MATHEMATICAL MODELLING AS AN ELEMENT OF PLANNING RAIL TRANSPORT STRATEGIES

Anna BORUCKA[1], Dariusz MAZURKIEWICZ[2*], Eliza ŁAGOWSKA[3]

[1,3]*Faculty of Security, Logistics and Management, Military University of Technology, Warsaw, Poland*
[2]*Faculty of Mechanical Engineering, Lublin University of Technology, Lublin, Poland*

Abstract. Effective planning and optimization of rail transport operations depends on effective and reliable forecasting of demand. The results of transport performance forecasts usually differ from measured values because the mathematical models used are inadequate. In response to this applicative need, we report the results of a study whose goal was to develop, on the basis of historical data, an effective mathematical model of rail passenger transport performance that would allow to make reliable forecasts of future demand for this service. Several models dedicated to this type of empirical data were proposed and selection criteria were established. The models used in the study are: the seasonal naive model, the Exponential Smoothing (ETS) model, the exponential smoothing state space model with Box–Cox transformation, ARMA errors, trigonometric trend and seasonal components (TBATS) model, and the AutoRegressive Integrated Moving Average (ARIMA) model. The proposed time series identification and forecasting methods are dedicated to the processing of time series data with trend and seasonality. Then, the best model was identified and its accuracy and effectiveness were assessed. It was noticed that investigated time series is characterized by strong seasonality and an upward trend. This information is important for planning a development strategy for rail passenger transport, because it shows that additional investments and engagement in the development of both transport infrastructure and superstructure are required to meet the existing demand. Finally, a forecast of transport performance in sequential periods of time was presented. Such forecast may significantly improve the system of scheduling train journeys and determining the level of demand for rolling stock depending on the season and the annual rise in passenger numbers, increasing the effectiveness of management of rail transport.

Keywords: time series, mathematical modelling, forecasting of transport performance, rail transport management, optimization.

Notations

Abbreviations:

AIC – Akaike information criterion;
AICC – corrected AIC;
ARIMA – autoregressive (AR) integrated moving average (MA);
ARMA – autoregressive (AR) moving average (MA);
BIC – Bayesian information criterion;
BP – backpropagation;
DEA – data envelopment analysis;
ETS – exponential smoothing;
GSO – glow-worm swarm optimization;
LSTM – long short-term memory network;
MAE – mean absolute error;
MAPE – mean absolute percentage error;
ME – mean error;

MPE – mean percentage error;
MSE – mean square error;
RMSE – root mean square error;
SARIMA – seasonal ARIMA;
TBATS – exponential smoothing state space model with Box–Cox transformation, ARMA errors, trigonometric trend and seasonal components.

Variables and functions:

b – trend coefficient/value;
d – order of integration;
h – periods;
k – number of parameters;
l – level term;
l_t – local level of the phenomenon under study at time t;
L – likelihood function;

*Corresponding author. E-mail: *d.mazurkiewicz@pollub.pl*

m – seasonal period;

N – number of seasons in a year;

p – number of autoregressive (AR) parameters;

q – number of moving average (MA) parameters;

$s_{j,t}^{(i)}$ – value of the seasonal component at time t;

$s_t^{(i)}$ – sum of the value of the seasonal compoent at time t;

y_t – the time series value at time t.

T – time;

α – smoothing parameter;

β – trend;

γ – sensonality;

φ – trend damping weight;

ε_t – the random variable white noise at time t;

ϕ – damping parameter;

λ – Box–Cox transformation.

Introduction

Rail transport capability is a critical indicator of the competitiveness of a country's economy and the possibilities of its development, which is why it is important to carry out analyses to assess the functioning and development of this mode of transport and to indicate possible and directions of change and expansion (Konowrocki, Chojnacki 2020; Kang *et al.* 2019; Markovits-Somogyi 2011; Baležentis, A., Baležentis, T. 2011). Information about a potential increase in transport performance (Schütze *et al.* 2020), which is the basic element of the analysis presented in this paper, is key to the process of planning, modernization and development of railways. The research hypothesis put forward in this paper assumes that it is possible to properly identify the parameters of a time series model describing rail transport performance and to reliably forecast the value of this performance in subsequent periods of time, in order to effectively support the planning of a development strategy for this branch of transport. Identification of the parameters of the model, viewed as an element of an expert system, and forecasting of transport performance permits to estimate future demand and adjust the number and frequency of journeys to the actual transport needs.

The negative impact of transport on the natural environment also needs to be considered. Effective determination of future transport performance allows to increase energy efficiency and reduce harmful emissions in the entire transport system by effectively planning the use of accompanying infrastructure or other, related types of transport. This issue was analysed, among others, by Jarašūnienė *et al.* (2019), who demonstrated that efficient interoperability of railway and maritime transport depended, among others, on coordination of large numbers of participants, strict schedule observance, and processing of large amounts of information. They found that the most important of these factors was the development of an effective information system by integrating individual elements and data regarding rail and sea transport. This shows that expert systems, including integrated systems, need to be supported in effective planning of transport performance.

The mathematical model developed in this paper is an important component of a rail transport decision support system. Additionally, it can be used as a tool to minimize the waste of resources (improve fleet management) and eliminate low efficiency caused by inaccurate forecasting of the number of passengers in railway traffic (Jarašūnienė *et al.* 2020; Zhou *et al.* 2020), i.e. to avoid situations in which a large number of carriages are used to transport a small number of passengers. This is particularly important when passenger numbers are growing and rolling stock resources are depleting, with trains constantly ageing and being decommissioned. By adjusting the number of carriages and passenger facilities to the demand, railway companies can avoid empty runs and unnecessary work, minimize energy consumption, reduce noise, extend the service life of machinery and equipment, and increase safety (Bureika *et al.* 2017).

