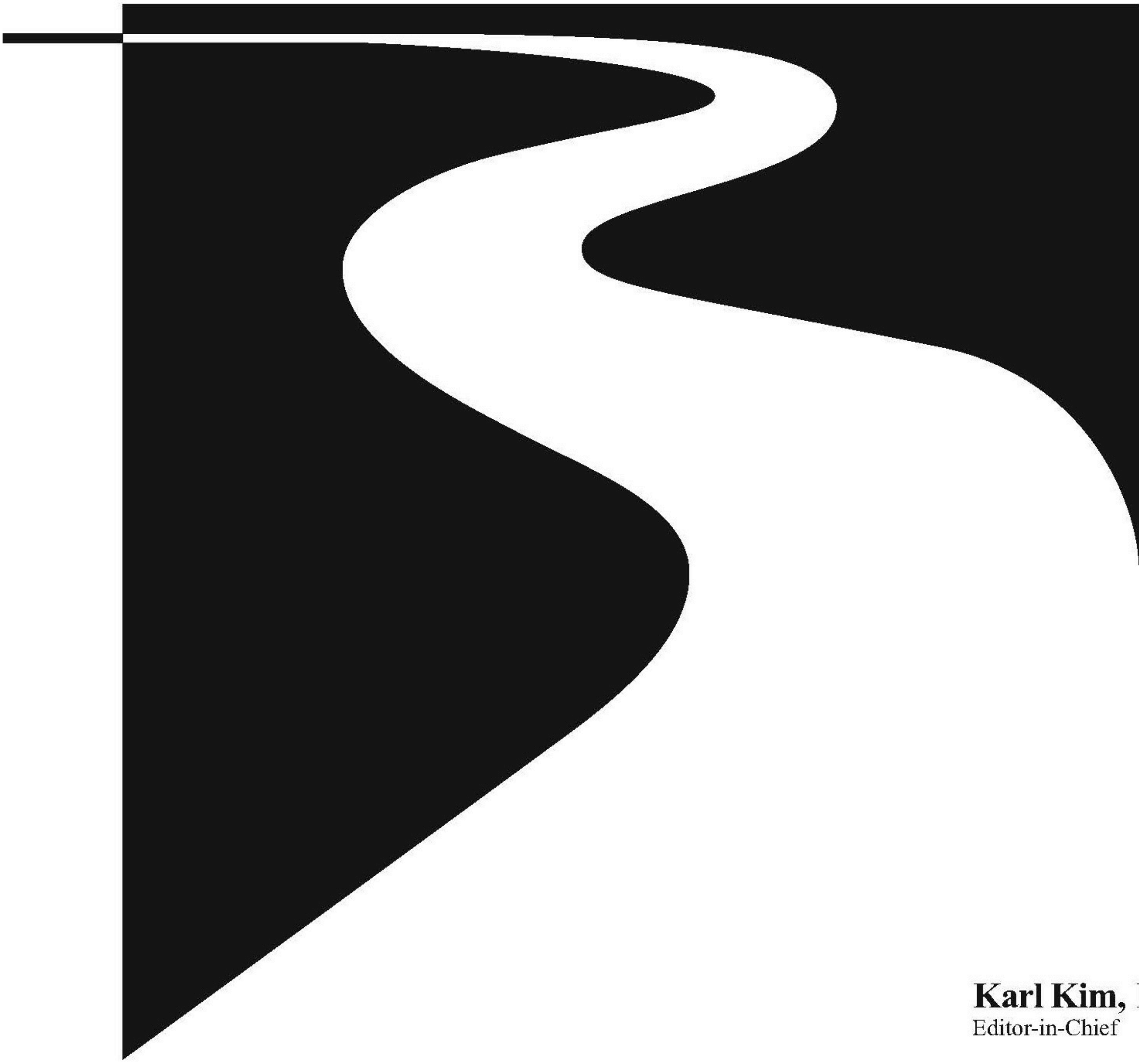
Volume 23, January 2024
ISSN: 2590-1982

TRANSPORTATION RESEARCH INTERDISCIPLINARY PERSPECTIVES



TRANSPORTATION RESEARCH INTERDISCIPLINARY PERSPECTIVES

Contents

Volume 23, January 2024

(Contents continued on inside back cover)

ELSEVIER

ISSN 2590-1982
23 (2024)

(Contents continued from outside back cover)

Transportation Research Interdisciplinary Perspectives 23 (2024) 101028

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives

Forecasting the demand for railway freight transportation in Kazakhstan: A case study

Madiyar Sultanbek [a,*,1], Nazdana Adilova [a,2], Aleksander Sładkowski [b,3], Arnur Karibayev [c,4]

[a] Al-Farabi KazNU, Al-Farabi Av. 71, 050040 Almaty, Kazakhstan
[b] Silesian University of Technology, Krasińskiego 8, 40-019 Katowice, Poland
[c] JSV "National Company "Kazakhstan Temir Zholy", Kunayev 6, 010000 Astana, Kazakhstan

ARTICLE INFO

Keywords:
Demand forecasting
Railway freight transportation
Volume and turnover
ARIMA
Forecast quality assessment

ABSTRACT

Effective resource planning in railway freight transportation necessitates precise demand forecasting, shaped by a complex array of dynamic variables. In Kazakhstan, the National State-owned Railway Company (KTZ) underwent a transition from traditional expert-based forecasting methods to the application of mathematical models, notably the Autoregressive Integrated Moving Average (ARIMA) model. Employing historical data spanning from 2012 to 2016, this shift signifies a significant evolution in KTZ's approach to anticipating freight demand.

