ISBN 978-83-959742-6-7

Transport Problems 2023
Conference proceedings

Konferencja dofinansowana ze środków budżetu państwa
w ramach programu Ministra Edukacji i Nauki pod nazwą
"Doskonała Nauka" nr projektu DNK/SP/549716/2022
kwota dofinansowania 145 200,00 zł
całkowita wartość projektu 299 000, 00 zł.

Ministerstwo Edukacji i Nauki

Doskonała Nauka

UNDER THE HONORARY PATRONAGE OF

Silesian University of Technology

Krajowa Reprezentacja Doktorantów KRD

XV INTERNATIONAL SCIENTIFIC CONFERENCE
21-23.06 2023
Katowice - Częstochowa

19-20.06 2023
Katowice
XII INTERNATIONAL SYMPOSIUM OF YOUNG RESEARCHERS

Silesian University of Technology
Faculty of Transport and Aviation Engineering

Transport Problems 2023

Proceedings

XV International Scientific Conference
XII International Symposium of Young Researchers

UNDER THE HONORARY PATRONAGE OF

Silesian University of Technology

Krajowa Reprezentacja Doktorantów KRD

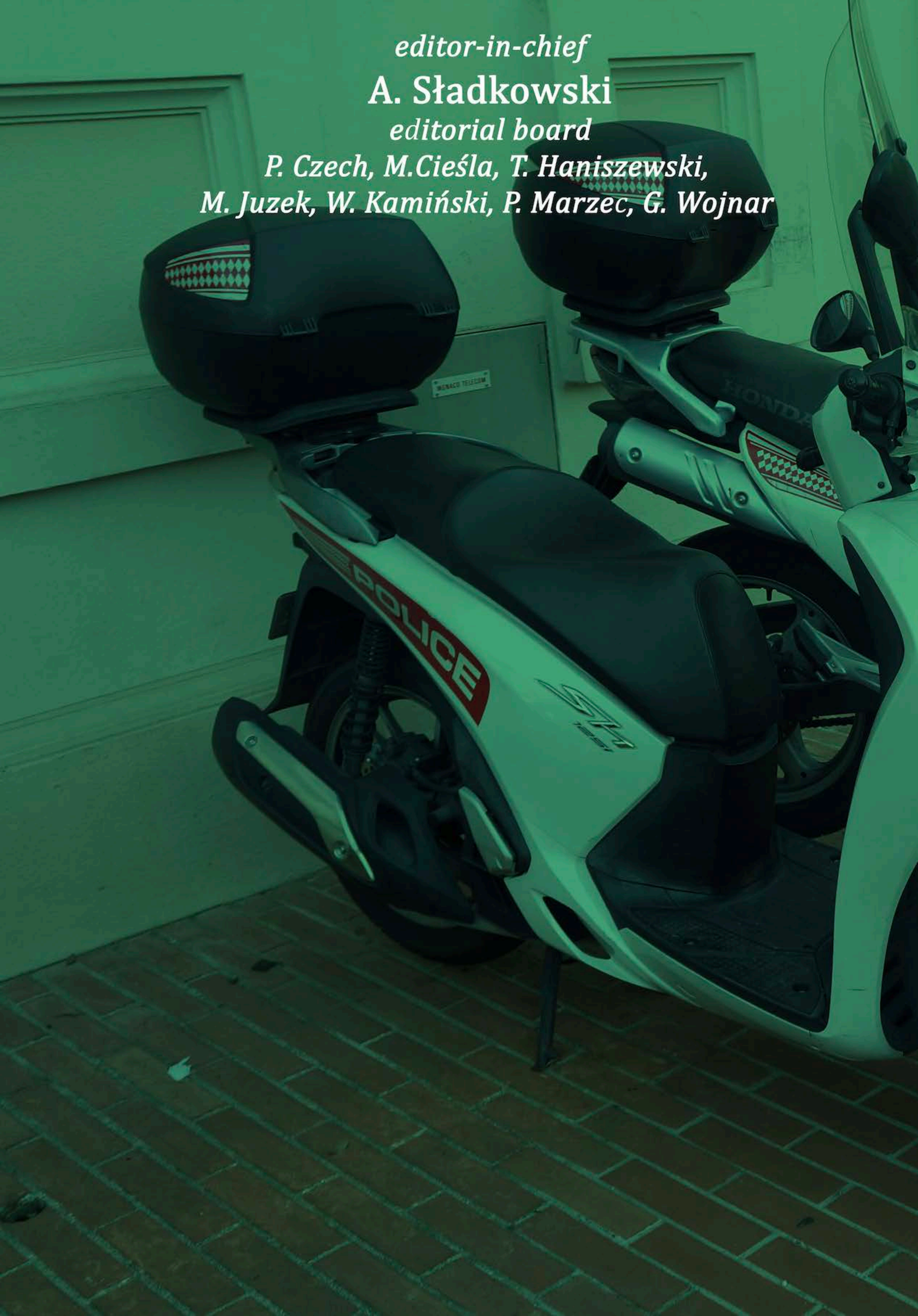
Media patronage:
Transport Problems International Scientific Journal
ISSN 1896-0596, The Silesian University of Technology,
Faculty of Transport and Aviation Engineering

Transport Problems
International Scientific Journal

Ministerstwo Edukacji i Nauki

Doskonała Nauka

Konferencja dofinansowana ze środków budżetu państwa
w ramach programu Ministra Edukacji i Nauki pod nazwą
"Doskonała Nauka" nr projektu DNK/SP/549716/2022
kwota dofinansowania 145 200,00 zł
całkowita wartość projektu 299 000, 00 zł.



Keywords: demand estimation, railway freight transportation, volume and turnover, ARIMA, forecast quality assessment

Madiyar SULTANBEK*, Nazdana ADILOVA
Al-Farabi KazNU
Al-Farabi av. 71, 050040 Almaty, Kazakhstan
Aleksander SŁADKOWSKI
Silesian University of Technology
Krasińskiego 8, 40-019 Katowice, Poland
Arnur KARIBAYEV
JSV "National company "Kazakhstan Temir Zholy"
Kunayev 6, 010000 Astana, Kazakhstan
**Corresponding author*. E-mail: sultanbekmadiyar@gmail.com

ESTIMATING THE DEMAND FOR RAILWAY FREIGHT TRANSPORTATION IN KAZAKHSTAN: A CASE STUDY

Summary. Rational resource planning of the company requires demand forecasting, as numerous factors can impact the future volume and turnover of railway freight transportation. Expert methods of forecasting no longer meet current requirements. To improve the accuracy of demand forecasting, this article presents the first application of mathematical models in National State-owned Railway Company in Kazakhstan, specifically the autoregressive integrated moving average (ARIMA), based on previous data from 2012 to 2016. The purpose of the study is to compare the accuracy of two different methods of forecasting with actual data from 2017, using mean absolute error (MAE) and mean absolute percentage error (MAPE) as quality assessment measures. The findings suggest that time series analysis improves the quality of railway freight demand forecasts in Kazakhstan. The article also includes a forecast for the volume of railway freight transportation in Kazakhstan for 2022, which was made using the ARIMA method and compared to actual data, showing high accuracy.

