applied sciences

MDPI

Article

Multi-Indicator Forecasting of Road Freight Transport Workload for Operational Planning

Jakub Konwerski * and Jarosław Ziółkowski

Faculty of Civil Engineering and Geodesy, Military University of Technology, gen. S. Kaliskiego 2, 00-908 Warsaw, Poland; jaroslaw.ziolkowski@wat.edu.pl

* Correspondence: jakub.konwerski@wat.edu.pl

Abstract

This article presents a multi-indicator approach to forecasting the monthly workload of a military road freight transport system in support of operational planning. The empirical basis of the study consisted of real-world operational data from 2020 to 2025, aggregated into regular monthly time series. Four complementary workload indicators were analysed: the number of transport tasks, the number of vehicles assigned to task execution, the mass of transported cargo, and transport work expressed in tonne-kilometres. The research procedure comprised data preprocessing, indicator construction, seasonality analysis, time-series decomposition, comparison of classical forecasting models, and assessment of forecast uncertainty using prediction intervals. The forecasting models considered included the naive model, the moving-average model, Brown's and Holt's exponential smoothing models, ETS, and ARIMA. Model performance was evaluated using a rolling-origin validation procedure with an expanding training window, based on MAE, RMSE, MAPE, MASE, and Bias metrics. The results showed that the recommended model depends on the forecasted workload dimension: Brown's model performed best for the number of transport tasks, ETS for the number of vehicles and transport work, whereas the 12-month moving-average model was most effective for transported cargo mass. All recommended models achieved MASE values below 1, indicating improved predictive performance compared with the naive benchmark. The study demonstrated that point forecasts supplemented with 80% and 95% prediction intervals can support monthly planning of fleet resources, transport capacity reserves, and future workload levels. Although the empirical analysis concerns a military transport system operating under peacetime conditions, the proposed framework may be adapted to support monthly workload forecasting and operational planning in other freight transport systems.

Keywords: fleet planning; military transport; road transport; rolling-origin validation; time-series forecasting; tonne-kilometres

Academic Editor: Roland Jachimowski

Received: 28 May 2026
Revised: 16 June 2026
Accepted: 18 June 2026
Published: 26 June 2026

1. Introduction

Forecasting the workload of transport systems is one of the fundamental elements of resource planning, transport organisation, and the assessment of the future operational performance of a vehicle fleet. In road freight transport systems, transport demand is described not solely by the number of orders but also by the required level of vehicle deployment, the mass of cargo transported, and the transport work resulting from the combined effect of cargo mass and transport distance. The importance of forecasting becomes particularly evident when transport activity is subject to monthly, seasonal, and operational variability,

while decisions concerning fleet resources, personnel, maintenance activities and reserve transport capacity must be made in advance. In this context, forecasting should not be regarded merely as an outcome of statistical modelling, but as a decision-support tool for transport system management [1].

In the present study, the number of transport tasks is treated as a basic organisational indicator of system workload and is defined as the number of transport operations or orders assigned to a given period. This indicator reflects the organisational level of system activity, as each operation requires planning, coordination, preparation of documentation, allocation of a vehicle and driver, and supervision of task execution. In the literature on freight demand modelling, the number of deliveries, shipments or freight movements is often considered a separate object of forecasting and operational decision-making. Nuzzolo and Comi [2] emphasised the need to jointly consider the quantitative, delivery-related, and vehicle-related dimensions when forecasting urban freight transport. Similarly, Al Hajj Hassan et al. [3] used forecasts of freight movement volumes to support operational decisions for an intermodal carrier, while Grzelak et al. [4] demonstrated that forecasting the number of shipments handled by a logistics operator may support workforce and vehicle planning. Recent research on European road freight operations has likewise treated freight demand volume as a distinct forecasting target and compared econometric and artificial intelligence methods for operational use [5]. However, the number of transport tasks alone does not provide a complete characterisation of future transport workload, since a similar number of tasks may correspond to different numbers of vehicles, cargo masses, route lengths, and volumes of transport work performed.

The second planning-relevant indicator is resource demand, expressed as the number of vehicle assignments performing transport task execution. This quantity should not be treated solely as a proportional function of the number of orders because it also depends on cargo structure, route length, vehicle capacity, fleet availability, maintenance and repair cycles, transport organisation, and the required reserve of transport capacity. In fleet planning problems, fleet size and composition are treated as independent strategic, tactical, and operational decision variables. The approaches presented by Golden et al. [6] and Hoff et al. [7] confirm that fleet configuration is one of the key factors in aligning transport capacity with demand, while Koç et al. [8] further demonstrated that both the number and type of vehicles affect costs, transport organisation, and environmental impact.

The mass of transported cargo represents the mass-related dimension of transport system workload. Tonnage defines the scale of the material flow handled by the system and enables comparisons between periods with a similar number of operations but different material volumes. Freight flow modelling emphasises the need to distinguish between commodity flows and vehicle flows, because changes in the number of trips do not necessarily imply proportional changes in cargo mass [9]. Lindsey et al. [10] used origin tons, destination tons and ton-miles to analyse the relationship between freight transport activity and demand for industrial space. Sultanbek et al. [11], in a study of rail freight transport, treated freight volume in tonnes and transport work as basic indicators for demand and resource planning. More recently, Karasu et al. [12] examined whether national-accounts and agricultural-production data could improve forecasts of road freight volume, thereby confirming that tonnage can be treated not only as a reporting measure but also as a forecasting variable representing future material flows.

The most synthetic measure of the scale of transport activity is transport work expressed in tonne-kilometres, because it combines cargo mass with transport distance. This indicator enables the assessment of the actual work performed by the transport system, rather than merely the number of operations or the amount of cargo transported. Woodburn [13] indicated that tonne-kilometres are a basic measure of freight transport activity

and of comparisons of market share in freight transport. Dente and Tavasszy [14] treated tonne-kilometres as an indicator of logistics activity relevant to road-freight emissions, while Holden et al. [15] emphasised that tonne-kilometres are a traditional and widely reported measure of logistics activity, although they do not fully describe the operational structure of transport operations. For the purposes of this article, it is particularly important that tonne-kilometres may be used not only as an ex post measure in reporting [16,17], cost assessment [18,19], energy efficiency analysis [20,21], and analysis of the environmental impact of transport [22,23], but also as a forecasted variable supporting the evaluation of future operational workload in the system. Korkmaz and Akgungor [24] estimated passenger-kilometre and tonne-kilometre values for Turkey's road transport system. Their study confirms the usefulness of tonne-kilometres as a measure of transport demand. However, this study was based on annual data and a random split of observations into training and test sets. In the context of monthly operational planning, it is more appropriate to preserve the chronological order of observations and apply a validation procedure that reflects the sequential availability of data.

Previous transport studies have often focused on a single dimension of system activity, such as the number of deliveries, fleet size, cargo mass, costs, emissions, or transport work. Although such an approach is useful in specialised analyses, it may be insufficient for the operational planning of a transport system. The same number of transport tasks may generate different vehicle requirements, different tonnage levels, and different values of transport work expressed in tonne-kilometres. Recent studies have broadened the scope of transport demand forecasting by considering multiple distribution centres [25], developing indicator systems for regional logistics demand [26], and incorporating relevant influencing factors into models based on historical time-series data [27]. These studies demonstrate the value of integrating complementary information, but they generally forecast one principal output variable. The present study therefore adopts a coordinated multi-indicator framework in which four complementary workload indicators are forecast separately using a common data structure, forecast horizon, validation procedure, and set of error measures. This design does not constitute a multivariate joint-forecasting model; rather, it provides a consistent basis for identifying divergences between the organisational, resource-related, mass-related, and operational dimensions of transport. These indicators are therefore complementary rather than interchangeable and should be interpreted jointly within a coordinated planning process.

Since the four indicators represent complementary dimensions of transport workload, their coordinated interpretation enables the identification of whether changes in monthly demand are driven primarily by organisational workload, fleet resource requirements, transport capacity requirements, or overall transport intensity. Accordingly, one-month-ahead forecasts supplemented with prediction intervals can support rolling revisions of monthly transport plans and the determination of capacity reserves under forecast uncertainty.

Operational transport indicators should not, however, be confused with the complete set of variables that generate freight demand. Freight transport is a derived service whose scale and structure reflect industrial production, trade flows, sectoral economic activity, population and spatial development, fuel and labour costs, infrastructure accessibility, public policy, legal constraints, geopolitical disruptions, and emergency events. Recent research has incorporated national-accounts data into road freight forecasting [12], modelled economic, behavioural, policy, and technological developments through transport scenarios [28], and analysed the effects of GDP, population, urbanisation, and infrastructure on long-term transport demand [29]. Empirical evidence also confirms relationships between sectoral GDP and road freight indicators [30], while the COVID-19 pandemic demonstrated that external shocks and regulatory restrictions may cause substantial deviations from

historically expected freight volumes [31]. Policy-oriented transport outlooks further show that freight activity is conditioned by infrastructure and regulatory scenarios, not solely by internal transport characteristics [32].

The relevance of these external determinants depends on the forecasting purpose, planning horizon, system boundaries, and availability of consistent data. The transport system analysed in this study is a military road transport system operating as an internal, non-commercial component of the armed forces' logistics structure. Its monthly workload is generated primarily by centrally planned transport tasks arising from logistics plans, training activities, operational requirements, sustainment needs, and command priorities. In accordance with the principles of NATO logistics doctrine and operational logistics, military transport activity is therefore shaped mainly by mission requirements, resource availability, readiness objectives, and established planning procedures [33,34]. Economic, geopolitical, demographic, and legislative conditions may influence the scale and structure of military transport activity, but their effects are indirect and are translated into specific transport tasks through defence planning, operational decisions, resource allocation, and applicable regulations. The present study consequently focuses on short-horizon forecasting based on historical monthly workload indicators and is intended to support recurring operational planning within the defined military transport system.

Forecasting transport work may be performed using time-series models, as data describing transport activity are time-ordered and reflect changes occurring in successive observation periods. Both classical models and more advanced forecasting methods have been applied in the literature. Sultanbek et al. [11] used an ARIMA model to forecast demand in rail freight transport. Chuwang and Chen [35] applied Box–Jenkins, SARIMA, ARMA, and Prophet models to forecast passenger demand, while Mamede et al. [25] indicated that transport demand forecasting should be treated as a time-series analysis based on historical data. In road freight transport, Liachovičius et al. [5] compared econometric models, including ARIMA-class methods, with artificial neural networks for demand and freight-rate forecasting. Li and Wei [27] compared time-series and impact-factor variants of LSTM, grey, neural network, and regression models, while Schmid et al. [36] provided a recent benchmark comparison of statistical and machine learning methods for logistics time series.

This study adopts a comparative set of classical forecasting models, including the naive model, the moving-average model, Brown's and Holt's exponential smoothing models, ETS, and ARIMA. The inclusion of simple benchmark methods is important because comparative studies have shown that more complex models do not always significantly improve forecasting accuracy over baseline models [37]. Exponential smoothing models are justified for time series exhibiting changes in level, trend, or seasonality [38], whereas ARIMA is a classical approach to stochastic modelling of temporal dependencies. Using several methods allows assessment of whether increased model complexity translates into improved forecasting performance for individual transport workload indicators.

An important aspect of model evaluation is the choice of validation procedure. A one-off split of the dataset into training and test subsets may yield results that poorly reflect real-world planning conditions, in which new observations are added to the database over time and forecasts are updated in subsequent periods. Tashman [39] emphasised the importance of out-of-sample testing for assessing forecasting accuracy, while Bergmeir and Benítez [40] highlighted the limitations of classical cross-validation for time-dependent data. For this reason, rolling-origin validation with an expanding training window was used in this study. This procedure preserves the chronological order of observations and better reflects the practice of monthly transport planning.

The aim of this article is to assess the usefulness of classical time-series models for multi-indicator forecasting of the monthly workload of a road freight transport system and to demonstrate how such forecasts can support smart logistics and fleet transport capacity planning. The analysis covers four complementary indicators: the number of transport tasks, the number of vehicles assigned to task execution, tonnage, and transport work measured in tonne-kilometres. This study seeks to identify the models that deliver the best forecasting performance for each workload dimension and to assess whether a multi-indicator approach provides more comprehensive planning information than analysing a single demand variable.

