



# Determinants of irregular demand for regional rail passenger services – case study of High Tatras in Slovakia

Martin Kendra<sup>1</sup> · Oľga Blažeková<sup>1</sup> · Mária Vojteková<sup>1</sup>

Accepted: 18 March 2024  
© The Author(s) 2024

## Abstract

The demand for public transport by tourists increases significantly in tourist-attractive destinations. This is in addition to regular passengers commuting to school and work. The level of irregular demand is influenced by several factors related to the characteristics of the day of the week, the period of the year, and the current weather. The main goal of the paper is to verify which factors most influence the irregular demand for transport in a tourist-attractive area to ensure operational planning of public passenger transport. Thanks to this, it is possible to ensure sufficient capacity and, at the same time, the efficiency of the operation of public passenger transport. The paper analyzes the main determinants of the irregular demand for regional public rail passenger transport in the High Tatras region of Slovakia. Multiple linear regressions were used to model the number of irregular passengers. The variables representing the day of the week, the attractiveness of the period, and the holiday were found to be the most significant. The variables describing the weather such as maximum daily temperature, precipitation, clouds, and wind had less influence. The obtained mathematical models for forecasting the irregular demand for public passenger transport can help optimize the timetable's operational setting and the train sets' size.

## Highlights

- Demand for public transport grows the most on Saturdays and Fridays during the week.
- Sunday is not a statistically significant determinant affecting the transportation demand.
- Period attractiveness is the most important determinant influencing the demand.
- Clouds, wind and precipitation comparatively reduce the demand for public transport.
- Distrust of public transport persists for a long time after anti-pandemic measures ended.

**Keywords** Regional railway transport · Passenger demand · Day of week · Weather condition · Pandemic



## Introduction

The current problem facing mountain resorts around the world is how to manage the growing number of visitors with the least possible impact on the natural environment. Public transport services are becoming a key area, as are efforts to reduce environmentally unfriendly car traffic. This challenge is also faced by the resorts in the High Tatras, where the number of traffic-critical days increases every year along with the growth of traffic.

The High Tatras are the highest mountains in Slovakia. They are located in the northern part of Slovakia, on the border with Poland. In Slovakia, they cover an area of 260 km<sup>2</sup>. In 1949, they were declared a national park, and in 1993, together with the Polish part of the Tatras, they were declared a “Tatra Biosphere Reserve” by UNESCO. The High Tatras are the most attractive tourist area in Slovakia. It has all the necessary infrastructure and the facilities to provide a complex of services throughout the year (Holek 2019).

The number of visitors to the High Tatras is high and constantly growing, more than three million people visit this area every year. This influx of visitors, which includes both same-day visitors and overnight visitors, causes more and more traffic problems.

The main transport system in the High Tatras is the Tatra Electric Railway (Tatranské elektrické železnice – TEŽ), which connects the main centers of the High Tatras - Štrba, Štrbské Pleso, Starý Smokovec, Tatranská Lomnica and Poprad-Tatry. This main transport system consists of a single-track adhesive conventional railway (35 km long) and a single-track cog railway (5 km long)(High-Tatras. travel). TEŽ is connected to the national conventional railway network at Poprad-Tatry and Štrba stations, where passengers can change from national and international long-distance trains to TEŽ trains. All TEŽ trains provide regional transport in the region approximately every hour. The TEŽ trains are used by local residents to go to school and to work, but a much larger number of passengers are tourists who come to the High Tatras for recreation. Passenger transportation on the TEŽ railway lines is provided by 425.9 series electric carriages manufactured by the consortium of companies STADLER, Adtranz and ŽOS Vrútky a. s. The capacity of the carriages is 88 seated passengers each. The units can be rolled up during operation. The TEŽ lines also use electric units of the 495.95 series manufactured by the Stadler company with a capacity of 91 seated passengers. In times of increased demand for transport, the transport infrastructure allows for additional reinforcement trains to be added to the timetable.

Additional regional transport in the High Tatras region is also provided by buses. They connect local villages and settlements; therefore, bus transport is mainly intended for regular transport of residents to work and school. Tourists use this transport only in exceptional cases.

The High Tatras region tries to regulate the number of passenger cars entering its territory by providing a limited number of parking places. Tourists who have booked accommodation in the High Tatras are allowed to use individual transport, while tourists who come to the region by car park their car near their accommodation (hotel, pension). They use TEŽ mainly for further transport in the region.

Based on the above-mentioned facts, it can be concluded that TEŽ is the main transport system for transport of irregular passengers in the High Tatras region.

In order to improve the situation, it is necessary to provide visitors with sufficient quality and comfortable ecological public passenger transport services, which represents the regional railway passenger transport system in the High Tatras region (Tatry mountain

resorts, a.s 2018). A large number of visitors is also evidenced by the fact that 30,232 visitors visited the Tatra National Park in one day (August 2020), which is almost 4,000 more than the previous record from four decades ago (Dobré noviny 2020).

The public transport timetable is prepared for the whole year as a long-term operational plan. The problem is the uneven distribution of demand for public passenger transport on the days of the week and throughout the year, which complicates the short-term operational planning of the timetable and the number of vehicles needed for public transport. The paper analyzes the main factors influencing the irregular demand for public transport related to the day of the week, weather, and pandemic measures. The aim of the paper is to verify which of these factors most influence the irregular demand for public passenger transport. Based on the obtained input data, a mathematical model using multiple linear regression has been developed, which can be used to predict the expected demand for public transport based on the day of the week, weather forecast and current pandemic measures. Based on the results found in this paper, public transport operators can operationally set the capacity of public transport so that it is sufficiently attractive to passengers and at the same time economically efficient.

## State of the art

Effective management of rail systems is critical to both operational efficiency and passenger service satisfaction. Expert system models are built from demand forecasting rules based on the knowledge of a human expert, but it is difficult to transform the knowledge of an expert into mathematical models. Another option is to analyze the transportation demand using mode choice models as multinomial probit model (Bilal et al. 2023) or multinomial logit model (Shi et al. 2023). Regression models are powerful tools for characterizing the relationship between demands and other important factors. Passenger demand forecasting plays a critical role in decision-making and planning. Travel demand forecasting is an attempt to predict and quantify the future travel patterns. As a result of the literature review, it has been observed that there are only a few studies exist for rail travel demand. From the mathematical point of view, various methods have been used to predict passenger demand, such as regression analysis techniques (Odgers and Schijndel 2011), autoregressive integrated moving average models (Cyril et al. 2018), fuzzy set theory (Dou et al. 2013), artificial neural network and machine learning algorithm technique (Nar and Arslankaya 2022) or chaos theory (Picano et al. 2019). An interesting approach has been taken by Zhao et al. (2011), who used wavelet analysis and a neural network to propose a nonlinear model for predicting the rate of transit passenger flow. They regarded the transit passenger flow data as a signal with a given length and concluded that the considered data had chaos characteristics.