1. INTRODUCTION

In any company, demand estimates are crucial for effective planning and decision-making, serving as a primary input. Various departments such as marketing, production, distribution, and finance rely on short-to-long-term forecasts to make different decisions. Due to their critical role in business decision-making, the accuracy and quality of these forecasts are highly significant [1].

Forecasting demand is a crucial component of managing business processes. Although the methods for forecasting can vary across different businesses and can be quite intricate, the ultimate goal remains consistent: to obtain a reasonably precise prediction of future demand for a product or service by utilizing historical data and current environmental factors (such as political, social, and economic conditions) in order to plan and coordinate business operations appropriately [2].

Predicting the future demand for transportation services is a crucial factor for the success of a transportation company. This forecast also serves as fundamental information for the planning and management of functional areas such as transportation operation planning, marketing, and finance [3].

Joint Stock Company "National Company "Kazakhstan Temir Zholy" (hereinafter referred to as KTZ) is a transport and logistics holding engaged in rail transportation. The sole shareholder is "Samruk-Kazyna" joint-stock company, which delegates the general management of the group's activities to the KTZ's Board of Directors. The sole shareholder of "Samruk-Kazyna" JSC in turn is the Government of the Republic of Kazakhstan. KTZ's corporate portfolio of assets at the end of 2019 included 56 companies, including 1 organization in trust management. The main subsidiaries and structural companies of the KTZ operate in the segments "Main railway network services", "Rail freight transportation", "Passenger rail transportation" and "Freight cars operations".

KTZ's main sources of income are income from freight and passenger transportation. The share of income from freight transportation is 86% of total income and consists of all components that provide transportation activities: services of the main railway network, locomotive traction services, freight commercial work and the provision of wagons.

The length of railway lines (unfolded length) is more than 21 thousand km, the fleet of freight cars is about 54 thousand units, the fleet of passenger cars is more than 2 thousand units, the fleet of locomotives is more than 1.6 thousand units. KTZ is the country's largest employer (over 115,000 employees).

Transportation is the basis for obtaining the revenue of a freight carrier. In freight railway transport in Kazakhstan, the volume of applications is almost equal to the volume of traffic. Therefore, the volume of cargo transportation in tons assess the demand for the services of a railway carrier, as well as the volume of cargo transportation in tons, multiplied by the distance transported in kilometers, the so-called freight turnover. At the same time, when calculating demand, it is precisely the tariff freight turnover that is taken into account, in which the tariff (that is, the shortest) distance between the points of loading and unloading is taken and multiplied by the volume of traffic. Freight turnover, in turn, becomes the basis for calculating future revenues from freight traffic, or, in other words, income from the main activity of freight transportation.

Different techniques have been devised for forecasting, which are based on two main approaches: qualitative and quantitative. The qualitative methods, such as Executive opinions, Delphi technique, Sales force polling, and Customer services, rely on judgments or opinions to generate forecasts. On the other hand, quantitative methods can be categorized as historical data forecasts, such as Naive method, Trend Analysis, Time Series Analysis, Holt's and Winter's models, or associative forecasts, which identify causal relationships between variables using Simple, Multiple or Symbolic regression. Furthermore, there are mixed or combined models that enable the integration of both approaches.

Numerous research centers have conducted studies on constructing models to describe the demand for rail services. For instance, [3] used the seasonal ARIMA model to forecast monthly passenger flows on Serbian railways, while [4] discussed an approach to forecasting short-term passenger flows in the Parisian urban rail network based on dynamic Bayesian networks. [5] analyzed the transport performance of the urban rail transit network in Beijing using a long short-time memory network, and [6] presented an alternative method of analysis, which combined a backpropagation neural network and the glow-worm swarm optimization algorithm. In contrast, [7] described methods for forecasting passenger traffic in Moscow based on network topology analysis. However, most of these studies are limited to analyzing cities that develop dynamically due to increased demand for rail transport, caused by urban sprawl and deteriorating road transport conditions. Fewer models are available for assessing the functioning of larger national rail networks, such as those developed in Sweden [8] and India [9].

Moreover, some studies have focused on evaluating the performance of different companies and organizations' decision-making units using the Data Envelopment Analysis (DEA) tool. [10] reviewed the results of such research and compiled data from 69 DEA applications reported in the literature. The author investigated their characteristics, fields of application, and the inputs and outputs used. However, the use of DEA has a serious drawback of being sensitive to measurement errors and noise in data. There is limited research that compares multiple modelling approaches to determine the most effective one. [11], for instance, offer several models that could be utilized to predict demand in the regular passenger transport sector.

Demand analyses and forecasts are crucial for developing effective decisions, but demand data are not always available due to a lack of appropriate mathematical models for generating demand forecasts. Therefore, it is necessary to analyze railway systems of various countries to select appropriate forecasting methods, which can be used as a scientific database for research conducted in other countries or transport systems.

This study aims to identify the parameters of a mathematical model of rail freight transport performance that will allow making reliable forecasts of future demand for this service using historical data of the Kazakhstan National State-owned railway company. Several models dedicated to this type of empirical data were proposed and evaluated, and the best model was identified, its accuracy and effectiveness were assessed, and a forecast of transport performance in sequential periods of time was presented. The paper includes an introduction, a section on the data and methodology of the study, sections on the mathematical models constructed using the adopted methods, empirical data, and summary of results, conclusions, and future research directions.