Moreover, forecast results obtained using methods of efficient estimation of transport performance can also contribute to the reduction of investment risk, supporting decisions on the expansion of the current network with additional railroads, platforms or waiting rooms, which is particularly important given that the expansion of such systems is costly and time-consuming. Transport demand forecasts provide support not only with regard to the development of railway alone but also cooperative systems, being an important element of planning budgets and development strategies for the infrastructure of other modes of transport in compliance with the concept of intermodal transport (Bureika *et al.* 2017; Jarašūnienė *et al.* 2019, 2020). Reliable positive forecasts serve as an incentive for the expansion of interchange hubs, as well as other elements of transport infrastructure: both additional lines (connections/roads) and structures (stops/stations, parking lots). Accurate forecasting of demand also ensures accurate scheduling of journeys, regulating their throughput capacity, which also affects the comfort of travelling, the stability of running times and the overall level of satisfaction of rail customers. These issues were considered in a study by (Jarašūnienė *et al.* 2019, 2020). The authors of that publication analysed the importance of the system of management of internationalization processes in developing railway transport. Their paper highlights the synergistic effect of a multilevel management model in internationalization processes, which requires adequate data modelling methods. Based on the example of Lithuanian railways in the Baltic region, they developed a new railway transport model using the "alliance + cluster" system, which is based on the management of internationalization processes. One of the conclusions of that paper is that an adequately developed model could be used as a basis for further economic research and practical activities, and from the mathematical modelling perspective, as part of professional and effective expert systems.

1. Literature review

To date, studies on the construction of models describing the demand for rail services have been conducted at several research centres. For example, in a paper by Milenković *et al.* (2018), the authors use the SARIMA model to forecast monthly passenger flows on Serbian railways. Roos *et al.* (2017) discuss an approach to forecasting short-term passenger flows in the Parisian urban rail network based on dynamic Bayesian networks. Zhang *et al.* (2019) use a LSTM network to analyse the transport performance of the urban rail transit network in Beijing. Tang *et al.* (2017) present an alternative method of analysis, which combines a BP neural network and the GSO algorithm. Methods of forecasting passenger traffic in Moscow based on network topology analysis are described in research by Namiot *et al.* (2018). Most of these papers and the analyses they present are limited to the study of cities, which develop in the most dynamic way. The increasing demand for rail transport in cities is a consequence of both urban sprawl and the deteriorating road transport conditions (congestion, traffic jams, increased vehicle exhaust emissions leading to smog, etc.) – Dolinayova *et al.* (2018). There are much fewer models that allow to assess the functioning of larger scale, national rail networks. Such models have been developed in Sweden (Andersson *et al.* 2017) and India (Prakaulya *et al.* 2017). Markovits-Somogyi (2011) reviewed the results of research on the application of DEA in the transport sector with special emphasis on the inputs and outputs selected in the DEA models employed in different fields. The author has compiled data from 69 DEA applications reported in the literature and investigated their characteristics and fields of application along with the inputs and outputs used. DEA is a tool for evaluating the performance of different companies' and organizations' decision making units, which convert multiple inputs into multiple outputs. Unfortunately, DEA use results show that this method has a serious drawback of being sensitive to measurement errors and noise in data (Baležentis, A., Baležentis, T. 2011; Markovits-Somogyi 2011). Most research in this area concerns the application and use of one specific mathematical model, without considering alternative solutions (Gao *et al.* 2019; Liu, Wang 2007). This usually results in a low effectiveness of the proposed methods of analysis (Markovits-Somogyi 2011) or a high cost of obtaining minimally satisfactory results. Similar conclusions can be drawn from a study by Jonaitis (2007), which shows that methods relying on the evaluation of qualitative characteristics of rolling stock are relatively expensive. Only a few studies compare several modelling methods to choose the best one among them. For example, Banerjee *et al.* (2020) propose various models that can be used to forecast demand in the regular passenger transport industry.

Demand analyses and forecasts are important sources of information for developing transport policies. Unfortunately, demand data are not always available for making key decisions, as there are no appropriate mathematical models for generating demand forecasts (Jarašūnienė *et al.* 2020; Milenković *et al.* 2018). In view of this, it is necessary to analyse the railway systems of various countries (regions) in order to select appropriate methods of forecasting transport performance. Such analyses allow to create a scientific database that can be used as reference for research conducted in other countries or with regard to other transport systems.

Against this background, the goal of this present study is to identify, on the basis of historical data, the parameters of a mathematical model of rail passenger transport performance that would allow to make reliable forecasts of future demand for this service. Keeping in mind the fact that the literature offers very few comprehensive analyses of large-scale transport systems, in this paper, we investigated a national (Polish) railway system using transport performance data. Several models dedicated to this type of empirical data were proposed and selection criteria were established. Then, the best model was identified and its accuracy and effectiveness were assessed. Finally, a forecast of transport performance in sequential periods of time was presented. The paper consists of an introduction, a section describing the methodology of the study followed by sections characterizing the mathematical models constructed using the adopted methods. Empirical data on the passenger transport performance of the Polish railways was used. The final sections provide a summary of the results, conclusions, and directions of future research.

2. Polish railway rolling stock

The mathematical models presented in the paper can also be implemented in other countries, and the obtained results can be compared and conclusions drawn on this basis. Therefore, the introduction to the methods and models presented in the paper is a detailed analysis of the research subject. The analysis relates to the potential and capability of rolling stock in Poland. To ensure the transparency of analysis and facilitate the implementation of the proposed methods in other countries, we describe the current state of rail transport in Poland. The description will allow to compare Polish rail transport data with those for other countries and benchmark the possibility of direct use of the solutions proposed. It also shows that an important element of such research is the availability of empirical data, the quality of which strongly determines the possibilities of the analyses performed.

