Discussion Paper

Road Sensor Traffic Flow Density Estimation using Neural Networks

Dewi Peerlings, Jan van den Brakel and Nalan Baştürk

January 3, 2023

Explaining traffic flow is important for planning purposes and for developing intelligent transport systems. Such estimations are often based on road sensor data, which contain missing and erroneous data points due to e.g. sensor malfunctioning. We propose to apply a non-parametric probability density function (PDF) estimation based on Neural Networks (NN) to pre-process road sensor data and to report the properties of the traffic flow distribution. We extend the literature on traffic flow analysis in two ways. First, we propose a density estimation method instead of the standard point estimation method to account for uncertainty in raw data and estimation as well as to report the probability of e.g. extreme events in traffic flow. Second, we design NNs to specifically allow for correlations between adjacent road sensors and over time. We apply the proposed method to data obtained from highways in the Netherlands belonging to different sensors on the same highway section and coded as time series of vehicle counts. We show that the proposed model captures distributional properties of these data and can be used as a pre-processing and density estimation method.

Key words: road sensors, traffic flow data, multivariate density estimation, neural networks, unsupervised learning

The paper has been reviewed by Joep Burger (Methodology Department, Statistics Netherlands) and Rob Willems (Methodology Department, Statistics Netherlands).

The views expressed in this paper are those of the authors, and do not necessarily reflect the policies of Statistics Netherlands. The authors are grateful to the reviewers for careful reading of a former draft of the manuscript and providing constructive comments to further improve the paper.

1 Introduction

Over the last years, research in the Intelligent Transportation System (ITS) field has increased substantially due to the growing availability of data. Traffic flow analysis, which is important for planning purposes and for traffic management, is one component of ITS for which this applies. Traffic flow data, commonly measured by road sensors along highways, are affected by spatial and temporal factors as well as environmental factors such as weather conditions or accidents. These data can be very noisy due to malfunctioning sensors or due to factors that cause extreme outlying traffic behaviour such as black ice, heavy snow fall or car accidents. Particularly in the case of anomaly measurements or missing observations due to malfunctioning sensors, it is important to pre-process these data prior to using them for forecasting, see Ermagun and Levinson (2018) for an overview.

This paper proposes a Neural Network (NN) approach to estimate the probability density of traffic flow data to describe the characteristics of these data. The obtained density can be used to impute missing values and to assess and predict extreme events under the assumption of missing completely at random (MCAR) data. We design the proposed NNs specifically to estimate the inter-dependencies between road sensors that are on the same road section and time dynamics of vehicle counts. The proposed method relates to three strands of the literature, namely forecasting traffic flow, pre-processing of road sensor data and density estimation using NNs.

The first strand of the literature, forecasting traffic flow, is often performed either by the use of classical time series methods, such as autoregressive integrated moving average (ARIMA) models or state space models (SSM), see e.g. Duan et al. (2018), or by relatively new deep learning methods such as spatial-temporal dynamic networks (STDN) or long short-term memory models (LSTM), see e.g. Do et al. (2019); Fang et al. (2022). Particularly the latter, deep learning models, can account for the noise as well as the non-linear and non-stationary behavior of the underlying data generating process (DGP) of traffic flow, see Vlahogianni et al. (2014). However, deep learning methods may fail to capture all the dynamics of traffic flow despite their generality, see Ma et al. (2020) who use post-processing methods to remove remaining traffic flow features in the residuals of a neural network model. We propose an NN architecture to pre-process and impute traffic flow data prior to using one or more of the mentioned forecasting methods.

Prior to forecasting, the noisy data needs to be pre-processed which relates to the second strand of related literature. Such pre-processing can be based on the analysis of the underlying data DGP, see Fusco et al. (2016) and Lippi et al. (2013) among others. Other studies pre-process the data without assuming a specific DGP, e.g. Chen et al. (2018) and Yao et al. (2019) rescale the data, Fang et al. (2022) uses the wavelet de-noising method and Qu et al. (2019), Ma et al. (2020) and Duan et al. (2018) impute missing/removed observations by adjacent or the average of previous observations. Furthermore, it is possible to combine the two methods. Polson and Sokolov (2017) for example use median filtering as well as constructing bivariate probability mass functions (PMFs) to find bivariate relations. Thus, there is no single method of pre-processing these data and often data properties need to be taken into account. Our method differs from the above-mentioned literature by modeling the full vehicle count distribution

which can be used to impute missing values, report the uncertainty in vehicle counts as well as to report probabilities of extreme events. We further show that the proposed method can be used to join information from multiple sensors to strengthen the estimation and it can be used to estimate the time series properties of the data.

The third strand of the literature that our method relates to is density estimation using NNs. There are two different approaches within this literature, estimating the probability density function (PDF) directly by NNs as in Trentin et al. (2018), or estimating the CDFs by NNs from which the PDF hereafter is obtained as in Magdon-Ismail and Atiya (2002) and Peerlings et al. (2022). We extend the recent PDF estimation method in Peerlings et al. (2022) to analyze road sensor data taking into account correlations of vehicle counts between sensors and over time. This implies we extimate probability mass functions (PMFs) instead of PDFs. Peerlings et al. (2022) proposes a two-step approach in which firstly the cumulative distribution function (CDF) is estimated by neural networks and secondly the structure of the neural network is differentiated to obtain the corresponding PMF. The advantages of this approach, compared to standard parametric distributions, is that it is able to approximate a much wider range of distributions. The advantages of this two-step approach compared to directly approximating the PDF by a NN are two-fold. Firstly, constructing labels are more straightforward for CDF values than for PDF values as CDF values are monotonic and range between 0 and 1. Secondly, this method is more robust against statistical error as it does not need to specify any bandwidth parameter.

As an extension of Peerlings et al. (2022), we process the data and design the NN in four ways, denoted as different approaches. Firstly, the *intra-day approach* which comprises PMFs of 15 minute time windows over the course of the day, resulting in 96 different PMFs a day. In this way regular everyday traffic flow behavior of each 15 minutes of the day are found. This can be used to make more informed imputations for possible missing observations due to e.g. malfunctioning of the sensor or to improve short-term forecasting. This is in line with short-term forecasting literature which has been progressing over the years, see e.g. Manibardo et al. (2021) for an overview. Secondly, the *inter-day approach* which comprises PMFs of a full day, that is 1440 minutes, focuses on irregularities of traffic flow. This approach can be used to compare traffic flow behavior between days, for example to study effects of severe weather conditions or car accidents. The use of these contextual factors is in line with the literature such as Ma et al. (2020). Thirdly, the *cross-sectional approach* estimates bivariate PMFs taking into account different sensors. Bivariate densities account for the correlation between road sensors placed on the same road section, or closely located sensors, which can be used to obtain better imputations for missing observations. More recently, it has been shown that the use of multivariate data improves traffic flow forecasting by combining spatial as well as temporal relations, see Tak et al. (2016). Therefore our final approach, the *temporal approach*, qualifies the dependency of one sensor at different points of time. In this way the rate at which traffic flow evolves over time is analyzed. Habtemichael and Cetin (2016) use similar patterns over time by using k-nearest neighbors (kNN) to pick the optimal lag length to forecast traffic flow. In this paper, we focus on consecutive time windows to find temporal relations. The design of the proposed NNs and density estimation can be applied to different data sets with similar cross-sectional and temporal relations.

The empirical results illustrate that the proposed non-parametric method to estimate

PMFs using NNs is very flexible. The approximation errors of the CDFs are small. As a result our method captures a wide variety of empirical densities resulting in more realistic approximations of the DGP compared to more standard parametric density approximations. The resulting PMFs exhibit complex patterns and multi-modality. These approaches are therefore in particular promising for pre-processing road sensor data as well as other noisy big data, for example in multiple imputation methods (Rubin, 1996) to obtain analysis results that accommodate missing data patterns without underestimating the level of uncertainty.

The paper is outlined as follows. The methodology that has been used for density estimation is explained in Section 2. The road sensor data with the corresponding approaches are treated in Section 3. Then the results of these are shown in Section 4. Finally, a discussion is given in Section 5.

2 Methodology

The probability density function (PDF) of a data generating process (DGP) is estimated using a neural network, following the approach proposed by Peerlings et al. (2022). This approach has the advantage that it refrains from parametric assumptions about the distribution of the data and has the flexibility to capture a wide range of distributions. Therefore it can be interpreted as a non-parametric approach, despite a neural network model is applied to describe it. We apply the PDF estimation method in Peerlings et al. (2022) to traffic flow data retrieved from road sensors. A multilayer perceptron (MLP) is used to estimate the empirical cumulative distribution function (CDF) of the data. Then the MLP is differentiated with respect to the input variables in order to obtain the probability mass function (PMF) as the data we work with is count data. The input variables are allowed to be multivariate and dependent.

More specifically, the MLP consists of H hidden layers, M hidden neurons for each layer and N input layers collected in an $N \times T$ matrix X, where the rows $x_{n\cdot} = (x_{n1}, x_{n2}, \dots, x_{nT})$ for $n = 1, \dots, N$ consist of the n^{th} variable or feature and columns $x_{\cdot t} = (x_{1t}, x_{2t}, \dots, x_{Nt})^\top$ for $t = 1, \dots, T$ consist of the t^{th} observation of the N variables. The input layer, hidden layers and the output layers are connected by weight matrices and biases. The input layer is connected to the first hidden layer by an $(N \times M)$ matrix $W^{[0]}$ and $(M \times 1)$ bias vector $b^{[0]}$, $(M \times M)$ matrix $W^{[q]}$ together with $(M \times 1)$ bias vector $b^{[q]}$ connect hidden layers q and q+1 and $(M \times 1)$ vector $W^{[H]}$ with (1×1) bias $b^{[H]}$ connect the last hidden layer H to the output layer. Each hidden layer is composed of a non-linear function together with a linear combination of the input variables and corresponding weight matrices and biases. The non-linear function used in this paper is the logistic output function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \tag{1}$$

where $\sigma(\cdot)$ is an activation function that takes as input an $(M \times 1)$ vector, which implies $\sigma(h^{[q]}) : \mathbb{R}^M \to \mathbb{R}^M$, with $q = 1, \dots, H-1$ and $\sigma(h^{[H]}) : \mathbb{R}^M \to \mathbb{R}^1$. We chose this activation function as it is N-differentiable and has range $[0, 1]$ which is convenient

working with probabilities. Feeding forward through the MLP results in the output $\tilde{y}_t$ which represents the empirical CDF with target variable $\hat{y}_t$:

$$h_{\cdot t}^{[1]} = \sigma(W^{[0]\top} x_{\cdot t} + b^{[0]}) \tag{2}$$

$$h_{\cdot t}^{[q]} = \sigma(W^{[q-1]\top} h_{\cdot t}^{[q-1]} + b^{[q-1]}), \quad \text{for} \quad q = 2, \dots, H \tag{3}$$

$$\tilde{y}_t = \sigma(W^{[H]\top} h_{\cdot t}^{[H]} + b^{[H]}), \tag{4}$$

where $h_{\cdot t} = (h_{1t}, h_{2t}, \dots, h_{Mt})^\top$ for $t = 1, \dots, T$.

The NN approach considers several different MLPs which are trained by the mean squared error (MSE) loss function. We use the MSE function due to noisy data that contains many missing observations. For computational efficiency, these models, each with different number of hidden layers and hidden neurons, are combined and trained simultaneously in one large MLP, for details see Peerlings et al. (2022). For computational efficiency, these models, each with different number of hidden layers and hidden neurons, are combined and trained simultaneously in one large MLP. This achieved by connecting the input layer to each of the MLPs, for details see Peerlings et al. (2022). Model selection is based on the MSE loss function imposed on the resulting CDFs.

The MSE loss function calculates the difference between the target CDF $\hat{y}_t$ and the estimated CDF $\tilde{y}_t$ of each model for $t = 1, \dots, T$:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^{T} [\tilde{y}_t(x_{\cdot t}) - \hat{y}_t(x_{\cdot t})]^2 . \tag{5}$$

The target CDF $\hat{y}_t$ approximates the empirical CDF of input variables $x_{1t}, \dots, x_{Nt}$. This empirical CDF target variable can be approximated using grid-based methods over the multivariate input space, following Peerlings et al. (2022). For convenience, both definitions for the univariate and multivariate case are repeated. In case of a single input variable $N = 1$, we define the following grid values: $\gamma = (\gamma_1, \dots, \gamma_G)$ with $\gamma_g < \gamma_{g+1}$ for $g = 1, \dots, G-1$ where $\gamma_1, \dots, \gamma_G$ are evenly spaced, capturing the range of input variables. The target values $\hat{y} = (\hat{y}_1, \dots, \hat{y}_T)'$ are calculated as follows:

$$\gamma_t = \underset{\gamma_g \in \{\gamma_1, \dots, \gamma_G\}}{\arg,\min} \left(x_{1t} - \gamma_g \mid x_{1t} \geq \gamma_g\right) \tag{6}$$

$$\hat{y}_t = \frac{\sum_{\tau=1}^{T} I(x_{1\tau} \leq \gamma_t)}{T}, \tag{7}$$

for $t = 1, \dots, T$ and $I()$ is the indicator function that takes the value 1 if the argument is true and 0 otherwise. This implies for each observation we choose the highest grid value that is smaller than the observation in (6). The empirical CDF of each output point is then calculated using the number of observations below the grid point, as in (7). For the multivariate density problem the specified grid Γ is of dimension $\mathbb{R}^{N \times G}$ where a vector of grid values are constructed for each combination of input variables. For $\Gamma = (\gamma_{\cdot 1}, \dots, \gamma_{\cdot G})$ with $\gamma_{\cdot g} = (\gamma_{1g}, \dots, \gamma_{Ng})'$ for $g = 1, \dots, G$. The target values $\hat{y} = (\hat{y}_1, \dots, \hat{y}_T)'$ are calculated as follows:

$$\gamma_{\cdot t} = \underset{\gamma_{ng} \in \{\gamma_{n1}, \dots, \gamma_{nG}\}}{\arg,\min} \left(x_{nt} - \gamma_{ng} \mid x_{nt} \geq \gamma_{ng}\right) \quad \forall n \tag{8}$$

$$\hat{y}_t = \frac{\sum_{\tau=1}^{T} I(x_{n\tau} \leq \gamma_{nt}, \forall n)}{T}, \tag{9}$$

where the grids are defined for each input variable in (8). The empirical CDF of each output point, is calculated using the number of observations that are included in the N-dimensional grid box given in (9). Note that in (9), the indicator function takes the value of 1 for a certain t', only if $x_{nt'} \leq \gamma_{nt} \forall n$.

After obtaining the estimated empirical CDF $\tilde{y}_t$, the functional form using the selected MLP in equation (4) of this CDF is differentiated:

$$\tilde{y}_t' = \frac{\partial^N}{\partial x_{1t} \cdots \partial x_{Nt}} \tilde{y}_t(x_{\cdot t}), \tag{10}$$

where $\tilde{y}_t$ is composed of non-linear $\sigma(\cdot)$ and linear functions as stated above. Therefore this can be rewritten as nested functions:

$$\tilde{y}_t = g^{[H]} \circ \sigma \circ g^{[H-1]} \circ \sigma \circ \cdots \circ \sigma \circ g^{[1]} \circ \sigma \circ g^{[0]}, \tag{11}$$

where the linear functions take different length vectors for each layer, for the input layer $g^{[0]}(\cdot) : \mathbb{R}^N \to \mathbb{R}^M$, the hidden layers $g^{[q]}(\cdot) : \mathbb{R}^M \to \mathbb{R}^M$ and the output layer $g^{[H]}(\cdot) : \mathbb{R}^M \to \mathbb{R}^1$. Differentiation of the entire MLP is done by treating one composite function at a time starting with $\sigma \circ g^{[0]}$. Then the next composite function $g^{[1]} \circ \sigma \circ g^{[0]}$ is differentiated, where $\sigma \circ g^{[0]}$ is treated as one function, and so fort. For details of the derivation, see Peerlings et al. (2022).

To be able to compare densities, the two sample Kolmogorov-Smirnov (KS) goodness-of-fit test is used, see Kanji (2006). The KS test is a non-parametric test to assess whether two samples are retrieved from the same population. The two sample test uses as input variables two estimated empirical CDFs, where one of them is denoted F_1 and the other is denoted F_2. The test statistic is defined as:

$$D_{T_1,T_2} = \max_{x_{\cdot t}} |F_1(x_{\cdot t}) - F_2(x_{\cdot t})|, \tag{12}$$

where T_1 and T_2 are the sample sizes of two different samples of vehicle counts per minute, $\max_{x_{\cdot t}}$ is the maximum of all distances between two empirical CDFs of vehicle count $x_{\cdot t}$ and $x_{\cdot t}$ ranges between 0 and the maximum vehicle count observed in both samples. The null hypothesis of two samples being retrieved from the same distribution is rejected when this KS statistic D_T is larger than the corresponding critical values. The critical values are retrieved from Table 1 in Massey Jr (1951). For $T > 35$, we have the following values: $1.63/\sqrt{T}$, $1.36/\sqrt{T}$ and $1.22/\sqrt{T}$ for $\alpha = 0.10$, $\alpha = 0.05$ and $\alpha = 0.01$, respectively.

The straightforward benchmark to compare this method with is Kernel Density Estimation (KDE). A comparison between NN and KDE is provided by Peerlings et al. (2022) where it is shown that KDE obtains largest losses compared to NN density estimation around the modes of the distribution in several simulation settings. A detailed comparison with the benchmark is outside the scope of this paper, but we note that the KDE estimates for the road sensor data are worse than those of the proposed PDF estimation method around the modes as well as over the entire domain in the majority of the applications presented in the next subsections.

3 Data Preparation

The data used in this paper are analyzed to estimate the behavior of traffic flow on Dutch highways for regular weekdays as well as extreme days, by constructing PMFs of these DGPs. The data are retrieved from road sensors located on Dutch highways, which are depicted in Figure 3.1. These sensors count the number of passing vehicles per minute. Each sensor also tracks which type of vehicle is passing, i.e. passenger car, van or truck, on each lane of the highway. This paper analyzes data for the leap year 2016. All vehicles are counted, hence no distinctions between different vehicle types are made. See Figure 3.2 for an example of these minute data on a particular day.

Figure 3.1 Road sensors in the Netherlands.
Red dots are road sensors depicted on highways and blue dots are sensors depicted on main roads. The total number of road sensors located on the Dutch highways is around 20,000.

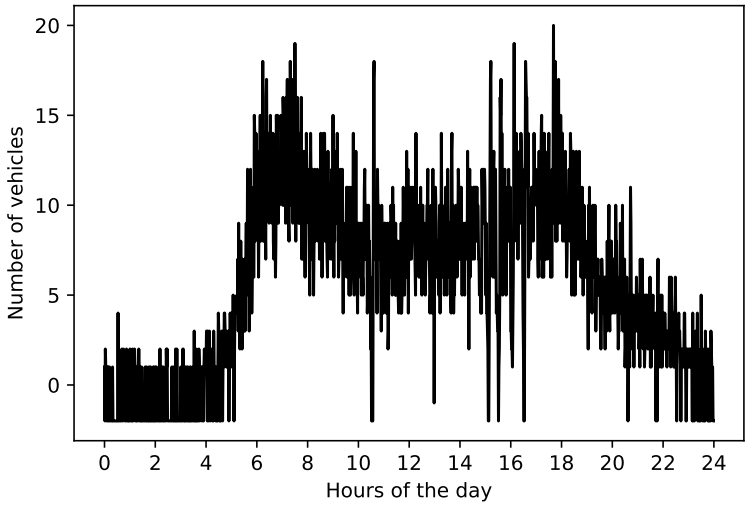


Figure 3.2 Road sensor data example.
The number of vehicle counts passing road sensor ***GEO02_R_RWSTI320***, lane 2 on the 26th of April 2016.

For each daily data, we store the minute data of a specific sensor and a specific lane consisting of 1440 measurements counting the number of passing vehicles. We aggregate these vehicle counts over the lanes. Negative measurements refer to malfunctioning of the sensor and are treated as missing. An exception are the emergency lanes, which might be used by traffic during rush hours. For this lane a negative value indicates that the emergency lane is closed and the negative value should be interpreted as a zero count. Periods with a large number of zero vehicle counts return PMFs that exhibit fat left tails not starting from zero probability values. This issue is circumvented by adding one vehicle count to each measurement. Additionally, holidays are deleted from the sample. Official holidays from the governmental site[1)] are removed, along with 'Sinterklaas' and New Year's Eve. Bridging days or other special days like Mother's Day are not treated as holidays.

A sufficient amount of data must be available to train a neural network. The above-mentioned minute data provides a large number of observations. There is a trade-off between the length of the time window and the amount of data used from the past. For short time windows, we can expect less complicated densities. Therefore a smaller amount of data might be required. On the other hand, short time windows contain a fewer number of minute counts and therefore more days from the past are required to fit a density. This might suppress trends and seasonal fluctuations of traffic behaviour. Larger time windows, on the other hand, contain a larger amount of minute counts. This makes it easier to use only very recent information that better reflect the actual level of the trend and seasonal effect. The complexity of the density, however, will also increase with the length of the time window. At the end of the day it is an empirical

1. https://www.rijksoverheid.nl/onderwerpen/arbeidsovereenkomst-en-cao/vraag-en-antwoord/officiele-feestdagen

question what the right trade-off is between the length of the time window and the amount of data from the past required to fit a density of sufficient quality.

In the following subsections four different applications are presented, which are univariate and bivariate applications with time windows of different length. We consider short time windows with a length of 15 minutes, which is further referred to as the *intra-day approach*. This approach estimates 96 densities independently from each other as this comprises a full day. Each univariate density uses the minute counts of one sensor observed for a time window of 15 minutes. As a result we obtain around 48 Mondays $\times 15$ minutes ≈ 720 observations for the sample size. This is repeated 96 times. We also consider long time windows of 1440 minutes that coincides with a full day, which is further referred to as the *inter-day approach*. One approach is a comparison of the distribution of a sample of regular Mondays with a sample based on a Monday with anomalous traffic flow behavior. The two univariate densities for this approach use 47 Mondays $\times 1440$ minutes $\approx 67,680$ observations for the regular Monday sample size and 1440 observations for the extreme Monday sample, respectively. The third application, which we refer to as the *cross-sectional (inter-day) approach*, uses data from two sensors combining vehicle records of two adjacent sensors with time windows of 1440 minutes. The bivariate density for this approach uses the vehicle counts from two different sensors where each sensor attains approximately 48 Mondays $\times 1440$ minutes $\approx 67,680$ observations. This combination of multiple sources of information has the advantage that this data sample is more appropriate to eliminate outliers and other sources of noise. For training and testing the model to estimate PMFs, vehicle count records are treated as paired observations or tuples, which occur less frequently if one source of information records noise. In this case, the tuple of that measurement is treated as missing and therefore removed. Once the PMF is estimated this information can be used for imputing the missing values of the malfunctioning sensor, conditional on the observations obtained with the other sensor. Our final application uses adjacent 15-minute time windows of the same sensor to observe temporal relations, referred to as the *temporal (intra-day) approach*. Here each tuple is constructed such that the time difference between the two measurements is equal to the chosen time window. In this case we estimate 96 distinct bivariate densities. For each density we use vehicle counts from two adjacent time windows of the same sensor where each time window attains approximately 48 Mondays $\times 15$ minutes ≈ 720 observations. The length of the short time windows of the intra-day and temporal approaches and the long time windows of the inter-day and cross-sectional approaches are kept similar to remain consistent in this application. Different length of time windows may be used in other situations. Longer time windows, e.g., may be used when the sample size is too small.