The main contributions of the paper are as follows:

- Simultaneous forecasting of four complementary indicators of monthly transport workload, covering the organisational, resource-related, mass-related, and operational dimensions of the analysed system;
- Treating transport work measured in tonne-kilometres as a forecasting variable, rather than solely as a reporting indicator used in ex post analyses;
- Comparison of forecasting models with varying levels of complexity within a unified rolling-origin validation procedure, ranging from the naive model and the moving-average model, through Brown's and Holt's exponential smoothing models and ETS, to ARIMA;
- Application of rolling-origin validation with an expanding training window, which enables assessment of forecasting performance under conditions similar to real-world sequential transport planning;
- Application of a multi-criteria assessment of forecasting performance using MAE, RMSE, MAPE, MASE, and Bias, which enables the evaluation not only of error magnitude but also of relative accuracy compared with the naive benchmark and the direction of systematic forecast deviation;
- Interpretation of the forecasting procedure as a decision-support module for smart logistics, in which empirical data are transformed into monthly workload indicators, models are evaluated using rolling-origin validation, and the recommended forecasts are translated into planning implications for fleet capacity, maintenance activities, and transport reserves.

Section 2 presents the Materials and Methods, including the characteristics of the dataset, data preprocessing, construction of transport workload indicators, exploratory and seasonal analysis, description of the forecasting models applied, the rolling-origin validation procedure, assessment of forecasting accuracy, and the method used to determine prediction intervals. Section 3 reports the empirical results, including analyses of indicator variability, seasonality assessment, time-series decomposition, model ranking, and prediction intervals for the recommended models. Section 4 discusses the results, their operational interpretation, implications for supporting resource and transport capacity planning in the transport system, research limitations, and directions for further analysis. Section 5 synthesises the main conclusions and highlights the practical relevance of the proposed forecasting approach.

2. Materials and Methods

The empirical basis of the study comprised operational records from an internal, non-commercial military road transport system responsible for the movement of material resources within Poland. The dataset covered transport tasks planned and performed between 1 January 2020 and 31 December 2025. Under peacetime conditions, the workload of the system results primarily from internally scheduled logistics, supply, training, relocation, and operational-support activities. The source records were subjected to preprocessing

and monthly aggregation, followed by the construction of transport workload indicators, exploratory analysis, and evaluation of forecasting models in accordance with the research framework presented in Figure 1. Data preprocessing, monthly aggregation, statistical analysis, forecasting-model evaluation, and preparation of charts were performed using Python 3.11 (Python Software Foundation, Wilmington, DE, USA).

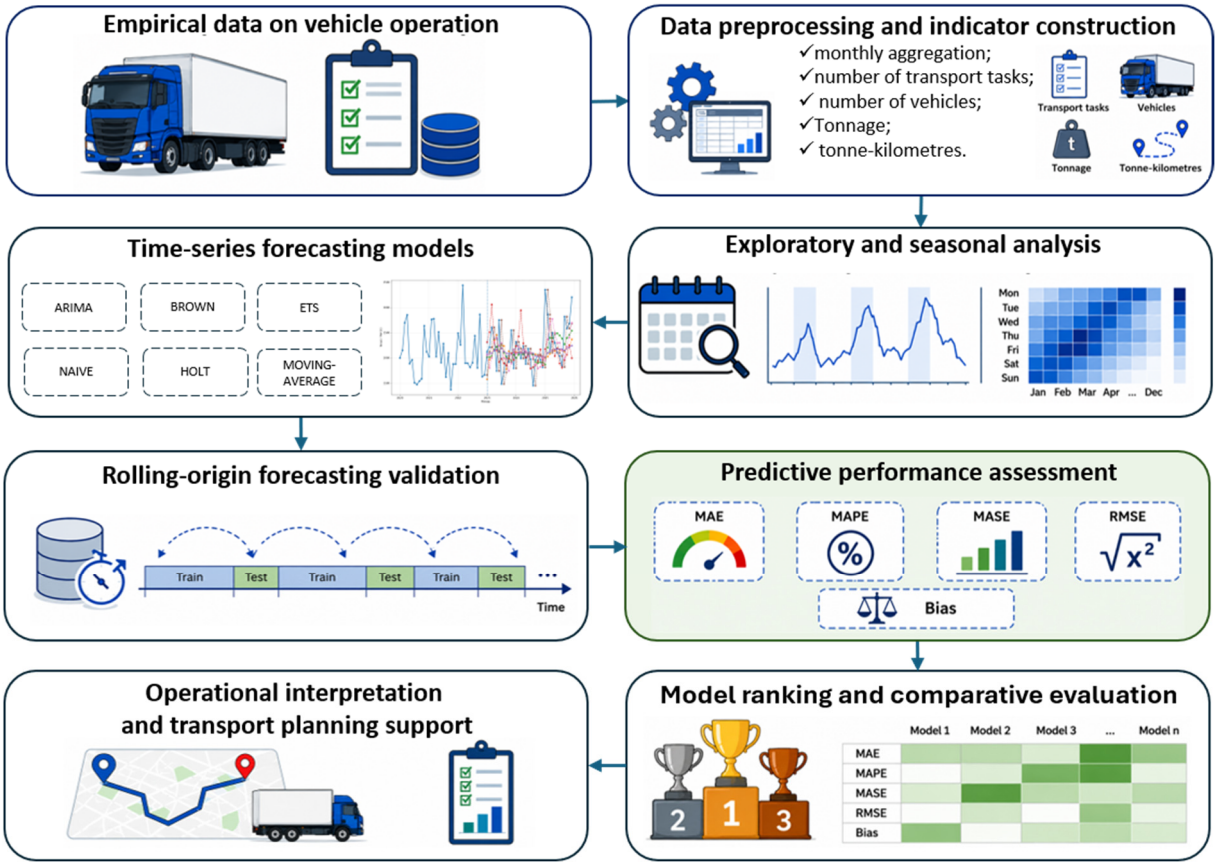

Figure 1. Methodological framework for data preprocessing, time-series forecasting and evaluation of transport workload indicators.

The analysed military transport system served as the empirical case study for the development and evaluation of the forecasting procedure. The methodological framework is not specific to military logistics, as it is based on generally applicable stages of monthly data aggregation, indicator construction, time-series modelling, rolling-origin validation, model comparison, and uncertainty assessment. Accordingly, the framework may be adapted to other freight transport systems, provided that the workload indicators and model inputs are adjusted to their organisational conditions, data availability, and demand-generation mechanisms.

The analysis included four basic indicators of transport workload: N_t—the number of transport tasks per month t, V_t—the number of vehicles assigned to transport task execution per month t, Q_t—the total tonnage scheduled for transport per month t, and TKM_t—transport work expressed in tonne-kilometres per month t.

The selection of these variables enables a multidimensional assessment of the workload of the analysed transport system. The number of transport tasks characterises the organisational aspect of system operation, including planning, coordination, and handling of individual transport operations. The number of vehicles reflects the level of demand for fleet resources, whereas tonnage defines the mass-related dimension of the tasks performed. Transport work expressed in tonne-kilometres combines cargo mass with transport distance and therefore constitutes a synthetic indicator of the scale of transport work performed.

The set of variables defined in this way makes it possible to analyse both the intensity of transport activity and its variability on a monthly and multi-year basis.

2.1. Data Preprocessing and Indicator Construction

The first computational stage involved verification of the source data and their transformation into a regular monthly structure. The data were aggregated by month, which made it possible to construct the time series required for further forecasting analysis.

The following equations constitute study-specific operational definitions developed by the authors to aggregate the source transport records into monthly workload indicators. They are not intended as new general mathematical methods, but formalise the procedure used to construct the analysed variables from the empirical dataset.

For each month t, the number of transport tasks was determined as the number of source records assigned to that month:

$$N_t = n_t \tag{1}$$

where N_t—the number of transport tasks in month t and n_t—the number of transport records assigned to month t.

The total number of vehicles assigned to transport task execution was calculated using the following formula:

$$V_t = \sum_{i=1}^{n_t} v_{i,t} \tag{2}$$

where $v_{i,t}$—the number of vehicles assigned to the i-th transport task in month t.

Monthly tonnage was defined as follows:

$$Q_t = \sum_{i=1}^{n_t} q_{i,t} \tag{3}$$

where $q_{i,t}$—the tonnage assigned to the i-th transport task in month t.

Monthly transport work expressed in tonne-kilometres was calculated using the following formula:

$$TKM_t = \sum_{i=1}^{n_t} tkm_{i,t} \tag{4}$$

where $tkm_{i,t}$—the tonne-kilometre value generated by the i-th transport task in month t.

2.2. Exploratory and Seasonal Analysis

For each analysed transport workload indicator, a monthly time series was defined as follows:

$$Y_t \in \{N_t, V_t, Q_t, TKM_t\},\ t = 1, 2, \ldots, T \tag{5}$$

Before the main forecasting modelling stage, each of these time series was subjected to exploratory and seasonal analysis. This stage was not intended to generate forecasts, but rather to formally identify the internal structure of the series, particularly its trend, seasonal, and irregular components. Accordingly, an additive time-series decomposition was adopted [41]:

$$Y_t = T_t + S_t + R_t \tag{6}$$

where T_t—the trend component, S_t—the seasonal component, and R_t—the irregular component.

The irregular component was determined as follows:

$$R_t = Y_t - T_t - S_t \tag{7}$$

For monthly data, annual seasonality was assumed, with $m = 12$. This means that under a recurring seasonal pattern the seasonal component repeats every twelve months:

$$S_t = S_{t-12} \tag{8}$$

This assumption follows from the natural calendar structure of monthly data and makes it possible to represent recurring fluctuations within the annual cycle.

2.3. Time-Series Forecasting Models

In the forecasting stage of the study, six classical time-series models were applied: the naive model, the moving-average (MA) model, Brown's model, Holt's model, ETS, and ARIMA. The model selection was designed for comparative purposes. The set of models included both simple benchmark methods and exponential smoothing approaches, as well as the ARIMA model. This approach is consistent with the findings of comparative forecasting studies, which indicate that simple methods should provide a benchmark for more complex models, since they often achieve competitive forecasting performance [37].

The naive model assumes that the forecast for the next month is equal to the most recently observed value of the series [41]:

$$\hat{Y}_{t+1|t} = Y_t \tag{9}$$

This model serves as a benchmark and provides a reference point for more complex forecasting methods. The inclusion of a baseline model is recommended in comparative forecasting studies, as it enables an assessment of whether more advanced models actually deliver additional predictive value [37,41].

The moving-average model determines the forecast as the average of the most recent observations k [37,41]:

$$\hat{Y}_{t+1|t} = \frac{1}{k}\sum_{j=0}^{k-1} Y_{t-j} \tag{10}$$

where k—the length of the moving-average window.

The value of this parameter was selected empirically using a rolling-origin validation procedure with an expanding training window. The tested values were $k \in \{2, 3, 4, 6, 12\}$, corresponding to two-month, three-month, four-month, semi-annual, and annual smoothing horizons. For each analysed transport workload indicator, the moving-average variant that achieved the lowest MASE value was adopted for the final comparison. The rolling-origin MASE values for all tested moving-average window lengths are reported in Appendix A, Table A1. The moving-average model was included as a simple method for smoothing short-term fluctuations in the time series. Its inclusion in the comparative set of models is justified by its low computational requirements and by the possibility of assessing whether more complex forecasting models outperform a basic smoothing procedure [37].

Brown's classical model was applied as a one-parameter double exponential smoothing method [42]:

$$S'_t = \alpha Y_t + (1-\alpha)S'_{t-1} \tag{11}$$

$$S''_t = \alpha S'_t + (1-\alpha)S''_{t-1} \tag{12}$$

$$a_t = 2S'_t - S''_t \tag{13}$$

$$b_t = \frac{\alpha}{1-\alpha}(S'_t - S''_t) \tag{14}$$

$$\hat{Y}_{t+h|t} = a_t + hb_t \tag{15}$$

where: S'_t and S''_t—the single- and double-smoothed values of the series, a_t—the local level of the series, b_t—the local trend increment, α—the smoothing parameter $(0 < \alpha < 1)$, and h—the forecast horizon.

