Count Data Prediction with Poisson Regressions on Poisson-Mixture Locations: Application to Traffic Counts in Prague Areas

Evženie Uglickich*[1], Ivan Nagy[1,2]

[1]*Department of Signal Processing, Institute of Information Theory and Automation, Czech Academy of Sciences, Pod vodárenskou věží 4, 18208 Prague, Czech Republic, uglickich@utia.cas.cz, ORCID: https://orcid.org/0000-0003-1764-5924*
[2]*Faculty of Transportation Sciences, Czech Technical University, Na Florenci 25, 11000 Prague, Czech Republic, nagy@utia.cas.cz, ORCID: https://orcid.org/0000-0002-7847-1932*

Abstract

The paper addresses the task of modeling and predicting count data, with the application to traffic counts at selected urban roads in Prague. The methodology proposed in the paper presents the following key ideas: (i) analysis of explanatory multiple counts and detection of their locations through recursive Bayesian estimation of Poisson mixtures; (ii) estimation of the target count model via local Poisson regressions at recognized locations; and (iii) prediction of target counts through real-time location detection. The algorithm's properties are first investigated using simulated data and then validated with real traffic counts. Experimental results demonstrate that the proposed algorithm outperforms alternative methods in predicting traffic count data across various quality metrics, even for weakly correlated data. The main contribution of the paper is the development of a novel approach for online target count prediction, which simultaneously analyzes the spatial locations and temporal evolution of multivariate explanatory count data.

Keywords: count data prediction, traffic counts, local Poisson regression, recursive Bayesian estimation of Poisson mixtures

1 Introduction

This paper presents an advanced modeling and prediction approach for analysis of count data applied to traffic counts at selected urban roads in Prague. Understanding traffic patterns is crucial for optimizing public transportation systems, controlling traffic flow, improving safety, urban planning, and assessing environmental impacts. Using the definition of traffic count data as the number of vehicles passing through a selected road section per unit of time, the framework built upon Poisson distribution models has been chosen to develop a progressive methodology, which was then compared with the existing original models.

Due to the nature of count data, the Poisson distribution can fit the data well for analysis purposes. The assumption of equidispersion limits the practical use of the Poisson distribution with empirical count data. When the assumption is violated, the Negative Binomial (NB) distribution [1], Generalized Poisson models (GPM) [2, 3] or mixtures of Poisson distributions [4] become key tools for describing univariate under/over-dispersed counts. Zero-inflated

*Corresponding author. Department of Signal Processing, Institute of Information Theory and Automation, Czech Academy of Sciences, Pod vodárenskou věží 4, 18208 Prague, Czech Republic. Tel: +420 266 052 358, email: uglickich@utia.cas.cz, ORCID: https://orcid.org/0000-0003-1764-5924

Poisson (ZIP), zero-inflated NB (ZINB) and compound Poisson (CP) models are commonly employed to handle the excess zeros often observed in count data. Zero-truncated Poisson (ZTP) and zero-truncated NB (ZTNB) models are suitable for conducting a targeted analysis of non-zero observations within count data.

Predictive models aimed at understanding the relationship between multiple counts cannot use these Poisson-based distributions because the Poisson distribution lacks a suitable general conditional form. Multivariate Poisson-based distributions [5] for multidimensional count data are less developed for practical use due to the limitation of independence consumption. In this area, [6] provide an advanced review on multivariate distributions derived from Poisson model dividing them among marginal Poisson distributions, Poisson mixtures and conditional Poisson-based distributions. According to [6], studies focusing on conditional Poisson generalizations are based on the use of Poisson graphical models or Markov random fields, specified by node-conditional distributions [7, 8]. Among them, the approach that is relatively close to the one described in this paper is to model multivariate count data by estimating local Poisson regressions conditioned by node-neighbors of the variables in the form of local Poisson graphical models [9–11].

Poisson regression is an important approach in predictive modeling of count data [12–17]. To address challenges such as over- or underdispersion of count data, the underlying alternative is NB regression [18–20], which introduces auxiliary dispersion parameters to be estimated. A valuable tool for predictive modeling of count data, which allows avoiding Poisson model assumptions, are mixtures of regression models, such as mixtures of Poisson and Poisson-based (ZIP/CP/Tweedie) regressions [21–28], NB/ZINB regressions [29–33] and Poisson-Gamma/Poisson-Gaussian models [34–36].

1.1 Poisson-related models in transportation domain

According to the literature, the use of Poisson models in transportation domain has a long history [37–40]. Relatively recent studies in the analysis of traffic counts have developed methods that use Poisson and NB models. In the paper of [41], the authors proposed a spatial-temporal NB regression with the temporal correlation of traffic volumes on multiple roads at previous time segments. [42] introduced a traffic flow prediction method for intercity roads using convolved bilinear Poisson regression with the incorporated latent factor model. Adopting the stochastic variational Bayes method, they continually updated model parameters online based on the most recent observations instead of all past data, which allowed them to enhance robustness against sparse observations and provide accurate predictions of intercity traffic flow with dynamically changing patterns. The authors of [43] explored a combination of the cosinor regression with Poisson, GPM, ZIP, NB and ZINB models for a developed software tool in application to traffic counts observed during the COVID-19 epidemic. In [44], the traffic flow state prediction model was proposed based on the relationship between traffic flow states in different urban locations estimated via recursive Bayesian Poisson mixture estimation using mixture pointers to identify clusters of actually measured traffic counts.

As far as can be seen from the literature, Poisson-related models are more popular in the analysis of count data in other transportation applications, such as vehicle crash count data [45–49], bicycle counts [50], and pedestrian traffic counts [51, 52].

1.2 Alternative traffic count data models

Numerous publications are based on different interpolation methods, such as e.g. spatial interpolation with Kriging-based methods [53], a hybrid ANN-fuzzy approach [54], combined pattern-matching and Bayesian statistics [55], etc. The extensive overview of traffic prediction methods can be found in [56], where existing algorithms are divided among time series models (both parametric and non-parametric), optimization approaches and neural networks.

Recent advances in traffic count data prediction have focused on improving model accuracy and accommodating complex patterns. [57] used fully-connected feedforward multi-layer ANN to address the problem of estimating historical hourly traffic volumes that transportation agencies need for planning and statewide performance measurement. In [58], a combination of bootstrap aggregation with parametric ARIMA models was explored to improve the traffic flow prediction accuracy. Hybrid methods based on ARIMA models were considered by [59–62]. [63] presented a spatio-temporal interpolation method that detects similar trends in the behavior of traffic counts at different locations. In the study [64], the authors considered traffic volume prediction based on vehicle detection and road characteristics extraction from Google aerial images via region-based convolutional neural networks. [65] focused on estimation of turning movement count data by means of a multi-output multilayer ANN with exogenous variables from intersection infrastructure point-of-interest data. [66] evaluated the temporal variation of traffic flows during the COVID-19 pandemic lockdown by means of comparison of the smartphone-based traffic count predictions with traffic quantification methods.

1.3 Main features of the presented paper

The author of [46] noted that although Poisson and NB models have been used extensively for the analysis of count data, the distributional assumptions make them more appropriate for cross-sectional count data rather than time series count data, where these assumptions limit the consideration of serial correlation of observations, which can be critical in the case of traffic counts. In this paper we present an approach to understanding the behavior of cross-sectional multiple count data by analyzing their locations recognized in the data space and capturing the evolution of the data over time with the *main aim of predicting the target count variable*. In the paper, these *locations* are defined as clusters of data. In general, they can be identified within the framework of preliminary offline data analysis, e.g., using clustering algorithms [67, 68], etc. However, the use of model-based location identification provides the pre-estimated parameterized models which, in the case of their recursive estimation, can be used for re-learning during the online prediction phase. The relationship between the count variable selected as the target and the rest of multiple counts utilized as explanatory data is described by estimating local Poisson regressions in the detected locations, which should be recognized online in order to use the pre-estimated (or re-learned) location models for prediction. Thus, the main features of the presented paper are as follows:

- Analysis of explanatory multiple counts and detection of their locations with the help of recursive Bayesian estimation of Poisson mixtures;
- Target count model estimation via local Poisson regressions at the locations;
- Target count prediction using online location detection;
- Application of the proposed approach to traffic counts at selected locations in Prague.

The proposed solution continues the line started in the paper by [69], who considered explanatory data of a mixed nature, including count and categorical data measured in real time, i.e., locations were directly indicated by the measurements. Here, this relatively trivial task is elaborated for recursive Poisson mixture estimation for unknown locations with subsequent construction of Poisson regressions on them.

The remainder of this paper is organized as follows: Section 2 briefly recalls basic facts about the models and algorithms used. In Section 3, the theoretical background of the proposed approach is presented, including the problem specification and a detailed description of the general solution to the problem. The properties and advantages of the proposed theoretical approach are then thoroughly investigated through experiments with simulations under various conditions and comparisons with Poisson regression. Section 4 evaluates the empirical performance of the proposed technique in application to traffic counts, comparing it with existing algorithms that also focus on addressing the

prediction task, and discusses potential applications and practical limitations of the approach. Finally, Section 5 concludes the paper by summarizing the main contributions of the approach and providing directions for future research.

2 Preliminaries

To facilitate the presentation of the main idea, basic facts about the Poisson model, Poisson regression, and recursive Bayesian mixture estimation are recalled in this section, along with a discussion of their limitations in the context under consideration.

2.1 Basis for Poisson model

The Poisson distribution describes the behavior of a discrete valued random variable x_t measured at discrete time instants $t = 1, 2, \ldots$, whose realizations represent a number of random independent events per time unit, i.e., x_t is a count variable and its Poisson model is given by the probability distribution

$$f(x_t|\lambda) = e^{-\lambda}\frac{\lambda^{x_t}}{x_t!}, \tag{1}$$

where the unknown non-negative parameter λ expresses both the expectation and the variance of x_t. The point estimate of the parameter λ is known to be obtained as the average of the count observations $\hat{\lambda} = 1/T\sum_{t=1}^{T} x_t$. The assumption regarding the parameter λ and the lack of the conditional form in (1) are known to be the main limitations of this model in practice.

