A Combined Approach for Predicting Operational Performance in Port Terminals

Ícaro Romolo Sousa Agostino, Satie L. Takeda-Berger and Enzo Morosini Frazzon
Industrial and Systems Engineering Department
Federal University of Santa Catarina
Florianópolis, SC 88040-900, Brazil
icaro.agostino@posgrad.ufsc.br, satie.takeda@posgrad.ufsc.br, enzo.frazzon@ufsc.br

Adriano Mendonça Souza
Statistics Department
Federal University of Santa Maria
Santa Maria, RS 97105-900, Brazil
amsouza.sm@gmail.com

Abstract

This paper proposes an approach to predict the operational performance of a port terminal using the combination of Exponential Smoothing, ARIMA and Artificial Neural Networks forecasting models. The empirical research was performed using data from the unloading operation of iron ore from the Carajás mine (Brazil), which is the largest open pit iron mine in the world. The results showed that the proposed combined forecasting approach generates better predictions than the univariates models when compared to other possible methods. The proposed approach can be applied to other complex decision problems that need to predict operational performance with high accuracy in the industrial and logistic context.

Keywords
Forecast combination; ARIMA; Artificial Neural Networks; Port Operations.

1. Introduction

Short-term decisions play an important role in the organizational context, especially in planning operations and resource allocation in productive processes (Vieira, Kück, Frazzon, and Freitag, 2017; Rashed, Meersman, Voorde, and Vanelslander, 2017; Choi, Wallace, and Wang, 2018; Leusin, Kück, Frazzon, and Freitag, 2018). In this context, forecasting techniques are presented with the aim of solving several problems involving risks, since they allow the generation of subsidies that assist decision makers in formulating strategies in productive processes (Wang and Chang, 2010; Bansal and Dyer, 2013; Amornpetchkul, Duenyas, and Şahin, 2015; Isasi, Frazzon, and Uriona, 2015; Boone, Ganeshan, Hicks, and Sanders, 2018). In general, forecasting in the organizational context uses quantitative methods, based on the existence of historical data records. Among the quantitative models for forecasting, the time series models stand out (Montgomery, Jennings, and Kulahci, 2015; Agostino, Silva, Veiga and Sousa, 2020).
In this scenario, port terminals are characterized as dynamic environments, depending on tools for planning and control of operations, aiming at minimizing efforts and increasing productivity (Gómez, Camarero, and Molina, 2016; Goedhals-Gerber, 2016; Agostino, Sousa, Frota, Daher, and Souza, 2019; Choi, Lee, Leung, Pinedo, and Briskorn, 2012). Among the works found in the literature, Fung (2001) developed a structured model of error corrections to predict the transfer rate of containers at the port terminals of Hong Kong. Vlahogianni, Karlaftis, and Golias (2005) used a nonlinear approach with application of artificial neural network (ANN) models for short-term prediction of flow and occupation of transport. Schulze and Prinz (2009) analyzed container movement at German ports using SARIMA models and Holt-Winters model, finding superior results for the adjusted SARIMA model. Pang and Gebka (2016) used the Holt-Winters models, ARIMA seasonal models and the Vectors Error Correction (VEC) model to

predict the container transfer rate of the Tanjung Priok port in Indonesia, finding superior results for the combination of models approach.

Studies involving quantitative forecasting models mostly employ models of exponential smoothing, ARIMA modeling and ANN models (Hyndman and Athanasopoulos, 2018; Martins and Werner, 2012). An alternative to obtaining more efficient forecasts is to combine individual forecasting techniques, resulting in predictions with statistically higher accuracy when compared to individual techniques (Hibon and Evgeniou, 2005). Several studies have shown that different combinations of methods have obtained more accurate results in relation to individual predictions (Clemen, 1989; Makridakis and Hibon, 2000; Stock and Watson, 2004; Amendola and Storti, 2008; Doganis, Aggelogiannaki, and Sarimveis, 2008; Jeong and Kim, 2009; Costantini and Pappalardo, 2010; Wallis, 2011; Azevedo and Campos, 2016; Wang and Petropoulos, 2016; Makridakis, Spiliotis, and Assimakopoulos, 2018).

Among combination of forecast methods, the arithmetic mean is one of the most popular methods employed, bringing satisfactory results in reducing predictions variability (Stock and Watson, 2004; Koning, Franses, Hibon, and Stekler, 2005; Martins and Werner, 2012). Another widely used approach is the minimum variance method, consisting in linear combination of individual forecasters with adoption of weights based on the prediction errors for the forecast combinations (Bates and Granger, 1969; Timmermann, 2006; Martins and Werner, 2012).

This research proposes the combination of the Holt-Winter, ARIMA and ANN models to predict the short-term behavior of the average unload time of an ore unload system composed of eight production lines. This unload time is the key performance indicator of the port terminal studied. For the forecast combination, the arithmetic mean and the minimum variance models were used. This study contributes to the research area since (i) it presents a robust approach for short-term forecasting, and (ii) proposed and approach that can be easily replicable in many contexts involving complex operational decision problems in industrial and logistic context.

The paper is organized as follows: Section 2 presents a review of exponential smoothing, ARIMA and ANN models, as well as forecast combination methods. Section 3 describes the methodological procedures adopted. Section 4 presents the case study. Section 5 presents the results. Finally, section 6 presents the discussion and the final considerations of the paper.

2. Theoretical Background

2.1 Forecasting Models

Exponential smoothing methods popularized by Holt (1957) and Winters (1960) constitute a family of mathematical models that aim to adjust a curve appropriate to the historical data of a time series, widely used in many areas, such as business and economics (Montgomery, Jennings, and Kulahci, 2015). The method called Holt-Winters has two distinct approaches: the multiplicative, in case the seasonal amplitude increases in relation to the time; and additive, when the series studied has a constant amplitude along its cycles (Hyndman and Athanasopoulos, 2018). The additive method is represented by Equation (1):

$$Z_{t+m} = (L_t + b_t m) + S_{t-s+m} \tag{1}$$

Where *Lt* represents the estimated value of the level; *bt* represents the trend estimate; *Z(t+m)* corresponds to the forecast in period *t+m*, *m* represents how many forward steps it wishes to predict; and *St* is the seasonal time series index.