Holt's linear trend model was applied as a two-parameter exponential smoothing method accounting for changes in the level and trend of the time series, without an explicit seasonal component [43]:

$$l_t = \alpha Y_t + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (16)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (17)$$

$$\hat{Y}_{t+1|t} = l_t + b_t \quad (18)$$

The seasonal exponential smoothing model denoted as ETS in the empirical comparison was specified using the additive Holt–Winters formulation [44] as follows:

$$\ell_t = \alpha(Y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1}) \quad (19)$$

$$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1} \quad (20)$$

$$s_t = \gamma(Y_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (21)$$

$$\hat{Y}_{t+h|t} = \ell_t + b_t + s_{t+1-m} \quad (22)$$

where ℓ_t—the level component of the time series at time t—, b_t—the trend component, α, β, and γ—the smoothing parameters, s_t—the seasonal component, and s_{t-m}—the seasonal component for the corresponding month of the preceding annual cycle, whereas s_{t+1-m}— the seasonal component corresponding to the forecast month. For monthly data, the seasonal period was set to $m = 12$.

The ARIMA model was applied in its non-seasonal form following the standard Box–Jenkins formulation [45]:

$$ARIMA(p, d, q) \quad (23)$$

For each transport workload indicator, candidate ARIMA specifications were screened over $p \in \{0, 1, 2, 3\}$, $d \in \{0, 1, 2\}$, and $q \in \{0, 1, 2, 3\}$, excluding $ARIMA(0, 0, 0)$. This yielded 47 candidate specifications. The order was selected separately for each indicator using the initial 36-month training sample, with the minimum Akaike information criterion (AIC) as the selection criterion. The selected order was retained throughout the expanding-window rolling-origin validation, while the model coefficients were re-estimated at each forecasting origin using all observations available up to that month. Consequently, no observations from the validation period were used to identify the ARIMA order. The selected orders and corresponding AIC values are reported in Appendix A, Table A2. AIC was used as an in-sample order-identification criterion, whereas the final assessment of the selected ARIMA specification was based on out-of-sample rolling-origin forecasting performance.

where p—the order of the autoregressive component, d—the degree of differencing, and q—the order of the moving-average component.

Formally, the model can be expressed using the lag operator B:

$$\phi(B)(1 - B)^d Y_t = \theta(B)\varepsilon_t \quad (24)$$

where $\phi(B)$—the autoregressive polynomial, $\theta(B)$—the moving-average polynomial, and ε_t—the random error term. These polynomials take the following form:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \quad (25)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q \quad (26)$$

2.4. Rolling-Origin Forecasting Validation with an Expanding Training Window

The use of rolling-origin validation results from the need to evaluate forecasting models under out-of-sample conditions. Tashman [39] emphasises that forecasting models should be assessed using data that were not used in the estimation process, whereas Bergmeir and Benítez [40] indicate that classical cross-validation may violate the assumptions arising from temporal dependence in time-series data. For this reason, a validation procedure that preserves the chronological order of observations was applied in this study.

On the basis of these methodological principles, Equations (27)–(30) provide a study-specific mathematical formalisation of the expanding-window rolling-origin validation procedure applied in this research.

Let $Y = \{Y_t : t = 1, 2, \ldots, N\}$ denote the monthly time series of a given transport workload indicator, where t is the time index and N denotes the total number of observations. The initial length of the training window was set to $L = 36$ months. The forecast horizon was $h = 1$, which means that a one-month-ahead forecast was generated. The number of forecasts obtained in the rolling-origin validation procedure was therefore equal to

$$R = N - L \quad (27)$$

For the r-th validation step, where $r = 1, 2, \ldots, R$, the training set was defined as follows:

$$\mathcal{D}_r = \{Y_1, Y_2, \ldots, Y_{L+r-1}\} \quad (28)$$

Based on the training set D_r, a forecasting model m was estimated, and then a one-step-ahead forecast was generated for the observation in period $L + r$:

$$\hat{Y}^{(m)}_{L+r|L+r-1} \quad (29)$$

For model m, the forecast error at the r-th validation step was defined as follows:

$$e_r^{(m)} = Y_{L+r} - \hat{Y}^{(m)}_{L+r|L+r-1} \quad (30)$$

Once the actual value for month $L + r$ had become available, the corresponding observation was incorporated into the training set in the subsequent validation step. Thus, the training window expanded sequentially, whereas the forecast horizon remained fixed.

2.5. Assessment of Forecasting Performance

Forecasting accuracy was assessed using five error metrics consistent with the research framework presented in Figure 1: MAE, RMSE, MAPE, MASE, and Bias. The mathematical expressions of these comparison criteria are given in Equations (31)–(35). The literature emphasises that the choice of error metric may significantly affect the assessment of forecasting performance and the final ranking of the forecasting methods compared [46]. For this reason, the use of a single measure is not sufficient to fully characterise model accuracy. Therefore, the parallel use of absolute, percentage, and scaled measures is justified, as each of these groups describes a different aspect of forecast error.

The Mean Absolute Error was calculated as follows [41,46]:

$$MAE^{(m)} = \frac{1}{R}\sum_{r=1}^{R} \mid e_r^{(m)} \mid, \quad (31)$$

where R—the number of one-step-ahead forecasts obtained in the rolling-origin validation procedure, and $e_r^{(m)}$—the forecast error of model m at validation step r.

The Root Mean Square Error was determined as follows [41,46]:

$$RMSE^{(m)} = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(e_r^{(m)}\right)^2}. \quad (32)$$

The Mean Absolute Percentage Error was defined as follows [41,46]:

$$MAPE^{(m)} = \frac{100}{R}\sum_{r=1}^{R}\left|\frac{e_r^{(m)}}{Y_{L+r}}\right|. \quad (33)$$

The Mean Absolute Scaled Error was calculated as follows [47]:

$$MASE^{(m)} = \frac{\frac{1}{R}\sum_{r=1}^{R}\left|e_r^{(m)}\right|}{\frac{1}{L-s}\sum_{t=s+1}^{L}\left|Y_t - Y_{t-s}\right|}, \quad (34)$$

where $s = 12$—annual seasonality, whereas $L = 36$—the initial length of the training window.

As an auxiliary indicator of the direction of forecast error, Bias was used, understood as the mean signed forecast error, i.e., the Mean Error, according to the following formula [48]:

$$Bias^{(m)} = ME^{(m)} = \frac{1}{R}\sum_{r=1}^{R} e_r^{(m)}. \quad (35)$$

A Bias value greater than zero indicates a tendency towards underestimation of the analysed indicator, whereas a negative value indicates a tendency towards overestimation.

2.6. Determination of Prediction Intervals

For each recommended model m, one-month-ahead prediction intervals were constructed from the empirical distribution of the 36 forecast errors obtained during expanding-window rolling-origin validation. These quantities are out-of-sample forecast errors, because the observation evaluated at each validation step was not available when the corresponding forecast was generated. They are therefore distinct from classical in-sample residuals obtained from fitting a model to the same observations [40,41].

The normality of the forecast errors was assessed using the Shapiro–Wilk test [49], supplemented by skewness, excess kurtosis, and normal Q–Q plots. In the Shapiro–Wilk test, the null hypothesis states that the forecast errors follow a normal distribution. At validation step r, the forecast error was defined as follows:

$$e_r^{(m)} = Y_{L+r} - \hat{Y}_{L+r|L+r-1}^{(m)}, \qquad r = 1, \ldots, R. \quad (36)$$

Equation (36) restates the study-specific rolling-origin error definition introduced in Equation (30) for the recommended model.

For $p \in (0, 1)$, the empirical p-th quantile of the R validation errors was defined as follows [50]:

$$Q_p^{(m)} = Quantile_p\left(e_1^{(m)}, \ldots, e_R^{(m)}\right). \quad (37)$$

To provide a consistent method applicable to all recommended models without imposing normality or symmetry on the forecast-error distributions, the prediction intervals were constructed non-parametrically using empirical quantiles of the rolling-origin forecast errors. Following the approach proposed by Lee and Scholtes [51], the lower and upper limits were obtained by adding the corresponding empirical forecast-error quantiles to the point forecast:

$$PI_{1-\alpha,t+1}^{(m)} = \left[\hat{Y}_{t+1|t}^{(m)} + Q_{\alpha/2}^{(m)},\ \hat{Y}_{t+1|t}^{(m)} + Q_{1-\alpha/2}^{(m)}\right]. \quad (38)$$

For the 80% interval, alpha was set to 0.20 and the 0.10 and 0.90 empirical quantiles were used. For the 95% interval, alpha was set to 0.05 and the 0.025 and 0.975 empirical quantiles were used. Since all analysed workload indicators are non-negative, lower interval limits below zero were truncated at zero.

Interval performance was assessed using Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW). PICP measures empirical coverage, whereas MPIW measures average interval width [52,53]. To reduce optimistic evaluation, both measures were calculated in leave-one-out form: for validation step r, the interval quantiles were estimated from the remaining R − 1 forecast errors.

$$PICP_{1-\alpha}^{(m)} = \frac{1}{R}\sum_{r=1}^{R} c_r^{(m)}, \qquad MPIW_{1-\alpha}^{(m)} = \frac{1}{R}\sum_{r=1}^{R}\left(U_r^{(m)} - L_r^{(m)}\right). \tag{39}$$

where $c_r^{(m)}$ equals 1 when the observed value lies within the leave-one-out interval and 0 otherwise, while $L_r^{(m)}$ and $U_r^{(m)}$ denote the corresponding lower and upper limits. The study used two standard levels for reporting forecast uncertainty: 80% and 95% [54].

3. Results

3.1. Preliminary Data Processing and Variability of Transport Workload Indicators

In accordance with the research framework presented in Figure 1, the source data were aggregated into monthly time series covering the period from January 2020 to December 2025. For each month, four transport workload indicators were calculated: the number of transport tasks (N_t), the number of vehicles assigned to task execution (V_t), tonnage (Q_t), and transport work expressed in tonne-kilometres (TKM_t). Ultimately, 72 monthly observations were obtained for each analysed indicator.

In the first stage, changes in the indicator values were assessed on an annual basis. The results presented in Table 1 indicate an increase in the organisational, resource-related, and mass-related dimensions of the transport system workload. Between 2020 and 2025, the number of transport tasks increased from 941 to 1413, i.e., by 50.2%. Over the same period, the number of vehicles assigned to task execution rose from 1480 to 2016, i.e., by 36.2%, while tonnage increased from 24,401.6 t to 30,334.6 t, i.e., by 24.3%. A different pattern was observed for transport work. The value of tonne-kilometres in 2025 was close to the 2020 level, amounting to 9.03 million tkm in both years. The highest value of transport work was recorded in 2023 (10.15 million tkm), whereas the lowest value was observed in 2024 (8.52 million tkm). This indicates that the increase in the number of tasks and vehicles did not translate proportionally into an increase in transport work performed.

Table 1. Annual values of transport workload indicators for 2020–2025.

Year	V_t [no.]	N_t [no.]	Q_t [t]	TKM_t [tkm]
2020	1480	941	24,401.6	9,032,312
2021	1365	979	25,118.1	8,906,382
2022	1569	1241	26,073.4	9,928,057
2023	1810	1307	25,283.3	10,146,837
2024	1969	1406	23,992.6	8,522,762
2025	2016	1413	30,334.6	9,026,003
Changes 2020–2025 [%]	36.2	50.2	24.3	−0.1

Next, the monthly variability of the analysed indicators was assessed. The descriptive statistics summarised in Table 2 show the range, average level, dispersion of monthly

values, and relative variability of the indicators. To compare the variability of variables expressed in different units, the coefficient of variation was calculated [55]:

$$CV = \frac{\sigma}{\overline{\mu}} \cdot 100\% \quad (40)$$

where σ—the standard deviation of the monthly values of a given indicator and μ—arithmetic mean.

Table 2. Descriptive statistics for monthly transport workload indicators.

Indicator	μ	Median	Minimum	Maximum	σ	CV [%]
N_t [no.]	101.21	105.00	63.00	129.00	19.87	19.6
V_t [no.]	141.79	145.00	95.00	195.00	27.50	19.4
Q_t [t]	2155.60	2048.50	1370.00	3437.00	458.44	21.3
TKM_t [tkm]	771,699.37	722,551.40	430,675.80	1,377,523.80	189,537.25	24.6

The number of transport tasks ranged from 63 to 129 tasks per month, with a mean of 101.21 and a median of 105 tasks. Figure 2d shows an increase in the level of this variable from 2022 onwards, with higher values maintained in the period 2023–2025.