According to Boston Consulting Group's – *The 2017 European Railway Performance Index* report (Duranton *et al.* 2017), Poland ranks 22 out of 25 countries (data for 2017), with a score of zero for safety due to the large number of accidents per train kilometre travelled. Polish railways do not come out well compared to other European railway networks. They are characterized by a low quality of service, unsatisfactory technical condition of the rolling stock and railway infrastructure as well as frequent delays resulting from repair works, accidents, high vehicle failure rates, as well as improper scheduling of journeys and

delays of other trains. Polish rolling stock is characterized by relatively outdated machinery, a high wear rate, low traction parameters, a high failure rate and unreliability. The passenger carriages and electric locomotives are on average around 30 years old, and diesel locomotives are around 40 years old. According to the data for the end of 2018 (Redakcja 2020), the Polish rolling stock consists of 314 electric locomotives, 109 diesel locomotives, 2047 passenger carriages with seating areas, 195 couchettes and sleepers, 1247 electric multiple units, and 269 diesel multiple units. In the same year, over 80% of diesel locomotives and 5% of electric locomotives were more than 40 years old (Redakcja 2020). The Polish railway rolling stock is ageing, and the vehicles that have been decommissioned are not replaced with new ones. For example, in 2018, operators used 314 electric vehicles whose average age was 33.8 years. When this figure is compared with data for 2015, one can see that 18 electric locomotives, i.e. 5.4% of all these vehicles, had been withdrawn from use. In the case of diesel locomotives, the number of available vehicles in 2018 was 109, and their average age was 41.7 years. Since 2015, as many as 43 locomotives had been decommissioned, which constituted 32.2% of these vehicles. Given that the average age of the Polish railway rolling stock exceeds 40 years, the vehicles have a high probability of failure and an increased susceptibility to damage and wear of components. Constantly exposed to difficult weather conditions, locomotives quickly lose their original efficiency. The age and failure rate of the vehicles, as well as the need to restore their rail-worthiness, generate higher and higher operating costs. In addition, diesel locomotives are a major contributor to air pollution, which is not in harmony with the paradigm of sustainable transport development.

Nevertheless, rail transport is a branch that should be developed dynamically, mainly because, right next to inland and maritime transport, it is one of the most environmentally friendly ways of transporting people and goods. A study of the causes and effects of environmental hazards in the EU (EEA 2021) shows that transport (mainly car and air transport) is one of the main sources of pollution, with almost 30% of total carbon dioxide emissions responsible for the greenhouse effect originating from transport sector, 72% of which come from road transport. In Poland, CO_2 emissions from transport account for approximately 24% of total CO_2 emissions (Rabiega, Sikora 2020). This is why it is important to undertake initiatives and research to promote the development of ecological forms of transport, including rail transport, and improve its availability and reliability (Konowrocki, Chojnacki 2020; Song, Schnieder 2018). Such projects should first be conducted on the national scale, and the experiences acquired should then be implemented more locally or in other regions. This is particularly important from the point of view of planning and forecasting, as the interest in rail transport in Poland is constantly growing. In 2019, as many as 335.9 million passengers travelled by rail. Comparing this figure with 2010, the number of passengers increased by 74.1 million

(28.3%). Owing to gradual modernization, Polish railways are gaining in popularity, and the growing numbers of motor vehicles and increasing road congestion encourage people to look for alternative modes of transport, including travelling by train.

The constantly growing interest in rail transport requires the introduction of effective development strategies for this mode of transport. To properly define those strategies, though, one needs to have access to comprehensive and transparent information on and reliable forecasts of the demand for rail transport services. The methods proposed in this study allow to make effective forecasts in this area.

3. Methods

The following models were used in the study: the seasonal naive model, the ETS model, the TBATS model, and the ARIMA model. The proposed time series identification and forecasting methods (Kozłowski *et al.* 2019a, 2020) are dedicated to the processing of time series data with trend and seasonality. It is these features that characterize the analysed transport performance observations. The theoretical foundations of these methods are discussed below.

3.1. Naive methods

The simplest method of estimating future values is naive forecasting, which assumes that if a forecast is made for period $T + 1$, the most adequate information will be that regarding the last observation, i.e. the change that occurred over time T. The advantage of this method is the simplicity of forecasting and the low requirements regarding availability of information. A disadvantage is that it does not allow to estimate the ex-ante error (Kozłowski *et al.* 2018). There are many types of naive forecasting, depending on the nature of the time series analysed. This paper uses a seasonal naive model. This type of model takes into account the number of seasons in a year ($N = 12$ in the case of monthly seasonality, $N = 4$ in the case of quarterly seasonality, etc.). A forecast is determined using the following formula (Kozłowski 2015):

$$y^*_{T+1} = y_{T-N+1}. \tag{1}$$

Each forecast is set to be the last observed value from the same season of the year.

As the simplest method of estimating future values, naive forecasting is worth using as a benchmark against other, more advanced methods of identifying and forecasting time series. Comparisons with the naive method allow to evaluate the effectiveness of more complex models.

3.2. ETS models

ETS models provide an alternative to naive forecasting. The basic idea behind ETS is to (exponentially) assign declining weights to historical observations when predicting a future observation. Among ETS methods, a special place belongs to the general class of ETS models (ETS

state space models). The individual letters of the acronym stand for error, trend and seasonality. These components can be combined additively, multiplicatively or in a mixed manner. The trend in ETS decomposes into two components: (1) a level term l and (2) a growth term b, which can be combined taking into account the damping parameter $\phi \in [0,1]$ to give five future trend types. Let T_h denote the forecast trend over the next h periods. Then we get the following five trend types:

»» N – none: $T_h = l$;
»» A – additive: $T_h = l + b \cdot h$;
»» Ad – additive damped: $T_h = l + \left(\phi + \phi^2 + \ldots + \phi^h \right) \cdot b$;
»» M – multiplicative: $T_h = l \cdot b^h$;
»» Md – multiplicative damped: $T_h = l \cdot b^{\left(\phi + \phi^2 + \ldots + \phi^h \right)}$.

If we select the seasonal component (no seasonality, the additive variant, and the multiplicative variant) and the trend component alone, we obtain 15 ETS models. When we additionally take into account random additive or multiplicative errors, we obtain 30 different models, as shown in Table 1 (Hyndman *et al.* 2008).

The type of an ETS model may be optimized by minimizing a selected information criterion (AIC, AICC, BIC) or forecast error (MSE, MAPE) (Kozłowski *et al.* 2019b). In this paper, AIC was used to compare the models. This criterion is commonly employed to compare statistical models as it allows to measure and test the relative predictive power of models (Kosicka *et al.* 2018).