The autoregressive integrated moving average (ARIMA) models were developed by Box and Jenkins (1970). The ARIMA modeling follows four steps: (i) identification, consisting of the determination of the model that best describes the series behavior, through the analysis of autocorrelation functions (ACF) and partial autocorrelation (PACF); (ii) estimation, making the estimation of autoregressive parameters, as well as moving averages; (iii) validation, consisting of the analysis of the adequacy of the model adjusted to the real series behavior, in which model residuals are analyzed, (iv) prediction, which is only performed when the previous steps are satisfactory (Montgomery, Jennings, and Kulahci, 2015; Hyndman and Athanasopoulos, 2018). Usually, ARIMA models (p, d, q) are represented by equation (2):

$$\phi(B)\Delta^d Z_t = \theta(B)\varepsilon_t \tag{2}$$

Where: Z_t represents the time series modeled, B represents the retroactive operator, d represents the order of integration, ϕ is the term that represents the autoregressive parameter of order p, θ represents the parameter of moving average of order q and ε_t represents the sequence of errors, denoted white noise when the average of errors is zero and the variance is constant with homoscedasticity ~ (0, σ2). The best fit model selected by the Akaike criterion (AIC) (Akaike, 1974) is selected for analysis and forecasting.

Artificial neural network (ANN) models were developed inspired by the human brain information process, through training algorithms, being considered a powerful approach for modeling time series, capable of capturing complex and non-linear patterns. An ANN structure is composed of four elements: neurons, type of model, networks and learning, being neurons the main element of an ANN (Agami, Atiya, Saleh, and El-Shishiny, 2009; Guresen, Kayakutlu, and Daim, 2011).
Among the main architecture models used to carry out predictions from ANN, the Multilayer Perceptron model (MLP) is the most widely used because it allows predictions with univariate models, using the lagged periods of the series itself in the input layer (Agami, Atiya, Saleh, and El-Shishiny, 2009; Ahmed, Atiya, Gayar, and El-Shishiny, 2010; Jacobs, Souza, and Zanini, 2016). The model uses non-linear sigmoid functions according to the structuring of the built network and can be considered a generic non-linear autoregressive model (Hyndman and Athanasopoulos, 2018; Zhang, Patuwo, and Hu, 2001). The generic model MLP/ANN can be represented by Equation (3):

$$Z_t = b_i + \sum_{j=1}^{N} w_{ij} x_j \tag{3}$$

Where: x_j represents the input signals; w_{ij} represents the weights given to each input signal; b_i represents the bias; and N is the total of input signals in the model.
The definition of the structure to be used in an ANN for forecasting does not follow a single rule, with several approaches in the literature. (Zhang, Patuwo, and Hu, 2001) suggest the construction of the ANN structure based on the autocorrelation structure of the time series to be modeled.

2.2 Combination of Forecasts

The theoretical aspects of the approaches for the forecast combination adopted are presented in this section. The Minimum Variance method was proposed by (Bates and Granger, 1969) and is one of the most widespread approaches for forecast combination (Timmermann, 2006; Martins and Werner, 2012). This method consists in combining forecast in a linear and objective way, assigning a weight "*w*" for the first forecast and a complementary weight (*1-w*) for the second forecast. The weights value is based on the observation of the variances of the predictions errors to be combined. Thus, it should be attributed less weight to the prediction with greater variability in the errors, in order to obtain greater accuracy, as can be seen in Equation (4):

$$\mathrm{F_c} = \mathrm{wF_1} + (1 - \mathrm{w})\mathrm{F_2} \tag{4}$$

Where: F_c is the combined forecast; *w* is the weight assigned to the forecast and F_1 and F_2 are the original predictions to be combined. The weights are calculated by the Equation (5):

$$w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \tag{5}$$

Where: w is the weight assigned to the forecast; ρ corresponds to Pearson's linear correlation coefficient between the errors from the forecasts; σ_1^2 and σ_2^2 corresponds to the error variance value of the forecasts F_1 and F_2.
A simplified approach suggested by (Makridakis and Winkler, 1983) refers to the use of the simple average of the forecasts, since according to the authors the average can provide greater accuracy of the forecast, reducing its variability, since there is no dependency information between the forecasts, it becomes reasonable to use it, and may give rise to better results than more sophisticated methods (Stock and Watson, 2004; Koning, Franses, Hibon, and Stekler, 2005).

3. Methodology

The methodology is divided into 2 sections: (1) description of the research data; (2) presentation of the proposed approach for individual modeling and forecasts combination and evaluation.

3.1 Survey data

The data used for modeling were collected from the database of notes of the industrial automation system of a large Brazilian company of the mining and logistics sector, and is the main production control system of the organization. The data refer to a month of operation, with hourly observations, generating eight-time series with 720 observations

each. A logarithmic transformation in the time series was applied in order to stabilize the data variance, as suggested by Montgomery, Jennings and Kulahci (2015) and Hyndman and Athanasopoulos (2018) in cases of data with non-constant variances over time.
The time series were divided into two parts, the first corresponding to the training period with about 80% of the observations and the second corresponding to test phase with about 20% of the observations. The training phase was used to adjust the models, while the test phase was used to evaluate the accuracy measurements.