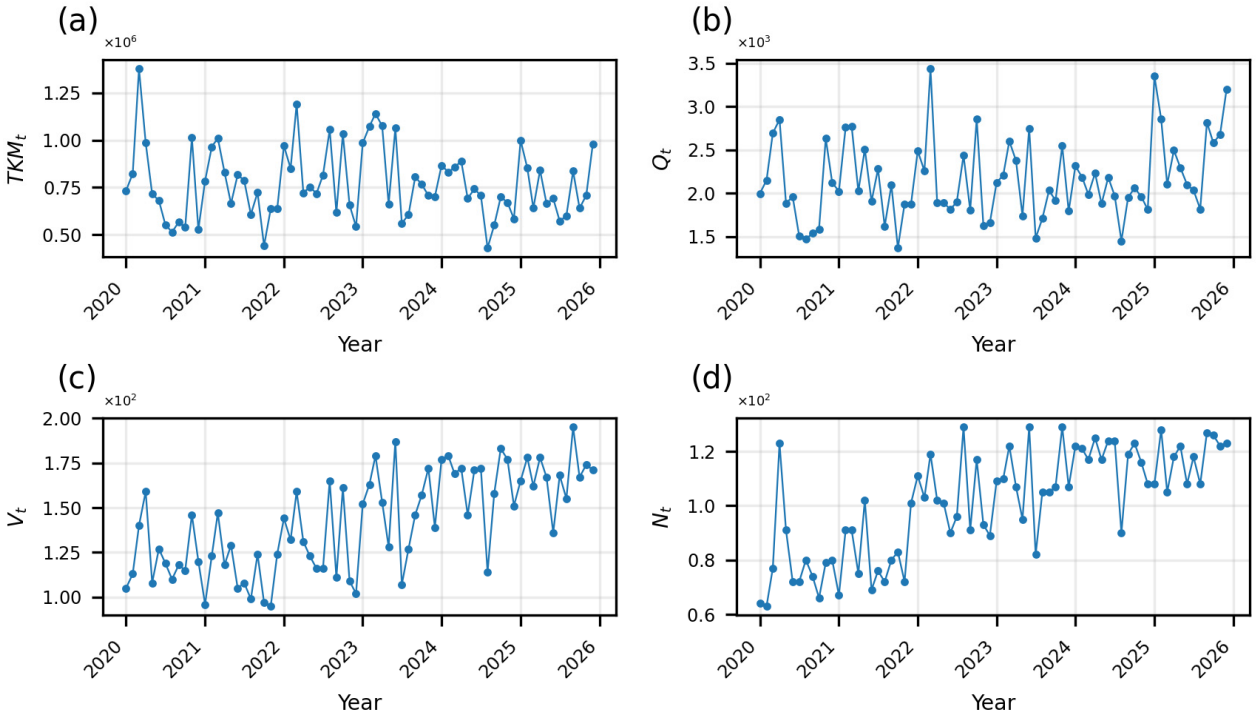

Figure 2. Monthly variability of transport workload indicators: (**a**) transport work; (**b**) tonnage; (**c**) number of vehicles assigned to task execution; (**d**) number of transport tasks.

The number of vehicles assigned to task execution ranged from 95 to 195 vehicles per month. The monthly mean was 141.79 vehicles, while the median was 145 vehicles. The pattern shown in Figure 2c indicates an increase in fleet resource demand in the final part of the series, particularly in 2024–2025. This suggests that the increase in the number of tasks was associated with the growing deployment of vehicle resources.

Tonnage was characterised by greater monthly variability than the number of tasks and the number of vehicles. Monthly values ranged from 1370.00 t to 3437.00 t, with a mean of 2155.60 t and a median of 2048.50 t. The variability shown in Figure 2b indicates that the mass of transported cargo was strongly dependent on the structure of the tasks performed, rather than solely on the number of transport operations.

Transport work exhibited the greatest amplitude of variation. Monthly values ranged from 430,675.80 tkm to 1,377,523.80 tkm, with a mean of 771,699.37 tkm and a median of 722,551.40 tkm. The pattern shown in Figure 2a confirms the high variability of this indicator, resulting from the combined effect of cargo mass and route length.

The CV values confirm that the greatest relative variability was observed for transport work ($CV = 24.6\%$), followed by tonnage ($CV = 21.3\%$). Lower values were obtained for the number of transport tasks ($CV = 19.6\%$) and the number of vehicles assigned to task execution ($CV = 19.4\%$). This means that mass-related and operational indicators were more susceptible to variability in the transport structure than organisational and resource-related indicators.

The comparison of the results presented in Tables 1 and 2 and Figure 2 confirms the validity of a multi-indicator approach to assessing transport workload. Table 1 shows the direction of annual changes, Table 2 summarises the level and range of monthly variability, while Figure 2 enables the temporal patterns of the indicators to be assessed. These results provide the basis for further analysis of seasonality, time-series decomposition, and evaluation of forecasting model performance.

3.2. Analysis of Monthly Seasonality

The distributions of the analysed indicators across individual months were presented using box plots, in which both the median and the mean were indicated (Figure 3a–d).

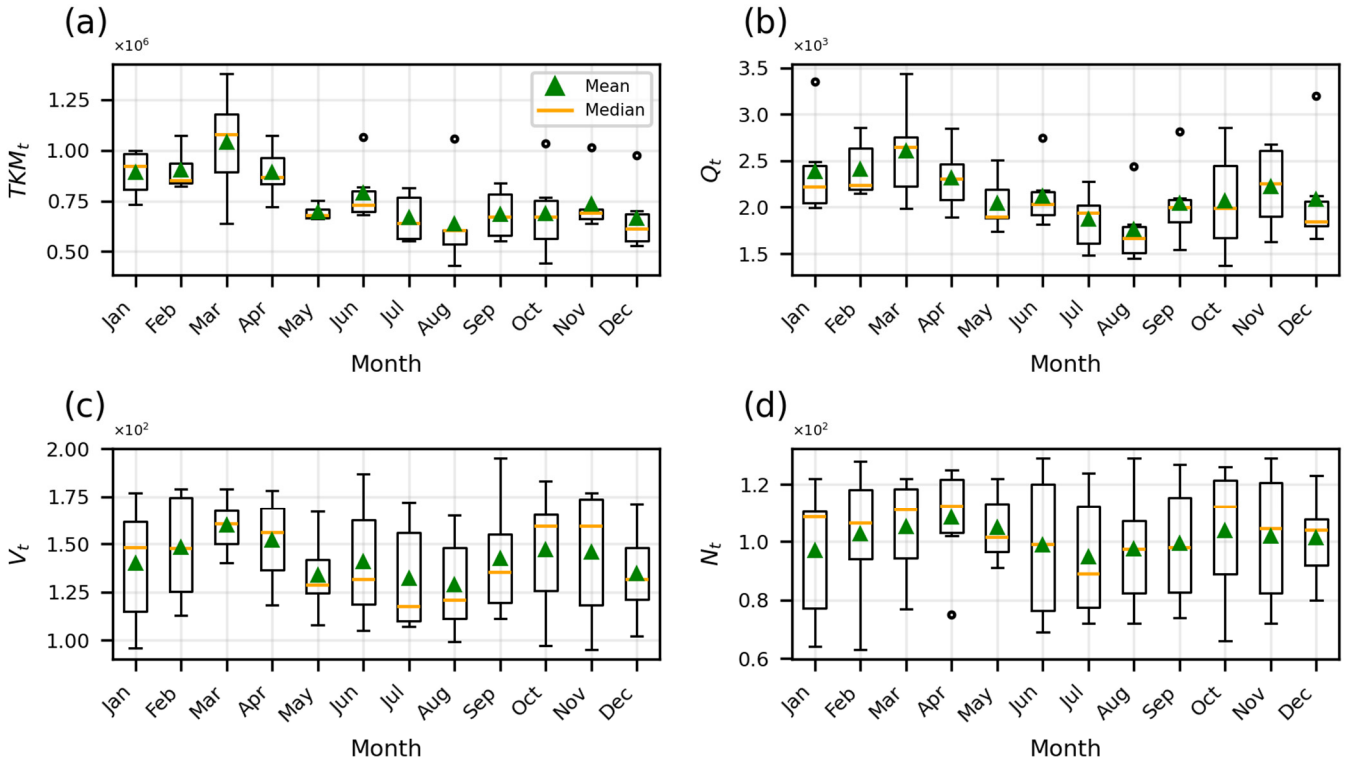

Figure 3. Monthly seasonality of transport workload indicators: (**a**) transport work; (**b**) tonnage; (**c**) number of vehicles assigned to task execution; (**d**) number of transport tasks.

The seasonality analysis revealed differences in the monthly level of transport workload. For the number of transport tasks, the highest mean value was recorded in April, at 108.33 tasks, whereas the lowest value was recorded in July, at 94.67 tasks. This indicates a moderate decrease in organisational workload during the summer period and a higher level of activity in the spring months.

For the number of vehicles, the highest mean value was recorded in March, at 159.33 vehicles, whereas the lowest value was recorded in August, at 128.33 vehicles.

This means that the system's resource demand was highest in the first part of the year and lower during the summer period. From the perspective of fleet operation planning, this implies the need to ensure greater vehicle availability in months with increased workload.

For tonnage, the highest mean value was also recorded in March, at 2599.05 t, whereas the lowest value was recorded in August, at 1751.42 t. A similar seasonal pattern was observed for tonne-kilometres, for which the highest monthly mean occurred in March and amounted to 1,036,011.30 tkm, whereas the lowest value occurred in August and amounted to 635,958.82 tkm.

These results indicate that the spring months generated higher transport workload, particularly in the resource-related and operational dimensions. At the same time, the width of the boxes and the presence of outliers indicate that seasonality does not fully explain the variability of the analysed indicators. This is particularly evident for tonnage and tonne-kilometres, for which substantial between-year variability within the same months can be observed, indicating a significant contribution of the irregular component.

3.3. Time-Series Decomposition

To provide a more comprehensive identification of the structure of the analysed time series, additive decomposition was applied, separating each series into trend, seasonal, and residual components (Figure 4a–d).

The decomposition of the number of transport tasks confirms an upward trend in the period 2021–2024. The trend component increased from approximately 79.0 tasks per month in January 2021 to 106.7 tasks in January 2023 and 115.2 tasks in January 2024. The highest trend level was recorded in May 2024, at 117.6 tasks per month. In the final part of the series, the trend stabilised at a similar level, reaching 114.0 tasks in December 2025. The seasonal component had a moderate peak-to-trough range, from −6.7 tasks in July to +8.2 tasks in April. This indicates that variability in the number of tasks was shaped mainly by an increase in the long-term level of system activity, with a smaller contribution from regular monthly fluctuations. The standard deviation of the residual component was 11.7 tasks, indicating the presence of additional irregular variability.

For the number of vehicles assigned to task execution, the upward trend was more pronounced. The minimum trend value was recorded in June 2021, at 113.6 vehicles per month. The trend then increased to 143.5 vehicles in January 2023, 157.9 vehicles in January 2024, and 167.7 vehicles in December 2025. The increase relative to the minimum observed in 2021 was therefore approximately 54 vehicles. The seasonal component ranged from −14.3 vehicles in August to +19.4 vehicles in March, corresponding to a peak-to-trough range of 33.7 vehicles. Compared with the number of transport tasks, seasonality had a greater contribution to shaping vehicle demand. The standard deviation of the residuals was 17.1 vehicles, while the maximum absolute residual deviation was 46.6 vehicles, indicating a significant influence of operational factors not captured by the trend and seasonal components.

The decomposition of tonnage indicates a relatively stable trend in the first part of the series and an increase in the final observation period. The trend component was approximately 1978.6 t in January 2020, 2103.8 t in January 2021, 2156.8 t in January 2023, 2203.7 t in January 2025, and 2558.1 t in December 2025. The highest trend level therefore occurred in the final part of the analysed period. The seasonal component ranged from −425.3 t in August to +475.3 t in March, corresponding to a peak-to-trough range of approximately 900.6 t. The seasonality of tonnage was therefore considerably stronger than that observed for the number of transport tasks and the number of vehicles. At the same time, the standard deviation of the residuals was 363.8 t, while the maximum absolute

residual deviation was 887.5 t. This result indicates that the mass of transported cargo was largely dependent on the irregular structure of the tasks performed.