Let k be the number of estimated parameters and L the maximum of the model's likelihood function. The smaller the value of the information criterion, the better the fit of the model.

The AIC value is given by:

$$AIC = -2 \cdot \ln L + 2 \cdot k. \tag{2}$$

3.3. TBATS model

The TBATS model has the following structural form (Kozłowski 2015):

$$TBATS\left(\lambda, \Phi, p, q, \{m_1, k_1\}, \{m_2, k_2\}, \ldots, \{m_T, k_T\} \right) \tag{3}$$

and is given by the following equations:

Table 1. Potential ETS models

ETS models		
ETS(M, M, N)	ETS(A, M, A)	ETS(M, N, M)
ETS(M, A, N)	ETS(A, Md, N)	ETS(M, N, A)
ETS(M, A, M)	ETS(A, Md, M)	ETS(M, N, N)
ETS(A, M, N)	ETS(A, N, A)	ETS(M, A, A)
ETS(A, N, N)	ETS(M, Ad, M)	ETS(A, Ad, M)
ETS(A, A, M)	ETS(M, Ad, N)	ETS(M, M, A)
ETS(M, M, M)	ETS(M, Md, M)	ETS(A, A, A)
ETS(A, N, M)	ETS(A, Ad, N)	ETS(A, Ad, A)
ETS(A, A, N)	ETS(M, Md, A)	ETS(M, Ad, A)
ETS(A, M, M)	ETS(M, Md, N)	ETS(A, Md, A)

$$y_t^{(\omega)} = l_{t-1} + \varphi \cdot b_{t-1} + \sum_{i=1}^{T} s_{t-1}^{(i)} + \alpha \cdot d_t; \tag{4}$$

$$b_t = b_{t-1} + \beta \cdot d_t; \tag{5}$$

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}; \tag{6}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cdot \cos \lambda_j^{(i)} + s_{j,t-1}^{*} \cdot \sin \lambda_j^{(i)} + \gamma_1^{(i)} \cdot d_t; \tag{7}$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \cdot \sin \lambda_j^{(i)} + s_{j,t-1}^{*} \cdot \cos \lambda_j^{(i)} + \gamma_2^{(i)} \cdot d_t; \tag{8}$$

$$\lambda_j^{(i)} = \frac{2 \cdot \pi \cdot j}{m_i}, \tag{9}$$

where: $i = 1, \ldots, T$.

This modelling method takes into account the Box–Cox transformation, seasonal and trend components, and the autocorrelation of model residuals through the ARMA process. The Box–Cox transformation is one of the transformations used in time series analysis to stabilize variance. It is also used to transform the continuous distribution of a random variable into a normal distribution (a normalizing transformation). The Box–Cox transformation is a family of transformations (Kozłowski 2015; Kozłowski *et al.* 2019a) given by:

$$f_\lambda(y) = \begin{cases} \dfrac{y^\lambda - 1}{\lambda}, & \lambda \neq 0; \\ \ln y, & \lambda = 0, \end{cases} \tag{10}$$

for $y > 0$.

A time series Y_1, Y_2, …, Y_n can be transformed using a Box–Cox transformation to series $\tilde{Y}_1 = f_\lambda(Y_1)$, $\tilde{Y}_2 = f_\lambda(Y_2)$, …, $\tilde{Y}_n = f_\lambda(Y_n)$. Parameter λ can take any real value. In practice, $\lambda = 0$ (logarithmic transformation) or $\lambda = 1/2$ (elemental transformation) are often used (Hyndman *et al.* 2008).

3.4. ARMA models

Stationary models are also popular time series models. Depending on the structure of the temporal correlation, two models are distinguished (Kozłowski 2015; Kozłowski *et al.* 2019b):

»» autoregressive model of order p $AR(p)$;
»» moving average model of order q $MA(q)$.

The autoregressive process uses process memory, with each specific observation is a combination of past observations. An important variable here is correct determination of order, which describes the number of past observations (lag terms) that affect the current observation. This process is denoted by $AR(p)$, where p is the order of autoregression. The autoregressive process is given by the following equation (Kang *et al.* 2019; Kozłowski 2015):

$$y_t = \varphi \cdot y_{t-1} + \varepsilon_t, \tag{11}$$

where: y_t – the value of the time series at time t; y_{t-1} – the value of the time series at time $t - 1$; ε_t – the random variable white noise at time t; φ – the strength of the relationship between the lag terms of the series and the current observation.

The moving average process assumes that the value of the series depends on the current and past white noise error terms. The value of parameter $MA(q)$ is defined for this process, which is the number of white noise error terms considered. This can be given by the following formula (Kang *et al.* 2019; Kozłowski 2015):

$$y_t = \beta \cdot \varepsilon_{t-1} + \varepsilon_t, \qquad (12)$$

where: y_t – the time series value at time t; ε_{t-1}, ε_t – a white noise error term at time t–1 and at time t; β – the effect of previous white noise error terms on the current value.

The moving average process can be defined as a weighted average of the last random disturbances.

ARMA class models are used with stationary time series. If the assumption of stationarity is not met, mathematical transformations must be used to stationarize the time series. One of the methods of transforming nonstationary time series into stationary ones is differencing of time series, given by (Kozłowski *et al.* 2020):

$$\Delta y_t = y_t - y_{t-1}. \qquad (13)$$

The values obtained by differencing are mostly stationary or at least represent an average constant level over a given period of time. If first differencing is insufficient, then second differencing must be performed on the differences series. The degree of integration of process "*I*" is described by parameter *d*, which is the order of integration of the time series that denotes the minimum number of differences required to obtain a stationary series. In this way, an ARIMA model and a SARIMA model, an extension of ARIMA that supports the direct modelling of seasonality, are obtained. These models use differencing at lag *d*.