The primary objective of this study is the empirical assessment of the efficacy of the ARIMA model in comparison to conventional qualitative forecasting techniques. Through the analysis of actual data from 2017, utilizing established metrics such as the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE), this research substantiates the utility of time series analysis through ARIMA. The findings not only confirm the model's effectiveness but also emphasize its superiority in refining the precision of railway freight demand forecasts, particularly within the unique context of Kazakhstan.

Beyond the validation of methodologies, this research serves as a precursor to advanced forecasting practices, offering the potential to redefine resource planning in the railway industry. By extending the predictive horizon to 2024, the manuscript aligns with contemporary standards, providing nuanced insights for operational and developmental considerations in Kazakhstan's railway freight sector. This expansion positions the study within the evolving landscape of the industry, ensuring a comprehensive and forward-looking contribution to efficient resource allocation and modernized planning practices.

1. Introduction

Accurate and high-quality demand estimations are indispensable for effective planning and decision-making within any organization. These estimations serve as a primary input for multiple departments, including finance, marketing, distribution, and production, influencing various decision-making processes. Given their critical role in business decision-making, ensuring the precision and quality of these forecasts is of utmost importance (Punia and Shankar, 2022).

Forecasting demand is a crucial aspect of managing business operations. While the methods for forecasting may vary across different businesses and can be quite complex, the ultimate objective remains consistent: obtaining a reasonably accurate prediction of future product demand by leveraging historical data and various environmental factors (such as political, social, and economic conditions) to plan and coordinate business operations effectively (Merkuryeva et al., 2019).

Predicting the future demand for transportation services is a vital 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,

* Corresponding author.
E-mail addresses: cultanbek_madiyar@live.kaznu.kz (M. Sultanbek), adilova.nazdana@kaznu.kz (N. Adilova), aleksander.sladkowski@polsl.pl (A. Sładkowski), karibayev_ar@railways.kz (A. Karibayev).
[1] orcid.org/0000-0002-4408-3552.
[2] orcid.org/0000-0002-8184-6993.
[3] orcid.org/0000-0002-1041-4309.
[4] orcid.org/0009-0006-7296-3502.

https://doi.org/10.1016/j.trip.2024.101028
Received 5 June 2023; Received in revised form 20 January 2024; Accepted 27 January 2024

and finance.

The "National Company Kazakhstan Temir Zholy" joint-stock company (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 only shareholder of "Samruk-Kazyna" JSC is, in turn, 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 KTZ operate in the segments "Main railway network services," "Rail freight transportation," "Passenger rail transportation," and "Freight cars operations."

KTZ's primary sources of income come from freight and passenger transportation. Income from freight transportation accounts for 86 % of the total income and encompasses all components that support transportation activities, including main railway network services, locomotive traction services, freight commercial work, and the provision of freight cars.

The railway network has an unfolded length of more than 21,000 km, with approximately 54,000 freight cars, over 2,000 passenger cars, and more than 1.6 thousand locomotives. KTZ is the country's largest employer, with over 115,000 employees.

In freight railway transport in Kazakhstan, the volume of applications closely matches the volume of traffic. Therefore, the volume of cargo transportation in tons assesses the demand for the services of a railway carrier. Additionally, the cargo transportation volume in tons, multiplied by the distance transported in kilometers, known as freight turnover, plays a significant role in calculating future revenues from freight traffic or income from the primary activity of freight transportation services.

Forecasting techniques can be categorized into two main approaches: quantitative and qualitative. Within these approaches, various methods have been developed. Qualitative methods, including the Delphi technique, Executive opinions, Customer services, and Sales force polling, generate forecasts based on judgments or opinions. On the other hand, quantitative methods can be further classified into historical data forecasts and associative forecasts. Historical data forecasts involve techniques like Time Series Analysis, Trend Analysis, the Naive method, Holt's and Winter's models. In contrast, associative forecasts determine causal relationships between variables using Symbolic, Multiple, or Simple regression. Additionally, there are mixed or combined models that integrate both qualitative and quantitative approaches.

Numerous researchers have explored the realm of modeling demand for rail services, each offering distinct innovations and contributions when compared to the existing body of references. For instance, Serbian railways employed the seasonal ARIMA model to predict monthly passenger flows, as demonstrated by Milenković et al. (2015). Meanwhile, another study (Roos et al., 2017) investigated the use of dynamic Bayesian networks to forecast short-term passenger flows within the urban rail network of Paris. Study (Zhang et al., 2019) leveraged a long short-term memory network to analyze the performance of Beijing's urban rail transit network, and study (Tang et al., 2017) introduced a method that amalgamated a backpropagation neural network with the glow-worm swarm optimization algorithm for analysis. On the other hand, НаМиот et al. (2018) detailed methods for forecasting Moscow passenger traffic, rooted in network topology analysis.

However, it is worth noting that most of these studies have a common focus on cities with dynamic development and an escalating demand for rail transport, typically attributed to urban expansion and worsening road transport conditions. Consequently, fewer models have been developed to assess the functioning of railway networks on a national scale, as exemplified in Sweden (Andersson et al., 2017) and India (Prakaulya et al., 2017).

Additionally, some research has centered on evaluating the effectiveness of decision-making units in various organizations, utilizing the Data Envelopment Analysis (DEA) tool. Study Markovits-Somogyi (2011) comprehensively reviewed these DEA applications, collecting and comparing data from 69 published cases. The researcher examined attributes, areas of utilization, as well as inputs and outputs used for the analysis. Nevertheless, it is important to acknowledge that DEA has a significant drawback, given its sensitivity to measurement errors and data noise. Surprisingly, limited research exists that conducts comparative assessments of multiple modeling approaches to identify the most effective one, as can be observed in Study Banerjee et al. (2020), which offers a range of models for predicting demand for passenger transportation services.