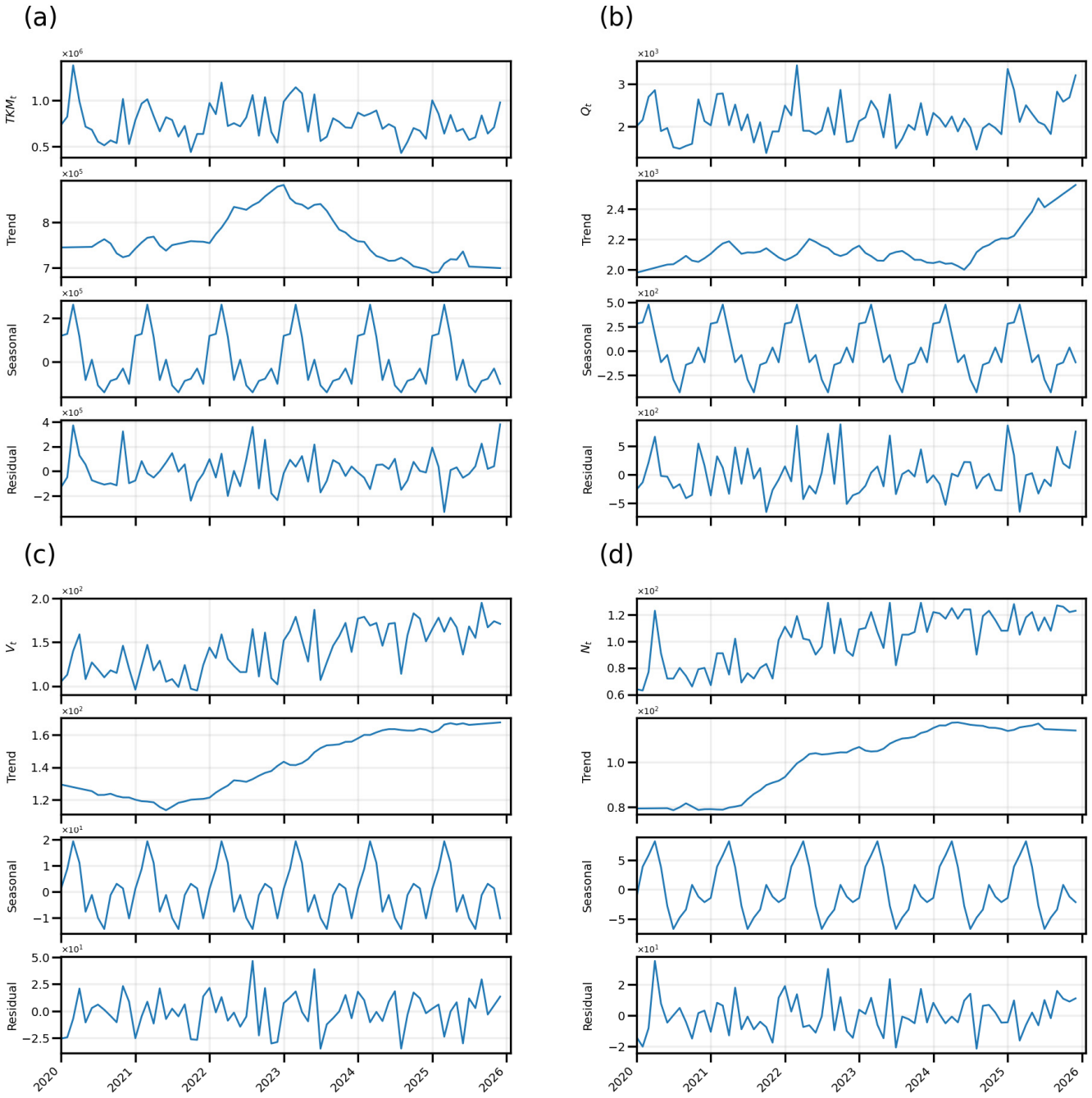

Figure 4. Additive seasonal decomposition of transport workload indicators: (**a**) transport work; (**b**) tonnage; (**c**) number of vehicles assigned to task execution; (**d**) number of transport tasks.

The tonne-kilometre series was characterised by the greatest irregularity. The trend component increased from approximately 742.1 thousand tkm in January 2021 to 882.8 thousand tkm in January 2023, after which it decreased to 758.3 thousand tkm in January 2024 and 689.0 thousand tkm in January 2025. In December 2025, the trend amounted to 699.2 thousand tkm, indicating stabilisation at a level lower than that observed during the peak period. The seasonal component ranged from −140.4 thousand tkm in August to +261.7 thousand tkm in March, corresponding to a peak-to-trough range of approximately 402.1 thousand tkm. The standard deviation of the residuals reached 139.9 thousand tkm, while the maximum absolute residual deviation was 379.5 thousand tkm. This confirms

that transport work was the most sensitive indicator to irregular operational changes, as it depends simultaneously on cargo mass and the length of transport routes.

3.4. Assessment of Forecasting Performance and Model Ranking

The forecasting models were evaluated using a rolling-origin procedure with an expanding training window. The initial training set covered 36 months, i.e., the period 2020–2022, whereas validation was performed for the subsequent 36 months, from January 2023 to December 2025. At each validation step, a one-month-ahead forecast was generated and, once the actual value had become available, the corresponding observation was added to the training set. This approach made it possible to evaluate the models under conditions similar to the real transport planning process, in which forecasts are updated as new operational data become available.

After the forecast errors had been calculated, a model ranking was prepared separately for each analysed transport workload indicator. The primary criterion for selecting the best model was the MASE value, as this measure enables forecasting performance to be assessed relative to the adopted benchmark model and allows forecast accuracy to be compared across series with different value scales [47]. On this basis, Equation (41) constitutes a study-specific decision rule used to identify the recommended model rather than a new forecasting or accuracy measure. The best model for a given variable was defined as follows:

$$m^* = \arg\min_m MASE^{(m)} \tag{41}$$

The remaining metrics, i.e., MAE, RMSE, MAPE, and Bias, were treated as supplementary criteria relative to MASE. MAE was used to assess the average absolute magnitude of the error in the units of the analysed variable, whereas RMSE additionally accounted for the influence of larger deviations between forecasts and actual values. MAPE enabled comparison of the relative error magnitude between indicators, while Bias was used to identify the direction of systematic forecast deviation. A positive Bias value indicated a tendency to underestimate transport workload, whereas a negative value indicated overestimation.

The results summarised in Table 3 indicate that the effectiveness of the forecasting models depended on the characteristics of the analysed time series. No single dominant model was identified for all transport workload indicators. For the number of transport tasks, Brown's model achieved the lowest MASE value. For the number of vehicles assigned to task execution and for transport work expressed in tonne-kilometres, the ETS model achieved the lowest MASE value, whereas for tonnage the best result was obtained using the moving-average model with a window length of $k = 12$.

For the number of transport tasks, Brown's model achieved an MAE of 8.666 tasks, an RMSE of 11.537 tasks, and a MAPE of 7.802%. The MASE value was 0.438 and was the lowest among the analysed indicators, indicating that the number of transport tasks was the relatively easiest variable to forecast. The Bias value of -1.103 indicated a slight tendency to overestimate the forecasts.

For the number of vehicles assigned to task execution, the ETS model achieved the best result. The MAE was 17.147 vehicles, the RMSE was 22.310 vehicles, the MAPE was 11.119%, and the MASE was 0.828. The positive Bias value of 3.370 vehicles indicates a tendency to underestimate fleet demand.

For tonnage, the best-performing model was the moving-average model with a window length of $k = 12$. The MAE was 312.851 t, the RMSE was 421.575 t, and the MAPE was 13.865%. The MASE value was 0.693, confirming the usefulness of this model relative to the adopted reference scale. However, the positive Bias value of 84.714 t indicates a tendency to underestimate the monthly cargo mass.

Table 3. Ranking of forecasting models according to MASE values.

Indicator	Model	MAE	RMSE	MAPE [%]	MASE	Bias
N_t [no.]	Brown	8.666	11.537	7.802	0.438	−1.103
	Holt	8.869	11.852	8.301	0.465	−3.401
	MA (k = 6)	8.968	12.077	8.102	0.470	2.190
	ARIMA	10.149	12.846	9.313	0.532	−0.429
	ETS	10.598	14.457	9.798	0.555	−4.665
	Naive	12.611	16.844	11.596	0.661	0.944
V_t [no.]	ETS	17.147	22.310	11.119	0.828	3.370
	ARIMA	17.235	22.666	11.715	0.832	−0.058
	MA (k = 12)	17.475	21.718	11.207	0.844	7.623
	Brown	17.530	23.086	11.591	0.847	3.584
	Holt	17.532	23.317	11.591	0.847	3.606
	Naive	23.694	29.508	15.845	1.144	1.917
Q_t [t]	MA (k = 12)	312.851	421.575	13.865	0.693	84.714
	Brown	324.332	432.167	14.453	0.718	75.812
	Holt	329.137	440.372	14.534	0.729	96.396
	ETS	345.213	467.480	14.843	0.765	66.942
	ARIMA	359.061	466.459	16.139	0.795	102.709
	Naive	422.038	542.753	19.361	0.935	42.806
TKM_t [tkm]	ETS	120,841.868	169,088.647	16.478	0.719	−28,739.924
	MA (k = 12)	132,015.532	163,860.731	18.134	0.786	−7571.233
	Holt	140,812.622	177,068.263	18.930	0.838	17,271.527
	Brown	143,878.970	172,577.361	19.928	0.856	−16,456.472
	Naive	156,426.101	206,672.619	21.300	0.931	12,067.633
	ARIMA	173,328.304	234,157.830	23.242	1.031	−990.273

For transport work expressed in tonne-kilometres, the ETS model achieved the best results. The MAE was 120,841.868 tkm, the RMSE was 169,088.647 tkm, the MAPE was 16.478%, and the MASE was 0.719. The negative Bias value of −28,739.924 tkm indicates a tendency to overestimate transport work.

A supplementary interpretation of the error measures confirms the varying levels of predictability of the analysed indicators. The lowest relative error was obtained for the number of transport tasks, whereas the highest was observed for transport work expressed in tonne-kilometres. This indicates that the organisational dimension of system workload was characterised by the relatively highest forecasting stability, while the mass-related and operational dimensions showed greater variability. In all cases, the RMSE values were higher than the MAE values, indicating the occurrence of individual months with larger deviations between forecasts and actual values. The direction of forecast error was not uniform: the best models for the number of vehicles and tonnage tended to underestimate the analysed indicators, whereas the best models for the number of transport tasks and tonne-kilometres tended to overestimate them.

The analysis of the full validation results showed that, for the number of transport tasks, Holt's model and the moving-average model ranked after Brown's model. For the number of vehicles assigned to task execution, after the ETS model, the ARIMA and

Brown's models achieved similar forecasting performance. In the case of tonnage, the moving-average model with a window length of $k = 12$ was the most effective, followed by Holt's and Brown's models. For transport work expressed in tonne-kilometres, the ETS model achieved the best result, followed by the moving-average model and Holt's model.

The comparison of the results indicates that, in the analysed dataset, greater formal complexity of a model did not guarantee improved forecasting performance. The ARIMA model did not rank first for any indicator, and for tonne-kilometres it was the least favourable method according to MASE. This result indicates that in short monthly transport time series characterised by a substantial irregular component simpler or adaptive models may achieve results comparable to, or better than, ARIMA-class models.

The MASE values for the best models were below 1 for all analysed indicators. This means that the models applied achieved favourable forecasting performance relative to the adopted reference scale. The lowest MASE value was recorded for the number of transport tasks, whereas the higher values obtained for the number of vehicles, tonnage, and tonne-kilometres indicate greater variability in these series and a stronger influence of operational factors not fully explained by historical monthly observations.

Figure 5 presents the rolling-origin forecasts generated by all analysed models for the four transport workload indicators. This visual comparison complements the quantitative ranking presented in Table 3 and shows how individual models reproduced changes in the validation period.