2.DATA AND METHODS

Current planning process in KTZ lacks of any modern tool. The paper [12] employs various techniques such as seasonal naive model, exponential smoothing model (ETS), exponential smoothing state-space model with Box-Cox transformation, ARMA (AutoRegressive Moving Average) errors, trigonometric trend and seasonal components (TBATS) model, to forecast demand in Polish Railways. The researchers conclude that the ARIMA (AutoRegressive Integrated Moving Average) method exhibits the least error. Consequently, this study introduces the initial implementation of the ARIMA model to produce a forecast for rail freight transportation in the case of KTZ. The current process of rail freight demand estimation in KTZ completely depends on a person - an expert in the field of freight transportation marketing who makes estimations using MS Excel. KTZ's Marketing and Tariff Policy Department (hereinafter, MTPD) is responsible for freight demand estimation in KTZ. MTPD uses following methods for freight demand estimation:

1) Expert estimates based on an assessment of the current moment and development prospects. MTPD experts analyze historical data of transportation for several years, studying the factors that have influenced freight transportation in the past. They also use forecasts of major shippers (if available), and opinions of leading experts in different industries related to transported cargo types.

2) Extrapolation - a method used to estimate the distribution of past trends for a future period, commonly used in transportation demand calculations for consignors not included in surveyed groups. The primary scientific method employed by MTPD staff for demand estimation is extrapolation, which is a form of approximation where the function is approximated outside a given interval instead of between given values. Linear extrapolation is the most frequently used method. Expert judgment and experience are also used, but extrapolation is the preferred method.

However, the MTPD's use of extrapolation has significant limitations. External environmental changes and the influence of external factors on demand estimates are not considered. For example, changes in exchange rates can have a substantial effect on transportation volume and geography, but the extrapolation method does not account for this. Extrapolation involves transferring conclusions made about a portion of objects or phenomena to the entire set or another section of them.

Based on the above, the following conclusions can be drawn:

1) MTPD experts are the key links in the process at all stages of forming the freight demand estimation. In the scientific literature, this method of forecasting is called an expert method. "Expert" in Latin means "experienced". The demand estimations made by an expert or team of experts based on their professional, scientific, and practical experience and opinion. Expert methods are normally applied in the following cases: if the object of research is extremely simple or, on the contrary, in case of extreme complexity of the object of estimation, its novelty, uncertainty of formation of some essential features, insufficient completeness of data and impossibility of complete mathematical formalization of the process of solving the problem set. The main principle underlying the methods of individual expert evaluations is the maximum possibility of using individual abilities of the expert. Since MTPD experts have an access to a vast amount of available digital historical data on transportation, the expert method of forecasting, as follows from the previous narrative, is not rational.

2) MTPD experts spend most of their time on operations like downloading data from KTZ systems, uploading to personal computers, generating summary tables, preparing data, generating reports, graphs,

tables, preparing paper questionnaires for shippers, and manually processing survey results. That is, most of the MTPD expert's working time spent on routine operations.

3) Processing of large data sets from various KTZ systems carried out in MS Excel, whose capabilities for processing large data sets are severely limited. For example, MS Excel unable to create tables with more than 1 048 576 rows and 16 384 columns. The analysis of thousands cargo types codes from the unified tariff and statistical nomenclature of goods (further - UTSNG) by hundreds of stations of departure and destination and by hundreds of shippers may potentially create the need for tables with tens of millions of rows and columns. In addition, MS Excel is limited in the number of available libraries for forecasting, so MTPD experts use the linear extrapolation method only, available in MS Excel. The lack of technical capability of the MTPD experts to perform a more detailed analysis of the input historical data and the inability to use many other methods of mathematical or statistical analysis besides linear extrapolation leads to simplifications, averaging and, consequently, to a deterioration in the quality of demand estimations.

There is a logical conclusion - it is necessary to partly automate the process of demand estimations based on modern software and thus accelerate the process of getting and loading data, analysis of large sets of data using a variety of methods, but not to replace the marketing experts with the program, but to increase productivity of the marketing expert. It is necessary for the marketing expert to spend more time on the analysis interpretation, rather than on the compilation of statistics. This requires the use of special software products to prepare and analyze data sets due to the high performance of database management systems and built-in libraries of algorithms, in which computing, and processing of data sets occurs in a matter of seconds much faster than manually.

KTZ decided to conduct a pilot research or experiment on the freight transportation demand estimations using specialized software that processes and analyzes data sets and compare the quality of the estimation results with the MTPD experts' estimations (marketing experts' opinion and linear extrapolation).

The experiment divided into several stages:

1. The monthly historical data on the rail freight volume and turnover from 2012 to 2016 for each nomenclature of goods UTSNG (unified tariff and statistical nomenclature of goods) and 13 aggregated nomenclatures of goods and for all types of communication (export, import, transit, and domestic transportation) were loaded from the KTZ systems into a specialized program for analysis, data science and forecasting.

2. Macroeconomic indicators (predictors) were found and loaded into a specialized program (in fact, 260 indicators in the appropriate format and periodicity of data were collected), that potentially correlate with the historical volumes of transportation or freight turnover. To assess the correlation level on the historical freight volume and freight turnover over a five-year period on monthly basis with predictors, it is vital that all predictors uploaded in appropriate granularity.

3. A model was created and tested on test data for 2012-2016, and then the model automatically generated a monthly forecast for 2017, and then compared with the 2017 actuals and with the estimations made by MTPD experts in 2016 for 2017. The comparison made for each aggregated nomenclature of cargo and for each type of communication (internal, export, import and transit) separately.

4. The assessment of the quality of the forecast carried out according to MAPE - mean absolute percentage error or MAE - mean absolute error, because these are the most common methods of assessment used in forecasting and checking the quality of demand estimation models. Formulas for calculating MAPE and MAE are presented below, where Z(t) is the actual value of the time series and X(t) is the forecast value. MAE is applied if the actual value of the indicator is zero. We compared the forecast with the fact and derived the MAPE/MAE indicator both for the manual forecast made by MTPD experts and for the forecast made by specialized software.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|Z(t)-X(t)|}{Z(t)} * 100\% \tag{1}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|Z(t)-X(t)| \tag{2}$$

IBM SPSS (Statistical Package for the Social Sciences) Modeler 18.0 (hereinafter referred to as SPSS) - visual data science and machine learning (ML) solution by IBM Company chosen as a specialized software for data analysis and demand estimation, because IBM product was ranked first in the category of Data Science platforms in the Gartner ranking in 2017 [13].

Since many statistical and mathematical forecasting methods are available in SPSS, there was an additional task to choose the best method for demand estimation among those available for use in SPSS. As the result, forecasting was done using three models: ARIMA model (autoregressive integrated moving average) - ARIMA is amongst the most generally used to forecast transport demand [14, 3]; neural net model (Nnet, neural network) and Auto Classifier (a combination of neural network methods, C&R tree, CHAID model, linear regression, support vector mechanism and others). The best model out of three mentioned was selected further.