2.2 Basis for Poisson regression

The Poisson regression models the relationship between the target count Poisson-distributed variable y_t and a vector of independent explanatory variables[1] $x_t = [x_{1;t}, x_{2;t}, \ldots, x_{N;t}]'$ via the logarithm of its conditional expectation $\ln(E[y_t|x_t])$ equal to a linear combination of unknown regression coefficients $\theta = [\theta_0, \theta_1, \ldots, \theta_N]'$

$$\ln(E[y_t|x_t]) = \theta_0 + \theta_1 x_{1;t} + \theta_2 x_{2;t} + \ldots + \theta_N x_{N;t}, \tag{2}$$

which can also be written inversely

$$\lambda_y \equiv E[y_t|x_t] = e^{\theta_0+\theta_1 x_{1;t}+\theta_2 x_{2;t}+\ldots+\theta_N x_{N;t}} \equiv e^{\theta' x_t}, \tag{3}$$

where λ_y stands for the expectation and the variance of y_t and the vector x_t is concatenated to 1. The well-known maximum likelihood estimation of the Poisson regression parameters is obtained by maximizing the log-likelihood function of θ with substituted training data

$$\ln L(\theta|y_t, x_t) = \ln \prod_{t=1}^{T} \frac{1}{y_t!} e^{-e^{\theta' x_t}} e^{y_t \theta' x_t}, \tag{4}$$

where T corresponds to the capacity of the training data set. Since the maximization of $\ln L(\theta|y_t, x_t)$ has no close-form solution, the point estimates of the regression coefficients θ are computed numerically as

$$\hat{\theta} = \arg\max_{\theta} \ln L(\theta|y_t, x_t). \tag{5}$$

Practical limitations of Poisson regression are related to the Poisson assumption of equidispersion of the target variable and to offline parameter estimation.

[1] In general, independent variables in Poisson regression need not be count data, but here we assume that x_t is a vector of count variables.

2.3 General scheme of recursive Bayesian mixture estimation

The general scheme of the recursive Bayesian mixture estimation theory [70–73] is used in this paper. According to this theory, data behavior is described by a mixture of parameterized components in the general form $f_i(\text{data}_t \mid \mathcal{P}_c)$, $i = \{1, 2, \ldots, N_c\}$, representing specific either static or dynamic distributions with their parameters $\mathcal{P}_c$. An essential part of the adopted methodology's mixture model is a pointer [70], which is an unmeasurable discrete random variable with realizations $i \in \{1, 2, \ldots, N_c\}$ indicating the component active at time t. Generally, the parameterized model of the pointer takes the form of a static/dynamic categorical distribution $f(\text{pointer}_t \mid \mathcal{P}_p)$ with the parameters $\mathcal{P}_p$ representing probabilities of component activity. The primary objectives of the recursive Bayesian mixture estimation are (i) component parameter estimation (clustering) and (ii) pointer estimation (classification), both based on continuously measured data and performed online in real time.

By applying the Bayes and chain rules [74] to the joint probability density function (pdf) of all unknown parameters and the pointer, the general scheme for deriving their posterior pdf [70, 71] is

$$\begin{aligned} f(\mathcal{P}_c, \mathcal{P}_p, \text{pointer}_t \mid \text{all data}) &\propto f(\text{data}_t, \mathcal{P}_c, \mathcal{P}_p, \text{pointer}_t \mid \text{past data}) \\ &= f_i(\text{data}_t \mid \mathcal{P}_c) \underbrace{f(\mathcal{P}_c \mid \text{ past data})}_{\text{conjugate prior}} \times f(\text{pointer}_t \mid \mathcal{P}_p) \underbrace{f(\mathcal{P}_p \mid \text{ past data})}_{\text{conjugate prior}}, \end{aligned} \tag{6}$$

where appropriate conjugate prior pdfs are selected for the distributions of the components used $\forall i \in \{1, 2, \ldots, N_c\}$, and the conjugate prior Dirichlet pdf is applied for estimating the parameters of the pointer model [71].

According to [70, 71, 73], the relation (6) is marginalized over the unknown parameters. In one-pass estimation, this gives the proximities of the data value measured at time t to the components [44, 73, 75]. In the normalized form (generally multiplied by the prior estimate of $\mathcal{P}_p$), the proximities provide weights of the components. The weights are incorporated into recursive updates of the statistics of the involved conjugate pdfs. The updated statistics are then used to recompute the prior point estimates of $\mathcal{P}_c$ and $\mathcal{P}_p$. This part of the scheme is utilized to update the location of the components. The maximum weight provides the point estimate of the pointer at time t, which classifies the actual data value into the active component. More details can be found in the mentioned sources.

Among the specific limitations of this methodology is the necessity to operate with distributions that have reproducing statistics to ensure closed-form computations. The methodology has been well-studied for mixtures of normal and categorical components [70–72]. The general approach for different types of components with reproducing statistics was presented by [73]. Subsequently, it has been applied to exponential [76, 77], uniform [78], binomial [79], and Poisson [44] components. A recursive algorithm for a mixture of Poisson regressions using the mentioned approach is not feasible due to the lack of a closed-form estimator. This motivated the development of a technique presented in this paper, which describes the relationship between multiple counts in detected locations.

The general theory of the presented technique is presented in the section below.

3 Online Count Prediction with Local Poisson Regressions on Poisson-Mixture Locations

3.1 General problem specification

The problem formulation is based on the need to predict the values of the target count variable y_t under the condition of continuously available multiple independent explanatory counts x_t, i.e., *the main aim* is the online prediction of y_t. Let us consider the vector $[y_t, x_t]' = [y_t, x_{1;t}, x_{2;t}, \ldots, x_{N;t}]'$ with all realizations measured for time $t \leq T$,

but only x_t observed for time $t > T$. It implies that through analysis of the explanatory count vector x_t measured simultaneously with the target y_t the behavior of y_t has to be explained. To avoid the unfulfilled single-distribution assumptions, an approach based on local models is appropriate, as it enables the capture of the data behavior under different conditions. Consequently, the locations for constructing the local models need to be determined, which means the data should be clustered. Moreover, in order to utilize the explanatory counts x_t online for time $t > T$, the solution is expected to facilitate online identification of locations so that the data can be classified to them later in the prediction phase. In the considered context, assuming the independence of individual explanatory counts and recalling the general definition of locations as clusters, we propose achieving this by combining local Poisson regressions on Poisson locations. The subgoals of the paper are specified as follows:

For $t = 1, 2, \ldots, T$ (*offline analysis of training data*)

1. $\forall j \in \{1, 2, \ldots, N\}$ and $\forall i \in \{1, 2, \ldots, N_c\}$, identify the i-th location of the explanatory count $x_{j;t}$ as the Poisson component (1) and recursively (pre-)estimate the component parameters via (6);
2. Estimate the Poisson mixture pointer in the joint explanatory data space;
3. $\forall i \in \{1, 2, \ldots, N_c\}$, at the i-th location indicated by the pointer, estimate the parameters of the Poisson regression (3) to explain the dependence between y_t and all $x_{j;t}$ classified to this location;

For $t = T + 1, T + 2, \ldots$ (*online prediction with testing data*)

1. Estimate the Poisson mixture pointer and detect the active location of the count vector x_t;
2. Optionally re-estimate the component parameters;
3. Predict the expectation of the target y_t using the local pre-estimated Poisson regression (3) at the active location by incorporating the actual measured explanatory counts x_t.

3.2 General solution steps

3.2.1 Training data analysis

Mixture initialization Identifying the locations in the data space of the explanatory counts x_t for $t = 1, 2, \ldots, T$ via the general scheme (6) inspired by [70, 71] requires the initialization of the recursive mixture estimation algorithm, namely, the number N_c of Poisson components (1) and their initial statistics should be set before the algorithm start. In this paper, independent explanatory counts $x_{j;t}$ $\forall j \in \{1, 2, \ldots, N\}$ are modeled individually. This essentially facilitates the mixture initialization procedure through heuristic techniques of prior count histogram analysis, which involves guessing the number of single Poisson distributions along with prior point estimates of their parameters (it will be demonstrated in Section 4). Denoting the prior point estimate of the i-th component parameter of each count $x_{j;t}$ $\forall j \in \{1, 2, \ldots, N\}$ and $\forall i \in \{1, 2, \ldots, N_c\}$ by $\hat{\lambda}_{i;j;t}$, the initial statistics is straightforward

$$\mathcal{S}_{i;j;t} = \hat{\lambda}_{i;j;t} \kappa_{i;j;t}, \quad \text{for } t = 0, \tag{7}$$

where $\kappa_{i;j;t}$ is the initial counter initialized to 1.

Location search in explanatory data space The scheme (6) applied to the Poisson mixture with the gamma prior conjugate pdf with the initialized statistics $\mathcal{S}_{i;j;t}$ and $\kappa_{i;j;t}$ (see derivations in [80]) suggests the recursive update of the statistics in the form

$$\mathcal{S}_{i;j;t} = \mathcal{S}_{i;j;t-1} + w_{i;t}x_{j;t}, \quad \kappa_{i;j;t} = \kappa_{i;j;t-1} + w_{i;t}, \quad \text{for } t = 1, 2, \ldots, T, \tag{8}$$

where the weight $w_{i;t}$ expresses the probability of activity of the i-component at time t in the joint data space of explanatory data counts. The weights are computed using the entry-wise product of the proximities $\mathcal{M}_{i;j;t}$ [73] of the counts $x_{j;t}$ to the i-th components $\forall i \in \{1, 2, \ldots, N_c\}$, which provides the joint proximity $\mathcal{M}_{i;t}$ (recalling the explanatory count independence assumption) as follows:

$$\mathcal{M}_{i;t} \propto \underbrace{\int_{\lambda^*_{i;1}} \underbrace{f_i(x_{1;t} \mid \lambda_{i;1})}_{\text{Component (1)}} \underbrace{f(\lambda_{i;1} \mid \{x_{1;t}\}_{t=0}^{T-1})}_{\text{gamma conjugate prior}} \, d\lambda_{i;1}}_{\mathcal{M}_{i;1;t}} \odot \ldots \odot \underbrace{\int_{\lambda^*_{i;N}} \underbrace{f_i(x_{N;t} \mid \lambda_{i;N})}_{\text{Component (1)}} \underbrace{f(\lambda_{i;N} \mid \{x_{N;t}\}_{t=0}^{T-1})}_{\text{gamma conjugate prior}} \, d\lambda_{i;N}}_{\mathcal{M}_{i;N;t}}, \tag{9}$$

where the prior point estimate of the pointer in (6) is assumed to be uniform and omitted for simplicity. The results $\mathcal{M}_{i;j;t}$ of individual integrals in (9) are called the proximities and represent the computed values of the Poisson probability functions (1) with the incorporated realizations of the count $x_{j;t}$ at time t and the prior point estimates $\hat{\lambda}_{i;j;t-1}$ of the component parameters. The weights at time t are thus simply given by normalizing the joint proximities

$$w_{i;t} = \frac{\mathcal{M}_{i;t}}{\sum_{i=1}^{N_c} \mathcal{M}_{k;t}} \tag{10}$$

and utilized in (8). The updated statistics (8) naturally provide the re-computed point estimates of the Poisson component parameters

$$\hat{\lambda}_{i;j;t} = \frac{\mathcal{S}_{i;j;t}}{\kappa_{i;j;t}} \tag{11}$$

$\forall j \in \{1, 2, \ldots, N\}$ and $\forall i \in \{1, 2, \ldots, N_c\}$ at time t, which are substituted into the relation (9) instead of $\lambda_{i;j}$ at the next time instant under the adopted recursive methodology. This part of the solution corresponds to step 1 in the time loop for $t = 1, 2, \ldots, T$ of *analysis of training data* in Section 3.1. Subsequently, the locations of the explanatory independent counts are identified through their models in the form of Poisson components, which are pre-estimated for later use in the prediction phase and can also be optionally re-estimated.

