Estimating city-wide hourly bicycle flow using a hybrid LSTM MDN

Myhrmann, Marcus Skyum; Mabit, Stefan Eriksen

Published in:
Transportation Research Part A: Policy and Practice

Link to article, DOI:
10.1016/j.tra.2023.103783

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Transportation Research Part A 176 (2023) 103783

Contents lists available at ScienceDirect

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

Estimating city-wide hourly bicycle flow using a hybrid LSTM MDN

Marcus Skyum Myhrmann, Stefan Eriksen Mabit *

Transport Division, DTU Management, Technical University of Denmark, Bygningstorvet Building 116B, Kgs. Lyngby, 2800, Denmark

ARTICLE INFO

Keywords:
Bicycle flow estimation
Long Short-Term Memory
Mixture Density Network
Deep Learning
Aggregation bias

ABSTRACT

Cycling can reduce greenhouse gas emissions and air pollution and increase public health. Hence, policymakers in cities worldwide seek to improve bicycle mode shares. Efforts to increase the bicycle's mode share involve many measures, one of them being the improvement of cycling safety often requiring an analysis of the factors surrounding accidents. However, meaningful analysis of cycling safety requires accurate bicycle flow data that are generally sparse or only available at the aggregate level. Therefore, safety engineers often rely on aggregated variables or calibration factors that fail to account for variations in the cycling traffic relevant to policymaking.

This paper illustrates how machine learning can support policy analysis by delivering detailed bicycle flow predictions. The illustration applies a Deep Learning approach, the Long Short-Term Memory Mixture Density Network (LSTMMDN), to estimate hourly bicycle flow in Copenhagen, conditional on weather, temporal and road conditions at the segment level. The method addresses some shortcomings in the calibration factor method resulting in 66–77% more accurate bicycle traffic estimates.

To quantify the impact of more accurate bicycle traffic estimates in cycling safety analysis, we test the effect of different flow estimates in a bicycle crash risk model, i.e. the models are identical except for the exposure variables. One model is estimated using the LSTMMDN estimates, one using the calibration-based estimates, and one using yearly mean traffic estimates. The results show that investing in more advanced methods for obtaining bicycle volume estimates can improve the quality of safety analyses and other performance measures.

1. Introduction

In a world where sustainable transport is increasingly essential, policy-makers in cities seek to increase the mode share of the bicycle. Not only is cycling emission free, but it also improves the health and wellbeing of its users (Mueller et al., 2015) and leads to improved livability of cities. However, a frequently reported barrier to increasing the mode share of cyclists is the fear of traffic-related injury (Horton, 2016; Transport for London, 2014; Vejdirektoratet, 2018).

To tackle the above problems, safety engineers, transport agencies, and researchers have investigated various aspects of bicycle accidents to identify the factors associated with bicycle crash occurrence (Boele-Vos et al., 2017; Janstrup et al., 2019; Aldred et al., 2018; Vandenbulcke et al., 2014; Dozza, 2017; Rossetti et al., 2018; Twisk and Reurings, 2013; Morrison et al., 2019; Kaplan and Prato, 2015; Ji et al., 2021; Saha et al., 2018; Raihan et al., 2019), and injury outcome (Myhrmann et al., 2021; Kaplan et al., 2014; Fountas et al., 2021; Kim et al., 2007; Behnood et al., 2014; Thomas and De Robertis, 2013; Chen et al., 2017; Samerei et al., 2021) to make informed mitigating efforts. However, as highlighted by Dozza (2017) and Thomas and De Robertis (2013), many such investigations do not account for cyclist exposure. This is primarily due to a lack of exposure data, as bicycle monitoring is often

* Corresponding author.
E-mail addresses: mskyum@dtu.dk (M.S. Myhrmann), smab@dtu.dk (S.E. Mabit).

https://doi.org/10.1016/j.tra.2023.103783

Available online 16 August 2023

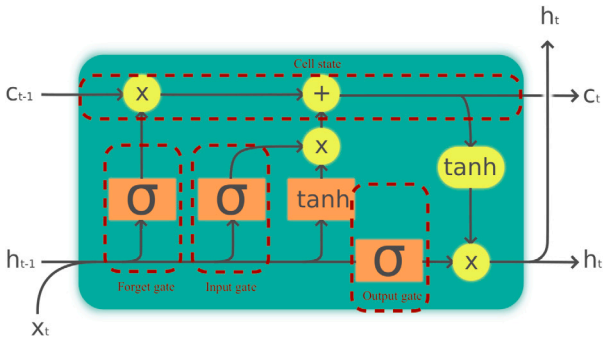

Fig. 1. Graphical illustration of LSTM cell structure[1]: Orange squares mark layers with respective activation functions, yellow circles mark point-wise operation, where x and + icons indicate respective element-wise multiplication and addition of tensors in the LSTM cell.

infrequent, only conducted at a few locations, or even entirely unavailable. The studies that account for cyclist exposure often rely on highly aggregated exposure measures such as the annual average daily cycling traffic (AADCT), annual average weekday cycling traffic (AAWCT), or population and commuter indicators whereas transport agencies use calibration factors applied to AADCT or AAWCT to obtain reasonable hourly volume profiles (Schrank, 2021). However, the hourly cycling volumes derived from calibration factors do not reflect variations in traffic due to weather, temporal effects, and other external factors. This presents a significant issue considering the impact weather and other factors have on cyclist ridership (Böcker et al., 2013; Nankervis, 1999; Nosal and Miranda-Moreno, 2014). Thus, to provide more informed policy and accident analyses, there appears to be a need for more detailed bicycle volume estimates.

Recent advances for the estimation of traffic volumes have primarily focused on the prediction of the short-term traffic state/flow (Lv et al., 2015; Polson and Sokolov, 2017; Du et al., 2021; Chen et al., 2018). Short-term traffic prediction is mainly relevant for segments and networks already subject to good monitoring, where detailed short-term forecasts can be applicable for congestion easing. This is because the short-term traffic predictions are conditioned on the previous traffic flow data. An especially popular model to aid this task is the Long Short-Term Memory (LSTM) neural network (Hochreiter and Schmidhuber, 1997) which has been adopted for many recent models for traffic forecasting (Ma et al., 2015; Duan et al., 2016; Chen et al., 2016; Cui et al., 2020; Zhao et al., 2017).

However, to aid safety engineers in improving bicycle accident- and safety analysis, historic bicycle flow estimates are needed. Therefore our study focuses on improving the estimation of historic bicycle volumes derived from mean daily exposure measures, which is also of interest for recently developed large scale models (Davies, 2017; Kjems and Paag, 2019). The model is intended to estimate historic bicycle flow, where only the mean-expected daily traffic is available.

To accomplish this, we apply a novel neural network approach to estimate hourly bicycle volumes conditional on weather conditions, temporal effects, and road conditions. This framework is a hybrid of an LSTM and a Mixture Density Network (Bishop, 1994), which introduces a Gaussian Mixture Model (GMM) extension to the traditional LSTM. This hybridisation enables the model to estimate a conditional bicycle flow distribution in contrast to the conventional conditional mean estimation. Furthermore, the MDN extension improves the estimates of hourly bicycle volumes by treating them as random draws from a distribution, thus introducing variation across the network even on measurably similar roads. Finally, we quantify the effect of improved bicycle exposure estimates by contrasting crash frequency models using different exposure variables.

2. Methodology

The following section describes the methodological approach employed to estimate the bicycle flow.

2.1. LSTM

LSTMs have exhibited a superior capability of handling nonlinear time series problems (Hochreiter and Schmidhuber, 1997). The LSTM learns to represent temporal data by introducing a memory cell and sub-processes, referred to as gates. There are three such gates in the LSTM cell: the input gate, the forget gate, and the output gate. The gates each handle different tasks, involving what information to keep from the previous cell state, what new input to consider, and which to include in the cell state. In Fig. 1, there is a visual illustration of the LSTM cell layout, and the computations performed in the LSTM cell for each time step in the temporal sequence are shown in Eqs. (1) to (6).

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \tag{1}$$

[1] https://upload.wikimedia.org/wikipedia/commons/3/3b/The_LSTM_cell.png

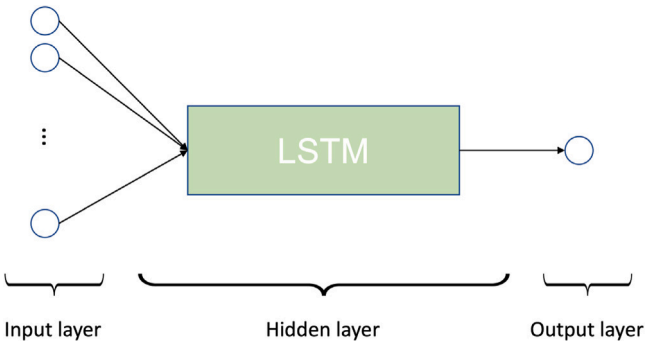

Fig. 2. Example structure of simple LSTM neural network.

