GAM MODELLING OF DAILY NUMBER OF TRAFFIC ACCIDENTS AS A FUNCTION OF METEOROLOGICAL VARIABLES IN THE CZECH REPUBLIC

Petra Kolísková[1]✉, Jiří Neubauer[1]

[1] *University of Defence, Brno, Czech Republic*

EUROPEAN JOURNAL OF BUSINESS SCIENCE AND TECHNOLOGY

Volume 11 Issue 1
ISSN 2694-7161
www.ejobsat.com

ABSTRACT

Meteorological conditions exert a considerable influence on traffic patterns. This paper examines the influence of meteorological variables on the daily number of traffic accidents requiring fire brigade intervention. The influence of meteorological variables, including maximum temperature, wind speed, air pressure, precipitation, snow cover and sunshine, was examined. A Generalized Additive Model for variables with a Poisson distribution was employed for modelling purposes, as this allows for the representation of non-linear dependencies. The analysis demonstrates that the lowest incidence of accidents occurs at temperatures approximating 10 °C. The average daily number of accidents increases with windy weather, the minimum number of accidents occurs at zero precipitation, and the accident rate rises with higher levels of sunshine. In the Czech Republic, the period of greatest risk in terms of road traffic accidents is the summer and winter months. The findings may have several practical applications, for example, in the improvement of meteorological warnings in traffic.

KEY WORDS

traffic accidents, integrated rescue system, weather, Poisson distribution, generalised additive model

JEL CODES

C020, C510

1 INTRODUCTION

The utilisation of personal vehicles for the purpose of transportation is an integral aspect of contemporary society. On the one hand, contemporary automobiles are equipped with an array of sophisticated technology and are manufactured with a heightened emphasis on safety compared to previous eras. Conversely, however, automobiles are becoming increasingly affordable, traffic congestion is intensifying, and drivers are often encouraged to utilise the full potential of their vehicles, which can have adverse consequences and potentially contribute to accidents.

KOLÍSKOVÁ, Petra, and NEUBAUER, Jiří. 2025. GAM Modelling of Daily Number of Traffic Accidents as a Function of Meteorological Variables in the Czech Republic. *European Journal of Business Science and Technology,* 11 (1): 23–38. ISSN 2694-7161, DOI 10.11118/ejobsat.2024.014.

The occurrence of traffic accidents can be attributed to a multitude of factors. Police records indicate that such causes include driver inattention, driving under the influence of addictive substances, and failure to comply with regulations. It appears that the influence of human personality can be negated by operating a vehicle with the assistance of artificial intelligence, as autonomous vehicles represent the future of transportation. However, it should be noted that there are also numerous external factors that can influence the vehicle, some of which are challenging to anticipate, while others can be predicted based on external conditions, such as the impact of weather. Even autonomous vehicles are susceptible to adverse weather conditions and require various sensors to monitor external factors such as temperature, pressure, and humidity. Consequently, there is a potential for unexpected hazards on the road.

The relationship between meteorological parameters and traffic is dependent on the specific characteristics of the local climate and the particularities of regional traffic patterns. The more densely populated regions of the country are distinguished by a more extensive and frequent transportation infrastructure, including long-distance highways. Conversely, mountainous border regions, which are sparsely populated, are also highly attractive to tourists during peak tourist seasons. Furthermore, traffic volume is indicative of the prevailing meteorological conditions. It is generally accepted that higher temperatures encourage people to engage in outdoor activities. However, research indicates that temperatures above 25–30 °C may have a deterring effect (Thorsson et al., 2004; Aultman-Hall et al., 2009). In the context of low temperatures, a reduction in vehicle traffic is observed. However, commercial traffic remains operational. Wind speeds may also present a challenge, particularly in regions with elevated altitudes or along stretches of highways where forestland gives way to open countryside. Extreme weather conditions represent a significant challenge to the uninterrupted flow of traffic, although such occurrences are not a common feature of the Czech Republic. In Brázdil et al. (2021), an analysis of traffic accidents in the Czech Republic is conducted, examining the impact of various meteorological conditions, including windstorms, convective storms, rain, snow, ice, frost, heat, fog and floods. The greatest influence was found to be that of frost (31%), followed by that of ice, rain and snow. High wind speeds present a significant hazard, particularly for trucks, and also for the collision of any vehicle with a flying object (Becker et al., 2022a). The transportation of goods and passengers over longer distances is frequently constrained by precipitation, particularly in the form of rain and snow (Roh, 2020). The occurrence of severe weather in conjunction with drivers' risky driving behaviour frequently results in an elevated incidence of traffic accidents (Brázdil et al., 2022). In urban areas, precipitation has been observed to increase vehicular traffic, which does not necessarily imply an elevated risk of accidents. A 21-year observation of daily traffic accidents in Athens (Yannis and Karlaftis, 2010) revealed that precipitation is associated with a reduction in accidents. This phenomenon can be attributed, at least in part, to drivers demonstrating heightened levels of attention and adherence to safety regulations. In general, the most recent findings indicate a declining trend in the impact of weather on fatal road accidents in the country (Brázdil et al., 2023).

The objective is to provide a quantitative description of the manner in which meteorological parameters can enhance the predictive capacity of models designed to anticipate traffic accidents. Such applications can be found in areas where accurate estimation of traffic flow is of importance, for example, in road traffic accident models, where traffic flow represents the dominant factor in the assessment of risk, and in the field of traffic management, as well as in the planning of routes for navigation systems (Becker, 2022b). The findings may have implications for real-time traffic management and the implementation of crash warning systems. Furthermore, based on these findings, weather-dependent driving restrictions could be issued with the aim of enhancing road safety (Becker et al., 2022a).

2 THEORETICAL BACKGROUND

The relationship between the number of traffic accidents and meteorological conditions and other explanatory variables can be examined through the application of diverse investigative techniques. The GAM (Generalized Additive Model) method is frequently contrasted with machine learning-based methods (Anderson et al., 2015). Both methods have been employed to describe non-linear relationships between variables in the tsunami generation area, and both have yielded comprehensive models. Despite achieving relatively low mean errors, the machine learning-based methods did not demonstrate a superior ability to describe the risk event model in comparison to GAM. However, the interpretation of smoothing splines may present a potential challenge with GAM. Despite the fact that machine learning methods did not yield optimal predictions (Cerna et al., 2020; Guyeux et al., 2020), they demonstrated superior performance in accommodating unanticipated surges in rescue operations due to difficult-to-forecast natural disasters in comparison to classical models. The effectiveness of GAMs has been demonstrated in the context of lightning-induced forest fires (Rodríguez-Pérez et al., 2020). The application of GAM models to traffic accident data is a common occurrence (Li et al., 2011; Zhang et al., 2012). When compared to GLM models, GAM displays greater flexibility in describing the effect of changes in independent variables and produces superior results. The relationship between hourly vehicle departures and weather conditions has been investigated by Lepage and Morency (2020), who employed GAM and ARIMA to model this variable. In comparing the models, it was found that ARIMA performed better in short-term forecasts, while GAM proved to be more suitable for long-term forecasts, where ARIMA exhibited a significant mean error. In light of the aforementioned literature and further study, it can be posited that GAM is an appropriate method for describing a sparse phenomenon such as the daily number of traffic accidents involving firefighter call-outs and its dependence on weather.

