FA-NBI presented better means's performance metrics (MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1 and U2) and better medians's performance metrics (MdAPE, MdAPE and sMdAPE) than all the individual methods chosen to combine - DES, WM, $ARIMA_{(1,1,1)}$ and $ARIMA_{(2,2,3)}$. In comparison with the other weighting methods, it obtained better performance in means' metrics than all - AVG, RB, BG/D with and without COV, only being surpassed in medians' metrics by RB and BG/D with COV. In addition to good performance in numerical terms presented by FA-NBI, it was proved in Table 14 that these gains were statistically significant based on the DM performed tests. The DM tests proved that the FA-NBI exhibited a significant statistic improvement of accuracy (performance) compared to individual and weighting mentioned methods. In relation to the precision metrics (SD and VAR of errors), FA-NBI was better than WM, $ARIMA_{(1,1,1)}$ and BG/D with COV, but it was surpassed by DES, AVG, RB and BG/D without COV. But it was demonstrated in the case that, in view of the joint optimization of multiple objectives (metrics), to achieve more precision it would be necessary to sacrifice performance.

Another objective of the paper was to test the applicability and feasibility of FA-NBI in relation to other real and simulated time series with different characteristics from the coffee demand series. Regarding the two real series, the coffee production and export were chosen to make joint analysis with the internal coffee demand. The Fig. 24 exhibited that the trend of coffee production growth is higher than that of domestic and export consumption, based on the forecasts made by FA-NBI, pointing to a low possibility of both internal and external market shortages of coffee. The results presented in Table 16, for coffee production and exportation series, and in Table 20 for the six simulated series showed that FA-NBI, on average, obtained very good results in terms of performance and precision when compared to individual methods combined and other weighting methods. As an additional advantage, FA-NBI adds information from different combined forecasting methods, and not just one, allowing the identification of the underlying process, since these different forecasting models were able to capture different information's aspects available for prediction. Another advantage is that weights of FA-NBI were selected based on 14 objectives, thus making the best use of available information in each of the 14 metrics. Therefore, the results showed that the approach followed in this work can be an effective way to find the weights for a combination of time series methods.

Table 21
Comparison between individual and other weighting methods with FA-NBI considering simulated series in out-of-sample analysis.

Series	Means' perforfance metrics	Median's performance metrics	Precisions' metrics
MA (1)	Better than all individual and weighting methods, being surpassed in MAPE only by RB and BG/D without COV.	Better than MDEC, AVG, RB and BG/D with and without COV.	Better than MDEC.
		Worse than $ARIMA_{(1,1,1)}$ and $ARMA_{(2,1)}$.	Worse than $ARIMA_{(1,1,1)}$, $ARMA_{(2,1)}$ and all methods to find weights' combination.
MA (2)	Better than all individual methods.	Better than all individual and weighted methods.	Better than $ARIMA_{(1,1,1)}$, RB and BG/D with COV.
	Worse than AVG, RB and BG/D with and without COV in MSE, RMSE, MAPE and sMAPE.		Worse than WM, MDEC, AVG and BG/D without COV.
AR (1)	Better than all individual and weighting methods.	Better than WM, $ARIMA_{(1,1,1)}$, RB and BG/D with COV.	Better than WM, MDEC, AVG and BG/D without COV.
		Worse than MDEC, AVG and BG/D without COV.	Worse than $ARIMA_{(1,1,1)}$, RB and BG/D with COV.
AR (2)	Better than all individual methods.	Better than MDEC and WM.	Better than all individual methods, with MDEC exception.
	Better than RB and BG/D with COV in MSE, RMSE and RMSPE, and worse in the others.	Worse than DES.	Better than all weighting methods
	Worse than AVG and BG/D with COV on all metrics.	Worse than all weighting methods.	
ARMA (1,1)	Better than WM, QTM, AVG and BG/D without COV.	Better than all individual and weighting methods.	Better than QTM, RB and BG/D with COV.
	Better than $ARMA_{(1,1)}$, RB and BG/D with COV in MAE, MASE, MAPE and sMAPE.		Worse than WM, $ARMA_{(1,1)}$, RB and BG/D with COV.
	Worse than $ARMA_{(1,1)}$, RB and BG/D with COV in MSE, RMSE, U1 and U2.		
SARIMA (1,1,1) $(1,1,1)_{12}$	Better than all individual methods and BG/D with and without COV.	Better than WM, EGTM and RB.	Better than WM, EGTM, AVG and BG/D with and without COV.
	Worse than AVG only in MAE, MASE, RMSPE and MAPE.	Worse than $ARIMA_{(1,1,1)}$, AVG and BG/D with and without COV.	Worse than $ARIMA_{(1,1,1)}$ and RB.
	Worse than RB just in sMAPE.		