For step 2 of the above time loop, the pointer of the Poisson mixtures in the joint explanatory data space is estimated as the argument of the maximum value of the whole weighting vector

$$\arg\max w_t \equiv \arg\max [w_{1;t}, \ldots, w_{N_c;t}]' \tag{12}$$

at each time t. The obtained index $i \in \{1, 2, \ldots, N_c\}$ of the maximum weight denotes the component where the value observed at time t should be classified.

In this way, steps 1 and 2 via (9), (10), (8), (11), and (12) are prepared to be used as long as new measurements of the explanatory counts arrive.

Local Poisson regressions on Poisson-mixture locations As stated in step 3 of the time loop for $t = 1, 2, \ldots, T$ of *analysis of training data* in Section 3.1, the pointer estimated in (12) is necessary to model the spatio-temporal relationships of the explanatory counts x_t and the target count y_t measured simultaneously at the same locations

indicated at time t. This is achieved by estimating the regression coefficients of N_c Poisson regressions (3) at the N_c locations. This procedure is specified as the search for N_c sets of time indices τ_i such that

$$\tau_i = \{t | \underbrace{\arg\max w_t = i}_{\text{via (12)}}\},\ t = 1, 2, \ldots, T,\ \forall i \in \{1, 2, \ldots, N_c\}. \tag{13}$$

Following this, the data set $\{y_{\tau_i}, x_{1;\tau_i}, x_{2;\tau_i}, \ldots, x_{N;\tau_i}\}$ for each i includes the target and multiple explanatory counts in the i-th Poisson-mixture location detected by the above algorithm. These sets are then incorporated into the log-likelihood function (4) to compute the point estimates of the regression coefficients $[\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_N]'_i$ at each i-th location numerically, as defined by (5). This part of the solution completes the analysis of the training data.

3.3 Online testing data prediction

Online location recognition At this stage of the solution for the time $t = T + 1, T + 2, \ldots$ (refer to steps 1 and 2 of *online prediction with testing data* in Section 3.1), the currently measured explanatory counts $x_{j;t}$ are used for the online prediction of the target count y_t. There are two options for using the pre-estimated Poisson models obtained during the training data analysis:

1. **Use of the final point estimates of the Poisson component parameters** Here, the point estimates $\hat{\lambda}_{i;j;t=T}$ are continuously substituted into the relation (9) along with the values of of the explanatory counts $x_{j;t}$ observed at the actual time instant $t = T + 1, T + 2, \ldots$, each into their component (1). This computes the real-time proximity $\mathcal{M}_{i;t}$, which, via (10) and (12), provides the required point estimate of the real-time pointer, allowing the detection of the active location.

2. **Re-estimation of the Poisson component parameters** In this case, the above option is enhanced by the recursive update of statistics (8) to re-compute (11) and utilize the new point estimates $\hat{\lambda}_{i;j;t}$ in (9) at each time t or for occasional re-learning. This enables refining weights of the components during the online prediction.

At each time, both options provide the current value of the pointer (12), indicating the active location where the current explanatory counts $x_{j;t}$ belong.

Target count prediction At the final step 3 of the time loop for the time $t = T + 1, T + 2, \ldots$, the vector of the currently measured explanatory counts x_t, classified in the previous step, is utilized in the local Poisson regression (3) for the online prediction of the target count y_t. Embedding the current realizations into (3) allows computation of the expectation of the target count

$$\hat{y}_t = \exp\{[\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_N]'_{i=\arg\max w_t} x_t\} \tag{14}$$

with the point estimates of the regression coefficients $[\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_N]'_{i=\arg\max w_t}$ corresponding to the location $i = \arg\max w_t$ indicated by the current pointer value (12) at time t.

3.4 Properties and limitations of the proposed approach

The properties and limitations of the proposed approach were first investigated through simulations using Scilab (www.scilab.org), a free and open-source programming environment for engineering and scientific computations. In the local-model-based approach, the relationship between the quality of the target count prediction and the accuracy of classification of explanatory counts can have a significant impact. To investigate this relationship, we explored

the dependence between the coefficient of determination R^2 and the accuracy of classification $\mathcal{A}_{cc}$. Here, R^2 is defined as

$$R^2 = 1 - \frac{\sum_{t=T+1}^{K}(y_t - \hat{y}_t)^2}{\sum_{t=T+1}^{K}(y_t - \bar{y})^2} \tag{15}$$

where $K - T$ represents the capacity of testing simulated data sets, and $\hat{y}_t$ is the predicted value of the target count from (14). Additionally, $\mathcal{A}_{cc}$ is defined as the ratio of the number of correct pointer estimates (12) to the total number of pointer estimates calculated during the prediction part of the algorithm.

For this part of the validation experiments, randomly generated parameters were used for the four Poisson components of explanatory counts, and the mixture initialization, identical for all experiments, was set to exclude the influence of experimental settings. Each simulated data set contained 1000 realizations of two explanatory counts $x_{1;t}$ and $x_{2;t}$ as well as the target count y_t.

To examine whether lower accuracy in classifying explanatory count data x_t affects the prediction quality of the target count variable y_t, three types of experiments were conducted. They involved simulations from:

1. Distant Poisson components with highly different parameters for both $x_{1;t}$ and $x_{2;t}$;
2. Close Poisson components with parameters close to each other for both $x_{1;t}$ and $x_{2;t}$;
3. A mix of both, featuring distant Poisson components for $x_{1;t}$ and close components for $x_{2;t}$.

The distance between the components was defined as the range of absolute differences between the component parameters of each explanatory count

$$\mathcal{R}_j = \max(|\lambda_{i;j} - \lambda_{k;j}|) - \min(|\lambda_{i;j} - \lambda_{k;j}|), \;\; \forall i, k \in \{1, 2, \ldots, N_c\}, \;\; \forall j \in \{1, 2, \ldots, N\}. \tag{16}$$

Each type of the experiments was conducted 30 times with a different choice of standard deviations used for random generation of the Poisson component parameters. Table 1 shows the average values of the ranges $\mathcal{R}_1$ and $\mathcal{R}_2$ of the component parameters of the explanatory counts, as well as the average values of R^2 and $\mathcal{A}_{cc}$ in percent obtained for each type of experiment, in comparison with the R^2 of the classical Poisson regression denoted by R^2_{PR}. The

Table 1: Explanatory data classification vs. target data prediction

Experiment type	$\mathcal{R}_1$	$\mathcal{R}_2$	$\mathcal{A}_{cc}$	R^2	R^2_{PR}
Distant components	15.318	17.917	95.854	93.798	86.258
Close components	5.209	5.721	61.924	49.047	39.358
Mixed components	39.712	4.850	98.093	92.334	89.364

table shows that $\mathcal{A}_{cc}$ is naturally higher for distant components, which are generally more easily recognizable, although classifying Poisson components with at least approximate equidispersion is not trivial, unlike Gaussian components; see histograms of distant and mixed components in Figure 1 (left). The closer the components are, the lower the accuracy $\mathcal{A}_{cc}$, especially for small values of the parameters and consequently with small variances. As expected, the target prediction quality, determined through $\mathcal{R}_2$, is affected by the lower $\mathcal{A}_{cc}$ and decreases as well (note the y-axis values in the plots of Figure 1 (right)). However, for all types of experiments shown in Figure 1 (right) and in Table 1, it holds that $\mathcal{R}_2 > R^2_{PR}$, meaning that the quality of prediction with the proposed method is higher than with classical Poisson regression. This holds true even for the case of close components, where the simulations have been prepared with higher standard deviations during generation, causing the data to almost lose their multimodality. This facilitates the prediction for Poisson regression, and the difference between $\mathcal{R}_2$ and R^2_{PR}

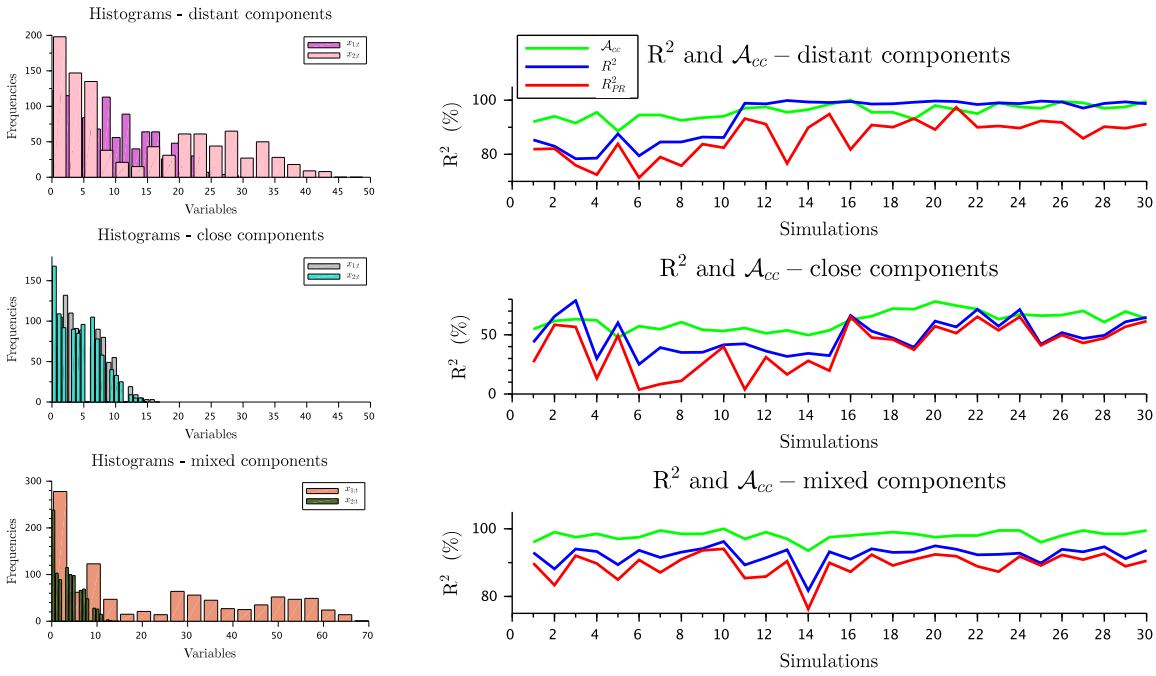

Figure 1: Explanatory data classification accuracy impact on target data prediction

is small for some of the simulations in Figure 1 (right). The improvements in prediction resulting from the proposed method are clearly evident.

