



Forecast combination using Optimization techniques

Marisol Valencia Cárdenas, Juan Carlos Correa Morales

ISBN: 978-958-48-8507-4

Abstract

Currently diverse forecasting methodologies emerge, based on the empirical knowledge, innovative methods, individual or combined, demonstrating optimal results. This document is derived from a research process, and presents alternatives related to forecast combinations, using metaheuristics, for example, by using Tabu search and Evolutionary programming to optimize forecasting. One of the designed process consists of creating combination forecasts based on evolutionary programming using, first, a mixture of Bayesian regression models and, second, a mixture of the classical linear regression model, the autoregressive integrated moving average model, exponential smoothing and Bayesian regression.

This document presents two papers derived from the research about forecast combination and optimization techniques based on simulation and statistical processes. The first research compares the novel combined algorithm with the individual results of these individual models and with the Bates and Granger combination using an error indicator and the symmetrical mean absolute error value. Those models and the novel design were applied to time series simulation and to a real case of dairy products sales, thus generating multiproduct combination forecasts for both the simulation and the real case. The novel combination combined with the evolutionary metaheuristic showed better results than those of the others that were used. The second research uses simulated time series and other metaheuristic that learns from the data and statistical behavior, comparing the combined forecasts with individual prediction results.

Table of Contents

Forecast combination based on evolutionary programming and a mixture of Bayesian and classical models for seasonal time series.....	5
Combination Forecasting algorithm with cluster Analysis, in a simulation scenario.....	28

Forecast combination based on evolutionary programming and a mixture of Bayesian and classical models for seasonal time series

Marisol Valencia Cárdenas¹, Juan Carlos Correa Morales²

- 1- *Professor. I.U. Tecnológico de Antioquia, Administrative and Economic Faculty, Medellín, Antioquia, Colombia. mvalencia@unal.edu.co, solmarival@gmail.com. <https://orcid.org/0000-0003-3135-3012>*
- 2- *Associate Professor. Universidad Nacional de Colombia, Medellín, Statistis. Medellín, Antioquia, Colombia. jccorrea@unal.edu.co*

ABSTRACT

This research proposes an evolutionary programming algorithm to perform combination forecasts for multi-item time series responses based on individual forecasts. The designed process consists of creating combination forecasts based on evolutionary programming using, first, a mixture of Bayesian regression models and, second, a mixture of the classical linear regression model, the autoregressive integrated moving average model, exponential smoothing and Bayesian regression. This research compares the novel combined algorithm with the individual results of these individual models and with the Bates and Granger combination using an error indicator and the symmetrical mean absolute error value. Those models and the novel design were applied to time series simulation and to a real case of dairy products sales, thus generating multiproduct combination forecasts for both the simulation and the real case. The novel combination combined with the evolutionary metaheuristic showed better results than those of the others that were used. The research facilitates a novel and practice form of the forecasting process for the industry, students, or research community who require multiproduct planning of demand. The research value is the novel accurate combination of Bayesian models and other mixtures of classical and Bayesian methods with the individual results applied for multi-item or multiproduct forecast generation. In addition, the program is designed based on the R software packaged, and it can also be reproduced using any other statistical programming software.

¹ Author for correspondence: Marisol Valencia Cárdenas. mvalencia@unal.edu.co.

Keywords: (B) Simulation; (O) Metaheuristics; (T) Forecasting; (D) Bayesian Models; (D) Analytics.

1. Introduction

Forecast communities, academics and companies are seeking more accurate alternatives that make good predictions, such as for multiple product demands in order to better plan inventories. The final sales are a determinant variable for the storage of manufacturing companies, and the variable can be found to be both deterministic and stochastic. Currently, more industries that use inventory controls also use statistical models to predict sales to improve the efficiency of production, as has been shown by different authors (Valencia, Díaz, & Correa 2015); in addition, statistical models can represent the uncertainty of these variables in inventory models to produce orders and stocks to correctly assess supply and try to cover market needs (Simchi-Levi, Kaminski, & Simchi-Levi, 2008).

Uncertainty is frequent due to many kinds of variations that occur in industrial processes due to, e.g., internal logistics, and these variations can affect the service. In this sense, when the data are not well fit with a normal distribution, such as for countable responses, there are few statistical models that provide acceptable predictions, and some of them are based on the Poisson distribution (Kolassa, 2016). Appropriate forecast models are important to make accurate projections of variables in different kinds of areas besides industry final demands, including population increases, flow growth, and energy prices, among others, and innovative forms that do this analysis are always valuable. There are two methods to perform forecasts: (i) an individual forecast and (ii) a combination of individual forecasts (Guo et al. 2017).

Classical regression with time series components, autoregressive integrated moving average (ARIMA) models and exponential smoothing are some of the most-used models for individual forecasts, but there are other techniques, such as neural networks and other more recent models based on Bayesian techniques that are also useful for forecasts (Petris, Petrone, & Campagnoli, 2009).

Bayesian forecasting models are often used by starting with the definition of a data distribution and prior distributions for the parameters, and not much data should be used to estimate models with these theories (Valencia-Cárdenas, 2016); these techniques have been proved to be accurate in comparison with other classical models, and they are applicable to practical fields, such as the agro-industrial sector. Agarwal et al., (2005) proposed a Bayesian regression in order to make better decisions

regarding deforestation practices. Ferragina et al. (2015) designed a prediction of fatty acids in the dairy sector with Bayesian models.

Dynamic updating is frequent in statistical forecasting models, such as dynamic linear Bayesian models (Valencia et al., 2015) and other Bayesian techniques. Valencia (2016) presented an updating process for the design matrix of a Bayesian regression model (BRM) for every forecasted period, and comparison with other models using a simulation showed that the proposed method was more accurate than classical regression, ARIMA, exponential smoothing, and the dynamic Bayesian models defined by Petris, Petrone and Campagnoli (2009).

Forecast combinations are based on individual models that do not always follow the same theoretical assumptions for estimations. A forecast combination based on individual models, such as ARIMA, neural networks, and regressions, was presented in Melo, Loaiza and Villamizar-Villegas (2016); the authors designed a combination Bayesian technique that improved the accuracy of the other models, and since neural networks are not based on the normal distribution assumption, as with ARIMA and Gaussian regression, the method has flexibility in the individual model that can be incorporated into the combination.

This flexibility in a forecast combination method helps improve the accuracy of the estimations, as has been shown in recent years. Combination techniques to perform forecasts are beginning to show increasing success (Cang & Yu, 2014) in looking for better accuracy in the predictions of different kinds of variables. Barrow and Kourentzes (2016) affirm that such techniques have applications in multiple fields, such as the forecasts for sales products to be considered in this work.

Different structures of forecast combinations are found in the literature (Hyndman, Ahmed, Athanasopoulos, & Shang, 2011; Miller, Berry, & Lai, 2007; Zotteri & Kalchschmidt, 2007). These combinations use weights such as the inverse of the root mean square error (RMSE) or random variables that are found after an optimization problem's formulation, and the weights are used in order to apply a linear combination of individual forecasts. These random variables can also be estimated with Bayesian techniques, such as by using the posterior Bayesian distribution presented by Li, Shi and Zhou (2011).

Many authors use Bayesian theory to improve the accuracy of the forecasts, thus creating innovative methods in order to find more precision (Cang & Yu, 2014; Hsiao & Wan, 2014; Kociecki, Kolasa,

& Rubaszek, 2012). Changing the prior distribution in the Bayesian method is also possible, and expert judgement can be added to a forecasting approach when there are no historical data (Kociecki et al., 2012).

The application of heuristic methods to build combination forecasts that find optimal weights can also provide more accurate results, according to Cang and Yu, (2014), because it can search the most approximate values of these in order to minimize an indicator, such as the RMSE.

These mixed techniques are associated with Salas (2008), who used a Bayesian mixed model based on alfa-stable distributions. The author affirms that the mixture requires independent and identically distributed random variables. Moreover, the Gaussian is among the most-used mixed models, but the weights must equal 1, which is a constraint that is not always considered in a metaheuristic process.

Chahkoutahi & Khashei (2017) and Gao, Sarlak, Parsaei, & Ferdosi (2017) propose the use of heuristics and metaheuristics in combination forecasts and affirm that a heuristic provides a reasonable solution to a problem (Silver, 2004); however, the authors do not guarantee that the heuristic will produce a mathematically optimal solution. A metaheuristic is an iterative master process that guides and modifies heuristic operations to produce efficient solutions. It means that metaheuristic processes use heuristics and other techniques, such as simulation or mathematics, to produce better solutions than those of a simple heuristic process.

There are many articles associated with metaheuristics applications (Silver, 2004), and among the most-used techniques are particle swarm optimization (PSO), the genetic algorithm (GA), ant colony optimization (ACO), the artificial bee colony (ABC), and differential evolution (DE). Successful applications of these techniques in combination forecasts use time series information and searching strategies, along with different kinds of models for the individual predictions to find better solutions (Chahkoutahi & Khashei, 2017; Gao et al., 2017).

The load forecast combination is another technique consisting of the formulation or searching process of weighted random values in a linear combination with the forecasted values. Nowotarski, Liu, Weron, and Hong (2016) affirm that this area requires more attention and development. Chahkoutahi and Khashei (2017) present some recent works related to hybrid techniques that are used to find the weights of a combination technique and affirm that one of the disadvantages of these methods is that they find a local optimum that is not always the real one. The authors use models such as multilayer perceptron (MLP) neural networks, adaptive network-based fuzzy inference systems (ANFIS), and

the seasonal autoregressive integrated moving average (SARIMA) to design the combination based on the individual models, thus creating a novel direct optimum parallel hybrid (DOPH) as the metaheuristic for the optimization. Other metaheuristics such as the genetic algorithm process are used in Sermpinis et al. (2015), and they produced accurate results for the forecast combination.

This paper presents a novel forecasting combination model that uses a metaheuristic technique based on evolutionary programming. In addition, it also uses the evolutionary programming metaheuristic proposal (EVOL) for multi-item time series, which is tested by comparing an EVOL combination of the Bayesian regression model developed by Valencia-Cárdenas (2016) to perform individual forecasts that consider the seasonal behaviour of the time series. The EVOL combination is then applied by using the individual models of linear regression, exponential smoothing, ARIMA, and the Bates and Granger (BG) combination result, thus finding the best possible model. In this sense, this research compares the novel algorithm with the results of all these individual forecasts by using an error indicator called the symmetrical mean absolute percentage error (SMAPE) value. The EVOL first applies the forecast combination to multivariate simulated time series data by using a non-normal distribution and non-stationary behaviour, but with a seasonal pattern and autocorrelation. Second, the method is applied to a real case consisting of three dairy product sales of a Colombian company, thus providing the forecasting combination for multiple products.

2. Methods

2.1. Bayesian process

Bayesian analysis uses Bayes' theorem to make inferences. It assumes some probability distributions for the implicit parameters in order to provide estimations for these or forecasts for the response, as in this case. The assumptions of these techniques are different with respect to classical models since the incorporation of prior information is quantified in a probability distribution (Gill, 2007).