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{2}$$

$$\tilde{C}_t = tanh(W_{Cx}x_t + W_{Ch}h_{t-1} + b_C) \tag{3}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{4}$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{5}$$

$$h_t = tanh(C_t) \times o_t \tag{6}$$

Here $\sigma(\cdot)$ and $\tanh(\cdot)$ are the respective activation functions, W_{fx}, W_{fh}, W_{ix}, W_{ih}, W_{ox}, W_{oh}, W_{Cx} and W_{Ch} are the weight matrices of the respective gates f_t, i_t, o_t and the memory cell C_t, in the LSTM cell. b_f, b_i, b_o and b_C are intercept/bias terms of the respective gates, and h_t represents the hidden state at time step t.

2.2. LSTM MDN

An illustration of a simple LSTM regression network is shown in Fig. 2. It has an input layer, a single LSTM cell with k computational nodes in each gate, and a single-node output layer. The optimisation of such a model by minimising the mean squared error (MSE) has been shown to approximate the conditional average of the target data (i.e. bicycle flow) (Bishop, 1994). However, we wish to allow for similar weather and seasonal conditions not necessarily yielding the same bicycle flow across roads. Therefore, we introduce a level of randomness in the bicycle flow estimation. Namely, through the Mixture Density Network (Bishop, 1994).

The original Mixture Density Network proposed by Bishop (1994) is a combination of an artificial neural network (ANN) and a mixture model (McLachlan and Basford, 1988). This combination provides the flexibility to model a general distribution and, as such, enables the estimation of the conditional density function of the target data, in contrast to the conditional average.

In the mixture model, the probability density of the target data is specified as a linear combination of kernel functions,

$$p(y|X) = \sum_{i=1}^{A} \alpha_i(X)\phi_i(y|X) \tag{7}$$

where A is the number of mixture components, $\alpha_i(X)$ are the mixing coefficients dependent on the input data, and $\phi_i(\cdot|X)$ are Gaussian probability density kernels:

$$\phi_i(t|x) \propto \frac{1}{\nu_i(x)^{1/2}} e^{-\frac{(t-\mu_i(x))^2}{2\nu_i(x)^2}} \tag{8}$$

where $\mu_i(x)$ is the centre of the kernel, i.e. the conditional average, and $\nu_i(x)$ the associated variance.

We refer to the combination of the LSTM and the MDN as the LSTMMDN. This model varies from the LSTM regression shown in Fig. 2 only in the output layer. The task of the LSTM in the LSTMMDN is to estimate the mixing coefficients $\alpha_i(X)$, the means $\mu_i(X)$ and the variances $\nu_i(X)$, conditional on input X, subsequently fed into the GMM.

To ensure the properties of the conditional density of the target data $p(y|X)$, the mixing coefficients $\alpha_i(X)$ need to satisfy the constraint in Eq. (9)

$$\sum_{i=1}^{A} \alpha_i(X) = 1 \tag{9}$$

which can be solved by connecting the mixing coefficients to the feed-in network using a softmax/multinomial logit regression (Bishop, 1994), i.e.

$$\alpha_i = \frac{\alpha_i(o_i^{\alpha})}{\sum_j^A o_j^{\alpha}}. \tag{10}$$

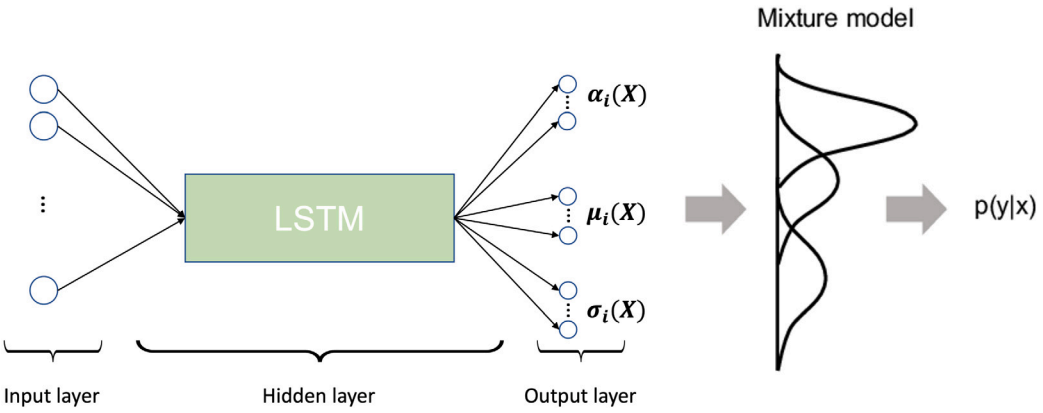

Fig. 3. Graphical illustration of the LSTMMDN setup used in the study.

where o_i^{α} is the network output relevant to the mixing coefficients. It is convenient to avoid the conditional variances tending to zero. Therefore we parameterise the conditional variance in terms of the exponential of the network output, see Eq. (11). This parametrisation also corresponds to choosing an un-informative Bayesian prior in a Bayesian framework, assuming that the output o_i^{v} has a uniform probability distribution (Bishop, 1994).

$$v_i = e^{o_i^{v}} \tag{11}$$

Finally, the centres/conditional means μ_i, shown in Eq. (12), are represented by location parameters that depend directly on the network outputs o_i.

$$\mu_i = o_i^{\mu} \tag{12}$$

To optimise the weights and biases $\Theta = \{W_{fx}, W_{fh}, W_{ix}, W_{ih}, W_{ox}, W_{oh}, W_{Cx}, W_{Ch}, b_f, b_i, b_o, b_C\}$ we use a method called backpropagation. This is an iterative process in which data are "fed forward" until a prediction is made upon which the deviation/error from the actual result is computed using a loss function. Based on the deviation/error of the result and the impact on the output of the individual weights (Θ), the error is then backpropagated using a gradient scheme to update the individual weights Θ. For further explanation of the statistics of backpropagation, see, e.g. Hastie et al. (2009), Bishop (1994). The loss function we use for the parameter optimisation is the log-likelihood as shown in Eq. (13).

$$\mathcal{LL}(\Theta) = \frac{1}{N}\sum_{j}^{N} -\ln\{\sum_{i=1}^{A} \alpha_i(X_j)\phi_i(y_j|X_j)\} \tag{13}$$

2.3. Model configuration

To sum up the model configuration: The LSTMMDN applied in this study is configured as shown in Fig. 3 using the activation functions described in Eqs. (1) to (6). We apply an internal dropout rate of 20% dropout in the LSTM cell to avoid overfitting/over-specification of the model (Srivastava et al., 2014). Meanwhile, we use a linear $a(x) = x$ in the output layer, along with the relevant transformations leading into the GMM, as described in Eqs. (10) to (12). The specific backpropagation scheme used for the parameter optimisation is the Adaptive Moment Estimation (Adam) algorithm (Kingma and Ba, 2015). This is a stochastic gradient-based optimisation approach, which provides some advantages over traditional gradient-based schemes (Hastie et al., 2009). Finally, we employ an early stopping criterion. This criterion is evaluated on a small validation sample of the data and ends the model estimation if no performance improvement is achieved over L model updates to avoid overfitting where we set $L = 100$.