For the intra-day and inter-day approach vehicle counts per minute from Sensor 1N are used, where N indicates that the sensor monitors the northern driving direction. This is the sensor placed on the right lane of highway A13 heading from Rotterdam to The Hague. For the cross-sectional approach Sensors 1S and 2S, heading from The Hague to Rotterdam, are used (S indicates that these sensors monitor the southern driving direction). For the temporal approach Sensor 1S is used. These sensors are located on the left line of highway A13 heading from The Hague to Rotterdam. See Figure 3.3 for the location of above mentioned sensors depicted by blue flags. We used in particular

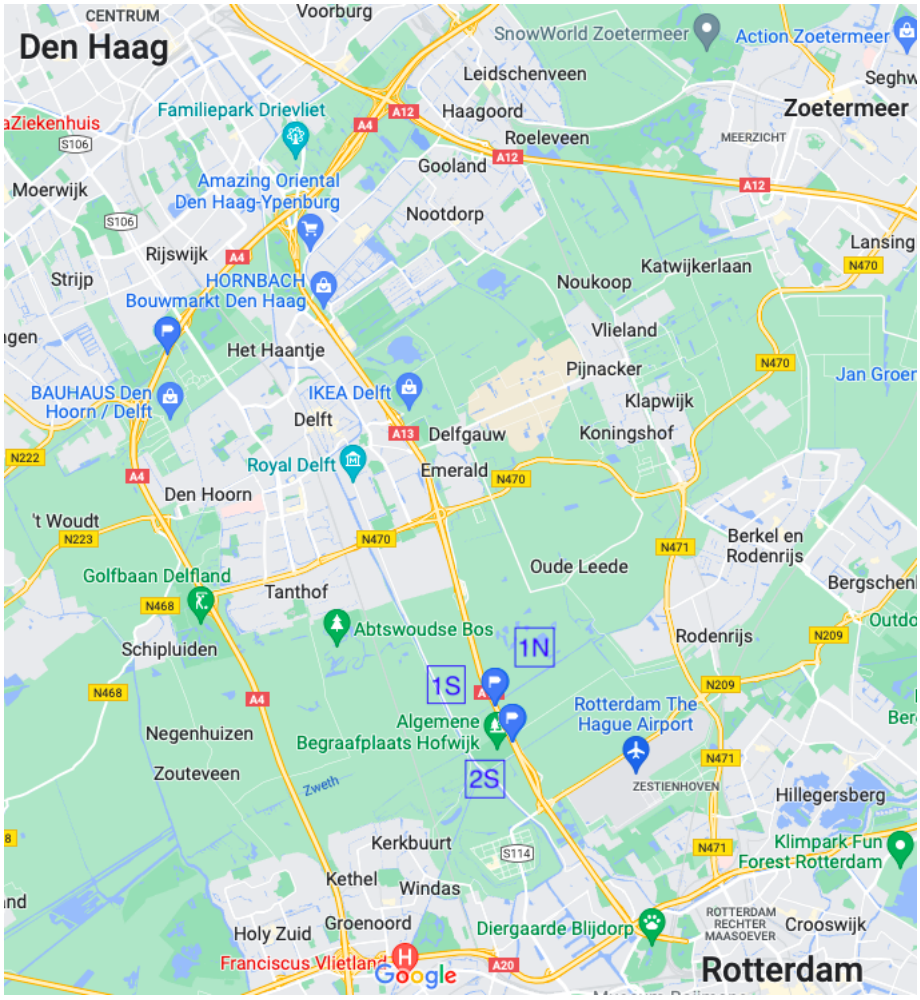


Figure 3.3 The highway A13 between The Hague and Rotterdam with three Sensors 1N, 1S and 2S depicted.
Sensor 1N measures the number of vehicles driving from Rotterdam to The Hague. Sensors 1S and 2S measure the number of vehicles driving in the opposite direction.

these sensors as they are placed in the top 3 of most busy highways according to CBS [2)].

4 Results

In this section we present four applications of neural network density estimation for road sensor traffic flows. These four applications are the intra-day approach, the inter-day approach, the cross-sectional approach and the temporal approach, explained in detail in Section 3. We use the validation sample for model selection in the k-fold learning phase. Then we retrain the the optimal model together with all other models and evaluate these on the test sample in the train-test learning phase. The entire data sample consists of a train sample and a test sample where the train sample is 5 times

2. https://www.cbs.nl/nl-nl/visualisaties/verkeer-en-vervoer/verkeer/rijkswegen

larger than the test sample implicating that each fold of the train sample is approximately equal to the size of the test sample. Firstly, during the k-fold learning phase training is done on 4 folds of the train sample and using the fifth fold as a validation sample. This is done five times, validating on a different fold each time. Subsequently, the average performance for the train sample and the validation sample is calculated. Secondly, during the train-test learning phase training is done using the full train sample and evaluate the models on the test sample. MSE losses are reported for these training, validation and test samples. For the k-fold learning phase the average over the MSE values over the 5 folds used for train and validation samples are reported. Due to missing observations from one or more of the sensors, training, validation and test samples can have slightly different sample sizes. Therefore we use the MSE loss function which is invariant to the sample size.

In all applications, we define a 'typical weekday' as a Monday to avoid the need to model heterogeneity between workdays. Our analysis can be applied to data from other days of the week without loss of generality. We remove the following holidays from the sample, since the behavior of traffic flows is expected to be atypical: 28-03-2016: Easter Monday, 16-05-2016: Whit Monday, 05-12-2016: Saint Nicolas day and 26-12-2016: Boxing Day. This results in analyzing 48 Mondays in the year 2016 instead of 52 Mondays. Model selection is based on the MSE loss of the validation sample. The reported MSE values for the test sample are only used to evaluate the uncertainty of the selected models. The MSEs of the test data provide the most fair and realistic uncertainty approximation, since these data have not been used for training, validation and model selection. The model obtaining the smallest MSE loss for the validation sample is the optimal one for which figures are shown. More specifically, figures of CDF and PMF predictions are given. For the intra-day approach evaluation on the validation and test samples are averages over 96 PMFs. Furthermore, the best, average and worst estimates, based on the MSE loss, of a typical day and night PMF are shown. We use the terms daytime and nighttime as 5:30-20:00 and 20:00-5:30, respectively, regardless of the time of the year or sunshine and sunset times. Since the time windows for the inter-day approach are very large and the data is rather noisy, it is tested to which extent the results are sensitive for randomly splitting the sample in 5 folds during the k-fold learning phase. To this end the above described k-fold learning phase and train-test learning phase are applied to 50 replications where we randomly split the data in train, validation and test samples. This shows the robustness against randomly splitting the sample in 6 sub-samples, 5 sub-samples for the k-fold learning phase and 1 sub-sample for the train-test learning phase. The MSE losses are averaged over these 50 sample replicates. For the cross-sectional and temporal approaches, similar approaches are followed, averages over 50 replications and averages over 96 PMFs of an entire day are given, respectively. Standard errors of the MSE losses are calculated by dividing the standard deviation by the square root of the average sample sizes:

$$\text{SE} = \sqrt{\frac{\frac{1}{z}\sum_{i=1}^{z}(e_i - \overline{e})^2}{\overline{T}}}, \quad \text{with} \quad \overline{e} = \frac{1}{z}\sum_{i=1}^{z} e_i, \tag{13}$$

where z are the 96 time windows for the intra-day approach and the temporal approach or the 50 replications for the inter-day approach and the cross-sectional approach, e_i is the MSE of the PMF estimate of i-th time window/sample split and $\overline{T}$ is the average sample size used for the train, validation and test samples. For the train and validation samples during the k-fold learning phase, e_i is the mean of 5 folds for the PMF estimate

of time window/sample split i.

Training is done using $10,000$ epochs and a batch size as large as the validation and test samples, with early stopping and corresponding patience parameter 250. The logistic output function is used as activation function in all layers. Furthermore, the Adaptive Moment Estimation (Adam) optimizer is used with learning rate 0.005 for the intra-day and inter-day approaches and 0.0025 for the cross-sectional and temporal approaches. A smaller learning rate for the latter two approaches is chosen because these applications are bivariate opposed to the other two approaches that are univariate. We follow Peerlings et al. (2022) in defining an MLP containing sub-MLPs corresponding to different numbers of layers and neurons which are defined in each subsection.

4.1 Intra-day approach

In this Section we use the NN method for density estimation to analyze the vehicle counts of Sensor 1N on Highway A13, going from Rotterdam to The Hague, recorded on Mondays in 2016. We use 15-minute time windows for 48 weekdays over the year which leads to 96 densities for a regular Monday. Hence to construct each density, $15 \times 48 = 720$ measurements are used. On average, each density is reconstructed by a slightly smaller number, 713 measurements due to missing values. The model specifications of the MLP range from 1 to 3 layers and 15 to 50 neurons, as summarized in Table 4.1.

model	1	2	3	4	5	6	7	8	9
# hidden layers H	1	1	1	2	2	2	3	3	3
# hidden neurons M	15	25	50	15	25	50	15	25	50

Table 4.1 Overview of models considered for the intra-day approach and the inter-day approach.

Table 4.2 presents the results for the train samples, validation sample and test sample of Sensor 1N. The mean and standard error of the MSE loss for the train and validation samples during the k-fold learning phase, are calculated as explained by Formula (13). Models 1, 2 and 3, which have a single hidden layer, have generally larger MSE loss values than the larger models in training, validation and test samples. This implies that we need at least 2 hidden layers to estimate these densities. Model 8 with 3 hidden layers and 25 hidden neurons is the best performing model obtaining the smallest MSE loss for the validation sample.

Figures 4.1 and 4.2 present the CDF and PMF estimates for all 96 time windows for the test sample using Model 8 of Sensor 1N. Estimated vehicle count density, reported as CDF and PMF estimates, is clearly changing over time. During the night, from 20:00 to 5:30, the CDF estimates are peaked around measurements of only a few vehicle counts. The CDF estimates from 5:30 to 20:00 obtain relatively high probabilities on a wide range of values around the mode ranging over more vehicle counts. All CDF estimates closely fit the target CDFs, including the tails of the distributions. For space considerations these results are reported in Appendix A, Figure A.1.

Model	1	2	3	4	5	6	7	8	9
Train (5 fold)	11.904	13.340	17.594	11.984	10.894	10.464	11.263	8.785	8.921
	(0.347)	(0.419)	(0.688)	(0.465)	(0.410)	(0.414)	(0.427)	(0.316)	(0.334)
Validation	12.103	13.062	16.843	12.283	10.909	9.990	11.413	8.831	9.388
	(0.718)	(0.850)	(1.252)	(1.020)	(0.913)	(0.814)	(0.946)	(0.694)	(0.921)
Train	12.142	13.454	17.268	12.884	10.888	9.208	10.557	8.877	9.340
	(0.302)	(0.371)	(0.577)	(0.496)	(0.424)	(0.382)	(0.383)	(0.475)	(0.448)
Test	14.807	16.457	20.347	15.635	13.933	13.549	14.996	15.529	14.214
	(0.948)	(1.174)	(1.551)	(1.182)	(1.246)	(1.355)	(1.401)	(1.928)	(1.539)
									Values in the table scaled by 10^{-5}

Table 4.2 MSE losses for the intra-day approach, Mondays 2016 (1N).
The table reports the mean (standard error in parentheses) of the average MSE losses using the difference between the target, and estimated CDF obtained for the train, validation and test samples.

To make a more precise comparison, the differences between the CDF estimates and target CDFs are depicted in Figure 4.3. Largest errors are found throughout daytime, especially during rush hours. The NN tends to over estimate the lower counts in the left tail and under estimate the high counts in the right tail of the CDF. The CDF estimates around the mode of the distributions are alternately larger or smaller than the empirical CDF. In absolute terms, these differences are larger during rush hours, e.g. between 16:00 and 19:00, as well as periods during the day, e.g. between 10:00 and 15:00. For the CDFs during the early morning from 0:00 to 7:00 and in the late evening from 21:00 to 0:00, loss values are relatively smaller obtaining bright coloring. The PMF estimates corresponding to the CDF estimates, depicted in Figure 4.2, are smoother and wider from 5:30 to 20:00 than for the other time windows. During the night only a few vehicles are counted resulting in peaked PMFs. The peaks between 20:00 and 1:00 are smoother than the peaks between 1:00 and 5:30. To investigate this further, the best, average and worst CDF and PMF estimates, in terms of MSE loss, between 5:30 and 20:00 and between 20:00 and 5:30 are analyzed below.

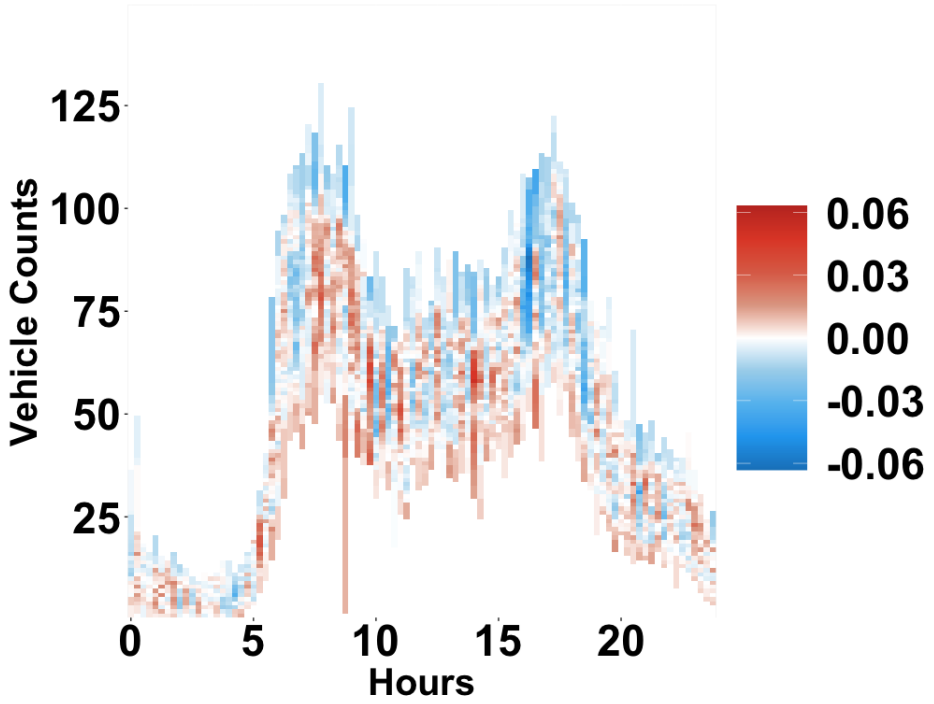
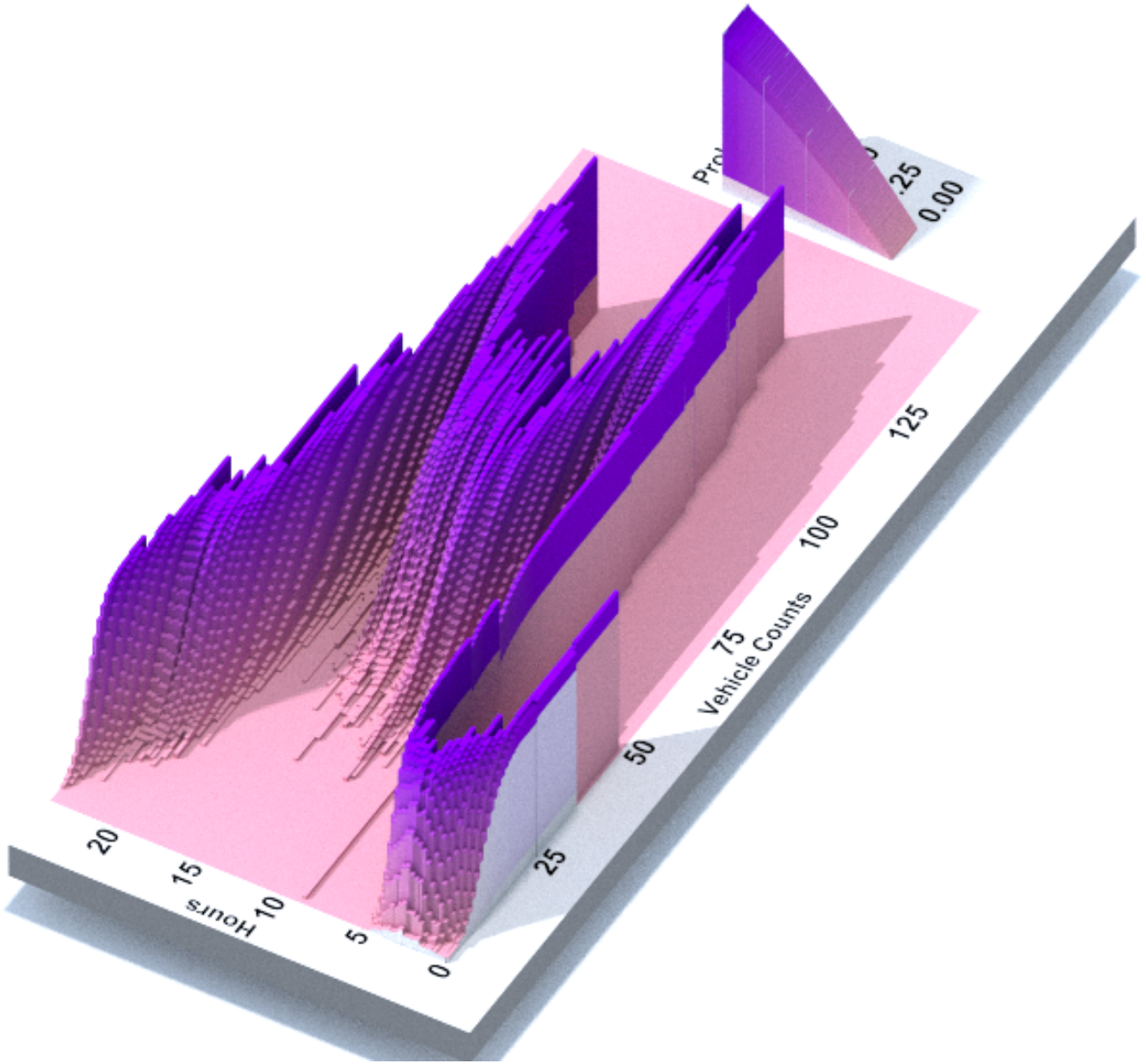


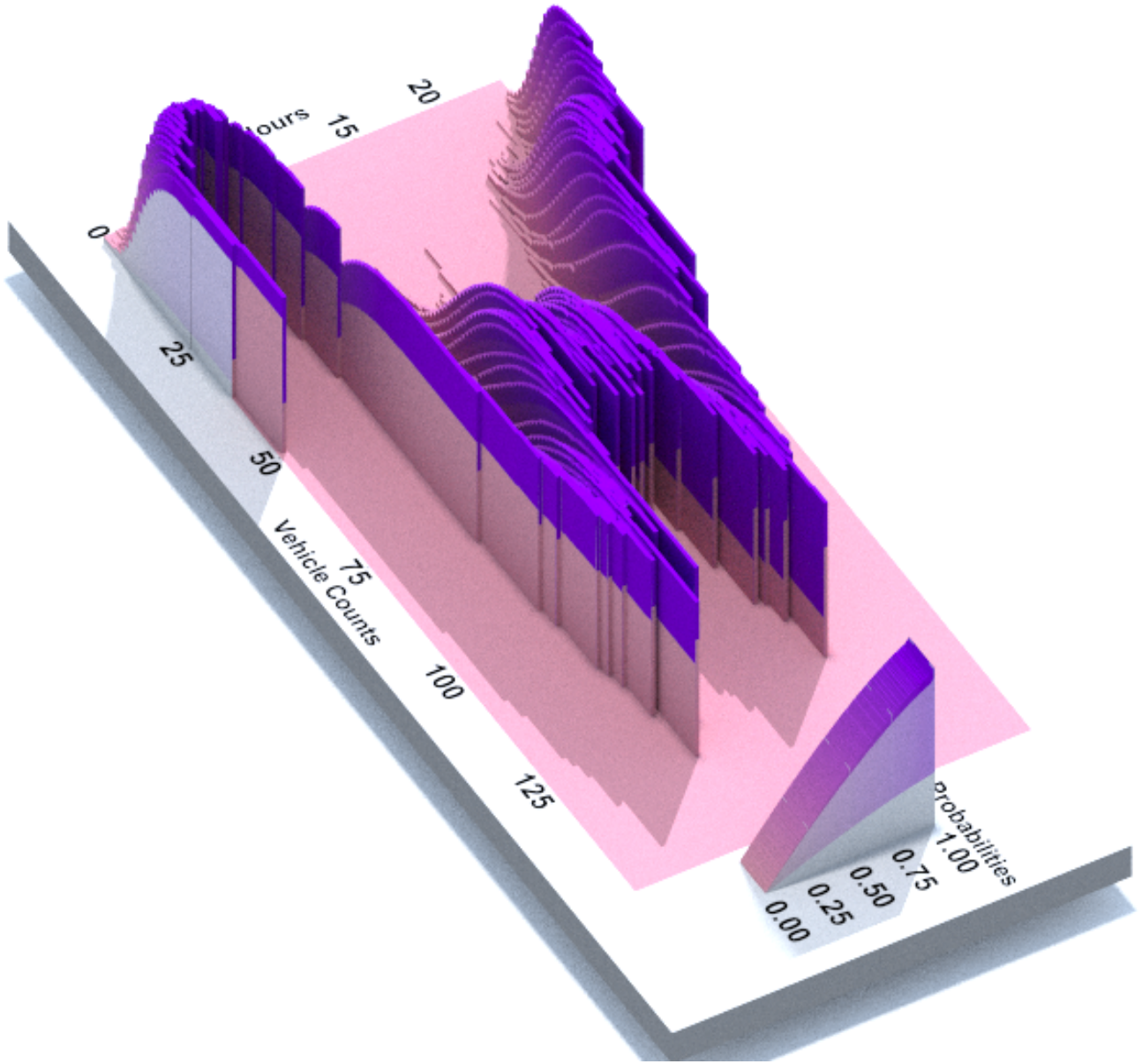
Figure 4.3 The differences between the CDF estimates and target CDFs of the intra-day approach for Sensor 1N.
The x-axis depicts the 96 time windows, the y-axis depicts the number of vehicle counts which is the domain of the CDF estimates and the colorbar returns the values of the differences between the target CDFs and the CDF estimates. The differences are shown for the test sample.