The data information must assume other probability distributions, which are represented by y_1, \dots, y_n . This information is included in the likelihood function $L(y_1, \dots, y_n, \beta_j)$, which is calculated with the product of the probability distribution of the data.

The form to estimate the posterior distribution is explained as follows. The prior distribution times the likelihood equals the posterior distribution. Then, the predictive distribution results from an integral of the distribution of the variable to be predicted times the posterior (Gill, 2007). Then, the forecast values are calculated with this function.

Bayesian inference requires the use of prior information for the parameter(s) and a probability distribution for the data, but there is often insufficient information, which is the reason that expert

knowledge is sometimes required. Valencia et al. (2015) presented a review of the use of dynamic demands in inventory planning, and they mention the application of Bayesian analysis in the ARIMA, state-space models, and regression models, among others.

2.2. Forecast combination

The combination establishes a new forecast based on individual forecasts of other models. As is explained in Chan et al. (1999), the forecast combinations begin with Bates and Granger (Weiss & Roetzer, 2016), starting from two individual forecasts. The authors show that an adequate linear combination of these two sets can yield a better result than that of the individuals because they found a lower variance error in the combination. This study has also motivated many other works with linear and nonlinear combinations (Chen, 2011).

The classic form of the Bates and Granger combination (BG) is found using an R software package called *GeomComb* (Weiss & Roetzer, 2016), which uses the weights equation given by (1).

$$W_i^{BG} = \frac{\hat{\sigma}^{-2}(i)}{\sum_{i=1}^k \hat{\sigma}^{-2}(i)} \quad (1)$$

where $\hat{\sigma}^{-2} = MSE$, and the final combined forecast \hat{f} is given by (2).

Moreover, the combined final forecast \hat{f} will be given by the following:

$$\hat{f} = \sum_{i=1}^k W_i^{BG} \hat{Y}_i \quad (2)$$

\hat{Y}_i is the vector of individual forecasted data, and k is the number of models used in the combination.

2.3. Heuristic and metaheuristic algorithms to perform forecast combination

Metaheuristic approaches facilitate the combination of different techniques, such as Tabu list with the Nelder Simplex (Chelouah & Siarry, 2005; Valencia, González, & Cardona, 2014). Guerrero et al. (2016) present a comparison among three techniques, including ant colony optimization (ACO), the genetic algorithm (GA) and the evolutionary programming (EP) metaheuristic, and they design an algorithm applied to an inventory problem; the authors obtain good performance with the ACO technique.

The evolutionary programming metaheuristics are a class of algorithms based on iterative evaluation of a large set of solutions that are continuously improved through local and global optimization techniques. According to Allahverdi and Al-Anzi (2006), an example of this class is the GA, but it has differences from EP. For example, EP does not use binary variables, but it combines the

populations of parents and sons (Guerrero et al., 2016). These algorithms have a simulation exploration that will be applied to the combination process of this research and find the appropriate weights to generate the combined forecast.

2.4. *Linear regression model*

The linear model (LM) is based on a univariate continuous response, and it uses the `lm` function in the R software package in this research (R Core Team, 2017). The linear model for the time series is also considered econometric because it uses chronological variables, and the covariables are incorporated as autoregressive endogenous, as are others with time dependence (Bowerman, Koehler, & O'Connell, 2007; Caridad-y-Ocerin, 1998); this research considers equation (3).

$$z_t = \beta_0 + \dots \beta_k z_{t-k} + \dots \beta_i X_i + \varepsilon \quad (3)$$

This permits one to analyse the relations among the response z_t with covariables in the dependent variables of a lag of the series z_{t-1} , due to the autocorrelation on the time series simulation, the time and the indicators for the seasonality.

2.5. *Exponential smoothing*

Bowerman et al. (2007) illustrate simple exponential smoothing for the fitting process; here, the alpha parameter α is a constant between 0 and 1 and is susceptible to change according to an optimization of the sum of squared errors of prediction (SSE). This model facilitates the univariate estimation according to equation (4).

$$\hat{Y}_t = \alpha Y_t + (1 - \alpha) \hat{Y}_{t-1} \quad (4)$$

where

- \hat{Y}_t = forecast for the next period
- α = smoothing constant
- Y_t = real value of the time series for period t
- \hat{Y}_{t-1} = forecast of t-1 period

2.6. *Autoregressive integrated moving average (ARIMA) models*

Traditional forecasting models have been used for the decision-making process in production enterprises. ARIMA models were developed in the seventies by George Box and Gwilym Jenkins and were used by other authors, such as Meinhold and Singpurwalla (1983); accuracy comparisons

among different models have also been performed (Wang & Hu, 2015). A variation of these models that incorporates the characteristics from the endogenous response is the seasonal ARIMA (SARIMA) model, but these models require large amounts of data to perform a correct estimation. An ARIMA model will be used in this work, according to the Wang and Hu (2015) formulation, and with the arima function provided in the R software package.

2.7. Generalized linear mixed model (GLMM)

GLMM is a classical model that considers a fixed and a random component, such as that expressed in the equation (5), that presents both components for every individual i -th (Gómez-Restrepo & Cogollo-Flórez, 2012).

$$y_i = x_i\boldsymbol{\beta} + z_i\alpha_i + \varepsilon_i \quad (5)$$

Here, \mathbf{y}_i is the response vector for i -th individual, $\boldsymbol{\beta}$ is a parameter vector $p \times 1$ corresponding to the fixed effects, \mathbf{x}_i is the covariables matrix, α_i is the i -th individual effect, \mathbf{z}_i are the covariables of the random effect, and ε_i is intra-individual error. The package lme4 from the R software package estimates this model. The GLMM performs an estimation that facilitates the exploration of the behaviours of the multivariate responses for the application of dairy products, but not for the simulated data due to the continuous nature of the simulated response.

2.8. Bayesian regression model (BRM)

This is a regression model with a univariate natured response. The BRM has very similar equations to the classical regression model, but its response forecasts use a predictive Bayesian distribution. The BRM for this research was shown to yield good performance in Valencia (2016) compared to that of other models, such as the classical regression, ARIMA, exponential smoothing and the Bayesian dynamic linear model, based on a simulation process. The process to build the BRM begins with assigning the normal as the priori distribution (an informative distribution) for the parameter vector $\boldsymbol{\beta}$, but it constitutes a joint distribution with the standard deviation parameter. The normal is also the distribution for the data. The model follows a dynamic updating process for the initial parameter vector $\boldsymbol{\beta}_0$ as an innovation process, but the initial variance is fixed to $\tau_0 = I/\sigma$.

The product of the multiple normal distributions of the independent responses generates the likelihood function of the data, as shown in equation (6); the a priori distribution of the normal for the $\boldsymbol{\beta}$ parameter vector is shown in equation (7), and the posteriori distribution is shown in (8). The distributions are obtained after the product of the a priori times the likelihood and the algebraic process. Here, $A = \boldsymbol{\beta}'_0\tau_0\boldsymbol{\beta}_0 + Y'Y$, and $\tilde{\boldsymbol{\beta}} = (X'X + \tau_0)^{-1}(X'X\hat{\boldsymbol{\beta}} + \boldsymbol{\beta}_0\tau_0)$.

$$L(y_t|y_0, \beta) \propto \tau^T \exp^{-\frac{\tau}{2}(Y-X\beta)'(Y-X\beta)} \quad (6)$$

$$\xi(\beta, \tau) \propto \tau \tau_0 \exp^{-\frac{\tau \tau_0}{2}(\beta - \beta_{oi})'(\beta - \beta_{oi})} \quad (7)$$

$$\xi(\beta, \tau | \beta_0, \tau_0, y_0, y_t) \propto \tau^{\frac{T+1}{2}} \exp^{-\frac{\tau}{2}[A(\beta - \tilde{\beta})' A^{-1}(X'X - \tau_0)(\beta - \tilde{\beta}) + 1]} \quad (8)$$

As previously stated, the predictive distribution is the integral of the product between the posterior and future distribution for the data (Valencia, 2016). The final result is presented in equation (9):

$$f(Y_+ | y_0, Y) = \left[((Y_+ - Y_n)' A^{-1} (Y_+ - Y_n) + 1) A \right]^{\frac{T+4}{2}} \quad (9)$$

The mean is $Y_n = X_+ \tilde{\beta}$, and it has v degrees of freedom. The variance is given by (10):

$$V = \frac{v}{v-2} A = \frac{v}{v-2} (Y_+ - Y_n)' (Y_+ - Y_n) \quad (10)$$

Finally, the forecast combination process includes the individual models that are estimated for simulated time series, a process that constitutes one of the principal differences from Valencia's work since there is not a combination application.

2.9. First comparative approach

The designed process creates functions using the R programming language by following the steps of Figure 1. The process starts by reading the data.

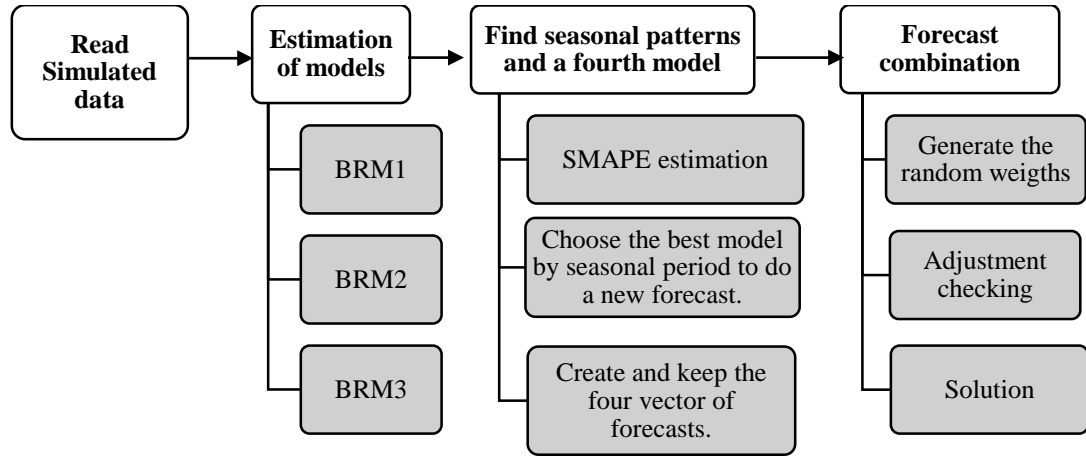


Figure 1. Second process. The selection of the best model for every seasonal period.