Estimating and forecasting demand is of paramount importance for making well-informed decisions. Unfortunately, the availability of suitable mathematical models for demand estimation and forecasting remains a challenge, hindering access to reliable demand data. Hence, it is imperative to conduct a comprehensive analysis of railway systems across various countries to identify appropriate forecasting methods. These can serve as a valuable scientific database for research conducted in other countries or within different transport systems.

The primary objective of this study is to pinpoint the most suitable mathematical model from an array of options for accurately predicting the future demand for railway freight transportation services in Kazakhstan. Historical data from the Kazakhstan National State Railway Company serves as the cornerstone for this endeavor. The article introduces and assesses multiple models, culminating in the identification of the optimal one. The paper encompasses an introductory section, a segment detailing the study's data and methodology, sections delving into the mathematical models developed through the employed techniques, empirical data analysis, and a concluding summary presenting the results obtained, conclusions drawn, and potential directions for future research. Importantly, the research has extended its predictions to 2024, providing a forward-looking perspective and enabling the provision of relevant operational or developmental suggestions based on the forecasted railway freight volume in the country.

2. Data and methods

Current planning process in KTZ lacks of any modern tool. The paper Borucka et al. (2021) employs various techniques such as seasonal exponential smoothing model (ETS), naive model, ARMA (Auto Regressive Moving Average) errors, exponential smoothing state-space model with Box-Cox transformation, 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.

The information provided above leads to the following deductions:

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 cargoes (further - UTSNC) 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 UTSNC (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 (Bermúdez et al., 2007).

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 (autoregessive integrated moving average) - ARIMA is amongst the most generally used to forecast transport demand (Bermúdez et al., 2007; Milenković et al., 2015); 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. The neural network depicted in Fig. 2 has the capability to approximate a wide range of predictive models without relying heavily on specific structures or assumptions. Its learning process determines the relationships between the target and predictors. If a linear relationship is suitable, the neural network's outcomes 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.

While the neural network offers modeling flexibility, it lacks interpretability, which makes it difficult to explain the underlying process that generates the relationship between the target and predictors. In such cases, a traditional statistical model would be more suitable for interpretation. However, if interpretability is not crucial, the neural network can provide accurate predictions.

To construct a neural network model, at least one target and one input field are required. There are no restrictions on the measurement levels of targets or predictors (inputs). During model construction, the initial weights assigned to the neural network and the final models generated depend on the field order in the data. However, Cloud Pak for Data sorts the data by field name before presenting it to the neural network for training. Therefore, explicitly changing the field order upstream will not affect the generated neural network models when a random seed is set in the model builder. However, modifying the input field names in a way that alters their sort order will produce different neural network models, even with a random seed set in the model builder. Nevertheless, changing the field names' sort order will not significantly affect the quality of the model.

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.

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.

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, and were loaded into SPSS. For each aggregated nomenclature of goods combinations were selected - Unified Tariff and Statistical Nomenclature of Cargoes (UTSNC) 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.

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.

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 %.

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.

Using the selected ARIMA method, in 2023, a forecast was formed for the volume of traffic in thousands of tons for 2024 based on five years' worth of historical data.

KTZ has consistently employed the ARIMA methodology for forecasting forthcoming cargo transportation volumes over an extended period. Illustrated in Figs. 6 and 7 are the actual monthly transportation volumes throughout 2022 and 2023 for diverse cargo types, alongside transportation forecasts extending into 2024. Fig. 8 encapsulates the comprehensive forecast of transportation volumes for all cargo types and directions, encompassing export, import, transit, and domestic categories. The discernible upward trajectory in total transportation volumes across the KTZ network implies escalating loads on the primary network, potentially resulting in diminished transportation speed and constraints in specific network segments.

A Fig. 9 delineates the actual transit traffic volumes for all cargo types, inclusive of both freight cars and containers, exhibiting an evident linear trend indicative of sustained growth. Nevertheless, the existing capacities and transshipment sites at Kazakhstan's borders with key trading partners such as China, Russia, and Uzbekistan have neared saturation, necessitating investments in infrastructure development to accommodate burgeoning demand. Utilizing mathematical models,

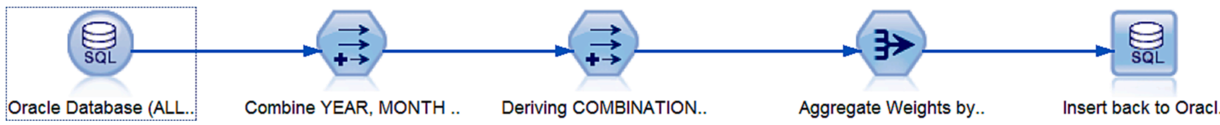


Fig. 1. Data aggregation algorithm in IBM SPSS Modeler.