2.1. Models used in the design study

Neural network. A neural network (Nnet in Fig.2) has the ability to approximate a wide range of predictive models without much reliance on specific model structures or assumptions. The learning process determines the form of the relationships between the target and predictors. If a linear relationship is appropriate, the neural network's results will closely resemble those of a traditional linear model. However, if a nonlinear relationship is more appropriate, the neural network will automatically approximate the correct model structure.

Although the neural network provides flexibility in modeling, it is not easily interpretable, making it challenging to explain the underlying process that produces the relationship between the target and predictors. In such cases, a more traditional statistical model would be better suited for interpretation. However, if interpretability is not critical, the neural network can provide good predictions.

To build a neural network model, at least one target and one input field are necessary. Measurement level restrictions do not apply to targets or predictors (inputs). During model building, the initial weights assigned to the neural network, and the final models produced, depend on the field order in the data. However, Cloud Pak for Data sorts data by field name before presenting it to the neural network for training, so explicitly changing the field order upstream will not affect the generated neural net models when a random seed is set in the model builder. Nevertheless, modifying the input field names in a manner that changes their sort order will produce different neural network models, even with a random seed set in the model builder. Nonetheless, changing the field names' sort order will not significantly affect the model's quality [15].

ARIMA. The ARIMA (Arima in Fig.2) technique enables the development of autoregressive integrated moving average models that can be used to refine time series simulations. ARIMA models offer more advanced techniques for modeling trend and seasonal factors than exponential smoothing models, and they also have the added benefit of allowing predictor variables to be incorporated into the model. By specifying the autoregressive, differential, and moving average order, as well as their seasonal counterparts, the ARIMA approach can be used to fine-tune the model. However, determining the rational values for these components through trial and error can be a time-consuming process [16].

Auto Classifier. The Auto Classifier node (Auto in Fig.2) evaluates and contrasts models by utilizing various techniques. This allows to experiment with different approaches within a single modeling run. For instance, instead of selecting one method for an SVM, such as Radial Basis Function, polynomial, sigmoid, or linear methods, you can try them all. The node explores all potential combinations of options, ranks each model based on the metric you specify, and saves the best models for scoring or further analysis. The supported model types include Neural Net, C&R Tree, QUEST, CHAID, C5.0, Logistic Regression, Decision List, Bayes Net, Discriminant, Nearest Neighbor, SVM, XGBoost Tree, and XGBoost-AS [17].

2.2. Description of Experiment Progress

Historical rail transportation volume and freight turnover data for 13 aggregated cargo nomenclatures over a five-year period (2012 to 2016), presented on a monthly scale, were loaded into SPSS. For each aggregated nomenclature of goods combinations were selected - Unified Tariff and Statistical Nomenclature of Goods (UTSNG) cargo type code, country of origin and country of destination. Then macroeconomic indicators from countries that have strong trade relations with Kazakhstan were found and loaded into the system, such as the volume of production of coal, oil, ore, electricity, etc.; the volume of export/import of goods; prices for various types of raw materials; exchange rates against local currency. All macro indicators in an appropriate granularity of monthly format for the same period as the historical data on rail freight transportation loaded into SPSS. Using special tools in SPSS, the correlation between macro indicators (predictors) and historical data analyzed, and the influence of predictors on historical data estimated. Then the model was trained on the training sample and tested on the test data of 2012-2016, and then formed a forecast by month for 2017 for all thirteen cargo nomenclatures by rail transportation volume using three different methods (ARIMA, neural net and auto classifier), of which the best was the ARIMA forecast. The best forecast generated by SPSS (ARIMA) compared with the 2017 actual freight transportation performance and the MTPD estimates generated using methods described in the Introduction section of the paper.

The Fig.1 and Fig.2 show the algorithm of data aggregation and modelling algorithm in SPSS. Fig.1 shows the algorithm used to aggregate all historical data into one new dataset. Fig.2 shows algorithm for creating, training and testing the models on new dataset, and forecasting results of the models.

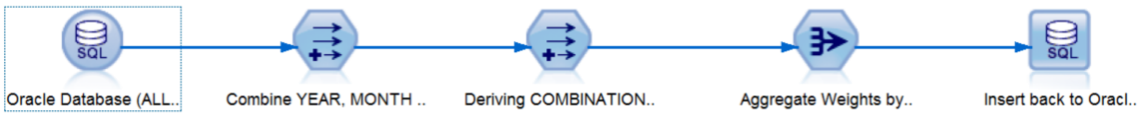


Fig. 1. Data aggregation algorithm in IBM SPSS Modeler.

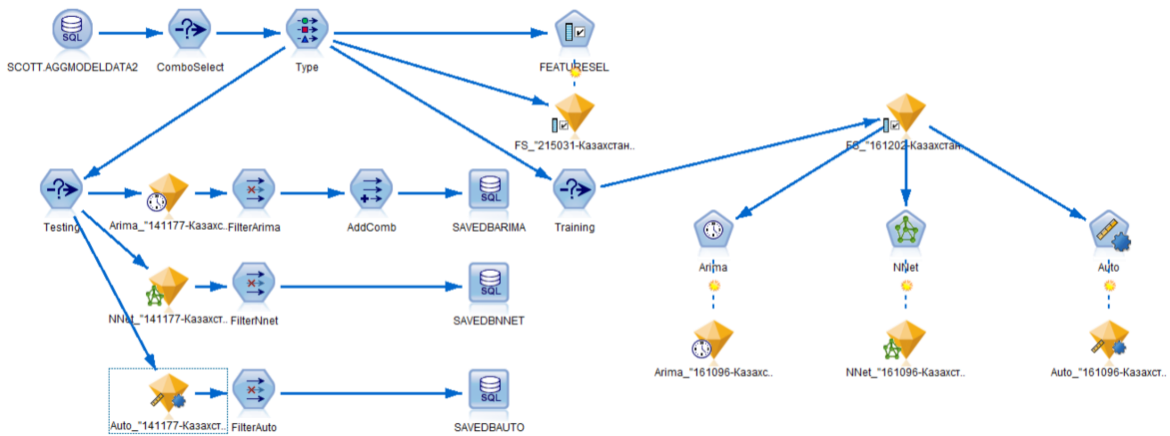


Fig. 2. Algorithm in SPSS using three different models (ARIMA, Nnet, Auto)

3. RESULTS

The main results of the research work and the experiment in graphical form presented below. In Fig. 3, we see three lines on the graph that show a comparison of freight traffic for all nomenclatures of goods and all types of cargo: the forecast of experts of the MTPD KTZ (green line), the forecast using ARIMA (blue line) and the actual volume of traffic in 2017 (red line). The Fig. 1 clearly shows that the blue line of the ARIMA monthly forecast for the volume of freight transportation and the red line of the actual volume of freight transportation practically coincide starting from the third month of 2017. At the same time, the green line differs significantly from the fact. The value of the MAPE indicator for the forecast of experts of the MTPD KTZ was 9.2%, while for ARIMA - 2.0%, which indicates a significant excess of the quality of the ARIMA forecast over the expert forecast.