4. Construction of mathematical models

To achieve reliable data for modelling rail passenger transport performance observations recorded in the years 2014–2019 (UTK 2020) were analysed. The data was archived on a monthly basis. The collected set of observations was divided into two subsets: a training set and a test set. The training set was used to build time series identification models and to make a forecast for the next period. The test set was used to validate the models built. The training set contained observations from January 2014 to December 2018, while the test set included data from January to December 2019. The sets are shown in Figure 1. A systematic increase in transport performance in the sequential years is noticeable, as well as a clear seasonality.

4.1. Naive model

In line with the adopted methodology, we first constructed a seasonal naive model. A graph comparing the seasonal naive forecast with empirical observations is shown in Figure 2.

As the graph (Figure 2) shows, the model does not satisfactorily fit the time series. All the forecast observa-

Table 2. Selected forecast errors of the naive model

	ME	RMSE	MAE	MPE	MAPE
Training set	103.46	122.01	107.35	6.36	6.647

tions are underestimated, which shows that the model has a poor predictive quality. In line with the adopted assumption, the forecast errors of the naive model given in Table 2 will be used as reference for the remaining models.

4.2. ETS model

ETS models are generally interpreted using four parameters: α, β, γ, ϕ where α is the smoothing parameter, β is trend, γ is seasonality, ϕ is trend damping weight. For the purposes of the present analysis, we estimated the parameters of various ETS models. Out of these, we selected the one with the lowest values of the AIC and MAPE, i.e. the ETS (AAA) model, which takes into account additive trend, seasonality and error. The parameters of this model are as follows: $\alpha = 0.8872$, $\beta = 1 \cdot 10^{-4}$, $\gamma = 3 \cdot 10^{-4}$. A graph comparing ETS (AAA) forecast and actual observations is shown in Figure 3.

The ETS (AAA) model (Figure 3) fits the time series well; only the maximum values of the test data are higher than the forecast, which may lead to an underestimation for periods of peak transport demand.

4.3. TBATS model

An alternative, complex model that takes seasonality into account is the TBATS model, which additionally permits to model seasonal effects of non-integer lengths. Not all parameters were estimated for the investigated time series. The calculated value of the Box–Cox transformation parameter λ is $\lambda = 0.000195$, which means the transformation is logarithmic: $Y(\lambda) = \log(Y)$. The AR and MA parameters were not estimated (Table 3), which means that apart from the logarithmic transformation, there was no need to correct the theoretical TBATS equations using an additional ARMA model. The damping parameter ϕ was not estimated either, which means that the short-term trend was not weakened and its value in period t was the same as the value of the previous period's trend.

Approximation with trigonometric Fourier series yielded four pairs of series with periodicity $m = 12$. The calculated ETS parameters were $\alpha = 0.6261$, $\beta = 0.1217$, $\gamma_1 = 0.0014$, and $\gamma_2 = 0.01463$. A graph of the TBATS (0, {0, 0}, –, {<12, 4>}) forecast and real observations is presented in Figure 4.

The graph (Figure 4) shows that the TBATS model (0, {0, 0}, –, {<12, 4>})reflects well the nature of real observations, especially in the first months of the forecast; the values for later months deviate more clearly from the measured values, which suggests the forecast may have a poorer predictive quality as the forecasting period becomes longer.
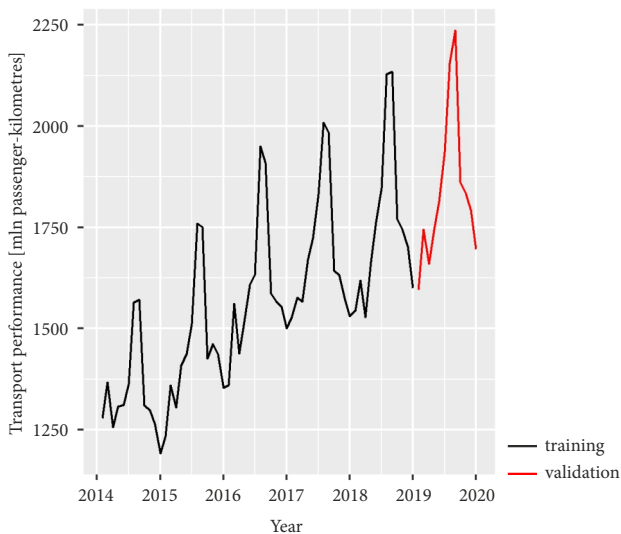
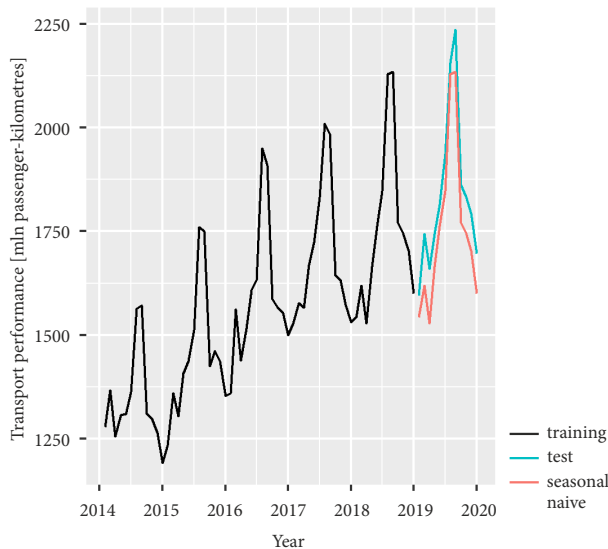Figure 1. Observations divided into a test set and a training set
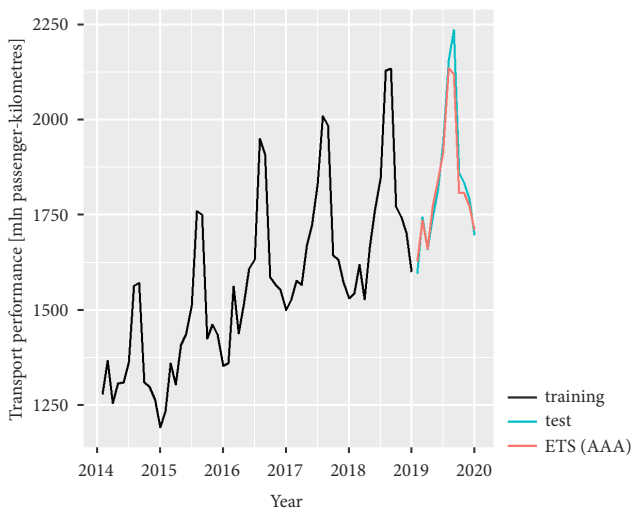