3.2 Individual modeling and combination of forecasts

The individual forecasts were obtained from the Exponential Smoothing, Box-Jenkins and ANN models, being selected as the best models in each forecast the ones that presented the best adjustments. The steps of each individual modeling are described below:
Exponential Smoothing: verification of the presence of trend and seasonality components; model identification (Exponential Smoothing, linear of Holt, Holt-Winters); estimation of parameters; forecasting and calculation of accuracy measurements.
Box-Jenkins: verification of the stationarity of the series by the ADF (Said and Dickey, 1984) and KPSS tests (Kwiatkowski, Phillips, Schmidt, and Shin, 1992) both using the significance level of 5%; model identification (p, d, q); estimation of parameters; validation of the model (verification of noise and parameters); forecasting and calculation of accuracy measurements.
ANN: definition of the number of neurons in the input layer; training a set of ANN; choosing the best ANN according to criteria of accuracy; forecasting and calculation of accuracy measurements.
After obtaining the individual forecasts, combinations were made in three different ways: using the minimum variance method and using simple arithmetic mean. For all combinations, it was used two individual forecasters that presented smaller quadratic errors. Figure 1 illustrates the proposed approach.

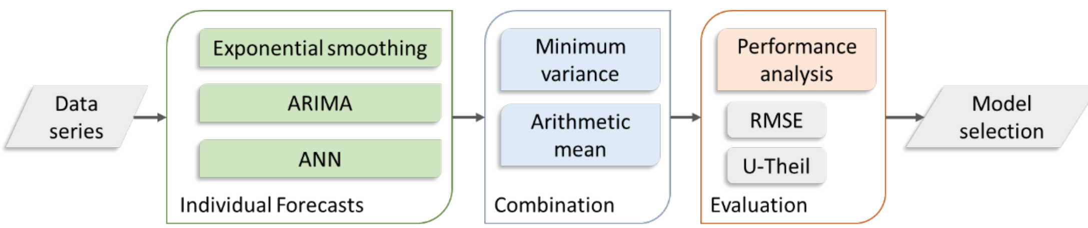

Figure 1. Proposed combined approach

The performance analysis of the proposed models, both individual and combined, were performed by comparing the accuracy measures *root mean squared error* (RMSE) and the *U-Theil* coefficient. The use of RMSE as a model assessment measure is suggested by Chai and Draxler (2014) and Hyndman and Athanasopoulos (2018) when analyzing forecasts in the same unit of measurement. The U-Theil coefficient is an important measure for forecast evaluation, the measure evaluates forecast performance regarding naive or trivial forecast (Hyndman and Koehler, 2006; Theil, 1966). A model is considered adequate when $U < 1$, indicating that the error of the adjusted model is smaller than a naive forecast, when $U \geq 1$, the error of the fitted model is bigger than or equal to a naive forecast. The U-Theil coefficient can be calculated by Equation (6):

$$U = \frac{\sqrt{\sum_{t=1}^{n}(\varepsilon_t)^2}}{\sqrt{\sum_{t=1}^{n}(Z_t - Z_{t-1})^2}} \quad (6)$$

Where: ε_t represents the forecast error at time t; n is the number of observations: Z_t is the real value at time t.
The forecast model selected was the one that presented the smallest errors for the proposed measures in the test phase of the modeled series. In this way, at the end of the research, it was possible to evaluate the application of the individual forecast models, as well as the combined models, allowing a comparison and discussion of the applicability of the methods proposed in time series. The methodology established in this paper allowed an objective and quantitative evaluation based on scientific criteria, enabling a discussion of the empirical evidence resulting from this study.
As a computational resource for treatment, analysis and modeling of data, the 'R' (R Core Team, 2018) software was used together with the 'tseries' packages (Trapletti and Hornik, 2018), 'forecast' (Hyndman and Khandakar, 2008).

4. Case Study

The study was developed in a port operation management, in which part of an integrated logistic system of iron ore handling (containing mine, railroad and port) was modeled, responsible for the unloading of all production of the Carajás mine (Brazil), which is currently considered the world's largest open pit mine (Ministry of Infrastructure, 2019). The material transportation, after extraction, is carried out via rail system by composition of three batches of 110 wagons forming a composition of 330 wagons. Such composition is dismembered via Rail Terminal and each lot is handled and positioned in a Wagon tippler to carry out the unload of the ore. Figure 2 illustrates the movement process studied.

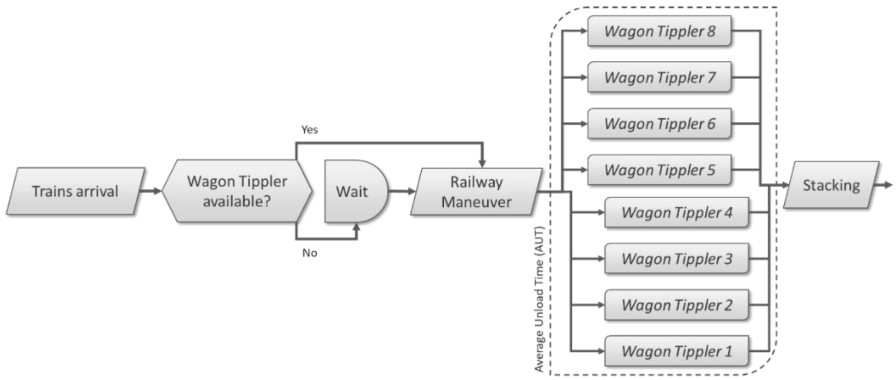

Figure 2. Ore unload system by wagon tippler

The initial port process is characterized by the ore unload using eight Wagon Tippler (WT), in which each industrial equipment turns one pair of wagons at a time, directing the material to a system of conveyor belts. The proposed model contemplates the ore unload operation. The modeled variable represents the Average Unloading Time (AUT) of the wagon tippler system every hour, the indicator is represented by Equation (7):

$$AUT = TUT/Nº \text{ batches} \quad (7)$$

Where: TUT represents the total unload time; Nº batches represents the total number of lots operated in the period, each lot consisting of 110 wagons.
The AUT variable is considered a key indicator of production, is the main control parameter of the port unload system. Due to the unloading process to attend to the subsequent port processes of stacking and shipment, the control over the variable AUT becomes an important factor to be approached in the decision processes, such as the production dimensioning and the planning of maintenance routines, in this way the adjusted model will be used as a forecaster of future behavior of the variable, serving as a tool to support decision making.