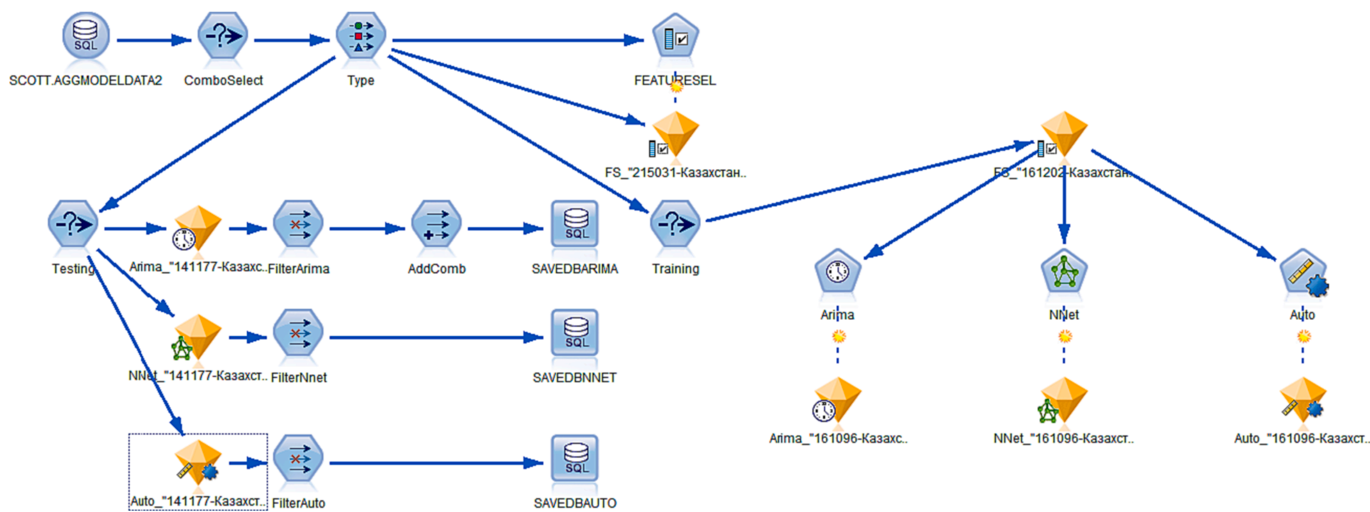


Fig. 2. Algorithm in SPSS using three different models (ARIMA, Nnet, Auto).

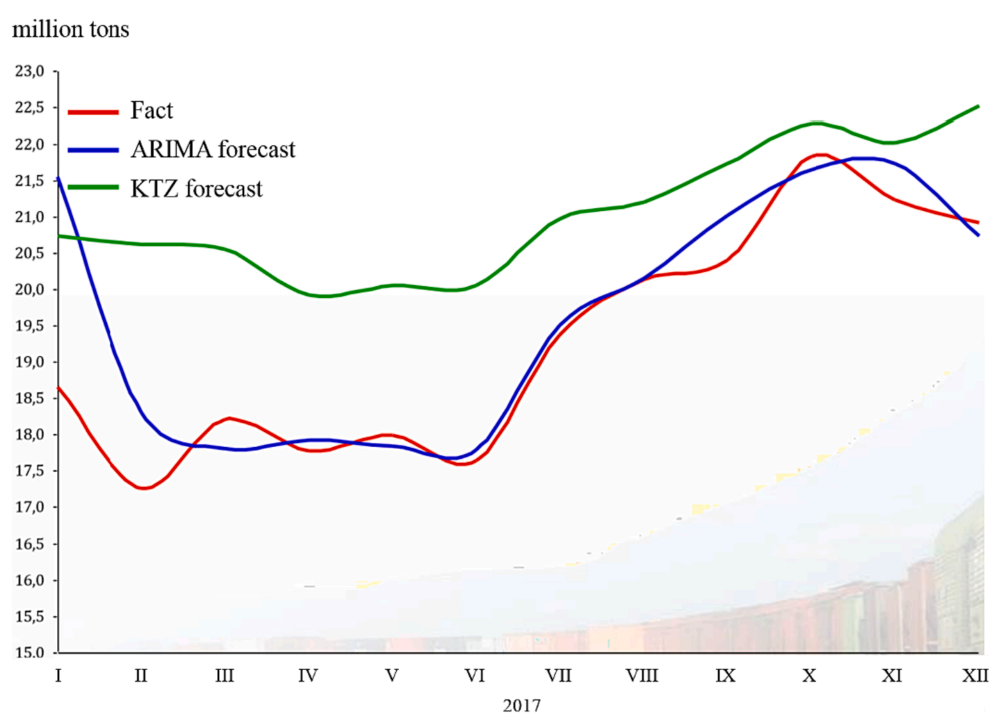


Fig. 3. Aggregate the assessment of rail freight volume across all cargo categories and communication methods in contrast to the 2017 data and the estimations provided by experts from MTPD KTZ.

notably ARIMA, facilitates the formulation of more realistic feasibility studies for infrastructure investments. These traffic volume forecasts have been instrumental in instigating pivotal infrastructure projects, including the construction of second tracks on the Dostyk-Moyynty section spanning 836 km. This initiative aims to significantly augment capacity at the China border.

A substantial undertaking involves the construction of a bypass line circumventing Almaty, presenting the opportunity to redirect transit freight traffic along a more efficient route, enhance traffic speed, and alleviate congestion at Almaty-1 station by an estimated 40 %. Additionally, the Darbaza – Maktaaral railway line project aims to mitigate congestion at Saryagash station and the Saryagash - Tashkent section, responding to the nearing operational limits of the existing infrastructure. The envisioned Darbaza – Maktaaral line will facilitate the

