A comparison between ARIMA, LSTM, ARIMA-LSTM and SSA for cross-border rail freight traffic forecasting: the case of Alpine-Western Balkan Rail Freight Corridor

Miloš Milenković (✉ m.milenkovic@sf.bg.ac.rs)

University of Belgrade - Division for Management in Railway

Miloš Gligorić

University of Belgrade

Nebojša Bojović

University of Belgrade - Division for Management in Railway

Zoran Gligorić

University of Belgrade

Research Article

Abstract

Excessive delays of railway traffic at border crossing points as a consequence of poor planning of border crossing procedures lower the performance of train service, increase its cost and reduce the satisfaction of shippers. Mid-term prediction of traffic flows may improve the process of planning border-crossing activities. In this paper, we model the intensity of cross-border railway traffic on the Alpine-Western Balkan Rail Freight Corridor (AWB RFC). For each of the four border crossing points: Dimitrovgrad, Presevo, Sid and Subotica, time series composed of 102 monthly export and import railway freight traffic observations are used for training and testing of alternative forecasting models. Traditional ARIMA, Long-Short-Term Memory (LSTM) neural network, hybrid ARIMA-LSTM and Singular Spectrum Analysis (SSA) models, are fitted to each of the eight time series. To enable the practical applicability of the proposed approach the "Best fit forecast" tool is developed.

1. Introduction

Rail freight transport represents the most environmentally friendly mode of transport and plays an important role in the freight transport market. Besides its environmental advantage rail freight transport can provide more reliable, safer, cheaper and faster transport service under the higher level of harmonisation of transport and technological processes. Increasing requirements in terms of quality and availability of rail freight services in Europe have led to the need for the creation of a single European rail area by establishing international rail corridors for competitive freight.

The Alpine-Western Balkan Rail Freight Corridor (AWB RFC) belongs to a recently extended network of TEN-T corridors. This corridor connects central Europe and South-East Europe and together with branches Xb (Nis-Sofia-Istanbul) and Xc (Belgrade-Novi Sad-Budapest) brings significant improvements to railway transport in the direction from Central Europe to Turkey. The AWB RFC is the first RFC to include Serbia in the European rail network for competitive freight. However, to utilize the potential of AWB RFC it is needed to improve services and infrastructure along the RFC. One of the main bottlenecks is inefficient the border-crossing process. The average stopping times of freight trains at the AWB RFC border crossing points (BCP) are around several hours. These times can be much longer leading to disturbances in train traffic from both directions. The main reasons lie in an insufficient number of locomotives, an insufficient number of train crews and a low level of synchronization between neighbouring railway infrastructure managers. Besides the technical means (such as equipping the BCPs with communication technologies) for eliminating these bottlenecks, technological improvements may enable more efficient and proactive decision-making facilitators of technological improvements can be advanced forecasting of import and export train flows that are passing through a BCP. Forecasting of railway border crossing traffic may represent an essential component of planning and control of the border crossing process. Accurate prediction of import and export train traffic may contribute to reduced transit time on corridors, increased reliability of rail service, reduced cost for the operator and shipper and higher satisfaction of shipper which will ultimately lead to a higher share of rail transport.

Several forecasting approaches are applicable to forecasting train traffic through BCPs. In the context of this paper, all forecasting approaches can be classified as linear, nonlinear and hybrid. Traditional linear models include historical average (Smith and Demetsky 1997; Stephanedes et al. 1981), smoothing techniques (Williams et al. 1998) and Autoregressive Integrated Moving Average (ARIMA) models (Box and Jenkins 1976; Box et al. 2008; Milenkovic et al. 2016). Purely linear models have limited performance in real-world time series modelling which is commonly characterized by a mix of linear and nonlinear temporal patterns. Artificial Neural Networks (ANN) represent the most popular nonlinear forecasting models. ANNs can handle complex patterns and generate models which adequately reflect nonlinear relationships (Milenkovic et al. 2019). ANNs represent non-parametric models, therefore, there is no need to define an explicit model form. The model is adaptively determined based on the characteristics of the time series. However, the adoption of single ANN may not be sufficient for modelling both linear and nonlinear patterns well due to problems with misspecification, under-fitting or overfitting of the model (Santos Junior, de Oliveira, de Mattos Neto, 2019). Therefore, since the real-world time series are almost always linear and nonlinear in terms of their correlation structure, neither ARIMA nor ANN can be solely applied to adequately model the time series. In that scenario, hybrid approaches that combine classical statistical models and ANNs are providing a good choice in terms of accuracy. Hybrid ARIMA-Long Short-term Memory (LSTM) neural network which hybridizes the ARIMA model and LSTM model to obtain the linear tendency and nonlinear tendency represents one very competitive alternative in comparison with other approaches based on many recent contributions (Khozani et al. 2022; Deng et al. 2020; Fan et al. 2021; Manowska et al. 2021; Abebe et al. 2022).

In this paper, we model cross-border train traffic flow time series by using the traditional ARIMA approach, Long Short Term Memory (LSTM), hybrid ARIMA-LSTM and Singular Spectrum Analysis (SSA) approach. The traditional Box-Jenkins method is used for fitting the ARIMA models. In comparison with ARIMA models, LSTM models are capable to look for nonlinear, non-stationary and intermittent or transient behaviour in an observed time series. The third used approach is based on combining the ARIMA with LSTM neural network to capture the linear and nonlinear dynamics in time series. This process is composed of three levels: linear modelling, non-linear modelling and prediction of future values (Phan and Nguyen 2020; Khozani et al. 2022). SSA represents an additional non-parametric technique based on the concept of separability between signal and noise components (Golyandina et al. 2001). Comparisons of proposed approaches are performed for Dimitrovgrad, Presevo, Sid and Subotica border crossing points, all within the railway network of Serbia. Performances of approaches were evaluated according to a set of relevant criteria independently for export and import train traffic flows. At the time of our analysis, a time series of monthly train traffic flows from January 2013 to June 2021 (102 monthly observations) were available. To make proposed techniques usable for the end-users (planners and managers at BCPs) a forecasting software application is developed. This application is based on a "best-fit forecasting" principle. It compares all considered forecasting models, automatically calculates the error for each model and assigns the forecast model to the forecasted train traffic flow.

The paper is organized as follows. In the next section, a review of relevant literature is given. Section 3 describes the methodology used in the modelling and forecasting of train traffic flows through selected

border crossing points. In Section 4 proposed models have been tested on eight different time series related to the export and import flows of Dimitrovgrad, Sid, Presevo and Subotica border crossing points. Performances of proposed models are compared in Section 4. Section 5 describes the best fit forecasting tool for the practical use of developed approaches. Concluding remarks and future research directions are given in the last section.

2. Literature Review

There are different classifications of quantitative forecasting approaches. All quantitative approaches can be decomposed into univariate or projective and multivariate or causal (Milenkovic et al. 2019). Both categories can be decomposed into parametric and non-parametric (Milenkovic et al. 2016) whereas within these two groups there are linear, nonlinear and hybrid approaches (Petropoulos et al. 2022). There is a plethora of contributions in literature which address the traffic flow forecasting problems. Excellent reviews can be found in Medina-Salgado et al. (2022), Kashyap et al. (2022), Liu et al. (2020) and Vlahogianni et al. (2004). Relevant transport flow forecasting contributions are discussed in this section.

Traditional linear or parametric forecasting techniques have been widely used for forecasting transport flows over the last few decades. Li (2013) compared exponential smoothing and brown exponential smoothing for freight turnover forecasting. Moiseev (2021) applies an exponential smoothing model in the oil shipping market forecasting. Alhindawi et al. (2020) applied double exponential smoothing for the projection of GHG emissions from the road transport sector.

In the domain of passenger traffic, Grubb and Mason (2001) used the Holt-Winters method for a very long time series about air passenger traffic. Jighjigh et al. (2021) formulated a multiplicative Holt-Winters method to forecast the volume of passenger traffic in Nigerian airports in the future. Dantas et al. (2017) combined the Bootstrap aggregating (Bagging) method with the exponential smoothing Holt Winters to predict future demand for passenger air transportation. Ge et al. (2013) compared exponential smoothing with the trend moving average method for bus passenger traffic prediction. Zhi-Peng et al. (2008) proposed an improved adaptive exponential smoothing model for short-term travel time forecasting of the urban arterial street. Sitzimis (2022) compares Winters' multiplicative method, simple seasonal model, decomposition multiplicative trend and seasonal model with Box-Jenkins ARIMA approach for forecasting passenger traffic in Greek coastal shipping. Milenkovic et al. (2016) developed Seasonal Autoregressive Integrated Moving Average (SARIMA) model for forecasting railway passenger demand. Williams et al. (2003), model short-term traffic condition data streams as SARIMA processes. Ding et al. (2011) propose a space-time autoregressive integrated moving average (STARIMA) model to predict the traffic volume in urban areas. Chen et al. (2009) proposed the Holt-Winters method, the seasonal ARIMA (SARIMA) model, and the GM(1,1) grey forecasting model to replicate monthly inbound air travel arrivals to Taiwan and to compare the models' forecasting performance. Kim et al. (2011) forecasted future customer volume by using the ARIMA model, time series analysis and system dynamics. Chen et al. (2009) applied the Holt-Winters method, the seasonal ARIMA (SARIMA) model, and the GM(1,1) grey forecasting model to replicate monthly inbound air travel arrivals to Taiwan and to compare the models'

forecasting performance. Miller (2018) applies the autoregressive integrated moving average (ARIMA) methodology to develop forecasts for three time series of monthly archival trucking prices. Tang and Deng (2016) applied ARIMA(1,1,8) model to adequately fit the civil aviation passenger turnover. Kumar and Vanajakshi (2015) proposed a prediction scheme using the Seasonal ARIMA (SARIMA) model for short-term prediction of road traffic flow using only limited input data.