Figure 2. Naive forecast

Figure 3. ETS (AAA) forecast

Figure 4. TBATS forecast
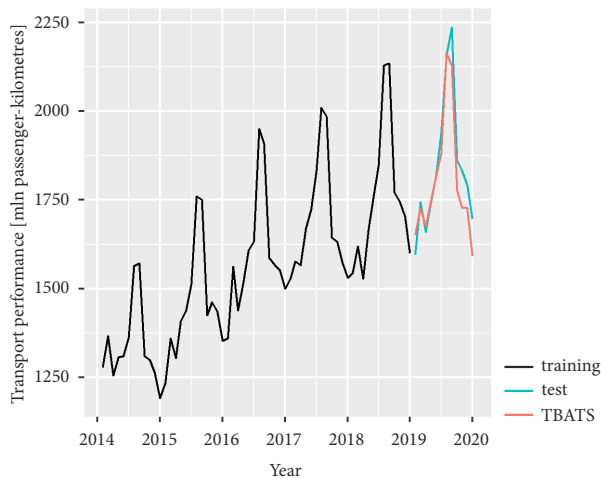
Table 3. Parameters of TBATS model (0, {0, 0}, –, {<12, 4>})

Component	Meaning
0	Box–Cox transformation parameter
{0, 0}	ARMA error
–	damping parameter
{<12, 4>})	seasonal period

Table 4. Estimates of parameters of ARIMA model ((1, 0, 0) (0, 1, 0) [12])

Parameter	Estimate	Standard error	z-value	$\Pr(>|z|)$
$ar1$	0.6521	0.1156	5.6274	$1.829 \cdot 10^8$
$drift$	8.0717	1.6967	4.7588	$1.947 \cdot 10^6$

4.4. ARIMA

Taking into consideration the relationships occurring in the tested time series, i.e. value increase over time and seasonality, an ARIMA model with drift and trend was constructed – ARIMA ((1, 0, 0) (0, 1, 0) [12]). The parameters of this model and their estimates are shown in Table 4.

In this case, only the autoregressive parameter and the drift parameter (Table 4) were estimated. A graph of the ARIMA ((1, 0, 0) (0, 1, 0) [12]) forecast and actual observations is shown in Figure 5.

The ARIMA model ((1, 0, 0) (0, 1, 0) [12]) best reflects the character of the collected empirical data. The forecast fits well with the data for both maximum and minimum observations.

5. Results discussion

In this paper, we analysed a set of transport performance data for passenger rail transport in Poland and compared different models to select one with the best prediction accuracy. The calculated values of selected forecast errors (Table 5) were used as the evaluation criterion.

The smallest forecast errors were obtained when using the ARIMA model ((1, 0, 0) (0, 1, 0) [12]), which confirmed our earlier remarks on the good predictive quality of this type of model. Clear differences in the prediction performance of the individual models can be seen in Figure 6, which plots the forecasts for all the models constructed in this study.

A comparative analysis of the data presented in Table 5 and Figure 6 shows that the data set considered in this paper is best described by the ARIMA model ((1, 0, 0)
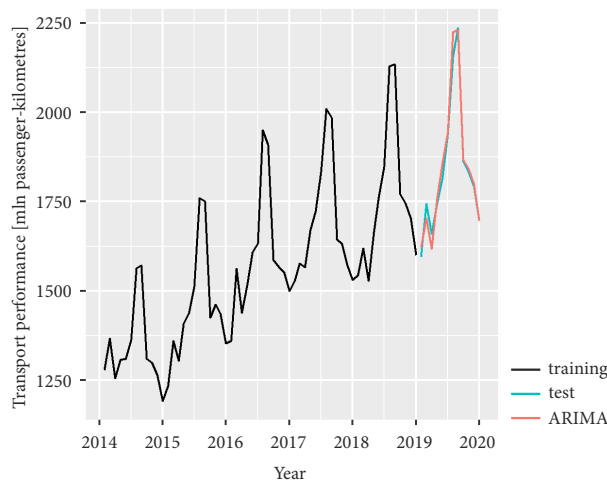

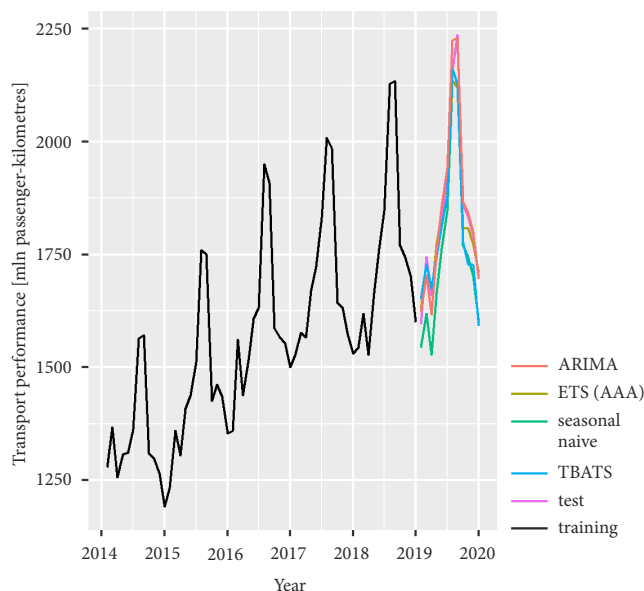Figure 5. ARIMA ((1, 0, 0) (0, 1, 0) [12]) forecast

Figure 6. Forecasts from all models constructed in this study

(0, 1, 0) [12]) with drift. The investigated time series is characterized by strong seasonality and an upward trend. This information is important for planning a development strategy for rail passenger transport, because it shows that additional investments and engagement in the development of both transport infrastructure and superstructure are required to meet the existing demand. The forecast we obtained may also significantly improve the system of scheduling train journeys and determining the level of demand for rolling stock depending on the season and the annual rise in passenger numbers, increasing the effectiveness of management of rail transport. Efficient planning will also reduce the number of overloaded and empty runs during a given season. It is worth emphasizing that the presented methods are universal and can also be applied in other countries. On the basis of benchmarking, decisions can be made regarding forecasting methods as well as strategic activities in the area of rail transport.