Figure 4.4 present the CDF estimate and CDF target for the best, average, and worst CDF estimates in terms of MSE loss for Model 8 for daytime observations, i.e. between 5:30 and 20:00 for Sensor 1N. In general the CDF estimate and the CDF target are almost identical, even for the worst case. Therefore it is difficult to see the differences between the two, but it implies that CDF estimates obtained with the neural net work are generally very accurate. The corresponding time slots are 18:00-18:15, 18:45-19:00 and 15:45-16:00, respectively. Figure 4.5 presents the corresponding PMF estimates. All CDF



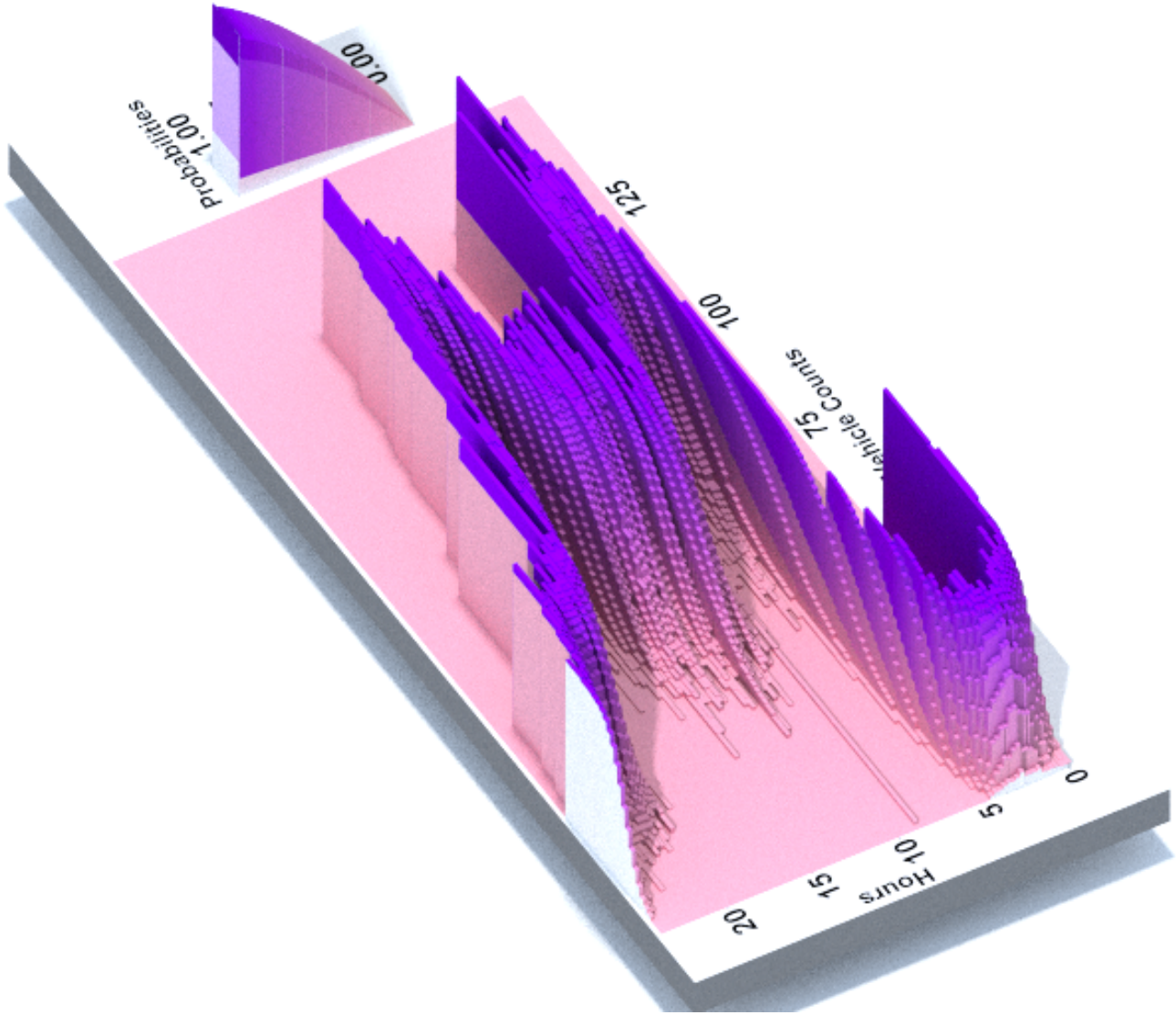
(a) CDF: front angle

Figure 4.1 The CDF estimates of the intra-day approach in different angles for the test sample (1N).



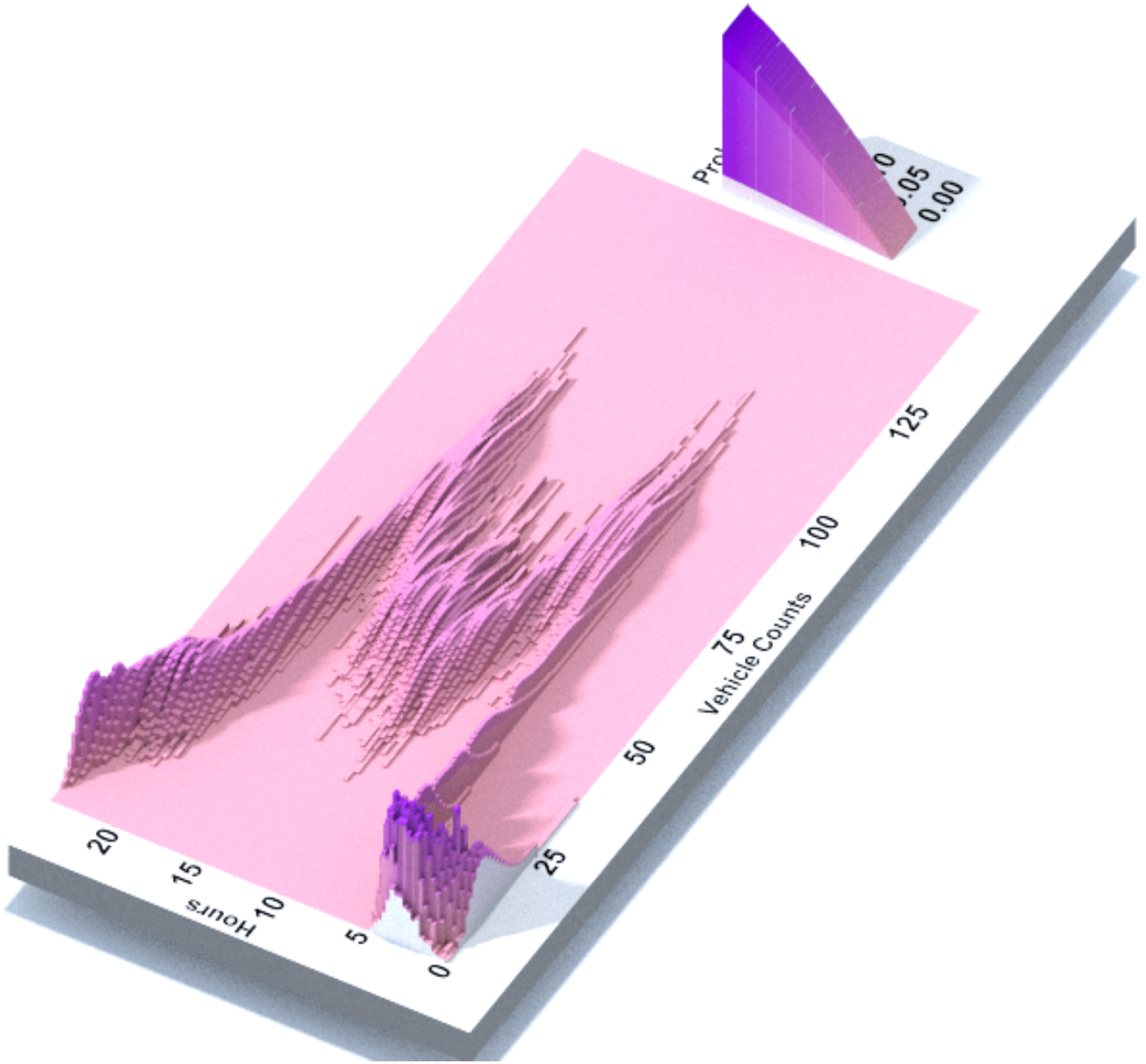
(b) CDF: left angle

Figure 4.1 The CDF estimates of the intra-day approach in different angles for the test sample (1N).



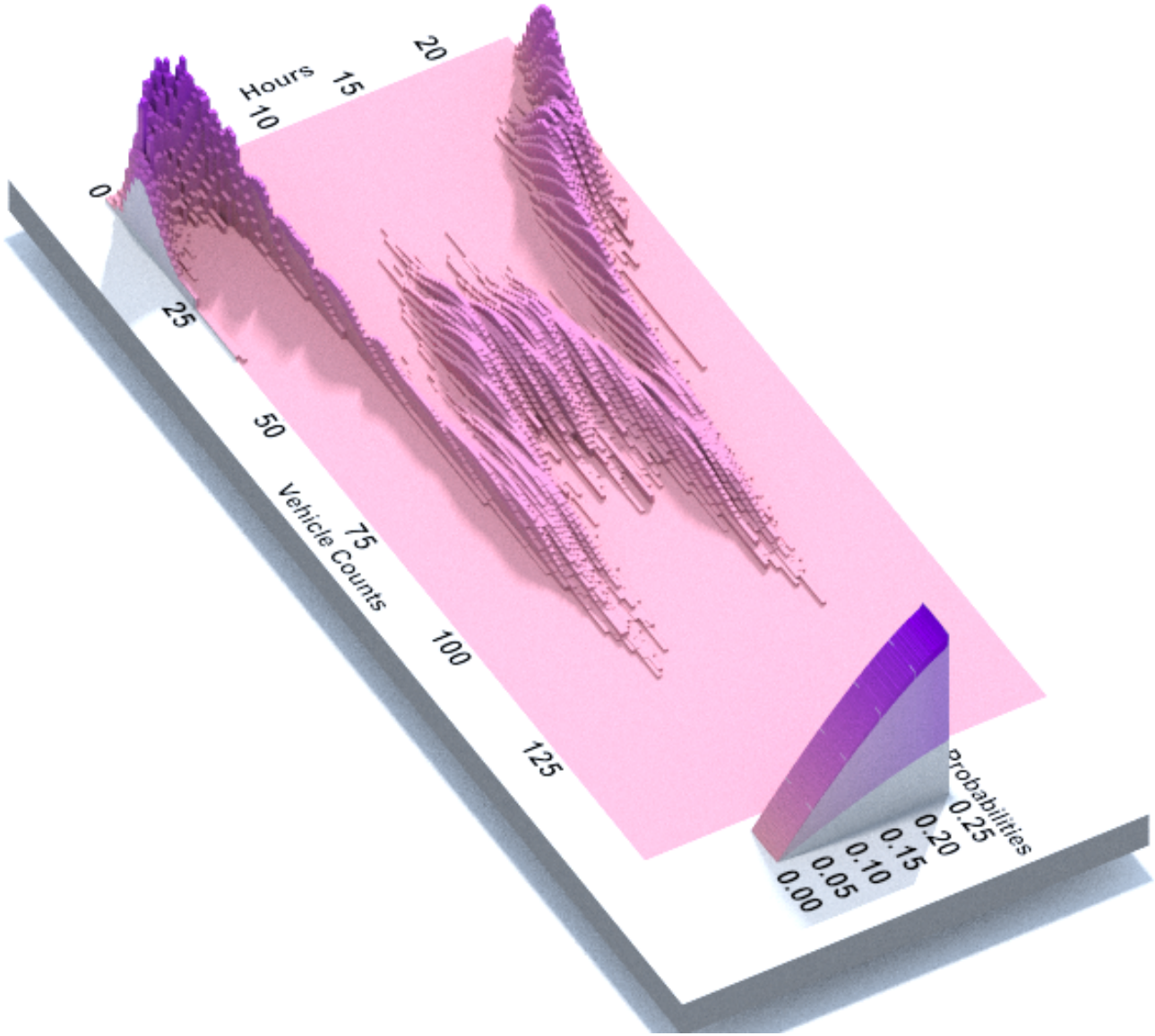
(c) CDF: right angle

Figure 4.1 The CDF estimates of the intra-day approach in different angles for the test sample (1N).



(a) PMF: front angle

Figure 4.2 The PMF estimates of the intra-day approach in different angles for the test sample (1N).



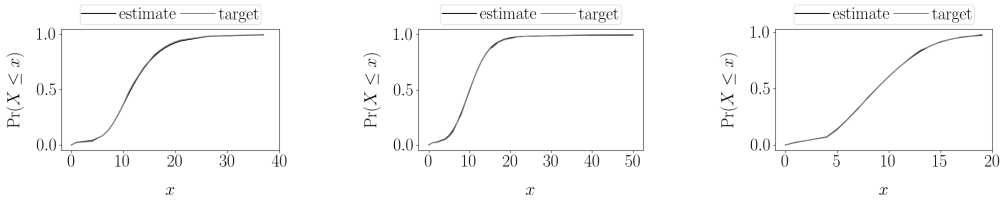
(b) PMF: left angle

Figure 4.2 The PMF estimates of the intra-day approach in different angles for the test sample (1N).

(c) PMF: right angle

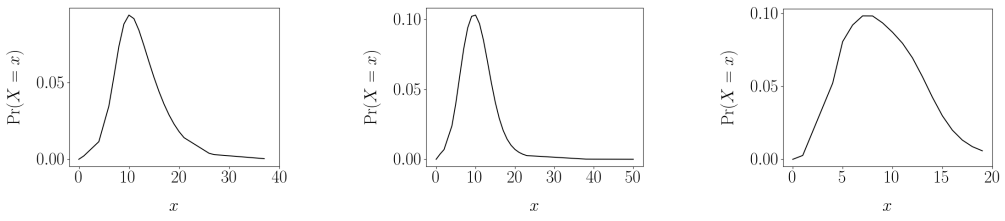
Figure 4.2 The PMF estimates of the intra-day approach in different angles for the test sample (1N).

estimates follow the target CDF closely. The shapes of the target CDFs among the best, average and worst CDF estimates are different. The main difference appears in the large domain area on the right hand side where the CDF values approach 1 and the best and average CDF estimates contain only a few vehicle counts higher than 20. This is also apparent in the corresponding PMF estimates with a relatively long right tail. The worst CDF estimate obtains no vehicle counts larger than 20 at all. Furthermore, this CDF has a kink at vehicle count 5. The corresponding PMF has a wider distribution around the mode than the other PMF estimates. The best PMF estimate, between 18:00-18:15, obtains a mode which is centered around 10 vehicle counts with a fat right tail during rush hour. Around 18:45-19:00 the mode is more tight and concentrated around 10 vehicle counts. Now less vehicles drive on the road but at a usual pace, counting more vehicles per minute than between 18:00-18:15. The PMF between 15:45-16:00 has relatively high probabilities on a wide range of values around the mode, potentially due to the start of the rush hour where in the beginning there is no traffic jam and around 13 vehicles are counted towards the end of the time window where typically a traffic jam starts and less vehicles are counted per minute.



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure 4.4 CDF estimates between 5:30 and 20:00 for the intra-day approach, Mondays 2016 (1N).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 3 hidden layers and 25 neurons are given for the test sample.

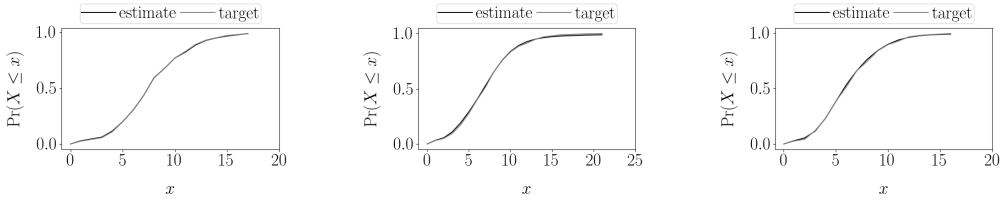


(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure 4.5 PMF estimates between 5:30 and 20:00 for the intra-day approach, Mondays 2016 (1N).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 3 hidden layers and 25 neurons are given for the test sample.

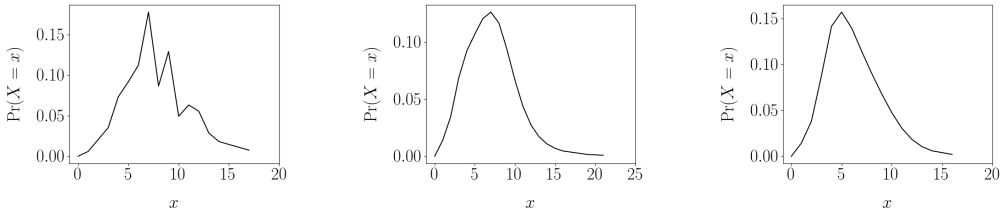
Figure 4.6 present the CDF estimate and CDF target for the best, average, and worst CDF estimates in terms of MSE loss for Model 8 for nighttime observations, i.e. between 20:00 and 5:30 for Sensor 1N. The corresponding time slots are 1:45-2:00, 22:00-22:15 and 20:45-21:00, respectively. In general, these are almost identical and therefore it is difficult to see the differences between the two. Figure 4.7 presents the corresponding PMF estimates. The average CDF estimate contains a few observations in the right tail,

relatively far away from the mode of the distribution. This results in a large domain area on the right hand side where the CDF is close to 1. Furthermore, all CDF estimates are not very smooth due to the small number of different vehicle counts during the night. This is also reflected in the corresponding PMF estimates showing high probabilities on a narrow range of values around the mode forming a peak. All PMF estimates show a skewed fat right tail. The average and worst PMF estimates during the evening between 22:00-22:15 and between 20:45-21:00 have a wider distribution around the mode compared to the more concentrated and tighter distribution around the mode in early morning between 1:45-2:00 as shown before in Figure 4.2. The PMF estimates in the evening are quite similar and very different from the PMF estimate in the early morning. In the early morning between 1:45-2:00, a more concentrated and tighter distribution appears around the mode around vehicle count 8, which are probably trucks driving in the early morning.



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure 4.6 CDF estimates between 20:00 and 5:30 for the intra-day approach, Mondays 2016 (1N).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 3 hidden layers and 25 neurons are given for the test sample.



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure 4.7 PMF estimates between 20:00 and 5:30 for the intra-day approach, Mondays 2016 (1N).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 3 hidden layers and 25 neurons are given for the test sample.

Comparing the day and night PMFs for Sensor 1N, the PMF estimates during the day are more smooth than during the night. This is probably due to the small number of different vehicle counts during the night. During the evening the PMFs are smoother than during the early morning. The lack of different number of vehicle counts does not affect the capability of estimating these PMFs as the best performing PMF in terms of MSE loss is selected during the early morning.

A number of additional results are presented in the appendices. Appendix B shows the analysis of the intra-day approach using more complex MLP models where the

complexity increases with the number of hidden neurons. It is shown that the PMF estimates of more complex models result in more erratic patterns opposed to models with less hidden neurons. This might be a sign of overfitting when complex models are applied to noisy data. In addition, Appendix C presents the sensitivity analysis against different time windows of 10 and 20 minutes in order to show the robustness against slightly different short time windows. As expected, the shorter time window of 10 minutes obtains larger MSE loss values and the larger time window of 20 minutes returns smaller MSE loss values than the 15 minute time window shown in this Section due to the smaller/larger sample sizes.

4.2 Inter-day approach

In this section we apply the NN method to estimate the inter-day density explained in Section 3. We also showcase that the estimated density can be used to assess the differences of vehicle count densities on a regular Monday and on a possibly extreme Monday. Vehicle counts of Sensor 1N on Highway A13, going from Rotterdam to The Hague recorded on Mondays in 2016 are analyzed. The extreme day is the 22^{nd} of August when the metro station in The Hague opened. For this day there are only 865 measurements available. Data for the regular Mondays consist of the entire year of 47 weekdays with 1440 measurements per day, resulting in 67,680 measurements. There are some missing observations during these Mondays which results in 67,583 measurements. The model specifications of the MLP range from 1 to 3 layers and 15 to 50 neurons, as summarized in Table 4.1.

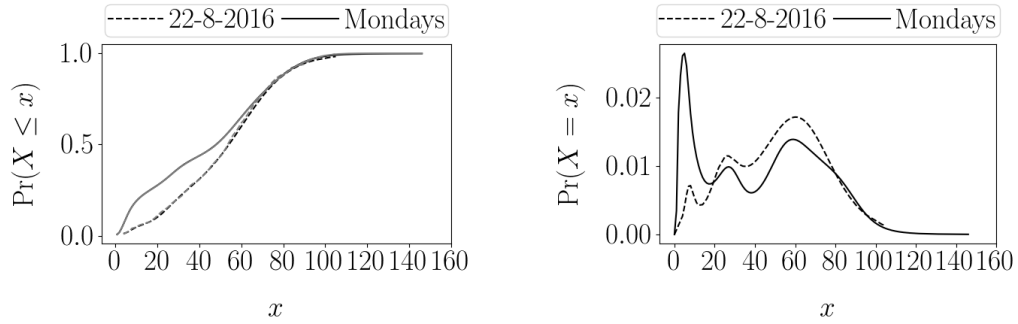
In order to compare a regular and possibly extreme traffic density, we apply the methodology to observations classified as 'regular Mondays' and the 'extreme Monday' separately and compare the estimated densities. For each estimation, the five-fold learning phase and train-test learning phase are applied to 50 replicates that randomly split the data into train, validation and test samples. Table 4.3 contains the average MSE losses over the 50 replicates obtained with the models from Table 4.1 for regular Mondays as well as the extreme Monday dated 22-8-2016. The mean and standard error of the MSE loss for the train and validation samples during the k-fold learning phase, are calculated as explained by Formula (13). For regular Mondays, models with at least two hidden layers have a substantial smaller MSE loss in all samples with optimal Model 5, containing 2 hidden layers and 25 hidden neurons. For the extreme Monday, MSE losses are similar for all models in all samples with optimal Model 1.

The differences between regular Mondays and the extreme Monday can be related to the sample size.As the extreme Monday has a smaller sample size, these MSE losses are larger than for the regular Monday sample. See Figures D.1 and D.2 in Appendix D for the separate MSE losses of optimal Models 1 and 5 over these 50 replicates for both samples. Hence for some of the replicates of the extreme Monday model selection is less robust than that of regular Mondays. The MSE value is relatively high and might result in different model selection results if the selection was based on this replicate only, see Appendix D for more details. Resampling the train, validation and test samples multiple times results in more accurate MSE approximations and makes the procedure more robust against one unfavourable random split of the sample. Furthermore, the MSE values are more affected by randomly splitting the sample in sub-samples when the sample size is smaller.

Model	1	2	3	4	5	6	7	8	9
Regular Mondays 2016 (T ≈ 67680)									
Train (5 folds)	2.259	1.487	3.998	0.646	0.532	0.809	0.599	0.688	0.890
	(0.002)	(0.003)	(0.014)	(0.003)	(0.003)	(0.005)	(0.003)	(0.004)	(0.006)
Validation	2.256	1.391	3.941	0.661	0.594	0.691	0.603	0.740	0.845
	(0.004)	(0.006)	(0.029)	(0.006)	(0.007)	(0.009)	(0.006)	(0.010)	(0.011)
Train	1.978	1.589	4.530	0.557	0.596	1.459	0.443	0.542	0.545
	(0.003)	(0.005)	(0.024)	(0.004)	(0.005)	(0.011)	(0.002)	(0.005)	(0.006)
Test	1.967	1.465	3.813	0.553	0.540	1.542	0.513	0.588	0.864
	(0.006)	(0.009)	(0.042)	(0.008)	(0.008)	(0.026)	(0.005)	(0.012)	(0.020)
Extreme Monday 22-8-2016 (T ≈ 1440)									
Train (5 folds)	3.757	4.787	5.951	3.949	3.944	4.190	4.409	3.877	3.941
	(0.063)	(0.101)	(0.144)	(0.074)	(0.068)	(0.085)	(0.087)	(0.072)	(0.078)
Validation	3.801	4.807	6.024	4.068	4.059	4.377	4.340	3.838	3.849
	(0.153)	(0.229)	(0.333)	(0.175)	(0.166)	(0.214)	(0.194)	(0.160)	(0.178)
Train	4.146	5.284	6.924	4.082	4.564	4.417	5.263	4.750	4.675
	(0.101)	(0.179)	(0.210)	(0.078)	(0.103)	(0.106)	(0.206)	(0.191)	(0.140)
Test	6.074	8.060	9.181	7.183	7.784	7.409	9.120	8.602	7.809
	(0.384)	(0.520)	(0.649)	(0.496)	(0.645)	(0.491)	(0.982)	(0.950)	(0.478)
									Values in the table scaled by 10^{-5}

Table 4.3 MSE losses for the inter-day approach (1N).
The table reports the mean (standard error in parentheses) of the average MSE losses using the difference between the target, and estimated CDF obtained for the 50 replicates that randomly split the sample in the train, validation and test samples.

The CDF and PMF estimates of one randomly chosen replicate are shown for the regular Mondays and the extreme Monday in Figure 4.8. Figure 4.8(a) shows that the CDF estimate follows the target CDF very closely for both cases. On regular Mondays vehicle counts above 100 occur, contrary to the extreme Monday where the vehicle counts don't exceed 100. Furthermore, the empirical CDFs differ substantially between vehicle count 1 and 70. This is reflected in the PMF estimates depicted in Figure 4.8(b). Both PMF estimates are three-modal, but for the regular Monday sample higher probabilities (on a narrow range of values) around the first mode are found than for the extreme Monday sample. This implies that especially small vehicle counts between 1 and 20 occur substantially less on the extreme Monday. The probabilities around the other two modes are slightly higher for the extreme Monday sample compared to the regular Monday sample.