Nonlinear dynamics forecasting models represent data-driven self-adaptive methods capable to learn from examples and capture complex functional relationships within the time series. Jiang and Luo (2022) made a comprehensive review of the application of Graphical Neural Networks for road traffic forecasting. Wei and Chen (2012) combined empirical mode decomposition (EMD) and back-propagation neural networks (BPN) is developed to predict the short-term passenger flow in metro systems. Murat and Ceylan (2006) developed an ANN model for transport energy demand forecasting. Dia (2001) developed an object-oriented neural network model for freeway short-term traffic forecasting. Peng et al. (2020) proposed a spatial-temporal incidence dynamic graph neural networks framework for urban traffic passenger flow prediction. Jiang and Adeli (2005) developed a nonparametric dynamic time-delay recurrent wavelet neural network model is developed for forecasting road traffic flow. Huang et al. (2021) proposed the Grey model GM (1,1) and Back Propagation (BP) neural network model for simulation and forecasting the logistics demand of Guangdong province from 2000 to 2019. Gallo et al. (2019) proposed Artificial Neural Networks (ANNs) approach for forecasting metro on-board passenger flows as a function of passenger counts at station turnstiles. Mostafa M. M. (2004) investigated the Suez Canal traffic forecasting and compared the performance of ARIMA models with that of neural networks on an example of a large monthly dataset. Blinova (2007) proposed a neural network model to forecast air intraregional and interregional passenger traffic flows. Jiang et al. (2014) developed a short-term demand forecasting approach by combining the ensemble empirical mode decomposition (EEMD) and grey support vector machine (GSVM) models. Zhang and Liu (2009) proposed least squares support vector machines (LS-SVMs) to forecast the travel time index. Wang and Shi (2013) proposed a short-term traffic speed forecasting hybrid model (Chaos−Wavelet Analysis-Support Vector Machine model to model the real traffic speed data. Cong et al. (2016) combined the least squares support vector machine (LSSVM) fruit fly optimization algorithm (FOA), to study the potential of traffic flow forecasting. Wang et al. (2018) proposed a novel hybrid model combining the support vector machine overall online (SVMOOL) model and the support vector machine partial online model (SVMPOL) model for short-term metro ridership forecasting. Bao et al. (2012) proposed an ensemble empirical mode decomposition (EEMD) based support vector machines (SVMs) modelling framework incorporating a slope-based method (EEMD-Slope-SVMs) to model the monthly air passenger traffic series including six selected airlines in USA and UK. Bin et al. (2006) analysed the feasibility and applicability of SVM in bus arrival time prediction. Syriopoulos et al. (2021) developed a support vector regression (SVR) model for ship price predictions for different vessel types and shipping markets. Zhongzhen et al. (2011) proposed a wavelet transform-SVM combined model to forecast the freight index of Panamax bulk carriers. Based on the Baltic Supermax Index and historical decision data of different companies, Guan et al. (2019) used the support vector machine model to predict the dry bulk carrier route selection. Cheu et al. (2006) compared two trip

forecasting approaches, namely neural networks and support vector machines, for a multiple-station shared-use vehicle system. Yang et al. (2016) developed the prediction model of bus arrival time based on a Support Vector Machine with a genetic algorithm (GA-SVM). Milenkovic et al. (2019) proposed a fuzzy neural network prediction approach based on metaheuristics for container flow forecasting. Glisovic et al. (2016) presented a hybrid model based on the integration of the genetic algorithm (GA) and the artificial neural networks (ANN) for forecasting the monthly volume of passengers on the Serbian railways.

Nonlinear models such as ANN are usually difficult to interpret and test for the statistical significance of the parameters (Medeiros and Veiga 2000). Therefore, over recent years, new formulations appeared that combine the traditional linear time series models and nonlinear models to handle both linear and nonlinear structures in time series equally well.

For the prediction of the number of goods subject to inspection at European Border Inspections Post Ruiz-Aguilar et al. (2014) applied a hybrid two-step procedure based on integrating the data obtained from autoregressive integrated moving averages (SARIMA) model in the artificial neural network model (ANN). Shahriari et al. (2020) combined bootstrap with the conventional parametric ARIMA model to improve the prediction accuracy of traffic count on four main arterial roads in Sydney. A hybrid model combining symbolic regression and Autoregressive Integrated Moving Average Model (ARIMA) was proposed by Li et al. (2017) for metro passenger flow forecasting. Xie et al. (2014) proposed two hybrid approaches based on seasonal decomposition and the least squares support vector regression (LSSVR) model for short-term forecasting of air passengers. Xu et al. (2019) proposed the SARIMA-SVR model is proposed to forecast statistical indicators in the aviation industry. Ge et al. (2021) proposed a hybrid of ARIMA and fuzzy support vector regression machine (FSVR) to predict the passenger flow at the Shanghai-Guangzhou high-speed railway.

Based on this detailed and comprehensive review of relevant literature we may draw the following conclusions:

- Forecasting models were developed mainly for non-rail and/or passenger transportation;
- To our knowledge, there are no attempts in the past that paid particular attention to forecasting traffic flow on railway border crossings;
- None of the reviewed contributions do consider the practical usability of the proposed approach.

Therefore, to eliminate these research gaps we model the cross-border train traffic flows by applying ARIMA, LSTM, ARIMA-LSTM and SSA approaches and propose a best-fit forecasting tool, desktop client application, as a solution for practical use of developed approaches. Comprehensiveness as an additional feature of this approach is based on separate analysis, assessment and forecasting of import and export train flow for each of the four border crossing points which produces a more detailed input to the corridor managers as support for solving a diversity of corridor-related decision-making problems.

3. Methodology

The flow chart in Fig. 1. illustrates the methodology applied in this paper. Each forecasting approach requires a set of steps to be conducted. ARIMA modelling is based on Box and Jenkins methodology that includes identification of a suitable model, estimation of parameters and diagnostic checking. LSTM neural network includes the transformation of a sequence of observations and the definition of the model. ARIMA-LSTM is composed of filtering the linear tendencies in the data and passing on the residual values to the LSTM model. Decomposition and reconstruction represent the main steps of the SSA approach. Included forecasting models are fitted on a training data sample and their optimal configurations are selected and compared on a test data sample. The model which produces the best performances can be used for forecasting freight train flows at a specific border-crossing. Included approaches are packed into a desktop software application named as Best-Fit Forecasting (BFF) tool to provide its practical usability.

The following subsections contain a brief description of included approaches.

3.1. ARIMA

ARIMA models represent one of the most widely used linear univariate time series modelling techniques. ARIMA models are composed of the Autoregressive (AR) model, the Moving Average (MA) model and ARMA as a combination of AR and MA models. AR model includes a linear combination of past values of the variable. AR model of order p (AR(p)) can be expressed as:

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t$$

1

Where c is a constant and ε_t is a white noise sequence assumed to be a normal random variable with zero mean and variance σ^2.

MA model uses past forecast errors as forecast variables. MA model of order q (MA(q)) can be expressed as:

$$Y_t = c + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

2

To apply the ARIMA models the time series needs to be stationary. The letter "I" (integrated) means that the first-order difference is applied to transform a considered time series into stationary. The full ARIMA model can be written as follows:

$$Y'_t = c + \sum_{i=1}^{p} \phi_i Y'_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

3

Y'_t is differenced time series. Equivalent integrated form of any ARIMA model looks as follows:

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B)\varepsilon_t$$

4

B represents the backshift operator, whose effect on a time series Y_t can be summarized as:

$$B^d Y_t = Y_{t-d}$$

5

Seasonal ARIMA models (SARIMA) represent an extension to cover seasonal variations in the time series. The seasonal part of the model consists of terms similar to the non-seasonal part, the difference is that the seasonal part includes backshifts of the seasonal period (not the first period like in the ARIMA model). In integrated form SARIMA model can be expressed as follows:

$$\phi_p(B)\Phi_P(B^S)(1 - B)^d(1 - B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

6

Equation (5) represents a $\mathrm{SARIMA}(\mathrm{p,d,q}) \times (\mathrm{P,D,Q})$ model where Φ and Θ are the seasonal ARMA coefficients and seasonal differencing operator $(1 - B^S)^D$ of order is applied to eliminate seasonal patterns. The modelling procedure of an ARIMA (SARIMA) model can be summarized in Fig. 2. Arima modelling was conducted in Python 3.10 using the pmdarima package (Smith et al., 2017).

3.2. Long Short Term Memory (LSTM) models

LSTM represents a type of Recurrent Neural Network (RNN). RNNs, due to their looped architecture, are capable to recognize sequential characteristics in data and therefore very suitable for sequence prediction problems (Hewamalage 2021). However, RNNs do suffer from a long-term dependency problem. Due to vanishing and exploding gradient issues RNNs, have difficulty in learning long-term dependencies (Somu et al. 2020). Another issue with standard RNN architecture is that the training of RNN model requires a predetermined delay window length, but it is difficult to automatically obtain the optimal value of this parameter (Xu et al. 2022).

LSTM provides a solution for long term dependency problems thanks to improved recursive neural network architecture with feedback so it can process not only individual data points but entire sequences as well (Sepp and Schmidhuber 1997; Manowska et al. 2021). A LSTM neural network is composed of

one input layer, one recurrent hidden layer and one output layer. Improvement in architecture relates to replacing the hidden layer of RNN cells with LSTM cells (hereafter memory blocks) to achieve long-term memory. Self-connected LSTM memory blocks enable the model to learn the long-term dependencies while handling sequential data (Somu et al. 2020). In contrast to RNNs which have only one hidden state, in LSTM neural network to each cell two states are transferred, the cell state and the hidden state. The cell state enables long-term memory capability, whereas the hidden state enables a working memory capability that contains only near-past information and uncontrollably overwrites at every step.

Memory blocks are responsible for memorizing, and manipulations between blocks are done by special multiplicative units called gates. Gates control the flow of information (Ma et al. 2015; Hrnjica and Mehr 2020). The input gate controls the flow of input activations into the memory cell. The output gate controls the output flow of the cell activation. Besides these two gates, there is also a forget gate which filters the information from the input and previous output and decides which one should be remembered, forgotten and dropped (Hrnjica and Mehr 2020). The adaptive gating mechanism enables that whenever the content of the cell is outdated, the forget gate resets the cell state, so the input and the output gates control the input and the output, respectively. In essence, the gates represent different neural networks that decide which information is allowed in the cell state. Besides the gates, the core of the memory cell is a recurrently self-connected linear unit-Constant Error Carousel (CEC), whose activation represents the cell state. Due to the presence of CEC the problem of vanishing or exploding gradient is solved since multiplicative gates can learn to open and close enabling the LSTM cell state to enforce the constant error flow.

Figure 3. provides an insight into the internal architecture of LSTM. Symbols , , and represent the input gate, forget gate the cell state vector and the output gate, respectively. σ and tanh are sigmoid and hyperbolic tangent activation functions, respectively. The elementwise multiplication of two vectors is denoted by $\otimes$.

For the tth data, an LSTM takes as input x_t, h_{t-1}, C_{t-1} and produces the hidden state h_t as well as the cell state C_t based on the following formulas:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

7

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

8

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \underline{C}_t$$

9

$$\underline{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

10

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

11

$$h_t = o_t \otimes \tanh(C_t)$$

12

where $W_k \in R^{n \times m}$, $U_k \in R^{n \times n}$ are weight matrices and $b_k \in R^n$ are bias vectors, for $k = \{i, f, c, o\}$. The symbols $\sigma(\cdot)$ and $\tanh(\cdot)$ refer to element-wise sigmoid and hyperbolic target functions. Element-wise multiplication is represented by the symbol $\otimes$.

In this paper, the LSTM network was built based on the Keras framework of the Python 3.10 platform. Before modelling each training data that belong to a time series for each border crossing is normalized or rescaled from the original range so that all values are within the range of 0 and 1. Then the methods and parameters of the LSTM model need to be configured. Depending on the time series, the hidden layer was built from 100 to 200 LSTM cells, the number of iterations varied from 100 to 300 and the batch size spanned from 2 to 12. The activation function was set to rectified linear activation function (ReLu), the loss function was MSE, and the optimizer was stochastic gradient descent (SGD).

3.3. ARIMA-LSTM models

ARIMA models can only recognize linear patterns of the time series whereas LSTM models are capable to mine the nonlinear relationships. Since it is proven that some of the time series related to rail cross-border traffic have some remaining nonlinearity, a hybrid ARIMA and LSTM model was proposed as an alternative model for forecasting cross-border traffic.

The modelling flow chart of the ARIMA-LSTM approach is shown in Fig. 4. In essence, the residuals of ARIMA models are regarded as the input of the LSTM model, and the LSTM model is utilized to train the nonlinear tendency by modelling the residual series (Deng et al. 2020). The linear part (L_t) and nonlinear part (N_t) were combined to obtain the prediction results (Y_t) of the hybrid ARIMA-LSTM model:

$$Y_t = L_t + N_t$$

13

3.4. Singular Spectrum Analysis (SSA) algorithm

Singular Spectrum Analysis algorithm is a very useful tool to forecast different phenomena in the field of transportation. Many of authors use the SSA algorithm to filter monitored time series i.e. to separate signal component from noisy components. After that, the signal component is used as a principal component to forecast future values of time series and different methods can be used to forecast these values. Such approaches are known as hybrid forecasting algorithms.