The simulated data consider a seasonal pattern and autocorrelation with the order one. Then, the seasonal variables are the trigonometrical $\sin\left(\frac{2\pi t}{L}\right)$ and the indicators I_t . BRM1, BRM2, BRM3, are models according to the equations (11), (12), (13), respectively.

$$1. \quad y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 \text{Sin}\left(\frac{2\pi t}{L}\right) + \sum_{i=1}^{L-1} \gamma_i I_i \quad (11)$$

$$2. \quad y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 t + \sum_{i=1}^{L-1} \gamma_i I_i \quad (12)$$

$$3. \quad y_t = \beta_0 + \beta_1 y_{t-1} + \sum_{i=1}^{L-1} \gamma_i I_i \quad (13)$$

where y_{t-1} is a lag with order one of the time series, t is the time, I_i is an indicator variable according to the seasonal periods, β_0 is the intercept for every model, β_i ($i=1, 2$) are the effects applied to specific covariables, and γ_i are the effects applied to indicator variables. The variable selection process considers Valencia's (2016) equations, which exhibited good performance in representing the seasonal patterns.

The BRM estimation applies for every time series simulated, as was illustrated in Figure 1, and uses every one of the three equations proposed (11, 12, and 13). After this, the algorithm finds the fourth model according to the selection of the best model for every seasonal period with the minimum SMAPE value. This algorithm was designed using the R programming language (R-Development-Core-Team, 2014). The last step is a summary for the EVOL process, explained as follows.

2.9.1. Evolutionary programming metaheuristic proposal (EVOL)

The combination method designed in this research starts by reading the forecasted data from every individual model. Then, a simulation process creates a population of five hundred (500) parents and two hundred fifty (250) sons of weights. Every vector of parents has four components that are simulated by using a trapezoidal distribution with no binary variables, and the sons are mixtures of every two parents (2 for 1). The final population consists of seven hundred fifty (750) individuals, but it adds a selection of the best two vectors of the population in order to create new sons for every population. If the algorithm finds negative vectors of adjusted values, it creates a mutation of new weights with the normal distribution to estimate the positive adjusted and forecasted values. The combination forecasts use equation (14) to calculate the linear combination among the individual forecasts and the weights.

$$Y_i = \sum_{i=1}^n w_i \hat{Y}_i \quad (14)$$

A selection of the best weights is performed to produce another son, and the new vectors of the weights are aggregated to the population of all the weights. The algorithm is repeated iteratively until it reaches the minimum possible SMAPE with a convergence criterion.

After finding the best Y_i with the minimum SMAPE values of the forecasts, it generates all the results. The algorithm repeats the simulations of the weights by estimating the adjusted and forecasted values of the time series and the SMAPE of the respective type of values. Every time, the algorithm compares the solution and provides the best possible solution.

The SMAPE indicator is the measure to evaluate the different models that will be used, searching for the minimum possible value, equation (13). Here, F_t is the observed value for every product in the time t , A_t is the forecasted value by the respective model or the forecast combination, and n is the number of data.

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{((A_t + F_t)/2)} \quad (13)$$

2.9.2. Bates and Granger combination (BG)

The BG is found in an R software package called *GeomComb* (Weiss & Roetzer, 2016), which uses the individual forecasts and the weights as it was explained before, where $\hat{\sigma}^{-2} = MSE$, and the final combined forecast \hat{f} is given by (2).

2.10. Second comparative approach

The second approach is a programming algorithm designed in R that compares the univariate LM, ARIMA model, exponential smoothing (SE), the Bayesian regression model (BRM) (Valencia-Cárdenas, 2016), and the result of the forecast combination Bates and Granger used in Valencia, Osorno, and Salazar (2017) with the novel EVOL combination designed in this research.

The algorithm comprises the following steps.

- Partition the response for every simulation of the time series, cut $n-k$ periods for an adjustment, and forecast the other k periods.
- Estimate the individual models for the simulated data: the LM, ES, ARIMA and BRM. Keep the adjusted, forecasted data for every model and their respective SMAPE values.
- Estimate the EVOL combination with the LM, ES, ARIMA, and BRM individual adjustment and forecasts.
- Estimate the Bates and Granger combination technique (BG) with the *GeomComb* package (Valencia et al. 2017) and estimate SMAPE.

- Compare and select the best model for every time series among the individual models and the BG and EVOL combinations according to the minimum SMAPE value of the forecasted values.

2.11. Time series simulation

The simulation of the time series permits an efficiency comparison of the algorithms. The simulation applies here for data that do not follow the normal distribution and that can represent a non-stationary process that happens in many real situations. The distribution for the simulation is a skewed normal (Valencia & Bedoya, 2014) with the skew parameter $\lambda=2$ as a low level and $\lambda=30$ as a high level. The simulation of the three time series introduces random fluctuations, seasonal behaviours, and significant autocorrelation. Figures 2 and 3 shows the behaviours of one case simulation and the respective autocorrelation.

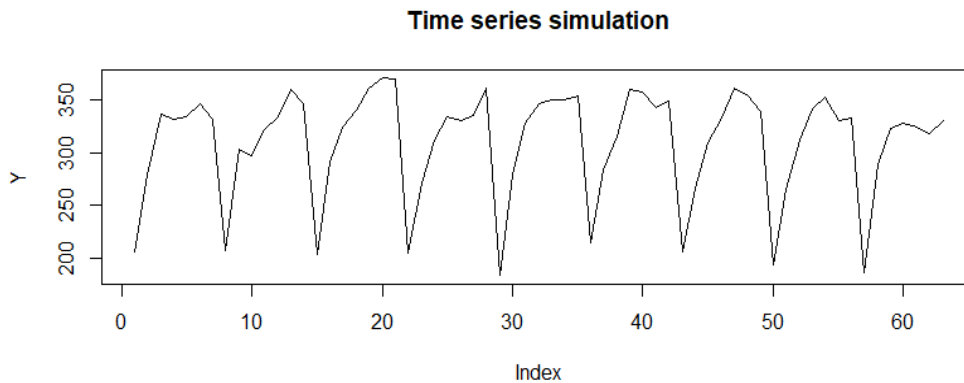


Figure 2. Simulated time series. The X-axis is the time, and the Y-axis is the values of the time series.

Figure 3 shows the autocorrelation function (ACF) and the partial autocorrelation function (PACF), where the blue dotted lines represent the normal bands to test if the autocorrelation values are significant in the case that they are outside of them. In the simulated case, the vertical lines outside the blue dotted horizontal exhibit significant dependence in a seasonal order since the autocorrelations are outside the bands every seven periods.

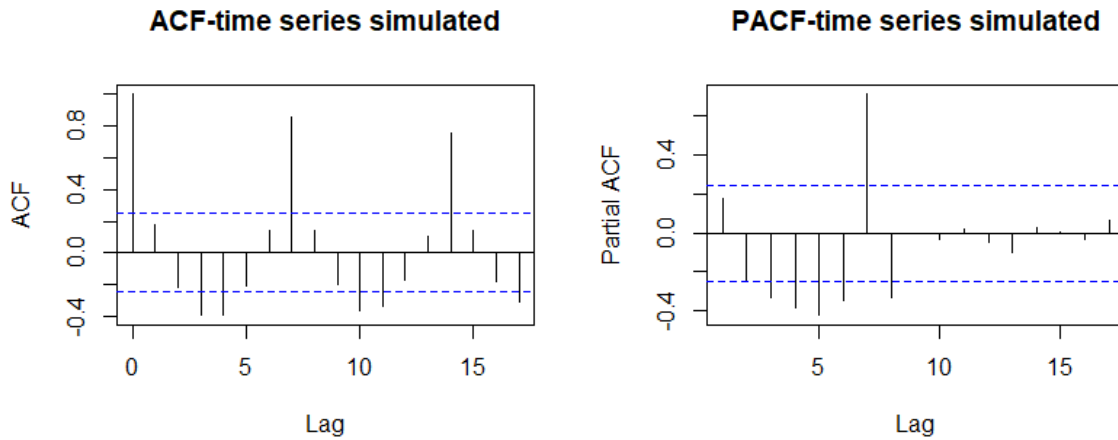


Figure 3. Autocorrelation of the time series simulation. The X-axis is the lag order.

Table 1. Autocorrelation tests for the simulated time series with the skewed normal distribution.

Box-Pierce test		
data: series 1		
X-squared = 68.61	df = 7	p-value = 2.82e-12
data: series 2		
X-squared = 71.713	df = 7	p-value = 6.652e-13
data: series 3		
X-squared = 39.753	df = 7	p-value = 1.404e-06

Table 1 presents the significance of the autocorrelation of the simulated time series. Since the p-values are less than 5%, the null hypothesis about no correlation is rejected. It means that the dependence simulation process of time series is correct.

Figure 3 also shows that the time series are non-stationary and exhibit high variations. The skewed normal distribution was fixed to have two skew parameters, 2 and 30, in order to prove the behaviour of the combination proposal in a low and a high value of the skewedness. The other two simulated time series data have similar behaviours.

The seasonal behaviour and autocorrelation demonstrated before with the ACF and PACF show that the explanatory variables to incorporate in the BRM are the one lag order of the same time series, y_{t-1} , and an indicator or dummy variable for the seasonal periods, I_i .

3. Results

3.1. Results of the first comparative approach

After simulating the time series data once with skewed normal distribution, and after calculating the forecasts for the BRM, the BG combination, and the EVOL combination, the EVOL showed a better performance according to the comparison of the SMAPE values of the forecasts (Tables 2 and 3).

Table 2. SMAPEs for one replication of a skewed normal time series with skew=2.

Time Series	BRM	BG	EVOL Combination
1	3.85%	3.78%	3.49%
2	8.34%	8.54%	8.24%
3	7.99%	7.97%	7.86%

According to the results, the proposed evolutionary algorithm EVOL that uses the seasonal information of the time series and that performs a simulation for the weights with the trapezoidal distribution searches for the best solution after creating the population of parents and sons. The results of that algorithm are better than the individual results of the Bayesian regression model (BRM) and the Bates and Granger combination (BG) in almost all the cases, as indicated in Tables 2 and 3, when the simulation was run once.

Table 3. SMAPEs for one replication of the skewed normal time series, with skew=30.

Time Series	BRM	BG	EVOL Combination
1	3.96%	3.96%	3.87%
2	4.00%	4.03%	3.98%
3	10.11%	9.98%	9.93%

In Figure 4, the blue dotted line is the EVOL combination forecast, the black line corresponds to the real data, and the red represents the BG combination. This figure shows that the EVOL combination is close to the direction of the data.