(a) CDF (b) PMF

Figure 4.8 The CDF and PMF estimates of the inter-day approach (1N). (a) The CDF estimates and target CDF of extreme day 22-8-2016 are depicted in dashed black and gray and of regular Mondays in 2016 in solid black and gray. The approximations closely resemble the target CDFs and therefore the difference between these are difficult to detect. (b) The PMF estimates of extreme day 22-8-2016 and regular Mondays in 2016 are depicted in dashed and solid black, respectively. The estimates are shown for the test sample.

In order to formally compare the regular Mondays and the extreme Monday, we apply the two-sample Kolmogorov-Smirnov (KS) test and compare the estimated CDFs. Model 5 of the regular Monday sample and Model 1 of the extreme Monday sample are used to construct the empirical CDFs. The domain values used as input for these empirical CDFs are integers going from the minimum of both samples to the maximum of both samples. From these two empirical CDFs the KS statistic is calculated for all 50 replicates. The average of this KS statistic is 0.157. The critical value at $\alpha = 0.01$ equals $\frac{1.22}{\sqrt{163}} = 0.096$. Hence the null hypothesis of two samples being retrieved from one population is rejected. This implies that the two empirical CDFs are significantly different from each other. A possible explanation of this finding is the opening of the metro station in The Hague heavily affects the distribution of vehicle counts on highway A13 between The Hague and Rotterdam on the 22^{nd} August which results in a different distribution than the distribution of a regular Monday.

4.3 Cross-sectional approach

In this section, we apply the NN density estimation method to cross-sectional, more specifically bivariate, data of vehicle counts. We analyze the inter-day data on 'regular Mondays' where the data are obtained from two separate sensors. This density therefore captures spatial relations between these sensors which can be used to impute missing values of other periods of time where both sensors were operating or other nearby sensors, particularly under the assumption of MCAR.

Vehicle counts of Sensors 1S and 2S on Highway A13 going from The Hague to Rotterdam, recorded for a typical Monday in 2016 are analyzed. Data of the entire year consists of 48 weekdays with 1440 measurements per day, resulting in 69,120 paired observations or tuples. The construction of tuples is explained in Section 3. The pairs of both Sensors 1S and 2S of which at least one measurement is missing are removed. This results in 54,630 tuples on average for training and 9,718 tuples on average for testing. The train and validation samples for k-fold training consists of 43,703 and 10,926 tuples on average, respectively. These are averages over the 50 replicates that split the sample in 6 sub-samples for the train, validation and test samples. In these samples, tuples where at least one measurement is missing are removed. The model specifications of the MLP range from 1 to 3 layers and 25 to 75 neurons, as summarized in Table 4.4.

model	1	2	3	4	5	6	7	8	9
# hidden layers H	1	1	1	2	2	2	3	3	3
# hidden neurons M	25	50	75	25	50	75	25	50	75

Table 4.4 Overview of models considered for the cross-sectional approach and the temporal approach.

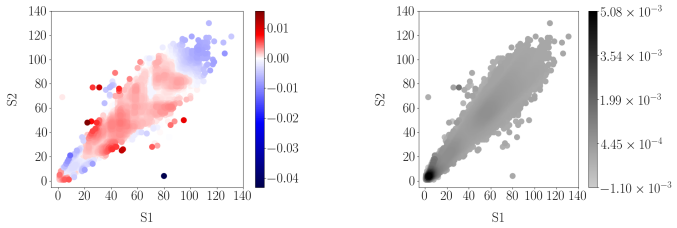
The average MSE losses over 50 replicates are given in Table 4.5. The mean and standard error of the MSE loss for the train and validation samples during the k-fold learning phase, are calculated as explained by Formula (13). Models 1, 2, and 3 with 1 hidden layer obtain a larger average MSE loss than the other models with more hidden layers. Model 5 with 2 hidden layers and 50 hidden neurons has the smallest MSE loss in the validation sample and is selected as the optimal model. See appendix E for the separate MSE loss values for the 50 replicates with Model 5. Most of the MSE values out of these 50 replicates are centered around the average MSE loss of Model 5 reported in table 4.5 with only a few outliers containing higher MSE values. Some of these outliers have a similar MSE value as the average MSE value of the smallest three models 1, 2 and 3, containing only 1 hidden layer, reported in table 4.5. For the train and test samples there are more outliers found than for the k-fold train and validation samples. This illustrates that using only one unfavourable replicate might affect the model selection. This is circumvented by resampling the train, validation and test samples multiple times. The results obtained with Model 5 are further analysed in the remainder of this Subsection.

Figure 4.9 presents the difference between the CDF estimate and CDF target (a) and the PMF estimate (b) for a random replicate concerning Sensors 1S and 2S using Model 5. Further details of the estimated CDF and the target CDF are shown in Figure F.1 in Appendix F. In general, the target and estimated CDFs are very similar. Hence, we present the differences between these values in Figure 4.9. The differences in CDFs are

Model	1	2	3	4	5	6	7	8	9
Train (5 folds)	2.674	3.574	4.679	1.664	0.901	1.211	1.098	1.296	1.237
	(0.003)	(0.011)	(0.017)	(0.004)	(0.004)	(0.006)	(0.004)	(0.006)	(0.007)
Validation	2.750	3.486	4.738	1.722	0.992	1.178	1.078	1.427	1.309
	(0.007)	(0.021)	(0.035)	(0.009)	(0.009)	(0.011)	(0.008)	(0.015)	(0.015)
Train	2.382	2.844	4.225	1.143	1.079	1.415	0.854	1.531	0.792
	(0.004)	(0.010)	(0.020)	(0.004)	(0.009)	(0.009)	(0.003)	(0.010)	(0.005)
Test	2.614	2.791	3.985	1.190	1.052	1.472	1.177	1.330	1.232
	(0.009)	(0.016)	(0.057)	(0.009)	(0.015)	(0.020)	(0.018)	(0.027)	(0.024)
									Values in the table scaled by 10^{-5}

Table 4.5 MSE losses for the cross-sectional approach, Mondays 2016 (1S and 2S).
The table reports the mean (standard error in parentheses) of the average MSE losses using the difference between the target, and estimated CDF obtained for the 50 replicates that randomly split the sample in the train, validation and test samples.

correlated across sensors, with respect to the number of vehicles, indicating that the NN approximation has similar properties for the two sensors. Therefore, we conjecture that the vehicle count densities for the two sensors are similar, hence our method can be used to impute possible missing values in one sensor. For example, the marginal distribution of one sensor conditional on the observed value of the other sensor can be used to report the full density of the missing value or a point estimate such as the mean.



(a) Estimate-Target (b) PMF

Figure 4.9 CDF and PMF estimates, Mondays 2016 (1S and 2S).
(a) difference between estimated and target CDF and (b) PMF estimate. The x-axis and y-axis labels refer to the number of vehicle counts per minute for the sensor of the label. The estimates are shown for the test sample.

Figure 4.9(b) presents the PMF estimates for Sensors 1S and 2S using Model 5. High probabilities on a narrow range of values that form a peak are found around zero vehicle counts. Hence the occurrences of zero or a few vehicles passing Sensor 1S and Sensor 2S are at the same time periods. On the upper right side of this peak the probability mass is concentrated around the diagonal. In general all the measurements are highly correlated as the PMF has an ellipse shape with the peak and the other mode in the diagonal middle of this area.

An observation in Figure 4.9(b) is the occurrences of negative PMF estimates. As explained in Peerlings et al. (2022), the NN density estimation method can lead to negative PMF values due to the non-monotonicity of the corresponding CDF estimate. This problem can be avoided by adding penalties to the NN loss function. Such additions, however, do not necessarily lead to stable estimates. We therefore do not consider such an extension particularly since the negative PMF estimates are very close to 0.We instead report these negative values in more detail: slightly negative values are found with lowest value -1.101×10^{-3} and median value -9.286×10^{-5}. More specifically, 212 out of 9,539 point estimates are negative and we consider this ratio negligible.

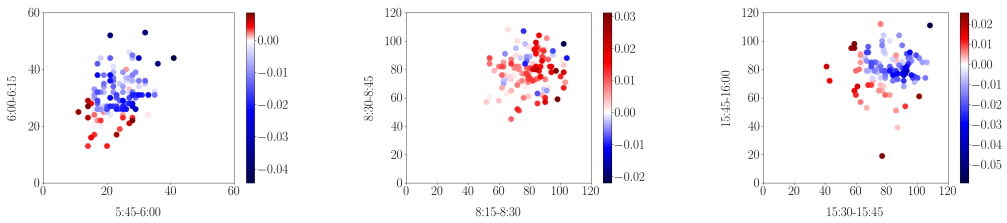
4.4 Temporal approach

In this Section, we apply the NN PMF estimation method in a setting where temporal properties of the data are taken into account. Similar data as for the intra-day approach are used where 15-minute time windows for a typical Monday over the year, that is 96 densities, are estimated. In this application, we estimate a bivariate PMF for consecutive time periods of traffic flow in order to take into account the dependence between consecutive time periods. This approach also implies that missing values in a time period can be imputed using the vehicle counts in earlier time periods. Vehicle counts of Sensor 1S on Highway A13, going from The Hague to Rotterdam recorded for regular Mondays

in 2016 is analyzed. The model specifications of the MLP range from 1 to 3 layers and 25 to 75 neurons, as summarized in Table 4.4.

Table 4.6 presents the results for the train samples, validation sample and test sample for Sensor 1S. The mean and standard error of the MSE loss for the train and validation samples during the k-fold learning phase, are calculated as explained by Formula (13). Model 1 with 1 hidden layer and 25 hidden neurons obtains the smallest loss for the validation sample, closely followed by models 2, 3 and 4. As a result Model 1 is selected and further investigated in the remainder of this Subsection.

Figure 4.10 present the differences between the CDF estimate and CDF target for the best, average, and worst CDF estimates in terms of MSE loss for Model 1 for daytime windows for Sensor 1S. The corresponding time slots are 5:45-6:00 & 6:00-6:15, 8:15-8:30 & 8:30-8:45, and 15:30-15:45 & 15:45-16:00, respectively. Figure 4.11 presents the corresponding PMF estimates. The separate CDF estimates and target CDFs are shown in Appendix G. Figures 4.10(a) and 4.10(c) obtain a large area of negative error values opposed to Figure 4.10(b) that obtains mainly positive error values. Extreme measurements depicted in Figures 4.10(a) and 4.10(c) result in largest absolute error values. In Figure 4.10(b) there are no such extreme measurements. This is possibly explained by the time windows between 8:15-8:30 and 8:30-8:45 around rush hour.



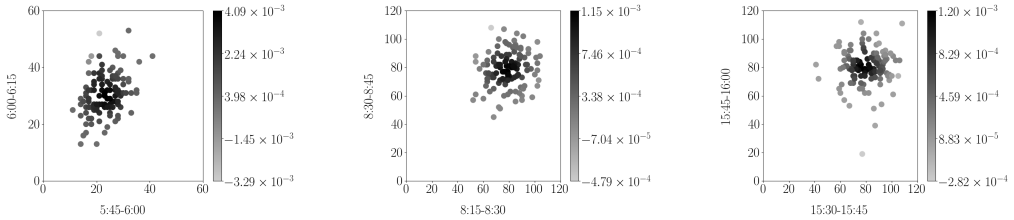
(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure 4.10 CDF estimates between 5:30 and 20:00 for the temporal approach, Mondays 2016 (1S).
The difference between estimated and target CDF of the best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons are given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.

The corresponding PMF estimates have 3, 4 and 5 negative estimates in total with lowest values -3.295×10^{-3}, -4.788×10^{-4} and -2.821×10^{-4} for Figures 4.11(a), 4.11(b) and 4.11(c), respectively. All negative estimates are found in the tails of the vehicle distribution. The corresponding PMF estimates show strong correlation between two time periods for Figure 4.11(a) as the bivariate PMF has an elliptic shape, particularly compared to Figures 4.11(b) and 4.11(c). Furthermore, Figure 4.11(a) has fatter tails opposed to Figures 4.11(b) and 4.11(c).

Model	1	2	3	4	5	6	7	8	9
Train (5 folds)	11.554	11.470	11.923	11.674	13.288	14.110	15.295	14.894	16.485
	(0.118)	(0.180)	(0.181)	(0.174)	(0.268)	(0.361)	(0.393)	(0.380)	(0.476)
Validation	12.605	13.180	13.934	13.416	14.864	16.319	17.459	17.151	19.383
	(0.390)	(0.594)	(0.597)	(0.557)	(0.787)	(1.017)	(1.091)	(1.149)	(1.445)
Train	11.789	12.501	13.626	13.683	14.997	15.298	17.706	14.450	16.196
	(0.294)	(0.363)	(0.470)	(0.517)	(0.617)	(0.606)	(0.813)	(0.640)	(0.659)
Test	17.484	16.406	16.215	17.887	16.273	16.862	20.053	15.199	16.671
	(1.172)	(1.142)	(1.201)	(1.327)	(1.504)	(1.415)	(1.724)	(1.392)	(1.299)
									Values in the table scaled by 10^{-5}

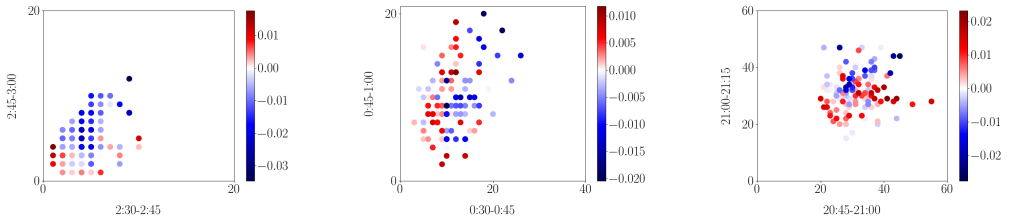
Table 4.6 MSE losses for the temporal approach, Mondays 2016 (1S).
The table reports the average MSE losses using the difference between the target, and estimated CDF obtained for the validation and test samples from data of Sensor 1S using 15 minute windows during the entire day.



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure 4.11 PMF estimates between 5:30 and 20:00 for the temporal approach, Mondays 2016 (1S).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons are given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.

Figure 4.12 presents the differences between the CDF estimate and CDF target for the best, average, and worst CDF estimates in terms of MSE loss for Model 1 for nighttime windows for Sensor 1S. The corresponding time slots are 2:30-2:45 & 2:45-3:00, 0:30-0:45 & 0:45-1:00, and 20:45-21:00 & 21:00-21:15, respectively. Figure 4.13 presents the corresponding PMF estimates. The separate CDF estimates and target CDFs are shown in Appendix G. Figures 4.12(a) and 4.12(b) that present early in the morning time slots measure lower vehicle counts than the time slots during the evening depicted in Figure 4.12(c). There are no extreme measurements depicted, as the measurements are more widespread especially for Figures 4.12(a) and 4.12(b).

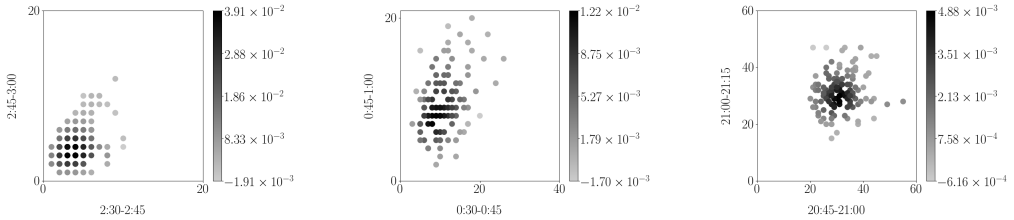


(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure 4.12 CDF estimates between 20:00 and 5:30 for the temporal approach, Mondays 2016 (1S).
The difference between estimated and target CDF of the best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons are given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.

The PMF estimate represented by Figure 4.13(a) has a more horizontal shape for the mode and PMF estimate represented by Figure 4.13(b) has a more vertical shape. Hence between 2:30-2:45 and 2:45-3:00 it is getting more quiet on the road and between 0:30-0:45 and 0:45-1:00 it is getting more crowded. This is probably a reflection on the behavior of truck drivers during the night. Both PMF estimates have a very long slim right-upper tail. The PMF estimate represented by Figure 4.13(c) is quite compact compared to the other two and has a round shape for the mode. This implies that the behavior in the early morning is very different from the evening.

Similar to the earlier applications, it is possible that the PMF estimates obtained by the proposed method lead to some negative values. For the temporal data application, the PMF estimates in Figure 4.13(a-c) have a few negative values. Specifically, we obtain 1, 6 and 2 negative estimates, with lowest values of -1.909×10^{-3}, -1.696×10^{-3} and -6.159×10^{-4} for Figure 4.13, respectively. All negative estimates are found in the tails of the PMF estimates as in the previous applications. This problem can be avoided by adding a penalty term in the loss function of the NN.



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure 4.13 PMF estimates between 20:00 and 5:30 for the temporal approach, Mondays 2016 (1S).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons are given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.

5 Discussion

Road sensors placed on highways to track the number of vehicles per minute are subject to malfunctioning due to various reasons resulting in missing observations and noisy data. Pre-processing is therefore necessary to analyze these data. Discarding the noisy nature or imputing missing values without accounting for the nature of these data can undermine the analyses. Examples are using averages of past observations even though highways are renovated ever since or imputing missing data on an extra-ordinary day with measurements of a regular weekday. We propose to approximate the underlying data generating process (DGP) of traffic flow data by estimating their probability mass functions (PMFs). These PMFs can be used to perform informed (real-time) multiple imputations and forecasts to obtain proper point and uncertainty estimates of raw data or analysis results. This assumes that the data is completely missing at random (MCAR).

To estimate the underlying DGP of traffic flow, this paper proposes an NN approach to account for the noisy and non-standard density characteristics contained in these data. The non-parametric two-step procedure of Peerlings et al. (2022) is applied and generalized. In this procedure, we first estimate the cumulative distribution function (CDF) with a neural network and, as the second step, we differentiate this neural network to obtain the PMF. Four different applications are presented in this paper of which two are univariate PMFs and the other two are bivariate PMFs. In two approaches PMFs of time windows of 15 minutes are estimated for all 96 windows of an entire day. In two other approaches the PMF of a much longer time window of 1440 minutes, i.e. a

full day, is estimated. Depending on the research question, time windows with different lengths can be chosen. The empirical results show that the univariate and bivariate CDF estimates closely resemble the empirical CDFs and can capture a wide variety of densities, since we use a non-parametric approach that does not assume a specific DGP or parametric distribution for the data.

It is proposed to resample the train, validation and test samples multiple times. Using one random replicate of the sample, makes the outcome vulnerable for an accidental unfavourable compositions of the subsamples. Resampling this process instead, results in more accurate average MSE approximations and makes the model selection more robust against accidental unfavourable compositions of the sub-samples. In this application this was in particular important for the large time windows, since they contain large variation of the vehicle counts.

One of our main contributions estimating the complete density of traffic flows by carefully designing NNs which specifically allow for correlations between adjacent road sensors and over time. This design should take into account the implicit assumption that all observations in the training sample of the neural network come from the same distribution. For this reason there is an implicit trade-off between the length of the time window and the amount of data used from the past. This trade-off affects the extent of averaging out seasonal properties and trends, with the assumption that the DGP is the same throughout the specific time intervals across days. Short time windows consist of only a few data points, which implies many days of the past are necessary to form a sufficiently large sample size. Larger time windows, on the contrary, attain many data points which make it unnecessary to use many days in the past. The implicit assumption of the same underlying DGP was also considered to choose a specific day of the week, Monday, for the analysis, to avoid weekly seasonality such as a Friday or weekend effect. In future work, weekdays can be clustered and data for a relatively shorter and longer time windows can be combined to improve the sample size.

So far, bivariate approaches are shown across two different road sensors and two adjacent time windows. A combination of the two, relations between two different road sections of consecutive time windows would be a natural extension to this paper. Moreover, the bivariate approaches may be extended to multivariate approaches which is left for future research.

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the policy of Statistics Netherlands.

Reviewed by Rob Willems and Joep Burger.

References

Chen, M., X. Yu, and Y. Liu (2018). Pcnn: Deep convolutional networks for short-term traffic congestion prediction. *IEEE Transactions on Intelligent Transportation Systems 19*(11), 3550–3559.

Do, L. N., H. L. Vu, B. Q. Vo, Z. Liu, and D. Phung (2019). An effective spatial-temporal attention based neural network for traffic flow prediction. *Transportation research part C: emerging technologies 108*, 12–28.

Duan, P., G. Mao, W. Liang, and D. Zhang (2018). A unified spatio-temporal model for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems 20*(9), 3212–3223.

Ermagun, A. and D. Levinson (2018). Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews 38*(6), 786–814.

Fang, W., W. Zhuo, J. Yan, Y. Song, D. Jiang, and T. Zhou (2022). Attention meets long short-term memory: A deep learning network for traffic flow forecasting. *Physica A: Statistical Mechanics and its Applications 587*, 126485.

Fusco, G., C. Colombaroni, and N. Isaenko (2016). Short-term speed predictions exploiting big data on large urban road networks. *Transportation Research Part C: Emerging Technologies 73*, 183–201.

Habtemichael, F. G. and M. Cetin (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation research Part C: emerging technologies 66*, 61–78.

Kanji, G. K. (2006). *100 statistical tests*. Sage.

Lippi, M., M. Bertini, and P. Frasconi (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems 14*(2), 871–882.

Ma, D., X. Song, and P. Li (2020). Daily traffic flow forecasting through a contextual convolutional recurrent neural network modeling inter-and intra-day traffic patterns. *IEEE Transactions on Intelligent Transportation Systems 22*(5), 2627–2636.

Ma, T., C. Antoniou, and T. Toledo (2020). Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. *Transportation Research Part C: Emerging Technologies 111*, 352–372.

Magdon-Ismail, M. and A. Atiya (2002). Density estimation and random variate generation using multilayer networks. *IEEE transactions on neural networks 13*(3), 497–520.

Manibardo, E. L., I. Laña, and J. Del Ser (2021). Deep learning for road traffic forecasting: Does it make a difference? *IEEE Transactions on Intelligent Transportation Systems*.

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association 46*(253), 68–78.

Peerlings, D. E. W., J. A. van den Brakel, N. Baştürk, and M. J. Puts (2022). Multivariate density estimation by neural networks. *IEEE Transactions on Neural Networks and Learning Systems*.

Polson, N. G. and V. O. Sokolov (2017). Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies 79*, 1–17.

Qu, L., W. Li, W. Li, D. Ma, and Y. Wang (2019). Daily long-term traffic flow forecasting based on a deep neural network. *Expert Systems with applications 121*, 304–312.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American statistical Association 91*(434), 473–489.

Tak, S., S. Woo, and H. Yeo (2016). Data-driven imputation method for traffic data in sectional units of road links. *IEEE Transactions on Intelligent Transportation Systems 17*(6), 1762–1771.