Unlike the time series models which use past demand data to forecast demand for future periods, regression analysis is a statistical forecasting method that uses the relationship between variables. Many empirical studies have analyzed the rail travel demand as a function of economic conditions. The exponential function and its transformation to linear regression was used by Rahman and Balijepalli (2016) to estimate suburban rail fare elasticity on the determinants of public transport demand in Indian railways (fare; per capita net state domestic product; petrol price; population of the city area; road vehicle population; (rail) vehicle kilometers). Rail fare, GDP per capita, fuel price, and route-density were

found to be the main determining factors of rail travel demand in the long run in Pakistan by Zamir Khan and Naheed Khan (2021). The relationship between the passenger rail demand in Perth, Australia, and 6 factors (rail fare, per capita income, fuel price index, population, number of kilometers travelled per year, number of accidental deaths related to the rail sector) were expressed by exponential regression model by Wijeweera and Charles (2013). The main factors (population, operation mileage, travel speed, the number of railway staff, and ownership of cars) influencing rail passenger volume in China were described by Zhi (2018) and a linear regression model was used to forecast the future development. A regression model of transport requirements in railway transport depending on economic indicators in Slovakia was used by Danis et al. (2016) and the main factors: GDP, average monthly wage, car ownership, and price of railway transport services were found.

It is a general opinion that impact of climate change and weather conditions on the transport system has not received the necessary attention (Clifton et al. 2014). Most of the models used in transport planning and traffic management do not include parameters indicating the impact of weather conditions, so they are adjusted to ideal weather conditions during the spring and autumn (Petrović et al. 2020). A systematic and comprehensive overview concerning the impact of weather conditions on daily travel activities was worked out by Böcker et al. (2013). The influence of weather on travel behaviour depends on differences in transport infrastructure. Travel behavior in countries or regions that are predominantly car-dependent is less sensitive to daily weather variations compared with those countries or regions where people are more dependent on non-motorized transport (Sabir 2011). The majority of research papers analyze the weather impact on non-motorized transport due to its greatest exposure to weather conditions, e.g. (Petrović et al. 2020), (Böcker et al. 2013). The influence of extreme weather conditions, especially extreme heat, frost, storm, fog, rain, and snow was analyzed by Cools et al. (2010) and Tuan and Huong (2020). A large number of research papers deal with the impact of weather on street network capacity and probability of congestion, such as (Ivanovic and Jović 2017) and (Tao et al. 2018), but the relationship between weather and rail transport has been the subject of few studies. The influence of weather on urban rail transit ridership in four major cities in China was studied by Wang et al. (2020) and models with different combinations of temperature and weather type factors were created to determine the weather effect on daily ridership rate using the linear regression method. The results of weather effect on travel demand obtained in one region cannot be directly applied to another one due to different weather conditions, cultural and socio-economic characteristics, so it is important to look at local conditions. The factor analysis was used by Zhu (2018) to find factors affecting the demand for railway passenger transport in various provinces in the eastern part of China and the results showed that there is an imbalance in the demand for railway passengers and a large difference among different regions.

The interactions between time allocation (activity duration and travel time), travel demand (number of trips), and mode choice (slow-mode share) were investigated by Liu et al. (2014) using combined a weather and travel survey. The choice of destination and departure time is highly dependent on the season and weather conditions which makes the modeling of leisure travel more difficult than recurring work trips (Haberl and Neuhold 2012). In order to operate the transportation system efficiently, it is necessary to have knowledge about the impact of the weather. While the impact of weather is not expected to dominate travel demand (e.g. work trips cannot be easily omitted), trips may be delayed or different

modes may be chosen (Rudloff et al. 2015). The weather conditions can be studied from the point of view of forecasted ones - Bursa et al. (2022a) pointed out the low impact of forecasted bad weather on activities in the Austrian Alps and found out with their research using questionnaires that in 5.92% of cases in summer and 1.52% in winter, people were forced to choose an alternative activity due to bad weather. Another approach considers the influence of measured weather conditions on tourists' transport mode choices. Bursa et al. (2022b) investigated in the Austrian Alps the effects of weather elements. They constated that temperature, sky overcast, and snow cover had no impact on tourists' transport mode choices in any season. But precipitation did and it was observed for walking. It was understandably negative in winter, yet surprisingly positive in summer, which they attributed to changes in activity and destination choices when it was raining. Similarly unexpected was the positive effect of wind on "cycling" in summer, which the authors explained by the phenomenon of foehn wind that carries dry weather. The effects of climate change on tourist mobility in mountain areas (Autonomous Province of South Tyrol - Italy) were investigated by Cavallaro et al. (2017) distinguishing between infrastructure, transport operation and travel demand.

## Materials and methods

The research was developed in three main stages:

1. data collection and pre-processing for considered models,
2. modeling by multiple linear regression of various combinations of independent variables,
3. analysis of the results.

The raw data were acquired from TEŽ as the number of tickets sold in 2019 and 2020. These were single-ride, seven-day, three-day, and one-day tickets sold through electronic cash desks on a specific date. It did not include commuter tickets and tickets purchased without a specific date. It means our interest was to model and predict daily numbers of traveling visitors to a tourist-attractive area. The input data were numbers of passengers calculated from the counts of sold tickets based on the TEŽ experts' recommendation. It is assumed that three train journeys correspond to each day of ticket validity on average.

When assessing the influence of meteorological elements, it was based on data provided by the Slovak Hydrometeorological Institute (SHMÚ). The data are from the meteorological station in Tatranská Polianka, which characterizes the weather in the central part of the High Tatras. The data provided included:

- the minimum and maximum daily (from 0:00 to 24:00) air temperature [°C];
- the daily cloud cover at 07:00 a.m., 02:00 p.m. and 09:00 p.m. (indicated by a 10-degree scale, 0 being cloudless and 10 being completely overcast);
- the daily wind speed at 07:00 a.m., 02:00 p.m. and 09:00 p.m. [m.s-1];
- total daily precipitation [mm];
- daily statistics on the occurrence of storm phenomena (on the following scale: no storm, storm at the weather station, distant and very distant storm, lightning).