3. Results

3.1. Data and experimental setup

The current study use bicycle flow in Copenhagen, Denmark, as case. The bicycle volume data was obtained for the Mastra service provided by the Danish Road Directorate (Vejdirektoratet, 2020). Here we selected the data recorded from automated counting stations in Copenhagen and Frederiksberg. Such counting stations generally record the amount of cyclists by registering bicycles riding over two tubes with a constant airflow inside. Thus enabling the counting and potential speed assessment of the cycling flow. Thusly, the hourly bicycle volume data recorded by bicycle counting stations was acquired for the period 2017–2020. The selected bicycle counting stations are located at the blue dots in Fig. 4. There was considerable variation in active counting days of the

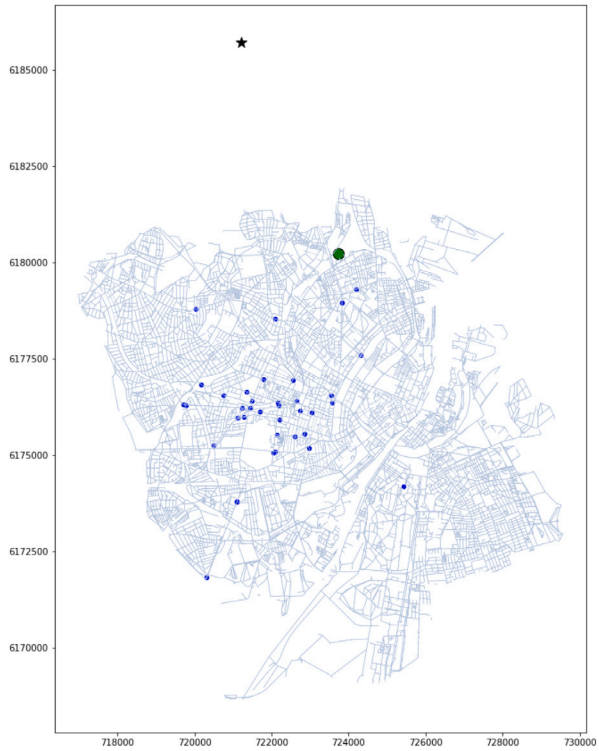

Fig. 4. Map of the measurement stations and weather stations in Copenhagen. Black star shows the location of the measuring station used to fill in missing wind data. The green dot shows the location of the HQ of the Danish Meteorological Institute. Blue dots are cycling counting stations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

counting stations in the four-year period from 2017 to 2020. The number of days counting stations were active ranged from 9 to 731 days. The median number of active counting days is 40. Overall, a total of 64,664 hourly bicycle volumes are recorded.

The reports provided by The Danish Road Directorate, which contain the hourly bicycle counts, also contain the AADCT and AAWCT for each bicycle counting station every year where activity is registered. The AADCT and AAWCT are computed by the Danish Road Directorate.

Using the Open Data API of the Danish Meteorological institute (DMI, 2021), located at the green dot in Fig. 4, we acquired weather data for Copenhagen in the period 2017–2020. The data are reported at 10-minute intervals and contain information on air temperature, pressure, wind speed, wind gusts, wind direction, precipitation levels, visibility and snow volume.

Finally, we also include time-related data such as an hour of the day, day of the week, week of the year, and indicators of public holidays, as this would be assumed to influence cyclist ridership strongly.

The data consist of 64,664 observations of hourly bicycle volumes and $\approx 388,000$ weather and temporal measurements at 10-minute intervals, totalling 17 features (i.e. explanatory variables). These 17 weather and time-related features, including the AADCT of the roads of the bicycle counters, are the predictors used to estimate the hourly bicycle volumes y_T. The observed 10-minute interval predictors $\boldsymbol{x}_t$ are grouped such that a sequence of six 10-minute interval observations $\boldsymbol{x}_t$ are paired with the matching response (i.e. the accumulated bicycle volume of the hour) as shown below.

These data are used to train and evaluate the LSTMMDN. The combined data of all bicycle count stations subsequently split into a training, validation and test data set containing 70%, 10% and 20% of the total data, respectively.

$$y_T(\boldsymbol{X}_T): \quad \boldsymbol{X}_T = \begin{pmatrix} \boldsymbol{x}_{t1} \\ \boldsymbol{x}_{t2} \\ \boldsymbol{x}_{t3} \\ \boldsymbol{x}_{t4} \\ \boldsymbol{x}_{t5} \\ \boldsymbol{x}_{t6} \end{pmatrix} = \begin{pmatrix} (x_{1,t1}, x_{2,t1}, \ldots, x_{D,t1}) \\ (x_{1,t2}, x_{2,t2}, \ldots, x_{D,t2}) \\ (x_{1,t3}, x_{2,t3}, \ldots, x_{D,t3}) \\ (x_{1,t4}, x_{2,t4}, \ldots, x_{D,t4}) \\ (x_{1,t5}, x_{2,t5}, \ldots, x_{D,t5}) \\ (x_{1,t6}, x_{2,t6}, \ldots, x_{D,t6}) \end{pmatrix} \tag{14}$$

Here $t1, t2, t3, t4, t5, t6$ index the six 10-minute intervals of the input sequence for the aggregated hourly cycling volume, T indicates the full hour, and D is the number of predictors.

Table 1
Goodness-of-fit (GOF) measures related to various LSTMMDN structures and the SVFs, determined on the test data.

Model specification	$-\log \mathcal{L}_\mu$	$-\log \hat{\mathcal{L}}$	MSE_μ	$\hat{MSE}$	Trainable parameters
MDN: G(A=6)	−6244	−6587	0.267	0.673	1962
ANN: (Sekuła et al., 2018)	5772	–	0.108	–	162,025
LSTM: $L(k=32)$	5851	–	0.129	–	6,561
LSTM: $L(k=64)$	5792	–	0.114	–	21,313
LSTM: $L(k=32)\times 6$	5832	–	0.128	–	6,733
LSTM: $L(k=64)\times 6$	5777	–	0.110	–	21,645
LSTMMDN: $L(k=32)\times G(A=6)$	5805	5929	0.119	0.226	7,122
LSTMMDN: $L(k=32)\times G(A=8)$	5811	5933	0.121	0.234	7,320
LSTMMDN: $L(k=64)\times G(A=6)$	5753	5864	0.102	0.195	22,418
LSTMMDN: $L(k=64)\times G(A=8)$	5777	5872	0.109	0.201	22,808
SVF-based estimates	6570	–	0.377	–	–

3.1.1. Data pre-processing

Missing data is an issue for the wind speed data, where 37% of the data were missing from the measurement station at DMI. The missing data are mainly centred around the winter months, potentially skewing the data representation and the model's ability to create accurate bicycle flow estimates. Therefore, we impute the missing data with wind speed observations from the nearest weather observation station, shown by the black asterisk in Fig. 4 . This station is approximately 7 kilometres removed from the DMI station. The imputed wind speeds only have 1.6% missing data. Thus we lose less data. The chosen approach may lead to a decreased precision in wind data for the bicycle volume estimation but presents a solid trade-off compared to missing 37% of the data. Any remaining 10-minute intervals of data with missing data, are omitted.

3.2. Bicycle flow estimation — model training and comparison

To assess the performance of the proposed LSTMMDN to estimate historical hourly bicycle volumes, we compare it against other neural network architectures and, most importantly, the method based on calibration factors currently employed by the Danish Road Directorate (the Seasonal Variation Factors, SVF). Five models for the estimation are being compared in various setups. A simple MDN with 1 hidden layer providing 18 outputs for 6 mixtures, the ANN setup proposed in Sekuła et al. (2018) to estimate hourly car traffic in Maryland, LSTM and LSTMMDN in similar setups, and the SVF-based method used by the Danish Road Directorate.

The ANN proposed by Sekuła et al. (2018) contains three hidden layers of each 258 neurons with 20% dropout in each layer and ELU activation (Clevert et al., 2015) in each hidden layer. The LSTM networks have a single LSTM cell with $k = 32$ or $k = 64$ computation nodes in the LSTM-gates. The output from the LSTM cell is passed through a single layer with $m = 6, m = 8$ computational nodes, and finally to a single node output layer. We also evaluate the performance of an LSTM network where the LSTM cell is connected directly to the single output node. The hidden layers all have linear $a(x) = x$ activation functions.

The LSTMMDN is set up similarly to the described LSTM networks with $k = 32$ or $k = 64$ nodes in the LSTM-gates, and either $A = 6$ or $A = 8$ mixture components. The latter means that the output layer will have $m = A \times 3$ as visualised in Fig. 3.

For better training efficiency, the flow data for model training is standardised to mean zero and standard deviation 1. The described models are all "trained" using the training data set and the validation set used for monitoring and early stopping. After the model training, the test data set is used to compare model performance. The models are compared based on the following measures:

- The average mse ($\hat{mse}$)
- The average negative log-likelihood $-\log \hat{\mathcal{L}}$
- The mse of the conditional average (mse_μ)
- The negative log-likelihood of the conditional average ($-\log \mathcal{L}_\mu$)

The first two are computed based on 100 posterior draws from the trained LSTMMDN, while the latter two are computed for the mean of the same 100 posterior draws. For the standard LSTM network, the ANN, and the calibration factor method, the last two measures are computed for one forecast as they all estimate the conditional average cycling flow already.

The results are shown in Table 1. Here $L(k)$ represents an LSTM cell with output dimension k, $\times \cdot m$ represents a connection with a hidden layer consisting of m computation nodes, and $\times G(A)$ refers to a connection with a GMM with A mixture components and therefore an output layer with $A \times 3$ computational nodes.