In the literature, a number of significant probability distributions for traffic accident data are identified. The impact of meteorological data, including precipitation, temperature, cloud cover, and wind speed, on the number of vehicles passing per hour was investigated by Becker et al. (2022b). The probability distribution was analysed as a Poisson distribution. The Poisson, Negative Binomial (NB), Zero Inflated Poisson (ZIP), and Zero Inflated Negative Binomial (ZINB) distributions are frequently employed for the description of traffic accidents (Lord et al., 2005). In Pop (2018), the Poisson and quasi-Poisson GLM model was employed for the analysis of fatal traffic accidents. It appears that overdispersion has a significant impact on the data distribution. In Basu and Saha (2017), it was demonstrated that a model based on a negative binomial distribution yielded superior results for real data exhibiting overdispersion in comparison to Poisson regression.

The objective of this paper is to investigate the influence of meteorological variables on the daily number of accidents. It is assumed that the variable follows a Poisson distribution, and a GAM approach is used for modelling. The quality of the prediction result is contingent upon the character of the data and the modelling method employed. The temporal and spatial relationships between traffic and meteorological data are not exact. It is challenging to establish such a link, as the meteorological measuring station is frequently situated at a distance from the location of the traffic accident. This is particularly evident in the case of precipitation, which varies significantly from place to place (Thorsson et al., 2004). Other potential limitations include the assumption that meteorological variables are equally significant across all regions, the use of data with outliers, and the overparameterisation of the model due to the smoothing.

3 METHODOLOGY AND DATA

3.1 Traffic Accident Data

In accordance with Czech legislation, a traffic accident is defined as an incident occurring on a road that endangers or threatens the life or health of individuals or causes damage to property, and which is subject to notification.

In the event of a traffic accident, the integrated rescue system is typically mobilised, with the fire brigade assuming a prominent role. In the event of vehicles obstructing traffic or liquids leaking, firefighters are equipped with the necessary apparatus to resolve the situation or assist an injured individual in extricating themselves from the wreckage.

The data on traffic accidents has been sourced from the database of the Czech Republic's Fire Rescue Service. It comprises information on accidents, primarily concerning the time and location of the incident. A concise overview of incidents involving firefighter call-outs in the 2012–2021 period is presented for each region in the form of boxplots in Fig. 1. The Czech Republic is comprised of 14 regions. The regions in question are Kralovehradecky (HKK), Jihocesky (JHC), Jihomoravsky (JHM), Karlovarsky (KVK), Liberecky (LBK), Moravskoslezsky (MSK), Olomoucky (OLK), Pardubicky (PAK), Hlavni mesto Praha (PHA), Plzensky (PLK), Stredocesky (STC), Ustecky (ULK), Vysocina (VYS), Zlinsky (ZLK).

From 1 January 2012 to 31 December 2021, a total of 198,773 traffic accidents occurred in the Czech Republic, which were attended to by fire rescue service units. The accident data were initially subjected to a series of adjustments. For instance, coordinates that were deemed to be meaningless were removed. Additionally, the redistribution of territory within municipalities due to administrative changes in the Czech Republic over the 10-year period for which the measurements were taken was taken into account. The dataset revealed 1,730 instances of exceptionally high daily accident counts across various regions. In the initial modelling stage, the outlier cases were included to the model of the average daily accident count. It is plausible that these outliers may represent the true data, rather than being the result of measurement error. Furthermore, they align with the inherent characteristics of the data.

The data were arranged in chronological order, with each day from 1 January 2012 to 31 December 2021 assigned the daily number of accidents and the categorical variables, including day of the week, month, region of the country, public holidays and state of emergency during a pandemic caused by the SARS-CoV-2 virus.

The data set was expanded with the inclusion of information pertaining to whether the day in question was a public holiday or a state of emergency. The following dates are observed as public holidays in the Czech Republic: 1 January (Day of the Independent Czech Republic), Good Friday and Easter Monday, 1 May (Labour Day), 8 May (Victory Day), 5 July (Cyril and Methodius Day), 6 July (Burnt Day of Master Jan Hus), 28 September (Day of Czech Statehood), 28 October (Day of the Establishment of Czechoslovakia), 17 November (Day of the Fight for Freedom and Democracy), 24–26 December (Christmas Day and the 1st and 2nd Christmas Holidays). The following dates were marked by the declaration of a state of emergency in response to the ongoing pandemic: The period between 12 March 2020 and 17 May 2020, between 5 October 2020 and 11 April 2021, and between 26 November 2021 and 25 December 2021 was characterised by the declaration of a state of emergency.

3.2 Meteorological Data

Furthermore, meteorological variables were incorporated into the data set, including maximum daily temperature, average daily wind speed, average daily air pressure, total daily precipitation, total daily snow height, and total daily sunshine.

The meteorological data is derived from a selection of weather stations located within each region. The Czech Meteorological Office

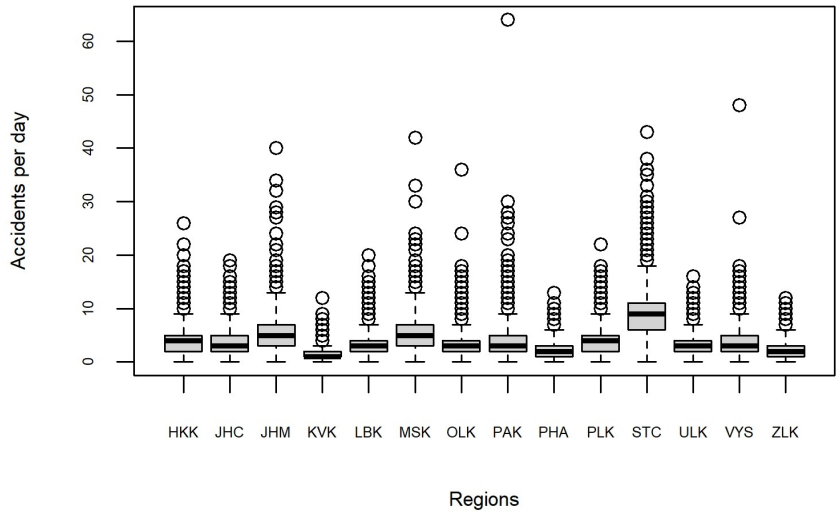

Fig. 1: Boxplots showing median, minimum, maximum and outliers of the daily number of accidents with the deployment of the Fire Rescue Service in the regions of the Czech Republic