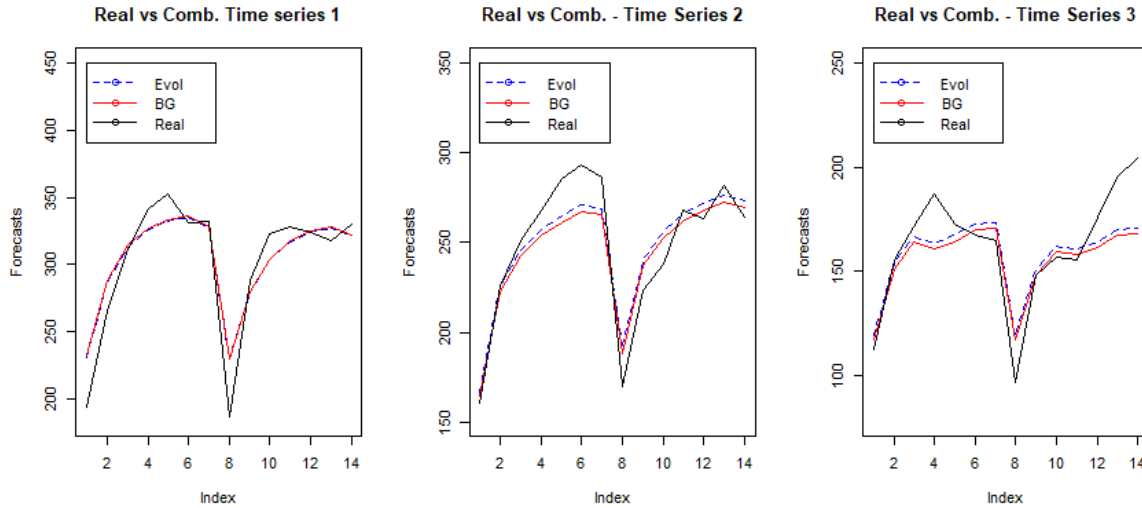


Figure 4. Simulated data vs combination techniques for one run of the time series with the SN and skew=2.

The simulation is applied 2000 times, and all the techniques are compared; the best performance is obtained with the EVOL technique (Table 4). The selection was deduced since EVOL is found as the best 100% of the times for series 2, and the lower quantity was found 94% of the times when EVOL was better than BG. The comparison of the respective SMAPE forecasting values was performed using a population of 100 vectors of weights and a cut of 2 seasonal periods of forecast (14 periods).

Table 4. Frequency of EVOL selection compared to that of BRM and BG.

Model	Distribution	Series 1	Series 2	Series 3
BRM	Skewed Normal S=2	98.78%	100%	100%
	Skewed Normal S=30	98.78%	100%	100%
BG	Skewed Normal S=2	100%	100%	94%
	Skewed Normal S=30	99%	100%	100%

Table 5 presents the differences between the proposed EVOL algorithm and the BRM model on the first line and between EVOL and the Bates and Granger combination on the second line.

Table 5. Differences in SMAPE values.

	Mean of differences		
EVOL Better than BRM	3.19%	3.51%	0.60%
EVOL Better than BG	3.96%	4.74%	2.74%

The proposed combination yields more accurate results than those of the BG combination for more distances than those for the BRM. This result means that despite the combination being better designed, the BRM still provides a good alternative for forecasting.

3.2. Results of the second comparative approach

The forecasted SMAPE values for the LM, ES, ARIMA, BRM, and the combination BG models and EVOL for the simulated time series are presented in Table 6.

Table 6. SMAPE comparisons among individual LM, ES, ARIMA, BRM, the BG and the EVOL combination.

	LM	ETS	AR	BRM	BG	EVOL
Time series 1	8.15%	10.30%	10.30%	3.05%	3.69%	2.44%
Time series 2	7.86%	15.51%	15.51%	7.08%	5.88%	5.03%
Time series 3	7.08%	12.60%	12.60%	4.01%	4.70%	3.44%

Table 6 indicates that the SMAPE values for the EVOL combination are less than those for the other models, even the combination BG. The BRM shows also an acceptable performance that is not far from that of the EVOL combination.

After 1000 repetitions of the simulations, the EVOL presents a better performance than those of the others by taking the average SMAPEs of the three models. EVOL is chosen as the best 80% of the time, the BG combination was the best 20% of the time, and no other model yielded a better SMAPE mean, as can be observed in Table 7.

Table 7. Selection of the best method with the minimum mean of SMAPE.

COMB_BG	EVOL
20%	80%

3.3. *Application of EVOL to a real case*

A Colombian dairy company provided the data that consists of three time series data with daily sales, non-stationary behaviours and seasonal behaviours. Table 8 presents the Box and Pierce test about the autocorrelation of the time series, thus confirming that the behaviours are not stationary.

Table 8. Autocorrelation tests for the real time series data.

Box-Pierce test		
data: Y1		
X-squared = 172.52	df = 7	p-value < 2.2e-16
data: Y2		
X-squared = 350.98	df = 7	p-value < 2.2e-16
data: Y3		
X-squared = 269.51	df = 7	p-value < 2.2e-16

Figure 5 shows the autocorrelation values in the ACF and PACF functions for the three time series data. Figure 5 confirms that there is a seasonal pattern of order 7 for every time series because of the values outside the limits.

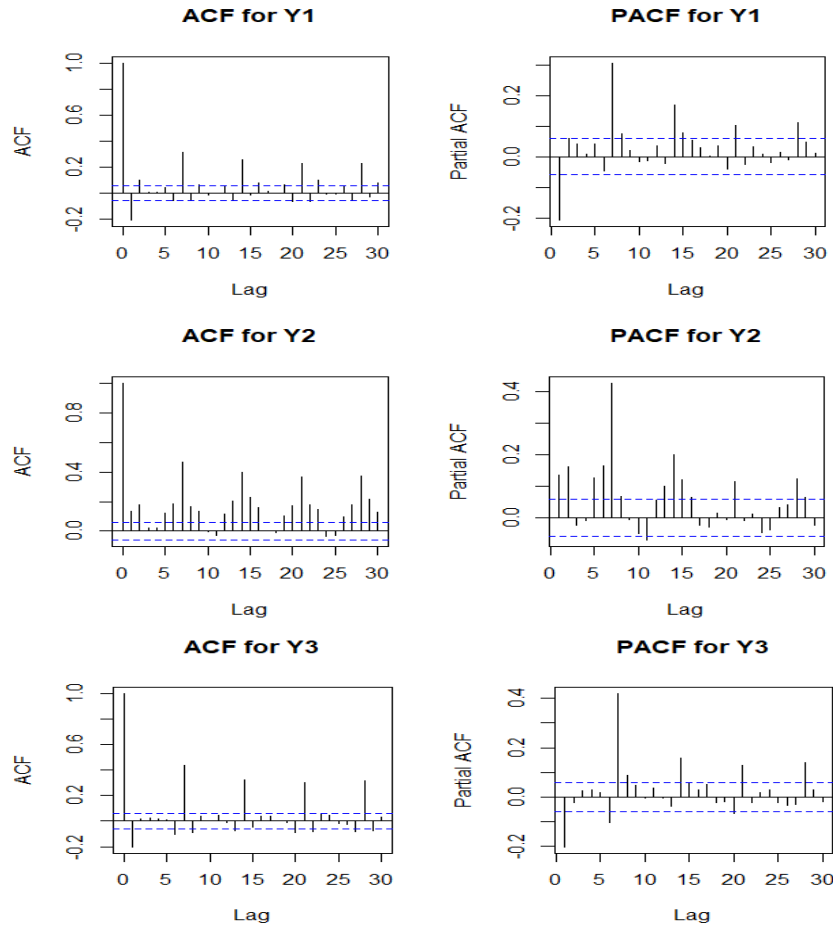


Figure 5. ACF and PACF functions for the real data. Y1, Y2 and Y3 are the sales of three dairy products.

The weights that can be found for every model and time series simulation are reported in Table 9.

Table 9. Weights for the final combination forecasts for the real case.

	Time series		
	1	2	3
Model 1	0.4199	0.02151	-0.00217
Model 2	0.21348	1.04577	0.343396
Model 3	0.4644	0.02201	-0.0796
Model 4	-0.1125	-0.06351	0.71161

Table 10 presents the results after the estimation of every model, beginning with a generalized linear mixed model (GLMM), a conventional model estimated with the R software package that uses the

function *glmer*. However, the GLMM's performance was the worst of all the models, and the EVOL combination technique yielded the best accuracy.

Table 10. SMAPE comparisons among individual BRM with the BG and EVOL combination for the real case.

TIME SERIES	MODEL						
	GLMM	BRM 1	BRM 2	BRM 3	MODEL 4	BG	EVOL
1	26.08%	24.023%	22.089%	22.305%	22.481%	22%	22.022%
2	26.6%	27.668%	33.513%	22.803%	27.554%	25.84%	22.156%
3	21.6%	19.980%	21.300%	20.678%	20.508%	20.37%	19.401%

The EVOL combination performs better than the other models for both the simulations and also for the real case. Despite this, the BG is still a good alternative to perform forecasts. The GLMM has generally poor performance. This result can lead us to infer that the designed EVOL technique is a very good alternative to elaborate forecasts, and therefore its use is recommended to improve predictions.

The algorithm provides multiproduct forecasts and adjusted values that facilitate forecasting in many fields, especially for planning sales and inventories in industry.

4. Discussion

The EVOL combination technique yielded the best performance for simulated data and the real case, thereby confirming similar results from authors such as Cang and Yu (2014), who express that combination techniques improve individual forecasts.

This result is important given that the simulated data were not stationary because the data exhibited seasonal patterns, non-normal fluctuations and autocorrelation, which are components that not all statistical models consider in their theoretical assumptions. Additionally, the combination is useful for multiproduct or multi-item approaches, thus facilitating the forecasting purposes and their probable inderence in planning processes or financial aspects. These aspects make the proposed EVOL a methodological approach that facilitates the difficult problem of forecasting.

The Bates and Granger combination and Bayesian regression are also adequate alternatives because their results do not differ so much from the EVOL result. These approaches generate other possibilities for multi-item forecasts.

The performance of the BRM is also good to estimate predictions for multiple variables for both the simulated time series data and for the real case. In this sense, the BRM also exhibited some of the advantages of Bayesian inferences. In this sense, other Bayesian alternatives can also be tested, or there can be a change in the distribution parameters, as is stated in Valencia (2016).

A novel algorithm based on evolutionary programming (EVOL) designed in the R programming language provides a more accurate forecast technique useful to many kinds of variables in addition to a multiple set of them. The algorithm can also be reproduced in any other programming language.

Acknowledgements

Thanks are given to the Colciencias scholarship (567-2012), which supported obtaining a Doctorate in Engineering-Industry and Organizations at the Universidad Nacional de Colombia, Sede Medellín.