Based on the GOF measures presented in Table 1, the LSTMMDN $L(k = 64) \times G(A = 6)$ appears to be the superior of the models. Noteworthy is that all the LSTMMDNs outperform their LSTM network counterpart with similar setups. Considering the less refined nature of the ANN (Sekuła et al., 2018) compared to the LSTM-based models, it is surprising that it is the second-best performing model on the test data. However, the ANN has ≈ 8 times as many estimable parameters as the second-largest model and takes significantly longer to train. It could be considered if a complicated model structure like that of the LSTMs is therefore necessary, but comparing the output with the simple MDN, which has 6 mixtures, we see that a simple ANN feeding into a mixture model fails to capture the temporal correlations present in the bicycle flow data. The model, in fact, is not much better than the SVF approach.

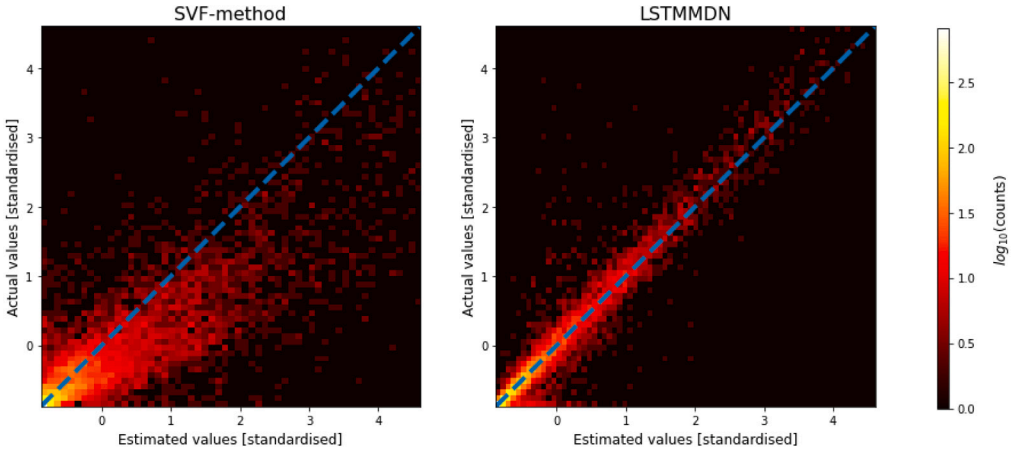

Fig. 5. Heat maps comparing the prediction accuracy of the SVF method (left) and the LSTMMDN method (right) based on the reserved test data set.

The most relevant comparison is the model-based approaches vs the SVF-based method currently used in road agencies and the Danish Road Directorate. Comparing all models-based estimates with the SVF-based estimates shows their superior performance when estimating the hourly bicycle flow while accounting for varying weather and time-related effects. With the improvements ranging from $66\% - 77\%$

For the remainder of this paper, we continue with the best performing LSTMMDN: $L(k = 64) \times G(A = 6)$, which will be compared further to the SVF-method for estimating hourly bicycle traffic.

3.3. LSTMMDN vs. SVF-calibration method

The LSTMMDN: $LSTM(64) \times G(6)$ performance on the test data yields an MSE_{μ} which is $\approx 77\%$ lower than that of the SVF method. To delve further into their respective ability to estimate hourly bicycle traffic accurately, we compare the fitted vs. actual bicycle flow using a heat plot in Fig. 5. The two heat plots reveal the superior ability of the LSTMMDN to estimate the hourly bicycle flow more accurately, as the concentration of estimates is much closer to the 45° line establishing a $x = y$ relation. It is similarly apparent from the heatplots in Fig. 5 that the SVF method tends to overestimate the bicycle volumes compared to the actual bicycle volumes. Meanwhile, we also see that the model-based approach seems to overestimate bike-flow in the low-bike-flow domain. This is primarily due to the few observations of low-exposure roads meaning that our model is not good at representing the bike flow on such roads and cycling lanes.

To look further into the discrepancies between the two bicycle flow estimation methods, we compare their representation of the bicycle flow over a continuous week. The plot in Fig. 6 shows a direct comparison of the standardised bicycle flow estimates from the LSTMMDN (orange) and the SVFs (green) for a continuous week at a monitoring station. The actual observations for the week at the station are shown in blue. Fig. 6 allows for a more fine-tuned diagnostic of the individual time-related effects learned by the LSTMMDN model vs the SVF calibration. Overall, the LSTMMDN estimates follow the observed values for the chosen road during this period much more tightly than the SVF estimates, coming closer in both the respective peaks and midday dips. However, the most significant difference between the LSTMMDN and the SVF estimates in Fig. 6 is the discrepancy of the SVF-based method to estimate the traffic at weekends accurately.

3.4. Impact of bicycle exposure on bicycle accident analyses

Having shown that our proposed method for estimating historical bicycle flow has potential compared to the calibration-factor method, we wish to assess the impact of more detailed bike-flow data on some city planners' and road agencies' tasks. One such task is to improve cycling safety, which involves accident analysis. As previous studies have highlighted the necessity of accounting for exposure in accident analysis (Vandenbulcke et al., 2014; Thomas and De Robertis, 2013; Aldred et al., 2018; Norros et al., 2016), we wish to quantify the impact of the quality of the exposure data further.

To do so, we estimate separate city-wide crash frequency models for Copenhagen with different exposure variables. Except for the exposure variable, all other exogenous variables are constant and equal across the different models. We consider three exposure variables: AAWCT, SVF-based hourly volume estimates, and LSTMMDN hourly volume estimates. The crash frequency model is a simple Poisson regression. This model has been used in many previous studies involving crash frequency (Lord and Mannering, 2010). Other slightly more general models were tested namely the Negative Binomial regression and the Zero Inflated Poisson regression. However, these models offered limited improvements in model fit to data and did not alter the conclusions based on Poisson models. The results in terms of Log Likelihood, AIC and BIC are shown in Table B.6.

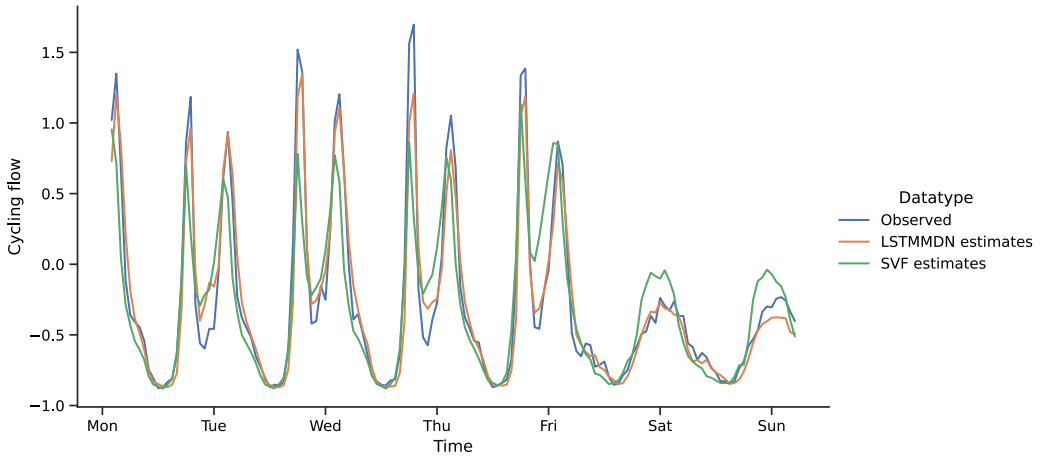

Fig. 6. Visualisation of the standardised bicycle flow over 148 continuously registered hours from a randomly picked counting station. Blue: Actual counts of bicyclists from random counting station, orange: LSTMMDN cycling flow estimates on the same road section and week, and green: SVF-based cycling flow estimates for same road section and week. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Summary statistics of the data used in the Poisson model.

Variable	Mean	Std. deviation
Visibility	27,519 m	14.266 m
Bank holiday	0.034	0.181
Exposure(LSTMMDN)	5430 cyclists/h	4221 cyclists/h
Exposure(SVF)	6051 cyclists/h	5594 cyclists/h
Exposure(AAWCT)	6079 cyclists/h	219 cyclists/h
Morning Peak/ Afternoon peak	0.089	0.285
Temperature < 0 °C	0.032	0.175
Temperature > 20 °C	0.075	0.262
Wind speed < 5 m/s	0.102	0.303
Wind speed > 9 m/s	0.008	0.087
Precipitation > 0 mm	0.429	0.471

We consider for the response variable the aggregated amount of bicycle crashes in Copenhagen during any given hour in the period 2017–2020. As exogenous variables, we consider the following: Visibility $[m]$, Temperature < 0 °C, Temperature > 20 °C, Morning peak hours (7–9 weekday), Afternoon peak (15–17 weekday), Wind speed < 5 m/s, Wind speed > 9 m/s, Precipitation (> 0 mm), Bank holidays, Bicycle flow.