5. Results and Discussions

The results are presented in 3 steps: (1) descriptive analysis of the series; (2) adjustment of the individual models; (3) combination of forecasts and comparison of the models performance outside of the sample.

5.1 Data Series

The descriptive statistics of time series are presented in Table 1. Among the unload lines, WT8 had the highest mean (119,177), while the lowest mean was obtained by WT3 (111,486). Regarding the dispersion, WT2 had the lowest C.V. (coefficient of variation) (6.0%), while the highest variability was observed in WT8 (12.7%). Figure 3 represents the behavior of the AUT indicator, the ACF and PACF for each unload line, collected during one month of operation with hourly observations, generating eight-time series with 720 observations.

Table 1. Descriptive statistics of AUT series

	WT1	WT2	WT3	WT4	WT5	WT6	WT7	WT8
Mean	115.508	112.406	111.486	112.97	113.581	113.911	114.489	119.177
C.V.	7.90%	6.00%	7.50%	9.30%	7.60%	9.90%	8.30%	12.70%

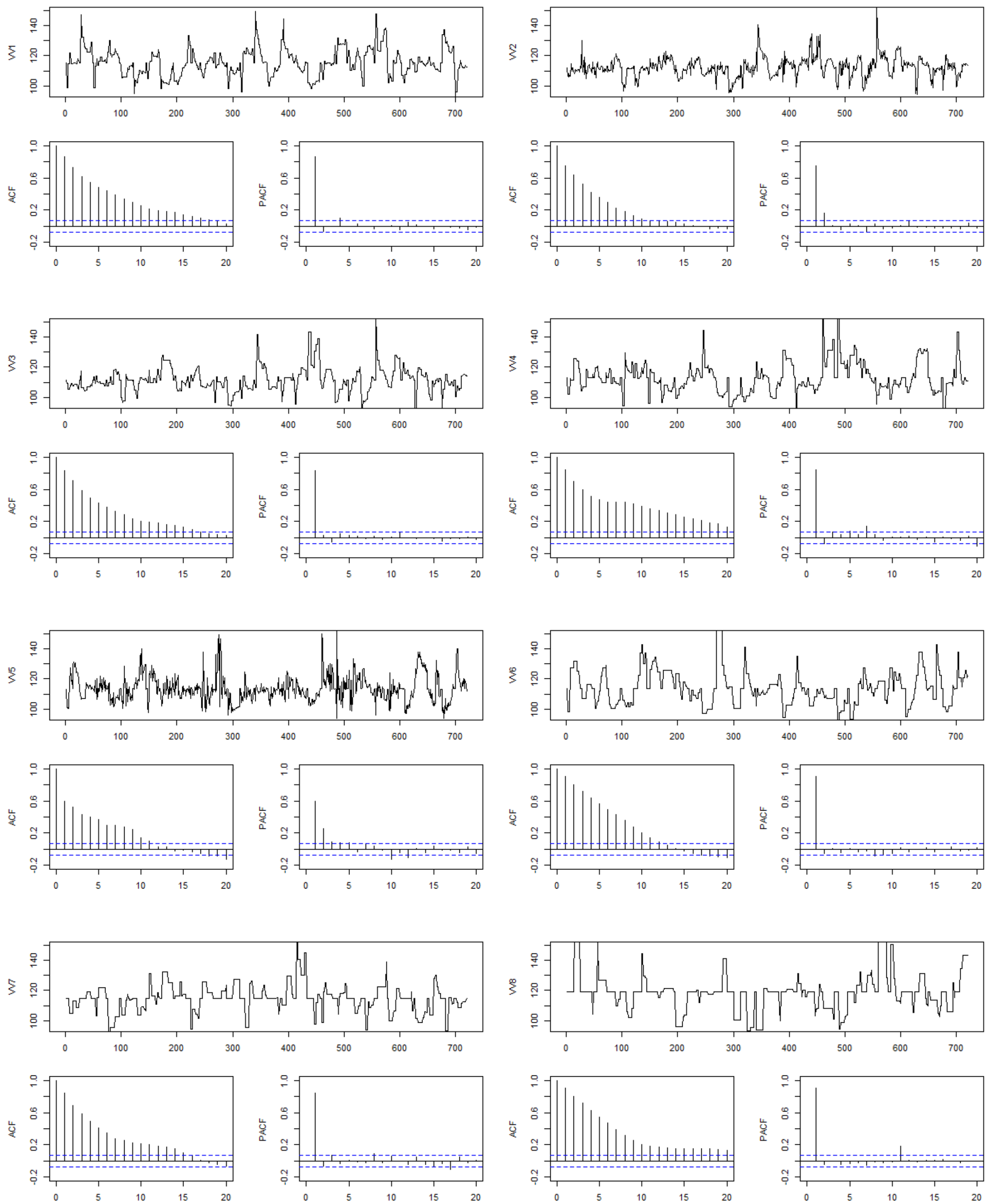

Figure 3. Time series and ACF and PACF correlograms.

In Figure 3, from the ACF it is possible to observe the presence of autocorrelation in all series between periods Y and Y_{t-1} indicating a short-term dependence between periods. In the PACF it is possible to identify autocorrelation between the residues, predominantly between the periods ε_t and ε_{t-1}, ..., ε_{t-10}. As mentioned in section 3.1, in this study, logarithmic transformation was applied in order to stabilize the variance of the data reducing problems regarding heteroscedasticity, in the next steps, the transformed data were used to adjust the models.

5.2 Individual Models

For the adjustment of all models, the AIC criterion was used. For the series WT1, WT2, WT3, WT4, WT5, WT5, WT6 and WT7 the exponential smoothing model with only one smoothing parameter obtained better performance. This occurred because the series were stationary and did not present seasonal behavior. Only the WT8 adjusted better to the additive model of Holt-Winters, without the presence of tendency. Table 2 contains the adjusted exponential smoothing models.