Kolidakis et al. (2020) developed the hybrid model which is composed of Singular spectrum analysis and Artificial neural networks to forecast intraday traffic volume. Shang et al. (2016) used Singular spectrum analysis to filter traffic flow data and Kernel extreme learning machine to make prediction of short-term traffic flow. Zhou et al. (2020) forecasted passenger flow in metro transfer station by novel model which is composed of Singular spectrum analysis and AdaBoost-Weighted Extreme Learning Machine. Shuai et al. (2021) applied Singular spectrum analysis, Long-short memory, and Support vector regression for the purpose of the traffic flow prediction of expressway.

This section explains the Singular Spectrum Analysis (SSA) algorithm for time series forecasting. For more information, readers are referred to the papers of Hassani and Zhigljavsky (2009), Hassani and Mahmoudvand (2013), Hassani (2007), Harris and Yan (2010). Briefly, the algorithm is composed of the two main stages: decomposition and reconstruction. In the first stage, monitored time series is represented as a spectrum of independent components such as trend, periodic oscillatory and noise components. In the second stage, monitored time series is reconstructed by using the less noisy components, i.e., principal components of the time series are created.

In the decomposition stage, there are two following steps: embedding and singular value decomposition. The embedding process transforms one-dimensional time series into a matrix of $L \times K$ dimension, where is the window length, equals $N + 1 - L$ and N is the amount of data. The size of the window length should be an integer within the following interval $2 \leqslant L \leqslant \frac{1}{2}N$ The concept of the window length is similar to the concept of the $k - th$ order autoregressive model of time series, but taking into account original data from $t = 1$ to $t = L$. Consider a stochastic process $\{x(t); t = 1, 2, \ldots, N\}$ and suppose there are realizations of $X(t) = \{x(1), x(2), \ldots, x(N)\}$ this process. Set $X(t)$ is time-invariant series and for simplicity, we can rewrite it as follows $X_N = \{x_1, x_2, \ldots, x_N\}$. An output of the embedding process is the trajectory matrix of the following form:

$$
H = |H_1, H_2, \ldots, H_K| = \left|x_{ij}\right|_{i,j=1}^{L,K} =
\begin{vmatrix}
x1 & x2 & \cdots & xK \\
x2 & x3 & \cdots & xK+1 \\
\vdots & \vdots & \ddots & \vdots \\
xL & xL+1 & \cdots & xN
\end{vmatrix}
$$

14

The trajectory matrix is known as Hankel matrix and all elements along diagonal are equal and $i + j = const$. Main objective of the singular value decomposition is to express Hankel matrix as a sum of weighted orthogonal matrices. Spectral decomposition is performed over the lag-covariance matrix $HH^T \in R^{L \times L}$. Let $\sigma_1, \sigma_2, \ldots, \sigma_L$ be eigenvalues (singular values) of HH^T arranged in decreasing order $\sigma_1 \geqslant 0, \sigma_2 \geqslant 0,\ldots,\sigma_L \geqslant 0$ and $U_1, U_2, \ldots, U_L$ be corresponding eigenvectors. If the number of nonzero eigenvalues equals the rank of matrix , then trajectory matrix can be represented as:

$$\widehat{H} = \sum_{i=1}^{r} U_i U_i^T H = \widehat{H}_1 + \widehat{H}_2 + \ldots + \widehat{H}_r$$

15

where is the rank of H.

The second stage, called reconstruction, is accomplished in two steps: grouping and diagonal averaging or Hankelization. In the grouping step, set of matrices $\{\widehat{H}_1, \widehat{H}_2, \ldots, \widehat{H}_r\}$ are divided in several disjointed subsets: $\{\widehat{H}_1, \widehat{H}_2, \ldots, \widehat{H}_r\} \to \{E_1, E_2, \ldots, E_m\}$, $m < r$. After that, all matrices within each subset are summed. The simplest case refers to the signal and noise component respectively: $\{\widehat{H}_1, \widehat{H}_2, \ldots, \widehat{H}_r\} \to \{E_1, E_2\}, m = 2, E_1 \cap E_2 = \emptyset$. In that case there are only two subsets, $E_1 = \sum_{i=1}^{d} \widehat{H}_i$ and $E_2 = \sum_{i=d+1}^{r} \widehat{H}_i$. Subset E_1 is associated with signal component while E_2 is associated with noise component. Selection of the appropriate value of is based on the plot of logarithms of singular values, $\log \sigma_1, \log \sigma_1, \ldots, \log \sigma_L, \forall \sigma > 0, \ \log \sigma_1, \ \log \sigma_2, \ldots, \ \log \sigma_L$. Point in the plot, where a significant drop in values occurs, can be treated as the start of noise components. Diagonal averaging represents the transformation of each reconstructed trajectory matrix (16) into new time series of length . Matrix elements over anti-diagonals are averaged, $i + j = k + 1$. Reconstructed trajectory matrix is of the following form:

$$\widehat{H} = UU^T H = |h_{ij}|_{i,j=1}^{L,K} = \begin{vmatrix} h_{11} & h_{12} & \cdots & h_{1K} \\ h_{21} & h_{22} & \cdots & h_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ h_{L1} & h_{L2} & \cdots & h_{LK} \end{vmatrix}$$

16

Elements of the new time series are extracted from $\widehat{H}$ by following calculations:

$$\widehat{X}_N = \{x_1, x_2, \ldots, x_N\} = \{h_{11}, \frac{h_{21} + h_{12}}{2}, \frac{h_{31} + h_{22} + h_{13}}{3}, \ldots, h_{LK}\}$$

17

Finally, original time series X_N is expressed as a sum of d principal vectors:

$$X_N = \sum_{i=1}^{d} \overrightarrow{\widehat{X}}_{iN} = \overrightarrow{\widehat{X}}_{1N} + \overrightarrow{\widehat{X}}_{2N} + \ldots + \overrightarrow{\widehat{X}}_{dN}$$

18

New time series is used to forecast future values for $N + 1, N + 2, \ldots, N + p$. Linear recurrent formulae is used to forecast future values of monitored time series. Eigenvector obtained by the singular value decomposition is as follows:

$$U = [u_1 \; u_2 \; \ldots \; u_{L\text{-}1} \; u_L]^T$$

19

Let $U^\nabla = [u_1 \; u_2 \; \ldots \; u_{L\text{-}1}]^T$ be the subset of composed of the first $L - 1$ coordinates, and $\pi = \{u_L\}$ be the last coordinate of the eigenvector. Accordingly, the verticality coefficient is defined as:

$$v^2 = \sum_{i=1}^{d} \pi_i^2 = \pi_1^2 + \pi_2^2 + \ldots + \pi_d^2$$

20

Condition $v^2 < 1$ meet if we want to use Singular spectrum analysis to forecast $p - values$ ahead. Obviously, the value of separates signal components from noise components, must be carefully selected to satisfy previous inequality as well.

The linear vector of coefficients $R = [\beta_{L-1}, \beta_{L-1}, \ldots, \beta_1]^T$ is calculated by the following equation:

$$R = \frac{1}{1 - v^2} \sum_{i=1}^{d} \pi_i U_i^\Delta$$

21

Future values forecasting is achieved by:

$$\{\tilde{x}(t)\}^T = \begin{cases} \{\tilde{x}(1), \tilde{x}(2), \ldots, \tilde{x}(t), t = 1, 2, \ldots, N\} \\ R^T X_p(t), t = N + 1, N + 2, \ldots, N + p \end{cases}$$

22

where

$$X_p(t) = \{\tilde{x}(N - L + p + 1), \ldots, \tilde{x}(N + p - 1)\}^T$$

23

Many time series encountered in the real-world exhibit nonstationary behaviour. The main reason is due to a presence of trend, seasonal variation, or a change in the local mean. For reducing a nonstationary series with trend to a stationary series (without trend) we apply the first differences of monitored time series as:

$$w(t) = x(t) - x(t-1), \ t = 2, 3, \ldots, N$$

24

The differenced data are often easier to model than the original data. Hence, Singular spectrum analysis is now performed over the differenced time series $W_N = \{w_2, w_3, \ldots w_N\}$ According to the Eqs. (9) and (10), forecasting of differenced data is as follows:

$$\{\widetilde{w}(t)\}^T = \begin{cases} \{w(2), w(3), \ldots, w(t)\}, \ t = 2, 3, \ldots, N \\ R^T W_p(t), \ t = N+1, N+2, \ldots, N+p \end{cases}$$

25

where

$$W_p(t) = \{\widetilde{w}(N - L + p + 1), \ldots, \widetilde{w}(N + p - 1)\}^T$$

26

Since we want to forecast the original values X_N not the differenced values W_N we must transform the model from the $\{\widetilde{w}(t)\}$ form to the $\{\tilde{x}(t)\}$ form. Recall that $w(t) = x(t) - x(t-1)$ forecasted values become:

$$\tilde{x}(t) = \tilde{x}(t-1) + \widetilde{w}(t), \ t = 2, 3, \ldots, N$$

27

4. Results And Discussion

European Union (EU) has established RFCs in order support the competitiveness of rail freight transport. Alpine-Western Balkan Rail Freight Corridor (AWP RFC) connects four EU member states (Austria, Slovenia, Croatia, Bulgaria) and fully integrates the EU candidate state Serbia. The corridor represents the shortest route from Central Europe to the Bulgarian/Turkish border for rail freight (EU and ITT, 2019). Within the territory of Republic of Serbia on AWP RFC there are four main BCPs with different operational conditions and volumes of freight traffic. These are Dimitrovgrad, Presevo, Sid and Subotica. This study focuses on the observation and prediction of import and export cross-border freight train flows for these BCPs. The historical data for import and export flows are presented in Fig. 5. and Fig. 6., respectively. The time series data are obtained from the Public Enterprise "Serbian Railways". The sample data are monthly observations of freight train flows on four BCPs covering the period from January 2013 to June 2021. The first 96 monthly observations are used as a training dataset whereas the remaining 6 observations served for verification of selected models. Both types of traffic (import and export) for each BCP are independently investigated and appropriate models are estimated. ARIMA, LSTM and ARIMA-LSTM are implemented by the use of R and Python software packages, and SSA is developed in Python.

4.1. ARIMA results

Table 1. summarizes the process of ARIMA model selection. In a preliminary step, the time series is visually examined, the procedure for detecting outliers is applied and variance (mean-variance relationship) is analyzed. Presevo Export and Sid Import train flow time series contain an outlier (observation no. 50 and 71, respectively) which are replaced by linearly interpolated values using the neighbouring observations (Hyndman and Athanasopoulos 2018).

In the first step, the ADF test is applied for detecting non-stationarity. Then in the second step, based on a visual plot of ACF and PACF, a general structure of the ARIMA model is proposed. The third step includes determining the best model based on the criteria of minimum AICc. In the last step, Ljung-Box statistics is applied to check if the selected model is correctly specified.