Hourly cycling exposure is estimated for the four years for each bicycle counting station using the LSTMMDN and calibration factors, respectively. Subsequently, the estimates are averaged over all stations to approximate city-wide cycling exposure (similar to Dozza, 2017). When using the AAWCT as an exposure variable, the volume is scaled to be the annual average hourly cycling traffic.

Due to missing weather data over the four years, leading to missing bicycle flow estimates, we undersample the bicycle flow to have three full years of data. The three years of data include 2104 bicycle crashes in 26,232 h, meaning 0.08 accidents per hour on average. The variance in the number of accidents is 0.085. Some summary statistics of the data used in the bicycle crash frequency model are shown in Table 2 while the resulting parameter estimates for the estimated crash models are shown in Table 3. Correlation matrices can found in Appendix A. We see that only peak hours (morning/afternoon) are even moderately correlated with the flow estimates (i.e. $0.3 < |\text{corr}| < 0.5$). Furthermore, no variables are strongly correlated (i.e. $0.5 < |\text{corr}|$ (Overholser and Sowinski, 2008).

We see from the results in Table 3 that Model 3 is the best based on the log-likelihood and Deviance. This leads us to conclude that the LSTMMDN estimates as exposure variables lead to superior model performance in crash risk analysis. The resulting log-likelihood is 5.5% higher in Model 3 than in Model 1. Having changed only the exposure variable across the three models, this indicates that the quality of exposure estimates used in models have a substantial impact on fit to data. In addition, we find that Model 2 (SVF-based hourly cycling volumes) fits the data more accurately than Model 1, with a 3.9% higher log-likelihood.

The results in Table 3 also reveal magnitude differences in parameter estimates and a difference in sign and significance of variable effects across the three models. We find that conclusions regarding various variable impacts on the bicycle crash frequency vary depending on which of the three model variations is employed. For example visibility, which is significantly positive in Model 1, becomes negative in Models 2 and 3 being significant in Model 2 and borderline insignificant in Model 3. Similarly, we observe that high wind speeds are found to be significantly associated with a decrease in the crash frequency in Model 1, while being

Table 3
Estimates for three Poisson regressions using similar input variables with the exception of varying the exposure variable (AAWCT, SVF-based hourly cycling volume and LSTMMDN-based hourly cycling volume).

	Model 1 (AAWCT based exposure)		Model 2 (SVF-based exposure)		Model 3 (LSTMMDN based exposure)	
No. Obs	26,232		26,232		26,232	
Estimated parameters	10		10		10	
$\log \mathcal{L}$	−7, 142		−6, 874		−6, 740	
Deviance	10,287		9,750		9,483	
Variables	Parameter estimates	*p*-value	Parameter estimates	*p*-value	Parameter estimates	*p*-value
Intercept	−23.390	< 0.001	−7.958	< 0.001	−10.740	< 0.001
Visibility (for one log change)	0.071	0.029	−0.068	0.037	−0.062	0.055
Bank holiday	−0.857	< 0.001	−0.809	< 0.001	−0.096	0.609
log(Exposure)	2.267	< 0.001	0.697	< 0.001	1.006	< 0.001
Morning peak	1.098	< 0.001	0.210	0.002	0.247	< 0.001
Afternoon peak	1.194	0.003	0.296	< 0.001	0.284	< 0.001
Temperature < 0 °C	−0.512	< 0.001	−0.113	0.519	−0.027	0.877
Temperature > 20 °C	0.422	0.774	0.075	0.270	0.224	0.001
Wind speed < 5 m/s	−0.023	0.946	0.181	0.025	0.136	0.090
Wind speed > 9 m/s	−0.018	0.001	−0.058	0.824	−0.030	0.908
Precipitation	0.147	0.001	0.095	0.031	0.124	0.005

insignificant in the two other regression models. The same is the case for low temperatures, which are similarly significant and negative only in Model 1. Meanwhile, high temperatures and low wind speeds are insignificant in Model 1, while these variables are associated with significant effects in Model 3 and Model 2, respectively.

4. Discussion

The current approach used by transport agencies to estimate hourly bicycle traffic relies on calibration factors and thus does not reflect variations in cycling exposure related to weather and other effects. This study aims to amend this issue by applying an LSTMMDN to estimate city-wide hourly bicycle volumes in Copenhagen based on the mean-daily traffic while accounting for weather and temporal effects. The results clearly show that the proposed LSTMMDN produces significantly more accurate estimates of hourly cycling flow than the calibration factor method. Conditional on the size of the applied network, the LSTMMDN yields 66% to 77% more accurate estimates of the hourly bicycle volume. The LSTMMDN contrasts models previously used in estimation and short-term traffic forecast and traffic estimation studies (Ma et al., 2015; Sekuła et al., 2018) as it estimates a conditional cycling distribution, compared to only estimating a conditional average cycling flow. As such, the LSTMMDN should provide a more realistic representation of cycling. As cycling flows are treated as draws from the conditional cycling distribution, there is randomness in the system. This means that cycling flows on otherwise measurably identical roads will be different. This would not be the case for models estimating conditional averages. The above highlights that machine learning methods are capable of improving the precision in flow estimates necessary for policy analysis. This should not be seen as a claim that the LSTMMDN is optimal. There could be other methods giving even higher precision. In addition, any decision on an optimal method for a given application includes a trade off between model simplicity and precision. However, the LSTMMDN serves to illustrate that there is room for improvement in precision, which can be achieved by well-established methods.

Several policy areas within transport could be impacted by improved bicycle volume estimates, e.g. recently developed large-scale models that include link-based cycling volumes (Kjems and Paag, 2019; Davies, 2017). Another area being cycling safety improvement, where many policy applications rely on exposure, i.e. bicycle flow. As an illustration of policy analysis, we quantify the potential impact of improving the quality of bicycle flow estimation. Specifically, we estimate three bicycle crash frequency models, which are identical except for the exposure variable. The results show that improving the accuracy of bicycle volume estimates result in better crash risk models. The best results are achieved from best to worst using the LSTMMDN estimates, the SVF estimates, and the modified mean daily aggregated cycling. These results add to previous research arguing that the accuracy of analyses is improved by including exposure, as opposed to no exposure (Thomas and De Robertis, 2013; Norros et al., 2016), by arguing that the detail and quality of the exposure variable affect the accuracy of inference. A potentially worrying notion was raised from the results of the crash frequency models. We observe both sign and significance level changes concerning the variable effects when comparing the models in Table 3. Looking at the simplest Model 1, policymakers could be concerned that having many cyclists is problematic since accidents increase more than proportionally to exposure. However, in Model 3 we find that accidents only increase proportionally with exposure. Similarly, peak hours are seen as relatively much more dangerous in Model 1 than in Models 2 and 3. This could influence the priorities placed on various policies towards improving cycling safety. This variation is evidence of aggregation bias influencing the estimation results when using aggregated exposure estimates and raises questions about the validity of results from studies using highly aggregated exposure in models. However, a definitive clarification would require an in-depth analysis and is the subject of future studies. Nonetheless, it presents a strong argument for the need for further bicycle observation/monitoring efforts and the methods for estimating bicycle flow.

4.1. Limitations

4.1.1. Model-based hourly cycling volumes

Several variables are not accounted for that could be considered when attempting to accurately estimate cycling flow. Accounting for the built environment, network structure and connected routes could improve the model's accuracy. Applying models similar to Bao et al. (2019), which combine network and temporal features, could be considered for future research. Previous research suggests that cyclists prefer bicycle paths separated from motorised traffic (Aldred et al., 2017; Broach et al., 2012; Transport for London, 2014), and future applications should seek to include car traffic as a predictor in the model. The specific strength of the LSTM is its ability to handle very long time-series data. In this study, however, the input sequences passed to the LSTM only contain sequences of six 10-minute weather intervals. Future research could concern itself with the topic and include extended series of weather data into the bicycle flow model and include previous bicycle counts to improve accuracy.