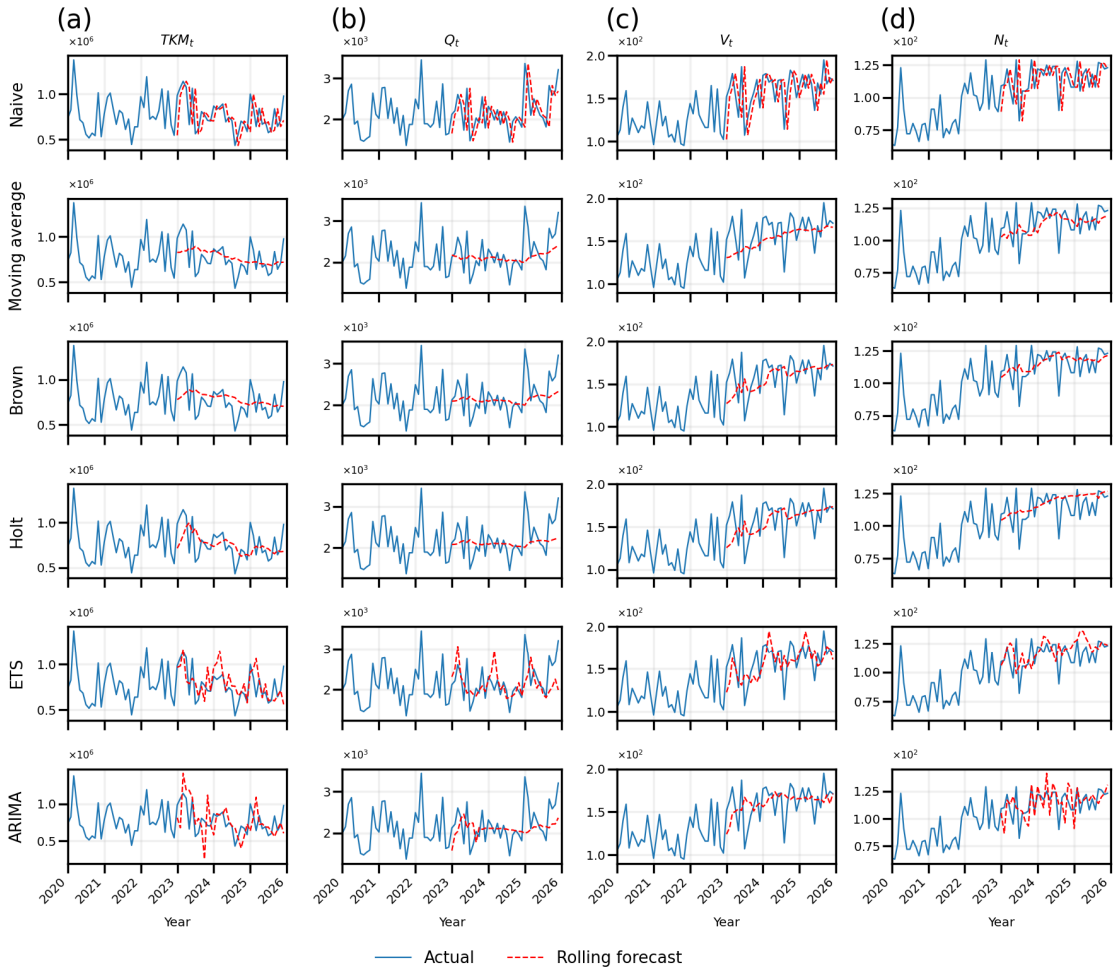

Figure 5. Rolling-origin forecasts for all analysed forecasting models and transport workload indicators: (**a**) transport work; (**b**) tonnage; (**c**) number of vehicles assigned to task execution; (**d**) number of transport tasks.

The blue line represents the actual values, while the red dashed line represents one-month-ahead rolling forecasts. For the number of transport tasks and the number of vehicles assigned to task execution, the forecasts generally followed the increasing level of the series in the validation period, although individual models differed in their responsiveness to monthly fluctuations. For tonnage and transport work, larger deviations between observed and forecasted values were visible, confirming the stronger influence of irregular operational factors on mass-related and operational workload indicators. The visual comparison presented in Figure 5 complements the quantitative ranking reported in Table 3. It shows that the analysed models differed in their ability to reproduce the observed trajectories during the rolling-origin validation period. The results also indicate that forecasting performance varied across workload indicators, which supports the selection of recommended models separately for each analysed variable.

3.5. Prediction Intervals for the Recommended Models

To supplement the assessment of point forecast accuracy, prediction intervals were determined for the recommended models. The intervals were calculated from the empirical distribution of the 36 one-month-ahead forecast errors obtained during expanding-window rolling-origin validation. Two nominal coverage levels were considered: 80% and 95%. The 80% interval was interpreted as a planning range representing typical forecast uncertainty, whereas the 95% interval provided a broader range for decisions requiring a greater safety margin.

The distributional diagnostics are reported in Appendix B, Table A3 and Figure A1. The Shapiro–Wilk test did not reject normality for the ETS model applied to the number of vehicle assignments ($p = 0.819$), the moving-average model with ($k = 12$) applied to tonnage ($p = 0.220$), or the ETS model applied to transport work ($p = 0.371$). For the Brown model applied to the number of transport tasks, normality was rejected ($W = 0.936$, $p = 0.039$).

Table 4 presents the empirical error-quantile offsets used to construct the 80% and 95% prediction intervals, together with the corresponding leave-one-out PICP and MPIW values. The leave-one-out PICP values were 72.2–75.0% for the nominal 80% intervals and 88.9–91.7% for the nominal 95% intervals. The empirical coverage was therefore moderately lower than the corresponding nominal levels. This difference should be interpreted in relation to the limited number of forecast errors (R = 36), which restricts the resolution of empirical quantile estimation, particularly in the tails of the distribution.

Table 4. Prediction interval parameters for the recommended models.

Indicator	Recommended Model	$MPIW_{95}$	$PICP_{95}$ [%]	95% Error-Quantile Offsets	$MPIW_{80}$	$PICP_{80}$ [%]	80% Error-Quantile Offsets
V_t	ETS	78.52	91.7	[−38.22; +39.94]	57.81	72.2	[−28.93; +29.61]
N_t	Brown	51.07	88.9	[−32.46; +19.56]	24.26	75.0	[−13.36; +10.99]
Q_t	MA (k = 12)	1484.09	91.7	[−621.50; +858.53]	992.21	75.0	[−379.00; +616.76]
TKM_t	ETS	624,333	91.7	[−339,017; +282,446]	440,498	72.2	[−248,313; +197,785]

Note: The offsets are empirical quantiles of the out-of-sample forecast errors and are added to each point forecast. PICP and MPIW were evaluated using leave-one-out quantile calibration. MPIW is expressed in the units of the corresponding indicator and should therefore not be compared directly across indicators. Lower limits were constrained to zero.

Because MPIW is expressed in the measurement units of the corresponding indicator, its numerical values should not be compared directly across the four workload variables. The MPIW results were therefore interpreted separately for each indicator and coverage level. The numerically wider intervals obtained for tonnage and transport work reflect the respective scales and units of these variables and should not be interpreted as direct

evidence of greater relative forecast uncertainty than that observed for the number of transport tasks or vehicle assignments.

Figure 6 presents the asymmetric empirical prediction intervals. The blue line represents the actual values, the red dashed line represents the one-month-ahead rolling forecast, and the shaded areas represent the 80% and 95% empirical uncertainty ranges.

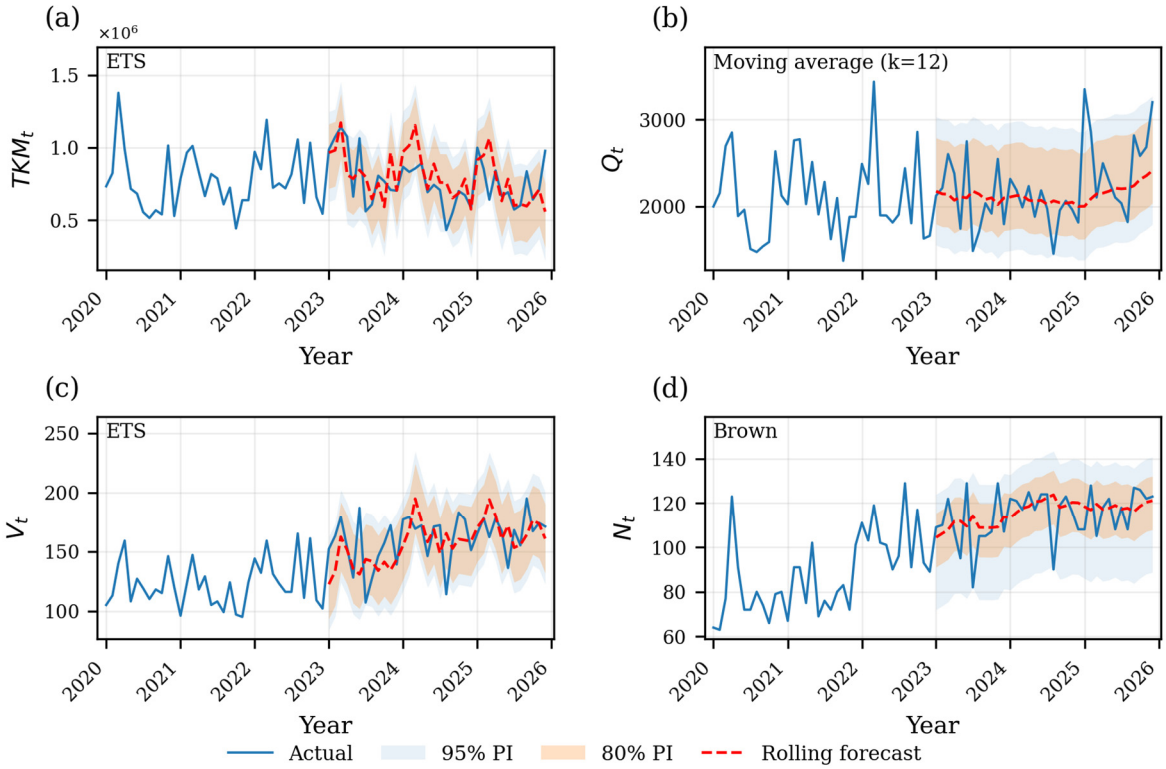

Figure 6. Rolling-origin forecasts with asymmetric empirical 80% and 95% prediction intervals: (**a**) transport work; (**b**) tonnage; (**c**) number of vehicles assigned to task execution; (**d**) number of transport tasks.

The results presented in Figure 6 complement the point forecast error metrics by showing the range and asymmetry of forecast uncertainty. Among the two count indicators, the number of transport tasks was characterised by narrower prediction intervals than the number of vehicle assignments. Tonnage and transport work exhibited broader uncertainty ranges in their respective units, consistent with their greater sensitivity to irregular changes in cargo mass and route length. For tonnage, the larger upper than lower error-quantile offsets indicate a greater risk of underestimating months with unusually high transport demand. From an operational perspective, the results support the use of flexible transport-capacity buffers and vehicle availability reserves, while the intervals should be interpreted together with the empirical coverage values reported in Table 4.

4. Discussion

4.1. Interpretation of Results and Implications for Transport Planning

The results indicate that monthly road freight workload should not be assessed using a single demand measure. The analysed indicators showed different temporal patterns, levels of dispersion, and sensitivity to irregular changes. This confirms that the multi-indicator approach is not only a descriptive extension of the analysis, but also a necessary condition for a more reliable interpretation of future transport workload.

The absence of a simple proportional relationship between the number of performed operations and the volume of transport work is particularly important. In operational prac-

tice, a similar number of transport tasks may correspond to different vehicle requirements, different cargo masses, and different tonne-kilometre values. Planning based only on the number of tasks could therefore lead to an incomplete assessment of future workload, because it would not capture changes in cargo structure, route length, and fleet utilisation. Conversely, transport work alone does not fully describe the organisational and resource-related burden on the system.

The seasonal analysis and decomposition results also show that the individual workload dimensions differ in terms of forecasting stability. The organisational and resource-related indicators were more stable, whereas tonnage and transport work were more sensitive to irregular changes in the structure of the transport operations performed. This is relevant for monthly planning, where decisions on vehicle availability, fleet reserves, driver allocation, and maintenance windows must be made before future demand is fully known.

The validation results demonstrate that no single forecasting model was optimal for all workload indicators. Brown's model was most effective for transport tasks, ETS for vehicles and transport work, and the 12-month moving average for tonnage. This result supports the conclusion that each workload dimension has its own time-series structure and should therefore be modelled separately. Selecting one universal model for the whole transport system would risk losing information about the specific behaviour of individual indicators. The rolling-origin forecast trajectories further confirm that the analysed models differed not only in aggregate error values, but also in their ability to reproduce short-term fluctuations and changes in the level of individual workload indicators.