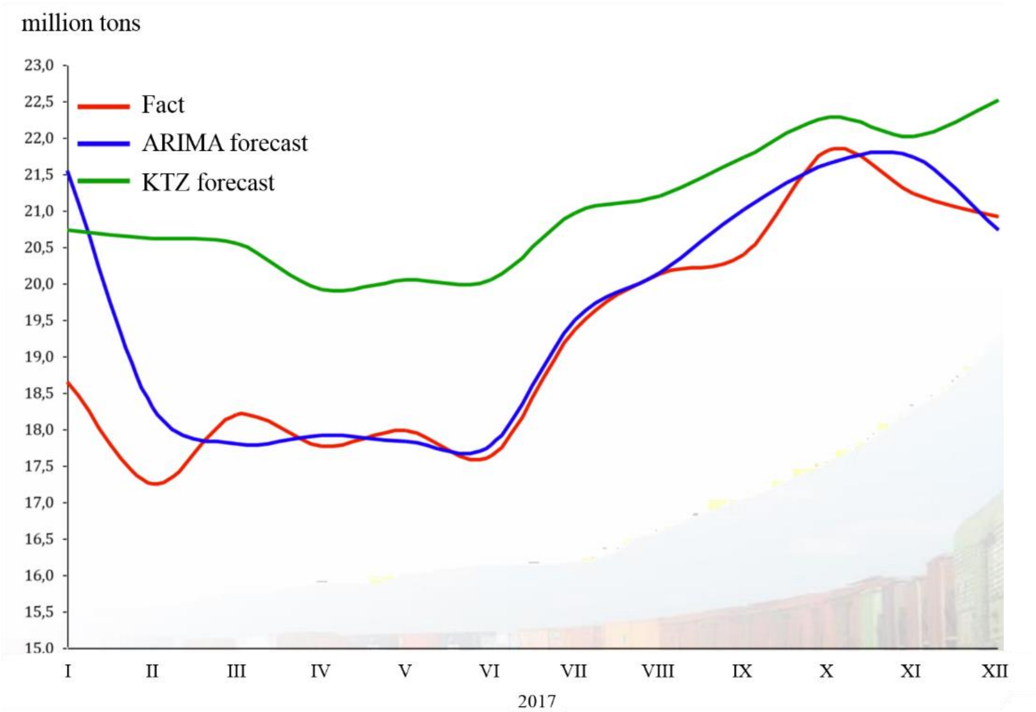


Fig. 3. Cumulative estimation of rail freight volume for all cargo types nomenclature and communication types compared to the 2017 fact and MTPD KTZ experts' estimations

As can be seen in Fig. 4, the forecast for total coal transportation volume, which is the main nomenclature of cargo type transported by KTZ (the share of coal in freight transportation volume exceeds 40%), was much better predicted by the ARIMA model. The value of the MAPE indicator for the forecast of experts of the MTPD KTZ was 6.6%, while for ARIMA - 2.6%.

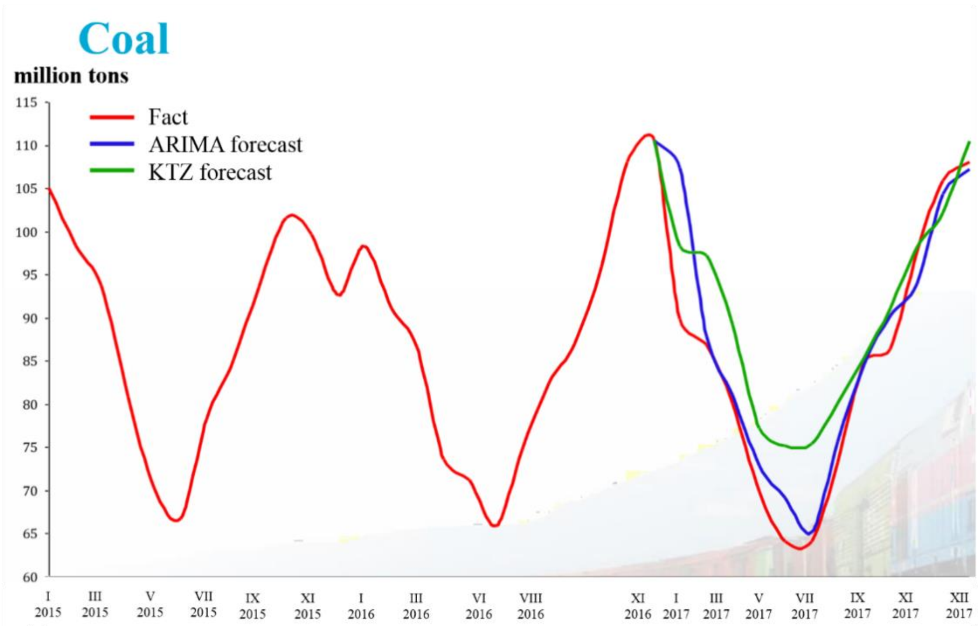


Fig. 4. Comparison of freight transportation volume in 2017 actually with MTPD experts' estimations and ARIMA model forecast on the total coal transportation volume in all types of communication (internal, export, import and transit)

One of the reasons the ARIMA model estimation for freight transportation volumes is so accurate – a strong correlation level of transportation volumes with the macro indicators (predictors) found during the research, as well as the high seasonality of coal transportation. The "predictor screening" feature in IBM SPSS Modeler allows selecting characteristics, helping to identify the fields most important in predicting certain outputs. From a set of hundreds or even thousands of predictors, the “feature selection” (See Fig.2) node ranks, and selects the predictors that are most important, and it helps to end up with a faster and more efficient model that uses fewer predictors types, runs faster, and is easier to interpret.