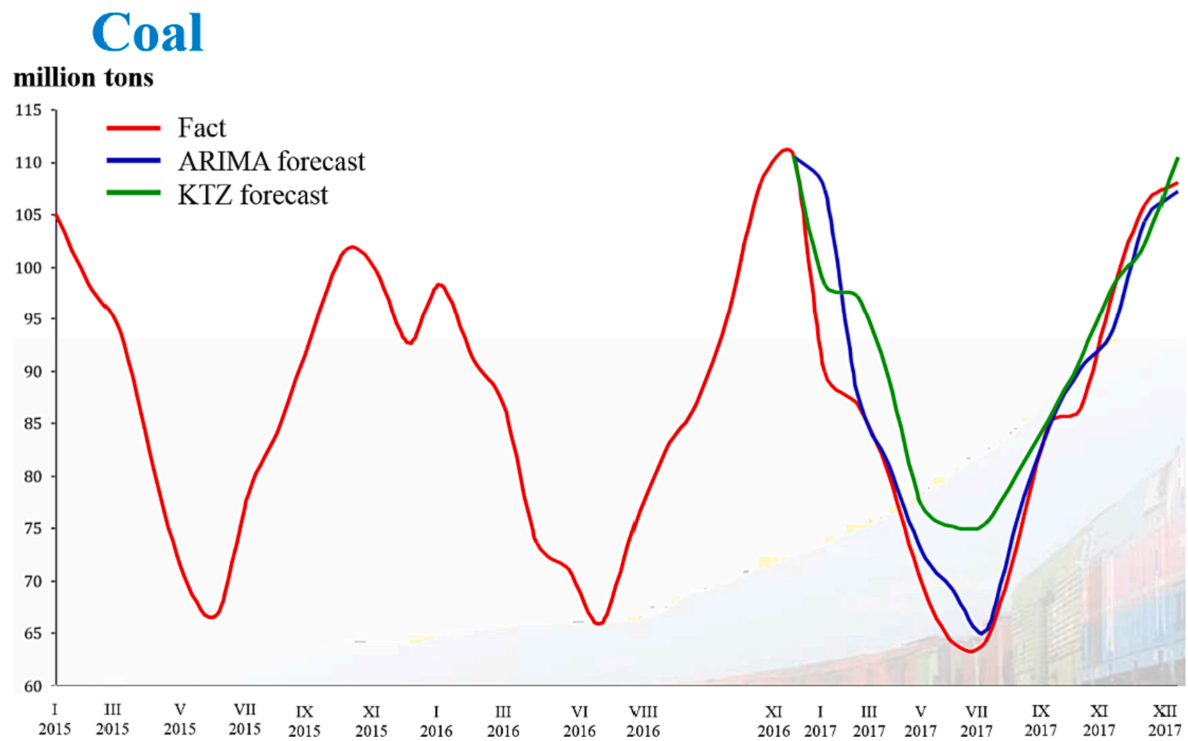


Fig. 4. 2017 Freight Transportation Volume: A Comparative Analysis of Actual Data, MTPD Expert Estimates, and ARIMA Model Forecasts for Total Coal Transportation Across Various Communication Modes (Internal, Export, Import, and Transit).

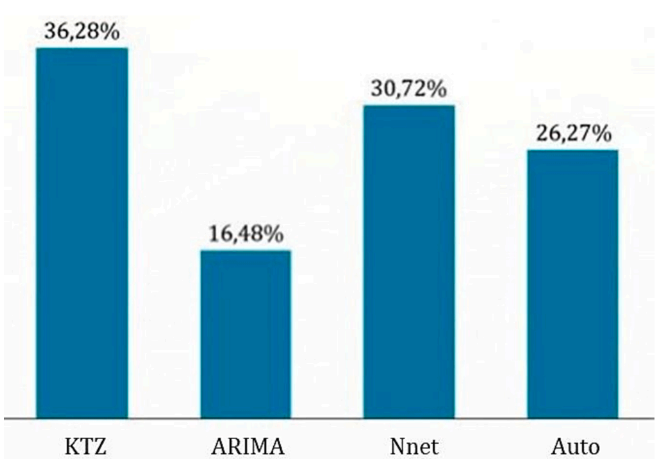


Fig. 5. The MAPE indicator across various goods nomenclatures, forecast types (volume and cargo turnover), and communication methods, employing a range of forecasting techniques within IBM SPSS Modeler.

redistribution of cargo traffic with Uzbekistan.

Another pivotal initiative encompasses the construction of a novel railway line, Bakhty – Ayagoz, spanning 272 km, designed to augment the transport and transit potential of Kazakhstan. Concurrently, the project involves the establishment of a third border crossing with China, Bakhty - Chuguchak, anticipated to double capacity between Kazakhstan and China. This strategic border crossing is poised to alleviate congestion at existing southern checkpoints and attract additional transit volumes. The multifaceted approach underscores KTZ's commitment to strategically leverage forecasting methodologies, specifically ARIMA, to inform and propel impactful infrastructure developments, ensuring alignment with the evolving demands of the cargo transportation landscape.

Ensuring the precise forecasting of transportation needs is paramount for optimizing the operational efficiency of the main railway network within KTZ. The efficient routing of train flows is a critical factor in minimizing operational costs and maintaining high service standards in the transportation of goods and passengers. Given the substantial role of KTZ's main railway network in connecting diverse regions and facilitating nationwide transportation, the development of a comprehensive methodology for accurate cost computation becomes essential. This methodology will serve as a crucial guide in decision-making processes, allowing for the strategic selection of cost-effective routing strategies to enhance overall operational efficiency.

4. Discussion

This scientific paper revolves around the investigation of demand forecasting in the transportation sector, with a specific focus on the case study of the joint stock company KTZ, which specializes in rail transportation and logistics. The primary objective of this study is to underscore the pivotal role of demand predictions in the efficient planning and decision-making processes within transportation firms, particularly for coordinating transportation operations, marketing strategies, and financial considerations. To achieve this, the study harnesses historical data and macroeconomic indicators to construct forecasting models specifically tailored to rail freight transportation. A variety of forecasting techniques are employed, including both qualitative methods such as expert opinions and extrapolation, and quantitative methods like ARIMA and neural network models. The results of this investigation indicate that the ARIMA model outperforms other forecasting methods, notably surpassing the expert-based estimation method currently employed by MTPD experts at KTZ.