The daily numbers of traveling visitors to TEŽ, depending on explanatory variables, were modeled by a linear regression model.

Linear regression model, describing the relationship between dependent variable  $Y$  and  $k$  independent variables  $X_j, j = 1, 2, \dots, k$ , is given by linear Eq. (1).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon. \tag{1}$$

The element  $\epsilon$  in formula (1) expresses random errors that represent a random component of the regression model. These errors include variable measurement errors, random factors, ...

Unknowns parameters  $\beta_0, \dots, \beta_k$  are found by data  $y_i, x_{i1}, \dots, x_{ik}, i = 1, 2, \dots, n$ ; where  $n$  is the number of observations, it means by data of variables  $Y, X_1, \dots, X_k$ . Estimates of these parameters are denoted  $b_0, \dots, b_k$  and the estimated regression model is in the form

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k. \tag{2}$$

In the formula (2) the coefficient  $b_0$  is the intercept, which often has no logical interpretation. It is a value of  $Y$  when all other parameters are set to 0, which is unrealistic in many cases.

Coefficients  $b_j, j = 1, 2, \dots, k$  are denoted regression coefficients. They indicate which increase ( $b_j > 0$ ) or decrease ( $b_j < 0$ ) of the mean value of the dependent variable  $Y$  corresponds to the unit increment of the explanatory variable  $X_j$  with unchanged values of the other explanatory variables (*ceteris paribus*).

For each variable included in the model, it is necessary to assess whether it is statistically significant or can be omitted from the model without affecting its quality. This significance is assessed using the significance test of the regression coefficient of a given variable, which is performed at the significance level  $\alpha$  and is evaluated based on the p-value of the test. If the p-value is less than  $\alpha$ , the coefficient is statistically significant, and the given explanatory variable is included in the model justifiably. Explanatory variables whose coefficients are not statistically significant must be dropped from the model.

The adjusted coefficient of determination  $R_a^2$  allows comparison of different regression models by considering that the coefficient of determination tends to increase as the number of variables increases (Rimarčík 2007)

$$R_a^2 = 1 - (1 - R^2) \frac{n - 1}{k - 1}, \tag{3}$$

where  $R^2$  is the coefficient of determination,  $n$  is the number of observations,  $k$  is the number of independent variables.

Mean absolute percentage error (MAPE) is commonly used as a loss function for regression problems and in model evaluation, because of its very intuitive interpretation in terms of relative error (Wikipedia 2023)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right| \cdot 100\% \tag{4}$$

In the formula (3), the  $x_i$  is the real value and  $\hat{x}_i$  is the predicted value,  $n$  is the number of observations. A lower value indicates more satisfactory predictive ability and higher accuracy.

Lewis (1982) interprets the MAPE results as a way to judge the accuracy of the forecast, where less than 10% is a highly accurate forecast; 10 – 20% is a good forecast; 20 – 50% is a reasonable forecast; and more than 50% is an inaccurate forecast.

## Results and discussion

The total number of observed passengers on TEŽ in 2019 was 1 247 854. Daily numbers of passengers showed considerable variability: the mean was 3 419 passengers with a standard deviation of 2 582; the minimum value was 505, and the maximal one was 10 498. The median was 2 554. Therefore, the goal was to find the factors that affect this variability.

Two models were created to calculate and forecast daily numbers of passengers:

- a model without weather,
- a model with weather.

### Model without weather

The first simple model consisted of a dependent variable of the daily number of passengers and explanatory variables representing the days of the week: “Friday”, “Saturday”, “Sunday”, and “Monday through Thursday”. This model was statistically significant (p-value  $6.27 \cdot 10^{-5}$ ) but explained only 8% of the variability in the number of passengers. Only the coefficient for the variable “Saturday” was statistically significant according to the p-value. The estimated daily numbers of passengers were only periodically repeated and did not take into account the season (period of the year). Therefore, additional variables were added to the model to account for holidays (“Holiday”) and seasonality (“Period attractiveness”).

The variable “Holiday” expressed whether it was a public holiday, a school holiday or it was a normal day. Initially, “School holiday” and “Public holiday” were marked as different variables, but at the significance level  $\alpha = 0.05$  the variable representing a public holiday was statistically insignificant, so the variables were merged into one dummy variable “Holiday”, with the value of 1 if the day was a school holiday or a public holiday and zero otherwise.

The “Period attractiveness” was coded on a scale of 1–5, with the value 5 for the most attractive period of the summer holidays and the period from Christmas to New Year’s Eve. We have built several models considering variations in passenger numbers at different periods during the year and depending on conditions for touring and skiing. We assigned numerical values to different periods of the year partly in a subjective manner (after consulting with experts in the field of tourism in the High Tatras), and partly based on the effects for the model fit (Table 1).

The considered model explained 84.66% ( $R^2 = 0.8466$ ) of the variability in the number of passengers. All included variables, except the variable “Sunday”, were statistically significant. Therefore, the model was subsequently adjusted for days of the week, with dummy

**Table 1** Assignment of period attractiveness values

Period	Value	Period	Value	Period	Value
1.1. – 6.1. New Year holiday	4	22.4. – 31.5.	2	16.9. –30.9.	3
7.1. – 10.3.	2	1.6. – 14.6.	3	1.10. –3.11.	2
11.3. – 15.4.	1	15.6. – 30.6.	4	4.11. –13.12.	1
16.4. – 18.4.	2	1.7. – 1.9. Summer holiday	5	14.12. –26.12.	2
19.4. – 21.4. Easter holiday	4	2.9. – 15.9.	4	27.12. –31.12. Christmas holiday	5

**Table 2** ANOVA values of the model without weather

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	2.05E+09	5.13E+08	492.4575	1.6E-144
Residual	360	3.75E+08	1,041,289		
Total	364	2.43E+09			

**Table 3** Statistical characteristics of coefficients for the model without weather

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>p-value</i>
Intercept	-869.142	121.6893	-7.1423	5.12E-12
Holiday	1147.368	143.358	8.0035	1.68E-14
Period attractiveness	1350.024	48.09462	28.0702	1.13E-92
Friday	587.285	154.9904	3.7892	0.000177
Saturday	1765.384	155.0339	11.3871	7.27E-26

variables “Friday” and “Saturday” and the reference variable being the group of days “Sunday to Thursday”.