Acknowledgement

The authors would like to express their gratitude to the Brazilian agencies of CAPES, CNPq, and FAPEMIG for supporting this research.

Appendix A. Forecast time series methods

Time series analysis requires only past values of the variable to be predicted. The pattern observed in the past values is expected to provide adequate information to predict future values. The forecast obtained by averaging methods takes into account the previous observed values of the variable that one requires to predict. They are methods that typically generate "adaptive" forecasts that adjust automatically to the most recent data available. Among the weighted averaging methods, Exponential Smoothing (ES) uses decreasing exponential weights from the more recent observations towards the older (Makridakis et al., 1998). According to Gaither and Frazier (2001), the Single Exponential Smoothing (SES) smooths the data by computing exponentially weighted averages and provides short-term forecasts. It takes the forecast for the previous period and adds an adjustment called smoothing constant or "smoothing" of the level or average (α), for the forecast for the next F_{t+1} period. Thus, the forecast equation for the F_{t+1} forward period is given by $F_{t+1} = F_t + \alpha\,(Y_t - F_t)$, where $Y_t - F_t$ is the forecast error of the period prior to one the be predicted. Unlike SES, the Double Exponential Smoothing (DES) is suitable for use in the presence of trend components and provides short-term forecasts. For this reason, besides the α parameter (for smoothing the average or level, as in SES), the DES method, also known as Browns' method, also uses the β parameter for smoothing the series trend (Gaither and Frazier, 2001). Equation (A.1) (for level), (A.2) (for trend) and (A.3) (for the forecast to m periods ahead) are required to calculate the forecast in DES method (Makridakis et al., 1998), where L_t is an estimate level of the time series in t period, and b_t is the estimate trend of the series in t period.

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} - b_{t-1}) \tag{A.1}$$

$$b_t = (L_t - L_{t-1}) + (1 - \beta)b_{t-1} \tag{A.2}$$

$$F_{t+m} = L_t + b_t m \tag{A.3}$$

Eq. (A.1) adjusts L_t by trend of the b_{t-1} previous period, adding the last smoothed L_{t-1} value to it. This helps to eliminate the delay and takes L_t to the approximate level of current data values. Eq. (A.2) updates the tendency that expresses the difference of the last two smoothed values. This is appropriate because there is a trend in the data, and the new values should be higher or lower than they are. Since there is some remaining randomness, the tendency is modified by smoothing, using β trend in the (L_t - L_{t-1}) last period adding to this last trend estimate multiplied by (1-β). Finally, Eq. (A.3) is used to calculate the forecast for m periods ahead, with the trend (b_t) being multiplied by m and added to the L_t base value. The initialization process requires two estimates, one for the first Y_1 smoothed value, and the other for the T_1 trend value. Alternatively, one can use $T_1 = Y_1$. The α and β weights may be chosen by trial, for example, as the values that together minimize the value of Mean Absolute Deviation (MAD) (Makridakis et al., 1998).

The Holt-Winters Exponential Smoothing (HW) method is an extension of DES method, and its forecast considers not only the smoothing constants for the level and trend but also the smoothing constant for seasonality (Newbold, 1994). The HW method, therefore, is based on three exponential smoothing equations: one to smooth the level, another one to smooth the trend, and the last one to smooth the seasonality. Therefore, besides the α parameter (for smoothing the average or level) and the β parameter (for smoothing to the trend), there is the γ parameter to smoothing the seasonality of the series. There are two different HW methods, depending whether seasonality is modeled multiplicatively or additively: The

Holt-Winter's multiplicative (WM) and the Holt-Winter's additive (WA). WM is based in Equation (A.4) (for level), (A.5) (for trend), (A.6) (for seasonality), and (A.7) (forecast for m periods ahead), where s is the size of seasonality (Makridakis et al., 1998):

$$L_t = \alpha\left(\frac{Y_t}{S_{t-s}}\right) + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{A.4}$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{A.5}$$

$$S_t = \gamma\left(\frac{Y_t}{L_t}\right) + (1-\gamma)S_{t-s} \tag{A.6}$$

$$F_{t+m} = (L_t + b_t m)S_{t-s+m} \tag{A.7}$$

Equation (A.4) differs from Eq. (A.1) due to its first term being divided by the (S_{t-s}) seasonal number to deseasonalize the Y_t series. Eq. (A.5) is the same as Eq. (A.2) of trend smoothing in DES. Eq. (A.6) is compared to a seasonal index because Y_t, which contains seasonality, is divided by current smoothed value of the series, where the division value may be higher or lower than 1 depending on the magnitude of Y_t. As Y_t also contains randomness, the Eq. (A.6) weights the newly computed seasonal factor with γ and the most recent seasonal number corresponding to the same season with 1-γ (Makridakis et al., 1998).