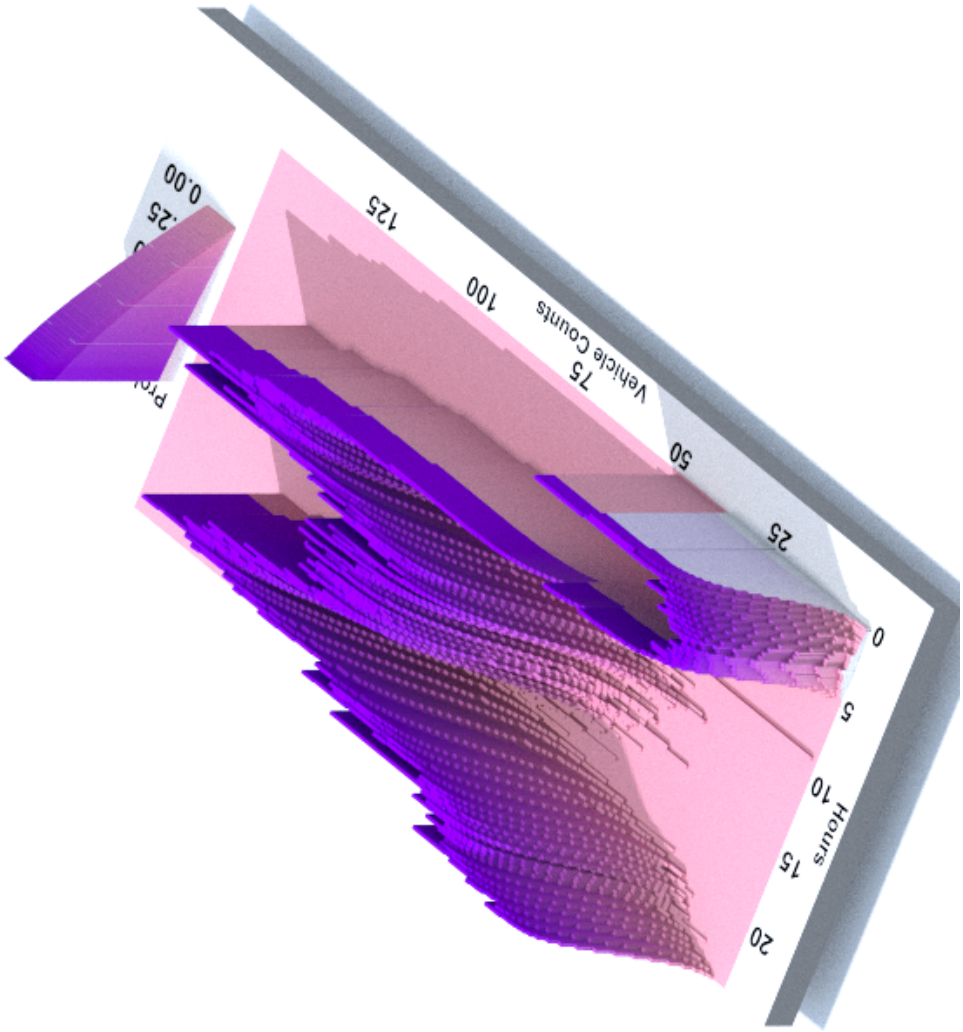
Trentin, E., L. Lusnig, and F. Cavalli (2018). Parzen neural networks: Fundamentals, properties, and an application to forensic anthropology. *Neural Networks 97*, 137–151.

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

Yao, H., X. Tang, H. Wei, G. Zheng, and Z. Li (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence*, Volume 33, pp. 5668–5675.

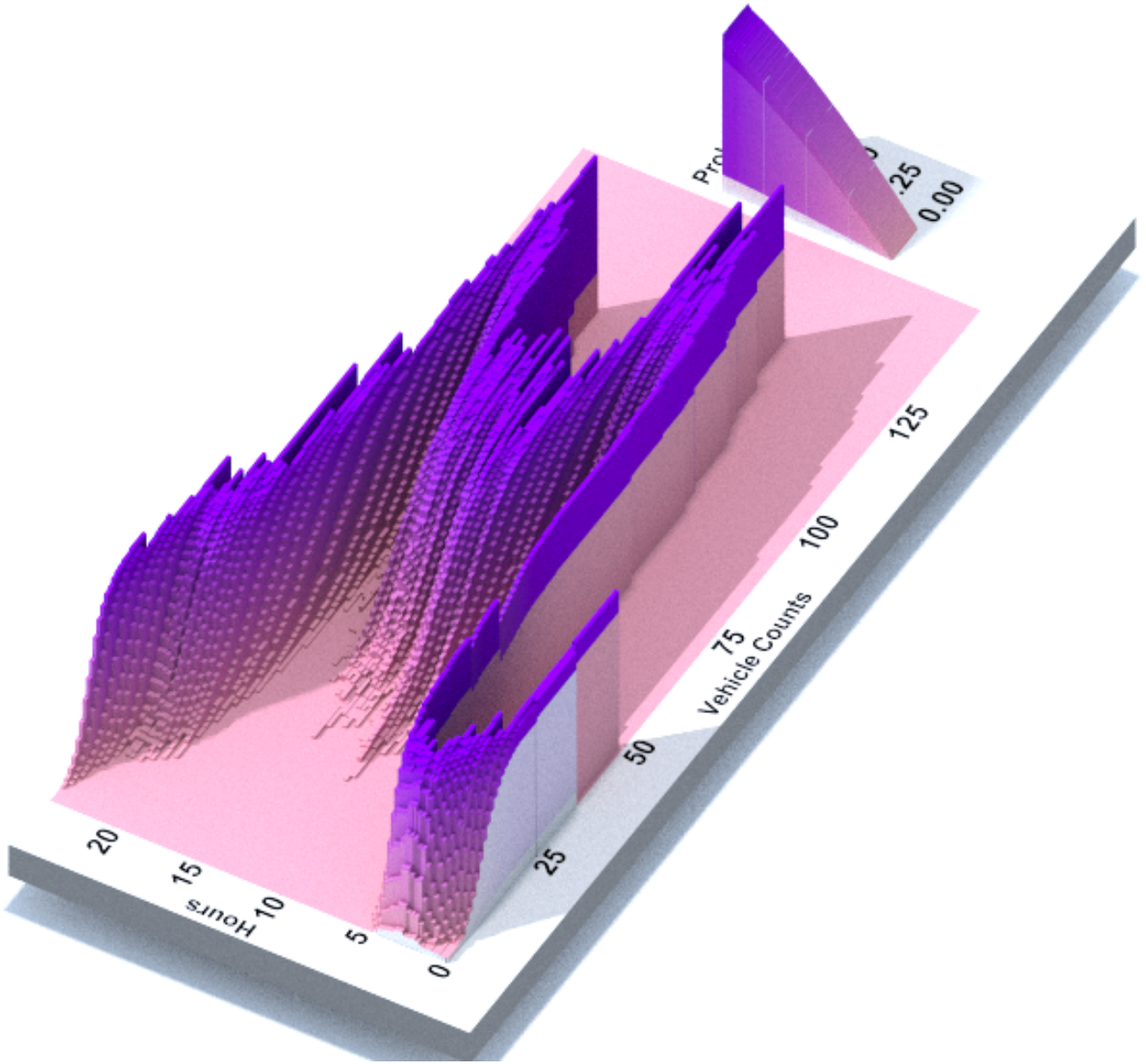
Appendix

A CDF estimates and targets of the intra-day approach



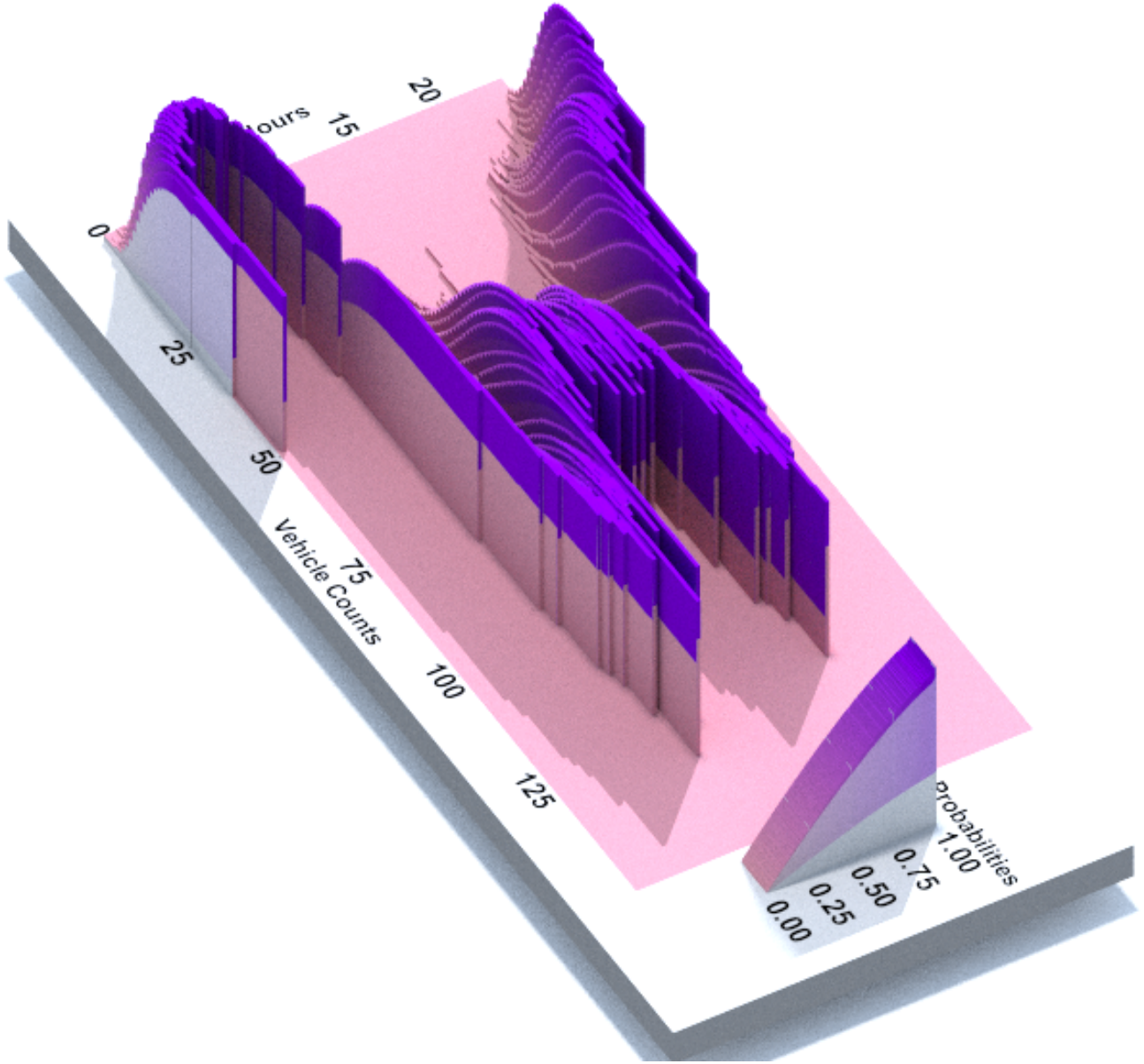
(a) CDF: front angle

Figure A.1 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1S).



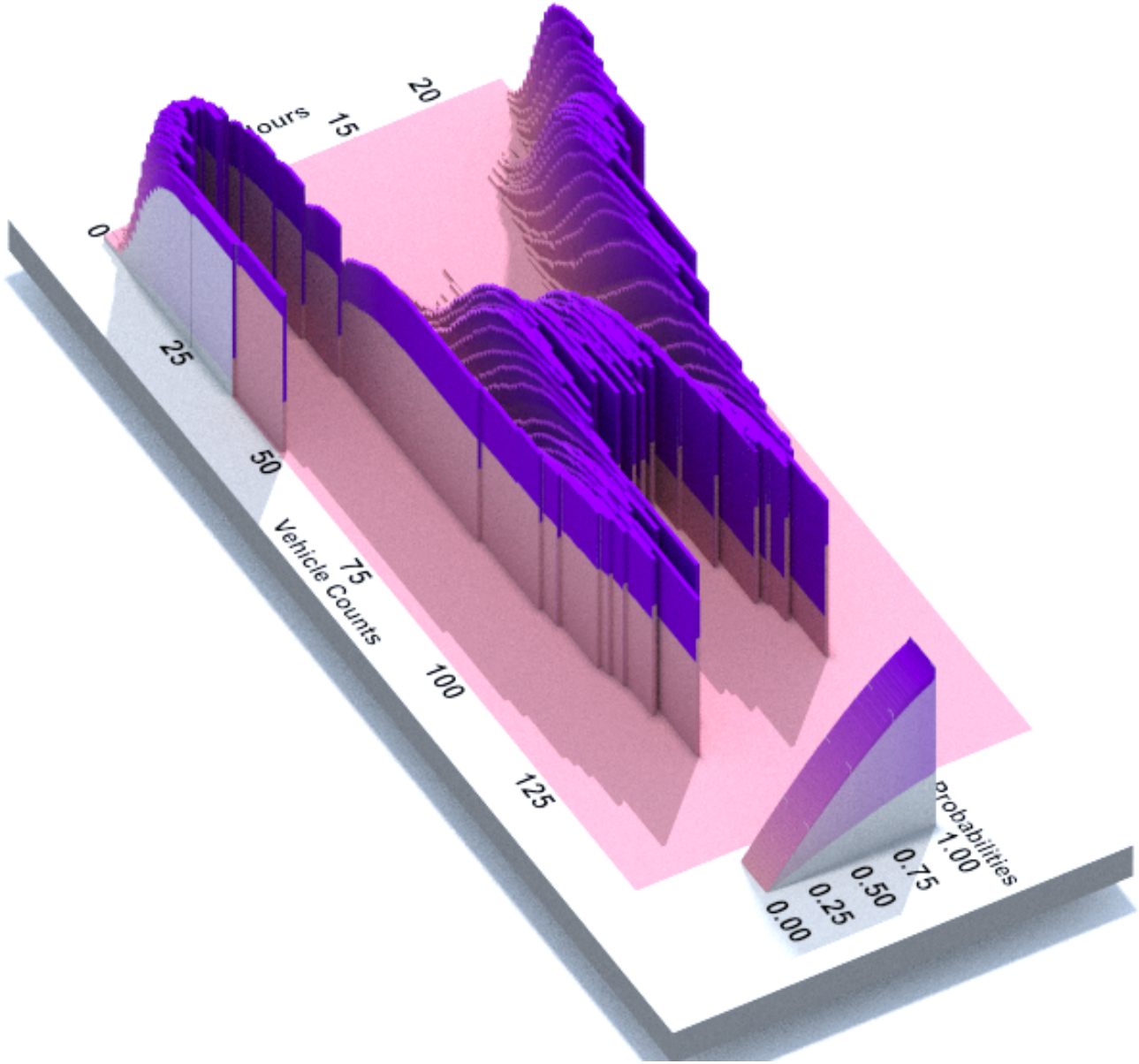
(b) CDF target: front angle

Figure A.1 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1S).



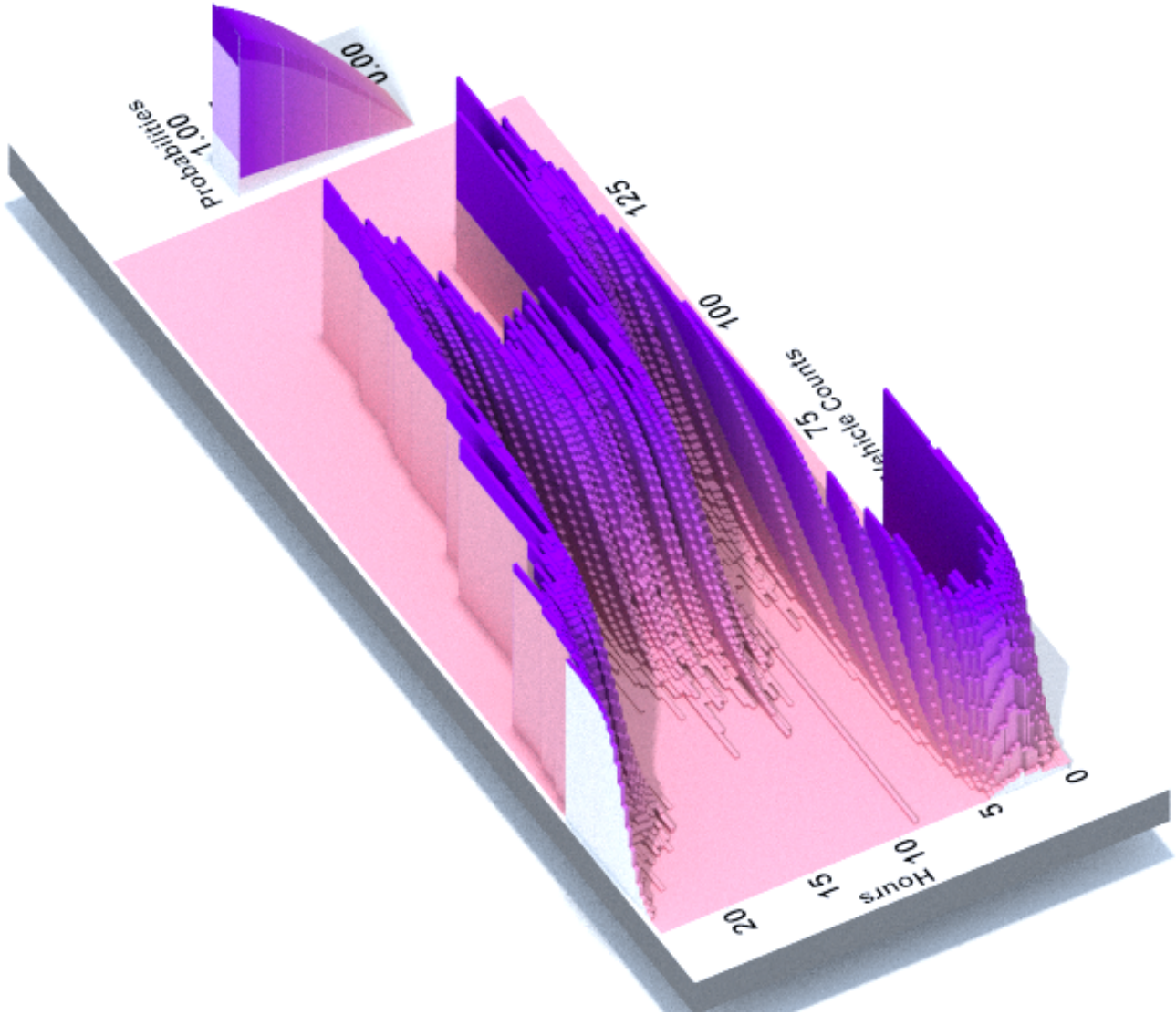
(c) CDF: left angle

Figure A.1 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1S).



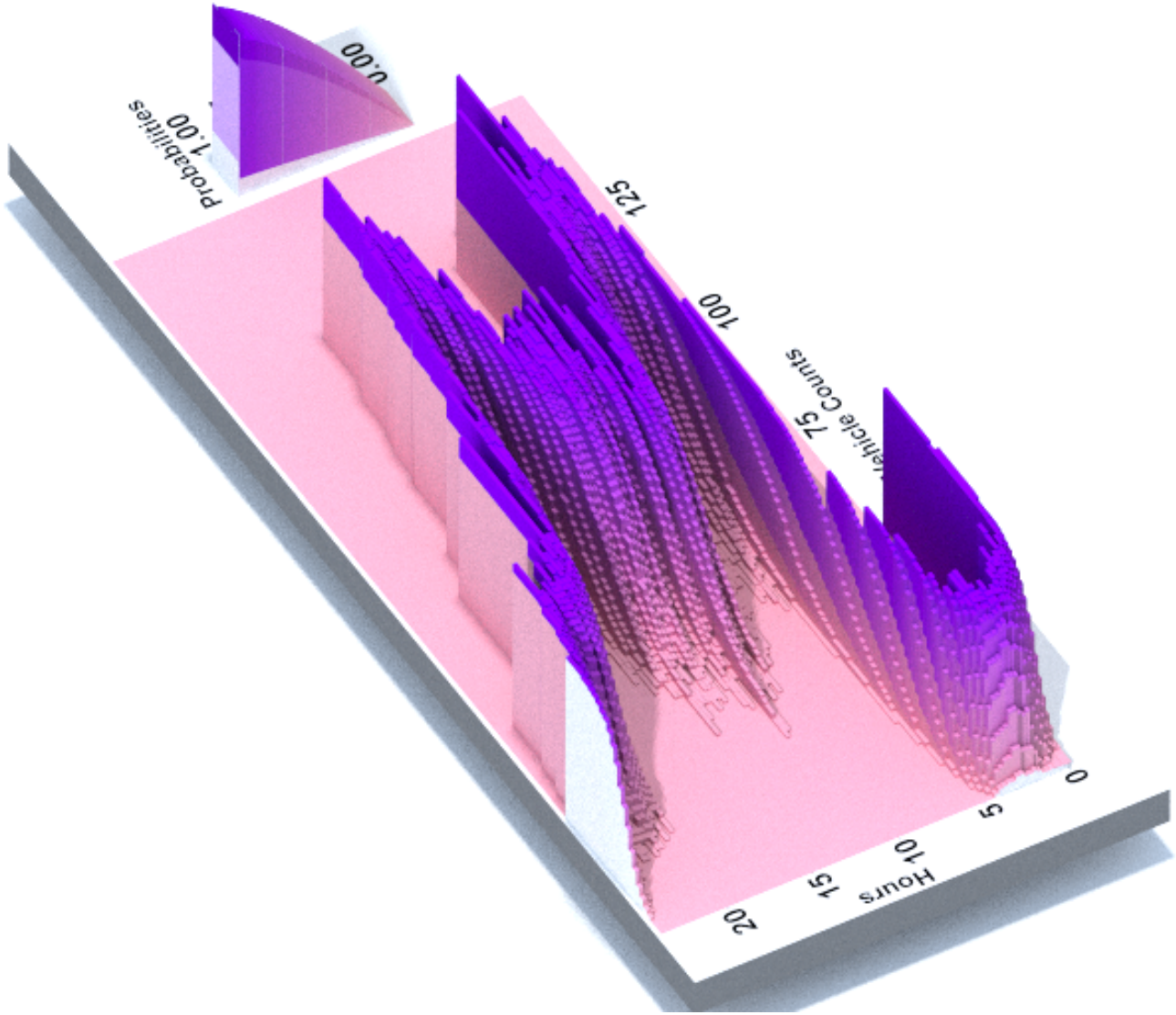
(d) CDF target: left angle

Figure A.1 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1S).



(e) CDF: right angle

Figure A.1 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1S).



(f) CDF target: right angle

Figure A.1 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1S).

B Sensitivity analysis of model complexity for the intra-day approach

In this Appendix more complex MLP models are applied to the intra-day data analysed in Subsection 4.1. The complexity is increased by the larger number of neurons. All other settings are exactly the same as in Subsection 4.1. The purpose of this analysis is to illustrate that too complex models tend to overfit the noisy data and return more ill-shaped PMFs. The model specifications of the MLP range from 1 to 3 layers and 30 to 150 neurons, as summarized in Table B.1.