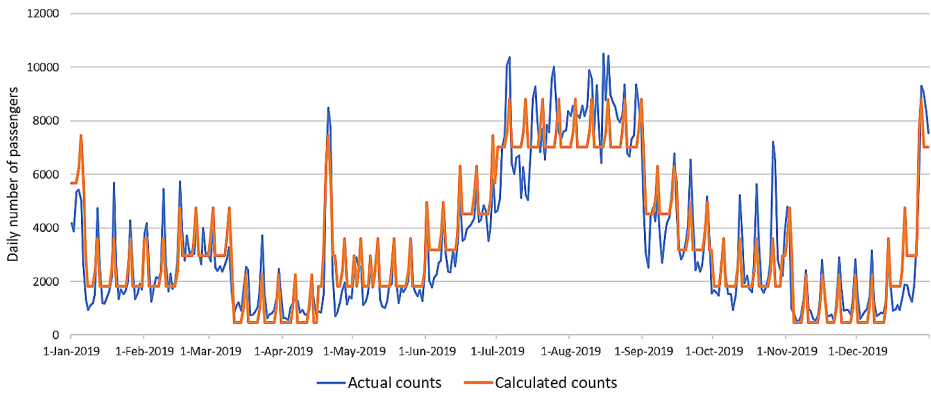
The  $R^2$  for the resulting model was 0.8454 and the value of the indicator MAPE is 29.79%. Detailed results of the model without weather are in Tables 2 and 3. In the p-value column (Table 3), all the values are less than the significance level of 0.05, so all the coefficients of the regression model are statistically significant.

The coefficient 1147.368 for the “Holiday” can be interpreted as 1 147 more passengers were traveling on a school holiday or national holiday compared to a normal day (*ceteris paribus*). The coefficient for “Period attractiveness” means that if the period had the value of the period attractiveness greater than one, an increase in the number of passengers was by 1 350 (*ceteris paribus*). The coefficient for the variable “Friday” means that on Friday the number of passengers was 587 higher compared to the number of passengers on other working days or on Sunday (*ceteris paribus*). Similarly, 1 765 more passengers traveled on Saturday compared to other working days or on Sunday (*ceteris paribus*).

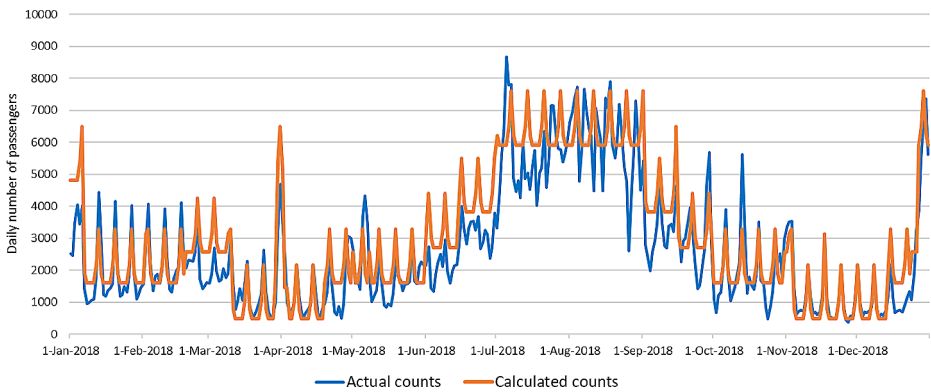
Figure 1 shows the differences between the real values of the daily number of passengers and values calculated by the model without weather during the year 2019.

To test our model without weather, we used data on the daily number of passengers in 2018 unaffected by Covid. Only the change in the Easter holiday date was taken considered for the “Season Attractiveness” variable. The value of the indicator MAPE is 34.05%. A





**Fig. 1** Time series of daily passenger flow in 2019 – model without weather



**Fig. 2** Time series of daily passenger flow in 2018 – model without weather

comparison of the actual counts and calculated ones using the model without weather can be seen in Fig. 2.

**Model with weather**

A model with weather initially contained 9 variables and was created from the model without weather by supplementing the influence of meteorological elements, with new variables expressing:

- clouds at 02:00 p.m.,
- wind speed at 02:00 p.m.,
- total daily precipitation,
- storms,
- maximum daily air temperature.

Data provided by the SHMU for cloudiness was in the form of 10 degrees. We considered several options for coding the variable as comparison with the monthly average or dummy variable with values of 1 - “small” cloudiness, 0 - “large” cloudiness. Statistically, the most convenient model was the model of three values: 0 - almost cloudless (values 0 and 1), 1 - average cloudiness (values 2 to 8), and 2 - large cloudiness (values 9 and 10 from the scale).

Cloud cover data provided by the SHMU contained ten levels. We considered several options for coding the variable: by comparing with the monthly average or as a dummy variable with values of 1 for “small” cloudiness, 0 for “large” cloudiness. We found that a transformation to three levels: 0 - almost cloudless (values 0 and 1), 1 - medium cloudiness (values 2 to 8) and 2 - heavy cloudiness (values 9 and 10 from the scale), provides the best fit to the data.

The dummy variable “Wind” was set to the values: 1 - strong wind (above  $5.5 \text{ m.s}^{-1}$ ), otherwise 0.

The dummy variable “Storm” takes the value 1 if a strong storm occurred during the day, otherwise, it is zero. The occurrence of storms recorded by SHMU was minimal for the observed period, so this variable is not statistically significant, which was confirmed by the p-value for this variable (0.175).

Temperature was the most problematic variable. The first reason was that only minimum temperatures and maximum temperatures measured in 24 h were available, and not the predicted values that tourists are interested in. The second reason was that it made no sense to take the temperature as a numerical variable since e.g.,  $12^\circ$  in February and June does not have the same effect. Likewise, we did not use data on the minimum daily temperature, as this is usually reached during the night. We created multiple models with different coding of the variable “Maximum daily temperature”. The resulting one was based on comparison of the given value with the average maximum daily temperature from the long-term observation (meteoblue 2022). The dummy variable “Maximum daily temperature” is 1 if the temperature is higher than the average maximum monthly temperature, otherwise, the value is zero. The variable was statistically significant (p-value 0.036). Another possibility was coding the maximum daily temperature into a dummy variable with values of 0 if the maximum daily temperature was within the interval of average minimum and maximum daily temperature in the given month, otherwise, the value was 1, but in this case, the variable was statistically insignificant (p-value 0.078).

To create a model with weather, the method of backward elimination of statistically insignificant variables based on the p-value was used.

The obtained model had a coefficient of determination  $R^2 = 0.8641$ , i.e., it explains up to 86.41% of the variability of the number of passengers. The value of the indicator MAPE was 28.97%. Detailed results of the model with respect to weather are in Tables 4 and 5. In the p-value column (Table 5), all the values are less than the significance level of 0.05, so all the coefficients of the regression model are statistically significant.