Autoregressive Integrated Moving Average (ARIMA) models were popularized by George Box and Gwilym Jenkins in the early 1970s. To carry out predictions from these models, it is important to know if the stochastic process, which generated the data set, varies or not in relation to time, i.e., if the process is stationary or not. In the case of linear stationary processes – where the data fluctuate around a constant mean, independent of time, and the variance of the fluctuation remains constant over time – the models used to represent them are the autoregressive models of order (p), AR (p); the moving average models of order (q), MA (q); and autoregressive models of order (p) coupled with the moving averages models of order (q), ARMA (p, q). In the event of non-stationary data series regarding the level and/or slope, the generating process of this series will be a linear homogeneous non-stationary process. In this case, to make the process stationary it will be necessary to promote d differentiations of this series. These processes will be represented by autoregressive integrated moving average models of order p, d and q (ARIMA (p, d, q)). After the time series are differentiated (d) times to make it stationary, the ARIMA (p, d, q) model can be represented by Eq. (A.8), that is, by the linear function of last known observed values of the variable being predicted and last prediction errors of an ARMA model:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + ,\ldots, + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - ,\ldots, - \theta_q\varepsilon_{t-q} \tag{A.8}$$

where Y_t is the forecast value in t period; ε_t is the random error at t period, that is, a random variable normally distributed, with a mean of zero, constant variance and covariance with value zero; φ_i ($i = 1, 2, 3, \ldots, p$) are the autoregressive parameters, θ_i ($j = 1, 2, 3 \ldots, q$) are the moving average parameters, and p and q are integers referred to as orders of the model.

Regarding the ARIMA model, the random process that generates Y_t series will be described by a weighted average of p past observations of the variable to be predicted, added to random disturbance in the current period (ε_t) coupled with a weighted average of the error series (random disturbances) that returns to q periods.

Appendix B. Statistical error measures

Metrics and equations	Definitions	
$MAE = \frac{1}{n}\sum_{t=1}^{n}\lvert e_t\rvert = \underset{t=1,\ldots n}{\text{mean}}\lvert e_t\rvert$	Mean absolute value of errors	(B.1)
$MdAE = \underset{t=1,\ldots n}{\text{median}}\lvert e_t\rvert$	Median absolute value of errors	(B.2)
$MSE = \frac{1}{n}\sum_{t=1}^{n}(e_t)^2 = \underset{t=1,\ldots n}{\text{mean}}(e_t^2)$	Mean squared value of errors	(B.3)
$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(e_t)^2} = \sqrt{\underset{t=1,\ldots n}{\text{mean}}(e_t^2)}$	Root mean squared value of errors	(B.4)
$MAPE = \frac{1}{n}\sum_{t=1}^{n}\lvert re_t\rvert = \underset{t=1,\ldots n}{\text{mean}}(re_t)$ Where: $re_t = \left(\frac{e_t}{y_t}\right)\times 100$	Mean absolute percentage value of errors	(B.5)
$RMSPE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(re_t)^2}$	Root mean square percentage error	(B.6)
$MdAPE = \underset{t=1,\ldots n}{\text{median}}\lvert re_t\rvert$	Median absolute percentage value of errors	(B.7)
$sMAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{\lvert\hat{y}_t - y_t\rvert}{(y_t + \hat{y}_t)/2}$	Symmetric mean absolute percentage value of errors	(B.8)
$sMdAPE = \underset{t=1,\ldots n}{\text{median}}\frac{\lvert\hat{y}_t - y_t\rvert}{(y_t - \hat{y}_t)/2}$	Symmetric median absolute percentage value of errors	(B.9)
$U1 = \frac{\left[\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2\right]^{\frac{1}{2}}}{\left[\frac{1}{n}\sum_{t=1}^{n}(y_t)^2\right]^{\frac{1}{2}} + \left[\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t)^2\right]^{\frac{1}{2}}}$	U1 Theil's statistic coefficient	(B.10)
$U2 = \frac{\left[\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2\right]^{\frac{1}{2}}}{\left[\frac{1}{n}\sum_{t=1}^{n}(y_t)^2\right]^{\frac{1}{2}}}$	U2 Theil's statistical coefficient	(B.11)
$MASE = \frac{MAE}{MAE_{RW}}$	Mean absolute scaled value of errors	(B.12)
$VAR = s^2 = \frac{\sum_{t=1}^{n}(e_t - \bar{e}_t)^2}{n-1}$	Variance of errors	(B.13)