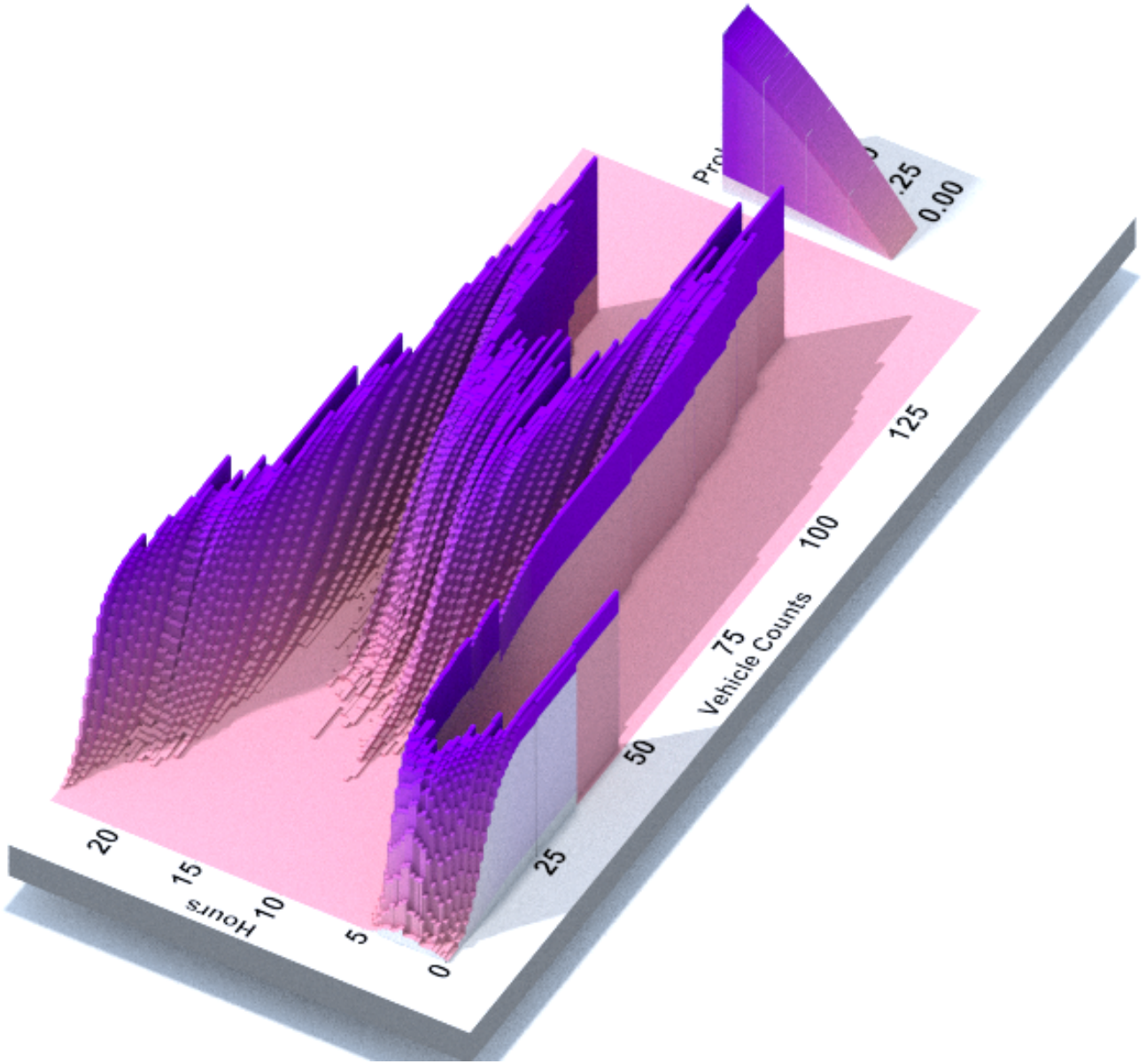
model	1	2	3	4	5	6	7	8	9
# hidden layers H	1	1	1	2	2	2	3	3	3
# hidden neurons M	30	90	150	30	90	150	30	90	150

Table B.1 Overview of more complex models considered for the intra-day approach.

Table B.2 presents the results for the train samples, validation samples and test samples of Sensor 1N. The mean and standard error of the MSE loss for the train and validation samples during the k-fold learning phase, are calculated as explained by Formula (13). Models 1, 2 and 3, which have a single hidden layer, have larger MSE loss values than the larger models. This implies that we need at least 2 hidden layers to estimate these densities. Model 6 with 2 hidden layers and 150 hidden neurons is the best performing model obtaining the smallest MSE loss for the validation sample.

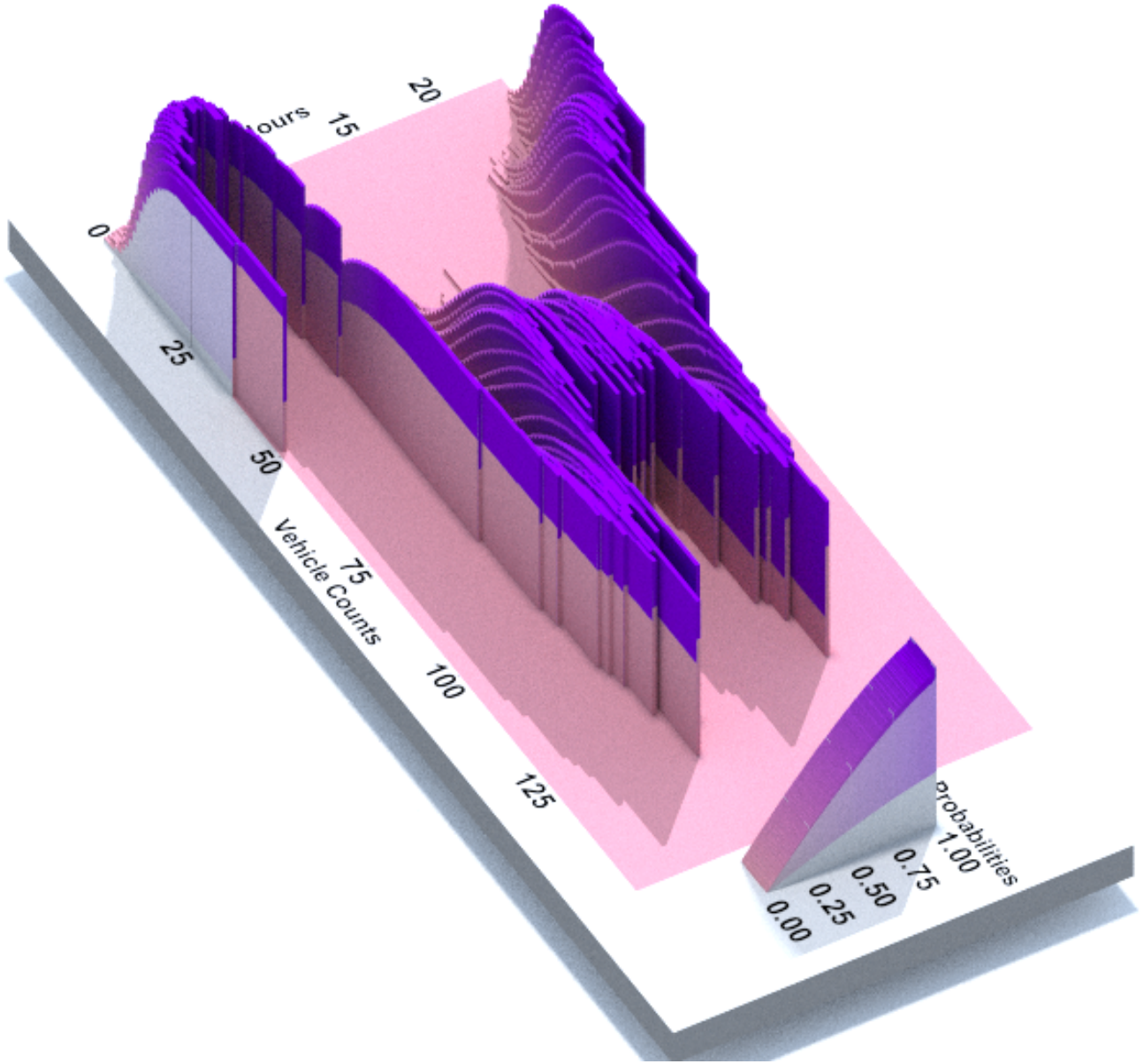
Model	1	2	3	4	5	6	7	8	9
Train (5 fold)	15.088	22.589	26.504	14.807	10.032	8.664	9.588	9.379	10.570
	(0.478)	(1.049)	(1.055)	(0.555)	(0.401)	(0.350)	(0.362)	(0.382)	(0.587)
Validation	15.047	22.686	25.186	14.871	10.103	8.529	9.837	9.413	8.840
	(0.971)	(2.630)	(2.006)	(1.131)	(0.854)	(0.728)	(0.808)	(0.909)	(0.810)
Train	15.529	26.070	27.401	13.384	8.708	8.025	9.890	8.371	7.750
	(0.562)	(1.273)	(1.084)	(0.464)	(0.351)	(0.323)	(0.353)	(0.396)	(0.395)
Test	18.100	24.648	26.162	16.056	12.374	12.867	13.617	13.511	15.024
	(1.135)	(2.138)	(2.238)	(1.110)	(1.119)	(1.292)	(1.174)	(1.225)	(1.614)
									Values in the table scaled by 10^{-5}

Table B.2 MSE losses for the intra-day approach, Mondays 2016 (1N).
The table reports the mean (standard error in parentheses) of the average MSE losses using the difference between the target, and estimated CDF obtained for the train, validation and test samples.



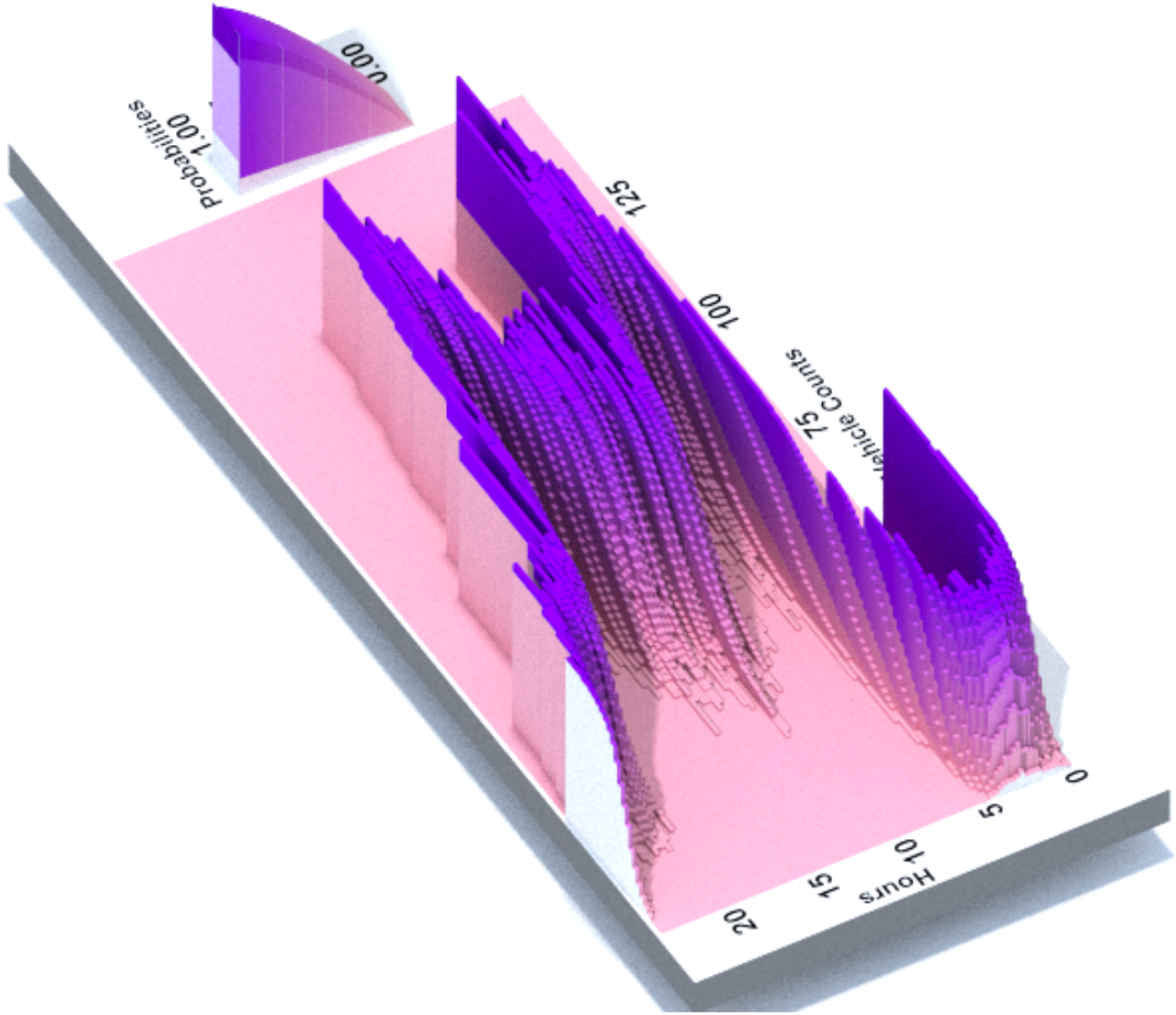
(a) CDF: front angle

Figure B.1 The CDF estimates of the intra-day approach in different angles for the test sample (1N).



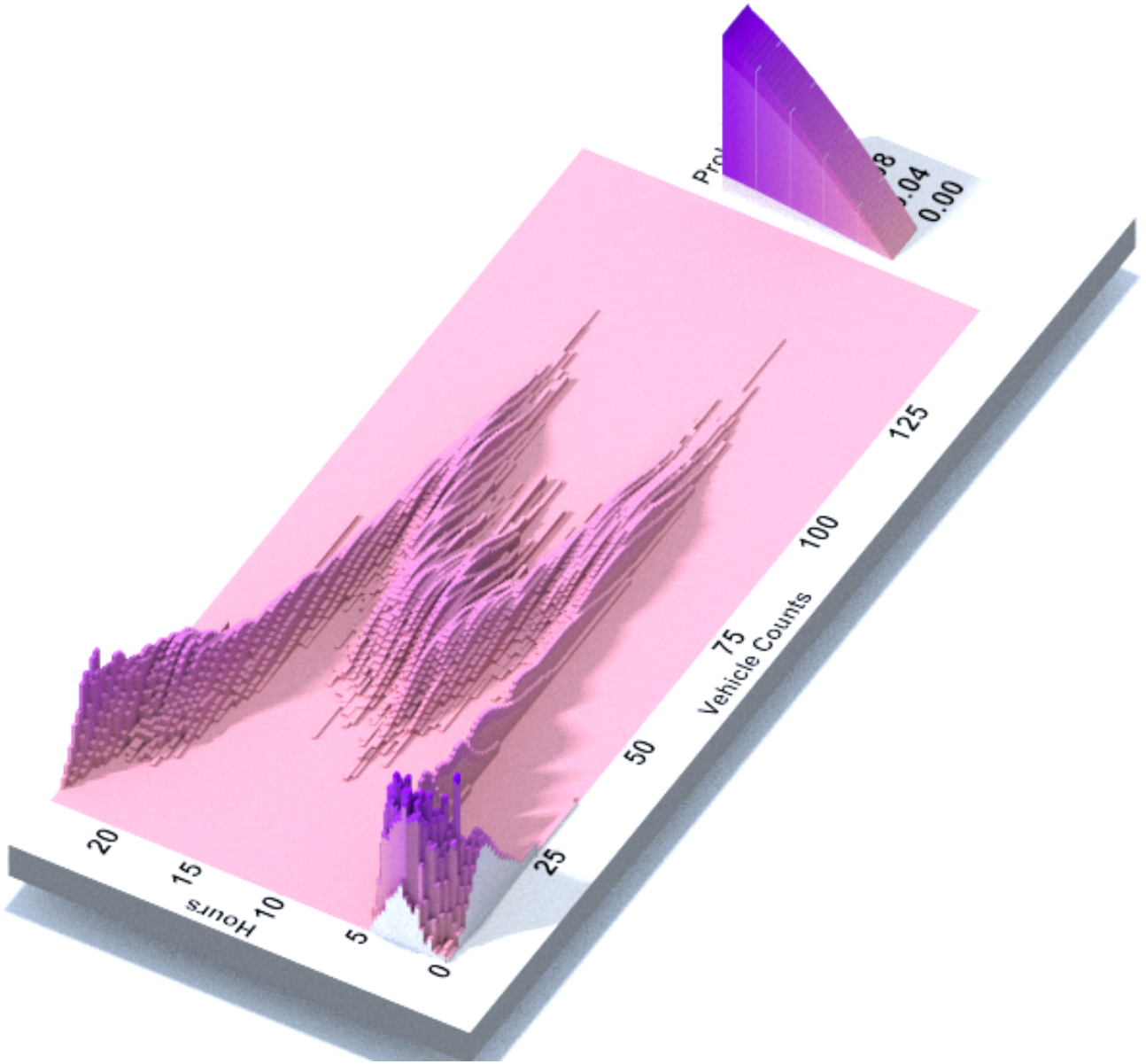
(b) CDF: left angle

Figure B.1 The CDF estimates of the intra-day approach in different angles for the test sample (1N).



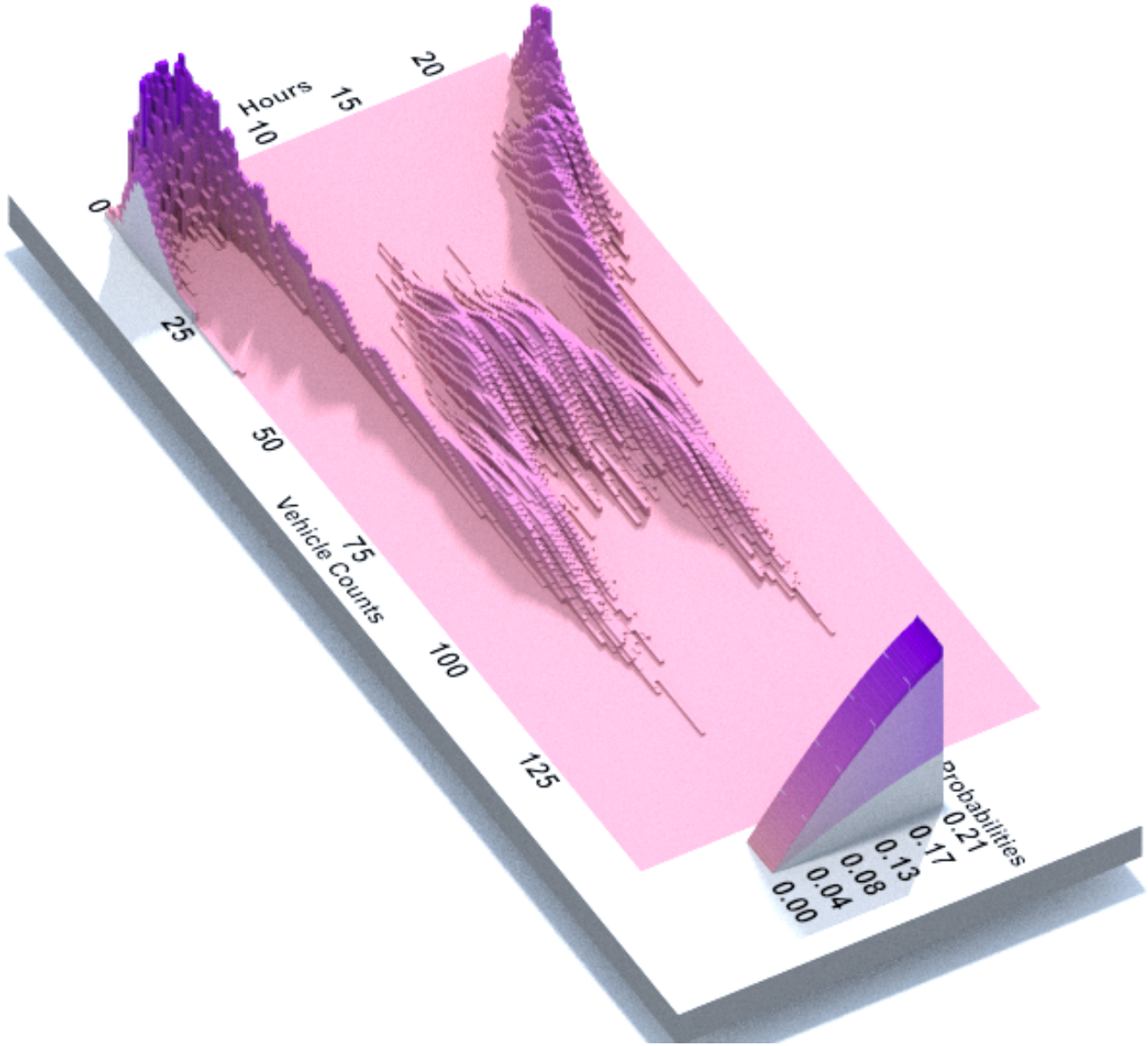
(c) CDF: right angle

Figure B.1 The CDF estimates of the intra-day approach in different angles for the test sample (1N).



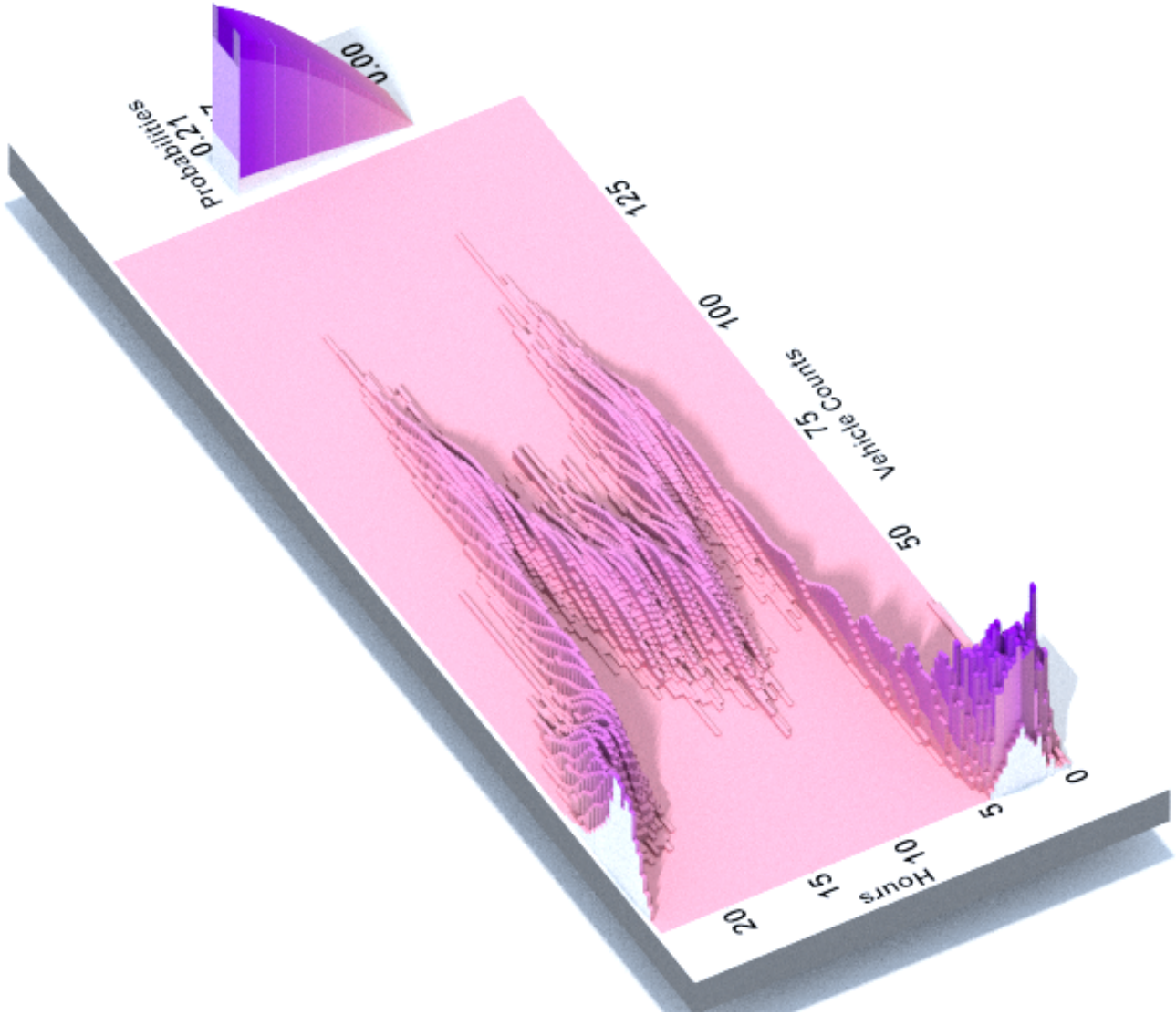
(a) PMF: front angle

Figure B.2 The PMF estimates of the intra-day approach in different angles for the test sample (1N).