Another important property of the proposed algorithm is that the target count variable y_t does not need to have a distinctly multimodal character to capture its relationship with the explanatory counts at their recognized locations. This can be seen in Figure 2, where the histograms of the target count y_t for each type of the above experiments lack distinct peaks. This indicates that clustering of the count variable y_t in the joint data space with explanatory variables $x_{1;t}$ and $x_{2;t}$ will not provide new information. Even in the case of such clustering, the mixture initialization would be a rather challenging task.

More experiments comparing the proposed approach with alternative existing algorithms applied to real traffic counts will be presented later in Section 4. Here, the general properties of the approach have been investigated and are summarized in Table 2.

Table 2: General properties of the approach compared to alternative methods

Alternative methods	The proposed approach
Poisson/ZIP/ZTP regression	No single-distribution assumptions
NB/ZINB/ZTNB regression, GPM	Local models for their later recognition in prediction
Mixture of Poisson/NB-related regressions, CP, Poisson-Gamma/Gaussian models	One-pass recursive estimation for online re-learning/prediction
Centroid-based clustering	Parametrized cluster models for prediction

Investigating the properties of the proposed algorithm, its limitations were also outlined. Naturally, a significant

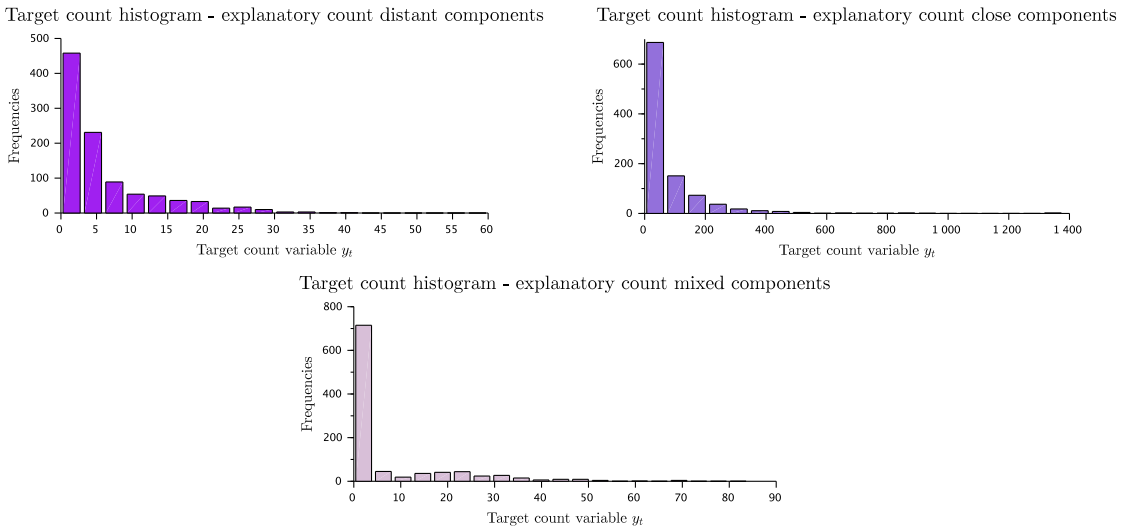

Figure 2: Histograms of target counts for various distances between explanatory data components

correlation between the target and explanatory variables is expected, but this holds for all predictive models. Examples of Spearman's test p-values are 2.58D-130 obtained for y_t and $x_{1;t}$, and 2.945D-78 for y_t and $x_{2;t}$, with Spearman's rank correlation coefficients 0.6073 and 0.4946, respectively. Additionally, the presence of at least slight multimodality in the explanatory counts is required to facilitate mixture initialization and detect locations.

Application of the proposed approach to real count data is presented in the next section.

4 Application to traffic counts in selected areas in Prague

Here, the presented approach was implemented using the Python programming language (www.python.org).

4.1 Data description and preliminary correlation analysis

The data sets were provided by the mobile measuring laboratory MobiLab from the Czech Technical University in Prague (https://mobilab.fd.cvut.cz) and can be downloaded from [81].

Traffic counts were measured every minute over a 24-hour period at the following four selected areas in Prague: Stodůlky (September 2021), Barrandov (June 2022), Radlice (September 2022), and Velká Chuchle (October 2022), resulting in four data sets. At each location, 3 to 6 points were chosen where traffic counts were collected in all possible directions along the corresponding road section/intersection. The measuring points are denoted by red circles in the map screenshots shown in Figure 3.

The four provided data sets were used for the experiments as follows: Within each data set from a specific area, 2 to 3 data samples were created, each utilizing one of the available traffic count variables as the target count variable y_t, indicating that the traffic counts in the chosen direction will be predicted, with the remaining traffic counts used as the explanatory variables $x_t = [x_{1;t}, x_{2;t}, \ldots, x_{N;t}]'$. Each resulting data sample has a different target count variable. In total, ten data samples have been prepared as follows: two for Stodůlky, with the target count variables at points S1 and S2 as shown in Figure 3; three for Barrandov (B1–B3), two for Radlice (RD1, RD2), and three for Velká Chuchle (VC1–VC3).

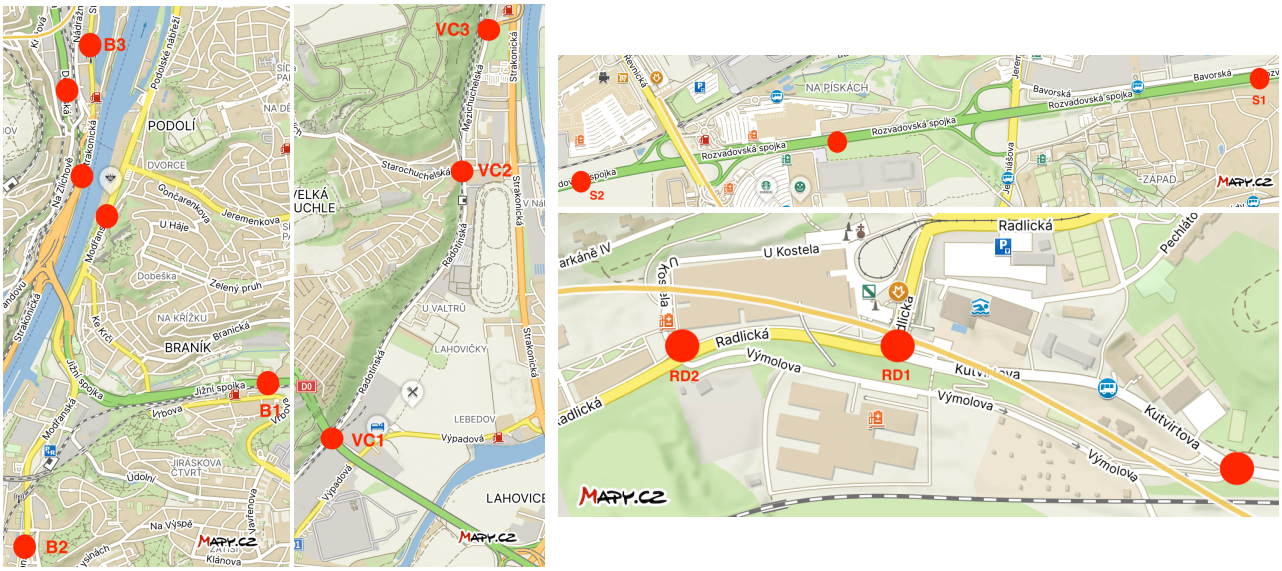

Figure 3: Map screenshots of four locations in Prague (www.mapy.cz ©Seznam.cz, a.s. 2024, ©AOPK ČR – ochrana přírody a krajiny, ©Přispěvatelé OpenStreetMap, ©Natural Earth, Tom Patterson)

The correlation between the target and individual explanatory counts within each data sample was tested using the Spearman's rank correlation test in the SciPy library [82]. All obtained p-values were lower than the significance level of 0.05, indicating a statistically significant dependence between the variables. The Spearman's rank correlation coefficients r_s between the target and individual explanatory counts are provided in Table 3, which also indicates that the number of explanatory counts N differed among the data samples chosen according to specific road sections.

Table 3: Spearman's rank correlation coefficients

Data sample	$r_s(y_t, x_{1;t})$	$r_s(y_t, x_{2;t})$	$r_s(y_t, x_{3;t})$	$r_s(y_t, x_{4;t})$
Stodůlky S1	0.856	0.949	–	–
Stodůlky S2	0.846	0.861	–	–
Barrandov B1	0.671	0.846	0.779	0.831
Barrandov B2	0.699	0.78	0.765	0.758
Barrandov B3	0.491	0.767	0.774	0.805
Radlice RD1	0.954	0.822	–	–
Radlice RD2	0.83	0.977	–	–
Velká Chuchle VC1	0.384	0.382	–	–
Velká Chuchle VC2	0.381	0.377	–	–
Velká Chuchle VC3	0.622	0.612	–	–

4.2 Poisson mixture initialization for explanatory traffic counts

Each of the ten data samples was split into 75% training and 25% testing data. The histograms of the explanatory training traffic counts were used to initialize the Poisson components of each mixture. To conserve space, Fig-

ure 4 displays histograms of the traffic counts for one data sample from each of the Prague areas used (S1, B1, RD1, and VC1). Lighter colors correspond to explanatory traffic counts, while brighter colors represent the target counts. Notice that only explanatory counts are used for initialization, while the histograms of the target data do not necessarily have to duplicate the form of the histograms of the explanatory counts. The number of Poisson