Table 2. Exponential smoothing models

Series	**Model**	**Parameters**	**Inicial states**	**AIC**	**MSE**
WT1	Simple Exponential smoothing	α 0.995	115.507	2.927	18.622
WT2	Simple Exponential smoothing	α 0.685	111.229	2.848	17.211
WT3	Simple Exponential smoothing	α 0.942	111.265	3.012	20.278
WT4	Simple Exponential smoothing	α 0.993	112.684	3.123	22.642
WT5	Simple Exponential smoothing	α 0.395	109.648	3.682	39.597
WT6	Simple Exponential smoothing	α 1.000	113.702	3.506	33.236
WT7	Simple Exponential smoothing	α 0.986	114.246	3.169	23.729
		α 0.995	113.090		
WT8	Additive Holt-Winters	β 0.000	-	3.284	26.472
		γ 0.965	5.946		

Among the adjusted exponential smoothing models, the model adjusted to the WT2 series presented lower MSE, as well as lower value for the AIC criterion. The reason for the superior performance of this model can be attributed to the lower variability of the series, since it presented a lower coefficient of variation (6.0%), indicating lower performance of the exponential smoothing models for series with greater variability.

AIC criterion was used for the adjustment of the ARIMA models, in addition to presenting white noise characteristics. All series were considered stationary at level ($d = 0$) by the ADF and KPSS tests at a significance level of 5%, not requiring application of differences. Table 3 contains the adjusted ARIMA models. All models presented autoregressive parameters and only the models adjusted to the WT3, WT7 and WT8 series did not present moving average parameters, all parameters of the adjusted models were significant (*p-value* <0.05). In all cases, the models confirm the short-term dependence between the periods of the series.

Table 3. ARIMA models

Series	**Model**	**AIC**	**MSE**
WT1	ARIMA (3,0,1)	2.808	16.396
WT2	ARIMA (4,0,2)	2.753	15.428
WT3	ARIMA (1,0,0)	2.924	18.506
WT4	ARIMA (3,0,2)	3.088	21.696
WT5	ARIMA (1,0,1)	3.616	36.87
WT6	ARIMA (2,0,1)	3.431	30.657
WT7	ARIMA (2,0,0)	3.170	23.683
WT 8	ARIMA (1,0,0)	3.241	25.423

In relation to the exponential smoothing models, the ARIMA model presented superior performance, since the adjusted models produced smaller squared errors, as well as being more parsimonious according to the AIC criterion, demonstrating that the relationship between quality and complexity of the adjusted models may justify the use of more robust models.

The ANN models, as well as the others, were adjusted and evaluated using the AIC criterion. For each variable, ten ANNs were trained, totaling 80 adjusted models. The structure and choice of networks for each model was defined according to the periodicity of the variables, using the 24-period cycle, and the autocorrelation structure of the same. Table 4 contains the final ANN models adjusted. For the input layer, the models obtained better performance using between 4 and 8 lagged periods, for the intermediate layer there was greater variability, WT3 presented only one neuron while the WT2, WT4, WT5 and WT7 presented 24 neurons.

Table 4. ANN models

Series	Model	AIC	MSE
WT1	ANN (6,05)	2.899	17.909
WT2	ANN (5,24)	2.701	14.690
WT3	ANN (5,01)	2.921	18.306
WT4	ANN (6,24)	3.111	22.067
WT5	ANN (8,24)	3.657	37.906
WT6	ANN (4,06)	3.444	30.951
WT7	ANN (4,24)	3.066	21.211
WT8	ANN (5,02)	3.236	25.073

In relation to the individual models of exponential smoothing and ARIMA, the ANN modeling presented a superior performance for the WT2, WT3, WT7 and WT8 series, both in terms of accuracy, producing smaller squared errors and lower values for the AIC criterion.
For the next stage of forecasts combination, ARIMA and ANN forecasts models were used, since a higher performance of the forecasts produced by these models was observed, presenting smaller squared errors, as proposed in section 3.2.

5.3 Combination of forecasts and comparison of the models (out of sample)

The results of the individual models forecasts, as well as the forecasts combinations by the methods of arithmetic mean and minimum variance are shown in Table 5. In previous steps, the AIC criterion was used as a performance evaluation measure to select the models. The AIC criterion is based on the relationship between complexity and quality of the adjusted model, but does not necessarily identify the most accurate model, capable of producing smaller errors (Azevedo and Campos, 2016). Thus, to compare models of different classes, RMSE and U-Theil accuracy measurements were used for the five final predictions obtained.

Table 5. Comparison of forecasts (out of sample)

	Individual models			Combined models	
	Exponential Smoothing	ARIMA	ANN	Arithmetic mean	Minimum variance
VV1					
RMSE	4.753730	4.511952	4.446045	4.438198	4.439137
U-Theil	0.020514	0.019483	0.019191	0.019159	0.019163
VV2					
RMSE	4.532623	4.387700	4.361616	4.351346	4.351005
U-Theil	0.020129	0.019507	0.019365	0.019332	0.019330
VV3					
RMSE	4.788464	4.607970	4.585916	4.588482	4.588376
U-Theil	0.021416	0.020625	0.020489	0.020518	0.020517
VV4					
RMSE	5.791635	5.542214	5.471289	5.474927	5.475237
U-Theil	0.025523	0.024449	0.024119	0.024144	0.024145
VV5					
RMSE	6.851115	6.682222	6.568798	6.586439	6.587198
U-Theil	0.030091	0.029357	0.028815	0.028910	0.028914
VV6					
RMSE	4.904011	4.784239	4.756972	4.749200	4.749182
U-Theil	0.021423	0.020917	0.020774	0.020751	0.020751
VV7					
RMSE	5.286164	5.292926	5.084330	5.094601	5.088796
U-Theil	0.023008	0.023040	0.022215	0.022218	0.022195
VV8					
RMSE	6.670583	6.488616	6.585766	6.516742	6.516494
U-Theil	0.027740	0.027027	0.027424	0.027140	0.027139