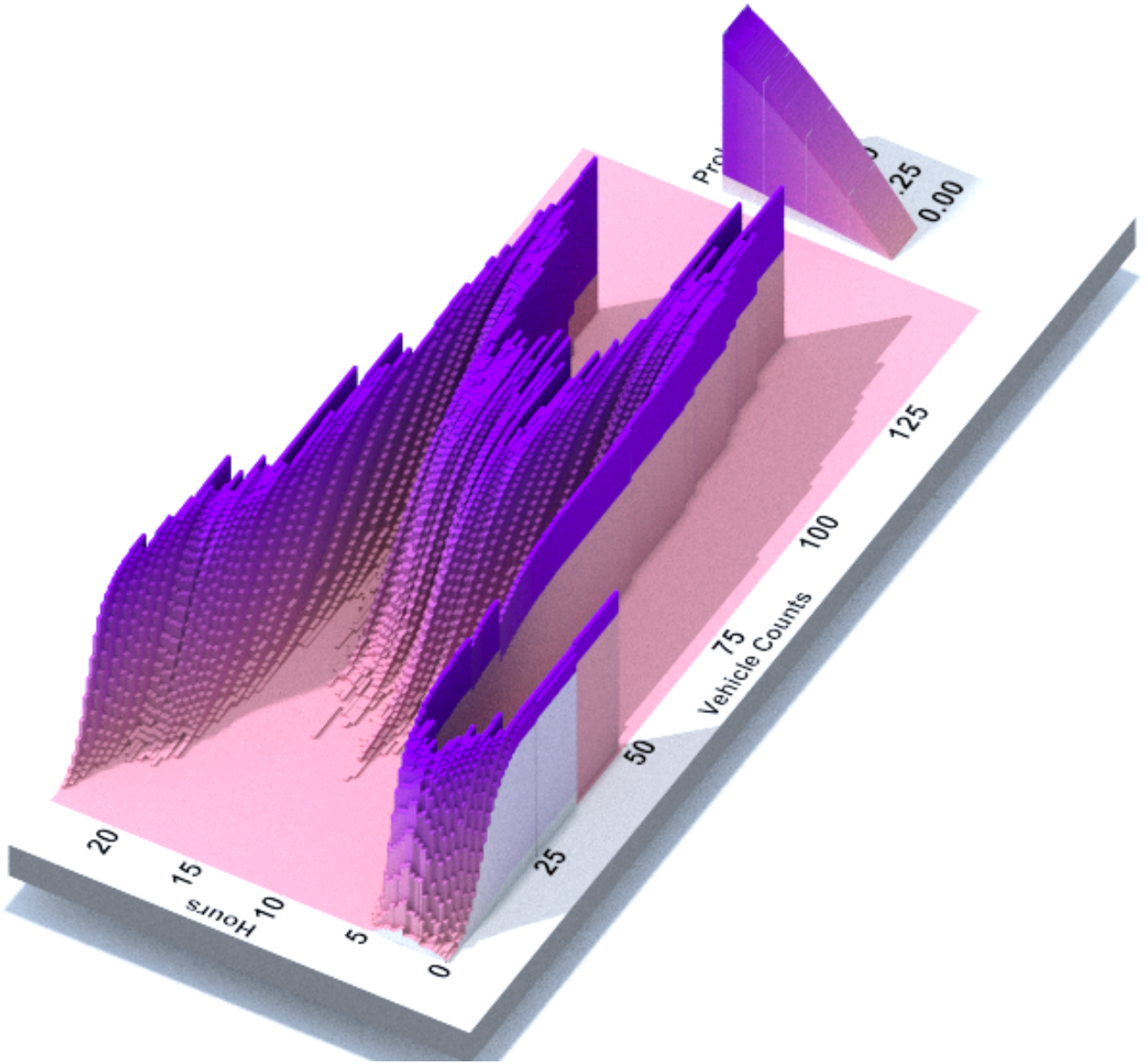
(b) PMF: left angle

Figure B.2 The PMF estimates of the intra-day approach in different angles for the test sample (1N).



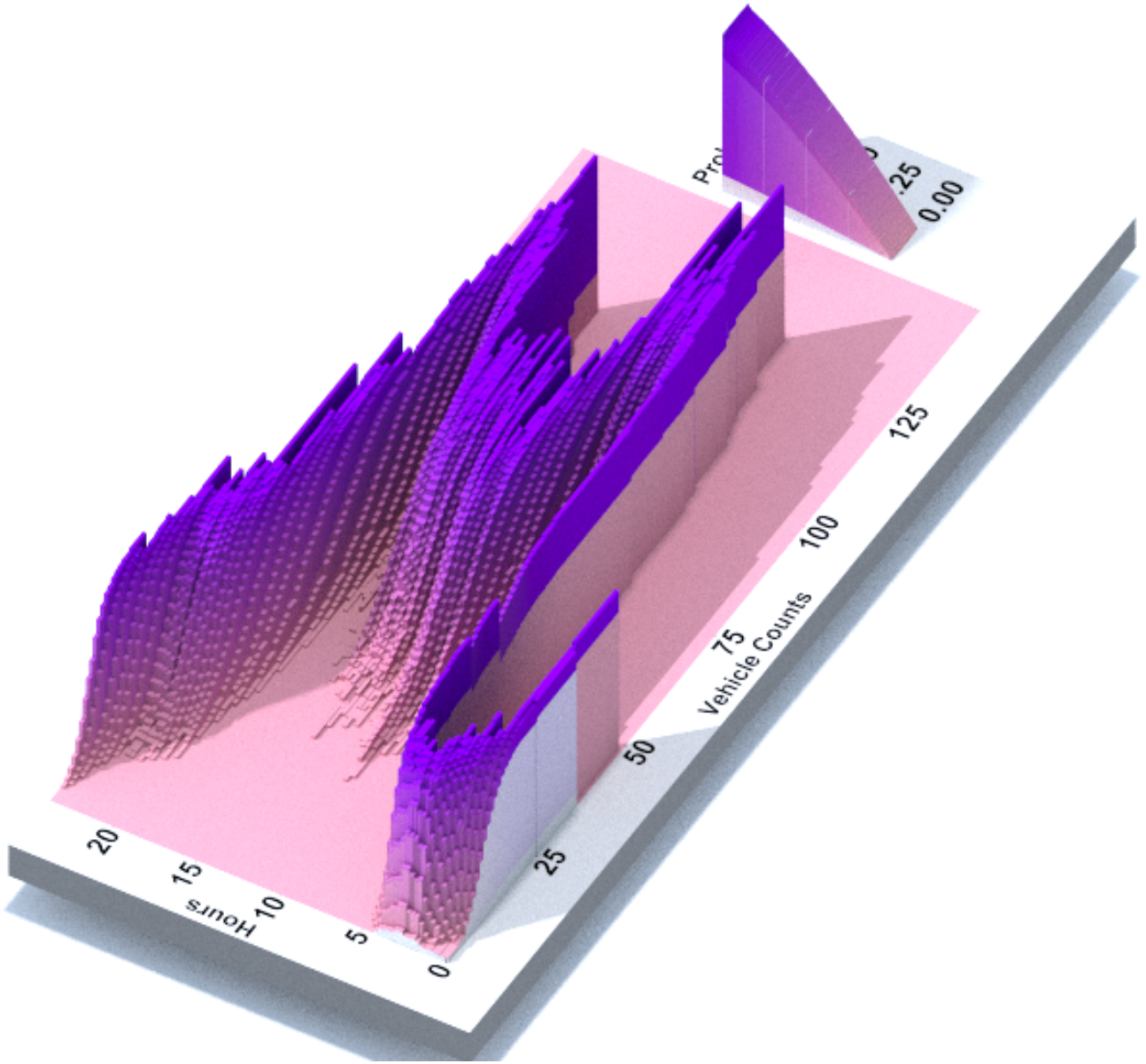
(c) PMF: right angle

Figure B.2 The PMF estimates of the intra-day approach in different angles for the test sample (1N).



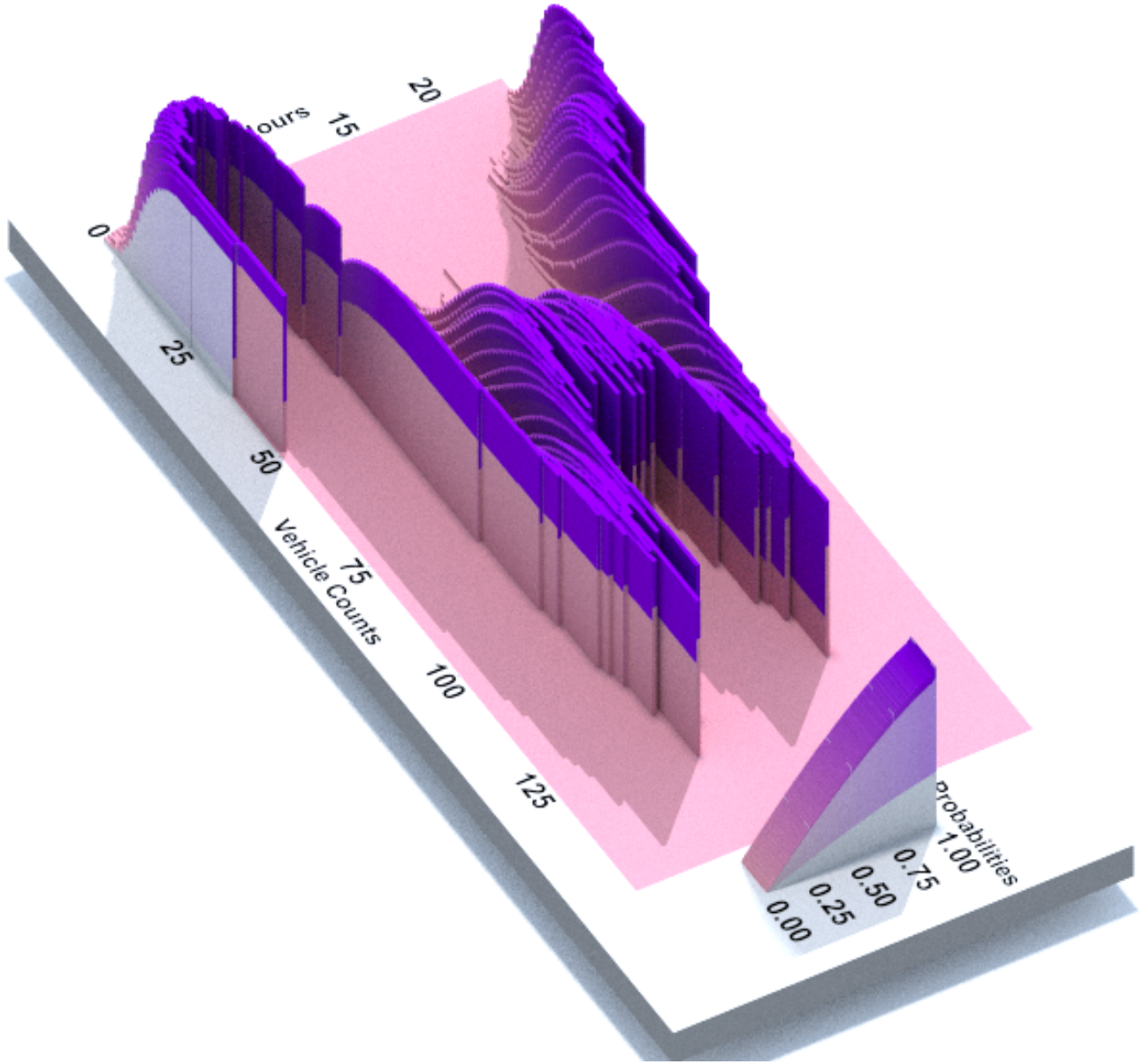
(a) CDF: front angle

Figure B.3 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1N).



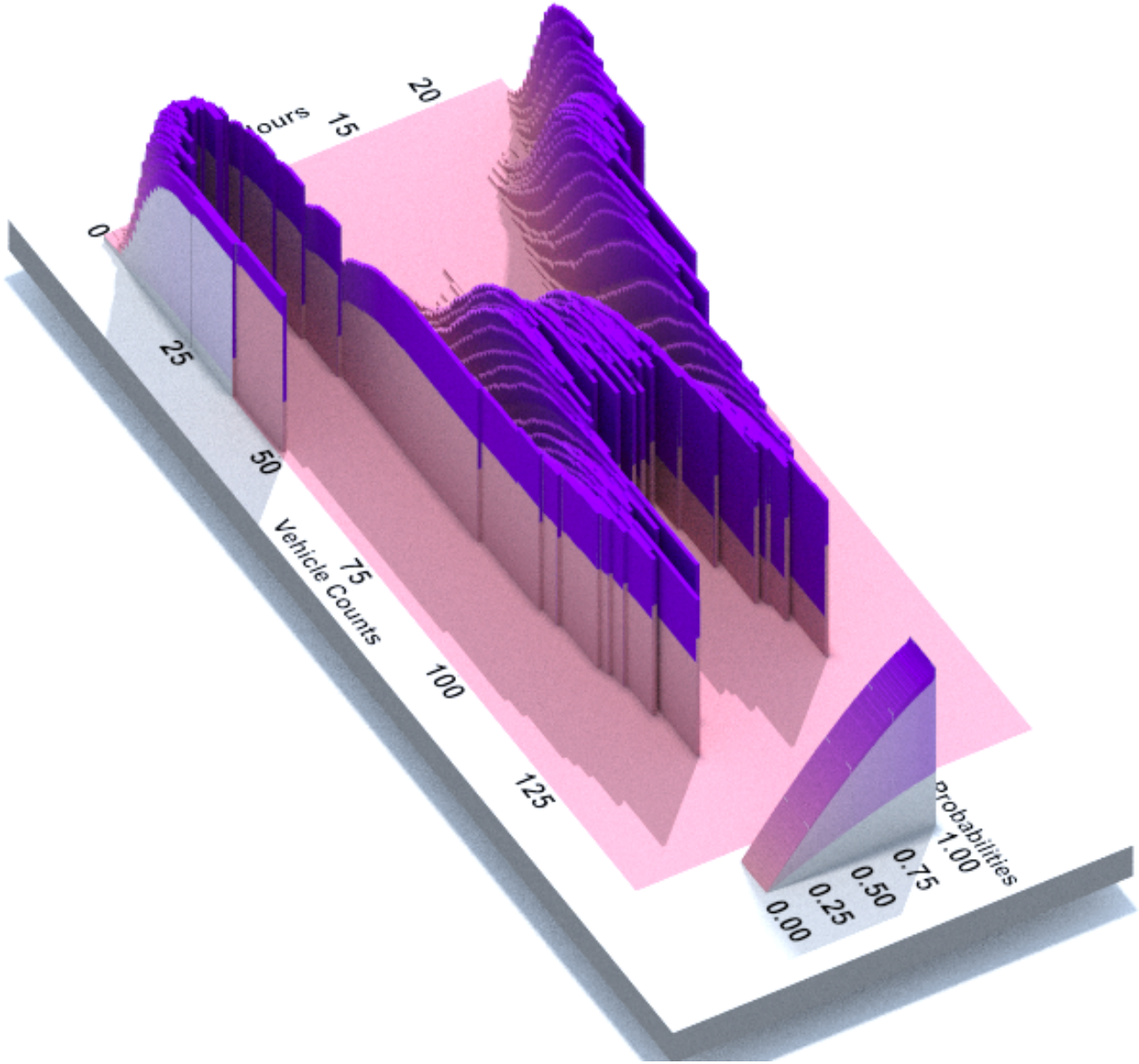
(b) CDF target: front angle

Figure B.3 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1N).



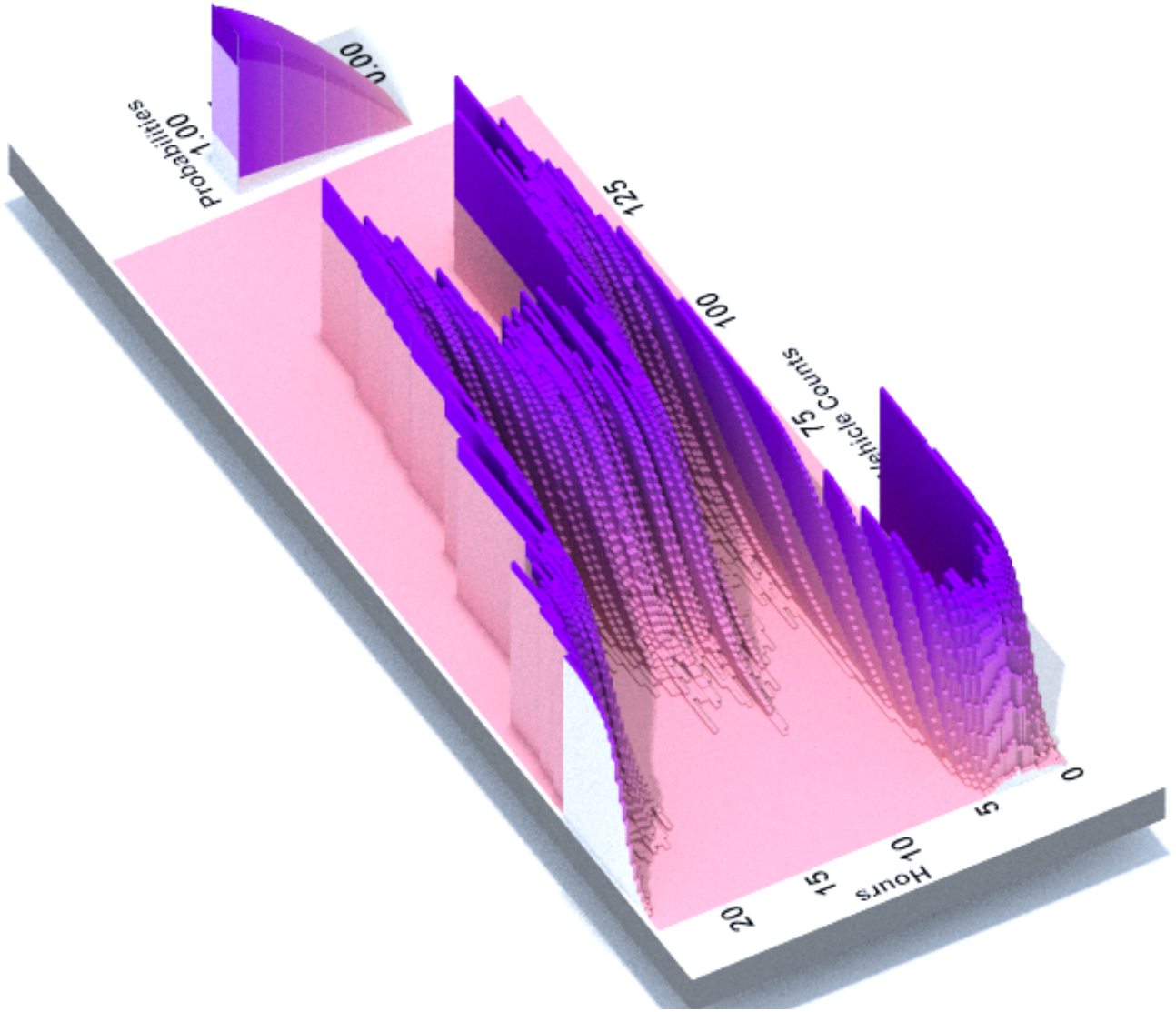
(c) CDF: left angle

Figure B.3 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1N).



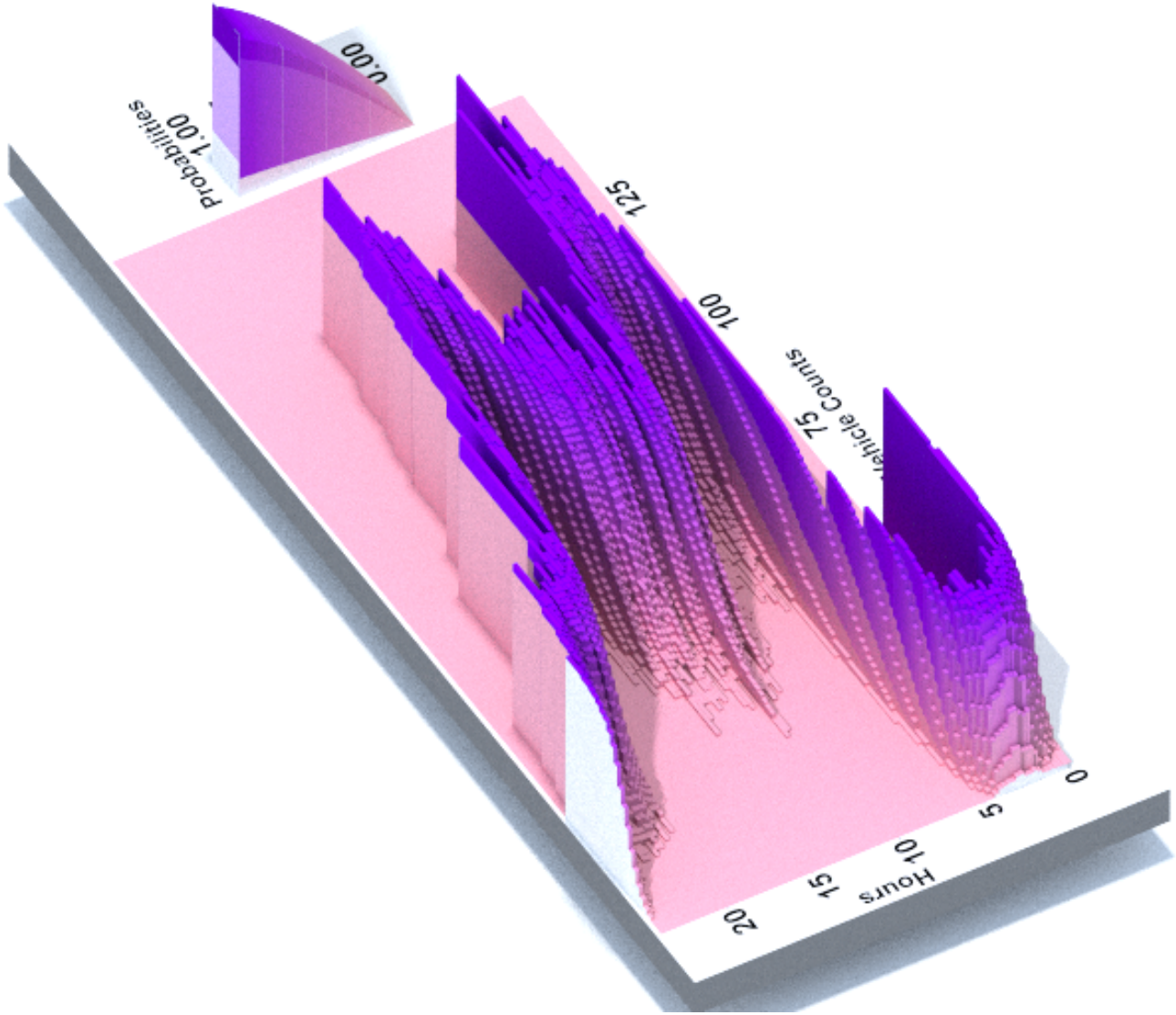
(d) CDF target: left angle

Figure B.3 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1N).



(e) CDF: right angle

Figure B.3 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1N).



(f) CDF target: right angle

Figure B.3 The CDF estimates and corresponding target CDFs of the intra-day approach in different angles for the test sample (1N).

Figures B.1 and B.2 present the CDF and PMF estimates for all 96 time windows for the test sample using Model 6 of Sensor 1N. Furthermore the target CDFs are shown in Figure B.3. To facilitate a more precise comparison, the differences between the target CDFs and CDF estimates are depicted in Figure B.4. Similar patterns as for Section 4.1 are found.