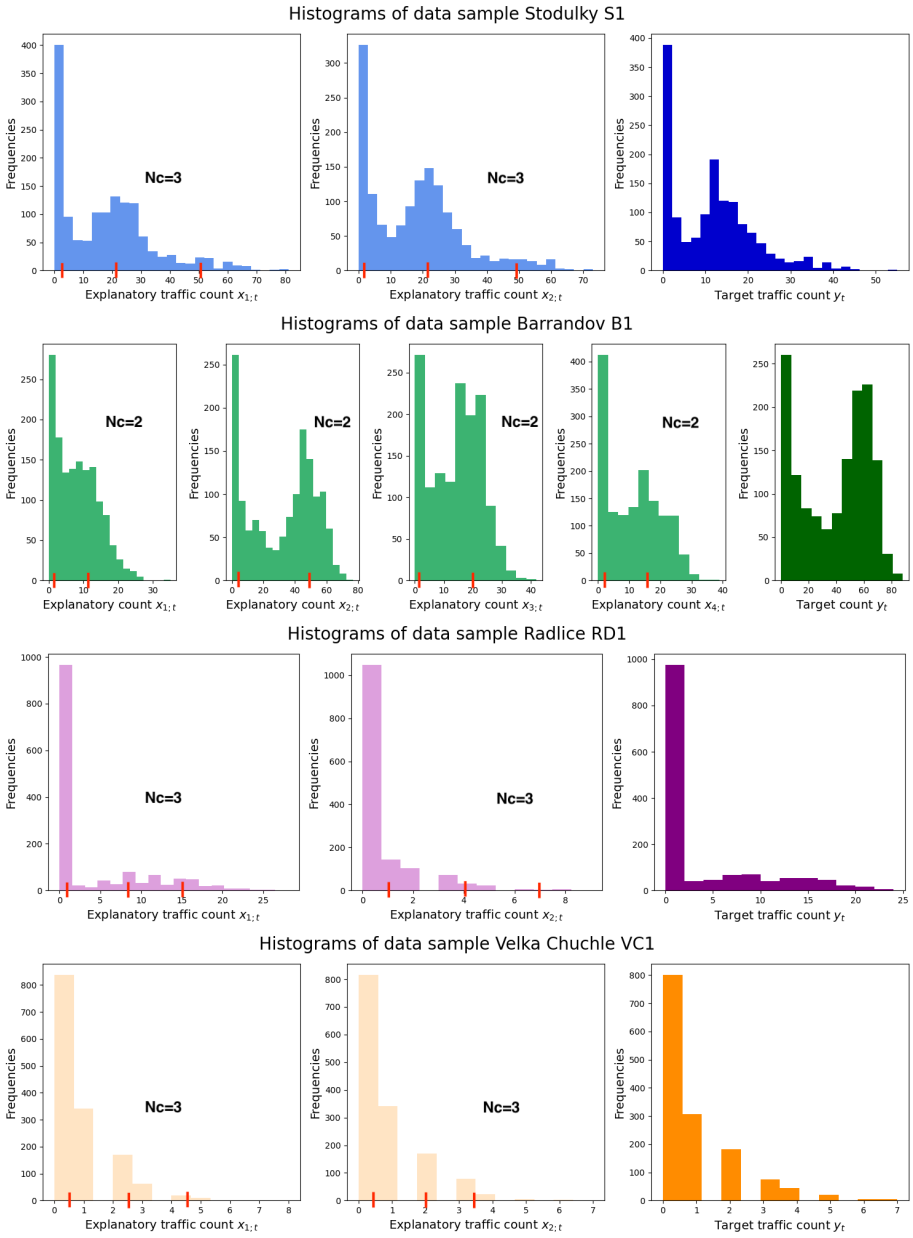

Figure 4: Histograms for one data sample from each of the Prague areas used for initialization

components N_c in the explanatory count mixtures was heuristically initialized based on the visually recognizable number of peaks in the histograms of the explanatory counts, as illustrated in Figure 4. The prior point estimates of the Poisson component parameters $\hat{\lambda}_{i;j;t}$ for $t = 0$ for individual explanatory traffic counts within each data sample were initialized based on the centers of the histogram peaks. For example, the data sample Stodůlky S1 includes two explanatory traffic counts, $x_{1;t}$ and $x_{2;t}$, each described by three components of the Poisson mixture. The initial

point estimates of their component parameters are $[3,\ 22,\ 50]'$ for $x_{1;t}$ and $[2,\ 25,\ 49]'$ for $x_{2;t}$, as indicated by the red lines in Figure 4. Similarly, for the data sample Barrandov B1, there are four explanatory traffic counts, each with two Poisson components. Their initialization values are $[1,\ 12]'$ for $x_{1;t}$, $[2,\ 50]'$ for $x_{2;t}$, $[1,\ 20]'$ for $x_{3;t}$, and $[2,\ 17]'$ for $x_{4;t}$. The initial statistics of the components were then calculated according to (7).

4.3 Explanatory traffic count locations and local target Poisson regressions

Figure 5 illustrates detected locations in the training explanatory traffic counts for each sample plotted against the values of the target traffic counts. In each plot, coefficients of local Poisson regression for recognized locations are positioned near their centers where it was possible due to overlapping clusters. To enhance clarity, the color themes of the plots align with those in Figure 4: blue denotes Stodůlky, green represents Barrandov, plum signifies Radlice, and orange indicates Velká Chuchle. It can be seen that the data samples from the areas of Stodůlky (blue), Radlice (plum), and Velká Chuchle (orange) each have three local Poisson regressions (3) with two explanatory counts, resulting in three point estimates of the regression coefficients $[\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2]'_i$ in each plot for each location, i.e.,

$$E[y_{\tau_i}|x_{\tau_i}] = e^{\hat{\theta}_{0;i}+\hat{\theta}_{1;i}x_{1;\tau_i}+\hat{\theta}_{2;i}x_{2;\tau_i}},\ \ i = \{1, 2, N_c = 3\}, \tag{17}$$

where the indices τ_i of the traffic counts belonging to the locations are defined in (13). The data samples from Barrandov (green B1 to B3) each have two local Poisson regressions with four explanatory counts, demonstrating five coefficients in each plot for each location, i.e.,

$$E[y_{\tau_i}|x_{\tau_i}] = e^{\hat{\theta}_{0;i}+\hat{\theta}_{1;i}x_{1;\tau_i}+\hat{\theta}_{2;i}x_{2;\tau_i}+\hat{\theta}_{3;i}x_{3;\tau_i}+\hat{\theta}_{4;i}x_{4;\tau_i}},\ i = \{1, N_c = 2\}. \tag{18}$$

Notice that the data samples from the busier roads (blue S1, S2, and green B1 to B3) with a traffic count range of up to 85 vehicles per minute have clearly recognizable locations in the plots, even when they overlap. The data samples from Radlice (plum RD1 and RD2), with a traffic count range of up to 28, have significantly overlapping locations. The data values from samples VC1 to VC3 (orange) should have been jiggled for plotting, as their ranges are up to 5 and the centers of the locations are close to each other. This indicates that searching for Poisson-based locations in the explanatory data space with a low number of realizations is not a trivial task. As discussed in Section 3.4, the clustering quality will impact the prediction error in the online target prediction phase of the algorithm, which utilizes the point estimates obtained in individual locations. The results of the target traffic count prediction are compared among all data samples and alternative well-known techniques in the next section.

4.4 Validation of target traffic count prediction

For the validation of the proposed algorithm (PA), each data sample was randomly shuffled during each execution of the algorithm and used for comparison with the following methods: (i) Poisson (PR) and NB regressions from `statsmodels.api` (www.statsmodels.org) [83]; (ii) decision tree classifier (DT) from `sklearn.tree`; (iii) random forest classifier (RF) from `sklearn.ensemble`; and (iv) multi-layer perceptron (MLP) available from `sklearn.neural_network`, all from www.scikit-learn.org [84].

The quality of the prediction of the traffic count variable y_t within each shuffle of each data sample was evaluated using root mean squared error (RMSE), mean absolute error (MAE), mean squared logarithmic error (MSLE)

$$\text{MSLE} = \frac{1}{K}\sum_{t=T+1}^{K}(\ln(1+y_t) - \ln(1+\hat{y}_t))^2, \tag{19}$$

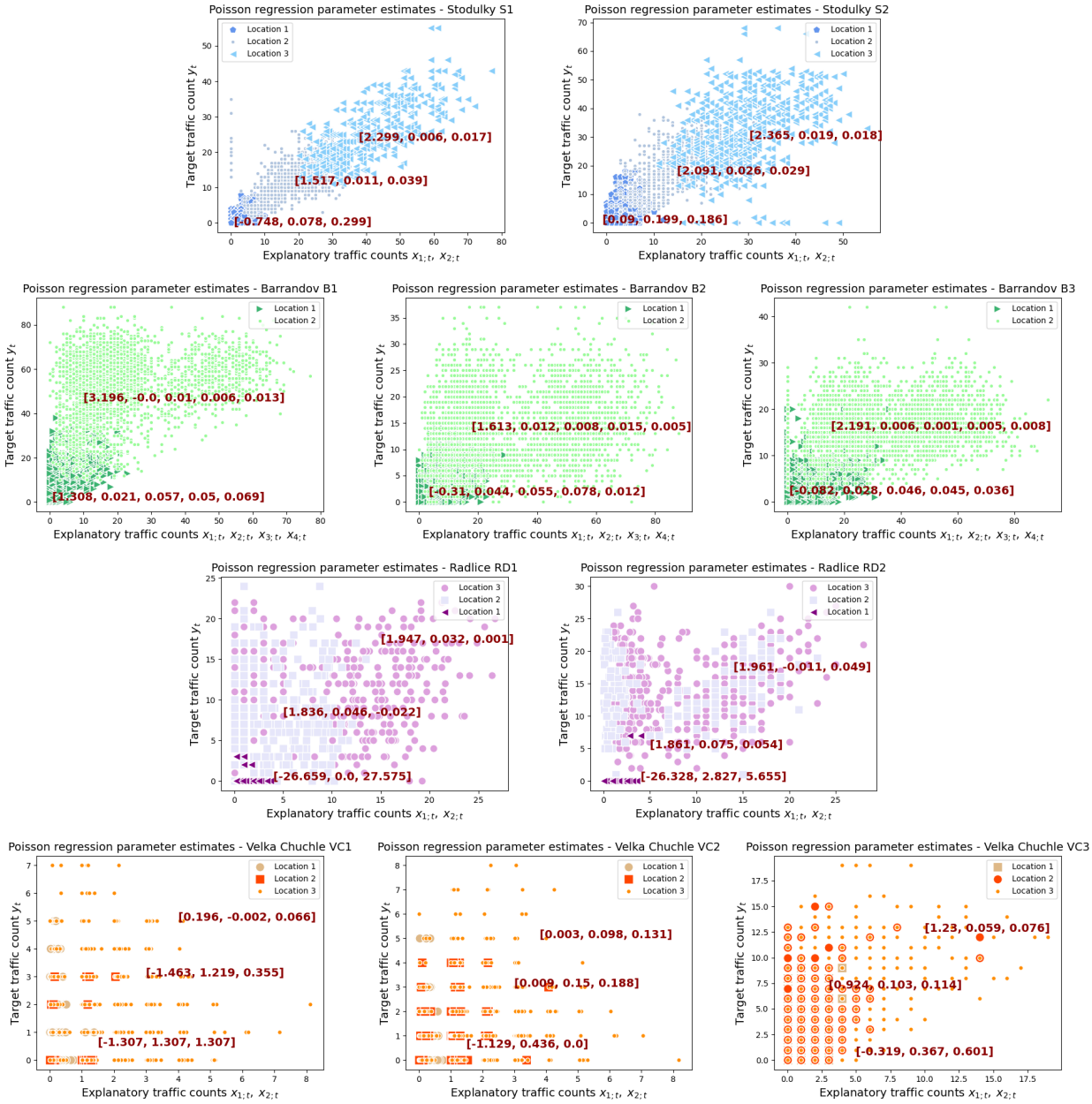

Figure 5: Detected explanatory traffic count locations and local target Poisson regression estimates

the negative log-likelihood (NLL) capturing the discrepancy between the observed target traffic counts and the predicted counts according to the Poisson distribution

$$\mathrm{NLL} = -\sum_{t=T+1}^{K} \ln\left(\frac{\hat{y}_t^{y_t} e^{-\hat{y}_t}}{y_t!}\right), \tag{20}$$

and the coefficient of determination R^2, all from `sklearn.metrics` (www.scikit-learn.org) [84]. For all metrics, $T = 1079$ indicates the capacity of training traffic data, while $K = 1440$ denotes the total capacity of each data sample, i.e., 360 data values were used for testing. Each data sample was shuffled 30 times. Tables 4–7 compare

prediction error metrics of all data samples, where PA denotes the row with results of the proposed algorithm. For each prediction metric, its average value obtained across 30 shuffles of the data sample, along with the standard deviation, is shown. The color themes of the tables (blue, green, plum, and orange) align with those in Figures 4 and 5, corresponding to Prague areas.