The results found for the out of sample forecasts showed that the forecasts combination obtained superior results in six of the eight variables modeled. For all predictions carried out, the U-Theil coefficient presented values lower than 1, indicating that the models perform better than the naive forecasts.
Among the individual forecast models, the exponential smoothing models obtained lower performance in the out of sample period in relation to the others. The ARIMA models presented a superior performance for the WT4 and WT8 indicators in relation to the other individual forecasts. The ANN models were superior in performance for the other indicators, showing that, for the data modeled, the ANN approach was more accurate in most cases for the individual forecasts.
In relation to the forecasts combination, the ANN model adjusted to the WT5 also presented better performance in the out of sample period, this can be partially explained by the number of layers, since among the adjusted ANN models, the WT5 obtained better adjustment with the largest number of parameters, indicating a high level of adjustment of the model to the series. However, the most robust analysis of the model is limited because, although it is possible to identify the hidden layers, the practical interpretation of its parameters and coefficients is difficult. In literature, such effect is called the "black box". However, the main focus of the model is the realization of forecasts, not making its use unfeasible due to its limitations (Gareta, Romeo, and Gil, 2006).
For the WT8, the adjusted ARIMA individual model presented better performance in the out of sample period, both compared to the individual and combined predictions, the ARIMA adjusted model (1,0,0) indicates that the series is influenced mainly by the immediately preceding period lagged, indicating a short-term memory in the indicator. For this variable, the forecasts combination was not able to reduce the forecast error, this fact can be explained by the greater variability of the series in relation to the others, making unstable the accomplishment of forecasts combination.
The combination methods adopted obtained better performance, since its execution was superior for the RMSE measure in relation to other methods. The minimum variance combination approach was superior, evidencing the method as an attractive approach to reduce errors of individual forecasting models. The final selection of models for the forecast of the key performance indicators considered the models more accurate according to the RMSE measurement. Among all the approaches adopted, the exponential smoothing method presented greater errors compared to the other models in the out of sample period, followed by ARIMA and ANN models. In general, the forecast combination was able to reduce the errors produced by the individual forecasts, increasing the precision of the combined forecasts that presented the best results in the majority of the cases.

6. Conclusions and outlook

The combination of the forecasting models is a widespread approach to improve forecast accuracy. The aim of this research was to obtain accurate forecasting models for eight key performance indicators of a port terminal, using the combination of the exponential smoothing, ARIMA and ANN models along with combination methods. The individual models were adjusted using the in-sample period containing 80% of the observations collected, being applied to the data a transformation by natural logarithm to stabilize the variance. The final evaluation of the models was performed by comparing the RMSE and U-Theil measurements for the predictions on the out of sample period.
This work presented a contribution to the literature combining three different models to predict the performance of an operation of a port terminal, obtaining relevant and original empirical results. The results showed that among the five final predictions obtained for each indicator, the ARIMA and ANN individual models were superior in accuracy in relation to the individual predictions in two of the eight cases and the combination methods were superior in the five other model indicators. These results are corroborated by the literature, as in the study performed by Martins and Werner (2012), which compared individual and combined forecasts for 50 real series of the industrial sector, obtaining superior results with combined forecasts. The study by Pang and Gekba (2016) that used the forecasts combination to obtain more accurate results applied to the transfer of containers in a port located in Indonesia, finding results superior to the individual forecasts. Another recent study by Jacobs, Souza and Zanini (2016) using ANN-based models, as well as ARIMA modeling in sets with several combination methods to predict the aggregate demand of a dairy industry, the results found by the authors demonstrated that for the case studied, the forecasts combination obtained the best accuracy, presenting smaller errors for the absolute and relative mean error statistics. Another recent study by Makridakis, Spiliotis and Assimakopoulos (2018), considering a hundred thousand real-time series of several sectors, showed that forecasting models with combination approaches, which jointly use computational and statistical methods, tend to perform better in terms of accuracy. In this way, the results corroborate with the literature confirming the importance of forecasts combination.
Although the adjusted final models present accurate results, the indicators management will require the constant exercise of modeling and forecast overtime to capture the variable nature of the data analyzed, however, the sample

considered allowed to conclude that models in the short term showed itself as accurate forecasters of the indicators. In addition, the proposed approach can be replicated in other studies involving complex decision problems that need to predict the behavior of variables with high precision. As a guide for future work, it is suggested the use of multivariate models to predict performance indicators, allowing the incorporation of the interrelation between variables such as vector autoregressive modeling and vectors error correction. Another important aspect to be investigated is the applicability of the forecast combination through other combination methods in different contexts and different forecast horizons.

Acknowledgements

The authors thank the Statistical Analysis and Modeling Laboratory of the Federal University of Santa Maria (LAME/UFSM) for the provision of equipment and infrastructure and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Financing Code 001.

References

Agami, N., Atiya, A., Saleh, M., and El-Shishiny, H. A neural network based dynamic forecasting model for Trend Impact Analysis. *Technological Forecasting and Social Change,* vol. 76, no. 9, pp. 952-962, 2009.

Agostino, I.R.S., da Silva, W.V., Veiga, C.P. and Souza, A.M., Forecasting models in the manufacturing processes and operations management: systematic literature review. *Journal of Forecasting*. accepted paper, 2020.

Agostino, I. Sousa, S., Frota, P., Daher R., and Souza, A. Modeling and Simulation of Operations: A Case Study in a Port Terminal of Vale S/A. In: *Lecture Notes in Management and Industrial Engineering.* s.l.:Springer International Publishing, pp. 91-99, 2019.

Ahmed, N. K., Atiya, A. F., Gayar, N. E., and El-Shishiny, H. An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*, vol. 29, no. 8, pp. 594-621, 2010.