$SD = s = \sqrt{\frac{\sum_{t=1}^{n}(e_t - \bar{e}_t)^2}{n-1}}$	Standard deviation of errors	(B.14)

Where *rw* is the random walk method; y_t is the value observed in period *t;* $\hat{y}_t$ is the predict value in the period *t*; ($e_t = y_t - \hat{y}_t$) is the error from the model-fitting process in period *t; n* is the number of observations; and $\bar{e}_t$ is the arithmetic mean of the errors series (*e*).

Appendix C. Experimental Designs of FS_1 and FS_2

Table C.1
Experimental Design of FS_1 and FS_2.

c	Weights DES	WM	ARIMA (1,1,1)	ARIMA (2,2,3)	Factors Scores FS_1	FS_2	c	Weights DES	WM	ARIMA (1,1,1)	ARIMA (2,2,3)	Factors Scores FS_1	FS_2
1	1.000	0.000	0.000	0.000	0.954	2.215	32	0.200	0.000	0.600	0.200	−0.175	0.778
2	0.800	0.200	0.000	0.000	0.536	1.009	33	0.200	0.000	0.400	0.400	−0.801	0.469
3	0.800	0.000	0.200	0.000	0.808	1.186	34	0.200	0.000	0.200	0.600	−1.133	0.884
4	0.800	0.000	0.000	0.200	0.259	1.878	35	0.200	0.000	0.000	0.800	−1.166	0.969
5	0.600	0.400	0.000	0.000	0.569	0.325	36	0.000	1.000	0.000	0.000	2.941	−0.523
6	0.600	0.200	0.200	0.000	0.537	0.431	37	0.000	0.800	0.200	0.000	2.180	−0.724
7	0.600	0.200	0.000	0.200	−0.271	0.154	38	0.000	0.800	0.000	0.200	0.974	−0.949
8	0.600	0.000	0.400	0.000	0.723	0.910	39	0.000	0.600	0.400	0.000	1.646	−1.449
9	0.600	0.000	0.200	0.200	0.007	0.946	40	0.000	0.600	0.200	0.200	0.532	−1.704
10	0.600	0.000	0.000	0.400	−0.287	1.251	41	0.000	0.600	0.000	0.400	−0.478	−1.854
11	0.400	0.600	0.000	0.000	0.964	−0.495	42	0.000	0.400	0.600	0.000	1.168	−0.361
12	0.400	0.400	0.200	0.000	0.687	−0.034	43	0.000	0.400	0.400	0.200	0.105	−0.987
13	0.400	0.400	0.000	0.200	−0.241	−0.473	44	0.000	0.400	0.200	0.400	−0.812	−1.444
14	0.400	0.200	0.400	0.000	0.565	0.887	45	0.000	0.400	0.000	0.600	−1.464	−2.424
15	0.400	0.200	0.200	0.200	−0.314	0.353	46	0.000	0.200	0.800	0.000	0.833	0.407
16	0.400	0.200	0.000	0.400	−0.933	−0.078	47	0.000	0.200	0.600	0.200	−0.125	−0.012
17	0.400	0.000	0.600	0.000	0.671	0.989	48	0.000	0.200	0.400	0.400	−0.922	−0.851
18	0.400	0.000	0.400	0.200	−0.128	0.663	49	0.000	0.200	0.200	0.600	−1.531	−1.031
19	0.400	0.000	0.200	0.400	−0.619	1.120	50	0.000	0.200	0.000	0.800	−1.790	−0.532
20	0.400	0.000	0.000	0.600	−0.782	1.200	51	0.000	0.000	1.000	0.000	0.762	0.903
21	0.200	0.800	0.000	0.000	1.744	−1.441	52	0.000	0.000	0.800	0.200	−0.148	0.681
22	0.200	0.600	0.200	0.000	1.271	−1.080	53	0.000	0.000	0.600	0.400	−0.859	−0.005
23	0.200	0.600	0.000	0.200	0.191	−1.264	54	0.000	0.000	0.400	0.600	−1.313	−0.003
24	0.200	0.400	0.400	0.000	0.854	0.092	55	0.000	0.000	0.200	0.800	−1.486	0.117
25	0.200	0.400	0.200	0.200	−0.099	−1.020	56	0.000	0.000	0.000	1.000	−1.416	0.531
26	0.200	0.400	0.000	0.400	−0.933	−1.439	57	0.250	0.250	0.250	0.250	−0.452	−0.301
27	0.200	0.200	0.600	0.000	0.672	0.490	58	0.625	0.125	0.125	0.125	0.090	0.485
28	0.200	0.200	0.400	0.200	−0.261	0.369	59	0.125	0.625	0.125	0.125	0.773	−1.825
29	0.200	0.200	0.200	0.400	−0.979	−0.445	60	0.125	0.125	0.625	0.125	0.082	0.826
30	0.200	0.200	0.000	0.600	−1.425	−0.186	61	0.125	0.125	0.125	0.625	−1.433	0.170
31	0.200	0.000	0.800	0.000	0.675	1.244							