4.5 Results in Prague areas with busier traffic flow

Analysis of the obtained results shows that in areas with busier traffic flow (Stodůlky, Barrandov and Radlice in Tables 4 (blue), 5 (green), and 6 (plum)), where the target traffic ranges are between 25 and 80 vehicles per minute (as shown on the y-axes of the blue, green, and plum plots in Figure 5), the R^2 of the proposed algorithm reaches values from 0.615 to 0.891. This signifies an accurate and reliable model, particularly when compared to alternative methods.

Table 4: Target count prediction errors – Stodůlky S1 and S2

	RMSE	MAE	MSLE	NLL	R^2
PA	**3.168±0.177**	**2.179 ±0.096**	**0.09±0.008**	**792.1± 15.92**	**0.891± 0.012**
PR	5.762±0.943	4.006±0.248	0.508±0.046	1136.08±25.85	0.634±0.113
NB	6.669±1.356	4.165±0.334	0.473±0.043	1138.32±33.16	0.505±0.194
DT	4.098±0.234	2.782±0.165	0.129±0.009	978.77±55.33	0.818±0.016
RF	4.061±0.223	2.745±0.144	0.124±0.01	948.9±49.84	0.821±0.016
MLP	3.484±0.257	2.31±0.145	0.093±0.007	914.1±96	0.868±0.014
	RMSE	MAE	MSLE	NLL	R^2
PA	**6.843±0.438**	**4.578±0.229**	**0.316±0.058**	**1151.75±38.29**	**0.771±0.03**
PR	8.531±0.516	6.209±0.262	0.665±0.064	1484.97±41.66	0.643±0.046
NB	8.703±0.55	6.256±0.278	0.652±0.064	1483.46±42.93	0.629±0.051
DT	9.623±0.441	6.388±0.294	0.642±0.074	2020.09±151.86	0.547±0.047
RF	9.199±0.529	6.186±0.31	0.535±0.079	1861.59±147.96	0.586±0.049
MLP	7.553±0.554	4.987±0.311	0.386±0.076	1623.15±226.35	0.72±0.042

The comparison of the average prediction quality, as shown in Tables 4-6, indicates that the RMSE, MAE, MSLE, and NLL of the proposed algorithm are the lowest among the compared methods, with a small standard deviation, while R^2 is the highest. This reflects that the predictions of target traffic counts obtained with the proposed algorithm were the most precise. The prediction quality produced by the alternative methods for these data samples is relatively high, except for NB, which exhibits higher deviations and even a negative R^2 value in Table 6. Nevertheless, it remains lower than that achieved by the proposed algorithm. For the data sample B2 in Table 5, where the R^2 of the proposed algorithm is only 0.615, some of the other methods yield significantly worse results. For example, DT and RF have average R^2 values of 0.223 and 0.321, respectively.

In this part of the validation experiments, the overall success of the prediction methods is influenced by the higher ranges of the traffic counts, where observations on busier roads resemble continuous data, and small deviations between predictions and data are not as significant. The higher correlation shown in Table 3 also impacts the predictions. For this reason, it is worth examining how the proposed algorithm performs with traffic counts characterized by small ranges and lower correlation coefficients shown in Table 3. These results will be presented in the next section.

Table 5: Target count prediction errors – Barrandov B1, B2, and B3

	RMSE	MAE	MSLE	NLL	R^2
PA	**8.676±0.345**	**6.515±0.289**	**0.134±0.021**	**1292.66±24.96**	**0.871±0.011**
PR	12.282±0.549	9.444±0.356	0.312±0.037	1668.82±39.68	0.741±0.025
NB	12.416±0.578	9.488±0.37	0.306±0.036	1668.85±40.07	0.736±0.027
DT	11.089±0.332	8.323±0.273	0.214±0.025	1608.94±53.7	0.789±0.017
RF	10.385±0.354	7.791±0.259	0.195±0.021	1523.64±43.59	0.815±0.015
MLP	10.803±0.506	7.971±0.365	0.185±0.022	1531.66±64.14	0.8±0.021
	RMSE	MAE	MSLE	NLL	R^2
PA	**4.965±0.195**	**3.64±0.14**	**0.248±0.016**	**1044.84±25.05**	**0.615±0.024**
PR	5.399±0.218	4.008±0.151	0.358±0.02	1137.6±29.43	0.544±0.037
NB	5.465±0.221	4.034±0.152	0.35±0.019	1139±29.65	0.532±0.04
DT	7.042±0.368	5.038±0.277	0.455±0.037	1749.9±83.91	0.223±0.084
RF	6.585±0.269	4.702±0.208	0.403±0.028	1694.89±72.77	0.321±0.054
MLP	6.017±0.328	4.319±0.233	0.372±0.027	1813.67±111.72	0.433±0.059
	RMSE	MAE	MSLE	NLL	R^2
PA	**4.247±0.133**	**3.154±0.112**	**0.207±0.017**	**964.89±18.89**	**0.735±0.016**
PR	5.263±0.202	4.012±0.155	0.379±0.033	1128.09±25.46	0.592±0.037
NB	5.374±0.213	4.06±0.163	0.365±0.031	1129.51±26.59	0.575±0.04
DT	5.652±0.179	4.116±0.129	0.332±0.03	1499.34±71.68	0.53±0.035
RF	5.16±0.229	3.747±0.176	0.301±0.025	1427.98±75.72	0.608±0.039
MLP	4.998±0.245	3.591±0.173	0.293±0.033	1509.3±106.08	0.633±0.031

Table 6: Target count prediction errors – Radlice RD1 and RD2

	RMSE	MAE	MSLE	NLL	R^2
PA	**2.903±0.154**	**1.4±0.103**	**0.123±0.025**	**406.22±26**	**0.74±0.022**
PR	4.784±0.615	2.668±0.172	0.536±0.019	871.41±38.69	0.287±0.179
NB	6.08±1.163	2.85±0.27	0.48±0.02	867.42±45.19	-0.173±0.44
DT	4.088±0.232	1.91±0.141	0.277±0.046	668.45±60.1	0.485±0.05
RF	4.045±0.241	1.891±0.153	0.229±0.037	585.99±51.04	0.495±0.062
MLP	3.504±0.242	1.647±0.127	0.162±0.028	497.45±44.16	0.621±0.047
	RMSE	MAE	MSLE	NLL	R^2
PA	**2.495±0.118**	**1.151±0.074**	**0.042±0.009**	**350.48±18.41**	**0.882±0.01**
PR	5.598±0.981	2.86±0.21	0.382±0.007	791.61±29.35	0.39±0.222
NB	6.908±1.519	3.06±0.322	0.331±0.007	791.71±34.73	0.056±0.43
DT	3.453±0.298	1.575±0.153	0.089±0.023	471.36±47.97	0.774±0.033
RF	3.266±0.283	1.49±0.138	0.069±0.012	426.37±33.6	0.798±0.031
MLP	2.826±0.148	1.296±0.091	0.053±0.008	383.38±23.97	0.849±0.017

4.6 Results in Prague areas with less busy traffic flow

In the areas with less busy traffic flow (Velká Chuchle), the target traffic ranges are 7 and 8 vehicles per minute for VC1 and VC2, respectively, and 18 for VC3. These ranges can be seen on the y-axes of the orange plots in Figure 5. Results obtained in these areas, as shown in Table 7 (orange), are not as successful when compared across all methods.

Analysis of the R^2 for data samples VC1 and VC2 shows that the proposed algorithm has the highest R^2 among the compared methods, although it also indicates a weak goodness of fit for the model, with average values of only 0.151 and 0.173. However, predictions from the remaining methods, except for PR, were poor, with negative R^2 values. This can be explained by the low correlation in the data.

Upon examining the prediction errors for data samples VC1 and VC2, it becomes apparent that the RMSE and MSLE of the proposed algorithm are the lowest, but its MAE is higher than those of DT, RF, and MLP. Attributed to the narrow range of counts in these data samples, this is explained by the sensitivity of MAE to small changes in data, highlighting its ability to detect even subtle variations in predictions. However, to assess the prediction quality from various perspectives, the NLL of the proposed algorithm is significantly lower than that of the other methods, indicating better predictive performance of the model.

When considering the results for the data sample VC3, which has a slightly wider target traffic count range of 18, they differ from those for VC1 and VC2. All the metrics in Table 7 (orange) for VC3 show the better performance of the proposed algorithm compared to other methods. There are no negative R^2 values in the average results; however, NB still exhibits very low R^2 values close to zero, with high standard deviation. The proposed algorithm achieves an R^2 value of 0.472, with MLP being the only method close to it. This difference in prediction quality for VC3 is also attributed to the higher correlation coefficients compared to VC1 and VC2 (refer to Table 3).