References

- Agarwal, D. K., Silander, J. A., Gelfand, A. E., Dewar, R. E., & Mickelson, J. G. (2005). Tropical deforestation in Madagascar: Analysis using hierarchical, spatially explicit, Bayesian regression models. *Ecological Modelling*, 185(1), 105–131. <http://doi.org/10.1016/j.ecolmodel.2004.11.023>
- Allahverdi, A., & Al-Anzi, F. S. (2006). Evolutionary heuristics and an algorithm for the two-stage assembly scheduling problem to minimize makespan with setup times. *International Journal of Production Research*, 44(22), 4713–4735. <http://doi.org/10.1080/00207540600621029>
- Barrow, D. K., & Kourentzes, N. (2016). Distributions of forecasting errors of forecast combinations: Implications for inventory management. *International Journal of Production Economics*, 177, 24–33. <http://doi.org/10.1016/j.ijpe.2016.03.017>
- Grothendieck, G. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <http://doi.org/10.18637/jss.v067.i01>

- Bowerman, B., Koehler, A., & O'Connell, R. (2007). *Pronósticos, series de tiempo y regresión: un enfoque aplicado*. Mexico, DF: CENCAGE Learning.
- Cang, S., & Yu, H. (2014). A combination selection algorithm on forecasting. *European Journal of Operational Research*, 234(1), 127–139. <http://doi.org/10.1016/j.ejor.2013.08.045>
- Caridad y Ocerin, J. M. (1998). *Econometria: Modelos Econométricos y series temporales*. (R. S.A., Ed.).
- Chahkoutahi, F., & Khashei, M. (2017). A seasonal direct optimal hybrid model of computational intelligence and soft computing techniques for electricity load forecasting. *Energy*, 140, 988–1004. <http://doi.org/10.1016/j.energy.2017.09.009>
- Chan, C. K., Kingsman, B. G., & Wong, H. (1999). The value of combining forecasts in inventory management ± a case study in banking. *European Journal of Operational Research*, 117, 199–210.
- Chelouah, R., & Siarry, P. (2005). A hybrid method combining continuous tabu search and Nelder–Mead simplex algorithms for the global optimization of multim minima functions. *European Journal of Operational Research*, 161(3), 636–654. <http://doi.org/10.1016/j.ejor.2003.08.053>
- Chen, K.-Y. (2011). Combining linear and nonlinear model in forecasting tourism demand. *Expert Systems with Applications*, 38(8), 10368–10376. <http://doi.org/10.1016/j.eswa.2011.02.049>
- Ferragina, A., de los Campos, G., Vazquez, A. I., Cecchinato, A., & Bittante, G. (2015). Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. *Journal of Dairy Science*, 98(11), 8133–8151. <http://doi.org/10.3168/jds.2014-9143>
- Gao, W., Sarlak, V., Parsaei, M.-R., & Ferdosi, M. (2017). Combination of Fuzzy Based on a Metaheuristic Algorithm to Predict Electricity Price in an Electricity Markets. *Chemical Engineering Research and Design*. <http://doi.org/10.1016/j.cherd.2017.09.021>
- Gill, J. (2007). *Bayesian methods: A social and behavioral sciences approach*. United States of America: Chapman & Hall. 2nd. Ed.
- Gómez-Restrepo, J., & Cogollo-Flórez, M. R. (2012). Detection of Fraudulent Transactions Through a Generalized Mixed Linear Models. *Ingeniería Y Ciencia*, 8(16), 221–237.
- Guerrero, L., Gómez, J. D., Zapata, D., & Valencia, M. (2016). Comparación de Tres Metaheurísticas para la Optimización de Inventarios con Estimación de Demanda, 51–68.
- Hsiao, C., & Wan, S. K. (2014). Is there an optimal forecast combination? *Journal of Econometrics*, 178(PART 2), 294–309. <http://doi.org/10.1016/j.jeconom.2013.11.003>

- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, 55(9), 2579–2589. <http://doi.org/10.1016/j.csda.2011.03.006>
- Kociecki, A., Kolasa, M., & Rubaszek, M. (2012). A Bayesian method of combining judgmental and model-based density forecasts. *Economic Modelling*, 29, 1349–1355. <http://doi.org/10.1016/j.econmod.2012.03.004>
- Kolassa, S. (2016). Evaluating predictive count data distributions in retail sales forecasting. *International Journal of Forecasting*, 32(3), 788–803. <http://doi.org/http://dx.doi.org/10.1016/j.ijforecast.2015.12.004>
- Li, G., Shi, J., & Zhou, J. (2011). Bayesian adaptive combination of short-term wind speed forecasts from neural network models. *Renewable Energy*, 36(1), 352–359. <http://doi.org/10.1016/j.renene.2010.06.049>
- Meinhold, R. J., & Singpurwalla, N. D. (1983). Understanding the Kalman Filter. *The American Statistician*, 37(2), 123–127.
- Melo, L. F., Loaiza, R. A., & Villamizar-Villegas, M. (2016). Bayesian combination for inflation forecasts: The effects of a prior based on central banks' estimates. *Economic Systems*. <http://doi.org/10.1016/j.ecosys.2015.11.002>
- Miller, J., Berry, C., & Lai, C. (2007). A comparison of alternative forecasting strategies for multistage production-inventory systems. *Decision Sciences*, 7(4), 714–724.
- Nowotarski, J., Liu, B., Weron, R., & Hong, T. (2016). Improving short term load forecast accuracy via combining sister forecasts. *Energy*, 98, 40–49. <http://doi.org/10.1016/j.energy.2015.12.142>
- Petris, G., Petrone, S., & Campagnoli, P. (2009). *Dynamic Linear Models with R*. Retrieved from <http://www.springer.com/statistics/statistical+theory+and+methods/book/978-0-387-77237-0>
- R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- R-Development-Core-Team. (2014). *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/>
- Salas, D. (2008). *Modelos de mezcla de distribuciones α -estables*. (Editorial de la Universidad de Granada, Ed.). Granada.
- Sermpinis, G., Stasinakis, C., Theofilatos, K., & Karathanasopoulos, A. (2015). Modeling, forecasting and trading the EUR exchange rates with hybrid rolling genetic algorithms -

- Support vector regression forecast combinations. *European Journal of Operational Research*, 247(3), 831–846. <http://doi.org/10.1016/j.ejor.2015.06.052>
- Silver, E. (2004). An overview of heuristic solution methods. *Journal of the Operational Research Society*, 55(9), 936–956. <http://doi.org/10.1057/palgrave.jors.2601758>
- Simchi-Levi, D., Kaminski, P., & Simchi-Levi, E. (2008). *Designing and Managing the Supply Chain*. (McGraw-Hill, Ed.) (3rd ed.). New York: McGraw-Hill.
- Valencia, M., & Bedoya, A. (2014). Prueba de sesgo sobre rendimientos financieros en el mercado colombiano. *Lecturas de Economía*, (80), 79–102.
- Valencia, M., Díaz, F. J., & Correa, J. C. (2015). Planeación de inventarios con demanda dinámica . Una revisión del estado del arte. *Revista DYNA*, 82(190), 182–191. <http://doi.org/http://dx.doi.org/10.15446/dyna.v82n190.42828>
- Valencia, M., González, D., & Cardona, J. (2014). Metodología de un modelo de optimización para el pronóstico y manejo de inventarios usando el metaheurístico Tabú. *Revista Ingeniería*, 24(1), 13–27. Retrieved from <http://www.revistas.ucr.ac.cr/index.php/ingenieria/article/view/13771>
- Valencia, M., Osorno, V., & Salazar, J. C. (2017). Comparative of forecasting models: Classical, Bayesian and Combination Techniques. *Revista Facultad de Ciencias*, 6, 124–140. <http://doi.org/https://doi.org/10.15446/rev.fac.cienc.v6n2.66085>
- Valencia-Cárdenas, M. (2016). Dynamic Model for the Multiproduct Inventory Optimization with Multivariate Demand. Universidad Nacional de Colombia, Sede Medellín. Retrieved from <http://www.bdigital.unal.edu.co/51516/1/43616826.2015.pdf>
- Valencia-Cárdenas, M. (2016). *Dynamic Model for the Multiproduct Inventory Optimization with Multivariate demand*. Universidad Nacional de Colombia.
- Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis - Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts usi. *Energy*, 93, 41–56. <http://doi.org/10.1016/j.energy.2015.08.045>
- Weiss, C., & Roetzer, G. R. (2016). Package “GeomComb .”
- Zotteri, G., & Kalchschmidt, M. (2007). A model for selecting the appropriate level of aggregation in forecasting processes. *International Journal of Production Economics*, 108(1-2), 74–83. <http://doi.org/10.1016/j.ijpe.2006.12.030>

Combination Forecasting algorithm with cluster Analysis, in a simulation scenario.

Marisol Valencia Cárdenas¹, Juan Carlos Correa Morales²

- 1- Professor. I.U. Tecnológico de Antioquia, Administrative and Economic Faculty, Medellín, Antioquia, Colombia. mvalencia@unal.edu.co, solmarival@gmail.com. <https://orcid.org/0000-0003-3135-3012>
- 2- Associate Professor. Universidad Nacional de Colombia, Medellín, Statistis. Medellín, Antioquia, Colombia. jccorrea@unal.edu.co

Abstract

Forecasting with combination of individual models has improved the results or future predictions, using one approach like the search of weights as the decision variables of a programming model, coordinated with techniques as statistical bases applied to a database. The algorithm includes a search based on different probability distributions, to generate the decision variables optimizing the error indicator for a forecasting combination process and cluster analysis, to include other quantitative parameter to improve predictions.

Keywords: Forecast Combination, metaheuristics.

1. Introduction

Statistical and mathematical models are helpful for decision making process, for example in forecasting, optimization strategies, or decision making process, reaching objectives as reducing errors in forecasting, reducing costs, among others (Riedl, Kaufmann, Zimmermann, & Perols, 2013; Valencia Cárdenas, Díaz Serna, & Correa Morales, 2015). Statistical modeling, mathematical procesing data are related to optimization processes, requiring information, as the strategies of Big Data Analysis do (Wolfert, Ge, Verdouw, & Bogaardt, 2017), using conectivity in order to facilitate the flow of information to share among partners involved in the operations relation. Stochastic models are common for many decision making processes, when probability distribution must be used, for example, industry (Sarimveis, Patrinos, Tarantilis, & Kiranoudis, 2008), or also, to do politics about the environment.

Also for pollution variables, researchers associate the relation among this and diseases, establishing reasons to increase politics against air contamination, and create some forecasts related (Dos-Santos, Constantino, & Lucio, 2018; Nascimento, Pompeo-Ferreira, & Cota, 2016)

Fluctuations and extreme dynamics in some random variables, frustrate the estimation for long trend forecasting affecting the performance of operations or decision making processes (Valencia-Cárdenas, Díaz-Serna, & Correa-Morales, 2016), generating mistakes around inferences for predictions. Constant terms are not enough to deal with many kind of dynamics caused by the high fluctuation of random variables (Fúquene, Álvarez, & Pericchi, 2015). None of these techniques have perfect fit, they fail in the goodness-of-fit or significance of variables and factors, perhaps influenced by the periodicity, the sample, seasonality, among others (Chen, 2011; Liu, Peng, Bai, Zhu, & Liao, 2014). Uncertainty in forecasts has been studied in the literature for many kind of variables, in especial, continuous time series data with autocorrelation or other components as the trend, seasonality, randomness, stationary process, among others, and according to these, the precision varies. Mistakes introduced by wrong decision-making in forecasting, causes problems in many areas, for example, in inventories coordination (Cai, Chen, Xiao, Xu, & Yu, 2013; Jedermann, Nicometo, Uysal, & Lang, 2018), but also, forecasting of air pollution has been of paramount importance recently.