Appendix D. The DM test

For one-step ahead forecasts, the DM test is computed considering the errors series of the two comparison models – (e_1) and (e_2) series. First, it is calculated the loss differential series (d_t) as Eq. (D.1), such as MAE error statistic, or as Eq. (D.2), such as MSE error statistic, for all *t* observations of e_1 and e_2, with $t = 1,2, \ldots n$.

$$d_t = |L(e_{1,t}) - L(e_{2,t})| \tag{D.1}$$

$$d_t = L(e_{1,t})^2 - L(e_{2,t})^2 \tag{D.2}$$

In Equations (D.1) and (D.2), L is the loss function of the prediction error, which is used to measure the forecasting accuracy of different models. The DM test compares the accuracy of two forecasts under the null hypothesis (H_0) which assumes that there is no significant difference about the prediction performance between the proposed model 1 and the compared model 2, i.e. all relevant information contained in 1 are also contained in 2. Thus, H_o is accept if E [L ($e_{1,t}$)] = E [L ($e_{2,t}$], and H_0 is rejected if E [L ($e_{1,t}$)] ≠ E [L ($e_{2,t}$].

Assuming that d_t has stationary covariance and DM test has an asymptotic standard normal distribution under the null hypothesis of equal predictive accuracy, in a second step it is calculated the mean value ($\bar{d}$) of the d_t series as ($\bar{d} = \frac{1}{n}\sum_{t=1}^{n} d_t$). The DM test is calculated as Eq. (D.3), where $\hat{V}(\bar{d})$ is a consistent estimate of the asymptotic variance of $\bar{d}$ calculated as Eq. (D.4):

$$DM = \bar{d}\Big/\sqrt{\hat{V}(\bar{d})} \tag{D.3}$$

$$\hat{V}(\bar{d}) \approx \frac{1}{n}\left(\gamma_0 + 2\sum_{i=1}^{\tau-1}\gamma_i\right) \tag{D.4}$$

where γ_i is the *i-th* autocovariance of $\overline{d,}$ estimated by Eq. (D.5) and is still assumed that τ-step-ahead forecasts exhibit dependence up to order τ −1 (Kisinbay, 2010).

$$\hat{\gamma}_i = n^{-1} \sum_{t=1+i}^{n} (d_t - \bar{d})(d_{t-i} - \bar{d}) \tag{D.5}$$

The null hypothesis will be rejected if $DM > Z_{\alpha/2}$, i.e. if P-value > 0.05 where $Z_{\alpha/2}$ is the critical (z – value) of the standard normal distribution, and α is the significance level (Xu et al., 2017; Du et al., 2017).

Appendix E. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpe.2019.03.001.