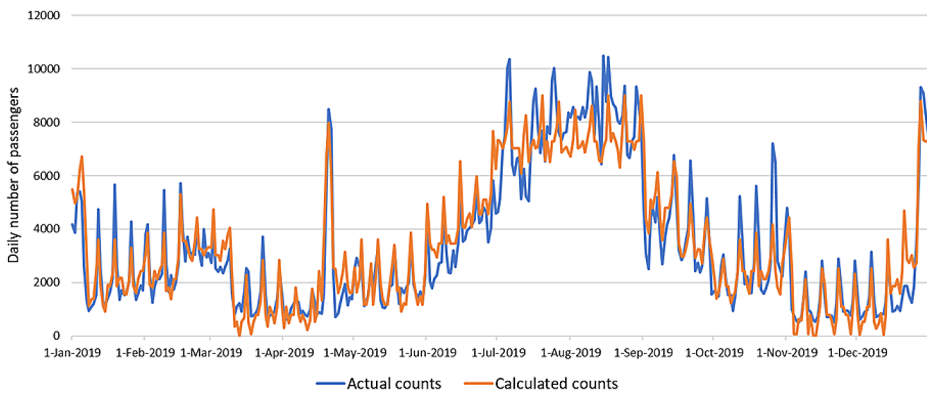
The interpretation of the results for the day of the week, period attractiveness, and holiday is the same as in the model without weather. The value of “clouds” 1 (moderate clouds)

**Table 4** ANOVA values of model with weather

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	8	2.1E+09	2.62E+08	282.9088	3.3E-149
Residual	356	3.3E+08	926,222		
Total	364	2.43E+09			

**Table 5** Statistical characteristics of coefficients for model with weather

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>p-value</i>
Intercept	-490.685	187.1279	-2.62219	0.009112
Holiday	1138.408	135.424	8.406256	1.03E-15
Period attractiveness	1337.658	46.00175	29.07843	4.63E-96
Friday	544.941	147.1422	3.703495	0.000246
Saturday	1737.632	146.4757	11.86294	1.4E-27
Clouds	-307.673	90.51511	-3.39914	0.000753
Wind	-244.776	117.843	-2.07713	0.038506
Precipitation	-202.285	78.53734	-2.57565	0.010408
Maximum daily temperature	245.662	116.6806	2.105424	0.035954



**Fig. 3** Time series of daily passenger flow in 2019 - model with weather

led to a decrease in the number of passengers by 308 compared to the value 0 (cloudless weather). Due to the value 2 (heavy clouds), a decrease of approximately 616 passengers compared to cloudless weather occurred (*ceteris paribus*). Strong wind (the variable “Wind” had a value of 1) reduced the number of passengers by 245. The coefficient for “Precipitation” said that increasing the value for the frequency of precipitation by one led to a decrease in the number of passengers by 202 (*ceteris paribus*). When the daily maximum temperature was higher than the average maximum monthly temperature (the value of “Maximum daily temperature” of 1) the number of passengers increased by 246 (*ceteris paribus*).

Figure 3 shows the differences between the real values of the daily number of passengers and values calculated by the model with weather during the year 2019.

The adjusted coefficient of determination  $R_a^2$  was used to compare models,  $R_a^2 = 0.8438$  for the model without weather and  $R_a^2 = 0.8610$  for the model with weather. The model with weather explains about 2% more of the variability of the number of passengers than the model without weather. This small difference is because vacationers book their stay in the High Tatras several months in advance regardless of the future weather, and the variables “Holiday” and “Period attractiveness” are of the main influence. A significantly bad weather forecast affects day trippers traveling at the last minute. Although the  $R^2$  for the model with weather was slightly greater than the  $R^2$  for the model without weather, they were both

indicative of a good fit. We used the F-test to compare the models. The obtained F-statistic of 12.1809, when compared to  $F_{crit}$  (5.65), indicates that the model with weather fits the data significantly better than the model without weather.

The demand for public passenger transport in 2020 was completely changed during the global pandemic of COVID-19. The total number of passengers observed on TEŽ in 2020 was only 759 721. This is a decrease of 488 133 passengers (39%) compared to 1 247 854 passengers in 2019. Daily numbers of passengers showed considerable variability: The mean was 2 076 passengers with a standard deviation of 2 135; the minimum value was 38, and the maximal one was 8 799. The median was 1 225.

On March 13, 2020, a state of emergency was declared in Slovakia. All schools were closed, international traffic was stopped, and ski resorts, wellness centers, spas, etc. were closed. As of March 15, a state of emergency was declared in hospitals. A ban on the presence of the public in public catering establishments was issued. On April 6, the Slovak government adopted resolutions on movement restrictions. On April 22, the first phase of relaxation began, outdoor sports grounds were opened. Accommodation and outdoor terraces for public catering were opened on May 6. Gradually, further releases came until the borders with other countries were opened.

From October 24, the curfew began to apply again, schools were closed, and nationwide tests were held. Movement restrictions lasted more or less until the end of 2020.

The model without weather was created for the year 2020. The model had the same variables as the model without weather created for the year 2019, with the season adjusted for the Easter holiday. The model created in this way explained only 64.44% ( $R^2 = 0.6444$ ) of the variability in the daily number of passengers, while the coefficient for “Friday” was statistically insignificant based on the p-value. After adding the new variable “Lockdown”, which took the value 1 during the period of strict restrictions (13/03 – 05/05/2020, 24/10 – 31/12/2020), otherwise, it was zero, the model explained 69.39% ( $R^2 = 0.6939$ ) of the variability of the number of passengers. The value of the indicator MAPE was up to 226.23%.

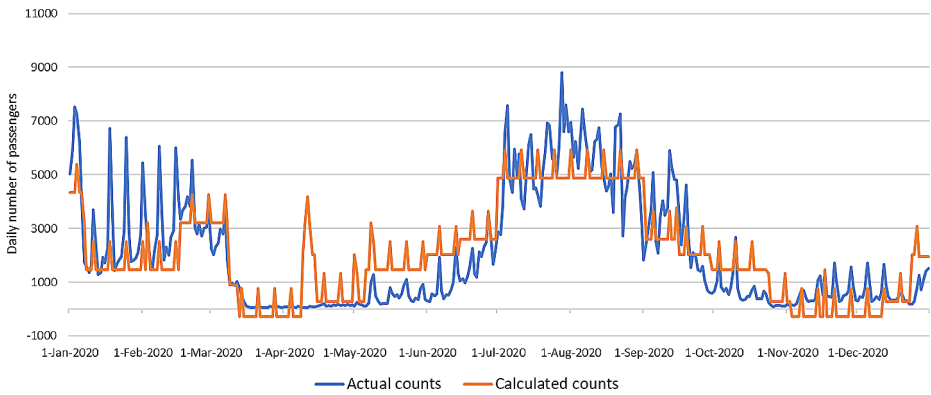
Figure 4 shows the differences between the real values of the daily number of passengers and the values calculated by the model without weather for the year 2020.