Akaike, H. A New Look at the Statistical Model Identification. In: *Springer Series in Statistics.* s.l.:Springer New York, pp. 215-222, 1974.

Amendola, A., and Storti, G. A GMM procedure for combining volatility forecasts. *Computational Statistics & Data Analysis*, vol. 52, no. 2, pp. 3047-3060, 2008.

Amornpetchkul, T. B., Duenyas, I., and Şahin, Ö. Mechanisms to Induce Buyer Forecasting: Do Suppliers Always Benefit from Better Forecasting?. *Production and Operations Management,* vol. 24, no. 3, pp. 1724-1749, 2015.

Azevedo, V. G., and Campos, L. M. S. Combination of forecasts for the price of crude oil on the spot market. *International Journal of Production Research*, vol. 54, no. 3, pp. 5219-5235, 2016.

Bansal, S., and Dyer, J. S. Updating Inventories of Substitutable Resources in Response to Forecast Updates. *Production and Operations Management,* vol. 23, no. 7, pp. 477-488, 2013.

Bates, J. M., and Granger, C. W. J. The Combination of Forecasts. *Journal of the Operational Research Society,* vol. 20, no. 12, pp. 451-468, 1969.

Boone, T., Ganeshan, R., Hicks, R. L., and Sanders, N. R. Can Google Trends Improve Your Sales Forecast?. *Production and Operations Management,* vol. 27, no. 2, pp. 1770-1774, 2018.

Box, G. E. P., and Jenkins, G. M. *Time series analysis. Forecasting and control.* San Francisco: Holden-Day, 1970.

Chai, T., and Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development,* vol. 7, no. 6, pp. 1247-1250, 2014.

Choi, B. C., Lee, K., Leung, J. Y. T., Pinedo, M. L. Container Scheduling: Complexity and Algorithms. *Production and Operations Management,* vol. 21, no. 3, pp. 115-128, 2012.

Choi, T.-M., Wallace, S. W., and Wang, Y. Big Data Analytics in Operations Management. *Production and Operations Management,* vol. 27, no. 2, pp. 1868-1883, 2018.

Clemen, R. T. Combining forecasts: A review and annotated bibliography. *International journal of forecasting,* vol. 5, pp. 559-583, 1989.

Costantini, M., and Pappalardo, C. A hierarchical procedure for the combination of forecasts. *International Journal of Forecasting,* vol. 26, no. 10, pp. 725-743, 2010.

Doganis, P., Aggelogiannaki, E., and Sarimveis, H. A combined model predictive control and time series forecasting framework for production-inventory systems. *International Journal of Production Research,* vol. 46, no. 12, pp. 6841-6853, 2008.

Fung, M. K. Forecasting Hong Kongs container throughput: an error-correction model. *Journal of Forecasting,* vol. 21, pp. 69-80, 2001.

Gareta, R., Romeo, L. M., and Gil, A. Forecasting of electricity prices with neural networks. *Energy conversion and management,* vol. 47, pp. 1770-1778, 2006.

Goedhals-Gerber, L. L. Predicting the throughput of grain products at the multipurpose terminal at the Port of Cape Town. *Journal of Transport and Supply Chain Management,* vol. 10, no. 1, 2016.

Gómez, R., Camarero, A., and Molina, R. Development of a Vessel-Performance Forecasting System: Methodological Framework and Case Study. *Journal of Waterway, Port, Coastal, and Ocean Engineering,* vol. 142, no. 3, p. 04015016, 2016.

Guresen, E., Kayakutlu, G., and Daim, T. U. Using artificial neural network models in stock market index prediction. *Expert Systems with Applications,* vol. 38, no. 8, pp. 10389-10397, 2011.

Hibon, M., and Evgeniou, T. To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting,* vol. 21, no. 1, pp. 15-24, 2005.

Holt, C. C. Forecasting trends and seasonal by exponentially weighted moving averages. *ONR Memorandum,* vol. 52, 1957.

Hyndman, R. J., and Athanasopoulos, G. *Forecasting: principles and practice.* OTexts, 2018.

Hyndman, R. J., and Khandakar, Y. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software,* vol. 27, 2008.

Hyndman, R. J., and Koehler, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting,* vol. 22, no. 10, pp. 679-688, 2006.

Isasi, N. K. G., Frazzon, E. M., and Uriona, M. Big data and business analytics in the supply chain: a review of the literature. *IEEE Latin America Transactions*, vol. 13, no. 10, pp. 3382-3391, 2015.

Jacobs, W., Souza, A. M., and Zanini, R. R. Combination of Box-Jenkins and MLP/RNA models for forecasting. *IEEE Latin America Transactions,* vol. 14, no. 4, pp. 1870-1878, 2016.

Jeong, D. I., and Kim, Y.-O. Combining single-value streamflow forecasts – A review and guidelines for selecting techniques. *Journal of Hydrology,* vol. 377, no. 10, pp. 284-299, 2009.

Koning, A. J., Franses, P. H., Hibon, M., and Stekler, H. O. The M3 competition: Statistical tests of the results. *International Journal of Forecasting,* vol. 21, no. 7, pp. 397-409, 2005.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics,* vol. 54, no. 10, pp. 159-178, 1992.

Leusin, M. E., Kück, M., Frazzon, E. M., Maldonado, M. U., and Freitag, M. Potential of a multi-agent system approach for production control in smart factories. *IFAC-PapersOnLine*, vol. 51, no. 11, pp. 1459-1464, 2018.

Makridakis, S., and Hibon, M. The M3-Competition: results, conclusions and implications. *International Journal of Forecasting,* vol. 16, no. 10, pp. 451-476, 2000.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting,* vol. 34, no. 10, pp. 802-808, 2018.

Makridakis, S., and Winkler, R. L. Averages of Forecasts: Some Empirical Results. *Management Science,* vol. 29, no. 9, pp. 987-996, 1983.

Martins, V. L. M., and Werner, L. Forecast combination in industrial series: A comparison between individual forecasts and its combinations with and without correlated errors. *Expert Systems with Applications,* vol. 39, no. 10, pp. 11479-11486, 2012.