References

Adhikari, R., Agrawal, R.K., 2014. Performance evaluation of weights selection schemes for linear combination of multiple forecasts. Artif. Intell. Ver. 42, 529–548.
Ahlburg, D.A., 1992. A commentary on error measures. Int. J. Forecast. 8, 99–111.
Ahmadi, A., Kaymanesh, A., Siano, P., Janghorbani, M., Nezhad, A.E., Sarno, D., 2015. Evaluating the effectiveness of normal boundary intersection method for short-term environmental/economic hydrothermal self-scheduling. Electr. Power Syst. Res. 123, 192–204.
Andrawis, R.R., Atiya, A.F., El-Shishiny, H., 2011. Combination of long term and short-term forecasts, with application to tourism demand forecasting. Int. J. Forecast. 27, 870–886.
Armstrong, J.S., Collopy, F., 1992. Error measures for generalizing about forecasting methods: empirical comparisons. Int. J. Forecast. 8, 69–80.
Barrow, D.K., Crone, S.F., 2016. Cross-validation aggregation for combining auto-regressive neural network forecasts. Int. J. Forecast. 32, 1120–1137.
Barrow, D.K., Kourentzes, N., 2016. Distributions of forecasting errors of forecast combinations: implications for inventory management. Int. J. Prod. Econ. 177, 24–33.
Bates, J.M., Granger, C.W.J., 1969. The combination of forecasts. Oper. Res. Soc. 20, 451–468.
Bjørnland, H.C., Gerdrup, K., Jore, A.S., Smith, C., Thorsrud, L.A., 2012. Does forecast combination improve norges bank inflation forecasts? Oxf. Bull. Econ. Stat. 74 (2). https://doi.org/10.1111/j.1468-0084.2011.00639. 0305-9049.
Bordignon, S., Bunn, D.W., Lisi, F., Nan, F., 2013. Combining day-ahead forecasts for British electricity prices. Energy Econ. 35, 88–103.
Brito, T.G., Paiva, A.P., Ferreira, J.R., Gomes, J.H.F., Balestrassi, P.P., 2014. A normal boundary intersection approach to multiresponse robust optimization of the surface roughness in end milling process with combined arrays. Precis. Eng. 38, 628–638.
Bunn, D.W., 1975. A bayesian approach to the linear combination of forecasts. Oper. Res. Q. 26 (2), 325–329 part 1.
Cang, S., Yu, H., 2014. A combination selection algorithm on forecasting. Eur. J. Oper. Res. 234, 127–138.
Chan, C.K., Kingsman, B.G., Wong, H., 1999. The value of combining forecasts in inventory management - a case study in banking. Eur. J. Oper. Res. 117, 199–210.
Clemen, R.T., 1989. Combining forecasts: a review and annotated. Int. J. Forecast. 5, 559–583.
Cornell, J.A., 2002. Experiments with Mixtures: Designs, Models and the Analysis of the Mixture Data, 3th. ed. John Wiley & Sons, New York.
Cornell, J.A., 2011. A Primer on Experiments with Mixtures, tenth ed. John Wiley & Sons, New Jersey.
Coronado, M., Segadães, A.M., Andrés, A., 2014. Combining mixture design of experiments with phase diagrams in the evaluation of structural ceramics containing foundry by-products. Appl. Clay Sci. 101, 390–400.
Crone, S.F., Hibon, M., Nikolopoulos, K., 2011. Advances in forecasting with neural networks? Empirical evidence from the NN3 competition on time series prediction. Int. J. Forecast. 27, 635–660.
Das, I., Dennis, J.E., 1998. Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems. Soc. Ind. Appl. Math. 8, 631–657.
Dekker, M., Donselaar, K.V.D., Ouwehand, P., 2004. How to use aggregation and combined forecasting to improve seasonal demand forecasts. Int. J. Prod. Econ. 90, 151–167.
Deutsch, M., Granger, C.W.J., Terasvirta, T., 1994. The combination of forecasts using changing weights. Int. J. Forecast. 10, 47–57.
Dickinson, J.P., 1973. Some statistical results in the combination of forecasts. Oper. Res. Q. 24 (2), 253–260.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13 (3), 253–263.
Diebold, F.X., Pauly, P., 1990. The use of prior information in forecast combination. Int. J. Forecast. 6, 503–508.
Dixon, W.J., 1950. Analysis of extreme values. Ann. Math. Stat. 21 (4), 488–506.
Du, P., Wang, J., Guo, Z., Yang, W., 2017. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. Energy Convers. Manag. 150, 90–107.
Du, P., Wang, J., Yang, W., Niu, T., 2018. Multi-step ahead forecasting in electrical power system using a hybrid forecasting system. Renew. Energy 122, 533–550.
Fang, Y., 2003. Forecasting combination and encompassing tests. Int. J. Forecast. 19, 87–94.
Faria, A.E., Mubwandarikwa, E., 2008. The geometric combination of Bayesian forecasting models. J. Forecast. 27, 519–535.
Fildes, R., Fotios, P., 2015. Simple versus complex selection rules for forecasting many time series. J. Bus. Res. 68, 1692–1701.