Table 1
ARIMA model selection for railway border crossing points

Time series	Stationarity (ADF test: p-value)	Correlogram - significant values		Selected model	MAPE	AICc	Adj R^2	Ljung-Box Q test (p-value)
		ACF	PACF					
Dimitrovgrad								
Import	0.016	1st lag (0.284) 5th lag (0.200)	1st lag (0.284)	ARMA(2,1)	8.10	800.88	15.82	0.912
Export	0.011	1st lag (0.386)	1st lag (0.386)	ARMA(1,0)	9.04	810.35	14.09	0.841
Presevo								
Import	0.039	1st lag (0.277) 11th lag (-0.246)	1st (0.277) 11th lag (0.287)	ARMA(1,0)	14.41	816.65	6.71	0.723
Export	0.013	1st lag (0.369)	1st (0.369) 18th lag (-0.234)	ARMA(1,0)	12.18	837.57	12.75	0.791
Sid								
Import	0.279	1st lag (0.648) to 5th lag (0.274) 11th lag (0.227) 12th lag (0.341) 13th lag (0.316) 14th lag (0.211)	1st lag (0.648) 12th lag (0.202)	ARIMA(1,1,1)	9.59	792.81	53.7	0.754

Time series	Stationarity (ADF test: p-value)	Correlogram - significant values		Selected model	MAPE	AICc	Adj R^2	Ljung-Box Q test (p-value)
		ACF	PACF					
Export	0.149	1st lag (0.577) - 4th lag (0.206) 12th lag (0.380) 24th lag (0.229)	1st lag (0.577) 12th lag (0.328)	SARIMA (2,1,1)× (1,1,3)	9.61	816.49	51.5	0.633
Subotica								
Import	0.561	1st lag (0.571) to 10th lag (0.221)	1st lag (0.571) 2nd lag (0.217)	ARIMA(1,1,1)	7.89	881.08	40.2	0.939
Export	0.487	1st lag (0.581) to 10th lag (0.247)	1st lag (0.581) 3nd lag (0.274)	ARIMA(0,1,1)	7.32	872.93	42.8	0.413

BDS test is applied to check if there is remaining nonlinearity in time series. According to BDS test, nonlinearity exists if residuals dataset after fitting a selected ARIMA model still contain nonlinear components (the null hypothesis of i.i.d. for the residuals can be not rejected at the 5% level of confidence). In that case there are p-values which are lower than 0.05. for any of embedding dimensions (m = 2,3) and epsilon (0.5–1.5). m represents the embedding dimension. ε is equal to 0.5, 1.0 and 1.5 times the standard deviation. The critical value for confidence level of 5% is 1.96. Table 6. contains the outputs of BDS test for all considered border crossing points. From the results of the BDS test applied to the residuals of chosen ARIMA models, it can be noticed that in five time series there is a remaining nonlinearity in the residuals (bold p-values in Table 2.). In case of Presevo Import, Sid Import and Subotica Import time series according to BDS test the hypothesis of the nonlinearity in the residuals of the ARIMA models can be rejected.

Table 2
Non-Linearity testing for ARMA residuals of time series

Time series	Parameter ε/σ	Dimension (m = 2)		Dimension (m = 3)	
		Statistic	Probability	Statistic	Probability
Dimitrovgrad					
Import	0.5	0.5073	0.6119	-1.4224	0.1549
	1.0	2.5701	**0.0102**	1.7656	0.0775
	1.5	2.1101	**0.0349**	1.3466	0.1781
Export	0.5	-1.3262	0.1848	-2.5074	**0.0122**
	1.0	-0.0664	0.9471	-1.2114	0.2258
	1.5	-0.5545	0.5792	-1.6564	0.0976
Presevo					
Import	0.5	0.5392	0.5897	0.0787	0.9373
	1.0	1.4816	0.1384	0..8427	0.3994
	1.5	1.9536	0.0508	1.4294	0.1529
Export	0.5	2.0909	**0.0365**	1.2162	0.2239
	1.0	2.3033	**0.0213**	1.4670	0.1424
	1.5	2.0196	**0.0434**	1.3052	0.1918
Sid					
Import	0.5	-1.2372	0.2160	1.0224	0.3066
	1.0	0.1405	0.8883	-1.1165	0.2642
	1.5	0.2050	0.8376	-1.2092	0.2266
Export	0.5	-2.4533	**0.0142**	-2.6731	**0.0075**
	1.0	-1.2745	0.2025	-1.6213	0.1050
	1.5	-0.8544	0.3929	-1.5801	0.1141
Subotica					
Import	0.5	0.6169	0.5373	-0.5381	0.5905
	1.0	-0.0609	0.9520	-0.2339	0.8151
	1.5	0.4575	0.6473	0.2993	0.7647
Export	0.5	3.4974	**0.0005**	3.1223	**0.0018**

Time series	Parameter ε/σ	Dimension (m = 2)		Dimension (m = 3)	
		Statistic	Probability	Statistic	Probability
	1.0	-0.5917	0.5541	-0.2711	0.7863
	1.5	-0.4945	0.6209	-0.2038	0.8385

4.2. LSTM results

The best LSTM model configurations for each of the time series with associated MAPE value for train and testing samples are given in Table 3.

Table 3
LSTM models configuration

Time series	LSTM model configuration			MAPE	
	Number of cells in hidden layer	Number of epochs	Batch size	Train	Test
Dimitrovgrad					
Import	150	170	12	4.281	4.221
Export	100	150	12	6.430	3.684
Presevo					
Import	200	170	12	9.033	5.391
Export	200	200	12	3.898	8.788
Sid					
Import	160	300	3	7.604	19.217
Export	200	300	3	7.923	10.601
Subotica					
Import	160	300	12	1.783	3.757
Export	180	350	2	6.597	5.640

4.3. ARIMA-LSTM results

For each time series, residuals from selected ARIMA model are forwarded to LSTM model. The best ARIMA-LSTM model configurations for residuals of each of the time series with associated MAPE value for train and testing samples are given in Table 4.

Table 4

ARIMA-LSTM models configuration

| Time series | ARIMA - LSTM model configuration | | | | MAPE | |
| | ARIMA model | LSTM model | | | | |
		Number of cells in hidden layer	Number of epochs	Batch size	Train	Test
Dimitrovgrad						
Import	ARIMA(2,1)	160	180	12	8.4845	5.474
Export	ARIMA(1,0)	100	150	12	9.050	5.192
Presevo						
Import	ARIMA(1,0)	200	170	12	14.416	6.742
Export	ARIMA(1,0)	200	200	12	12.187	8.257
Sid						
Import	ARIMA(1,1,1)	150	170	12	5.487	15.856
Export	SARIMA(2,1,1)(1,1,3)	180	170	12	10.267	23.911
Subotica						
Import	ARIMA(1,1,1)	150	170	12	8.167	3.700
Export	ARIMA(0,1,1)	150	170	12	7.075	6.377

4.4. SSA results

SSA parameters for monitored time series are presented in Table 5.

Table 5
SSA parameters and the accuracy of proposed SSA models for considered time series

Time series	Window length	Principal components d	MAPE (%) Training data	MAPE (%) Forecasting data
Dimitrovgrad-Export	10	9	1.129	6.309
Dimitrovgrad-Import	10	9	1.020	6.594
Presevo-Export	10	8	2.702	9.692
Presevo-Import	10	8	2.820	4.684
Sid-Export	10	9	1.611	18.358
Sid-Import	10	9	1.559	14.182
Subotica-Export	10	7	2.300	6.954
Subotica-Import	10	7	2.473	7.455

4.5. Discussion of results

In this section, prediction accuracy of each of the four forecasting methods is analyzed. Actual observations, as well as the predictions generated by each of the proposed models for training and testing samples of import and output flows for each border crossing point are graphically illustrated on Fig. 7. (Import freight train flows) and Fig. 8. (Export freight train flows). Vertical dashed line divides the training and testing samples.

To test forecasting performance, we applied the mean absolute error (MAE), mean average percent error (MAPE) and root mean squared error (RMSE) defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i|$$

28

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} |1 - \frac{\overline{\overline{Y_i}}}{Y_i}|$$

29

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \overline{\overline{Y}}_i)}{n}}$$

30

where Y_i and $\overline{Y}_i$ represent the actual and predicted values of the time series in period , respectively. MAE represents the mean of absolute errors. MAPE is one of the most commonly used criteria to measure forecast accuracy. It represents the sum of the individual absolute errors divided by the actual observation. RMSE represents a square root of the average squared error.

Figure 9. graphically illustrates comparison of forecasting accuracy of proposed methods for import freight train flows on all border crossing points. In terms of performances on a training sample SSA shows the lowest values of MAE, MAPE and RMSE. On the other side, SSA shows the lowest predictive performances (the highest errors on test sample). If we compare average ratio between training and testing performances, ARIMA and ARIMA-LSTM have the lowest ratio (< = 1) whereas SSA significantly exceeds one ($RMSE(train/test) = 7.09, MAE(train/test) = 7.30,$ $MAPE(train/test) = 5.05$). On test sample, LSTM has the best average performance for modelling of import flows of all border crossings (MAE = 11.595, MAPE = 7.853, RMSE = 13.201). The highest levels of errors are in case of Sid Import, where the lowest errors on test and train sample has SSA model. If we focus only on Presevo Import, Sid Import and Subotica Import, which do not contain remaining nonlinearities (according to BDS test) we may conclude that linear ARIMA, as well as ARIMA-LSTM provide competitive performances.

Figure 10. graphically illustrates comparison of forecasting accuracy of proposed methods for export freight train flows on all border crossing points. Again, SSA shows the best performances on training the model, but the worst on testing sample. In case of export time series, in terms of average ratio between training and testing performances, SSA again significantly exceeds one ($RMSE(train/test) = 8.428$,$MAE(train/test) = 8.342, MAPE(train/test) = 5.898$). On test sample, LSTM has again the best average performance (MAE = 12.866, MAPE = 7.864, RMSE = 13.141). The highest levels of errors are again in the case of Sid Export, where the lowest errors on test and train sample has SSA model.

5. Best Fit Forecasting Tool

To enable the practical use of proposed forecasting methods, a desktop client application "Best Fit Forecast" (BFF) is developed. BFF tool was built using the WPF development platform and C# programming language in Visual Studio 2019. A screenshot of the graphical user interface (GUI) is displayed in Fig. 11. Proposed GUI design fulfils the most important design principles (Lauesen 2005):

- Clarity: all the information can be accurately conveyed to the user and prevents the user from making mistakes;
- User centric: potential end users were involved in the design process), users' perspective of understanding is taken into account;
- KISS (Keep It Simple and Stupid): Users can use this tool without putting much effort into interactions;

- Readability of the content: Font selection as a crucial factor of content readability is selected to comply with preferences of potential user;
- Efficiency: Users can reach their goals quickly and easily;
- Well organized layout: Minimal effort that users have to spend in interaction with user interface.

User interface of BFF tool is composed of eight sections. The first section is for data manipulation, the user can select the data file, data range, divide time series on train and test (holdout) and specify the forecast horizon. Second section is related to model selection. User can choose between all methods used in this paper as well as to use Best Fit tool which provides the predictions by the method that generates the lowest errors. Third section is related to predictions, it contains predicted values as an output of a specific model, or Best Fit option. Forth section "Processing log" shows the info related to modelling process. "Plot original data" section contains the time series which is subject of analysis. Section "Plot predictions" plots the predicted and observed values of test sample in time series. Performance section displays MAPE, MAE and RMSE values. "Error log" provides details about potential error.