A major revelation emerging from this study is that the ARIMA model furnishes more precise demand forecasts in comparison to the traditional expert-based approach. This superior accuracy is attributed to the ARIMA model's capacity to account for fluctuations in the external environment and the influence of external factors, leading to an enhancement in forecasting precision. Moreover, the ARIMA model's predictive capability is further enriched by the correlation between

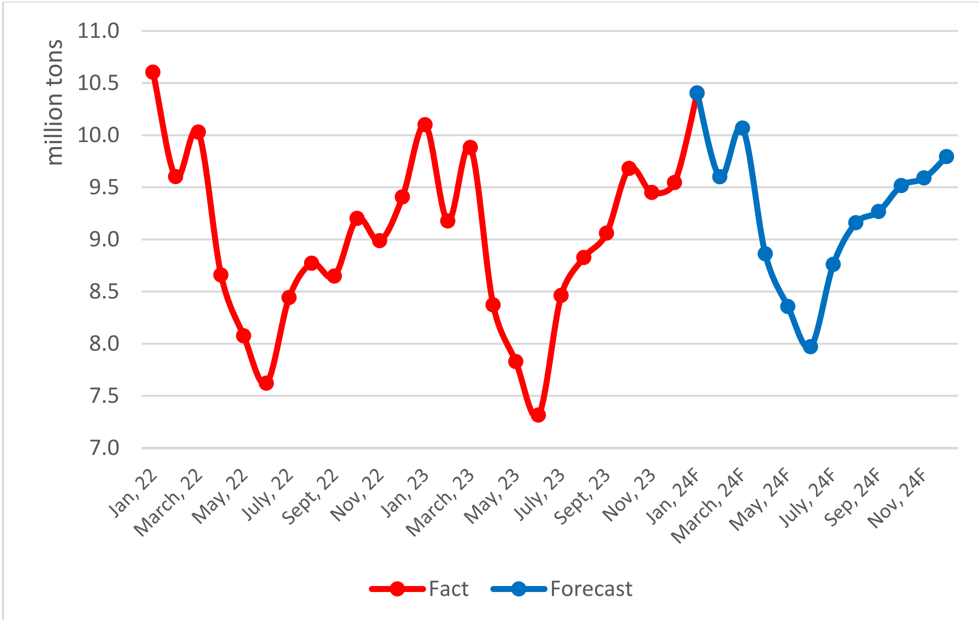


Fig. 6. Monthly trends in coal transportation volumes: actual data for 2022–2023 and forecast for 2024.

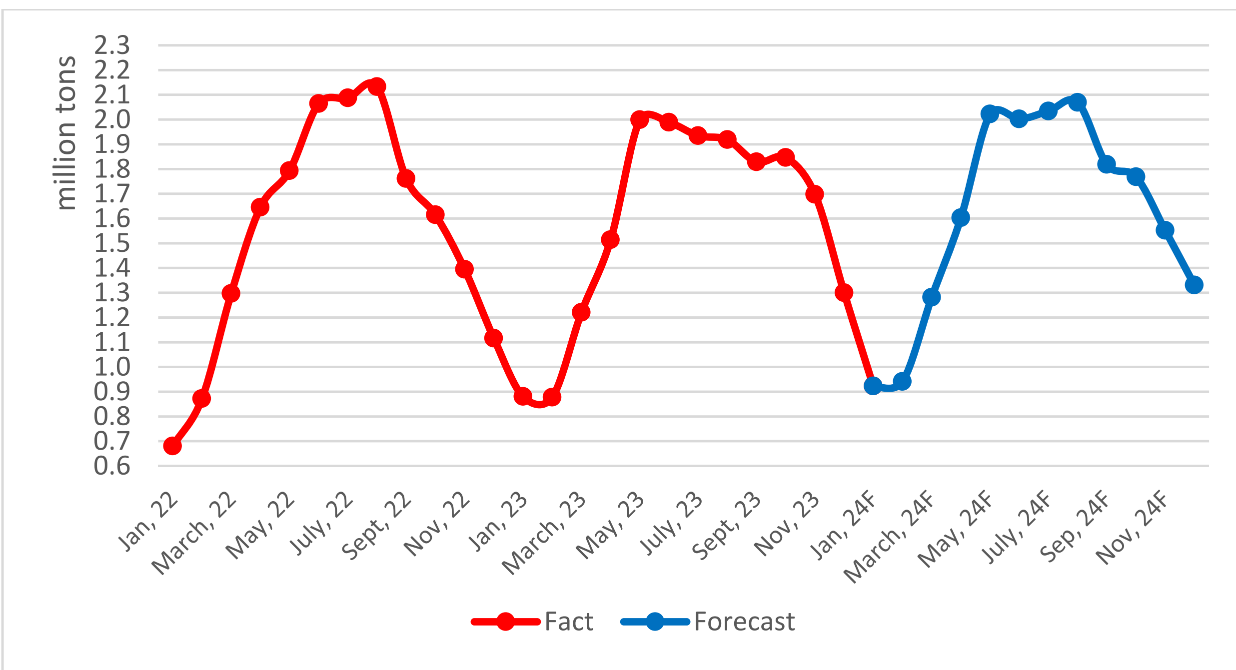


Fig. 7. Monthly trends in construction cargo transportation volumes: actual data for 2022–2023 and forecast for 2024.

transportation volumes and macroeconomic indicators.

This paper also draws attention to the constraints inherent in the current manual approach used by MTPD experts, primarily stemming from a heavy reliance on routine operations and the limited data processing capabilities of tools like MS Excel. As a remedy, the proposal is to automate parts of the demand estimation process through modern software, with the aim of augmenting the efficiency of marketing experts. This automated approach promises faster data analysis, greater flexibility in using diverse forecasting methods, and a reduction in the time allocated to routine tasks, ultimately culminating in improved demand estimation quality.

Furthermore, the study acknowledges certain aspects that require further refinement and examination. For instance, the ARIMA model exhibits a slight tendency to overestimate demand for domestic transportation and exports while underestimating demand for imports and transit. These patterns may be influenced by specific factors, and additional research is deemed necessary to fine-tune the models and account for these variances.