4.1.2. Data

The current model relies on the availability of monitoring of hourly bicycle volumes. Therefore, if no data on bicycle volumes is available other methods need to be explored. In terms of policy making it is interesting to know which variables are crucial and which are negligible in cases where one might want to estimate traffic. Therefore we investigated the impact of removing variables considered in the study by leaving out one variable at a time, while keeping all other variables. We investigate the relative change in the GOF-statistics reported in Table 1 compared to the exhaustive model. This identifies temporal effects such as day of week, week of year and mean daily cycling traffic at a station as very important factors for determining the cycling flow. Leaving any of these three factors out resulted in 23.3%, 17.8% and 201% increase in the MSE_{μ} respectively. Similarly, the amount of data needed for models is relevant, and in terms of policymaking, data costs money. We therefore investigated the performance drop of the best model identified in the study if only using randomly sampled $5\%, 10\%, 20\%$ and 30% of the data to train the model, while testing the performance on the same data set as used for the results in Table 1. The results suggest that even with only 5% of the data for training instead of the 70% we used, the model achieves improvements of 50%. The full set of variables and results for this can be found in Appendix B.

The increased availability of technologies to monitor cyclists and their behaviour via systems such as instrumented cyclists (Gustafsson and Archer, 2013; Roos and Lindqvist, 2020) also implies the potential for new ways and methods to estimate cycling flow that could turn cycling counting stations obsolete in the future. Meanwhile, the newer technologies enabling the monitoring of cyclist volumes and behaviour come from private companies and their products, making them potentially less viable for the transport agencies. Therefore, although sparse, the most easily accessible cycling monitoring data for transport agencies and safety engineers still stems from automated bicycle counting stations. Hence, it should still be in their interest to develop models to improve the hourly bicycle estimates based on what little information about the cycling exposure is currently available.

4.1.3. Bicycle crash models

The crash model and the variables used for those models are very simplistic and quantify the impact that better bicycle volume estimates can have on the accident models that are crucial to making informed decisions to increase cycling safety. Many studies have investigated the factors related to bicycle crashes and the outcomes thereof Janstrup et al. (2019), Myhrmann et al. (2021), Fountas et al. (2021), Schepers et al. (2020), Aldred et al. (2018), Kaplan and Prato (2013), Kim et al. (2007). This study illustrates that the conclusions reached in these studies may have a dependence on the detail of exposure used in the various studies. The three Poisson regressions results indicate that the bicycle flow estimates affect the crash analysis. This result makes it plausible that the same would be the case in more advanced models. However, this is left for future research. As previously mentioned, we tested a Negative Binomial regression and Zero Inflated Poisson regression to compare to the simplistic nature of the Poisson regression in this paper. In neither of the three exposure settings did we find evidence of a significant dispersion parameter. As is seen in Appendix C there is only a benefit to increasing the model complexity by applying NB or ZIP regressions, when using the simplest type of exposure measure, namely the AAWCT. Meanwhile, when using the SVF and LSTMMDN-based exposure measures the Poisson regression performs the best when evaluating the AIC and BIC measures.

5. Conclusion

This paper focuses on the application of machine learning for estimating historical bicycle flows that transport agencies use for various tasks such as improving cycling safety or as input in larger transport models. The proposed LSTMMDN estimates cycling traffic based on the mean daily traffic and accounts for weather and time-related factors. The proposed method significantly outperforms the current calibration factor method to obtain hourly bicycle flow estimates, and the cycling flow estimates are up to 77% more accurate in the studied example. The LSTMMDN also provides a more realistic cycling representation due to the model's built-in uncertainty. Assuming less monitoring of the cycling network and therefore less data for training still leads to model performances far superior compared to the current calibration factor method. Using only 5% of the original data for model estimation still leads to cycling flow estimates that are 50% more accurate in the studied example. Overall this suggests that traffic estimation efforts in transport agencies would benefit from accounting for weather and other influencing factors as well as embracing newer statistical frameworks. Also, the quality of exposure data could easily be thought to impact the results of their analyses. In line with this, the current study analyses the impact of the improved cycling flow data from the LSTMMDN in accident analysis. The results show how these models, generally used to make informed decisions regarding mitigating action, benefit from the more detailed and accurate bicycle flow data. Hence, the results suggest that highly aggregated exposure data could lead to erroneous conclusions.

Table A.4
Table showing the Pearson correlation values between the flow and other variables in the Poisson regressions. Bold indicates correlations that are significantly different from zero.

	log(vis)	Bank holiday	Morning peak	Afternoon peak	Temp. < 0 °C	Temp. > 20 °C	Wind < 5 m/s	Wind > 9 m/s	Precip.
log(SVF-flow)	**0.18**	**−0.01**	**0.32**	**0.36**	**−0.12**	**0.23**	**−0.14**	0.01	**−0.01**
log(LSTMMDN-flow)	**0.18**	**−0.09**	**0.23**	**0.31**	**−0.15**	**0.18**	**−0.10**	**0.00**	−0.01

Table A.5
Table showing the Pearson correlation values between exogenous variables (non-flow) in the Poisson crash models. Bold indicates correlations that are significantly different from zero.

	log(vis)	Bank holiday	Morning peak	Afternoon peak	Temp. < 0 °C	Temp. > 20 °C	Wind < 5 m/s	Wind > 9 m/s	Precip.
log(vis)	1.00	**0.06**	**−0.05**	**0.07**	**0.02**	**0.15**	**−0.12**	0.01	**−0.14**
Bank holiday	**0.06**	1.00	0.00	0.00	0.00	**−0.05**	**−0.02**	**0.04**	0.01
Morning peak	**−0.05**	0.00	1.00	**−0.10**	0.01	−0.01	**−0.03**	0.01	0.00
Afternoon peak	**0.07**	0.00	**−0.10**	1.00	**−0.03**	**0.09**	**−0.06**	0.01	0.00
Temp. < 0 °C	**0.02**	0.00	0.01	**−0.03**	1.00	**−0.05**	**0.02**	**−0.01**	**−0.08**
Temp. > 20 °C	**0.15**	**−0.05**	−0.01	**0.09**	**−0.05**	1.00	**−0.07**	**−0.02**	**−0.12**
Wind < 5 m/s	**−0.12**	**−0.02**	**−0.03**	**−0.06**	**0.02**	**−0.07**	1.00	**−0.03**	−0.01
Wind > 9 m/s	0.01	**0.04**	0.01	0.01	**−0.01**	**−0.02**	**−0.03**	1.00	0.01
Precip.	**−0.14**	0.01	0.00	0.00	**−0.08**	**−0.12**	−0.01	0.01	1.00

CRediT authorship contribution statement

Marcus Skyum Myhrmann: Conceptualization, Data management, Formal analysis, Visualisation, Writing – original draft, Writing – review & editing. **Stefan Eriksen Mabit:** Conceptualization, Formal analysis, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

We thank Filipe Rodrigues and Mads Paulsen for valuable feedback on an earlier version of the paper. We are also grateful to two anonymous reviewers for their useful comments that helped us improve the paper.

Appendix A. Correlation tables for regression

See Tables A.4 and A.5.

Appendix B. Comparing Poisson, negative binomial and ZIP models

See Table B.6.

Appendix C. Test of limited data and variables

See Tables C.7 and C.8.

Table B.6
Comparison of crash frequency models using Poisson, Negative Binomial (NB) and Zero-inflated Poisson (ZIP) regressions.

GOF	Poisson reg.	NB reg.	ZIP reg.
AAWCT-based exposure			
$\log \mathcal{L}$	−7142	−7141	−7122
AIC	14306	14305	14288
BIC	14396	14404	14468
SVF-based exposure			
$\log \mathcal{L}$	−6874	−6874	−6863
AIC	13770	13773	13771
BIC	13860	13871	13951
LSTMMDN-based exposure			
$\log \mathcal{L}$	−6740	−6743	−6734
AIC	13502	13509.11	13511.18
BIC	13591.92	13607.2	13691.03

Table C.7
Relative difference in GOF measures related to different variables left out in the training of the LSTMMDN: $L(k = 64) \times G(A = 6)$, compared to the exhaustive model, determined on the test data.

Left out variable	$\widehat{MSE}$ (%)	MSE_{μ} (%)	$-\log \hat{\mathcal{L}}$ (%)	$-\log \mathcal{L}_{\mu}$ (%)
Temperature	−1.82	1.46	0.05	0.08
Humidity	1.46	−1.54	0.03	−0.08
Precipitation	3.29	−0.29	0.30	−0.02
Pressure	3.75	1.83	0.20	0.10
Visibility	0.68	2.59	0.23	0.14
Visibility (past 10 min)	−1.02	−3.67	−0.11	−0.20
Holiday	0.37	2.50	0.15	0.14
Day of week	24.12	23.33	1.86	1.29
Week of year	16.55	17.76	1.12	0.98
Wind speed	−6.59	−2.14	−0.03	−0.12
Wind Gusts	−2.68	−1.30	0.05	−0.07
AAWCT	205.52	201.28	13.61	11.10

Table C.8
GOF measures related to the LSTMMDN: $L(k = 64) \times G(A = 6)$ trained on reduced amounts of data, determined on the test data.