Figure 12. shows an example of using BFF. The user selected a time series (Presevo Import) with 6 observations as a testing sample and LSTM as a modelling approach. Original time series is plotted in section "Plot original data", predictions are displayed in "Predictions" window and plotted on "Plot predictions" window. Performance section shows the values of relevant performance measures. Error log summarizes errors which in this case are not related to feasibility of tool execution.

6. Concluding Remarks

A detailed analysis of operational conditions and procedures for rail freight at the border crossings of AWB RFC showed that large improvements can be made by better planning of operations at such locations. Time series forecasting represents a powerful tool to improve planning processes on border crossings. In this paper we analyzed import and export freight traffic flows on four main border crossings on Serbian railway network. The four prediction models (ARIMA, LSTM, ARIMA-LSTM, SSA) are applied for modelling of considered eight time series. Three evaluation indices were selected to evaluate the prediction accuracy of considered prediction models in the training and testing sample. On average, the highest prediction accuracy have LSTM models (RMSE = 13.140, MAE = 12.866, MAPE = 7.864) then ARIMA (RMSE = 17.348, MAE = 15.308, MAPE = 11.760) and ARIMA LSTM models (RMSE = 18.298, MAE = 15.339, MAPE = 10.934). SSA models demonstrated the lowest prediction accuracy (RMSE = 22.303, MAE = 18.887, MAPE = 10.328). To provide practical usability of proposed approaches, a prototype desktop client application is developed. Application enables modelling of any univariate time series by the models included in this paper. Application also provides "Best Fit" feature, or prediction based on the model that has the best performances. Future research will be devoted to the development of models that will produce higher prediction accuracy, as well as to extending the features of Best Fit forecasting tool.

Declarations

References

1. Abebe, M., Noh, Y., Kang, Y.J., Seo, C., Kim, D., Seo, J.: Ship trajectory planning for collision avoidance using hybrid ARIMA-LSTM models. Ocean Engineering 256, 1-15 (2022)

2. Alhindawi, R., Nahleh, Y.A., Kumar, A., Shiwakoti, N.: Projection of Greenhouse Gas Emissions for the Road Transport Sector Based on Multivariate Regression and the Double Exponential Smoothing Model, *Sustainability* 12, 21, 1-18 (2020)

3. Bao, Y., Xiong, T., Hu, Z.: Forecasting air passenger traffic by support vector machines with ensemble empirical mode decomposition and slope-based method. Discrete Dynamics in Nature and Society, 1-13 (2012)

4. Bin, Y., Zhongzhen, Y., Baozhen, Y.: Bus arrival time prediction using support vector machines. Journal of Intelligent Transportation Systems 10, 4, 151-158 (2006)

5. Blinova, T.O., 2007. Analysis of possibility of using neural network to forecast passenger traffic flows in Russia. Aviation 11, 1, 28-34 (2007)

6. Box, G.E.P., Jenkins, G.M.: Time series analysis: forecasting and control. Holden Day San Francisco (1976)

7. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time series analysis forecasting and control. Wiley, New Jersey (2008)

8. Chen, C.-F., Chang, Y.-H., Chang, Y.-W.: Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan. Transportmetrica 5, 2, 125-140 (2009)

9. Cheu, R.L., Xu, J., Kek, A.G., Lim, W.P., Chen, W.L.: Forecasting shared-use vehicle trips with neural networks and support vector machines. Transportation research record 1968, 1, 40-46 (2006)

10. Cong, Y., Wang, J., Li, X.: Traffic Flow Forecasting by a Least Squares Support Vector Machine with a Fruit Fly Optimization Algorithm. Procedia Engineering 137, 59-68 (2016)

11. Dantas, T.M., Oliveira, F.L.C., Repolho, H.M.V.: Air transportation demand forecast through Bagging Holt Winters methods. Journal of Air Transport Management 59, 116-123 (2017)

12. Deng, Y., Fan, H., Wu, S.: A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits. Journal of Ambient Intelligence and Humanized Computing, 1-11 (2020)

13. Dia, H.: An object-oriented neural network approach to short-term traffic forecasting. European Journal of Operational Research 131, 2, 253-261 (2001)

14. Ding, Q., Wang, X., Zhang, X., Sun, Z.: Forecasting Traffic Volume with Space-Time ARIMA Model. Advanced Materials Research, 156-157, 979-983 (2011)

15. Fan, D., Sun, H., Yao, J., Zhang, K., Yan, X., Sun, Z.: Well production forecasting based on ARIMA-LSTM model considering manual operations. Energy 220, 1-13 (2021)

16. Gallo, M., De Luca, G., D'Acierno, L., Botte, M.: Artificial neural networks for forecasting passenger flows on metro lines. Sensors 19, 15, 1-14 (2019)

17. Ge, M., Junfeng, Z., Wu, J., Han, H., Shan, X., Wang, H.: ARIMA-FSVR Hybrid Method for High-Speed Railway Passenger Traffic Forecasting. Mathematical Problems in Engineering, 1-5 (2021)

18. Ge, S.Y., Zheng, C.J., Hou, M.M.: Forecast of Bus Passenger Traffic Based on Exponential Smoothing and Trend Moving Average Method. Applied Mechanics and Materials, 1374-1378 (2013)

19. Glišović, N., Milenković, M., Bojović, N., Svadlenka, L., Avramovic, Z.: A hybrid model for forecasting the volume of passenger flows on Serbian railways. Operations Research: An International Journal 16,271-285 (2016)

20. Golyandina, N., Nekrtutkim, V., Zhigljavsky, A.A.: Analysis of Time Series Structure. Routledge (2001)

21. Grubb, H., Mason, A.: Long lead-time forecasting of UK air passengers by Holt-Winters methods with damped trend. International Journal of Forecasting 17, 1, 71-82 (2001)

22. Guan, F., Shen, X., Wu, L., Yu, Y., Sun, D., Yang, Y.: Fleet route selection prediction problem based on support vector machine. Advances in Mechanical Engineering 11, 4, 1-11 (2019)

23. Harris, T.J., Yuan, H.: Filtering and frequency interpretations of Singular Spectrum Analysis. Physica D 239, 1958-1967 (2010)

24. Hassani, H., Mahmoudvand, R.: Multivariate Singular Spectrum Analysis: A general view and new vector forecasting approach. International Journal of Energy and Statistics 1, 55-83 (2013)

25. Hassani, H., Zhigljavsky, A.: Singular Spectrum Analysis: Methodology and Application to Economics Data. Journal of Systems Science and Complexity 22, 372-394 (2009)

26. Hassani, H.: Singular Spectrum Analysis: Methodology and Comparison. Journal of Data Science 5, 239-257 (2007)

27. Hewamalage, H., Bergmeir, C., Bandara, K.: Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting 37, 1, 388-427 (2021)

28. Hrnjica, B., Mehr, A.D.: Energy Demand Forecasting Using Deep learning In Smart Cities Performability Cognition Security, edited by Al-Turjman F. EAI/Springer Innovations in Communication and Computing, 71-104 (2020)

29. Huang, L., Xie, G., Zhao, W., Gu, Y., Huang, Y.: Regional logistics demand forecasting: a BP neural network approach. Complex Intelligent Systems, 1-16 (2021)

30. Hyndman, R.J., Athanasopoulos, G.: *Forecasting: Principles and Practice*. OTexts (2018)

31. Jiang, W., Luo, J.: Graph neural network for traffic forecasting: A survey. Expert Systems with Applications 207, 1-28 (2022)

32. Jiang, X., Adeli, H.: Dynamic wavelet neural network model for traffic flow forecasting. Journal of transportation engineering 131, 10, 771-779 (2005)

33. Jiang, X., Zhang, L., Chen, X.: Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. Transportation Research Part C: Emerging Technologies 44, 110-127 (2014)

34. Jighjigh, T.A., Oladejo, A., Michael, O., Ojowu, O.J.: The holt-winters multiplicative model of passengers' traffic forecast of the Nigeria airports. International Journal of Engineering in Computer Science 3, 1, 35-40 (2021)

35. Kashyap, A.A., Raviraj, S., Devarakonda, A., Nayak, S.R.K., Santhosh, K.V., Bhat, S.J.: Traffic flow prediction models - A review of deep learning techniques. Cogent Engineering, 1-24 (2022)

36. Khozani, Z.S., Banadkooki, F.B., Ehteram, M., Ahmed, A.N., El-Shafie, A.: Combining autoregressive integrated moving average with Long Short-Term Memory neural network and optimisation algorithms for predicting ground water level. Journal of Cleaner Production 348, 1-21 (2022)

37. Kim, J.K., Pak, J.Y., Wang, Y., Park, S.I., Yeo, G.T.: A Study on forecasting container volume of port using SD and ARIMA. Journal of Navigation and Port Research International Edition 35, 4, 343-349 (2011)

38. Kolidakis, S., Botzoris, G., Profillidis, V., Kokkalis, A.: Real-time Intraday Traffic Volume Forecasting - A Hybrid Application Using Singular Spectrum Analysis and Artificial Neural Networks. Periodica Polytechnica Transportation Engineering 48, 3, 226-235 (2020)

39. Kumar, S.V., Vanajakshi, L.: Short-term traffic flow prediction using seasonal ARIMA model with limited input data. European Transport Research Review 7, 21, 1-9 (2015)

40. Lauesen, S., 2005. User interface Design. Pearson (2005)

41. Li, L., Wang, Y., Zhong, G., Zhang, Y., Ran, B.: Short-to-medium Term Passenger Flow Forecasting for Metro Stations using a Hybrid Model. KSCE Journal of Civil Engineering 22, 1937-1945 (2019)

42. Li, X.: Comparison and Analysis Between Holt Exponential Smoothing and Brown Exponential Smoothing Used for Freight Turnover Forecasts. *Third International Conference on Intelligent System Design and Engineering Applications,* 453-456 (2013)

43. Liu, J., Wu, N., Qiao, Y., Li, Z.: A scientometric review of research on traffic forecasting in transportation. IET Intelligent Transport Systems 15, 1, 1-16 (2021)

44. Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y.: Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transportation Research Part C: Emerging Technologies 54, 187-197 (2015)

45. Manowska, A., Rybak, A., Dylong, A., Pielot, J.: Forecasting of Natural Gas Consumption in Poland Based on ARIMA-LSTM Hybrid Model. *Energies 14,* 1-16 (2021)

46. Medeiros, M.C., Veiga, A.: A hybrid linear-neural model for time series forecasting. IEEE Transactions on Neural Networks 11, 6, 1402-1412 (2000)

47. Medina-Salgado B., De la Cruz, E.S., Pozos-Parra, P., Sierra, J.E.: Urban traffic flow prediction techniques: A review. Sustainable Computing: Informatics and Systems 35, 1-16 (2022)

48. Milenkovic, M., Milosavljevic, N., Bojovic, N., Val, S.: Container flow forecasting through neural networks based on metaheuristics. Operations Research: An International Journal 21, 965-997 (2021)

49. Milenkovic, M., Svadlenka, L., Melichar, V., Bojovic, N., Avramovic, Z.: SARIMA modelling approach for railway passenger flow forecasting. Transport 33, 5, 1113-1120 (2016)

50. Miller, J.W.: ARIMA Time Series Models for Full Truckload Transportation Prices. *Forecasting 1,* 121-134 (2019)

51. Moiseev, G.: Forecasting oil tanker shipping market in crisis periods: Exponential smoothing model application. The Asian Journal of Shipping and Logistics 37, 3, 239-244 (2021)