In summation, this study imparts invaluable insights into the demand forecasting process within the transportation industry, particularly in the domain of rail freight transportation. By showcasing the efficacy of the ARIMA model, this research underscores the potential advantages of integrating modern software tools to enhance the precision of demand estimations. The findings not only provide a guiding

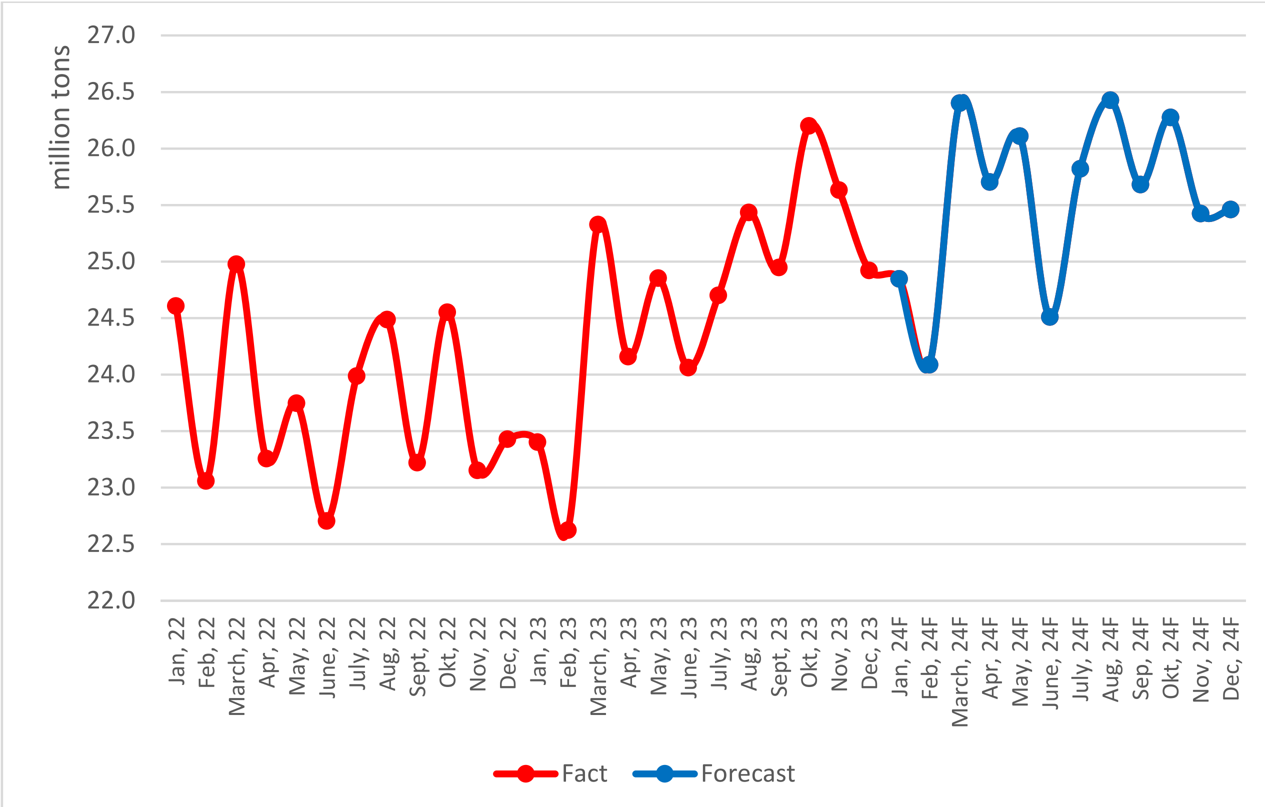


Fig. 8. ARIMA forecast of transportation volumes for all cargo types and directions (export, import, transit, and domestic) for 2024.

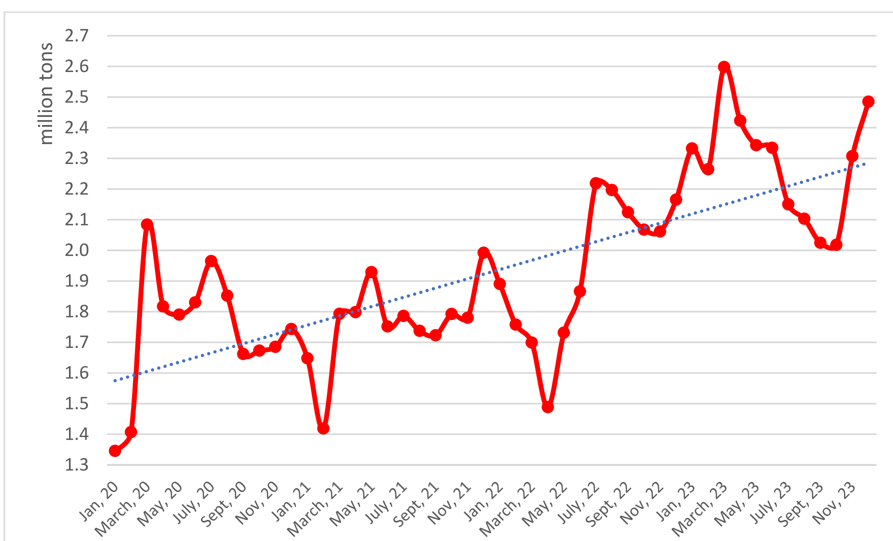


Fig. 9. Actual transit traffic volumes for all cargo types, including freight cars and containers, demonstrating a sustained linear growth trend.

framework for transportation companies, including KTZ, to refine their demand forecasting methodologies but also present an avenue to augment their planning and decision-making processes. The methodical structure of this paper facilitates a more comprehensive understanding and practical application of the discussed concepts for researchers and professionals operating within the field. Importantly, the extension of predictions to 2024 and the subsequent provision of operational and developmental suggestions based on forecasted railway freight volume further enhance the relevance and applicability of the study in navigating the evolving landscape of the transportation industry.