A relevant methodological finding is that greater formal complexity did not automatically improve predictive performance. In the analysed case, simpler or adaptive models were competitive with more complex stochastic models. This is important from an implementation perspective, because transparent models that are easier to audit and maintain may provide sufficient forecasting quality for short monthly transport series affected by irregular operational variation.

Based on the MAE values reported in Table 3, the recommended models reduced the mean absolute forecast error relative to the naive model by 31.3% for the number of transport tasks, 27.6% for the number of vehicle assignments, 25.9% for transported cargo mass, and 22.7% for transport work. This corresponded to an average monthly reduction in forecast error of 3.95 transport tasks, 6.55 vehicle assignments, 109.2 t, and 35.6 thousand tkm, respectively. These results quantitatively confirm the improvement in planning accuracy relative to a forecast based on the most recent observation. However, they should not be interpreted as direct measures of cost savings or fleet size reduction, because these effects were beyond the scope of the study.

Prediction intervals further extend the operational interpretation of forecasts. Instead of basing decisions solely on point estimates, planners can define safety buffers around the expected monthly workload. For example, wider intervals for tonnage and tonne-kilometres reflect greater irregularity and indicate a greater need to maintain transport capacity reserves in months with uncertain material flow intensity or route structure. For this reason, prediction intervals increase the practical utility of the models. They allow forecasts to be interpreted not as single expected values, but as ranges of possible workload that can be directly used in planning fleet reserves and transport capacity.

The proposed approach may serve as an interpretable analytical component in transport planning and fleet management systems. Its practical value results from the integration of operational data, rolling-origin validation, model ranking, uncertainty assessment, and interpretation of results in terms of planning decisions. In this context, the obtained results may constitute a decision-support module for cyclical planning of transport task execution, assessment of fleet resource requirements, and definition of transport capacity buffers. As a

result, the procedure can support operational planning without the need to apply complex or difficult-to-interpret computational methods.

4.2. Limitations

The analysis has several methodological limitations. The dataset comprises 72 monthly observations, which is sufficient for a comparative assessment using rolling-origin validation, but limits the identification of long-term structural changes and complex seasonal effects. The models use only historical values of the analysed indicators and do not include exogenous variables, such as route length, cargo type, calendar effects, vehicle category, fleet availability, or extraordinary operational decisions.

Furthermore, the proposed framework was developed and evaluated using data from a single military transport system. Its application to commercial and other public-sector freight transport systems therefore requires further empirical validation and adaptation concerning their organisational characteristics, data availability, demand-generation mechanisms, and relevant external determinants.

The empirical prediction intervals were constructed from 36 one-month-ahead out-of-sample forecast errors. This sample size provides limited resolution for estimating the 0.025 and 0.975 quantiles and makes the tails of the nominal 95% intervals sensitive to individual extreme errors. The PICP and MPIW values should therefore be interpreted as sample-specific measures of empirical coverage and interval width, respectively. Accordingly, the prediction intervals should be interpreted as practical uncertainty ranges supporting monthly operational planning. Their applicability to other fleets and transport systems should be assessed through external validation and, where necessary, adaptation to system-specific operating conditions and data characteristics.

5. Conclusions

This article proposes a multi-indicator framework for forecasting the monthly workload of a road freight transport system and interprets it as a transparent decision-support module for smart logistics and fleet transport capacity planning. The framework includes four complementary workload indicators: the number of transport tasks, the number of vehicles assigned to task execution, tonnage, and transport work expressed in tonne-kilometres.

The empirical results confirmed that, between 2020 and 2025, the analysed military transport system was characterised by a marked increase in organisational and resource-related workload. The number of transport tasks increased by 50.2%, the number of vehicles by 36.2%, and tonnage by 24.3%, whereas the value of tonne-kilometres remained close to the 2020 level. This means that the increase in the number of operations and vehicles did not translate proportionally into an increase in transport work, confirming the need for a multi-indicator approach to assessing transport workload.

Rolling-origin validation showed that there was no single universal model that performed best for all workload dimensions. Brown's exponential smoothing model was the most accurate for the number of transport tasks, the ETS model for the number of vehicles and tonne-kilometres, and the 12-month moving-average model for tonnage. All recommended models achieved MASE values below 1, confirming their usefulness relative to the adopted reference scale.

Compared with the naive model, the recommended models reduced the mean absolute forecast error by 22.7–31.3%, depending on the analysed dimension of transport workload.

The inclusion of prediction intervals increases the planning value of the proposed approach. The intervals provide information on the expected uncertainty of one-month-ahead forecasts and can support the planning of fleet reserves, transport capacity buffers,

and risk-adjusted interpretation of future workload. The greater uncertainty observed for tonnage and tonne-kilometres indicates that the intensity of material flows and the structure of routes are more irregular than the organisational number of transport tasks.

From a practical perspective, the proposed framework can support monthly planning of fleet availability, transport capacity reserves, maintenance windows, and anticipated staff workload. The practical value of the proposed approach therefore lies in supporting more accurate monthly planning and providing quantifiable ranges of forecast uncertainty for each dimension of transport workload. These uncertainty ranges enable planners to distinguish between expected workload and the additional capacity that may be required to accommodate forecast variability. The forecasts do not directly determine detailed vehicle, maintenance, or staff schedules, but provide input for such decisions. The contribution of this article lies in providing a repeatable, data-driven, and interpretable forecasting layer that can be integrated into smart logistics planning systems.

Further research should include additional explanatory variables, including route length, cargo type, fleet structure and availability, and calendar effects. Future studies should also examine multivariate and machine learning models, weekly or daily data, and the integration of forecasting with optimisation models for vehicle allocation, maintenance scheduling, and staff planning. The indirect effects of economic, geopolitical, demographic, and legislative determinants on military transport workload should also be examined, particularly through their translation into transport tasks by defence planning, resource allocation, operational decisions, and applicable regulations. Given that the empirical analysis was conducted using data from a single military road transport system, the proposed framework should also be validated using data from commercial freight transport systems and other public-sector transport systems.

Author Contributions: Conceptualization, J.K.; methodology, J.K.; software, J.K.; validation, J.K.; formal analysis, J.K.; investigation, J.K.; resources, J.K.; data curation, J.K.; writing—original draft preparation, J.K.; writing—review and editing, J.K. and J.Z.; visualisation, J.K.; supervision, J.Z.; project administration, J.K. and J.Z.; funding acquisition, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Military University of Technology under Grant 143/UGB.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We thank M. Oszczypała for his helpful feedback on this work.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike information criterion
k	Length of the moving-average window, $k \in \{2, 3, 4, 6, 12\}$
MA	Moving average
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MASE	Mean absolute scaled error
N_t	Number of transport tasks per month t
n_t	Number of transport records assigned to month t

RMSE	Root mean square error
TKM_t	Transport work expressed in tonne-kilometres per month t
$tkm_{i,t}$	Tonne-kilometre value generated by the i-th transport task in month t
V_t	Number of vehicles assigned to transport task execution per month t
$v_{i,t}$	Number of vehicles assigned to the i-th transport task in month t
Q_t	Total tonnage scheduled for transport per month t
$q_{i,t}$	Tonnage assigned to the i-th transport task in month t
PICP	Prediction interval coverage probability
MPIW	Mean prediction interval width

Appendix A. Model-Parameter Selection and Sensitivity Analysis

Table A1 presents the sensitivity of moving-average forecasting accuracy to the window length, evaluated using MASE in the same expanding-window rolling-origin validation procedure as the main model comparison. Table A2 reports the ARIMA orders selected from the 47 candidate specifications using the initial 36-month training sample.

Table A1. Sensitivity of moving-average forecasting accuracy to window length.

Indicator	k = 2	k = 3	k = 4	k = 6	k = 12
V_t	0.5611	0.4934	0.4909	0.4699	0.4858
N_t	1.0429	0.9412	0.8856	0.8829	0.8438
Q_t	0.8092	0.7727	0.7989	0.7977	0.6930
TKM_t	0.8762	0.8165	0.8703	0.9052	0.7856

Note: MASE values were calculated from 36 one-month-ahead forecasts. The lowest value in each row identifies the moving-average window retained for the final model comparison.

Table A2. ARIMA orders selected using the initial 36-month training sample.

Indicator	Selected ARIMA Model	p	d	q	AIC
V_t	ARIMA(2,2,3)	2	2	3	254.622
N_t	ARIMA(1,2,3)	1	2	3	279.771
Q_t	ARIMA(1,2,3)	1	2	3	473.274
TKM_t	ARIMA(3,2,3)	3	2	3	826.687

Note: All 47 candidate specifications were successfully estimated for each indicator. The selected order was retained during rolling-origin validation, while the coefficients were re-estimated after each expansion of the training window.

Appendix B. Distributional Diagnostics of Rolling-Origin Forecast Errors

Table A3 reports the distributional diagnostics for the 36 one-month-ahead out-of-sample forecast errors generated by each recommended model. Figure A1 presents the corresponding normal Q–Q plots.

Table A3. Distributional diagnostics of the rolling-origin forecast errors.

Indicator	Recommended Model	R	Normality at $\alpha = 0.05$	Excess Kurtosis	Skewness	p-Value	Shapiro–Wilk W
V_t	ETS	36	Not rejected	0.453	−0.250	0.819	0.982
N_t	Brown	36	Rejected	1.721	−0.909	0.039	0.936
Q_t	MA (k = 12)	36	Not rejected	1.225	0.780	0.220	0.960
TKM_t	ETS	36	Not rejected	0.975	0.166	0.371	0.968

Note: A p-value below 0.05 indicates rejection of the null hypothesis of normality. Positive excess kurtosis indicates heavier tails than the normal distribution.

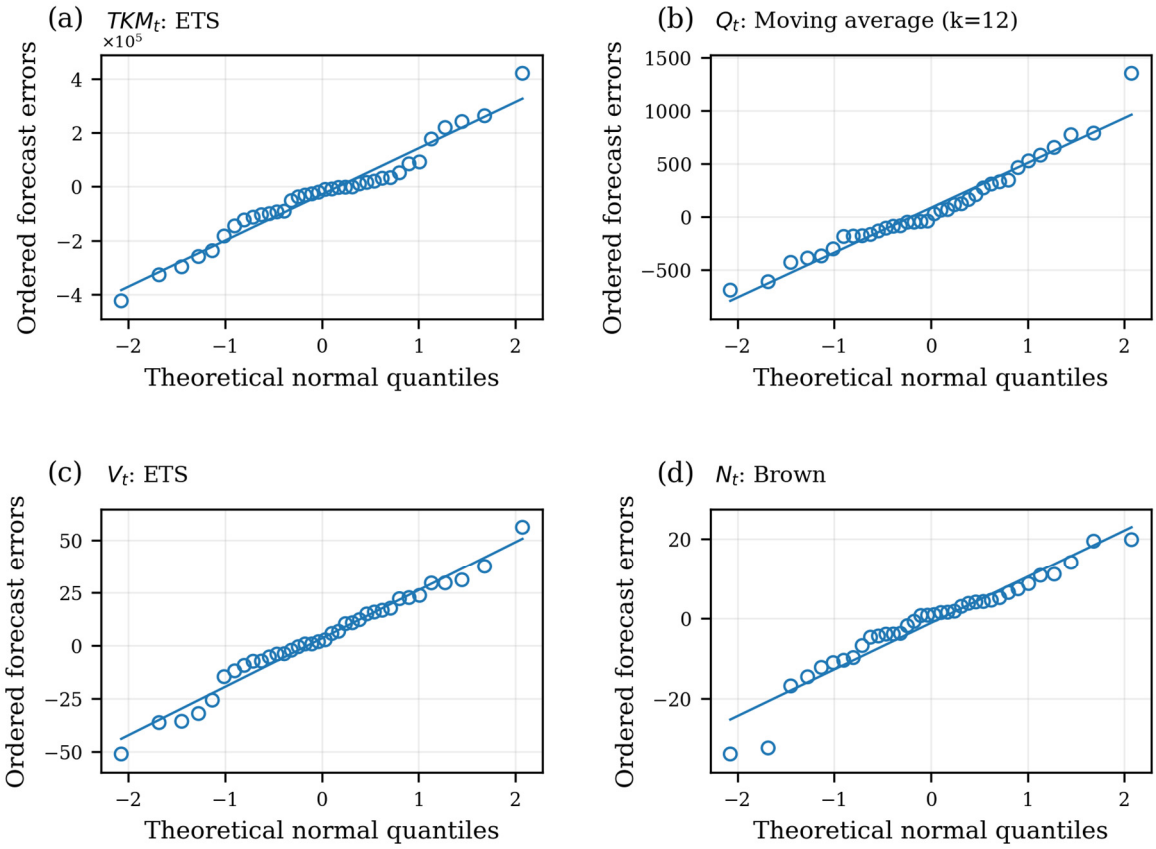

Figure A1. Normal Q–Q plots of the one-month-ahead rolling-origin forecast errors for the recommended models: (**a**) transport work; (**b**) tonnage; (**c**) number of vehicles assigned to task execution; (**d**) number of transport tasks.