If the pre-pandemic period was excluded from the data (01/01 – 12/03/2020) and the model was considered only for the pandemic period with the “Lockdown” variable, the model would already explain up to 75.30% ( $R^2 = 0.753$ ) of the variability in the daily number of passengers.

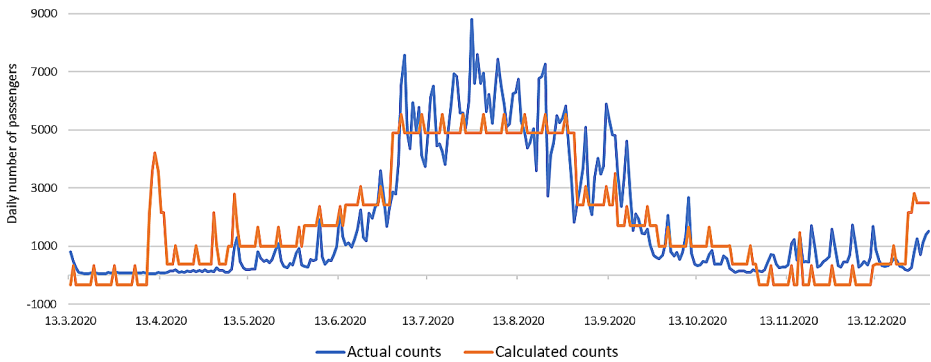
Figure 5 shows the differences between the real values of the daily number of passengers and values calculated by the model without weather during the pandemic period of the year 2020.

Relaxed restrictions in the summer months of July and August in 2020 brought a record number of one-day tourists in the High Tatras (Mrázik 2020). Nevertheless, a comparison of the number of passengers in these months in 2019 (number 488,634) and 2020 (number 338,408) shows that the return of non-regular passengers to public passenger transport did not take place immediately after the anti-pandemic measures were lifted, but the mistrust of this segment of transport in this regard still persisted for a long time.

In the available scientific literature, there are very few articles describing a similar approach to building a model and forecasting the number of passengers, that does not allow for a detailed comparison of the achieved results. For example, Nissen et al. (2020) analyzed



**Fig. 4** Time series of daily passenger flow in 2020 - model without weather



**Fig. 5** Time series of daily passenger flow during the pandemic period of the year 2020

the effect of weather on public transport usage in Berlin and showed that the most important factor influencing ticket sales was temperature. Temperatures below  $-5\text{ }^{\circ}\text{C}$  led to an increase in ticket sales by up to 30% on weekdays, while on hot days ( $>28\text{ }^{\circ}\text{C}$ ) the number of passengers decreased by up to 5%. Precipitation increases the number of sales on working days by up to 5%. Milenković et al. (2016) constated that the time series of passenger numbers realized on the Serbian railway network had a strong autocorrelation of seasonal characteristics and used Seasonal AutoRegressive Integrated Moving Average (SARIMA) method for fitting and forecasting the time series. Butkevičius et al. (2004) forecasted number of passengers carried by railway on the local (national) route by regression analysis taking time factor and national income factor as variables but without considering the quality of the model. Nar and Arslankaya (2022) used regression analysis and artificial neural network and machine learning algorithms technique to forecast the passenger demand on the Yenikapı–Kirazlı metro line. The values of MAPE were less than 2%.

We encountered several problems and limitations while in creating the models:

- Due to the available input data regular commuters to school or work and passengers who bought a ticket without an exact date of travel were not included in the models

described in this article. All models followed the behavior of irregular passengers, i.e., visitors to the High Tatras region. For comparison, Baro and Khouadjia (2021) used the dataset consisting of records of the passenger load collected via sensors in vehicles and calculated at each time a train stops in a station.

- The models created in this study were based on measured weather in a given region on specific days. However, tourists plan their trips based on the weather forecast, which may not always be the same as the real weather on a given day. The advantage of the High Tatras region is that there is a weather station right there, where the forecast is made for the High Tatras region itself, and therefore we can assume that the weather forecasts for this region for the next 24 h are relatively accurate and the differences with reality are minimal. Unfortunately, the authors of this study did not have the opportunity to compare the predicted and actual weather conditions. But it is certainly a challenge for further research that would also take this factor into account, as well as the interaction effects of weather phenomena.
- There are many ways to encode variables describing the weather, whether based on mathematical reasoning, experience, or subjectivity, and it is difficult to determine which method is the most appropriate. In the model without weather, the variable “Period attractiveness” was most affected by subjectivity. The variable “Holiday” could also include separate days when significant cultural or sports events took place in the High Tatras region. We also worked with a model including “bridge” day between a holiday and a weekend, but it was not statistically significant. The influence of the weather during the year is generally difficult to describe mathematically, which was most evident when encoding the variables “Maximum daily temperature” and “Precipitation”. Available data included only values of measured total daily precipitation, but it is difference if the light and persistent rainfall is forecasted or heavy short one, so this variable does not describe the influence of rain very well. It is also difficult to answer the question whether to take these variables as dummy variables, scale variables, or quantitative variables.

## Conclusions

Touristic interesting regions attract many tourists, which means that mobility must be ensured for all visitors to the region. Considering the negative external costs of transport, it is essential that tourists in these areas use public transport as much as possible for their mobility. In order for this public passenger transport to be sufficiently attractive to passengers, it must meet both qualitative and quantitative requirements, and its services must be sufficient to meet the demand for transport. However, the demand for transportation by non-regular passengers is very uneven and is influenced by several factors. With our model, we have proved that some basic factors influencing the demand for transportation by non-regular passengers in touristic attractive areas (days of the week, weather) can be mathematically described and quantified.

Based on the created models and their mathematical evaluation, it can be concluded that the increase in the demand of non-regular passengers for public passenger transport in a tourist-attractive region, in terms of days of the week, is most affected by Friday and Saturday. From this point of view, Sunday seems to be of little importance.