(CHMI) oversees the monitoring of climatic and meteorological phenomena in the Czech Republic through a network of various types of stations. The most significant of these are those with a professional meteorologist and an automated measuring system. The meteorological parameters on which the dependence of the explained variable was studied in this paper are as follows:

- *temperature.max* – The maximum daily air temperature in °C – is recorded at the hourly interval between 9 p.m. on the previous day and 9 p.m. on the current day.
- *wind.speed* – The average daily wind speed in m/s – is determined by calculating the mean of the measurements taken at 7 a.m., 2 p.m., and 9 p.m. hours local mean solar time.
- *air.pressure* – The average daily air pressure in hPa – is calculated as the mean of the pressures observed at 7 a.m., 2 p.m., and 9 p.m. local mean solar time.
- *precipitation* – The daily rainfall in mm – is measured at 7 a.m. for the previous 24 hours and recorded as of the previous day.
- *snow.height* – The total height of the snow cover in cm – measured at 7 a.m.
- *sunshine* – The daily total sunshine in hours – represents the time interval in whole hours between sunrise and sunset when the sun was not obscured by clouds or other obstructions, i.e., the intensity of the flow of direct solar radiation greater than 120 W/m^2.

The Czech Meteorological Office measured a number of variables, and those that contribute to the predictive ability of the model were selected. It was not possible to include all variables due to the resulting multicollinearity. The statistical characteristics of the meteorological variables during the period of interest are presented in Tab. 1, and the data are presented graphically in Fig. 2.

3.3 Methods

The linear regression model (LRM) is a well-established technique for expressing the explained variable Y_i as a linear combination of the independent variables X_{ij}, where $i = 1, \ldots, n$, $j = 1, \ldots, k$, and the corresponding parameters. In this context, n represents the number of observations and k denotes the number of parameters. The regressors may be either continuous or categorical. In the event that the explanatory variable is discrete, a Poisson regression model may be employed. This falls within the GLM family, with a logarithmic function serving as the link function.

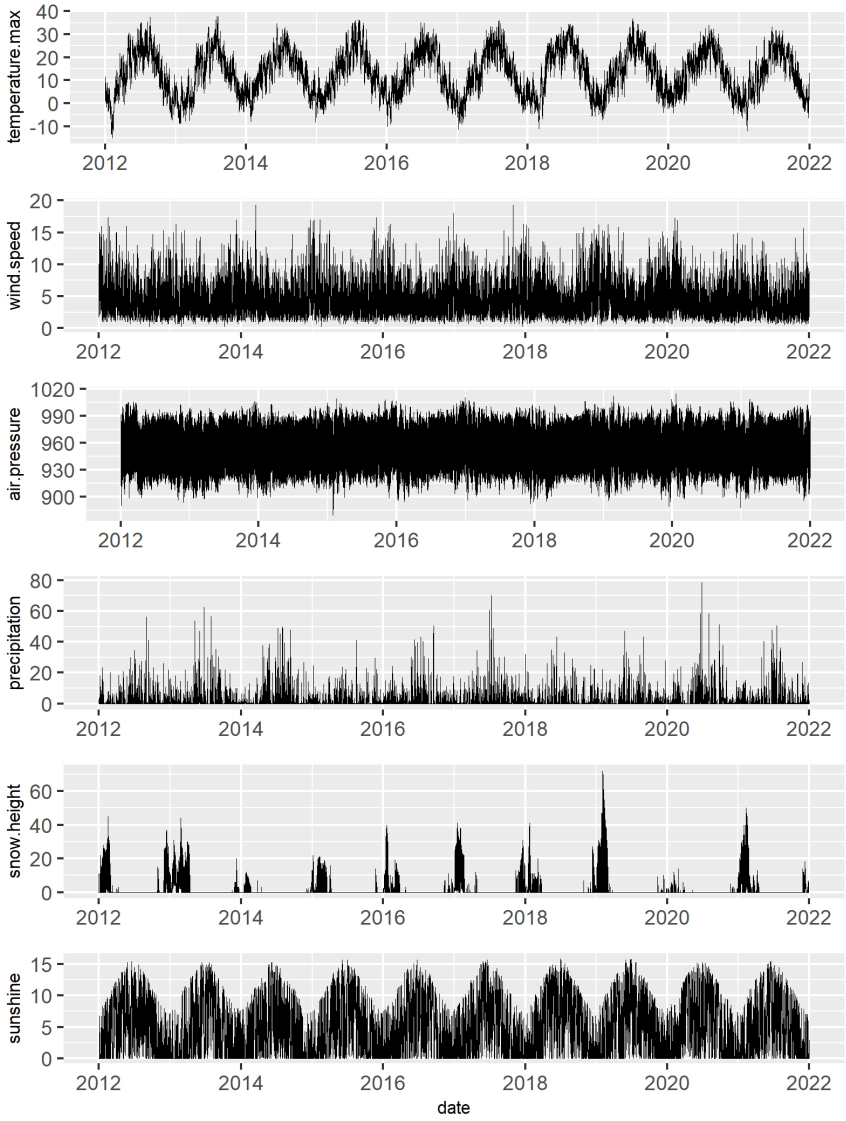

Fig. 2: Graphs of meteorological variables used for the period 2012–2021

Tab. 1: Statistical characteristics of meteorological variables for the period 2012–2021

	temperature.max	**wind.speed**	**air.pressure**	**precipitation**	**snow.height**	**sunshine**
Minimum	−14.90	0.300	879.6	0.000	0.000	0.000
1st Quartile	5.60	2.300	947.1	0.000	0.000	0.200
Median	13.40	3.300	966.8	0.000	0.000	3.600
Mean	13.39	3.857	962.3	1.747	1.268	4.697
3rd Quartile	21.00	4.800	981.3	1.300	0.000	8.300
Maximum	37.80	19.300	1015.0	78.700	72.000	15.900
Standard deviation	9.50	2.078	23.5	4.380	4.681	4.476

The generalised additive model (GAM) represents an extension of the generalised linear model (GLM), wherein the linear terms are substituted with non-parametric smooth functions (Anderson et al., 2015). The fundamental tenets of both models remain consistent: the quantity under investigation belongs to the exponential family of distributions, in this case the Poisson distribution, and its mean can be expressed in terms of regressors via a link function.