Gaither, N., Frazier, G., 2001. Operations Management. South-Western, Ohio.
Ganesan, T., Vasant, P., Elamvazuthi, I., 2013. Normal-boundary intersection based parametric multi-objective optimization of green sand mould system. J. Manuf. Syst. 32, 197–205.
Graefe, A., Armstrong, J.S., Jones, J.R., Randall, J., Cuzán, A.G., 2014. Combining forecasts: an application to elections. Int. J. Forecast. 30, 43–54.
Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. J. Forecast. 3, 197–204.
Gulledge, J.R., Thomas, R., Ringuest, J.L., Richardson, J.A., 1986. Subjective evaluation of composite econometric policy inputs. Soc. Econ. Plann. Sci. 20, 51–55.
Hair Jr., J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2010. Multivariate Data Analysis, 7 th ed. Prentice Hall, New York.
Hayton, J.C., Allen, D.G., Vida, S., 2004. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. Organ. Res. Methods 7, 191–205.
Hibon, M., Evgeniou, T., 2005. To Combine or not to Combine: selecting among forecasts and their combinations. Int. J. Forecast. 21, 15–24.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688.
International Cofee Organization (ICO). . http://www.ico.org, Accessed date: 23 June 2018.
Jia, Z., Ierapetritou, M.G., 2007. Generate Pareto optimal solutions of scheduling problems using normal boundary intersection technique. Comput. Chem. Eng. 31, 268–280.
Johnson, R.A., Wichern, D.W., 2007. Applied Multivariate Statistical Analysis, sixth ed. Pearson, New York.
Jolliffe, L.T., 2002. Principal Component Analysis, 2th ed. Springer series in statistics, New York.
Jose, V.R.R., Winkler, R.L., 2008. Simple robust averages of forecasts: some empirical results. Int. J. Forecast. 24, 163–169.
Kaiser, H.F., 1960. The application of electronic computers to factor analysis. Educ. Psychol. Meas. 20, 141–151.
Kang, H., 1986. Unstable in the combination of forecasts. Manag. Sci. 32, 683–695.
Kisinbay, T., 2010. The USE of encompassing tests for forecast combinations. J. Forecast. 29, 715–727.
Lam, K.F., Mui, H.W., Yuen, H.K., 2001. A note on minimizing absolute percentage error in combined forecasts. Comput. Oper. Res. 28, 1141–1147.
Lesage, J.P., Magura, M., 1992. A mixture-model approach to combining forecasts. J. Bus. Econ. Stat. 10, 445–452.
Leung, M.T., Daouk, H., Che, A.-S., 2001. Using investment portfolio return to combine forecasts: a multiobjective approach. Eur. J. Oper. Res. 134, 84–102.
Lopes, L.G.D., Brito, T.G., Paiva, A.P., Peruchi, R.S., Balestrassi, P.P., 2016. Robust parameter optimization based on multivariate normal boundary intersection. Comput. Ind. Eng. 93, 55–66.
Mahmoud, E., 1989. Combining forecasts: some managerial issues. Int. J. Forecast. 5, 599–600.
Makridakis, S., 1990. Sliding simulation: a new approach to time series forecasting. Manag. Sci. 36, 505–512.
Makridakis, S., 1993. Accuracy measures: theoretic and practical concerns. Int. J. Forecast. 9, 527–529.
Makridakis, S., Winkler, R.L., 1983. Averages of forecasts: some empirical results. Manag. Sci. 29, 987–996.
Makridakis, S., Wheelwrigt, S.C., Hyndman, R.J., 1998. Forecasting: Methods and Applications, 3th ed. John Wiley & Sons, New York.
Martínez-Rivera, B., Ventosa-Santaulària, D., Vera-Valdés, J.E., 2012. Spurious forecasts? J. Forecast. 31, 245–259.
Martins, V.L.M., Werner, L., 2012. Forecast Combination in Industrial Series: a comparison between individual forecasts and its combinations with and without correlated errors. Expert Syst. Appl. 39, 11479–11486.
Meade, N., 2002. A comparison of the accuracy of short term foreign Exchange forecasting methods. Int. J. Forecast. 18, 67–83.
Montgomery, D.C., Jennings, C.L., Kulahci, M., 2008. Introduction to Times Series Analysis and Forecasting. John Wiley & Sons, New York.
Moreno, B., López, A.J., 2013. Combining economic forecasts by using a maximum Entropy econometric approach. J. Forecast. 32, 124–136.
Myers, R.H., Montgomery, D.C., Anderson-Cook, C.M., 2009. Response Surface Methodology: Process and Product Optimization Using Design Experiments, 3th. ed. John Wiley & Sons, New York.
Newbold, P., 1994. Statistics for Business & Economics. Prentice Hall. Inc, New Jersey.
Newbold, P., Granger, C.W.J., 1974. Experience with forecasting univariate time series and the combination of forecasts. J. R. Stat. Soc. 137, 131–165.
Oliveira, F.A., Paiva, A. P. de, Lima, J.W.M., Balestrassi, P.P., Mendes, R.R.A., 2011.