52. Mostafa, M.M.: Forecasting the Suez Canal traffic: a neural network analysis. Maritime Policy Management 31, 2, 139-156 (2004)

53. Murat, Y.S., Ceylan, H.: Use of artificial neural networks for transport energy demand modeling. Energy Policy 34, 17, 3165-3172 (2006)

54. Peng, H., Wang, H., Du, B., Bhuiyan, M.Z.A., Ma, H., Liu, J., Wang, L., Yang, Z., Du, L., Wang, S., Yu, P.S.: Spatial temporal incidence dynamic graph neural networks for traffic flow forecasting. Information Sciences 521, 277-290 (2020)

55. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z.: Forecasting: theory and practice. International Journal of Forecasting 38, 3, 705-871 (2022)

56. Phan, T.T.H., Nguyen, X.H.: Combining statistical machine learning models with ARIMA for water level forecasting: the case of the Red river Advances in Water Resources 142, 1-15 (2020)

57. Ruiz-Aguilar, J.J., Turias, I.J., Jiménez-Come, M.J.: Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting. Transportation Research Part E: Logistics and Transportation Review 67, 1-13 (2014)

58. Sepp, H., Schmidhuber, J.: Long short-term memory. Neural computation 9, 8, 1735-1780 (1997)

59. Shahriari, S., Ghasri, M., Sisson, S.A., Rashidi, T.: Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction. Transportmetrica A: Transport Science 16, 3, 1552-1573 (2020)

60. Shang, Q., Lin, C., Yang, Z., Bing, Q., Zhou, X.: A Hybrid Short-Term Traffic Flow Prediction Model Based on Singular Spectrum Analysis and Kernel Extreme Learning Machine. PLoS ONE 11, 8, 1-25 (2016)

61. Shuai, C., Pan, Z., Gao, L., Zuo, H.W.: Short-Term Traffic Flow Prediction of Expressway: A Hybrid Method Based on Singular Spectrum Analysis Decomposition. Advances in Civil Engineering, 1-10 (2021)

62. Sitzimis, I.: An Optimal Forecasting Method of Passenger Traffic in Greek Coastal Shipping. International Journal of Business and Economic Sciences Applied Research, 1-16 (2022)

63. Smith, B., Demetsky, M.: Traffic flow forecasting: comparison of modelling approaches. Journal of Transportation Engineering 123, 4, 261-266 (1997)

64. Smith, T.G.: pmdarima: ARIMA estimators for Python Downloaded from: http://wwwalkaline-mlcom/pmdarima (2017)

65. Somu, N., Raman, G., Ramamritham, K.: A hybrid model for building energy consumption forecasting using long short term memory networks. Applied Energy 261, 1-20 (2020)

66. Stephanedes, Y.J., Michalopoulos, P.G., Plum, R.A.: Improved estimation of traffic flow for real-time control (discussion and closure). Transportation Research Record 795, 28–39 (1981)

67. Syriopoulos, T., Tsatsaronis, M., Karamanos, I.: Support vector machine algorithms: An application to ship price forecasting Computational Economics 57, 1, 55-87 (2021)

68. Tang, X., Deng, G., 2016. Prediction of Civil Aviation Passenger Transportation Based on ARIMA Model. Open Journal of Statistics 6, 824-834 (2016)

69. Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G.: Short-term traffic forecasting: Overview of objectives and methods. Transport Reviews: A Transnational Transdisciplinary Journal 24, 5, 533-557 (2004)

70. Vlahogianni, E.I., Golias, J.C., Karlaftis, M.G.: Short-term traffic forecasting: Overview of objectives and methods. Transport Reviews 24, 5, 533-557 (2004)

71. Wang, J., Shi, Q., 2013. Short-term traffic speed forecasting hybrid model based on Chaos-Wavelet Analysis-Support Vector Machine theory. Transportation Research Part C: Emerging Technologies 27, 219-232 (2013)

72. Wang, X., Zhang, N., Zhang, Y., Shi, Z.: Forecasting of short-term metro ridership with support vector machine online model. Journal of Advanced Transportation, 1-14 (2018)

73. Wei, Y., Chen, M.C., 2012. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C: Emerging Technologies 21, 1, 148-162 (2012)

74. Williams, B.M., Hoel, L.A., 2003. Modelling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. Journal of Transportation Engineering 129, 6, 664-672 (2003)

75. Williams, B.M., Durvasula, P.K., Brown, D.E.: Urban Freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. Transportation Research Record 1644, 132–141 (1998)

76. Xie, G., Wang, S., Lai, K.K.: Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches. Journal of Air Transport Management 37, 20-26 (2014)

77. Xu, D., Zhang, Q., Ding, Y., Zhang, D.: Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting. Environmental Science and Pollution research 29, 4128-4144 (2022)

78. Xu, S., Chan, H.K., Zhang, T.: Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. Transportation Research Part E: Logistics and Transportation Review 122,

169-180 (2019)

79. Yang, M., Chen, C., Wang, L., Yan, X., Zhou, L.: Bus arrival time prediction using support vector machine with genetic algorithm. Neural Network World 26, 3, 205-217 (2016)

80. Yang, Z., Jin, L., Wang, M.: Forecasting Baltic Panamax Index with Support Vector Machine. Journal of Transportation Systems Engineering and Information Technology 11, 3, 50-57 (2011)

81. Zhang, Y., Liu, Y.: Traffic forecasting using least squares support vector machines. Transportmetrica 5, 3, 193-213 (2009)

82. Zhi-Peng, L., Hong, Y.U., Yun-Cai, L., Fu-Qiang, L.: An Improved Adaptive Exponential Smoothing Model for Short-term Travel Time Forecasting of Urban Arterial Street. Acta Automatica Sinica 34, 11, 1404-1409 (2008)

83. Zhou, W., Wang, W., Zhao, D.: Passenger Flow Forecasting in Metro Transfer Station Based on the Combination of Singular Spectrum Analysis and AdaBoost - Weighted Extreme Learning Machine. Sensors, 1-23 (2020)

Figures
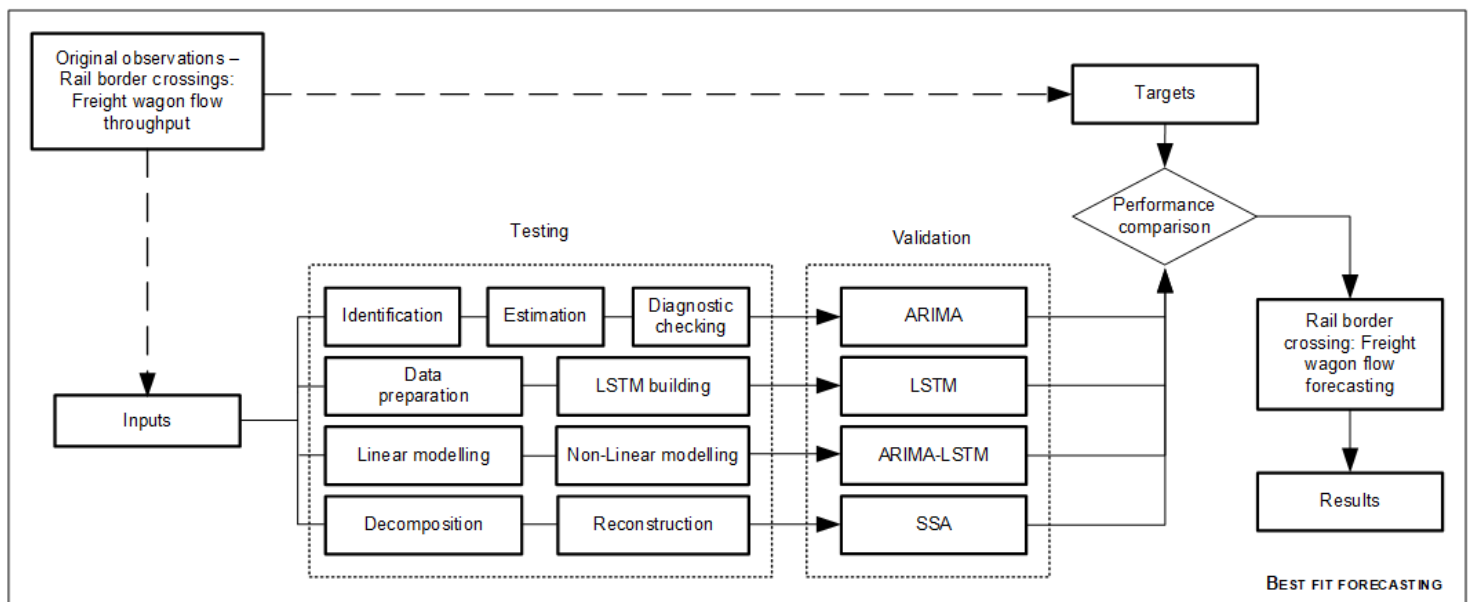
Figure 1

Comparative analysis of ARIMA, ARIMA-LSTM and SSA methods for modelling of freight wagon flows at border crossing points.
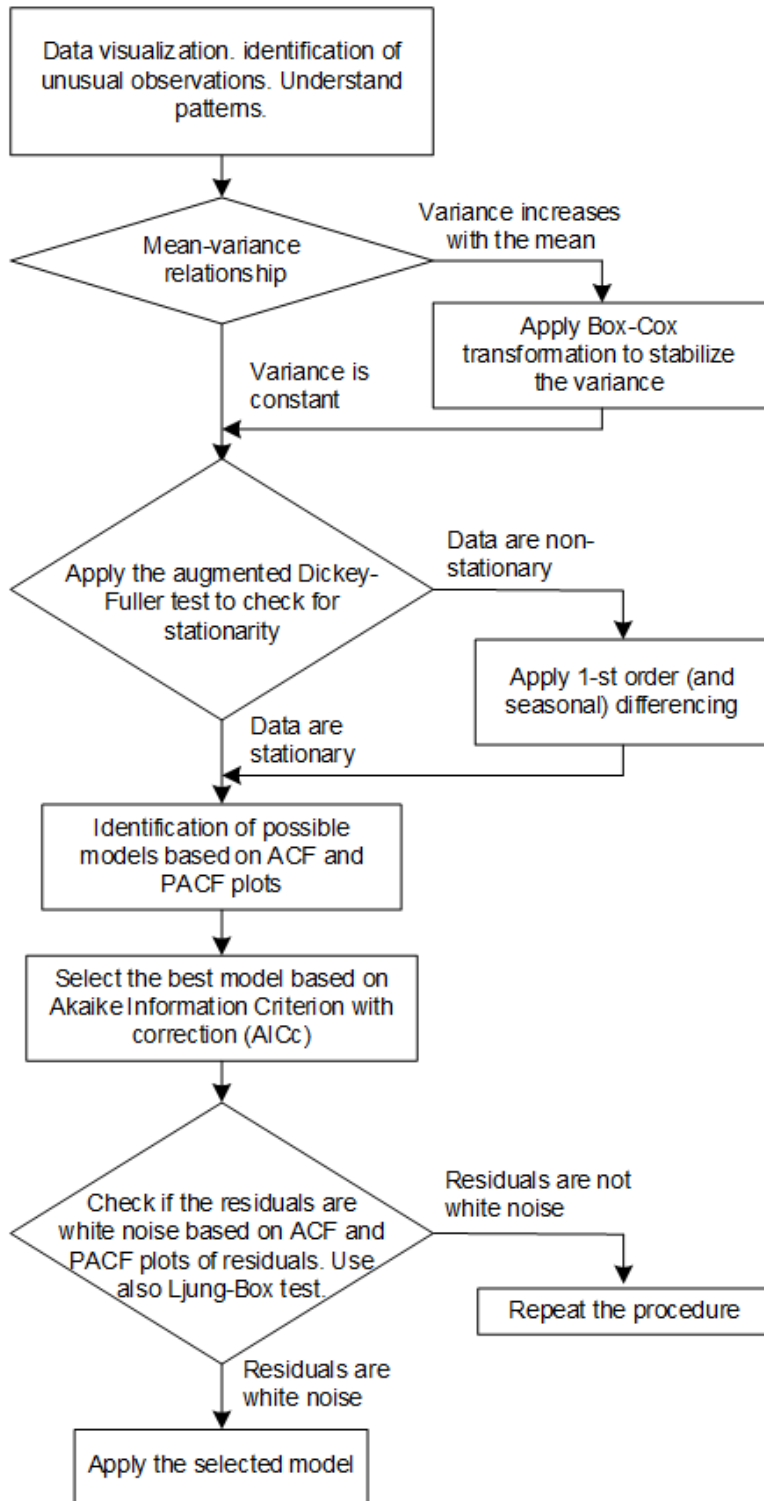
Figure 2