By using multiple linear regression of various combinations of independent variables, it is possible to predict the change in the level of the demand for public passenger transport depending on the importance of the day (for example ordinary day, holiday), period attractiveness (for example Easter, summer holiday, period from Christmas to New Year's Eve) and weather conditions (clouds, wind, precipitation, temperature).

The results of this model can be used directly in the operational planning of public passenger transport. Since the public transport operator can receive the characteristics of the parameters of the following days 1 to 2 days in advance, it can adjust the timetable, the capacity of the sets and plan the driving of vehicles and the need for personnel. This operational management can better ensure the rational use of vehicles and the carrier's employees, ensure sufficiently attractive public passenger transport and at the same time effective use of costs.

It is very difficult to forecast the demand for public passenger transport during the changing anti-pandemic measures (lockdown) because even after the measures are relaxed, potential passengers are likely to be afraid of public transport and prefer to use individual transport by private car.

**Author contributions** Conceptualization: M. Kendra, O. Blažeková, M. Vojteková; Methodology: M. Kendra, O. Blažeková, M. Vojteková; Formal analysis and investigation: O. Blažeková, M. Vojteková; Writing - original draft preparation: M. Kendra, O. Blažeková, M. Vojteková; Writing - review and editing: M. Kendra, O. Blažeková, M. Vojteková; Resources: M. Kendra; Supervision: M. Kendra.

**Funding** This publication was created thanks to support under the Operational Program Integrated Infrastructure for the project: Identification and possibilities of implementation of new technological measures in transport to achieve safe mobility during a pandemic caused by COVID-19 (ITMS code: 313011AUX5), co-financed by the European Regional Development Fund. The authors declare that they have no financial or non-financial interests that are directly or indirectly related to this work.

Open access funding provided by The Ministry of Education, Science, Research and Sport of the Slovak Republic in cooperation with Centre for Scientific and Technical Information of the Slovak Republic

## Declarations

**Competing interests** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Baro, J., Khouadjia, M.: Passenger flow forecasting on transportation network: sensitivity analysis of the spatiotemporal features. In: 2021 International Conference on Data Mining Workshops (ICDMW). pp. 734–741. IEEE, Auckland, New Zealand (2021)
- Bilal, M., Son, S., Jang, K.: Traveler's interactive decision-making behavior between itinerary and mode choice using Copula-based discrete-count joint modeling. *Transportation*. (2023). <https://doi.org/10.1007/s11116-023-10409-w>

- Böcker, L., Dijst, M., Prillwitz, J.: Impact of Everyday Weather on Individual Daily Travel Behaviours in Perspective: A literature review. *Transp. Rev.* 33 (2013). <https://doi.org/10.1080/01441647.2012.747114>
- Bursa, B., Mailer, M., Axhausen, K.W.: Intra-destination travel behavior of alpine tourists: A literature review on choice determinants and the survey work. *Transportation*. 49, 1465–1516 (2022). <https://doi.org/10.1007/s11116-022-10267-y> (a)
- Bursa, B., Mailer, M., Axhausen, K.W.: Travel behavior on vacation: Transport mode choice of tourists at destinations - ScienceDirect. *Transp. Res. Part. Policy Pract.* 234–261 (2022)(b)
- Butkevicius, J., Mazūra, M., Ivankovas, V., Mazūra, S.: Analysis and forecast of the dynamics of passenger transportation by public land transport. *TRANSPORT*. 19, 3–8 (2004). <https://doi.org/10.3846/16484142.2004.9637944>
- Cavallaro, F., Ciari, F., Nocera, S., Prettenthaler, F., Scuttari, A.: The impacts of climate change on tourist mobility in mountain areas. *J. Sustain. Tour.* 25, 1063–1083 (2017). <https://doi.org/10.1080/09669582.2016.1253092>
- Clifton, K., Chen, R., Cutter, A.: Representing Weather in Travel Behaviour models: A Case Study from Sydney, AUS. *ATRF 2011–34th Australas. Transp. Res. Forum* (2014)
- Cools, M., Moons, E., Creemers, L., Wets, G.: Changes in travel behavior in response to Weather conditions: Do type of Weather and trip purpose Matter. *Transp. Res. Rec J. Transp. Res. Board.* 2157, 22–28 (2010). <https://doi.org/10.3141/2157-03>
- Cyril, A., Mulangi, R.H., George, V.: Modelling and forecasting bus passenger demand using Time Series Method. In: *Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp. 460–466. IEEE, Noida, India (2018)
- Danis, J., Dolinayova, A., Camaj, J.: Prognosis of Rail Transport Demand in The Slovak Republic. Presented at the International Conference on Engineering Science and Management August (2016) (2016)
- Dobré noviny, T.A.S.R.: Počet turistov vo Vysokých Tatrách prekonal 40-ročný rekord: Za jeden deň ich napočítali viac ako 30-tisíc, (2020). <https://www.dobrenoviny.sk/c/188112/pocet-turistov-v-tatrach-prekonal-40-rocnny-rekord-za-jeden-den-ich-napocitali-viac-ako-30-tisic>, Accessed 10 June 2022
- Dou, F.D., Xu, J., Wang, L., Jia, L.: A train dispatching model based on fuzzy passenger demand forecasting during holidays. *J. Ind. Eng. Manag.* 6, 320–335 (2013). <https://doi.org/10.3926/jiem.699>
- Haberl, M., Neuhold, R.: Modelling the demand of leisure traffic in the province of Salzburg. In: *ETC Conference Papers 2012*. pp. 1–17., Glasgow, UK (2012)
- High-Tatras.travel: Transport in High Tatras, <https://www.high-tatras.travel/information/useful-information/>
- Holek, I.M.: Vysoké Tatry, <https://travelingtunas.com/sk/vysoke-tatry/> Accessed 1 August 2023
- Ivanovic, I., Jović, J.: Sensitivity of street network capacity under the rain impact: Case study of Belgrade. *Transport*. 33, 1–8 (2017). <https://doi.org/10.3846/16484142.2017.1283532>
- Lewis, C.D.: *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*. Butterworth Scientific, London (U. A.) (1982)
- Liu, C., Susilo, Y.O., Karlström, A.: Examining the impact of weather variability on non-commuters' daily activity–travel patterns in different regions of Sweden. *J. Transp. Geogr.* 39, 36–48 (2014). <https://doi.org/10.1016/j.jtrangeo.2014.06.019>
- meteoblue: Simulované historické údaje o klíme a počasí pre Tatranská Polianka, [https://www.meteoblue.com/sk/po%C4%8Dasie/historyclimate/climatemodelled/tatranska%3a1-polianka\\_slovensk%C3a1-republika\\_723396](https://www.meteoblue.com/sk/po%C4%8Dasie/historyclimate/climatemodelled/tatranska%3a1-polianka_slovensk%C3a1-republika_723396) Accessed 13 May 2022
- Milenković, M., Švadlenka, L., Melichar, V., Bojović, N., Avramović, Z.: SARIMA modelling approach for railway passenger flow forecasting. *Transport*. 1–8 (2016). <https://doi.org/10.3846/16484142.2016.1139623>
- Nar, M., Arslankaya, S.: Passenger demand forecasting for railway systems. *Open. Chem.* 20, 105–119 (2022). <https://doi.org/10.1515/chem-2022-0124>
- Nissen, K.M., Becker, N., Dähne, O., Rabe, M., Scheffler, J., Solle, M., Ulbrich, U.: How does weather affect the use of public transport in Berlin? *Environ. Res. Lett.* 15, 085001 (2020). <https://doi.org/10.1088/1748-9326/ab8ec3>
- Odgers, J., van Schijndel, L.: Forecasting annual train boardings in Melbourne using time series data. *World Transit. Res.* (2011)
- Petrović, D., Ivanović, I., Đorić, V., Jović, J.: Impact of Weather conditions on Travel demand – the most common research methods and Applied models. *Promet - TrafficTransportation*. 32, 711–725 (2020). <https://doi.org/10.7307/ptt.v32i5.3499>
- Picano, B., Chiti, F., Fantacci, R., Han, Z.: Passengers Demand Forecasting Based on Chaos Theory: 2019 IEEE International Conference on Communications, ICC 2019. 2019 IEEE Int. Conf. Commun. ICC 2019 - Proc. (2019). <https://doi.org/10.1109/ICC.2019.8762041>
- Rahman, S., Balijepalli, C.: Understanding the determinants of demand for public transport: Evidence from suburban rail operations in five divisions of Indian railways. *Transp. Policy*. 48, 13–22 (2016). <https://doi.org/10.1016/j.tranpol.2016.02.006>