Ministry of Infrastructure, *Carajás Railroad.* Available at: https://www.infraestrutura.gov.br/component/content/article/84-ferrovias/4593-estrada-de-ferro-carajas.html, 2019

Montgomery, D. C., Jennings, C. L., and Kulahci, M. *Introduction to time series analysis and forecasting.* s.l.:John Wiley & Sons, 2015.

Pang, G., and Gebka, B. Forecasting container throughput using aggregate or terminal-specific data? The case of Tanjung Priok Port, Indonesia. *International Journal of Production Research,* vol. 55, no. 8, pp. 2454-2469, 2016.

R Core Team. *R: A Language and Environment for Statistical Computing.* [Online] Available at: https://www.R-project.org/, 2018.

Rashed, Y., Meersman, H., Voorde, E. V., and Vanelslander, T. Short-term forecast of container throughout: An ARIMA-intervention model for the port of Antwerp. *Maritime Economics & Logistics,* vol. 19, no. 11, pp. 749-764, 2017.

Said, E., and Dickey, D. A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika,* vol. 71, pp. 599-607, 1984.

Schulze, P. M., and Prinz, A. Forecasting container transshipment in Germany. *Applied Economics,* vol. 41, no. 10, pp. 2809-2815, 2019.

Stock, J. H., and Watson, M. W. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting,* vol. 23, no. 9, pp. 405-430, 2004.

Theil, H. *Applied economic forecasting.* Chicago: Rand McNally, 1966.

Timmermann, A. Forecast combinations. *Handbook of economic forecasting,* vol. 1, pp. 135-196, 2006.

Trapletti, A., and Hornik, K. *tseries: Time Series Analysis and Computational Finance.* [Online] Available at: https://CRAN.R-project.org/package=tseries, 2018.

Vieira, G. E., Kück, M., Frazzon, E., and Freitag, M. Evaluating the robustness of production schedules using discrete-event simulation. *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 7953-7958, 2017.

Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach. *Transportation Research Part C: Emerging Technologies,* vol. 13, no. 6, pp. 211-234, 2005.

Wallis, K. F. Combining forecasts – forty years later. *Applied Financial Economics*, vol. 21, no. 12, pp. 33-41, 2011.

Wang, F.-K., and Chang, K.-K. Adaptive neuro-fuzzy inference system for combined forecasts in a panel manufacturer. *Expert Systems with Applications,* vol. 37, no. 12, pp. 8119-8126, 2010.

Wang, X., and Petropoulos, F. To select or to combine? The inventory performance of model and expert forecasts. *International Journal of Production Research,* vol. 54, no. 3, pp. 5271-5282, 2016.

Winters, P. R. Forecasting sales by exponentially weighted moving averages. *Management science,* vol. 6, pp. 324-342, 1960.

Zhang, G. P., Patuwo, B. E., and Hu, M. Y. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers & Operations Research,* vol. 28, no. 4, pp. 381-396, 2001.

Biographies

Ícaro Romolo Sousa Agostino is PhD student in Production Engineering (Federal University of Santa Catarina, Brazil (2019 - current) and Visiting Researcher at Bremer Institut für Produktion und Logistik GmbH, Germany (2019 - current). Has a Master degree in Production Engineering (Federal University of Santa Maria - UFSM/2017 - 2019) and B.S. in Production Engineering (Ceuma University/2012 - 2017). Research Assistence at the Production and Logistics Intelligent Systems (ProLogIS/UFSC). Was Research Assistance at the Laboratory of Statistical Analysis and Modeling (LAME/UFSM). Research Interests: Industrial Engineering, with emphasis on Forecasting, Statistical Modeling, Discrete-Event Simulation and Quantitative Methods for Decision Making.

Satie Ledoux Takeda-Berger is a Ph.D. student in Production Engineering at Federal University of Santa Catarina (UFSC), Brazil (2018-current). She has a Master's degree in Production Engineering at the Federal University of Santa Catarina (UFSC) and B.S. in Production and Systems Engineering at the State University of Santa Catarina (UDESC). Member of the Laboratory of Intelligent Productive and Logistics Systems (ProLogIS) and participates in the Productive and Logistics Systems CNPq research group (S-ProLog). Her research interests are in Industrial Engineering with an emphasis on Production Planning and Control, Discrete-Event Simulation and Lean Supply Chain.

Enzo Morosini Frazzon is Associate Professor at the Industrial and Systems Engineering Depart. of the Federal University of Santa Catarina, Florianópolis/Brazil (2010 - ...), has a Dr.-Ing. from the Univ. of Bremen/Germany (2006 - 2009), was Post-doctoral Researcher at BIBA/Bremen (2009 - 2010) and was Visiting Scholar at the University of Parma/Italy (2018 - 2019). After graduating in Mechanical Engineering at Federal University of Santa Catarina (1994 - 1998) he acquired very relevant practical experience as an employee of VWAudi (1998 - 2001) and Arcelor Mittal (2002 - 2003). Later on, in parallel with a Master in Production Engineering at Federal University of Santa Catarina (2003 - 2005). Lead researcher ProLogIS/UFSC (http://www.ProLogIS.ufsc.br).

Adriano Mendonça Souza has a degree in Mathematics from Imaculada Conceição College (1988), a specialization in Statistics and Quantitative Modeling from the Federal University of Santa Maria (1991), a Master in Production Engineering from the same institution (1993) and a PhD in Production Engineering from the Federal University of Santa Catarina (2000) and during this period he worked as a researcher at Texas A & M University, did post-doc at the Instituto Superior da Ciência do Trabalho e da Empresa - ISCTE - Business School - Lisbon Institute - Portugal. He is currently Professor at the Department of Statistics of the Federal University of Santa Maria. He has experience in applied statistics, quality control, time series analysis, multivariate analysis, operational research and production management.