ARIMA (SARIMA) modelling flow chart
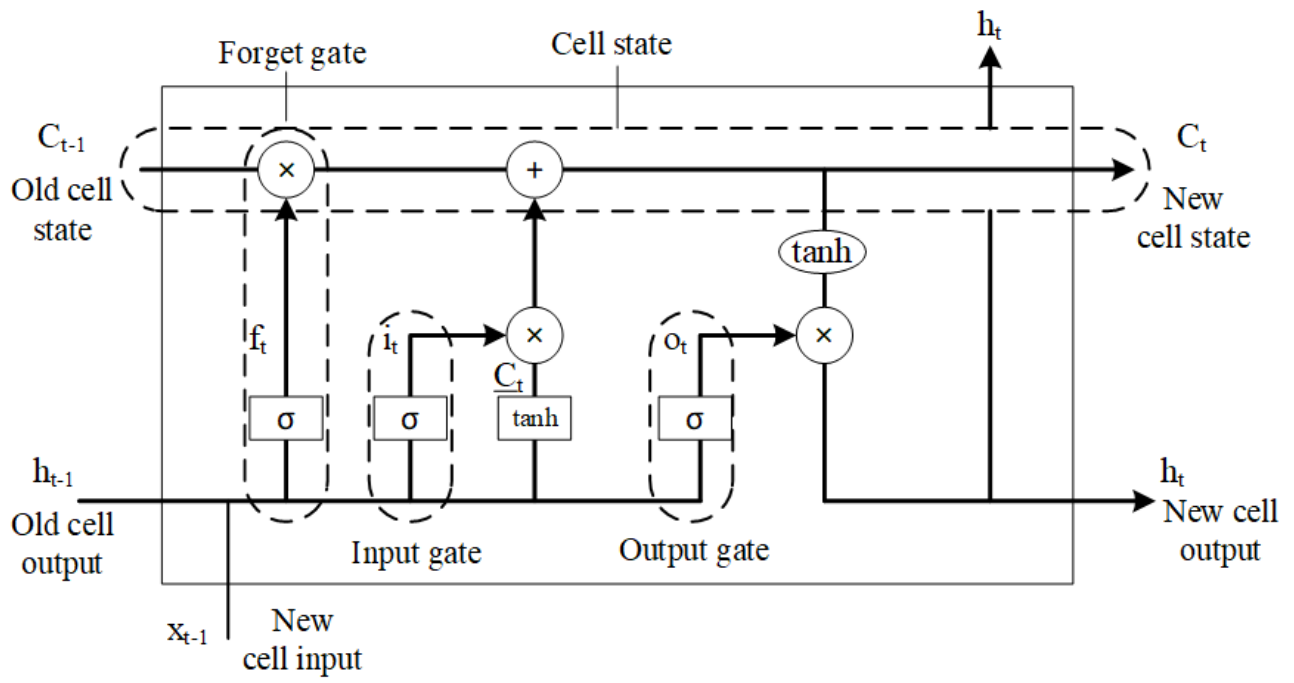
Figure 3

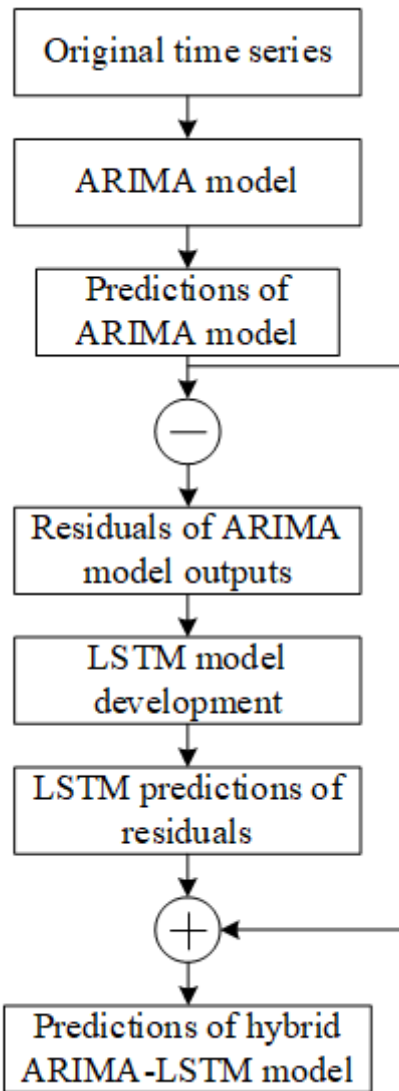Memory cell and gate units of LSTM memory block.

Figure 4

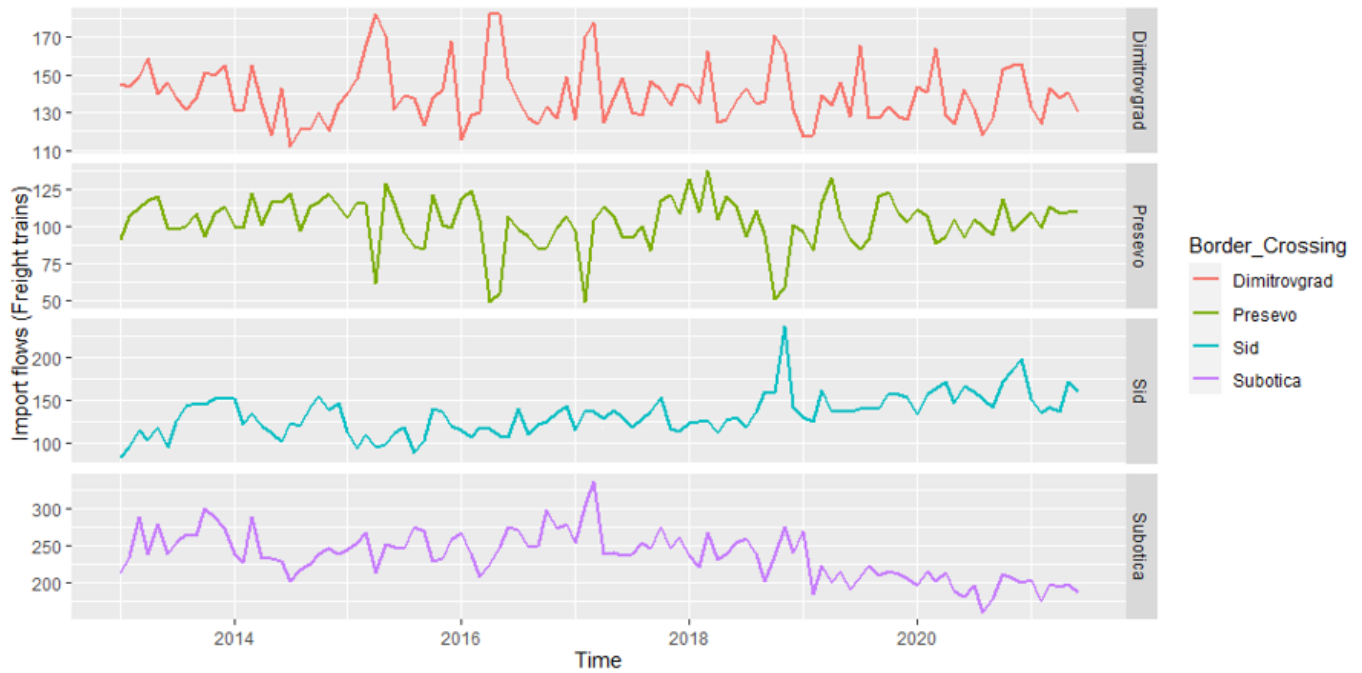Hybrid ARIMA-LSTM model flow chart

Figure 5

Border-crossing points: railway import flows expressed in the number of freight trains (January 2012 – June 2021).
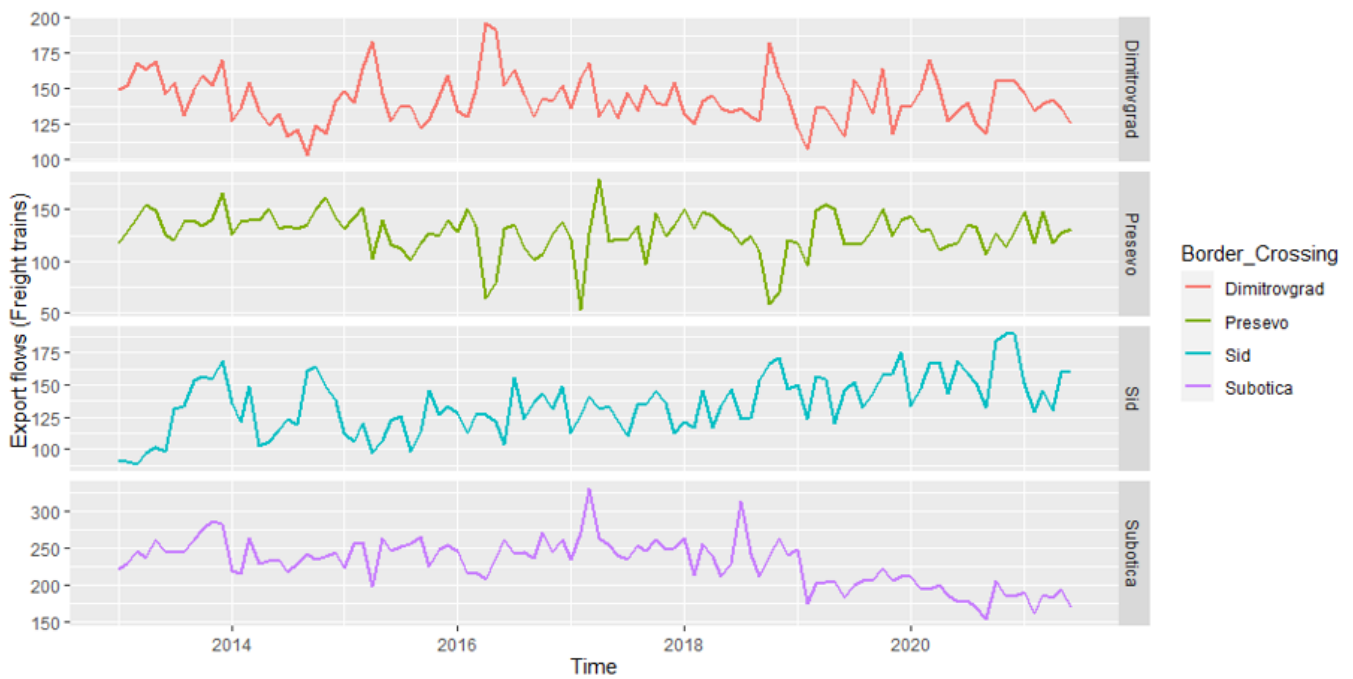

Figure 6

Border-crossing points: railway export flows expressed in the number of freight trains (January 2012 – June 2021).