$$E(y) = \mu$$
$$g(\mu) = b_0 + f(x_1) + f(x_2) + \ldots + f(x_l)$$

The principal distinction is that it incorporates smooth functions $f(x)$ of one or more regressors into the linear model. Smooth functions are typically nonlinear, thereby enabling the modelling of nonlinear relationships between regressors and the explained variable (Clark, 2024). The smooth function f can be written as

$$f(x) = \sum_{j=1}^{J} b_j(x)\beta_j,$$

where $b_j(x)$ is the j-th of some basis functions and β_j are some unknown parameters, which must be estimated (Becker et al., 2022a; Wood, 2017). The number of parameters affects the degree of change in the shape of the resulting dependent variable, or its "wiggliness". A function with a greater number of parameters is better able to capture finer patterns and has a more complex shape. However, when building a model, it is undesirable for the function to be over-fitting and to reproduce noise rather than the principal trend. A smoothing parameter λ is therefore introduced in order to achieve a balance between the degree of wiggliness and the likelihood of the model, which characterises its ability to describe the data accurately:

$$\text{Fit} = \text{Likelihood} - \lambda \cdot \text{Wiggliness}$$

Thus, GAMs are generalised linear models that are formed by the sum of smooth functions of variables instead of a linear combination of variables. The advantage of GAM over GLM is that they permit the description of nonlinear relationships between variables and the generation of accurate predictions. Like GLM, GAM models are capable of statistical inference and the explanation of the underlying structure of the models, as well as the justification of their predictions. Nevertheless, the process of smoothing variables can result in an increased number of degrees of freedom in the model, necessitating the use of more extensive data for GAM modelling. The use of a large number of parameters to generate spline functions in GAM models can result in estimated coefficients that are challenging to interpret (Li et al., 2011). The results of the GLM and GAM models are comparable when the regressors are genuinely independent and the variable being explained is a linear variable. In this context, GAM does not offer a notable advantage over GLM; instead, it can lead to overfitting of the data (Li et al., 2011).

The proportion (or percentage) of explained deviance serves as a measure of model quality. For both GAM and GLM models, the proportion of explained deviance can be determined using the expression

$$1 - \frac{D}{D_{\text{null}}},$$

where the deviance D_{null} corresponds to the model containing only the intercept (worst fit) and the residual deviation D is for the model containing the independent variables. An higher value of D indicates a greater discrepancy between the estimated and observed values, thereby signifying that the model is less capable of accurately representing the observed dependence. The proportion of explained deviance is an indicator of the accuracy of the model, with values ranging from 0 to 1. A value approaching 1 indicates a superior model that more closely aligns with the actual data. The statistical software typically expresses this in percentage terms.

In the event that a variable is modelled as a function of multiple variables, the graphical output of the GAM is then constituted by a set of graphs, each representing the partial contribution of a given variable to the overall prediction. The modelled smooth curve for a specific variable represents the predicted mean value of the explained variable, assuming that

the other dependent variables reach their mean. It is accompanied by a 95% confidence interval.

The data were initially subjected to a series of tests to ascertain their suitability for analysis. These tests included the Cramer-von Mises and Anderson-Darling goodness-of-fit tests, which indicated that the data exhibited the best fit with a Poisson distribution. Prior to investigating the impact of meteorological variables on traffic accidents, the potential for multicollinearity among the meteorological regressors was evaluated. This analysis revealed that there was no evidence of collinearity between the selected variables. The calculations were conducted using the R statistical software, version 4.3.1.

4 RESULTS

The initial model employed for the analysis of the dependence of daily accident rates on meteorological variables was a GLM model for Poisson distribution data with a logarithmic link function. All variables were found to be statistically significant, with a p-value of less than 0.01. In this case, the percentage of explained deviance was 34.6%. A negative binomial model is frequently employed for risk event count data (Basu and Saha, 2017; Pop, 2018), thus a GLM model was also conducted to assess the assumption of a negative binomial distribution of the data for comparison. However, this model did not yield any improvement, resulting in a lower percentage of explained variance, namely 33.3%.

The data demonstrate a clear non-linear relationship between meteorological variables and the number of traffic accidents resulting in firefighter deployments. In such cases, the GAM model is an advantageous statistical tool for analysis. The GAM model was estimated for the case of the same variables as the GLM model. Initially, the model was estimated without the use of splines, which resulted in a model that was almost indistinguishable from the GLM, as anticipated. Subsequently, the meteorological variables were smoothed, as illustrated in Tab. 2. The variables region, month, day, holiday and Covid emergency are categorical variables and are represented as parametric coefficients in the model. A treatment contrast (dummy coding) was employed, with the variables region (HKK), month (January), and day (Monday) serving as the reference category. The results of the model demonstrated that all variables were statistically significant, with the exception of month2 February and day Thursday, which exhibited a daily number of accidents that was nearly identical to the reference value. The table illustrates which regions, days and months exhibit a lower or higher number of accidents per day in comparison to the reference variables. The variables representing holidays and the state of the pandemic have a negative coefficient, indicating that they serve to reduce the number of accidents. The GAM model with a Poisson distribution and outliers included exhibited an explained deviance of 36.1%. The percentage of explained deviance for the GAM model assuming a negative binomial distribution is once again lower, at 34.8%.

The output of the GAM model is also a table of the smoothed functions, which are visualised by a series of graphs. The value edf represents the effective degrees of freedom and characterises the degree of wiggliness of the curve. A value of edf 1 represents a straight line, edf 2 a parabola, and so on. The remaining columns pertain to the results of the significance testing, which indicate that all variables are statistically significant. The plots for the meteorological parameters in Fig. 3 show the smoothed terms plots expressing the contribution of a given variable to the value of the modelled variable – daily number of accidents. The horizontal axis depicts the observed values of the variable, while the vertical axis illustrates the partial effect of the smoothed variable at the estimated number of degrees of freedom relative to the overall mean of the explained variable. The 95% confidence interval is indicated in grey. The wider confidence interval is attributable to the paucity of accident data for these values of the independent variable.

Tab. 2: R output for GAM with Poisson distribution, model for complete data and data without outliers