5. Conclusions

The primary aim of this research was to employ the ARIMA approach for predicting the demand for rail freight transportation services in Kazakhstan through time series analysis, a methodology not previously utilized by KTZ. In contrast to the reliance on expert forecasting methods, which had become outdated and failed to meet management expectations, this study sought to enhance the precision of demand estimation. The comparison of various data analysis methods, including ARIMA, neural net, and auto classifier, using IBM SPSS Modeler software, demonstrated the superior performance of the ARIMA model. The study spanned a five-year period (2012 to 2016 inclusive), leveraging actual historical data on transportation volume and freight turnover loaded into the specialized IBM SPSS Modeler. This comparison, encompassing all cargo types and communication channels, revealed that the demand estimation using the ARIMA model yielded comparable or even superior results than the manual process results of MTPD experts.

The application of the ARIMA model facilitated the forecasting of demand for freight transportation volume, extending predictions to 2024. This forward-looking perspective, based on forecasted railway freight volume, enabled the formulation of relevant operational and developmental suggestions for railway freight, presenting a comprehensive and strategic approach for KTZ.

In light of the research findings, several key conclusions emerge:

1. Integration of Modern Models and Software. Modern mathematical and statistical models, alongside advanced software, should be seamlessly integrated into the operational practices of major enterprises in Kazakhstan, with a specific emphasis on the transport industry. This integration will not only enhance efficiency but also contribute to more accurate and reliable demand forecasting.
2. Potential for Time and Resource Savings. The experimentation with various methods, including ARIMA, showcased the potential for substantial time and resource savings. The adoption of advanced methodologies and software tools for data analysis and forecasting could save up to 8,000 man-hours annually for MTPD experts. This efficiency gain allows experts to shift their focus from manual data processing to more nuanced analysis and interpretation.
3. Practical Implementation. The experimental approach, conducted on real historical data, facilitated a comprehensive comparative analysis of results for all cargo types and communication channels used in KTZ practice. The collaboration with MTPD experts and presentation of the findings to KTZ top management laid the groundwork for the launch of the "Integrated Planning System" project. The practical implementation of the methodology discussed in this article, along with the recommendations derived from the research results, has proven successful within the operational framework of KTZ.

In essence, this research not only contributes to the refinement of demand forecasting methodologies within the transportation sector but also underscores the transformative impact of advanced models and software tools. The incorporation of predictions for 2024 and the subsequent provision of relevant operational and developmental suggestions reinforce the practical relevance and applicability of the study, providing valuable insights for KTZ and other enterprises navigating the complex landscape of railway freight transportation.

Complementary data

The study is funded by grant #AP14871940 by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan.

CRediT authorship contribution statement

Madiyar Sultanbek: Data curation, Methodology, Writing – original draft, Writing – review & editing. **Nazdana Adilova:** Funding acquisition, Project administration, Resources. **Aleksander Sładkowski:** Data curation, Software, Supervision, Validation. **Arnur Karibayev:** Investigation, Software, Validation, Visualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nazdana Adilova reports financial support was provided by Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan. Madiyar Sultanbek reports a relationship with Kazakhstan National Railways Company that includes: employment. Arnur Karibayev reports a relationship with Kazakhstan National Railways Company that includes: employment.

Data availability

The data that has been used is confidential.

References

Andersson, M., Brundell-Freij, K., Eliasson, J., 2017. Validation of aggregate reference forecasts for passenger transport. Transp. Res. A Policy Pract. 96, 101–118. https://doi.org/10.1016/j.tra.2016.12.008.

Banerjee, N., Morton, A., Akartunalı, K., 2020. Passenger demand forecasting in scheduled transportation. Eur. J. Oper. Res. 286 (3), 797–810. https://doi.org/10.1016/j.ejor.2019.10.032.

Bermúdez, J.D., Segura, J.V., Vercher, E., 2007. Holt-winters forecasting: an alternative formulation applied to UK air passenger data. J. Appl. Stat. 34 (9).

Borucka, A., Mazurkiewicz, D., Lagowska, E., 2021. Mathematical modelling as an element of planning rail transport strategies. Transport 36 (4), 354–363. https://doi.org/10.3846/transport.2021.16043.

НаМиот, Д.Е., Покусаев, О.Н., Лазуткина, В.С. 2018. 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. *Int. J. Open Inf. Technol.*].

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.

Merkuryeva, G., Valberga, A., Smirnov, A., 2019. Demand forecasting in pharmaceutical supply chains: A case study. Procedia Comput. Sci. 149, 3–10. https://doi.org/10.1016/j.procs.2019.01.100.

Milenković, M., Švadlenka, L., Melichar, V., Bojović, N., Avramović, Z., 2015. SARIMA modelling approach for railway passenger flow forecasting. Transport 33 (2), 1–8. https://doi.org/10.3846/16484142.2016.1139623.

Prakaulya, V., Sharma, R., Singh, U., Itare, R., 2017. Railway passenger forecasting using time series decomposition model. In: International conference of Electronics, Communication and Aerospace Technology (ICECA), 20–22 April 2017, Coimbatore, India, pp. 554–558.

Punia, S., Shankar, S., 2022. Predictive analytics for demand forecasting: A deep learning-based decision support system. Knowl.-Based Syst. 258, 109956. https://doi.org/10.1016/j.knosys.2022.109956.

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

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].

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