Fraction of data	$\widehat{MSE}$	MSE_{μ}	$-\log \hat{\mathcal{L}}$	$-\log \mathcal{L}_{\mu}$
5%	0.316	0.188	−6221	−6010
10%	0.284	0.164	−6123	−5939
20%	0.266	0.152	−6058	−5903
30%	0.238	0.136	−5982	−5854

References

Aldred, R., Elliott, B., Woodcock, J., Goodman, A., 2017. Cycling provision separated from motor traffic: a systematic review exploring whether stated preferences vary by gender and age. Transp. Rev. 37 (1), 29–55. http://dx.doi.org/10.1080/01441647.2016.1200156.

Aldred, R., Goodman, A., Gulliver, J., Woodcock, J., 2018. Cycling injury risk in London: A case-control study exploring the impact of cycle volumes, motor vehicle volumes, and road characteristics including speed limits. Accid Anal Prev. 117, 75–84. http://dx.doi.org/10.1016/J.AAP.2018.03.003, URL: https://www.sciencedirect.com/science/article/pii/S0001457518301076?via%3Dihub.

Bao, J., Liu, P., Ukkusuri, S.V., 2019. A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data. Accid. Anal. Prev. http://dx.doi.org/10.1016/j.aap.2018.10.015.

Behnood, A., Roshandeh, A.M., Mannering, F.L., 2014. Latent class analysis of the effects of age, gender, and alcohol consumption on driver-injury severities. Anal. Methods Accid. Res. 3–4, 56–91. http://dx.doi.org/10.1016/J.AMAR.2014.10.001, URL: https://www.sciencedirect.com/science/article/pii/S2213665714000256.

Bishop, C.M., 1994. Mixture Density Networks. Technical Report, Department of Computer Science and Applied Mathematics, Aston University, URL: http://www.ncrg.aston.ac.uk/.

Böcker, L., Dijst, M., Prillwitz, J., 2013. Impact of everyday weather on individual daily travel behaviours in perspective: A literature review. Transp. Rev. 33 (1), 71–91. http://dx.doi.org/10.1080/01441647.2012.747114, URL: http://www.tandfonline.com/doi/abs/10.1080/01441647.2012.747114.

Boele-Vos, M., Van Duijvenvoorde, K., Doumen, M., Duivenvoorden, C., Louwerse, W., Davidse, R., 2017. Crashes involving cyclists aged 50 and over in the Netherlands: An in-depth study. Accid Anal Prev. 105, 4–10. http://dx.doi.org/10.1016/J.AAP.2016.07.016, URL: https://www.sciencedirect.com/science/article/pii/S0001457516302457.

Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transp. Res. A 46 (10), 1730–1740. http://dx.doi.org/10.1016/j.tra.2012.07.005.

Chen, C., Anderson, J.C., Wang, H., Wang, Y., Vogt, R., Hernandez, S., 2017. How bicycle level of traffic stress correlate with reported cyclist accidents injury severities: A geospatial and mixed logit analysis. Accid Anal Prev. 108, 234–244. http://dx.doi.org/10.1016/J.AAP.2017.09.001, URL: https://www.sciencedirect.com/science/article/pii/S0001457517303160.

Chen, J.F., Lo, S.K., Do, Q.H., 2018. Forecasting short-term traffic flow by fuzzy wavelet neural network with parameters optimized by biogeography-based optimization algorithm. Comput. Intell. Neurosci. 2018, http://dx.doi.org/10.1155/2018/5469428.

Chen, Y.Y., Lv, Y., Li, Z., Wang, F.Y., 2016. Long short-term memory model for traffic congestion prediction with online open data. In: IEEE Conference on Intelligent Transportation Systems, Proceedings. ITSC, IEEE, pp. 132–137. http://dx.doi.org/10.1109/ITSC.2016.7795543.

Clevert, D.A., Unterthiner, T., Hochreiter, S., 2015. Fast and accurate deep network learning by exponential linear units (ELUs). In: 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings. International Conference on Learning Representations, ICLR, URL: http://arxiv.org/abs/1511.07289.

Cui, Z., Ke, R., Pu, Z., Wang, Y., 2020. Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. Transp. Res. C 118 (March 2019), 102674. http://dx.doi.org/10.1016/j.trc.2020.102674.

Davies, A., 2017. Cynemon-Cycling Network Model for London. Technical Report, Transport for London, URL: https://www.ucl.ac.uk/transport/sites/transport/files/Davies_slides.pdf.

DMI, 2021. Danish Meteorological Institute - open data - DMI open data - confluence. URL: https://confluence.govcloud.dk/display/FDAPI/Danish+Meteorological+Institute+-+Open+Data.

Dozza, M., 2017. Crash risk: How cycling flow can help explain crash data. Accid Anal Prev. 105, 21–29. http://dx.doi.org/10.1016/J.AAP.2016.04.033, URL: https://www.sciencedirect.com/science/article/pii/S0001457516301464.

Du, W., Zhang, Q., Chen, Y., Ye, Z., 2021. An urban short-term traffic flow prediction model based on wavelet neural network with improved whale optimization algorithm. Sustainable Cities Soc. 69, 102858. http://dx.doi.org/10.1016/J.SCS.2021.102858.

Duan, Y., Lv, Y., Wang, F.Y., 2016. Travel time prediction with LSTM neural network. In: IEEE Conference on Intelligent Transportation Systems, Proceedings. ITSC, IEEE, pp. 1053–1058. http://dx.doi.org/10.1109/ITSC.2016.7795686.

Fountas, G., Fonzone, A., Olowosegun, A., McTigue, C., 2021. Addressing unobserved heterogeneity in the analysis of bicycle crash injuries in Scotland: A correlated random parameters ordered probit approach with heterogeneity in means. Anal. Methods Accid. Res. 32, 100181. http://dx.doi.org/10.1016/J.AMAR.2021.100181.

Gustafsson, L., Archer, J., 2013. A naturalistic study of commuter cyclists in the greater Stockholm area. Accid. Anal. Prev. 58, 286–298. http://dx.doi.org/10.1016/j.aap.2012.06.004.

Hastie, T., Tibshirani, R., Friedman, J., 2009. Neural networks. In: The Elements of Statistical Learning Data Mining, Inference, and Prediction. pp. 389–416. http://dx.doi.org/10.1007/978-0-387-84858-7{_}11, URL: http://link.springer.com/10.1007/978-0-387-84858-7_11.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735, URL: https://www.mitpressjournals.org/doi/abs/10.1162/neco.1997.9.8.1735.

Horton, D., 2016. Fear of cycling. In: Cycling and Society. Routledge, pp. 149–168. http://dx.doi.org/10.4324/9781315575735-13, URL: https://www.taylorfrancis.com/books/9781317155140/chapters/10.4324/9781315575735-13.

Janstrup, K.H., Møller, M., Pilegaard, N., 2019. A clustering approach to integrate traffic safety in road maintenance prioritization. Traffic Inj. Prev. 20 (4), 442–448. http://dx.doi.org/10.1080/15389588.2019.1580700, URL: https://www.tandfonline.com/doi/full/10.1080/15389588.2019.1580700.

Ji, S., Wang, Y., Wang, Y., 2021. Geographically weighted poisson regression under linear model of coregionalization assistance: Application to a bicycle crash study. Accid Anal Prev. 159, 106230. http://dx.doi.org/10.1016/j.aap.2021.106230, URL: https://linkinghub.elsevier.com/retrieve/pii/S000145752100261X.

Kaplan, S., Prato, C.G., 2013. Cyclist-motorist crash patterns in Denmark: a latent class clustering approach.. Traffic Inj. Prev. 14 (7), 725–733. http://dx.doi.org/10.1080/15389588.2012.759654.

Kaplan, S., Prato, C.G., 2015. A spatial analysis of land use and network effects on frequency and severity of cyclist–motorist crashes in the Copenhagen region. Traffic Inj. Prev. 16 (7), 724–731. http://dx.doi.org/10.1080/15389588.2014.1003818, URL: https://www.tandfonline.com/doi/full/10.1080/15389588.2014.1003818.

Kaplan, S., Vavatsoulas, K., Prato, C.G., 2014. Aggravating and mitigating factors associated with cyclist injury severity in Denmark. J. Saf. Res. 50, 75–82. http://dx.doi.org/10.1016/j.jsr.2014.03.012, URL: https://www.sciencedirect.com/science/article/pii/S0022437514000437?via%3Dihub.