	Complete data				Without outliers			
Parametric coefficients	**Estim.**	**Std. error**	***p*-value**		**Estim.**	**Std. error**	***p*-value**	
(Intercept)	1.363	0.015	< 0.001	***	1.297	0.016	< 0.001	***
regionJHC	−0.136	0.014	< 0.001	***	−0.087	0.014	< 0.001	***
regionJHM	0.353	0.014	< 0.001	***	0.337	0.014	< 0.001	***
regionKVK	−1.087	0.026	< 0.001	***	−1.097	0.029	< 0.001	***
regionLBK	−0.262	0.013	< 0.001	***	−0.289	0.013	< 0.001	***
regionMSK	0.301	0.014	< 0.001	***	0.301	0.014	< 0.001	***
regionOLK	−0.151	0.015	< 0.001	***	−0.206	0.015	< 0.001	***
regionPAK	−0.160	0.014	< 0.001	***	−0.119	0.015	< 0.001	***
regionPHA	−0.454	0.014	< 0.001	***	−0.474	0.014	< 0.001	***
regionPLK	−0.215	0.024	< 0.001	***	−0.041	0.026	0.112	
regionSTC	0.853	0.010	< 0.001	***	0.845	0.011	< 0.001	***
regionULK	−0.489	0.028	< 0.001	***	−0.284	0.030	< 0.001	***
regionVYS	−0.195	0.014	< 0.001	***	−0.159	0.015	< 0.001	***
regionZLK	−0.492	0.016	< 0.001	***	−0.516	0.016	< 0.001	***
month2	−0.020	0.012	0.094	.	−0.046	0.013	< 0.001	***
month3	−0.081	0.013	< 0.001	***	−0.093	0.014	< 0.001	***
month4	−0.028	0.014	0.050	*	−0.042	0.015	0.006	**
month5	0.049	0.015	0.001	**	0.032	0.016	< 0.001	*
month6	0.197	0.016	< 0.001	***	0.144	0.017	< 0.001	***
month7	0.140	0.017	< 0.001	***	0.096	0.018	< 0.001	***
month8	0.187	0.017	< 0.001	***	0.118	0.017	< 0.001	***
month9	0.246	0.015	< 0.001	***	0.180	0.016	< 0.001	***
month10	0.252	0.014	< 0.001	***	0.207	0.015	< 0.001	***
month11	0.152	0.012	< 0.001	***	0.136	0.013	< 0.001	***
month12	0.196	0.011	< 0.001	***	0.175	0.012	< 0.001	***
dayTuesday	−0.062	0.008	< 0.001	***	−0.060	0.009	< 0.001	***
dayWednesday	−0.044	0.008	< 0.001	***	−0.039	0.009	< 0.001	***
dayThursday	−0.020	0.008	0.015	*	−0.024	0.009	0.006	**
dayFriday	0.117	0.008	< 0.001	***	0.103	0.008	< 0.001	***
daySaturday	−0.071	0.008	< 0.001	***	−0.066	0.009	< 0.001	***
daySunday	−0.232	0.009	< 0.001	***	−0.208	0.009	< 0.001	***
holiday1	−0.243	0.014	< 0.001	***	−0.231	0.015	< 0.001	***
COVIDemergency1	−0.141	0.009	< 0.001	***	−0.123	0.010	< 0.001	***
Approximate significance of smooth terms:	**Edf**	**Ref. df**	***p*-value**		**Edf**	**Ref. df**	***p*-value**	
s(temperature.max)	8.911	8.997	< 0.001	***	8.439	8.887	< 0.001	***
s(wind.speed)	6.035	6.929	< 0.001	***	2.753	3.483	< 0.001	***
s(air.pressure)	8.307	8.831	< 0.001	***	6.970	7.975	0.011	*
s(precipitation)	8.652	8.956	< 0.001	***	8.324	8.837	< 0.001	***
s(snow.height)	8.325	8.780	< 0.001	***	3.929	4.774	< 0.001	***
s(sunshine)	5.251	6.332	< 0.001	***	4.216	5.172	< 0.001	***
	$R^2_{\text{adj}} = 0.379$				$R^2_{\text{adj}} = 0.439$			
	Deviance explained = 36.1%				Deviance explained = 39.3%			
	$n = 51141$				$n = 49408$			

Note: Significant levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

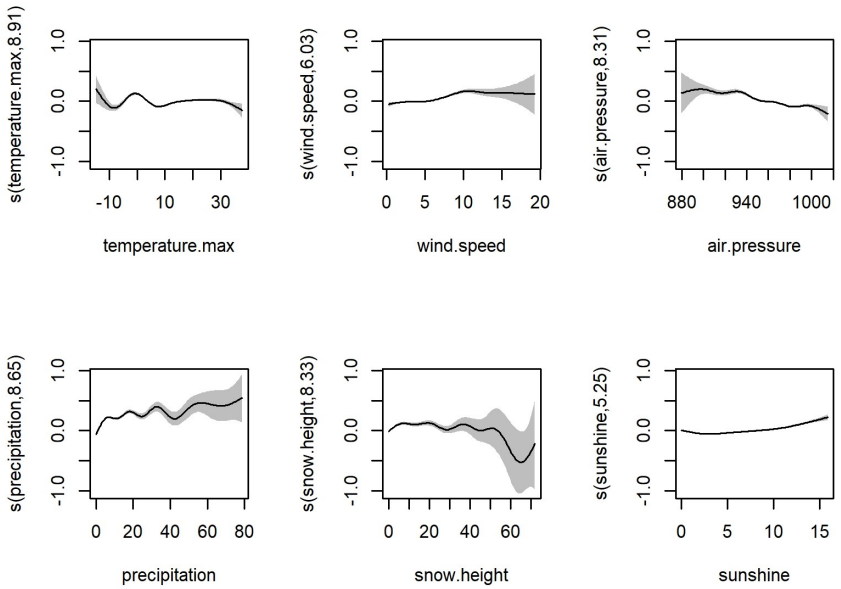

Fig. 3: Graphical output from R for GAM with Poisson distribution (partial effect plots show the component effect of each of the smooth terms in the model)

Fig. 3 illustrates the impact of specific meteorological variables on the incidence of accidents. A negative effect on the number of accidents is observed in temperatures between −5 and 0. The lowest incidence of accidents occurs at temperatures close to −10, which is likely to be a period with minimal traffic. Conversely, temperatures close to 10, which probably correspond to the spring season, are associated with a low number of accidents.

The occurrence of windy weather has been observed to result in an increase in the number of accidents. In our geographical conditions at the ground surface, the average wind speed is typically between 2 and 8 m/s, with rare instances of speeds exceeding 15 m/s (ČHMÚ, 2024).

The observed increase in air pressure is indicative of a corresponding decline in the number of accidents. In general, lower pressure is associated with the presence of precipitation, whereas higher pressure is associated with conditions of dry and sunny weather.

There are notable fluctuations in precipitation and snow depth. The period between 2014 and 2017 was one of the driest on record, with a number of concurrent hydrometeorological events, including the occurrence of weak winters, heat waves and rainless periods (ČHMÚ, 2018). Given the relatively dry conditions that have prevailed in recent years, inclement weather is an uncommon occurrence, and drivers are discouraged or more cautious. A lack of precipitation is associated with a reduction in the incidence of accidents.

The incidence of accidents tends to rise in line with an increase in levels of sunshine. The greatest number of accidents occurs outside of the winter months, specifically in the summer (Brázdil et al., 2021). This is attributable to a greater number of inexperienced drivers, an increase in the number of motorcyclists, cyclists and pedestrians on the roads, and the negative effects of high temperatures on the body, including fatigue and a decline in cognitive function, particularly in older vehicles lacking air conditioning (PČR, 2024).