This kind of forecast is the basis for taking pollution control measures, leading the topic to increase attention in accurate forecasting. The methods of air pollution forecasting consider different type of models, as classical statistical forecasting methods, artificial intelligence methods, and numerical. More recently, some hybrid models have been proposed, to improve accuracy, as other areas do (Bai, Wang, Ma, & Lu, 2018).

Improvement of forecasting can also be obtained with Bayesian Methods (Petris, 2010). Dynamic Linear Models consider variations in parameters across time, and other kind of theories for their estimations. Other models are Bayesian regression (Min & Zellner, 1993; Zellner, 1996), considering the relation among a response variable and covariables, with similar equation than the classical regression model (Ferragina, de los Campos, Vazquez, Cecchinato, & Bittante, 2015), with probability distributions that can have flexible modifications in the estimation process.

Efficiency of models is a high interest to forecasters, but it is not very common to find good accuracy when the response variables have special scale or behaviour (Wallström & Segerstedt, 2010). However, there are successful findings if it is used an approximation to the Normal distribution when the values have a high scale (Valencia, Vanegas, Correa, & Restrepo, 2017), a very known and used

distribution with many applications and accuracy results. Other lines conduce to find also efficiency in forecasting process, with statistical learning applications, that comprehends development process designing algorithms that can learn from the data (Bensoussan, Çakanyıldırım, Li, & Sethi, 2014; Bensoussan et al., 2014; Debnath & Mourshed, 2018; Puchalsky, Trierweiler, Pereira, Zanetti, & Coelho, 2018).

Forecast combination shows more accurate results than individual models (Barrow & Kourentzes, 2016), and also, because it can be also used for demand prediction, for example of food sales. Guo et al. (2017) affirm that there are two forms to make forecasts: one individual and the other, by combination of different models, and those aspects are applied by different authors (Hyndman, Ahmed, Athanasopoulos, & Shang, 2011; Miller, Berry, & Lai, 2007; Zotteri & Kalchschmidt, 2007). Combination can also be applied with the minimization of the error variance, or optimization of least squares, or Bayesian probabilities (Barroy, 2016), (Andersson & Karlson, 2007; Kociecki, Kolasa, & Rubaszek, 2012).

Melo et al. (2016) realizan una propuesta de combinación de pronósticos para tasas de inflación en bancos, usando una aproximación bayesiana para el banco central de Colombia, con estimaciones de la inflación. En sus resultados muestran que la combinación propuesta mejora los pronósticos individuales en cualquier horizonte de planeación. Dichas aproximaciones bayesianas de combinación de pronósticos están siendo utilizadas cada vez más (Bergman, Noble, Mcgarvey, & Bradley, 2017; Guo et al., 2017; Kociecki et al., 2012; Melo et al., 2016), pero no ha sido muy común el uso de modelos clásicos junto con bayesianos dentro de este tipo de mezclas, como se muestra en éste trabajo.

Forecast combinations (Barrow & Kourentzes, 2016; Hibon & Evgeniou, 2005), are recent alternatives created by the join of other individual models, as literature presents (L. Gao, 2015; W. Gao, Sarlak, Parsaei, & Ferdosi, 2018; Hsiao & Wan, 2014). These alternatives use individual forecasts in a linear combination equation, in order to produce a forecast, expected to generate one value better to the real one. Some authors have used also some special Metaheuristics (Gao et al., 2018) in the forecast combination, by the use of a set of operations directed to optimize an objective value, searching decision variables, by the use of methods as algorithms working on some statistical or mathematical language, as R (R Core Team, 2017). This paper establishes a comparison among different predicting models, a Bayesian Regression Model, and a forecasting combination model, designed by Bates and Granger, by using an algorithm designed in that program R (Weiss & Roetzer, 2016). Bayesian techniques and forecast combinations is also a recent method to improve forecasts (Andersson & Karlson, 2007; Melo et al., 2016, 2016; W. J. Wang & Xu, 2014).

The use of combinatorial techniques to make forecast have been growing due to their impact in precision (Kociecki et al., 2012; Nowotarski, Liu, Weron, & Hong, 2016), because they could provide better accuracy of predictions than other kind of individual models (Cang & Yu, 2014).

The rest of the document will be organized as follows. Section two addresses the methodologies, models and variables used, and a short explanation of the models. The designed algorithm applies to a study of case that consists in the estimation of the models Classical linear regression, exponential smoothing, ARIMA, a Bayesian regression model used in Valencia (2016) and a combination of such methods: Bates and Wranger, using R program. After this, the results of section 3 show the most relevant results. Finally, a discussion will be drawn.

2. Methods

The algorithm designed in R, compares the individual and univariate models: Linear Model (LM), Autorregressive Integrated Moving Average (ARIMA), Exponential Smoothing (SE), Bayesian Regression Model (BRM) (Valencia-Cárdenas et al., 2016) with the combination alternatives: the classical forecast combination by Bates & Granger (Weiss & Roetzer, 2016) using the R package called GeomComb, and TABUPRO metaheuristic designed in this document, using a process which is differenced in the cluster variable aggregated.

A summary of the process used is:

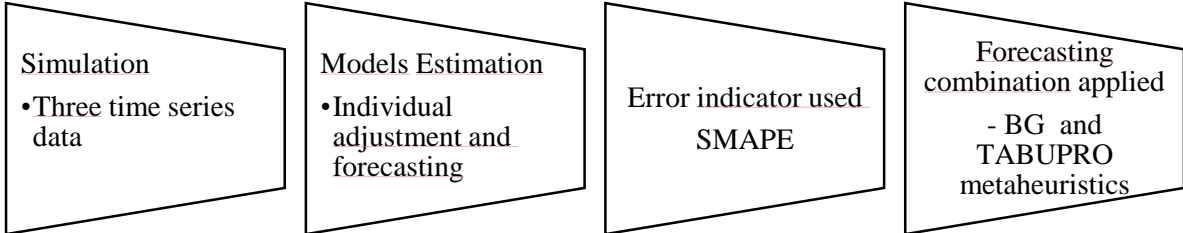


Figure 1. Summary explanation of the hole process.

2.3. Bayesian Regression Model (BRM)

The Bayesian Regression Model (BRM) works in univariate form, however Valencia-Cárdenas (2016) showed a better performance compared to linear mixed model, and better than a Bayesian dynamic linear model. In this work. the forecasts are estimated according with the next rules: For the BRM model, the process is: i) The Normal is the prior distribution for the parameters b and a non-informative for variance, being 1/s. ii) Fix Normal distribution for data. iii) Use Regression equation:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 t + \beta_3 \sin t + \sum_{i=4}^k \beta_i I_i$$

iv) The predictive distribution depends on a prior vector parameter β_0 that is updated for every period, this is T student distribution.

2.4. BG Combination

Bates and Granger (Bates & Granger, 1969), has designed in the R program, is based on the inverse of the RMSE indicator that every individual model has. The equations (2) and (3), show the application of the BG Combination.

$$\hat{f}_t = \sum_{i=1}^k W_i^{BG} \hat{Y}_t \quad (2)$$

$$W_i^{BG} = \frac{\hat{\sigma}^{-2}(i)}{\sum_{i=1}^k \hat{\sigma}^{-2}(i)} \quad (3)$$

Where: i is the i -th model, and $i=1,2, \dots, k$, there are the k models used ($k=4$). Y_i are the forecasts of the models. For every individual model, the SMAPE and a RMSE are also calculated and for combinations. In this work, the RMSE indicators of the forecasts will be used to compare.

2.5. Algorithm - Part I

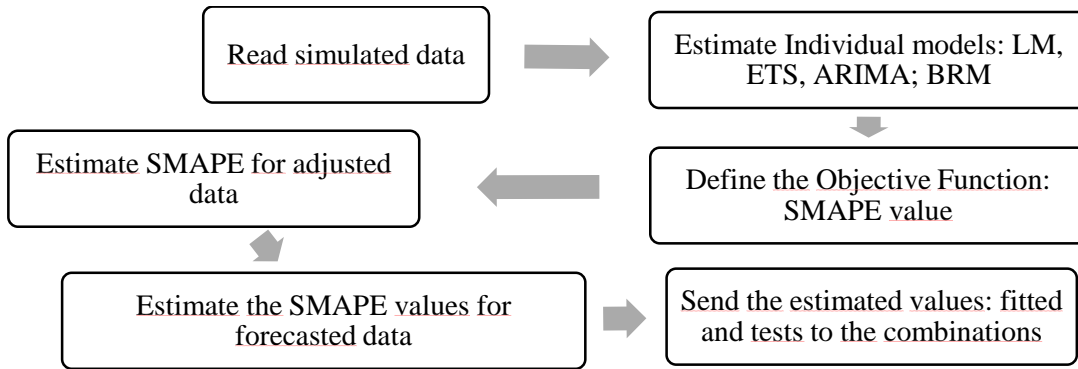


Figure 2. Part I- Algorithm.

2.6. Algorithm - Part II

The designed combination consists on a sum of weights per the predicted variables, according to equation (1):

$$\hat{f}_t = \sum_{i=1}^k W_i \hat{Y}_t \quad (1)$$

Where: i is the i -th model, and $i=1,2, \dots, k$. \hat{Y}_t are the forecasts of the individual models. The W_i are the decision variables to be found according to the simulation from the algorithm.

After this, the error indicator called SMAPE is calculated with the real and forecasted data, with: Y_t and \hat{Y}_t are, the vector of the real values, and the forecasted, according to equation (2):

$$SMAPE = \frac{1}{M} \sum_{T=1}^M \frac{|Y_t - \hat{Y}_t|}{(Y_t + \hat{Y}_t)/2} \quad (2)$$

Here M is the total of periods considered in the adjustment or the tests; Y_t and \hat{Y}_t are, the vector of the real values, and the adjustments.

Figure 3 shows a syntaxes of the process.

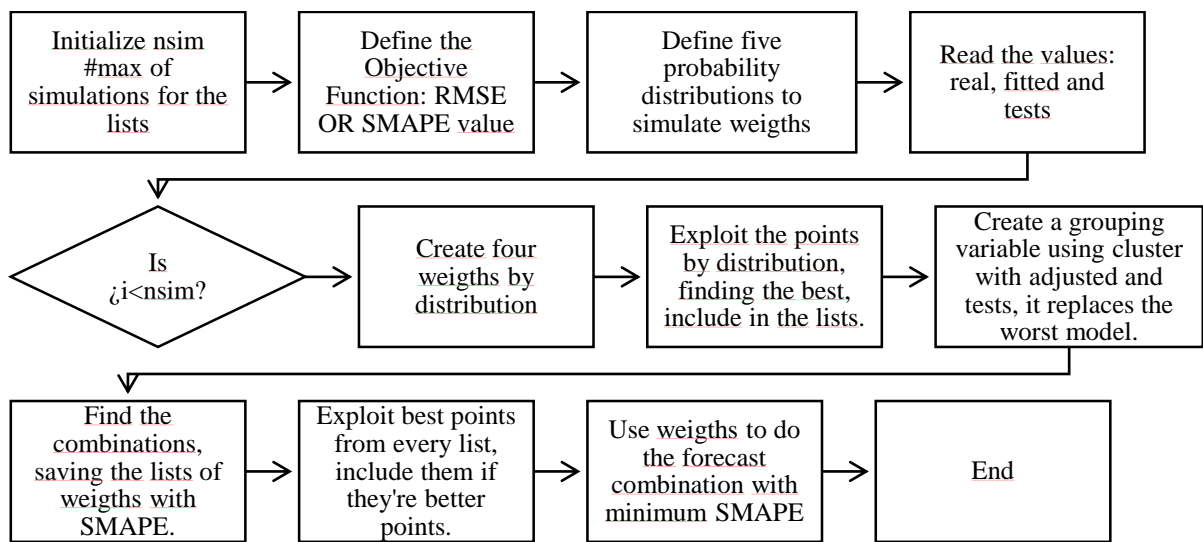


Figure 3. TABUPROC process explanation.