Fig. 5 shows that the ARIMA predictive model gives much better results compared to even the neural network model and auto classifier model (Auto in Fig.5). However, the neural network model (Nnet in Fig. 5) requires more “fine” tuning, and has the potential for improvement.

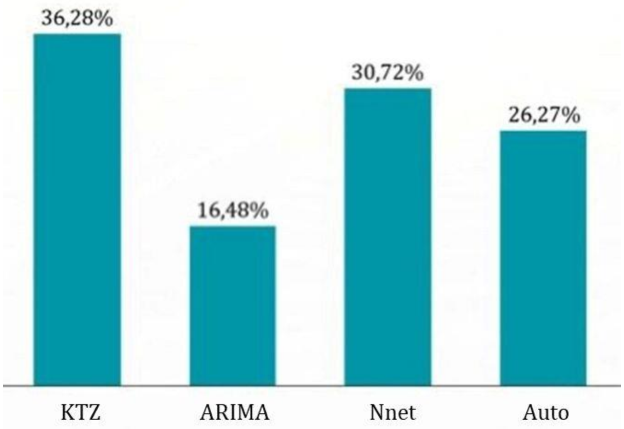


Fig. 5. The indicator of the MAPE for all nomenclatures of goods, all types of forecasts (volume and cargo turnover), all types of communication using various forecasting methods available in IBM SPSS Modeler

Using the selected ARIMA method in 2021, based on historical data for 5 years, a forecast was formed for the volume of traffic in thousands of tons for 2022. In the figures below you can see the results of comparing the forecast with the fact for 2022 for all types of traffic communication and for all total nomenclatures of goods.

As can be seen in Fig. 6, which shows a comparison of the forecast for 2022 in the export direction with the actual traffic volumes, the first three months of 2022 were forecasted quite well. We see how the forecast and fact lines practically coincide, although the forecast for 2022 is formed in June-July of the previous year in accordance with KTZ corporate procedures. However, since April 2022, export traffic has fallen sharply and deviated from the forecast values, which affected the value of the MAPE indicator. According to railway industry experts, a sharp decline in exports could be due to political reasons, including the imposition of sanctions against Russian companies, which, as a result, lost many sales markets and, accordingly, began to purchase less raw materials and components. However, in the last quarter of 2022, the forecast and actual traffic showed a similar trend line. Overall, the MAPE score for the export forecast was 12.4%, below the target threshold of 10%. It is necessary to conduct an additional study of the causes and possibly refine the model.

Fig. 7 shows the forecast of imports for all types of goods in 2022 compared with the actual volumes of imported freight traffic. We see significant fluctuations in the volume of imports in the first 4 months of 2022 and then moving towards more stable monthly traffic volumes in the subsequent months of 2022. In the period from month 05 to month 09, the forecast line and the actual value line are very close to each other, which affected the final value of the MAPE indicator of 5.3%, which fits into the target range of forecast quality. Abnormal values of imports at the beginning of 2022 could also be dictated by political and, as a result, foreign economic factors related to Kazakhstan's neighborhood with Russia. Given the sharp fluctuations at the beginning of the year, the import-forecasting model requires adjustments.

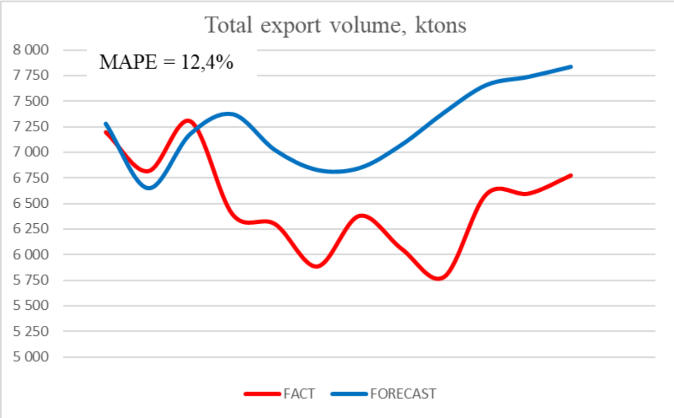


Fig. 6. Total 2022 export volume for all types of cargo nomenclature

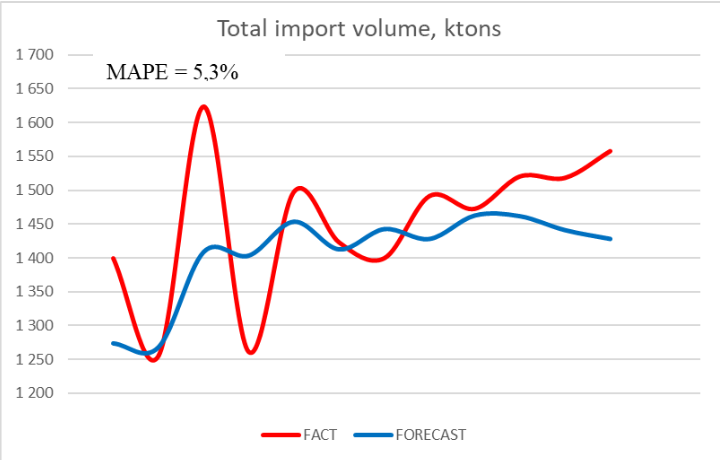


Fig. 7. Total 2022 import volume for all types of cargo nomenclature

Fig. 8 shows the forecast and fact lines for the transit direction of cargo transportation. We again see a dramatic drop in transit in the first 4 months of 2022 and a complete divergence from the forecast. Transit traffic is usually the most difficult to predict, as it is the movement of goods that originate outside our territory and are also sent to foreign customers, often located far beyond the national borders of Kazakhstan. Considering the fact that the main transit through Kazakhstan traditionally passed further through the territory of Russia, with which Kazakhstan has the longest land border in the world, the influence of political and economic factors on the volume of transit traffic at the beginning of 2022 is obvious and difficult to predict. However, due to the sharp growth and further relative stabilization of transit traffic volumes in the second half of 2022, the MAPE indicator for the whole year amounted to 8.1%. However, given what happened, the transit forecasting model also requires further study and tuning.

Fig. 9 compares the forecast of internal cargo traffic within the country's borders with the actual volumes of goods transported. The figure shows that the first 8 months of the forecast and fact lines are very close to each other, and the discrepancies are visible only in the final 4 months of 2022. Obviously, the model predicted future traffic volumes very well, although there is a tendency to somewhat overestimate the volume of traffic. The value of the MAPE indicator was 6.1%.