A diagnosis of deviance residuals was made, as illustrated in Fig. 4. The results of the analyses demonstrated that the residuals did not exhibit the assumed properties. The issue was identified in the potential outliers, and thus a process of identification and exclusion was initiated. An outlier was considered to be any observation that exceeded $x_{0.75}+1.5\cdot\text{IQR}$ ($x_{0.75}$ is the upper quartile, IQR is the interquartile

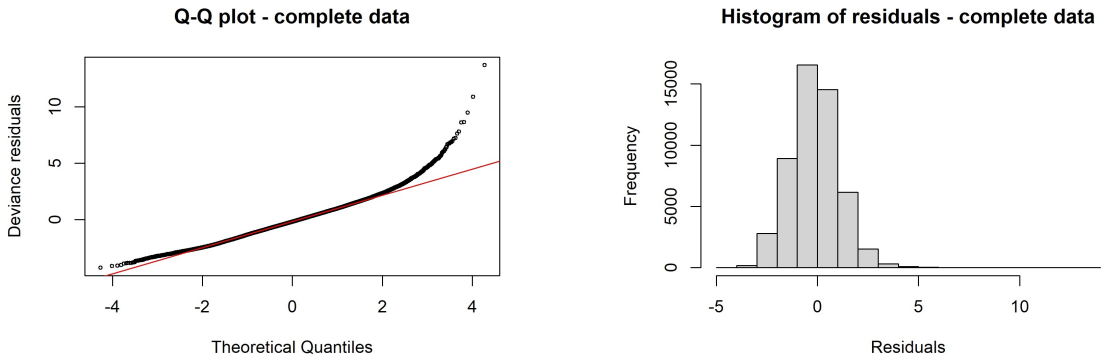

Fig. 4: Diagnostic graphs – model for complete data (the Q-Q plot of the residuals is displayed on the left, while the corresponding histogram is shown on the right)

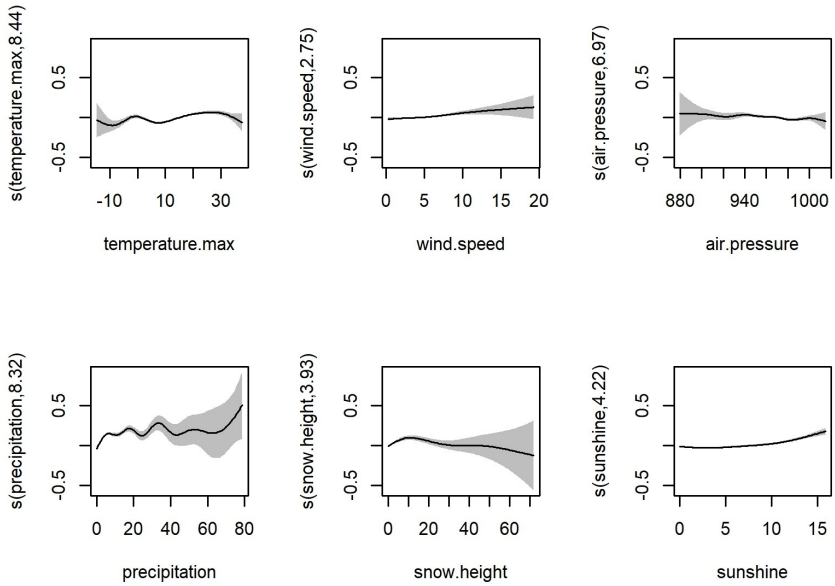

Fig. 5: Graphical output from R for GAM with Poisson distribution and outliers omitted (partial effect plots add up to the overall prediction)

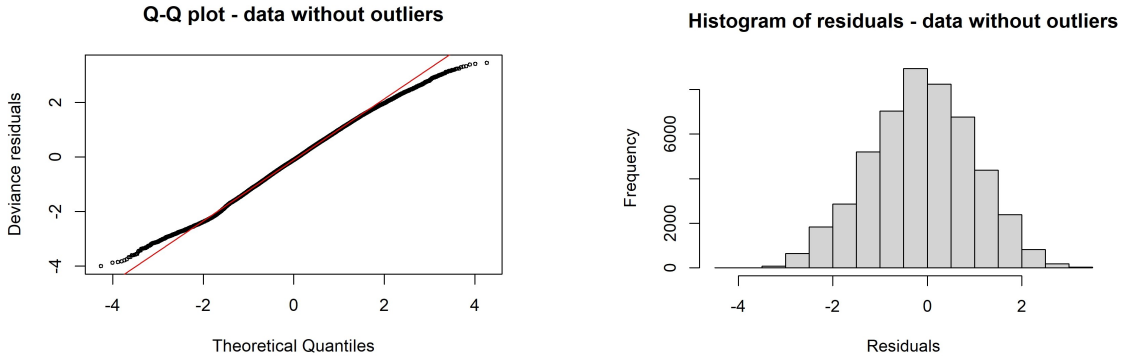

Fig. 6: Diagnostic graphs – model without outliers (the Q-Q plot of the residuals is displayed on the left, while the corresponding histogram is shown on the right)

range). The detection of outliers was performed in each region separately, resulting in a total of 1730 potential outliers being identified.

Additionally, the data were modelled after the exclusion of outliers. The principal distinction is the decline in the mean daily number of accidents by day, month, region, holiday and state of emergency. The model without outliers is better described by the probability distribution, with the GAM giving better results. The percentage of explained deviance increased to 39.3%. A comparison of the graphs in Fig. 3 and 5 shows that the number of accidents decreased during adverse weather conditions, including high maximum temperature, high wind speed, high or low air pressure, heavy precipitation and snow. It can therefore be hypothesised that these meteorological influences lead to the extreme daily number of accidents. A comparison of the plots of the smooth terms (Fig. 3 and 5) reveals that the influence of the air pressure variable diminished after the outliers were excluded. Nevertheless, the F-test continues to indicate statistical significance.

At last, the GAM model was validated. Tests of the GAM model for data with outliers included did not yield optimal results. This may be attributed to the fact that the data only approximately fit the theoretical probability distribution. For the model with outliers omitted, please refer to Fig. 6, which depicts a Q-Q plot comparing the model residuals to a normal distribution. Additionally, the histogram of the residuals exhibits a symmetric bell shape.

5 SUMMARY AND DISCUSSION

The objective of this study was to examine the relationship between the daily number of traffic accidents involving firefighter call-outs and meteorological measurements. The data from the period between 2012 and 2021 were subjected to analysis, with the objective of determining the dependence of the daily number of accidents in the country on meteorological variables, including temperature, pressure, wind, precipitation and sunshine. The present paper is loosely related to a study (Kolísková and Neubauer, 2024) in which the authors employ a GLM model to analyse the relationship between the number of accidents and the day of the week and month. In this paper, the GAM method was employed for modelling purposes. This entailed estimating the relationship between the number of accidents and meteorological variables, as well as the influence of additional factors such as the day of the week, month, region, holiday and the impact of the Coronavirus (Covid-19) emergency. Furthermore, the non-linear relationship between meteorological regressors and accidents was examined using smoothing splines.