Table 7: Target count prediction errors – Velká Chuchle VC1, VC2, and VC3

	RMSE	MAE	MSLE	NLL	R^2
PA	**1.166±0.063**	**0.83±0.032**	**0.268±0.015**	**445.29±22.58**	**0.151±0.032**
PR	1.23±0.067	0.903±0.033	0.291±0.013	470.92±22.95	0.054±0.044
NB	1.326±0.091	0.93±0.042	0.3±0.019	476.94±25.06	-0.104±0.145
DT	1.413±0.088	0.794±0.059	0.409±0.035	2091.02±190.12	-0.246±0.055
RF	1.408±0.09	0.794±0.061	0.404±0.036	2030.3±173.75	-0.238±0.064
MLP	1.373±0.09	0.762±0.061	0.377±0.035	1913.24±187.54	-0.177±0.051
	RMSE	MAE	MSLE	NLL	R^2
PA	**1.198±0.053**	**0.869±0.028**	**0.272±0.015**	**456.8±18.47**	**0.173±0.029**
PR	1.268±0.06	0.937±0.027	0.296±0.011	479.51±15.16	0.073±0.059
NB	1.369±0.107	0.959±0.035	0.303±0.016	484.07±18.18	-0.087±0.163
DT	1.488±0.073	0.873±0.048	0.455±0.036	2452.05±184.01	-0.277±0.071
RF	1.47±0.069	0.861±0.045	0.441±0.035	2353.73±178.5	-0.246±0.075
MLP	1.48±0.072	0.856±0.049	0.445±0.038	2419.27±218.4	-0.264±0.084
	RMSE	MAE	MSLE	NLL	R^2
PA	**2.468±0.111**	**1.798±0.065**	**0.399±0.028**	**752.64±22.24**	**0.472±0.036**
PR	2.894±0.217	2.13±0.079	0.544±0.032	850.01±24.96	0.271±0.107
NB	3.342±0.427	2.199±0.109	0.522±0.029	855.17±28.31	0.016±0.257
DT	3.116±0.171	2.002±0.128	0.614±0.058	2076.31±166.15	0.16±0.066
RF	3.024±0.142	1.945±0.105	0.57±0.046	1985.84±147.84	0.207±0.08
MLP	2.773±0.162	1.809±0.116	0.521±0.048	1923.28±144.66	0.334 ±0.062

4.7 Discussion

In the experimental part of the work, differently-aimed metrics were specially chosen to evaluate the prediction quality of the proposed algorithm from different perspectives. These include: (i) assessing differences between predicted and actual values via RMSE, MAE, and MSLE; (ii) measuring the proportion of variance in the target

variable explained by the explanatory variables, representing the overall goodness of fit of the model by R^2; and (iii) evaluating the model's ability to predict probabilities that align with the actual data distribution, providing insight into the model's predictive performance through NLL. To summarize the results of the experiments, it can be stated that traffic count data prediction with the proposed algorithm was more successful across all applied metrics compared to alternative methods, even in the case of weakly correlated data, with the exception of insignificant differences in MAE for two data samples used. The obtained results report that for traffic counts with higher ranges, corresponding to busier urban roads, the proposed approach provided the most accurate model among the methods used, yielding the lowest RMSE, MAE, MSLE, and NLL, as well as the highest R^2.

Concerning traffic counts with lower ranges, which in the used data samples indicated less busy urban roads, it is noted that despite the overall weak goodness of fit observed across all methods, the proposed algorithm performed more accurately even in this scenario. However, the application of all compared methods under these conditions may be limited, as low ranges of traffic data may also be caused by congestion. Additionally, it is important to consider the impact of correlation within the data sample. This underscores that, in addition to the limitations outlined in Section 3.4, low ranges of count data, particularly when accompanied by weak correlations, can further constrain the algorithm's performance.

Based on the above discussion, it can be concluded that the main aim of the paper, as stated in Section 1.3, to predict the target count variable by analyzing cross-sectional count data in identified locations and capturing data evolution over time, has been successfully achieved. The proposed approach contributes to understanding the spatio-temporal relationships between traffic counts across different locations on urban roads. This is believed to pave the way for its potential application in intelligent transportation systems, particularly in the areas of urban planning, traffic management, and infrastructure development. Such application offers enhanced decision-making capabilities and resource allocation strategies, ultimately aiming to optimize transportation efficiency.

5 Conclusion

This study aimed to address a task of modeling and prediction of count data with the application to traffic counts at selected urban roads in Prague. The key ideas of the approach included (i) exploring multiple explanatory counts and identifying their locations through recursive Bayesian estimation of Poisson mixtures; (ii) estimating the target count model using local Poisson regressions at specific locations; and (iii) predicting target counts through real-time location detection. The findings from the conducted experiments are promising, indicating that the proposed algorithm demonstrates improvements in prediction quality, outperforming the alternative methods used for comparison.

The main contribution of the paper is the development of a novel approach for online target count prediction by simultaneously analyzing the spatial locations and temporal evolution of multivariate explanatory count data. The approach brings new benefits to the area of count data models through (i) the use of local models for later recognition in prediction, (ii) one-pass recursive estimation for efficient online re-learning and prediction, and (iii) the implementation of parametrized cluster models for enhanced predictive performance.

To address the remaining challenges in models for count data, the proposed methodology can be extended to use dynamic local Poisson regressions at detected locations. This extension is expected to offer significant advantages, such as efficiently updating models in real-time, capturing spatiotemporal interactions, managing computational complexity, and ensuring robustness against noisy data. Addressing these issues is crucial for enhancing prediction accuracy and scalability in real-world applications.

Applied to traffic count data, the promising performance demonstrated by the proposed algorithm offers a hopeful vision for traffic prediction and urban planning, suggesting its potential as a valuable tool for enhancing transportation efficiency by optimizing the timing of city traffic lights to improve traffic flow.

Acknowledgements

European Funding: Under grant ECSEL 101007321, StorAIge is co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Chips Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them. The project is supported by the CHIPS JU and its members (including top-up funding by France, Belgium, Czech Republic, Germany, Italy, Sweden, Switzerland, Turkey).

National Funding: ÚTIA is partner in project StorAIge. ÚTIA has received national funding from the Ministry of Education, Youth and Sports of the Czech Republic (MEYS) under grant agreement MSMT 8A21009.

Declarations

Declarations of interest: none.

Authors' contributions: Evženie Uglickich: Conceptualization, Software, Investigation, Validation, Writing-Original draft preparation, Reviewing and Editing. Ivan Nagy: Methodology, Software, Writing- Reviewing and Editing.

References

[1] Fisher, P., 1941. Negative Binomial Distribution. Annals of Eugenics, 11, pp.182-787.

[2] Consul, P.C., Famoye, F., 1992. Generalized Poisson regression model. Commun Stat Theor Methods, 21, 89–109.

[3] Yadav, B., Jeyaseelan, L., Jeyaseelan, V., Durairaj, J., George, S., Selvaraj, K.G., Bangdiwala, S. I., 2021. Can Generalized Poisson model replace any other count data models? An evaluation, Clinical Epidemiology and Global Health, 11, 100774.

[4] Congdon, P., 2005. Bayesian Models for Categorical Data. John Wiley & Sons.

[5] Johnson, N.L., Kotz, S., Balakrishnan, N., 1997. Discrete Multivariate Distributions. Wiley Series in probability and Statistics. John Wiley &Sons, inc. New York.

[6] Inouye, D.I., Yang, E., Allen, G.I. and Ravikumar, P., 2017. A review of multivariate distributions for count data derived from the Poisson distribution. Wiley Interdisciplinary Reviews: Computational Statistics, 9(3), p.e1398.

[7] Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological), 36(2), pp.192-225.

[8] Yang, E., Ravikumar, P.K., Allen, G.I. and Liu, Z., 2013. On Poisson graphical models. Advances in neural information processing systems, 26, 1–9.

[9] Allen, G.I. and Liu, Z., 2013. A local Poisson graphical model for inferring networks from sequencing data. IEEE transactions on nanobioscience, 12(3), pp.189-198.

[10] Hadiji, F., Molina, A., Natarajan, S. and Kersting, K., 2015. Poisson dependency networks: Gradient boosted models for multivariate count data. Machine Learning, 100, pp.477-507.

[11] Han, S.W. and Zhong, H., 2016. Estimation of sparse directed acyclic graphs for multivariate counts data. Biometrics, 72(3), pp.791-803.

[12] Heeringa, S.G., West, B.T., Berglung, P.A., 2010. Applied Survey Data Analysis. Chapman & Hall/CRC.

[13] Falissard, B., 2012. Analysis of Questionnaire Data with R. Chapman & Hall/CRC, Boca Raton.

[14] Armstrong, B.G., Gasparrini, A., Tobias, A., 2014. Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis. BMC Medical Research Methodology, 14, 122.

[15] Long, J. S., Freese, J., 2014. Regression Models for Categorical Dependent Variables Using Stata. 3rd edn. Stata Press.

[16] Agresti, A., 2018. An Introduction to Categorical Data Analysis. 3rd Ed. Wiley, 2018.

[17] Diallo, A. O., Diop, A., Dupuy, J.-F., 2018. Analysis of multinomial counts with joint zero-inflation, with an application to health economics, Journal of Statistical Planning and Inference, 194, 85–105.

[18] Ver Hoef, J.M. and Boveng, P.L., 2007. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data?. Ecology, 88(11), 2766–2772.

[19] Berk, R. and MacDonald, J.M., 2008. Overdispersion and Poisson regression. Journal of Quantitative Criminology, 24, 269–284.

[20] Hilbe, J. M., 2011. Negative binomial regression. Cambridge University Press.

[21] Wang, K., Yau, K.K., Lee, A.H. and McLachlan, G.J., 2007. Two-component Poisson mixture regression modelling of count data with bivariate random effects. Mathematical and Computer Modelling, 46(11-12), 1468–1476.

[22] Cui, Y. and Yang, W., 2009. Zero-inflated generalized Poisson regression mixture model for mapping quantitative trait loci underlying count trait with many zeros. Journal of theoretical biology, 256(2), 276–285.

[23] Lim, H.K., Li, W.K. and Philip, L.H., 2014. Zero-inflated Poisson regression mixture model. Computational Statistics & Data Analysis, 71, 151–158.

[24] Papastamoulis, P., Martin-Magniette, M.L. and Maugis-Rabusseau, C., 2016. On the estimation of mixtures of Poisson regression models with large number of components. Computational Statistics & Data Analysis, 93, 97–106.

[25] Bao, J., Durango-Cohen, E.J., Levontin, L. and Durango-Cohen, P.L., 2022. Analysis of factors influencing recurring donations in a university setting: A compound poisson mixture regression model. Journal of Business Research, 151, 489–503.

[26] Abonazel, M.R., Alzahrani, A.R.R., Saber, A.A., Dawoud, I., Tageldin, E. and Azazy, A.R., 2023. Developing ridge estimators for the extended Poisson-Tweedie regression model: Method, simulation, and application. Scientific African, 23, e02006.

[27] Olivares, K.G., Meetei, O.N., Ma, R., Reddy, R., Cao, M. and Dicker, L., 2024. Probabilistic hierarchical forecasting with deep poisson mixtures. International Journal of Forecasting, 40(2), 2024, 470-489.

[28] Zeeshan, M., Khan, A., Amanullah, M., Bakr, M.E., Alshangiti, A.M., Balogun, O.S. and Yusuf, M., 2024. A new modified biased estimator for Zero inflated Poisson regression model. Heliyon. 10, e24225.

[29] Park, B.J., Lord, D. and Hart, J.D., 2010. Bias properties of Bayesian statistics in finite mixture of negative binomial regression models in crash data analysis. Accident Analysis & Prevention, 42(2), pp.741–749.

[30] Zou, Y., Zhang, Y. and Lord, D., 2014. Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models. Analytic methods in accident research, 1, 39–52.

[31] Tzougas, G. and di Cerchiara, A.P., 2021. The multivariate mixed negative binomial regression model with an application to insurance a posteriori ratemaking. Insurance: Mathematics and Economics, 101, 602–625.

[32] Hajihosseini, M., Amini, P., Saidi-Mehrabad, A. and Dinu, I., 2023. Infants' gut microbiome data: A Bayesian Marginal Zero-inflated Negative Binomial regression model for multivariate analyses of count data. Computational and Structural Biotechnology Journal, 21, 1621–1629.