After SMAPE estimations, the algorithm aggregates a cluster grouping variable, applied to the adjusted values, and it eliminates the worst model. Then, it starts the exploitation points from every distribution simulation of the weights applied to the predicted data from every model, finding the best in order to incorporate them to the list.

3. Simulation results

Case 1. Time series with Normal distribution simulation.

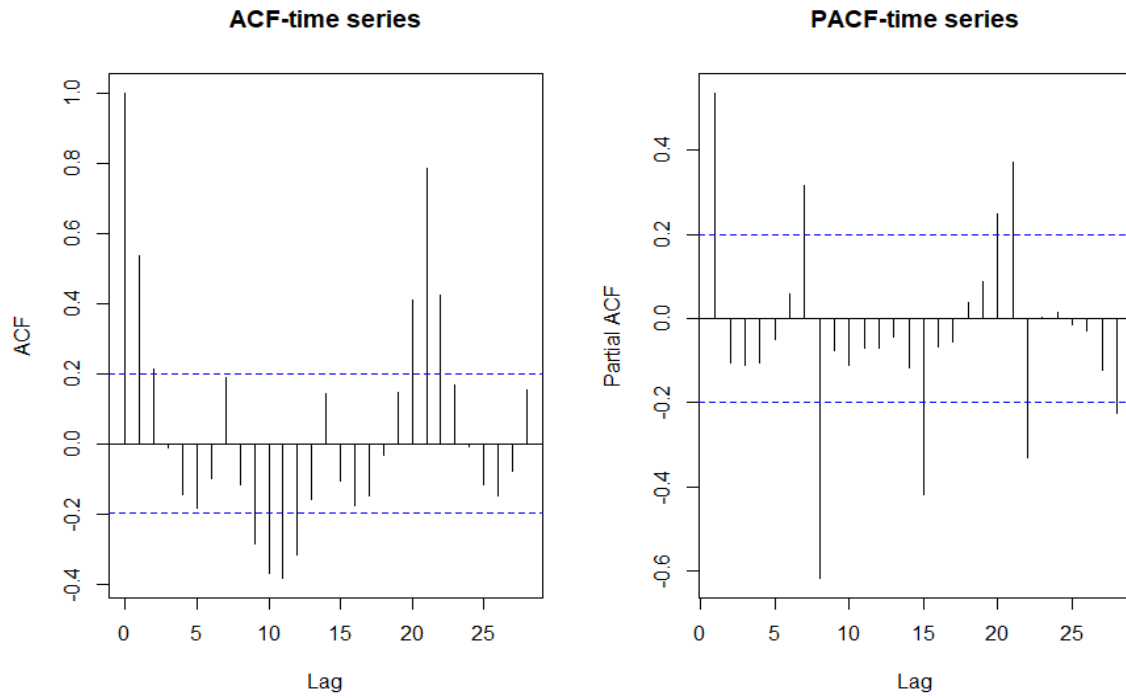


Figure 4. Time series autocorrelations-Skew Normal distribution simulation.

The time series presents autocorrelation, as the outside lines from the bands indicates.

Case 2. Time series with Skew Normal distribution simulation.

For the Skew Normal distribution simulations, the Box pierce tests were applied, checking the correct autocorrelation dependence, temporal depending, the P values are ranging from 1.8×10^{-7} to 4.11×10^{-8} , proximately.

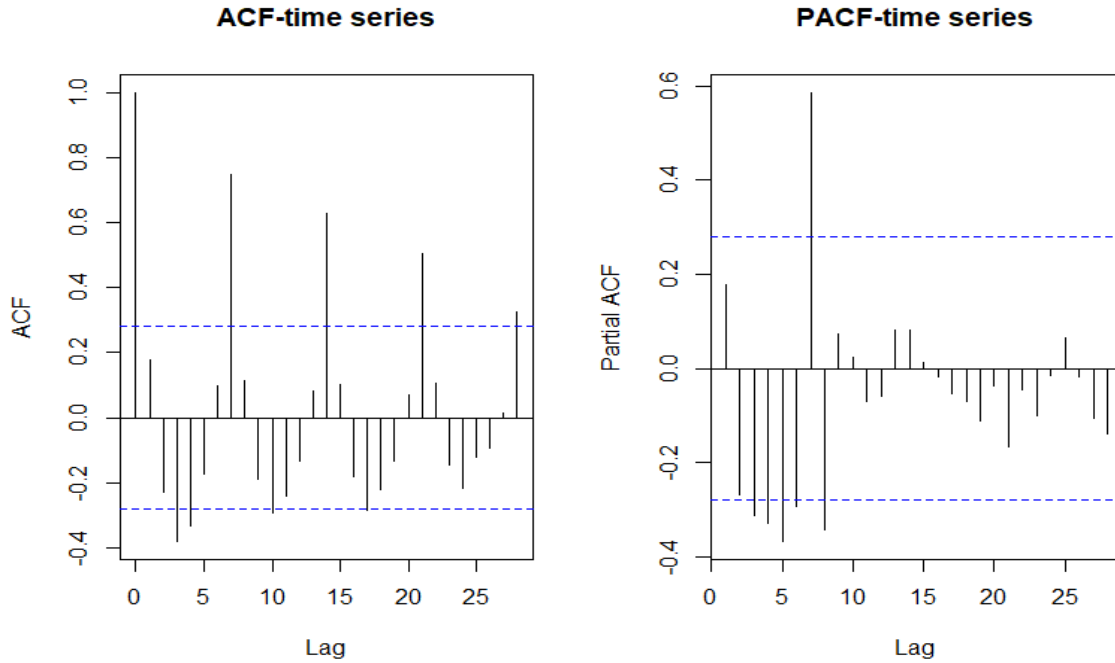


Figure 5. Time series autocorrelations – Skew Normal distribution simulation.

3.1. Comparison results

The SMAPE results are shown in table 1, where every column represents the results from every model and the last three, the BG combination, the metaheuristic without clusters (M), and the last one, the metaheuristic with the cluster variable (MV).

Simultaneously the algorithm projects the results for all the time series simulated, in this case, three results by every individual or combined method. The best results can be the BRM, the BG combination, and the metaheuristics combinations.

Table 1. Results by time series simulation.

SKEW NORMAL DISTRIBUTION							
	lm	ETS	AR	BRM	COMB_BG	COMB_M	COMB_MV
SERIE1	6,6%	9,0%	8,2%	5,0%	6,4%	6,5%	5,4%
SERIE2	4,0%	8,5%	7,9%	4,3%	4,1%	4,9%	4,9%
SERIE3	8,7%	10,3%	10,2%	6,0%	7,7%	7,6%	6,4%
NORMAL DISTRIBUTION							
	lm	ETS	AR	BRM	COMB_BG	COMB_M	COMB_MV
SERIE1	8,0%	8,6%	8,3%	5,0%	6,4%	4,4%	6,4%
SERIE2	4,0%	8,5%	7,9%	4,3%	4,1%	5,1%	4,8%
SERIE3	9,1%	9,9%	9,8%	6,0%	7,7%	8,8%	5,7%

The figure 6 shows the real data, forecasted in blue and red for metaheuristics, which seems to be close to the black line (real).

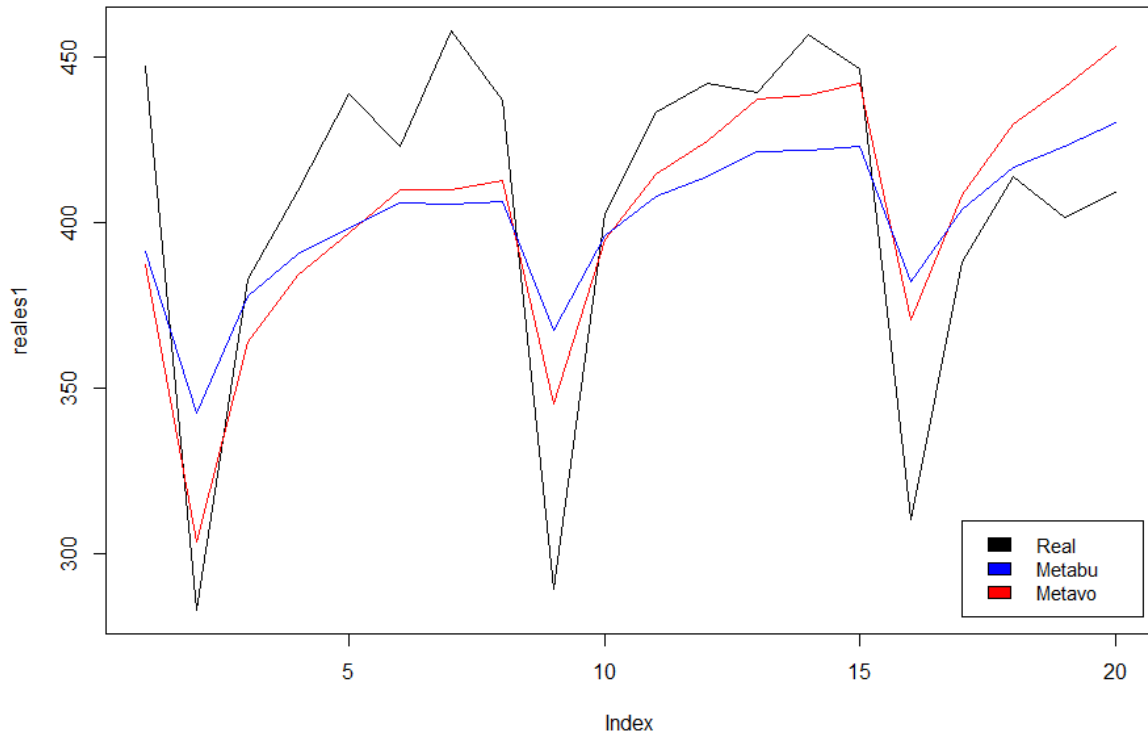


Figure 6. Real and Forecasted data.

The metaheuristics using statistical learning programming, find adequate forms to do predictions, by using statistical learning, aspects that are being programmed in R program.

Discussion

The proposed work compared Individual models with forecast combinations, finding good alternatives to do forecasting, as the BRM and also, the Metaheuristic with cluster analysis.