Fig. 10 shows the volume of freight transportation in total in all directions and for all types of cargo. Considering that the volume of domestic traffic significantly exceeds the volume of freight transportation in export, import and transit directions, the forecast and actual lines are very similar to what we saw in Fig. 9. Therefore, the MAPE indicator is similar to the value of 6.1% as in domestic freight transportation forecast.

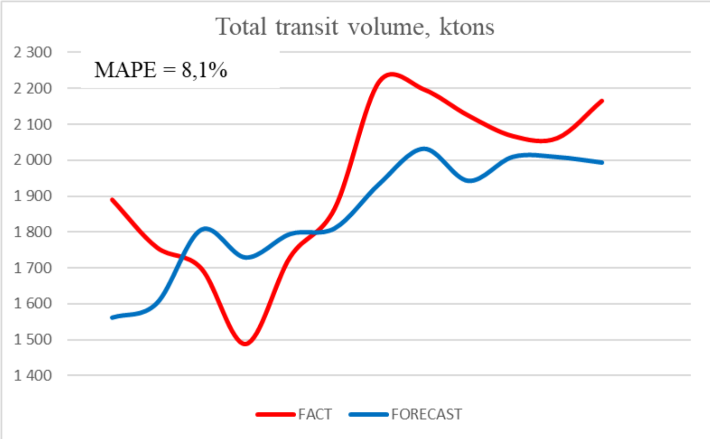


Fig. 8. Total 2022 transit volume for all types of cargo nomenclature

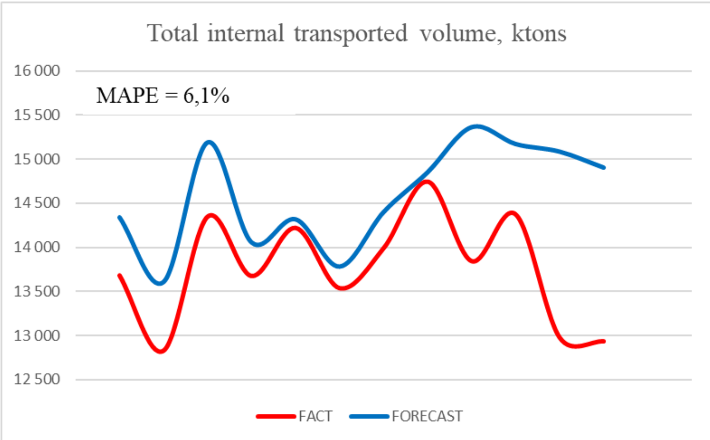


Fig. 9. Total 2022 internal volume for all types of cargo nomenclature

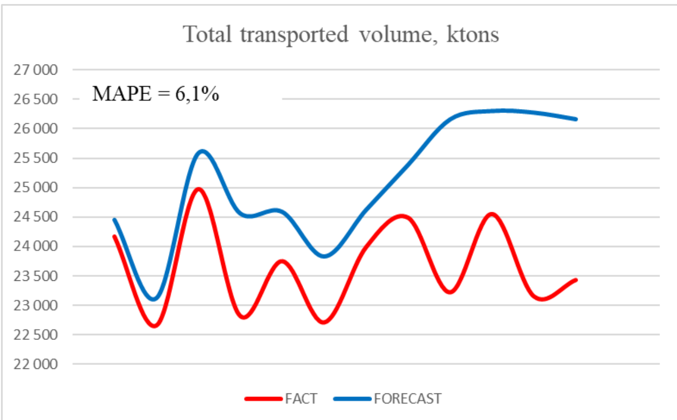


Fig. 10. Total 2022 transported volume for all types of communication and all types of cargo nomenclature.

In general, it can be seen that in domestic transportation and in exports, the model as a whole somewhat overestimates the forecast, while in imports and transit, on the contrary, it underestimates. These features of the models also need to be studied and refined.

4. DISCUSSION

The advantages of using the methods and models described above:

1. The assumption that the ARIMA model will perform well in predicting the demand for freight transport in KTZ, similarly to the fact that in study [12] ARIMA showed the best results in predicting passenger traffic in rail transport in Poland, turned out to be correct. ARIMA really showed the best results, not only in 2017, but also relatively good results in 2022, which, was full of political and economic events that were difficult to predict in advance. However, there is still a room for improvement in the model, as for export, for example, the quality of forecasting in 2022 turned out to be below the target level for the MAPE indicator (more than 10%).

2. The speed of freight demand estimations process reduces from the current 3 months of marketing experts' work to a maximum of 3–5 days, that saves more than 1,000 man-days or 8,000 man-hours for KTZ MTPD.

3. With this speed of processing of historical data and estimations' generation, there is an opportunity to develop more than three forecasting scenarios (MTPD experts generated only three scenarios - basic, optimistic, and pessimistic; now the number of scenarios is almost unlimited), considering various scenarios of economic sectors' indicators development.

4. There is a real opportunity to reduce the average monthly error in demand estimations from the current average of 36%, to a potentially ambitious 10% or even less. Achieving a MAPE of less than 10% would allow for much more accurate forecasting of both KTZ freight transportation revenues and variable costs, which are dependent on freight transportation performance. Forming more accurate revenue estimations on a monthly basis for the next year allows an optimization of the costs for credit lines and other borrowing instruments used by KTZ.

5. With more accurate freight turnover estimations, KTZ can more accurately calculate the volume of demand and the timing of purchasing and delivery of fuel for diesel locomotives and electricity for electric locomotives. Expecting peaks or increased demand for transportation of this or that type of cargo during certain periods during the next year, KTZ can rationally plan the operating fleet of both locomotives and cars and, if necessary, plan a request for car assistance from neighboring railway administrations.

6. When the advantages of partly automation and MTPD experts' knowledge in the field are combined, serious progress can be made in the speed and quality of demand estimations, and, consequently, in the overall business planning process at KTZ.