Portfolio optimization using Mixture Design of Experiments: scheduling trades within electricity markets. Energy Econ. 33, 24–32.
Osborn, J.W., 2015. What is rotation in exploratory factor analysis? Practical Assess. Res. Eval. 20, 1–7.
Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K., 2014. 'Horses for Courses' in demand forecasting. Eur. J. Oper. Res. 237, 152–163.
Reeves, G.R., Lawrence, K.D., 1991. Combining forecasts given different types of objectives. Eur. J. Oper. Res. 51, 65–72.
Reeves, G.R., Lawrence, K.D., Lawrence, 1982. Combining multiple forecasts given multiple objectives. J. Forecast. 1, 271–279.
Reeves, G.R., Lawrence, K.D., Lawrence, S.M., Guerard, J.B., 1988. Combining earnings forecasts using multiple objective linear programming. Comput. Oper. Res. 15, 551–559.
Rocha, L.C.S., Paiva, P.P., Balestrassi, P.P., Severino, G., Rotela Junior, P., 2015. Entropy-based weighting for multiobjective optimization: as application on vertical turning. Math. Probl Eng. 2015, 1–11.
Sankaran, S., 1989. A comparative evaluation of methods for combining forecasts. Akron Bus. Econ. Rev. 20, 33–39.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (379–423), 623–656.
Shukla, P.K., Deb, K., 2007. On finding multiple Pareto-optimal solutions using classical evolutionary generating methods. Eur. J. Oper. Res. 181, 1630–1652.
Simionescu, M., 2013. The performance of unemployment rate predictions in Romania. Strategies to improve the forecasts accuracy. Rev. Econ. Persp. Národohospodárský obzor 13 (4), 161–175. https://doi.org/10.2478/revecp-2013-0007.
Tselentis, D.I., Vlahogianni, E.I., Karlaftis, M.G., 2015. Improving short-term traffic forecasts: to combine models or not to combine? IET Intell. Transp. Syst. 9 (2), 193–201. https://doi.org/10.1049/iet-its.2013.0191.
Tseng, F.-M., Yu, H.-C., Tzeng, G.-H., 2002. Combining neural network model with seasonal time series ARIMA model. Technol. Forecast. Soc. Change 69, 71–87.
Ustun, O., Kasimbeyli, R., 2012. Combined forecasts in portfolio optimization: a generalized approach. Comput. Oper. Res. 39, 805–819.
Utyuzhnikov, S.V., Fantini, P., Guenov, M.D., 2009. A method for generating a well-distributed Pareto set in nonlinear multiobjective optimization. J. Comput. Appl. Math. 223, 820–884.
Vahidinasab, V., Jadid, S., 2010. Normal boundary intersection method for suppliers' strategic bidding in electricity markets: an environmental/economic approach. Energy Convers. Manag. 51, 1111–1119.
Wallstrom, P., Segerstedt, A., 2010. Evaluation of forecasting error measurements and techniques for intermittent demand. Int. J. Prod. Econ. 128, 625–636.
Wang, J., Du, P., Niu, T., Yang, W., 2017. A novel hybrid system based on a new proposed algorithm – multi-Objective Algorithm for wind speed forecasting. Appl. Energy 208, 344–360.
Wang, J., Yang, W., Du, P., Niu, T., 2018. A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm. Energy Convers. Manag. 163, 134–150.
Weatherford, L.R., Kimes, S.E., 2003. A comparison of forecasting methods for hotel revenue management. Int. J. Forecast. 19, 401–415.
Winkler, R.L., 1989. Combining Forecasts: a philosophical basis and some current issues. Int. J. Forecast. 5, 605–609.
Winkler, R.L., Makridakis, S., 1983. The combination of forecast. J. R. Stat. Soc. 146, 150–157.
Xu, Y., Yang, W., Wang, J., Yang, 2017. Air quality early-warming system for cities in China. Atmos. Environ. 148, 239–257.
Zhao, W., Wang, J., Lu, H., 2014. Combining forecasts of electricity consumption in China with time-varying weights updated by a high-order Markov chain model. Omega 45, 80–91.