Conclusions

The key to maintaining the efficiency of the processes executed in all modes of transport, including rail transport, is optimization of decisions and activities. In this context, various travel factors can be analysed, such as traffic volume, number of passengers, number of accidents, average travelling speed on a given section, or road capacity. Passenger rail transport is an attractive mode of transportation because of the comfort of travelling it offers to passengers. Other important advantages of this mode of transport include its environmental friendliness and the relatively low travelling costs, especially when long distances have to be covered. This provides plenty of room for manoeuvre in the development of transport, but also requires precise forecasting of demand. Accurate forecasts can support dynamic development of rail transport and contribute to the reduction of ecological waste. In view of the continuously growing interest in passenger rail transport, it is worth regularly making in-depth forecasts and predictions for rail journeys using appropriate forecasting models. The analysis presented in this paper, which uses naive, ETS, TBATS, and ARIMA models, demonstrates that it is possible to effectively and efficiently forecast transport performance. It also shows that different models can be used for different data sets, depending on the specific nature of the observations (seasonality, the period of time analysed, expected forecast accuracy, etc.).

The models, analysis and forecast results presented in this paper suffice to make a general assessment of the

Table 5. Comparison of selected forecast errors for all models

Test set	AIC	ME	RMSE	MAE	MPE	MAPE
Seasonal naive		85.24	90.00	85.24	4.71	4.71
ETS	713.69	13.92	42.02	30.66	0.61	1.59
TBATS (0, {0, 0}, –, {<12, 4>})	708.56	38.20	65.80	52.10	1.98	2.82
ARIMA ((1, 0, 0) (0, 1, 0) [12])	518.96	−7.25	30.32	22.16	−0.35	1.21

national demand for transport services and to establish the main goals and directions for the development of this industry. In our future research, we plan to analyse more detailed data, such as time series archived every week or even every hour, relating to specific routes (lines), which will enable detailed scheduling of transport operations adequate to the reported demand. In further analyses, additional variables influencing the transport performance will also be used. This will enable the extension of the presented analyses and the models included in this paper will constitute an excellent background for further considerations.

References

Andersson, M.; Brundell-Freij, K.; Eliasson, J. 2017. Validation of aggregate reference forecasts for passenger transport, *Transportation Research Part A: Policy and Practice* 96: 101–118. https://doi.org/10.1016/j.tra.2016.12.008

Baležentis, A.; Baležentis, T. 2011. Assessing the efficiency of Lithuanian transport sector by applying the methods of MULTIMOORA and data envelopment analysis, *Transport* 26(3): 263–270. https://doi.org/10.3846/16484142.2011.621146

Banerjee, N.; Morton, A.; Akartunalı, K. 2020. Passenger demand forecasting in scheduled transportation, *European Journal of Operational Research* 286(3): 797–810. https://doi.org/10.1016/j.ejor.2019.10.032

Bureika, G.; Gaidamauskas, E.; Kupinas, J.; Bogdevičius, M.; Steišūnas, S. 2017. Modelling the assessment of traffic risk at level crossing of Lithuanian railways, *Transport* 32(3): 282–290. https://doi.org/10.3846/16484142.2016.1244114

Dolinayova, A.; Danis, J.; Cerna, L. 2018. Regional railways transport – effectiveness of the regional railway line, in A. Fraszczyk, M. Marinov (Eds.). *Sustainable Rail Transport*, 181–200. https://doi.org/10.1007/978-3-319-78544-8_10

Duranton, S.; Audier, A.; Hazan, J.; Langhorn, M. P.; Gauche, V. 2017. *The 2017 European Railway Performance Index*. Boston Consulting Group, Boston, MA, US. Available from Internet: https://www.bcg.com/publications/2017/transportation-travel-tourism-2017-european-railway-performance-index

EEA. 2021. *National Emissions Reported to the UNFCCC and to the EU Greenhouse Gas Monitoring Mechanism*. European Environment Agency (EEA). Available from Internet: https://www.eea.europa.eu/data-and-maps/data/national-emissions-reported-to-the-unfccc-and-to-the-eu-greenhouse-gas-monitoring-mechanism-17

Gao, H.; Xu, J.; Li, S.; Xu, L. 2019. Forecast of passenger flow under the interruption of urban rail transit operation, *Lecture Notes in Electrical Engineering* 639: 283–291. https://doi.org/10.1007/978-981-15-2866-8_27

Hyndman, R.; Koehler, A.; Ord, K.; Snyder, R. 2008. *Forecasting with Exponential Smoothing: the State Space Approach*. Springer. 362 p. https://doi.org/10.1007/978-3-540-71918-2

Jarašūnienė, A.; Čižiūnienė, K.; Petraška, A. 2019. Research on rail and maritime transport interoperability in the area of information systems: the case of Lithuania, *Transport* 34(4): 467–475. https://doi.org/10.3846/transport.2019.11236

Jarašūnienė, A.; Sinkevičius, G.; Čižiūnienė, K.; Čereška, A. 2020. Adaptation of the management model of internationalization processes in the development of railway transport activities, *Sustainability* 12(15): 6248. https://doi.org/10.3390/su12156248

Jonaitis, J. 2007. Planning of the amount of trains needed for transportation by rail, *Transport* 22(2): 83–89. https://doi.org/10.3846/16484142.2007.9638104