Better alternatives are found with the forecast combinations, using metaheuristics.

It was possible to apply Tabu Search, but also, an exploitation process around the simulated points, finding a good approximation to the objective function.

The algorithm created is also flexible, in order to select different simulation number, and some parameters of the metaheuristics.

References

- Andersson, M., & Karlson, S. (2007). Bayesian forecast combination for VAR models *. *Sveriges Riskbank-Working Papers*, 1–17. Retrieved from http://www.riksbank.se/Upload/Dokument_riksbank/Kat_publicerat/WorkingPapers/2007/wp216.pdf
- Bai, L., Wang, J., Ma, X., & Lu, H. (2018). Air Pollution Forecasts : An Overview. *International Journal of Enviromental Research and Public Health*, 15(780), 1–44. <https://doi.org/10.3390/ijerph15040780>
- Barrow, D. K., & Kourentzes, N. (2016). Distributions of forecasting errors of forecast combinations: Implications for inventory management. *International Journal of Production Economics*, 177, 24–33. <https://doi.org/10.1016/j.ijpe.2016.03.017>
- Bates, J. M., & Granger, C. W. . (1969). The combination of forecasts. *Operational Research Quraterly*, 20(4), 451.
- Bensoussan, A., Çakanyıldırım, M., Li, M., & Sethi, S. (2014). Inventory management with overlapping shrinkages and demands. *Risk and Decision Analysis*, 5(4), 189–210. <https://doi.org/10.3233/RDA-150110>
- Bergman, J. J., Noble, J. S., Mcgarvey, R. G., & Bradley, R. L. (2017). A Bayesian approach to demand forecasting for new equipment programs. *Robotics and Computer Integrated Manufacturing*, 47, 17–21. <https://doi.org/10.1016/j.rcim.2016.12.010>
- Cai, X., Chen, J., Xiao, Y., Xu, X., & Yu, G. (2013). Fresh-product supply chain management with logistics outsourcing. *Omega (United Kingdom)*, 41(4), 752–765. <https://doi.org/10.1016/j.omega.2012.09.004>
- Cang, S., & Yu, H. (2014). A combination selection algorithm on forecasting. *European Journal of Operational Research*, 234(1), 127–139. <https://doi.org/10.1016/j.ejor.2013.08.045>
- Chen, K.-Y. (2011). Combining linear and nonlinear model in forecasting tourism demand. *Expert Systems with Applications*, 38(8), 10368–10376. <https://doi.org/10.1016/j.eswa.2011.02.049>
- Debnath, K. B., & Mourshed, M. (2018). Forecasting methods in energy planning models. *Renewable and Sustainable Energy Reviews*, 88(September 2017), 297–325. <https://doi.org/10.1016/j.rser.2018.02.002>

- Dos-Santos, A., Constantino, M., & Lucio, P. (2018). Dynamic Regression Model for Evaluating the Association Between Atmospheric Conditions and Deaths due to Respiratory Diseases in São Paulo , Brazil Modelo Dinâmico de Regressão Para Avaliar a Associação Entre as Condições Atmosféricas e as Mortes por Doen. *Revista Brasileira de Meteorologia*, 33(1), 1–10. <https://doi.org/10.1590/0102-7786331001>
- Ferragina, A., de los Campos, G., Vazquez, A. I., Cecchinato, A., & Bittante, G. (2015). Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. *Journal of Dairy Science*, 98(11), 8133–8151. <https://doi.org/10.3168/jds.2014-9143>
- Fúquene, J., Álvarez, M., & Pericchi, L. (2015). A robust Bayesian dynamic linear model for Latin-American economic time series: “the Mexico and Puerto Rico cases.” *Latin American Economic Review*, 24(1), 6. <https://doi.org/10.1007/s40503-015-0020-z>
- Gao, L. (2015). Collaborative forecasting, inventory hedging and contract coordination in dynamic supply risk management. *European Journal of Operational Research*, 245(1), 133–145. <https://doi.org/10.1016/j.ejor.2015.02.048>
- Gao, W., Sarlak, V., Parsaei, M. R., & Ferdosi, M. (2018). Combination of fuzzy based on a meta-heuristic algorithm to predict electricity price in an electricity markets. *Chemical Engineering Research and Design*, 131, 333–345. <https://doi.org/10.1016/j.cherd.2017.09.021>
- Guo, F., Diao, J., Zhao, Q., Wang, D., & Sun, Q. (2017). A double-level combination approach for demand forecasting of repairable airplane spare parts based on turnover data. *Computers & Industrial Engineering*, 110, 92–108. <https://doi.org/10.1016/j.cie.2017.05.002>
- Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: Selecting among forecasts and their combinations. *International Journal of Forecasting*, 21, 15–24. <https://doi.org/10.1016/j.ijforecast.2004.05.002>
- Hsiao, C., & Wan, S. K. (2014). Is there an optimal forecast combination? *Journal of Econometrics*, 178(PART 2), 294–309. <https://doi.org/10.1016/j.jeconom.2013.11.003>
- Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., & Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational Statistics and Data Analysis*, 55(9), 2579–2589. <https://doi.org/10.1016/j.csda.2011.03.006>
- Jedermann, R., Nicometo, M., Uysal, I., & Lang, W. (2018). Reducing food losses by intelligent

- food logistics. *The Royal Society*, 372(2017), 1–20.
- Kociecki, A., Kolasa, M., & Rubaszek, M. (2012). A Bayesian method of combining judgmental and model-based density forecasts. *Economic Modelling*, 29, 1349–1355.
<https://doi.org/10.1016/j.econmod.2012.03.004>
- Liu, X., Peng, H., Bai, Y., Zhu, Y., & Liao, L. (2014). Tourism Flows Prediction based on an Improved Grey GM(1,1) Model. *Procedia - Social and Behavioral Sciences*, 138, 767–775.
<https://doi.org/10.1016/j.sbspro.2014.07.256>
- Meinhold, R. J., & Singpurwalla, N. D. (1983). Understanding the Kalman Filter. *The American Statistician*, 37(2), 123–127.
- Melo, L. F., Loaiza, R. A., & Villamizar-Villegas, M. (2016). Bayesian combination for inflation forecasts: The effects of a prior based on central banks' estimates. *Economic Systems*.
<https://doi.org/10.1016/j.ecosys.2015.11.002>
- Miller, J., Berry, C., & Lai, C. (2007). A comparison of alternative forecasting strategies for multistage production-inventory systems. *Decision Sciences*, 7(4), 714–724.
- Min, C., & Zellner, A. (1993). Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, 56(1–2), 89–118. [https://doi.org/https://doi.org/10.1016/0304-4076\(93\)90102-B](https://doi.org/https://doi.org/10.1016/0304-4076(93)90102-B)
- Nascimento, L., Pompeo-Ferreira, L., & Cota, K. (2016). Air pollution and respiratory diseases : ecological time series Poluição do ar e doenças respiratórias : estudo ecológico de série temporal. *Sao Paulo Med*, 134(4), 315–321. <https://doi.org/10.1590/1516-3180.2015.0237250216>
- Nowotarski, J., Liu, B., Weron, R., & Hong, T. (2016). Improving short term load forecast accuracy via combining sister forecasts. *Energy*, 98, 40–49.
<https://doi.org/10.1016/j.energy.2015.12.142>
- Petris, G. (2010). An R Package for Dynamic Linear Models. *Journal of Statistical Software*, 36(12), 1–16. Retrieved from <http://www.jstatsoft.org/>
- Puchalsky, W., Trierweiler, G., Pereira, C., Zanetti, R., & Coelho, S. (2018). Agribusiness time series forecasting using Wavelet neural networks and metaheuristic optimization: an analysis of the soybean sack price and perishable products demand. *International Journal of Production Economics*, 203(June), 174–189. <https://doi.org/S0925527318302548>

- R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Riedl, D. F., Kaufmann, L., Zimmermann, C., & Perols, J. L. (2013). Reducing uncertainty in supplier selection decisions: Antecedents and outcomes of procedural rationality. *Journal of Operations Management*, *31*(1–2), 24–36. <https://doi.org/10.1016/j.jom.2012.10.003>
- Sarimveis, H., Patrinos, P., Tarantilis, C. D., & Kiranoudis, C. T. (2008). Dynamic modeling and control of supply chain systems: A review. *Computers & Operations Research*, *35*(11), 3530–3561. <https://doi.org/10.1016/j.cor.2007.01.017>
- Valencia-Cárdenas, M., Díaz-Serna, F. ., & Correa-Morales, J. . (2016). Multi-product inventory modeling with demand forecasting and Bayesian optimization [Modelo de inventario multi-producto, con pronósticos de demanda y optimización Bayesiana]. *DYNA (Colombia)*, *83*(198), 236–244. <https://doi.org/10.15446/dyna.v83n198.51310>
- Valencia-Cárdenas, Marisol. (2016). *Dynamic Model for the Multiproduct Inventory Optimization with Multivariate Demand*. Universidad Nacional de Colombia, Sede Medellín. Retrieved from <http://www.bdigital.unal.edu.co/51516/1/43616826.2015.pdf>
- Valencia Cárdenas, M., Díaz Serna, F. J., & Correa Morales, J. C. (2015). Inventory planning with dynamic demand. A state of art review. *DYNA*. <https://doi.org/10.15446/dyna.v82n190.42828>
- Valencia, M., Vanegas, J., Correa, J., & Restrepo, J. (2017). Comparación de pronósticos para la dinámica del turismo en Medellín , Colombia. *Lecturas de Economía*, (86), 199–230. <https://doi.org/10.17533/udea.le.n86a08> Primera
- Wallström, P., & Segerstedt, A. (2010). Evaluation of forecasting error measurements and techniques for intermittent demand. *International Journal of Production Economics*, *128*(2), 625–636. <https://doi.org/10.1016/j.ijpe.2010.07.013>
- Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis - Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts usi. *Energy*, *93*, 41–56. <https://doi.org/10.1016/j.energy.2015.08.045>
- Wang, W. J., & Xu, Q. (2014). A Bayesian Combination Forecasting Model for Retail Supply Chain Coordination. *Journal of Applied Research and Technology*, *12*, 315–324. [https://doi.org/10.1016/S1665-6423\(14\)72347-8](https://doi.org/10.1016/S1665-6423(14)72347-8)

- Weiss, C., & Roetzer, G. R. (2016). Package ‘GeomComb.’ Retrieved from <https://github.com/ceweiss/GeomComb>
- Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming – A review. *Agricultural Systems*, 153, 69–80. <https://doi.org/10.1016/j.agsy.2017.01.023>
- Zellner, A. (1996). *An introduction to bayesian inference in econometrics* (Second). Wiley Classics Library.
- Zotteri, G., & Kalchschmidt, M. (2007). A model for selecting the appropriate level of aggregation in forecasting processes. *International Journal of Production Economics*, 108(1–2), 74–83. <https://doi.org/10.1016/j.ijpe.2006.12.030>