5. CONCLUSIONS

The objective of the research was to employ the ARIMA approach for predicting the demand for rail freight transportation services in Kazakhstan using time series analysis, which had not been previously utilized by KTZ. Previously, KTZ relied solely on expert forecasting methods, but their accuracy no longer met management expectations. The study attempted to compare various data analysis methods, including ARIMA, neural net, and auto classifier, using IBM SPSS Modeler software, and to compare these mathematical models against each other, the expert method, and 2017 actual transportation data. For this purpose, a five-year period (from 2012 to 2016 inclusive) was selected and actual historical data on the volume of transportation and freight turnover were loaded into a specialized software product for data analysis and forecasting – IBM SPSS Modeler. Various macroeconomic indicators for the Republic of Kazakhstan and other trading partner countries of Kazakhstan were also loaded into the system. SPSS generated a forecast of transportation volume and freight turnover for all nomenclatures of goods and types of communication for 2017 and the IBM forecast compares with the official actual data of KTZ performance in 2017 and with the estimations made by MTPD experts in 2016 for 2017. Thus, forecasts compared with actual data for all cargo types and all types of communication in terms of tonnes of freight transportation volume and tonne-kilometres of freight turnover. The comparison revealed that demand estimation using ARIMA model showed quite comparable or even better results than three months manual process results of MTPD experts. Results that are more accurate were obtained for those nomenclatures of goods and types of communication, for which there were holistic data sets without gaps loaded to SPSS, and for which it was possible to find macro indicators correlating with the volume of transportation and freight turnover in the past. Using the aforementioned ARIMA model, the article presents the results of forecasting demand for freight transportation volume in 2022 and comparing the forecast with the fact based on MAPE. The forecast for 2022 for all types of rail traffic communication showed good forecast quality results, with the exception of exports. While export results may have been affected by political factors in 2022, a further investigation into the possible reasons for the deviation of the forecast from an acceptable quality range will nevertheless be carried out.

Brief conclusions from the research are as follows:

1. Modern mathematical and statistical models and software should be introduced into the practice of the largest enterprises in Kazakhstan, including the transport industry.

2. The experiment has shown the potential of using the above methods in practice for data sets analysis, forecasting and allows to save up to 8,000 man-hours annually for MTPD experts. The saved man-hours allow MTPD experts to focus not on manual processing of data, but on the analysis interpretation.

3. The experiment conducted on real historical data; a comparative analysis of the results was performed for all cargo types and types of communication used in KTZ practice together with the MTPD experts and presented to the KTZ top management. The results of the pilot experiment on the use of specialized forecasting software served as the basis for launching the "Integrated Planning System" project.

Thus, the methodology discussed in this article, the results of the scientific experiment and recommendations based on the results of the experiment have been put into practice in KTZ.

References

1. Punia, S. & Shankar, S. Predictive analytics for demand forecasting: A deep learning-based decision support system. *Knowledge-Based Systems.* 2022. Vol. 258. No. 109956. DOI: 10.1016/j.knosys.2022.109956.
2. Merkuryeva, G. & Valberga, A. & Smirnov, A. Demand forecasting in pharmaceutical supply chains: A case study. *Procedia Computer Science.* 2019. Vol. 149. P. 3-10. DOI: 10.1016/j.procs.2019.01.100.
3. Milenković, M. & Švadlenka, L. & Melichar, V. & Bojović, N. & Avramović, Z. SARIMA modelling approach for railway passenger flow forecasting. *Transport.* 2015. Vol. 33. No. 2. P. 1-8. DOI: 10.3846/16484142.2016.1139623.
4. Roos, J. & Gavin, G. & Bonnevay, S. A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data. *Transportation Research Procedia.* 2017. No. 26. P. 53-61. DOI: 10.1016/j.trpro.2017.07.008.
5. Zhang, J. & Chen, F. & Shen, Q. Cluster-based LSTM network for short-term passenger flow forecasting in urban rail transit. *IEEE Access*. 2019. Vol. 7. P. 147653-147671. DOI: 10.1109/ACCESS.2019.2941987.
6. Tang, Q. & Cheng, P. & Li, N. 基于GSO-BPNN方法的城市轨道交通客流短时预测[J]. 2017. 通信领域的技术与经济(1): 1–4. [In Chinese: Tang, Q., Cheng, P., Li, N. Short time forecasting of passenger flow in urban railway using GSO-BPNN method. *Technology & Economy in Areas of Communications.* 2017].
7. Намиот, Д.Е. & Покусаев, О.Н. & Лазуткина, В.С. О моделях пассажирского потока для городских железных дорог. *International Journal of Open Information Technologies*. 2018. Vol. 6. No. 3. P. 9-14. [In Russian: Namiot, D.E. & Pokusaev, O.N. & Lazutkina, V.S. On passenger flow models for urban railways. *International Journal of Open Information Technologies*. 2018].
8. Andersson, M. & Brundell-Freij, K. & Eliasson, J. Validation of aggregate reference forecasts for passenger transport. *Transportation Research Part A: Policy and Practice.* 2017. Vol. 96. P. 101-118. DOI: 10.1016/j.tra.2016.12.008.

9. Prakaulya, V. & Sharma, R. & Singh, U. & Itare, R. 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. P. 554-558.
10. Markovits-Somogyi, R. Measuring efficiency in transport: the state of the art of applying data envelopment analysis. *Transport*. 2011. Vol. 26. No. 1. P. 11-19. DOI: 10.3846/16484142.2011.555500.
11. Banerjee, N. & Morton, A. & Akartunalı, K. Passenger demand forecasting in scheduled transportation. *European Journal of Operational Research*. 2020. Vol. 286. No. 3. P. 797-810. DOI: 10.1016/j.ejor.2019.10.032.
12. Borucka, A. & Mazurkiewicz, D. & Lagowska, E. Mathematical modelling as an element of planning rail transport strategies. *Transport*. 2021. Vol. 36. No. 4. P. 354-363. DOI: 10.3846/transport.2021.16043.
13. *Downloading IBM SPSS Modeler 18.0*. Available at: https://www.ibm.com/support/pages/downloading-ibm-spss-modeler-180.
14. Bermúdez, J.D. & Segura, J.V. & Vercher, E. Holt–Winters forecasting: an alternative formulation applied to UK air passenger data. *Journal of Applied Statistics*. 2007. Vol. 34. No. 9. P. 1075-1090.
15. *The Neural Networks Model*. Available at: https://www.ibm.com/docs/en/spss-modeler/18.0.0?topic=networks-neural-model#neuralnet_model.
16. *ARIMA*. Available at: https://www.ibm.com/docs/en/spss-modeler/18.0.0?topic=series-arima.
17. *Auto Classifier node*. Available at: https://www.ibm.com/docs/en/cloud-paks/cp-data/4.6.x?topic=modeling-auto-classifier-node.