Kang, R.; Wang, J.; Cheng, J.; Chen, J.; Pang, Y. 2019. Intelligent forecasting of automatic train protection system failure rate in China high-speed railway, *Eksploatacja i Niezawodność – Maintenance and Reliability* 21(4): 567–576. https://doi.org/10.17531/ein.2019.4.5

Konowrocki, R.; Chojnacki, A. 2020. Analysis of rail vehicles' operational reliability in the aspect of safety against derailment based on various methods of determining the assessment criterion, *Eksploatacja i Niezawodność – Maintenance and Reliability* 22(1): 73–85. https://doi.org/10.17531/ein.2020.1.9

Kosicka, E.; Kozłowski, E.; Mazurkiewicz, D. 2018. Intelligent systems of forecasting the failure of machinery park and supporting fulfilment of orders of spare parts, *Advances in Intelligent Systems and Computing* 637: 54–63. https://doi.org/10.1007/978-3-319-64465-3_6

Kozłowski, E. 2015. *Analiza i identyfikacja szeregów czasowych.* Politechnika Lubelska. 273 s. (in Polish).

Kozłowski, E.; Kowalska, B.; Kowalski, D.; Mazurkiewicz, D. 2018. Water demand forecasting by trend and harmonic analysis, *Archives of Civil and Mechanical Engineering* 18(1): 140–148. https://doi.org/10.1016/j.acme.2017.05.006

Kozłowski, E.; Kowalska, B.; Kowalski, D.; Mazurkiewicz, D. 2019a. Survival function in the analysis of the factors influencing the reliability of water wells operation, *Water Resources Management* 33(14): 4909–4921. https://doi.org/10.1007/s11269-019-02419-0

Kozłowski, E.; Mazurkiewicz, D.; Kowalska, B.; Kowalski, D. 2019b. Application of a multidimensional scaling method to identify the factors influencing on reliability of deep wells, *Advances in Intelligent Systems and Computing* 835: 56–65. https://doi.org/10.1007/978-3-319-97490-3_6

Kozłowski, E.; Mazurkiewicz, D.; Żabiński, T.; Prucnal, S.; Sęp, J. 2020. Machining sensor data management for operation-level predictive model, *Expert Systems with Applications* 159: 113600. https://doi.org/10.1016/j.eswa.2020.113600

Liu, J.; Wang, H. 2007. Passengers and goods transportation volume forecast and SWOT analysis of Xiaoyong railway, *Journal of Lanzhou Jiaotong University* (4): 17–20. (in Chinese).

Markovits-Somogyi, R. 2011. Measuring efficiency in transport: the state of the art of applying data envelopment analysis, *Transport* 26(1): 11–19. https://doi.org/10.3846/16484142.2011.555500

Milenković, M.; Švadlenka, L.; Melichar, V.; Bojović, N.; Avramović, Z. 2018. SARIMA modelling approach for railway passenger flow forecasting, *Transport* 33(5): 1113–1120. https://doi.org/10.3846/16484142.2016.1139623

Namiot, D. E.; Pokusaev, O. N.; Lazutkina, V. S. 2018. O modeljah passazhirskogo potoka dlja gorodskih zheleznyh dorog, *International Journal of Open Information Technologies* 6(3): 9–14. (in Russian).

Prakaulya, V.; Sharma, R.; Singh, U.; Itare, R. 2017. Railway passenger forecasting using time series decomposition model, in *2017 International conference of Electronics, Communication and Aerospace Technology* (ICECA), 20–22 April 2017, Coimbatore, India, 554–558. https://doi.org/10.1109/ICECA.2017.8212725

Rabiega, W. P.; Sikora, P. 2020. *Ścieżki redukcji emisji CO_2 w sektorze transportu w polsce w kontekście "Europejskiego zielonego ładu"*. Centrum analiz klimatyczno-energetycznych, Warszawa, Polska. 48 s. (in Polish). Available from Internet: http://climatecake.pl/wp-content/uploads/2020/10/%C5%9A-

cie%C5%BCki-redukcji-emisji-CO2-w-sektorze-transpor-
tu-w-PL-w-kontek%C5%9Bcie-Europejskiego-Zielonego-
-%C5%81adu.pdf

Redakcja. 2020. Inwestycje są potrzebne. Polski tabor się starzeje, *Raport Kolejowy*, 11 stycznia 2020 (in Polish). Available from Internet: https://raportkolejowy.pl/inwestycje-sa-potrzebne-polski-tabor-sie-starzeje/

Roos, J.; Gavin, G.; Bonnevay, S. 2017. A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data, *Transportation Research Procedia* 26: 53–61. https://doi.org/10.1016/j.trpro.2017.07.008

UTK. 2020. *Statystyka przewozów pasażerskich*. Urząd Transportu Kolejowego (UTK). (in Polish). Available from Internet: https://utk.gov.pl/pl/raporty-i-analizy/analizy-i-monitoring/statystyka-przewozow-pa

Schütze, C.; Schmidt, N.; Liimatainen, H.; Siefer, T. 2020. How to achieve a continuous increase in public transport ridership? – A case study of Braunschweig and Tampere, *Sustainability* 12(19): 8063. https://doi.org/10.3390/su12198063

Song, H.; Schnieder, E. 2018. Modeling of railway system maintenance and availability by means of colored Petri nets, *Eksploatacja i Niezawodność – Maintenance and Reliability* 20(2): 236–243. https://doi.org/10.17531/ein.2018.2.08

Tang, Q.; Cheng, P.; Li, N. 2017. Short time forecasting of passenger flow in urban railway using GSO-BPNN method, *Technology* & *Economy in Areas of Communications* (1): 1–4. (in Chinese).

Zhang, J.; Chen, F.; Shen, Q. 2019. Cluster-based LSTM network for short-term passenger flow forecasting in urban rail transit, *IEEE Access* 7: 147653–147671.
https://doi.org/10.1109/ACCESS.2019.2941987

Zhou, W.; You, X.; Fan, W. 2020. A mixed integer linear programming method for simultaneous multi-periodic train timetabling and routing on a high-speed rail network, *Sustainability* 12(3): 1131. https://doi.org/10.3390/su12031131