Kim, J.K., Kim, S., Ulfarsson, G.F., Porrello, L.A., 2007. Bicyclist injury severities in bicycle–motor vehicle accidents. Accid Anal Prev. 39 (2), 238–251. http://dx.doi.org/10.1016/J.AAP.2006.07.002, URL: https://www.sciencedirect.com/science/article/pii/S000145750600128X.

Kingma, D.P., Ba, J.L., 2015. Adam: A method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. p. 15, International Conference on Learning Representations, ICLR.

Kjems, S., Paag, H., 2019. COMPASS: Ny trafikmodel for hovedstadsområdet. In: Proceedings from the Annual Transport Conference At Aalborg University. Copenhagen, p. 4, URL: www.trafikdage.dk/artikelarkiv.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transp. Res. A 44 (5), 291–305. http://dx.doi.org/10.1016/J.TRA.2010.02.001, URL: https://www.sciencedirect.com/science/article/pii/S0965856410000376.

Lv, Y., Duan, Y., Kang, W., Li, Z., Wang, F.Y., 2015. Traffic flow prediction with big data: A deep learning approach. IEEE Trans. Intell. Transp. Syst. 16 (2), 865–873. http://dx.doi.org/10.1109/TITS.2014.2345663.

Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y., 2015. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp. Res. C 54, 187–197. http://dx.doi.org/10.1016/j.trc.2015.03.014, URL: https://linkinghub.elsevier.com/retrieve/pii/S0968090X15000935.

McLachlan, G.J., Basford, K.E., 1988. Mixture models : inference and applications to clustering. MIT OpenCourseware 84 (2), 353–362, URL: http://www.ncbi.nlm.nih.gov/pubmed/20491729.

Morrison, C.N., Thompson, J., Kondo, M.C., Beck, B., 2019. On-road bicycle lane types, roadway characteristics, and risks for bicycle crashes. Accid. Anal. Prev. 123 (August 2018), 123–131. http://dx.doi.org/10.1016/j.aap.2018.11.017.

Mueller, N., Rojas-Rueda, D., Cole-Hunter, T., de Nazelle, A., Dons, E., Gerike, R., Götschi, T., Int Panis, L., Kahlmeier, S., Nieuwenhuijsen, M., 2015. Health impact assessment of active transportation: A systematic review. Prev. Med. 76, 103–114. http://dx.doi.org/10.1016/J.YPMED.2015.04.010, URL: https://www.sciencedirect.com/science/article/pii/S0091743515001164?via%3Dihub.

Myhrmann, M.S., Janstrup, K.H., Møller, M., Mabit, S.E., 2021. Factors influencing the injury severity of single-bicycle crashes. Accid Anal Prev. 149, 105875. http://dx.doi.org/10.1016/j.aap.2020.105875, URL: https://linkinghub.elsevier.com/retrieve/pii/S000145752031695X.

Nankervis, M., 1999. The effect of weather and climate on bicycle commuting. Transp. Res. A 33 (6), 417–431. http://dx.doi.org/10.1016/S0965-8564(98)00022-6.

Norros, I., Kuusela, P., Innamaa, S., Pilli-Sihvola, E., Rajamäki, R., 2016. The palm distribution of traffic conditions and its application to accident risk assessment. Anal. Methods Accid. Res. 12 (March), 48–65. http://dx.doi.org/10.1016/j.amar.2016.10.002.

Nosal, T., Miranda-Moreno, L.F., 2014. The effect of weather on the use of North American bicycle facilities: A multi-city analysis using automatic counts. Transp. Res. A 66 (1), 213–225. http://dx.doi.org/10.1016/j.tra.2014.04.012.

Overholser, B.R., Sowinski, K.M., 2008. Biostatistics primer: Part 2. In: Nutrition in Clinical Practice, Vol. 23, No. 1. John Wiley & Sons, Ltd, pp. 76–84. http://dx.doi.org/10.1177/011542650802300176, https://onlinelibrary.wiley.com/doi/full/10.1177/011542650802300176, https://onlinelibrary.wiley.com/doi/abs/10.1177/011542650802300176, https://aspenjournals.onlinelibrary.wiley.com/doi/10.1177/011542650802300176.

Polson, N.G., Sokolov, V.O., 2017. Deep learning for short-term traffic flow prediction. Transp. Res. C 79, 1–17. http://dx.doi.org/10.1016/J.TRC.2017.02.024.

Raihan, M.A., Alluri, P., Wu, W., Gan, A., 2019. Estimation of bicycle crash modification factors (CMFs) on urban facilities using zero inflated negative binomial models. Accid. Anal. Prev. 123, 303–313. http://dx.doi.org/10.1016/j.aap.2018.12.009.

Roos, J., Lindqvist, S., 2020. Identifiering av områden med förhöjd olycksrisk för cyklister baserad på cykelhjälmsdata. MalmÖ University Electronic Publishing, Technical Report, Malmö universitet/Teknik och samhälle, Malmö University, URL: http://muep.mau.se/handle/2043/32083.

Rossetti, T., Guevara, C.A., Galilea, P., Hurtubia, R., 2018. Modeling safety as a perceptual latent variable to assess cycling infrastructure. Transp. Res. A 111, 252–265. http://dx.doi.org/10.1016/j.tra.2018.03.019.

Saha, D., Alluri, P., Gan, A., Wu, W., 2018. Spatial analysis of macro-level bicycle crashes using the class of conditional autoregressive models. Accid Anal Prev. 118, 166–177. http://dx.doi.org/10.1016/J.AAP.2018.02.014, URL: https://www.sciencedirect.com/science/article/pii/S000145751830071X?via%3Dihub.

Samerei, S.A., Aghabayk, K., Shiwakoti, N., Mohammadi, A., 2021. Using latent class clustering and binary logistic regression to model Australian cyclist injury severity in motor vehicle–bicycle crashes. J. Saf. Res. http://dx.doi.org/10.1016/J.JSR.2021.09.005.

Schepers, P., de Geus, B., van Cauwenberg, J., Ampe, T., Engbers, C., 2020. The perception of bicycle crashes with and without motor vehicles: Which crash types do older and middle-aged cyclists fear most? Transp. Res. F 71, 157–167. http://dx.doi.org/10.1016/j.trf.2020.03.021.

Schrank, D., 2021. 2021 Urban Mobility Report – Appendix A: Methodology. Technical Report, Texas A&M Transportation Institute, URL: http://mobility.tamu.edu/umr/congestion-data/.

Sekuła, P., Marković, N., Vander Laan, Z., Sadabadi, K.F., 2018. Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A maryland case study. Transp. Res. C 97 (December 2017), 147–158. http://dx.doi.org/10.1016/J.TRC.2018.10.012, URL: https://www.sciencedirect.com/science/article/pii/S0968090X18314773?dgcid=rss_sd_all.

Srivastava, N., Hinton, G., Krizhevsky, A., Salakhutdinov, R., 2014. Dropout: A simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Thomas, B., De Robertis, M., 2013. The safety of urban cycle tracks: A review of the literature. Accid. Anal. Prev. 52, 219–227. http://dx.doi.org/10.1016/j.aap.2012.12.017.

Transport for London, 2014. Attitudes Towards Cycling Annual Report 2014 TfL. Technical Report, Transport for London, pp. 1–198, URL: https://tfl.gov.uk/cdn/static/cms/documents/attitudes-to-cycling-2014-report.pdf.

Twisk, D.A.M., Reurings, M., 2013. An epidemiological study of the risk of cycling in the dark: The role of visual perception, conspicuity and alcohol use. Accid. Anal. Prev. 60, 134–140. http://dx.doi.org/10.1016/j.aap.2013.08.015.

Vandenbulcke, G., Thomas, I., Int Panis, L., 2014. Predicting cycling accident risk in Brussels: A spatial case–control approach. Accid Anal Prev. 62, 341–357. http://dx.doi.org/10.1016/J.AAP.2013.07.001, URL: https://www.sciencedirect.com/science/article/pii/S0001457513002686.

Vejdirektoratet, 2018. Hvorfor stiller vi cyklen?. Technical Report, Vejdirektoratet, Copenhagen, p. 16.

Vejdirektoratet, 2020. Mastra, URL: https://www.vejdirektoratet.dk/side/viden-om-ydelser-mastra.

Zhao, Z., Chen, W., Wu, X., Chen, P.C., Liu, J., 2017. LSTM network: A deep learning approach for short-term traffic forecast. IET Image Process. 11 (1), 68–75. http://dx.doi.org/10.1049/iet-its.2016.0208.