[33] Pho, K.H. and Lukusa, T.M., 2024. Parameter Estimations on Zero-Inflated Negative Binomial Model with Incomplete Data. Applied Mathematical Modelling, 129, 207–231.

[34] Mothafer, G.I., Yamamoto, T. and Shankar, V.N., 2016. Evaluating crash type covariances and roadway geometric marginal effects using the multivariate Poisson gamma mixture model. Analytic methods in accident research, 9, 16-26.

[35] Yu, J., Gwak, J. and Jeon, M., 2016, October. Gaussian-Poisson mixture model for anomaly detection of crowd behaviour. In 2016 International Conference on Control, Automation and Information Sciences (ICCAIS), pp. 106-111. IEEE.

[36] Silva, A., Rothstein, S.J., McNicholas, P.D. and Subedi, S., 2019. A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data. BMC bioinformatics, 20(1), pp.1-11.

[37] Kinzer, J. P., 1934. Application of the Theory of Probability to Problems of Highway Traffic, thesis submitted in partial satisfaction of requirements for degree of B.C.E., Polytechnic Institute of Brooklyn. Abstracted in: Proceedings, Institute of Traffic Engineers, 5, pp. 118–124.

[38] Adams, W. F., 1936. Road traffic considered as a random series, Institution of Civil Engineers, 4, 121–130.

[39] Greenshields, B. D., Shapiro, D., Ericksen, E. L., 1947. Traffic Performance at Urban Street Intersections, Technical Report No. 1, Yale Bureau of Highway Traffic.

[40] Gerlough, D. L. and Schuhl, A., 1955. Use of Poisson Distribution in Highway Traffic. The Probability Theory Applied to Distribution of Vehicles on Two-Lane Highways. Eno Foundation for Highway Traffic Control. US Transportation Collection. https://rosap.ntl.bts.gov/view/dot/16299

[41] Daraghmi, Y.A., Yi, C.W. and Chiang, T.C., 2012, November. Space-time multivariate negative binomial regression for urban short-term traffic volume prediction. In 2012 12th International Conference on ITS Telecommunications, pp. 35-40. IEEE.

[42] Okawa, M., Kim, H. and Toda, H., 2017, May. Online traffic flow prediction using convolved bilinear Poisson regression. In 2017 18th IEEE International Conference on Mobile Data Management (MDM), pp.134–143. IEEE.

[43] Velikajne, N. and Moškon, M., 2022. RhythmCount: A Python package to analyse the rhythmicity in count data. Journal of Computational Science, 63, 101758.

[44] Uglickich, E. and Nagy, I., 2023. Using Poisson proximity-based weights for traffic flow state prediction. Neural Network World, 4, 291–315.

[45] Lord, D., Washington, S.P. 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.

[46] Quddus, M.A., 2008. Time series count data models: an empirical application to traffic accidents. Accident analysis & prevention, 40(5), 1732–1741.

[47] Liu, C., Zhao, M., Li, W. and Sharma, A., 2018. Multivariate random parameters zero-inflated negative binomial regression for analyzing urban midblock crashes. Analytic methods in accident research, 17, 32–46.

[48] Lukusa, M.T. and Phoa, F.K.H., 2020. A Horvitz-type estimation on incomplete traffic accident data analyzed via a zero-inflated Poisson model. Accident Analysis & Prevention, 134, 105235.

[49] Simmachan, T., Wongsai, N., Wongsai, S., Lerdsuwansri, R., 2022. Modeling road accident fatalities with underdispersion and zero-inflated counts. PLoS ONE 17(11): e0269022. doi: 10.1371/journal.pone.0269022

[50] Miah, M.M., Hyun, K.K. and Mattingly, S.P., 2024. A review of bike volume prediction studies. Transportation Letters, 1-28.

[51] Zaouche, M. and Bode, N.W., 2023. Bayesian spatio-temporal models for mapping urban pedestrian traffic. Journal of transport geography, 111, 103647.

[52] Strongylis, L., Petraki, V. and Yannis, G., 2023. Critical impact factors of pedestrians traffic combining multiple data sources in Athens. Transportation research procedia, 72, 3332–3339.

[53] Wang, X. and Kockelman, K. M., 2009. Forecasting Network Data: Spatial Interpolation of Traffic Counts from Texas Data. Transportation Research Record, 2105(1), 100–108. doi: 10.3141/2105-13

[54] Gastaldi, M., Gecchele, G. and Rossi, R., 2014. Estimation of Annual Average Daily Traffic from one-week traffic counts. A combined ANN-Fuzzy approach. Transportation Research Part C: Emerging Technologies, 47, 86–99.

[55] Bagheri, E., Zhong, M. and Christie, J., 2015. Improving AADT estimation accuracy of short-term traffic counts using pattern matching and Bayesian statistics. Journal of Transportation Engineering, 141(6), A4014001.

[56] Vlahogianni, E.I., Karlaftis, M.G. and Golias, J.C., 2014. Short-term traffic forecasting: Where we are and where we're going. Transportation Research Part C: Emerging Technologies, 43, 3–19.

[57] Sekuła, P., Marković, N., Vander Laan, Z. and Sadabadi, K.F., 2018. Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A Maryland case study. Transportation Research Part C: Emerging Technologies, 97, 147–158.

[58] Shahriari, S., Ghasri, M., Sisson, S.A. and Rashidi, T., 2020. Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction. Transportmetrica A: Transport Science, 16(3), 1552–1573.

[59] Yao, R., Zhang, W. and Zhang, L., 2020. Hybrid methods for short-term traffic flow prediction based on ARIMA-GARCH model and wavelet neural network. Journal of Transportation Engineering, Part A: Systems, 146(8), 04020086.

[60] Liu, B., Tang, X., Cheng, J. and Shi, P., 2020. Traffic flow combination forecasting method based on improved LSTM and ARIMA. International Journal of Embedded Systems, 12(1), 22–30.

[61] Lin, X. and Huang, Y., 2021. Short-term high-speed traffic flow prediction based on ARIMA-GARCH-M model. Wireless Personal Communications, 117(4), 3421–3430.

[62] Zhao, L., Wen, X., Wang, Y. and Shao, Y., 2022. A novel hybrid model of ARIMA-MCC and CKDE-GARCH for urban short-term traffic flow prediction. IET Intelligent Transport Systems, 16(2), 206–217.

[63] Ganji, A., Shekarrizfard, M., Harpalani, A., Coleman, J. and Hatzopoulou, M., 2020. Methodology for spatio-temporal predictions of traffic counts across an urban road network and generation of an on-road greenhouse gas emission inventory. Computer-Aided Civil and Infrastructure Engineering, 35(10), 1063–1084.

[64] Ganji, A., Zhang, M. and Hatzopoulou, M., 2022. Traffic volume prediction using aerial imagery and sparse data from road counts. Transportation research part C: emerging technologies, 141, 103739.

[65] Xu, P., Li, X., Wu, Y.J. and Noh, H., 2023. Network-level turning movement counts estimation using traffic controller event-based data. Journal of Intelligent Transportation Systems, 27(5), 677–691.

[66] Goenaga, B., Underwood, B.S., Castorena, C., Cantillo, V. and Arellana, J., 2023. Using continuous traffic counts extracted from smartphone data to evaluate traffic reductions during COVID-19 pandemic in North Carolina. Latin American transport studies, 1, 100005.

[67] Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.C., 2005. A possibilistic fuzzy c-means clustering algorithm. IEEE Transactions on Fuzzy Systems, 13(4), 517–530.

[68] Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), 651–666.

[69] Uglickich, E., Nagy, I., Reznychenko, T., 2023. Count predictive model with mixed categorical and count explanatory variables. In 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), p.51–56.

[70] Kárný, M., Kadlec, J., Sutanto, E.L., 1998. Quasi-Bayes estimation applied to normal mixture. In: Rojíček, J., Valečková, M., Kárný, M., Warwick, K. (eds.). Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing, Prague, CZ, pp. 77–82.

[71] Kárný, M., Böhm, J., Guy, T.V., Jirsa, L., Nagy, I., Nedoma, P., Tesař, L., 2006. Optimized Bayesian Dynamic Advising: Theory and Algorithms. Springer, London.

[72] Nagy, I., Suzdaleva, E., Kárný, M., Mlynářová, T., 2011. Bayesian estimation of dynamic finite mixtures. International Journal of Adaptive Control and Signal Processing. 25(9), 765–787.

[73] Nagy, I. and Suzdaleva, E., 2017. Algorithms and Programs of Dynamic Mixture Estimation. Unified Approach to Different Types of Components, SpringerBriefs in Statistics. Springer International Publishing, Heidelberg.

[74] Peterka, V., 1981. Bayesian system identification. In Eykhoff, P. (Ed.), Trends and Progress in System Identification. Oxford, Pergamon Press, pp. 239-304.

[75] Nagy, I., Suzdaleva, E., Pecherková, P., 2016. Comparison of various definitions of proximity in mixture estimation. In Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics (ICINCO), 1, pp. 527–534.

[76] Suzdaleva, E., Nagy, I., Petrouš, M., 2017. Recursive clustering hematological data using mixture of exponential components. In Proceedings of International Conference on Intelligent Informatics and BioMedical Sciences ICIIBMS, pp. 63–70. doi: 10.1109/ICIIBMS.2017.8279700

[77] Nagy, I., Suzdaleva, E., Petrouš, M., 2017. Clustering with a model of sub-mixtures of different distributions. In Proceedings of IEEE 15th International Symposium on Intelligent Systems and Informatics SISY, pp. 315-320. doi: 10.1109/SISY.2017.8080574

[78] Suzdaleva, E., Nagy, I., Pecherková, P., Likhonina, R., 2017. Initialization of recursive mixture-based clustering with uniform components. In Proceedings of the 14th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2017), pp. 449–458.

[79] Jozová, Š., Uglickich, E., Nagy, I., Likhonina, R., 2022. Modeling of discrete questionnaire data with dimension reduction. Neural Network World 32(1), 15–41.

[80] Uglickich, E. and Nagy, I., 2022. Recursive mixture estimation with univariate multimodal Poisson variable. Research Report 2394. ÚTIA AV ČR, v.v.i., Prague. http://library.utia.cas.cz/separaty/2022/ZS/uglickich-0557467.pdf

[81] Uglickich, E., Nagy, I., Kumpošt, P., Richter, P., 2024. Datasets: Traffic Counts in Prague, Mendeley Data, V1, doi: 10.17632/hck3d5cj6p.1. Accessed May 2024.

[82] SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., Mulbregt, P. van, & SciPy 1.0 Contributors. 2020. Nature Methods, 17(3), 261-272. doi: 10.1038/s41592-019-0686-2

[83] Skipper, S. a Perktold, J., 2010. Statsmodels: econometric and statistical modeling with Python. In Proceedings of the 9th Python in Science Conference, pp. 92–96.

[84] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. Scikit-learn: Machine Learning in Python, 2011. Journal of Machine Learning Research, 12, 2825–2830.