Figure 7

Import freight train flows: The fitting performances of ARIMA, LSTM, ARIMA-LSTM and SSA.

Figure 8

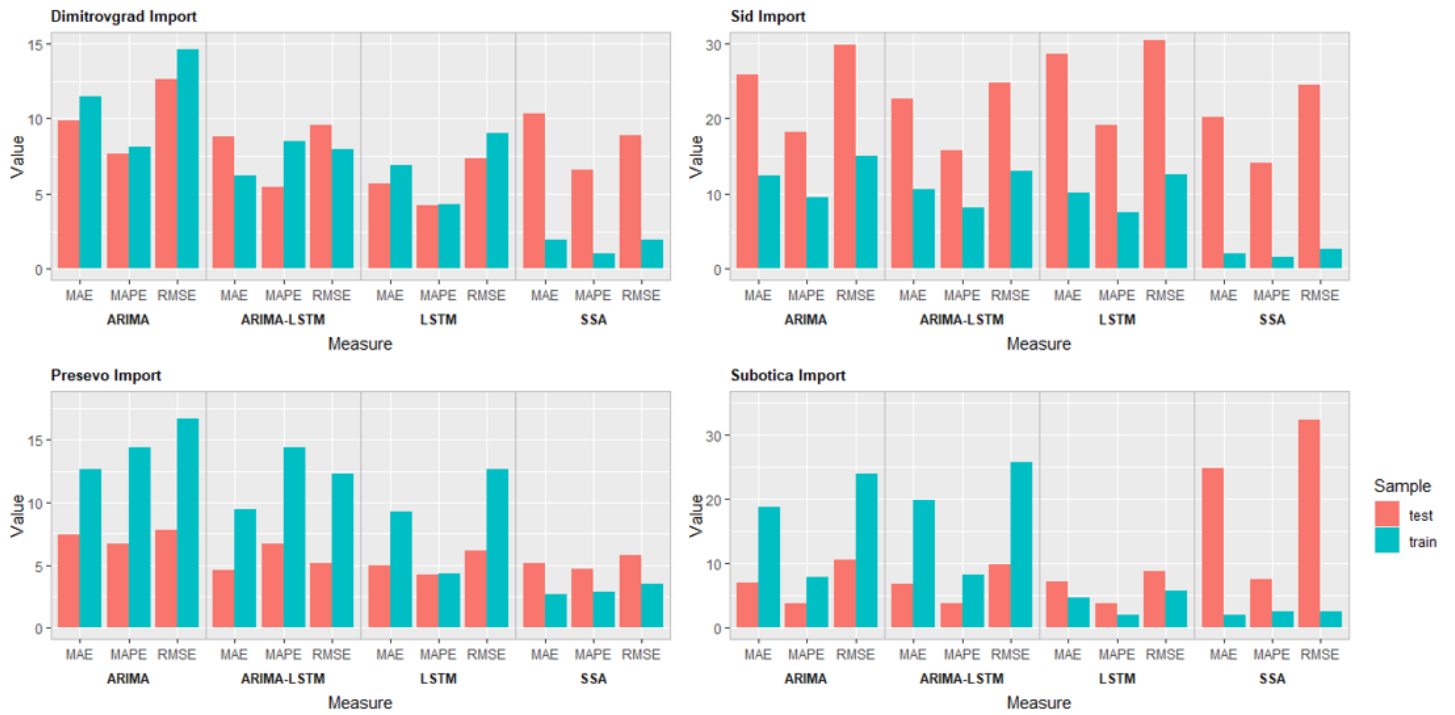Import freight train flows: The fitting performances of ARIMA, LSTM, ARIMA-LSTM and SSA

Figure 9

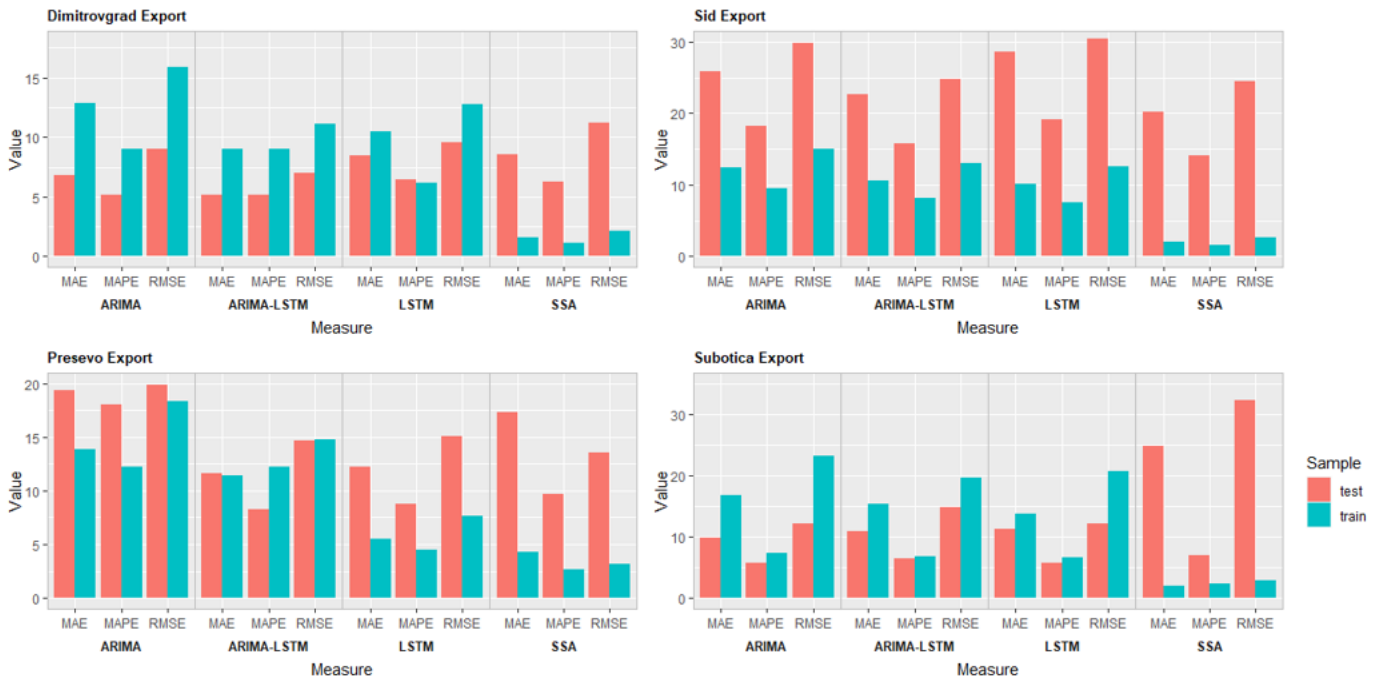Forecast Accuracy - Import freight train flows.

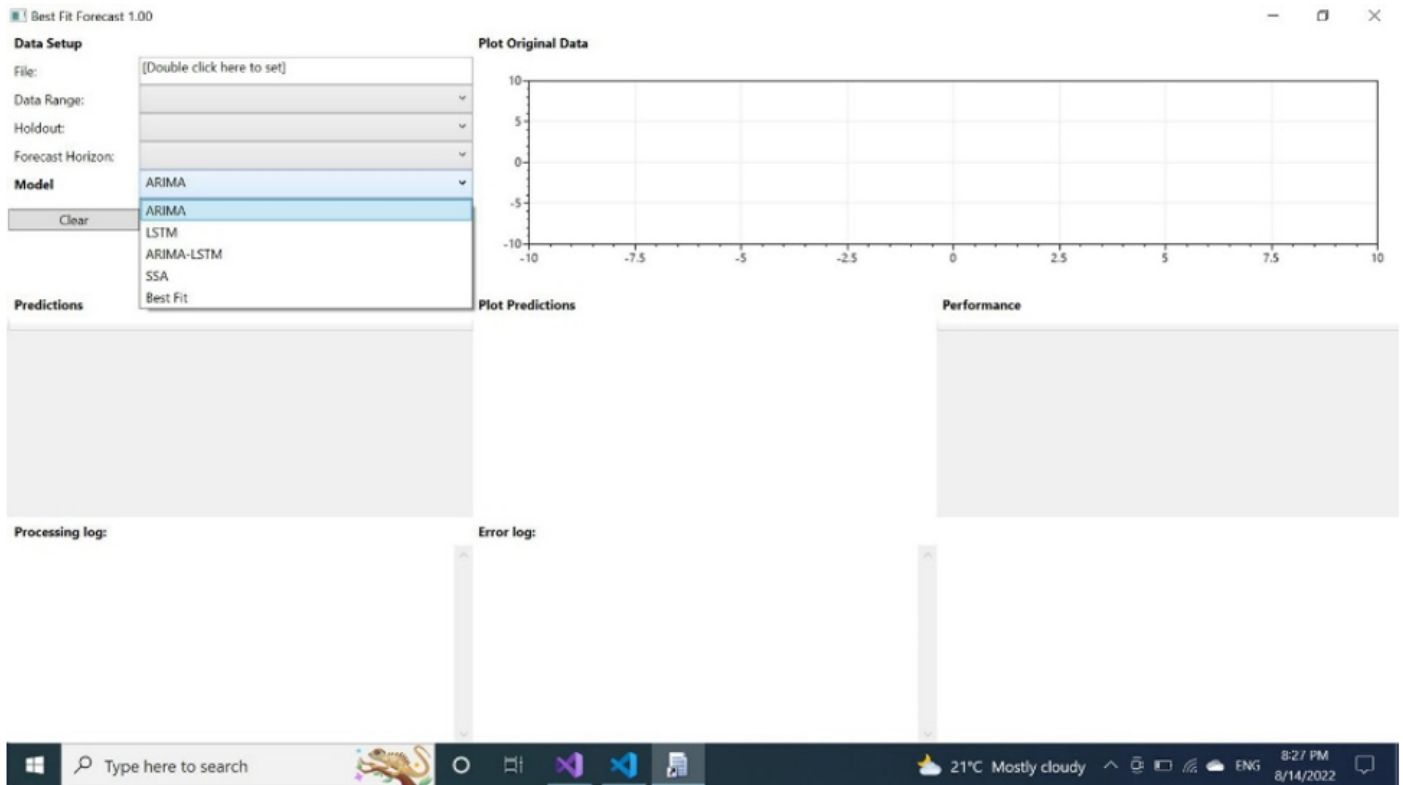## Figure 10

Forecast Accuracy – Export freight train flows

Figure 11

GUI of the BFF tool.

Figure 12

BFF tool - LSTM modelling.