- Rimarčík, M.: Štatistika pre prax. (2007). <http://rimarcik.com>
- Rudloff, C., Leodolter, M., Bauer, D., Auer, R., Brög, W., Kehnscherper, K.: Influence of Weather on Transport demand. *Transp. Res. Rec J. Transp. Res. Board.* 2482, 110–116 (2015). <https://doi.org/10.3141/2482-14>
- Sabir, M.: Weather and travel behaviour, (2011). <https://research.vu.nl/ws/portafiles/portal/42203407/complete+dissertation.pdf>, Accessed 10 June 2022
- Shi, Z., Pan, W., He, M., Liu, Y.: Understanding passenger route choice behavior under the influence of detailed route information based on smart card data. *Transportation.* (2023). <https://doi.org/10.1007/s11116-023-10432-x>
- Tao, S., Corcoran, J., Rowe, F., Hickman, M.: To travel or not to travel: Weather is the question. Modelling the effect of local weather conditions on bus ridership. *Transp. Res. Part. C-Emerg Technol.* 86, 147–167 (2018). <https://doi.org/10.1016/j.trc.2017.11.005>
- Tatry mountain resorts, a.s.: Ekologická doprava je šanca pre lepšiu budúcnosť Tatier, (2018). [https://tmr.sk/data/upload/short.messages/ekologicka-doprava-je-sanca-pre-lepsiu-buducnost-tatier-3\\_20180222\\_135418.pdf](https://tmr.sk/data/upload/short.messages/ekologicka-doprava-je-sanca-pre-lepsiu-buducnost-tatier-3_20180222_135418.pdf), Accessed 9 September 2022
- Tuan, V.A., Huong, N.T.T.: Analysis of Mode Choice Behavior under adverse Weather conditions using RA and SA surveys - A case study from Ho Chi Minh City (HCMC), Vietnam. In: Ha-Minh, C., Dao, D.V., Benboudjema, F., Derrible, S., Huynh, D.V.K., Tang, A.M. (eds.) *CIGOS 2019, Innovation for Sustainable Infrastructure*, pp. 959–964. Springer, Singapore (2020)
- Wang, X., Guo, Y., Bai, C., Liu, S., Liu, S., Han, J.: The effects of Weather on Passenger Flow of Urban Rail Transit. *Civ. Eng. J.* 6, 11–20 (2020). <https://doi.org/10.28991/cej-2020-03091449>
- Wijeweera, A., Charles, M.B.: Determinants of passenger rail demand in Perth, Australia: A Time Series Analysis. *Appl. Econom Int. Dev.* 13, 217–230 (2013)
- Wikipedia: Mean absolute percentage error, (2023). [https://en.wikipedia.org/w/index.php?title=Mean\\_absolute\\_percentage\\_error&oldid=1163966059](https://en.wikipedia.org/w/index.php?title=Mean_absolute_percentage_error&oldid=1163966059), Accessed 13 July 2023
- Zamir Khan, M., Naheed Khan, F.: A dynamic analysis of rail travel demand in Pakistan. *Case Stud. Transp. Policy.* 9, 860–869 (2021). <https://doi.org/10.1016/j.cstp.2021.04.006>
- Zhao, S.-Z., Ni, T.-H., Wang, Y., Gao, X.-T.: A new approach to the prediction of passenger flow in a transit system. *Comput. Math. Appl.* 61, 1968–1974 (2011). <https://doi.org/10.1016/j.camwa.2010.08.023>
- Zhi, X.: 2nd International Conference on Electrical, Engineering: Analysis and Forecast of Railway Passenger Volume Using System Dynamics. Presented at the and Automation (ICEEA 2018) March (2018) (2018)
- Zhu, F.: Factor Analysis of Railway Passenger Transport Demand in Eastern China. Presented at the 2018 2nd International Conference on Education, Economics and Management Research (ICEEMR 2018) June (2018)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Martin Kendra<sup>1</sup>  · Olga Blažeková<sup>1</sup>  · Mária Vojteková<sup>1</sup> 

✉ Martin Kendra  
martin.kendra@uniza.sk

Olga Blažeková  
olga.blazekova@uniza.sk

Mária Vojteková  
maria.vojtekova@uniza.sk

<sup>1</sup> Faculty of Operation and Economics of Transport and Communications, University of Žilina, Univerzitná 8215/1, Žilina 010 26, Slovakia