The model output verifies that the observed variables are statistically significant. The JHM, MSK and STC regions have a higher incidence of accidents than the reference region, HKK. Nevertheless, comparisons between the regions are challenging due to the discrepancies in their sizes, populations, traffic network densities, geographical characteristics, and other factors. However, the regions exhibit a similar pattern of fluctuations in accident rates throughout the year, as evidenced by the findings of Brázdil et al. (2021, 2023). The number of accidents per day is below average during the spring months, particularly in March. The number of accidents per day is more pronounced from November to February, with the highest number of accidents occurring in summer and early autumn, particularly in September and October. This trend is undoubtedly influenced by the variability of traffic and climatic characteristics typical of the season. During the summer months, there is a notable increase in travel, whereas the winter season is characterised by the occurrence of sudden and unforeseen circumstances, such as the formation of ice and the melting of snow, which can lead to adverse road conditions. The lowest number of accidents per day is recorded on Sundays. Conversely, Monday and Thursday are associated with a higher incidence

of accidents. Friday is identified as the riskiest day of the week in terms of accidents, which can be attributed to fluctuations in traffic volume throughout the week. The variables associated with holidays and the impact of the pandemic have an attenuating effect on the value of the daily number of accidents.

The models demonstrate that a maximum daily temperature of −5 to 0 exerts a negative influence on the number of accidents. The lowest incidence of accidents is observed at temperatures approaching −10, when traffic levels appear to be minimal, and at temperatures close to 10, which correlates with the spring period when the number of accidents is relatively low. The occurrence of accidents is found to increase in the presence of windy weather conditions. An increase in pressure, which is typically associated with dry and sunny conditions, is associated with a reduction in the number of accidents. The most pronounced fluctuations are observed in average daily precipitation and total daily snow depth, which are related to historical climatic circumstances and driver habits. The 2014–2017 period was one of the driest on record, and drivers are more likely to be deterred from travel or to exercise greater caution when the weather is significantly adverse. The absence of collisions is indicative of minimal accident occurrence. Conversely, the incidence of accidents is higher in areas with greater levels of sunshine. The effects of rain, drizzle, snow, ice, fog and wind were examined by Brázdil et al. (2022), and the findings yielded comparable results. The highest number of accidents occurs outside of the winter months, with the summer period seeing the greatest number of accidents. This is likely due to the increased fatigue experienced by drivers in high temperatures, as well as other risk factors associated with extreme weather.

In line with the findings of several recent studies, we employed the use of generalised additive models, which have been specifically designed for the analysis of non-linear relationships between meteorological regressors and the number of traffic accidents (Anderson et al., 2015; Lepage and Morency, 2020). The non-parametric GAM method was employed for modelling a variable with a Poisson probability distribution. The methodology employed was contrasted with that of the Poisson GLM approach. The two methods were compared on the basis of the percentage of explained deviance, which was found to be 36.1% for GAM and 34.6% for GLM. The potential benefits of a negative binomial distribution were also considered, but this did not result in an improvement to the model. This distribution is typically recommended for models with real (overdispersed) data (Basu and Saha, 2017). However, in this case, the predictive ability of the model deteriorated by 1–2%. In light of the aforementioned findings, it can be concluded that the Poisson model is an appropriate choice for these data. Ultimately, a model that excluded outliers was estimated, which demonstrated an enhancement in the percentage of explained variance, reaching 39.3%. The proportion of outliers in the analysed data is approximately 3%, which equates to approximately 10 outliers per region per year. It is evident that some cases are directly attributable to sudden changes in meteorological conditions. For instance, the formation of ice on the road or the occurrence of snow calamities in regions with higher elevations can be attributed to such changes.

A comparison of GLM and GAM methods revealed that GAM may result in over-parameterisation of the model. However, the computational software can be optimally adjusted if sufficient data is available. For the more extreme values of the meteorological variables, the confidence intervals of the smooth functions exhibited considerable width at their upper and lower limits. This was a direct consequence of the low occurrence of outliers in the data. The GAM model is likely to be unsuitable for use over a large area with disparate conditions due to the extent of the generalisation involved. In future work, it may be possible to carry out an analysis using GAM modelling for individual regions of the country, where it will also be possible to examine outliers in greater detail. The model incorporates a considerable number of variables, yet the interactions between meteorological variables

remain unexplored. For instance, the impact of current rain and wind, which are associated with an increased risk of accidents, has not been investigated. Similarly, the influence of sunshine and snow cover, which can result in glare from reflection, has not been considered. These factors represent promising avenues for future research.

6 CONCLUSIONS

It has been demonstrated that substituting the GLM model with the GAM model markedly enhances the model's capacity to forecast the daily number of traffic accidents with firefighter call-outs. The GAM employs smooth functions to identify the optimal functional relationships between the explanatory variable and the regressors, a strategy that has been demonstrated to be advantageous for this data set. The extent of the improvement is contingent upon the data distribution employed. The models that demonstrated the greatest efficacy were those based on the Poisson distribution. The GAM model with a Poisson distribution and outliers omitted exhibited a percentage of explained deviance of 39.3%.

The paper's conclusion is that weather exerts a significant influence on the number of accidents, and that the relationship between weather and accidents is non-linear. The number of accidents is found to depend significantly on the maximum daily temperature, to decrease slightly with increasing air pressure, and to increase with increasing sunshine. In the context of the relatively arid climate of the Czech Republic, precipitation appears to exert its influence on the number of accidents primarily through its impact on driver behaviour.

The analysis of the impact of meteorological conditions on accident rates reveals a number of potential applications for such models. Such models may also be applied to road traffic accident models, for example, to enhance accident warning systems. The integration of meteorological data into traffic forecasts can facilitate enhanced navigation systems, improved traffic management, the identification of regions experiencing a high frequency of accidents contingent on weather conditions, and the distribution of heavy traffic. Moreover, the introduction of weather-related driving restrictions is anticipated to contribute to enhanced road safety. The potential for innovative solutions arises from its application in real-time traffic management systems and predictive modelling for accident prevention. This could result in the implementation of more intelligent, data-driven decisions regarding the deployment of rescue services and the implementation of traffic safety measures. It is of significant importance to increase awareness of traffic issues, particularly among those who are most vulnerable. This encompasses the implementation of preventative programmes for groups such as inexperienced drivers, which address the challenges posed by adverse driving conditions. Such factors may include, for instance, road slipperiness, driver visibility, reaction time and braking distance. Such factors may be more accurately estimated by an AI system than by a human operator in an autonomous vehicle. Such precautions may prove effective in reducing both the frequency of accidents and their associated economic and human costs.

7 ACKNOWLEDGEMENTS

This paper was supported from the research project Conduct of Land Operations, LANDOPS, Ministry of the Defence of the Czech Republic.

8 REFERENCES

Anderson, D., Davidson, R., Himoto, K. and Scawthorn, C. 2015. Statistical Modeling of Fire Occurrence Using Data from the Thoku, Japan Earthquake and Tsunami. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, 36 (2), 378–395. DOI: 10.1111/risa.12455.