References

1. Sonnleitner, B.; Kourentzes, N.; Ehrig, C.; Pflaum, A. Forecasting for Optimization in Road Freight Transport: A Review. *Transp. Res. Part E Logist. Transp. Rev.* **2025**, *204*, 104378. [CrossRef]
2. Nuzzolo, A.; Comi, A. Urban Freight Demand Forecasting: A Mixed Quantity/Delivery/Vehicle-Based Model. *Transp. Res. Part E Logist. Transp. Rev.* **2014**, *65*, 84–98. [CrossRef]
3. Al Hajj Hassan, L.; Mahmassani, H.S.; Chen, Y. Reinforcement Learning Framework for Freight Demand Forecasting to Support Operational Planning Decisions. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *137*, 101926. [CrossRef]
4. Grzelak, M.; Borucka, A.; Buczyński, Z. Forecasting the Demand for Transport Services on the Example of a Selected Logistic Operator. *Arch. Transp.* **2019**, *52*, 81–93. [CrossRef]
5. Liachovičius, E.; Šabanovič, E.; Skrickij, V. Freight Rate And Demand Forecasting In Road Freight Transportation Using Econometric And Artificial Intelli-Gence Methods. *Transport* **2023**, *38*, 231–242. [CrossRef]
6. Golden, B.; Assad, A.; Levy, L.; Gheysens, F. The Fleet Size and Mix Vehicle Routing Problem. *Comput. Oper. Res.* **1984**, *11*, 49–66. [CrossRef]
7. Hoff, A.; Andersson, H.; Christiansen, M.; Hasle, G.; Løkketangen, A. Industrial Aspects and Literature Survey: Fleet Composition and Routing. *Comput. Oper. Res.* **2010**, *37*, 2041–2061. [CrossRef]
8. Koç, Ç.; Bektaş, T.; Jabali, O.; Laporte, G. The Fleet Size and Mix Pollution-Routing Problem. *Transp. Res. Part B Methodol.* **2014**, *70*, 239–254. [CrossRef]
9. Russo, F.; Comi, A. A Modelling System to Simulate Goods Movements at an Urban Scale. *Transportation* **2010**, *37*, 987–1009. [CrossRef]
10. Lindsey, C.; Mahmassani, H.S.; Mullarkey, M.; Nash, T.; Rothberg, S. Industrial Space Demand and Freight Transportation Activity: Exploring the Connection. *J. Transp. Geogr.* **2014**, *37*, 93–101. [CrossRef]
11. Sultanbek, M.; Adilova, N.; Sładkowski, A.; Karibayev, A. Forecasting the Demand for Railway Freight Transportation in Kazakhstan: A Case Study. *Transp. Res. Interdiscip. Perspect.* **2024**, *23*, 101028. [CrossRef]
12. Karasu, T.; Leviäkangas, P.; Edwards, D.J. Road Freight Demand Forecasting Using National Accounts' Data—The Case of Cereals. *Agriculture* **2024**, *14*, 1980. [CrossRef]

13. Woodburn, A. Appropriate Indicators of Rail Freight Activity and Market Share: A Review of UK Practice and Recommendations for Change. *Transp. Policy* **2007**, *14*, 59–69. [CrossRef]
14. Dente, S.M.R.; Tavasszy, L. Policy Oriented Emission Factors for Road Freight Transport. *Transp. Res. Part D Transp. Environ.* **2018**, *61*, 33–41. [CrossRef]
15. Holden, R.; Xu, B.; Greening, P.; Piecyk, M.; Dadhich, P. Towards a Common Measure of Greenhouse Gas Related Logistics Activity Using Data Envelopment Analysis. *Transp. Res. Part A Policy Pract.* **2016**, *91*, 105–119. [CrossRef]
16. Alises, A.; Vassallo, J.M. Comparison of Road Freight Transport Trends in Europe. Coupling and Decoupling Factors from an Input–Output Structural Decomposition Analysis. *Transp. Res. Part A Policy Pract.* **2015**, *82*, 141–157. [CrossRef]
17. Zhu, F.; Wu, X.; Gao, Y. Decomposition Analysis of Decoupling Freight Transport from Economic Growth in China. *Transp. Res. Part D Transp. Environ.* **2020**, *78*, 102201. [CrossRef]
18. Kotsios, P.; Folinas, D. Analysis and Comparison of Road Freight Transport Cost in 20 European Countries. *Int. J. Appl. Logist.* **2020**, *10*, 13–26. [CrossRef]
19. Camisón-Haba, S.; Clemente-Almendros, J.A. A Global Model for the Estimation of Transport Costs. *Econ. Res.-Ekon. Istraživanja* **2020**, *33*, 2075–2100. [CrossRef]
20. Milewski, D.; Milewska, B. Efficiency of the Consumption of Energy in the Road Transport of Goods in the Context of the Energy Crisis. *Energies* **2023**, *16*, 1257. [CrossRef]
21. Vajihi, M.; Ricci, S. Energy Efficiency Assessment of Rail Freight Transport: Freight Tram in Berlin. *Energies* **2021**, *14*, 3982. [CrossRef]
22. Kaack, L.H.; Vaishnav, P.; Morgan, M.G.; Azevedo, I.L.; Rai, S. Decarbonizing Intraregional Freight Systems with a Focus on Modal Shift. *Environ. Res. Lett.* **2018**, *13*, 083001. [CrossRef]
23. Jacyna, M.; Wasiak, M.; Kłodawski, M.; Lewczuk, K. Simulation Model of Transport System of Poland as a Tool for Developing Sustainable Transport. *Arch. Transp.* **2014**, *31*, 23–35. [CrossRef]
24. Korkmaz, E.; Akgungor, A.P. Estimation of Passenger-Kilometer and Tonnekilometer Values for Highway Transportation in Turkey Using the Flower Pollination Algorithm. *Sci. J. Silesian Univ. Technol. Ser. Transp.* **2018**, *98*, 45–52. [CrossRef]
25. Mamede, F.P.; Da Silva, R.F.; De Brito Junior, I.; Yoshizaki, H.T.Y.; Hino, C.M.; Cugnasca, C.E. Deep Learning and Statistical Models for Forecasting Transportation Demand: A Case Study of Multiple Distribution Centers. *Logistics* **2023**, *7*, 86. [CrossRef]
26. Ma, F.; Wang, S.; Xie, T.; Sun, C. Regional Logistics Express Demand Forecasting Based on Improved GA-BP Neural Network with Indicator Data Characteristics. *Appl. Sci.* **2024**, *14*, 6766. [CrossRef]
27. Li, Y.; Wei, Z. Regional Logistics Demand Prediction: A Long Short-Term Memory Network Method. *Sustainability* **2022**, *14*, 13478. [CrossRef]
28. Asgarpour, S.; Hartmann, A.; Gkiotsalitis, K.; Neef, R. Scenario-Based Strategic Modeling of Road Transport Demand and Performance. *Transp. Res. Rec. J. Transp. Res. Board* **2023**, *2677*, 1415–1440. [CrossRef]
29. Tjandra, S.; Kraus, S.; Ishmam, S.; Grube, T.; Linßen, J.; May, J.; Stolten, D. Model-Based Analysis of Future Global Transport Demand. *Transp. Res. Interdiscip. Perspect.* **2024**, *23*, 101016. [CrossRef]
30. Urazán-Bonells, C.F.; Rondón-Quintana, H.A.; Zafra-Mejía, C.A. Correlation between Sectoral GDP and the Values of Road Freight Transportation in Colombia. *Economies* **2024**, *12*, 205. [CrossRef]
31. Gonzalez, J.N.; Camarero-Orive, A.; González-Cancelas, N.; Guzman, A.F. Impact of the COVID-19 Pandemic on Road Freight Transportation—A Colombian Case Study. *Res. Transp. Bus. Manag.* **2022**, *43*, 100802. [CrossRef] [PubMed]
32. International Transport Forum. *ITF Transport Outlook 2023*; ITF Transport Outlook; OECD Publishing: Paris, France, 2023.
33. Kress, M. *Operational Logistics: The Art and Science of Sustaining Military Operations*; Management for Professionals; Springer International Publishing: Cham, Switzerland, 2016.
34. NATO Standardization Office. *Allied Joint Doctrine for Logistics*; B, Version 1; Standardization Office: Brussels, Belgium, 2018.
35. Chuwang, D.D.; Chen, W. Forecasting Daily and Weekly Passenger Demand for Urban Rail Transit Stations Based on a Time Series Model Approach. *Forecasting* **2022**, *4*, 904–924. [CrossRef]
36. Schmid, L.; Roidl, M.; Kirchheim, A.; Pauly, M. Comparing Statistical and Machine Learning Methods for Time Series Forecasting in Data-Driven Logistics—A Simulation Study. *Entropy* **2025**, *27*, 25. [CrossRef]
37. Makridakis, S.; Hibon, M. The M3-Competition: Results, Conclusions and Implications. *Int. J. Forecast.* **2000**, *16*, 451–476. [CrossRef]
38. Hyndman, R.J.; Koehler, A.B.; Snyder, R.D.; Grose, S. A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods. *Int. J. Forecast.* **2002**, *18*, 439–454. [CrossRef]
39. Tashman, L.J. Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review. *Int. J. Forecast.* **2000**, *16*, 437–450. [CrossRef]
40. Bergmeir, C.; Benítez, J.M. On the Use of Cross-Validation for Time Series Predictor Evaluation. *Inf. Sci.* **2012**, *191*, 192–213. [CrossRef]
41. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd ed.; OTexts: Melbourne, Australia, 2021.
42. Brown, R.G. *Statistical Forecasting for Inventory Control*; McGraw-Hill: New York, NY, USA, 1959.

43. Holt, C.C. Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. *Int. J. Forecast.* **2004**, *20*, 5–10. [CrossRef]
44. Gardner, E.S. Exponential Smoothing: The State of the Art—Part II. *Int. J. Forecast.* **2006**, *22*, 637–666. [CrossRef]
45. Box, G.E.P.; Reinsel, G.C.; Ljung, G.M.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; John Wiley and Sons Inc.: Hoboken, NJ, USA, 2015.
46. Makridakis, S. Accuracy Measures: Theoretical and Practical Concerns. *Int. J. Forecast.* **1993**, *9*, 527–529. [CrossRef]
47. Hyndman, R.J.; Koehler, A.B. Another Look at Measures of Forecast Accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]
48. Gorzelańczyk, P.; Sokolovskij, E. Using Neural Networks to Forecast the Amount of Traffic Accidents in Poland and Lithuania. *Sustainability* **2025**, *17*, 1846. [CrossRef]
49. Shapiro, S.S.; Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591–611. [CrossRef]
50. Hyndman, R.J.; Fan, Y. Sample Quantiles in Statistical Packages. *Am. Stat.* **1996**, *50*, 361–365. [CrossRef]
51. Lee, Y.S.; Scholtes, S. Empirical Prediction Intervals Revisited. *Int. J. Forecast.* **2014**, *30*, 217–234. [CrossRef]
52. Gneiting, T.; Balabdaoui, F.; Raftery, A.E. Probabilistic Forecasts, Calibration and Sharpness. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2007**, *69*, 243–268. [CrossRef]
53. Pearce, T.; Brintrup, A.; Zaki, M.; Neely, A. High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach. In Proceedings of the 35th International Conference on Machine Learning (ICML), Stockholm, Sweden, 10–15 July 2018; Volume 80, pp. 4075–4084.
54. Hyndman, R.J.; Khandakar, Y. Automatic Time Series Forecasting: The Forecast Package for R. *J. Stat. Soft.* **2008**, *27*, 1–22. [CrossRef]
55. Arachchige, C.N.P.G.; Prendergast, L.A.; Staudte, R.G. Robust Analogs to the Coefficient of Variation. *J. Appl. Stat.* **2022**, *49*, 268–290. [CrossRef] [PubMed]