Aultman-Hall, L., Lane, D. and Lambert, R. R. 2009. Assessing Impact of Weather and Season on Pedestrian Traffic Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, 2140 (1), 35–43. DOI: 10.3141/2140-04.

Basu, S. and Saha, P. 2017. Regression Models of Highway Traffic Crashes: A Review of Recent Research and Future Research Needs. *Procedia Engineering*, 187, 59–66. DOI: 10.1016/j.proeng.2017.04.350.

Becker, N., Rust, H. W. and Ulbrich, U. 2022a. Weather Impacts on Various Types of Road Crashes: A Quantitative Analysis Using Generalized Additive Models. *European Transport Research Review*, 14, 37 DOI: 10.1186/s12544-022-00561-2.

Becker, N., Rust, H. W. and Ulbrich, U. 2022b. Modeling Hourly Weather-Related Road Traffic Variations for Different Vehicle Types in Germany. *European Transport Research Review*, 14, 16. DOI: 10.1186/s12544-022-00539-0.

Brázdil, R., Chromá, K., Dolák, L., Řehoř, J., Řezníčková, L., Zahradníček, P. and Dobrovolný, P. 2021. Fatalities Associated with the Severe Weather Conditions in the Czech Republic, 2000–2019. *Natural Hazards and Earth System Sciences*, 21 (5), 1355–1382. DOI: 10.5194/nhess-21-1355-2021.

Brázdil, R., Chromá, K., Dolák, L., Zahradníček, P., Řehoř, J., Dobrovolný, P. and Řezníčková, L. 2023. The 100-Year Series of Weather-Related Fatalities in the Czech Republic: Interactions of Climate, Environment, and Society. *Water*, 15 (10), 1965. DOI: 10.3390/w15101965.

Brázdil, R., Chromá, K., Zahradníček, P., Dobrovolný, P. and Dolák, L. 2022. Weather and Traffic Accidents in the Czech Republic, 1979–2020. *Theoretical and Applied Climatology*, 149, 153–167. DOI: 10.1007/s00704-022-04042-3.

Cerna, S., Guyeux, C., Arcolezi, H. H., Couturier, R. and Royer, G. 2020. A Comparison of LSTM and XGBoost for Predicting Firemen Interventions. In Rocha, Á., Adeli, H., Reis, L., Costanzo, S., Orovic, I. and Moreira, F. (eds.). *Trends and Innovations in Information Systems and Technologies*. WorldCIST 2020. Advances in Intelligent Systems and Computing, vol. 1160, pp. 424–434. DOI: 10.1007/978-3-030-45691-7_39.

Clark, M. *Generalized Aditive Models* [online]. Available at: `https://m-clark.github.io/generalized-additive-models/introduction.html`. [Accessed 2024, April 4].

ČHMÚ. 2018. *Suché období 2014–2017: Vyhodnocení, dopady a opatření*. Praha: Český hydrometeorologický ústav. 1. vyd. ISBN 978-80-87577-81-3.

ČHMÚ. 2024. *Český hydrometeorologický ústav* [online]. Available at: `https://www.chmi.cz/files/portal/docs/meteo/om/sivs/vitr.html`. [Accessed 2024, July 15].

Guyeux, C., Nicod, J.-M., Varnier, C., Al Masry, Z., Zerhouny, N., Omri, N. and Royer, G. 2020. Firemen Prediction by Using Neural Networks: A Real Case Study. In Bi, Y., Bhatia, R. and Kapoor, S. (eds.). *Intelligent Systems and Applications*. IntelliSys 2019. Advances in Intelligent Systems and Computing, vol. 1037, pp. 541–552. Springer, Cham. DOI: 10.1007/978-3-030-29516-5_42.

Kolísková, P. and Neubauer, J. 2024. Analysis of Traffic Accidents and the Deployment of the Fire Rescue Service in the Czech Republic. *AD ALTA: Journal of Interdisciplinary Research*, 14 (1), 290–295.

Lepage, S. and Morency, C. 2020. Impact of Weather, Activities, and Service Disruptions on Transportation Demand. *Transportation Research Record: Journal of the Transportation Research Board*, 2675 (1), 294–304. DOI: 10.1177/0361198120966326.

Li, X., Lord, D. and Zhang, Y. 2011. Development of Accident Modification Factors for Rural Frontage Road Segments in Texas Using Generalized Additive Models. *Journal of Transportation Engineering*, 137 (1), DOI: 10.1061/(ASCE)TE.1943-5436.0000202.

Lord, D., Washington, S. and Ivan, J. N. 2005. Poisson, Poisson-gamma and Zero Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis & Prevention*, 37 (1), 35–46. DOI: 10.1016/j.aap.2004.02.004.

PČR. 2024. *Archiv zpravodajství* [online]. Available at: `https://www.policie.cz/clanek/dopravni-rizika-v-letnich-mesicich.aspx`. [Accessed 2024, July 15].

Pop, D. 2018. Generalised Poisson Linear Model for Fatal Crashes Analysis. *Romanian Journal of Automotive Engineering*, 24 (4), 131–134.

Rodríguez-Pérez, J. R., Ordóñez, C., Roca-Pardiñas, J., Vecín-Arias, D. and Castedo-Dorado, F. 2020. Evaluating Lightning-Caused Fire Occurrence Using Spatial Generalized Additive Models: A Case Study in Central Spain. *Risk Analysis*, 40 (7), 1418–1437. DOI: 10.1111/risa.13488.

Roh, H.-J. 2020. Assessing the Effect of Snowfall and Cold Temperatureon a Commuter Highway Traffic Volume Using Several Layers of Statistical Methods. *Transportation Engineering*, 2, 100022. DOI: 10.1016/j.treng.2020.100022.

Thorsson, S., Lindqvist, M. and Lindqvist, S. 2004. Thermal Bioclimatic Conditions and Patterns of Behaviour in an Urban Park in Göteborg, Sweden. *International Journal of Biometeorology*, 48 (3), 149–156. DOI: 10.1007/s00484-003-0189-8.

Wood, S. N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. New York: Chapman and Hall/CRC. DOI: 10.1201/9781315370279.

Yannis, G. and Karlaftis, M. G. 2010. Weather Effects on Daily Traffic Accidents and Fatalities: Time Series Count Data Approach. In *Proceedings of the 89th Annual Meeting of the Transportation Research Board*, Washington, 17 pp.

Zhang, Y., Xie, Y. and Li, L. 2012. Crash Frequency Analysis of Different Types of Urban Roadway Segments Using Generalized Additive Model. *Journal of Safety Research*, 43 (2), 107–114. DOI: 10.1016/j.jsr.2012.01.003.

AUTHOR'S ADDRESS

Petra Kolísková, University of Defence, Kounicova 65, Brno, Czech Republic, e-mail: petra.koliskova@unob.cz (corresponding author)

Jiří Neubauer, University of Defence, Kounicova 65, Brno, Czech Republic, e-mail: jiri.neubauer@unob.cz