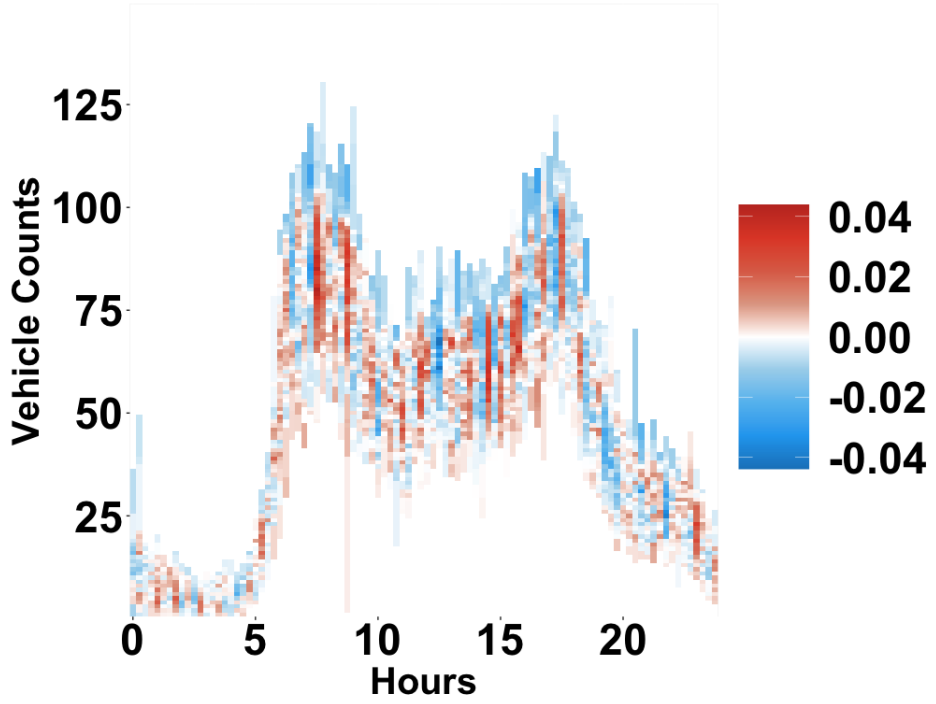
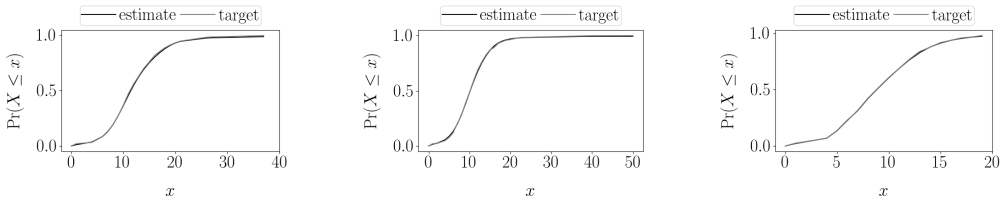


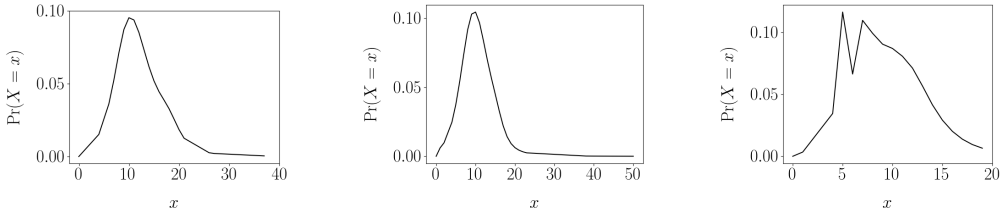
Figure B.4 The differences between the CDF estimates and target CDFs of the intra-day approach for Sensor 1N.
The x-axis depicts the 96 time windows, the y-axis depicts the number of vehicle counts which is the domain of the CDF estimates and the colorbar returns the values of the differences between the CDF estimates and the target CDFs. The differences are shown for the test sample.

Figure B.5 presents the differences between the CDF estimate and CDF target for the best, average, and worst CDF estimates in terms of MSE loss for Model 6 for daytime observations for Sensor 1N. The corresponding time slots are 10:15-10:30, 12:15-12:30 and 16:30-16:45, respectively. The differences between the CDF estimate and CDF target are minor and as a result difficult to detect in the Figure. Figure B.6 presents the corresponding PMF estimates. Different time slots are chosen compared to Section 4.1. The best and average CDF and PMF estimates are smooth functions similar to the CDF and PMF estimates in Section 4.1. However, the worst CDF and PMF estimate attains a kink at vehicle count 5. This is shown in the corresponding PMF estimate that is bi-modal at vehicle counts 5 and 8. The distribution around the mode is wide just as for the corresponding Figure in Section 4.1 but with a split this time. We anticipate that this is caused by the large number of neurons, that is 150 neurons opposed to 25 neurons. Too complex models applied to noisy data patterns might result in PMF estimates with erratic patterns.



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

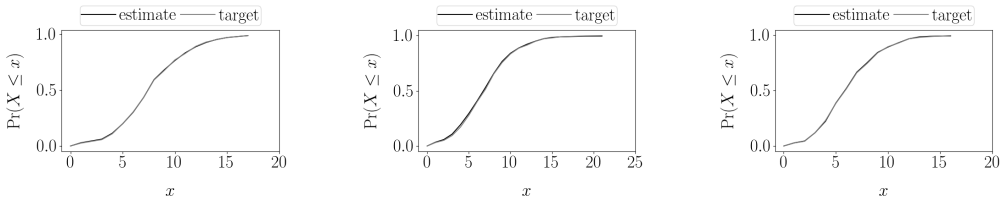
Figure B.5 CDF estimates between 5:30 and 20:00 for the intra-day approach, Mondays 2016 (1N).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 2 hidden layers and 150 neurons are given for the test sample.



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

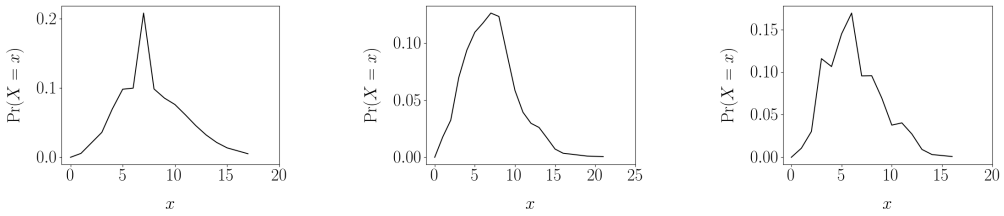
Figure B.6 PMF estimates between 5:30 and 20:00 for the intra-day approach, Mondays 2016 (1N).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 2 hidden layers and 150 neurons are given for the test sample.

Figure B.7 present the differences between the CDF estimate and CDF target for the best, average, and worst CDF estimates in terms of MSE loss for Model 6 for nighttime observations, i.e. between 20:00 and 5:30 for Sensor 1N. The corresponding time slots are 2:45-3:00, 23:30-23:45 and 20:45-21:00, respectively. These closely resemble one another. Figure B.8 presents the corresponding PMF estimates. Slightly different but contemporaneous time slots are selected as in Section 4.1. The best CDF and PMF estimates are erratic similar to the CDF and PMF estimates in Section 4.1 due to the small number of different vehicle counts during the early morning. The average CDF and PMF estimates are smoother than the best estimates which is also in line with Section 4.1. However, the worst CDF and PMF estimates in this Appendix are not as smooth as expected during the late evening, as shown in Figure 4.7(c) of Section 4.1. This is probably due to the complex model used in this Appendix similar to the worst PMF estimate during the day depicted in Figure B.6(c).



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure B.7 CDF estimates between 20:00 and 5:30 for the intra-day approach, Mondays 2016 (1N).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 2 hidden layers and 150 neurons are given for the test sample.



(a) smallest MSE loss (b) average MSE loss (c) largest MSE loss

Figure B.8 PMF estimates between 20:00 and 5:30 for the intra-day approach, Mondays 2016 (1N).
The best, average, and worst estimates in terms of MSE loss values out of 96 PMFs using 2 hidden layers and 150 neurons are given for the test sample.

Similar general conclusions as in Section 4.1 are made. The difference is that the smaller model with 25 neurons used in Section 4.1 obtains smoother PMFs for some estimates, especially for the worst PMF estimates, compared to the PMFs estimated by the model in this Appendix with 150 neurons.

C Sensitivity analysis of different time windows for the intra-day approach

In this Appendix average MSE losses over 50 replicates for 10 minute and 20 minute time windows are presented to show the robustness against short time windows other than the 15 minute time window used in Section 4.1. See Tables C.1 and C.2 for the average MSE losses using a time window of 10 minutes and 20 minutes, respectively. Similarly to the MSE losses using 15 minutes, Models 1, 2 and 3 obtain larger loss values than the models with more than one hidden layer. Note that the MSE losses for time window 10 are larger and the MSE losses for time window 20 are smaller than the 15 minute time window MSE losses. This is intuitive as the sample sizes with a 10 minute time window are smaller and the sample sizes with a 20 minute time window are larger than the sample sizes with a 15 minute time window.

Model	1	2	3	4	5	6	7	8	9
Train (5 fold)	15.115	16.274	21.027	14.575	13.228	13.147	13.383	11.168	11.010
	(0.516)	(0.579)	(0.880)	(0.680)	(0.600)	(0.618)	(0.623)	(0.502)	(0.492)
Validation	14.921	16.578	21.626	14.267	13.579	14.019	13.642	11.462	11.565
	(1.046)	(1.233)	(2.299)	(1.357)	(1.351)	(1.550)	(1.350)	(1.141)	(1.143)
Train	17.239	19.177	25.458	16.948	15.175	13.944	14.889	12.452	12.096
	(0.641)	(0.836)	(1.243)	(0.762)	(0.681)	(0.602)	(0.609)	(0.642)	(0.623)
Test	24.486	26.241	30.109	25.202	23.014	21.580	22.663	19.189	19.077
	(2.392)	(2.879)	(3.801)	(3.503)	(2.999)	(2.858)	(2.368)	(2.999)	(3.072)
									Values in the table scaled by 10^{-5}

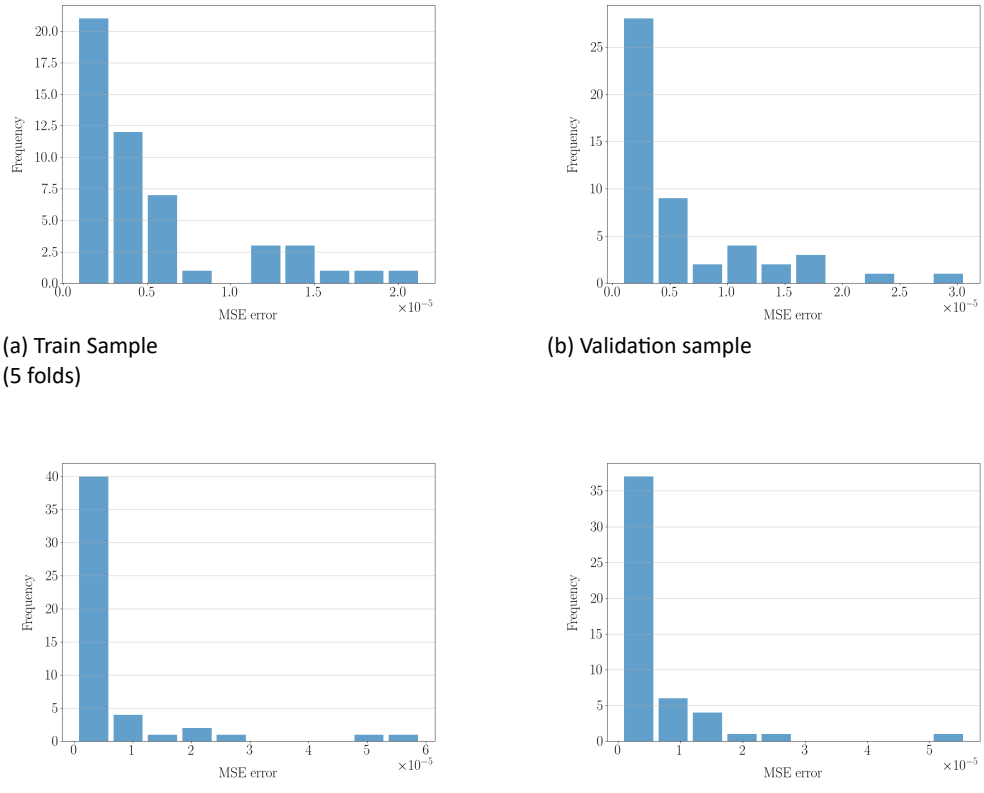
Table C.1 MSE losses for the intra-day approach with time window 10, Mondays 2016 (1N).
The table reports the mean (standard error in parentheses) of the average MSE losses using the difference between the target, and estimated CDF obtained for the train, validation and test samples.

Model	1	2	3	4	5	6	7	8	9
Train (5 fold)	9.891	10.868	14.477	9.263	8.038	7.634	8.472	6.737	6.749
	(0.241)	(0.296)	(0.423)	(0.313)	(0.272)	(0.250)	(0.277)	(0.214)	(0.214)
Validation	10.228	11.485	15.336	9.211	8.276	6.889	8.463	6.026	6.709
	(0.562)	(0.709)	(0.992)	(0.656)	(0.611)	(0.447)	(0.589)	(0.367)	(0.467)
Train	10.094	11.527	13.949	9.176	7.143	7.315	8.099	5.205	5.359
	(0.234)	(0.294)	(0.372)	(0.289)	(0.227)	(0.269)	(0.270)	(0.164)	(0.199)
Test	12.782	14.438	14.828	11.472	10.018	10.383	9.491	8.102	9.009
	(0.890)	(1.137)	(0.904)	(0.887)	(0.802)	(1.205)	(0.666)	(0.663)	(0.921)
									Values in the table scaled by 10^{-5}

Table C.2 MSE losses for the intra-day approach with time window 20, Mondays 2016 (1N).
The table reports the mean (standard error in parentheses) of the average MSE losses using the difference between the target, and estimated CDF obtained for the train, validation and test samples.

D Range of MSE errors for the inter-day approach

As explained in Section 4, the k-fold training phase and train-test learning phase for the intra-day data is repeated 50 times. For each replicate the sample is randomly split in 6 sub-samples for the train, validation and test samples. Figure D.1 shows a histogram of the 50 MSE loss values of estimating the inter-day densities using optimal Model 5. In all figures, most MSE errors obtain similar small values with a few larger values for some replicates. Note that the x-axis differs for the upper and lower figures. The MSE values of the k-fold train and validation samples shown in Figures D.1(a) and D.1(b) are less widespread than the train and test samples shown in Figures D.1(c) and D.1(d).



(a) Train Sample (5 folds)

(b) Validation sample

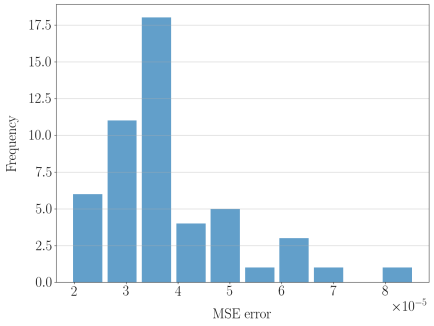
(c) Train sample

(d) Test sample

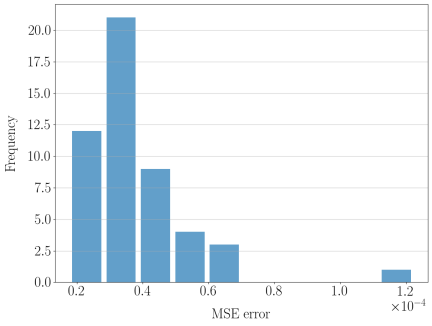
Figure D.1 The MSE errors of the inter-day approach for 50 replicates in a histogram for a regular Monday in 2016 for the test sample (1N).

Histograms of the 50 MSE loss values of the inter-day densities under Model 1 for the extreme Monday on the 22^{nd} of August are shown in Figure D.2. For all figures, the MSE values are more spread than for the regular Monday but still centered around a certain value. The MSE values of the k-fold train and validation samples shown in Figures D.2(a) and D.2(b) are more concentrated around the mode than the train and test samples shown in Figures D.2(c) and D.2(d). For the k-fold cross validation sample shown in

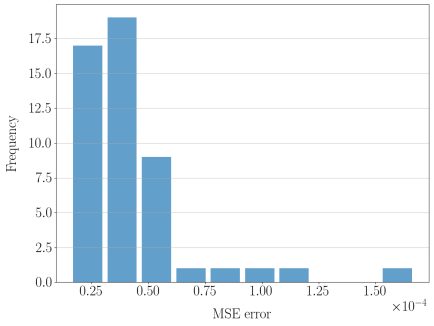
Figure D.2(a), the MSE values are most volatile over the domain though 10 times smaller than for the validation, train and test samples.



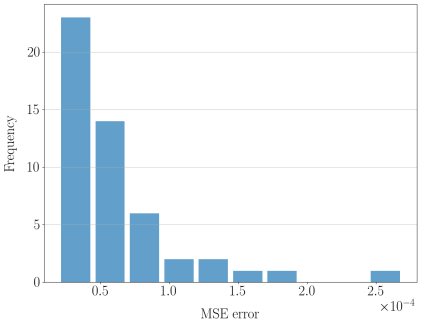
(a) Train Sample
(5 folds)



(b) Validation sample



(c) Train sample



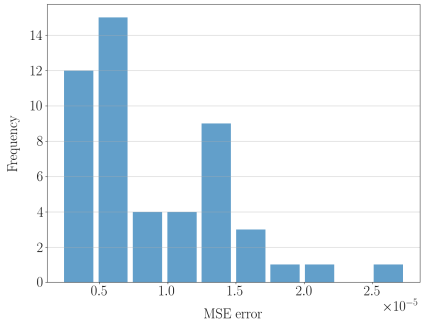
(d) Test sample

Figure D.2 The MSE errors of the inter-day approach for 50 replicates in a histogram for the extreme Monday which is the 22^{nd} of August in 2016 for the test sample (1N).

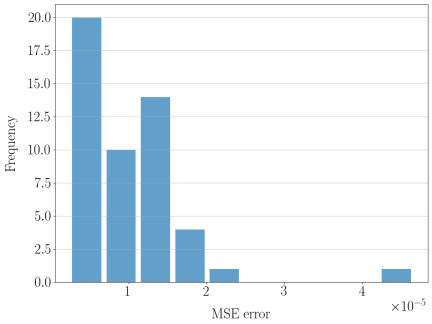
In general, the MSE values are more affected by randomly splitting the sample in sub-samples when the sample size is smaller.

E Range of MSE errors for the cross-sectional approach

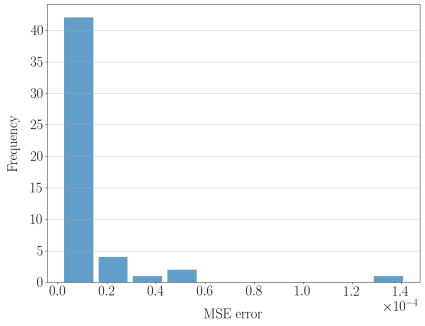
The k-fold training phase and the train-test learning phase for the cross-sectional data is repeated 50 times, using different samples for training and evaluation from the observed data of the regular Mondays. Figure E.1 shows a histogram of the 50 MSE loss values of estimating the cross-sectional densities using the optimal Model 5. The MSE loss values for k-fold train and validation samples shown in Figures E.1(a) and E.1(b) obtain smaller MSE values than for the train and test samples shown in Figures E.1(c) and E.1(d). The majority of MSE loss values are concentrated at a similar value.



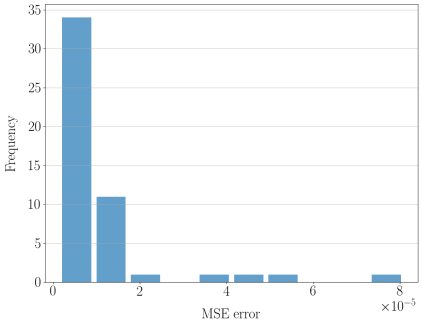
(a) Train Sample
(5 folds)



(b) Validation sample



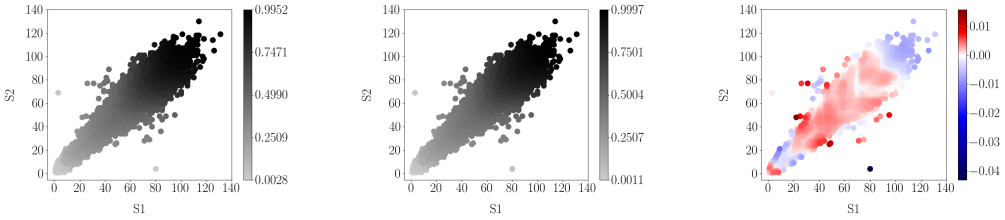
(c) Train sample



(d) Test sample

Figure E.1 The MSE errors of the cross-sectional approach for 50 replicates in a histogram for a regular Monday in 2016 for the test sample (1S and 2S).

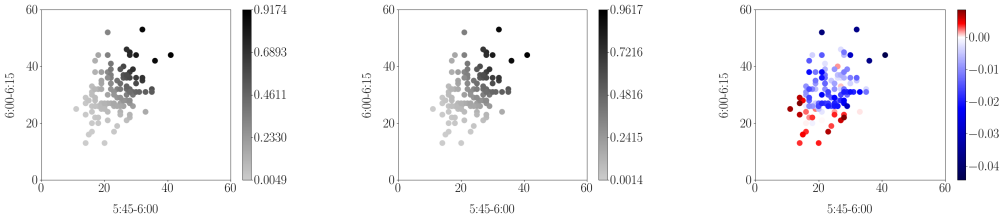
F CDF estimates and targets of the cross-sectional approach



(a) Estimate (b) Target (c) Estimate-Target

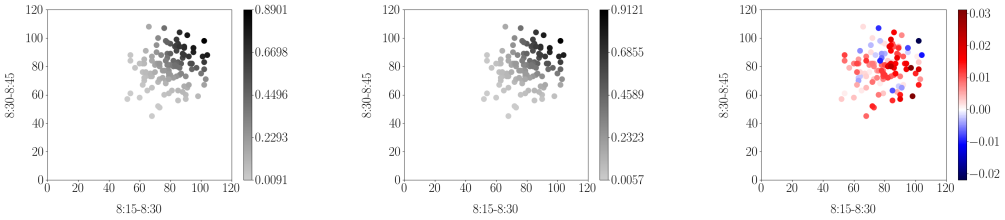
Figure F.1 CDF estimate, Mondays 2016 (1S and 2S).
(a) estimated CDF, (b) target CDF and (c) difference between estimated and target CDF. The x-axis and y-axis labels refer to the number of vehicle counts per minute for the sensor of the label. The estimates are shown for the test sample.

G CDF estimates and targets of the temporal approach



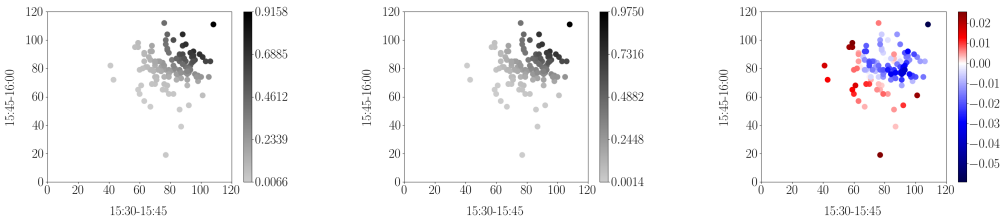
(a) Estimate (b) Target (c) Estimate-Target

Figure G.1 Best CDF estimate between 5:30h and 21:00h, Mondays 2016 (1S). (a) estimated CDF, (b) target CDF and (c) difference between estimated and target CDF. The best estimate in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons is given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.



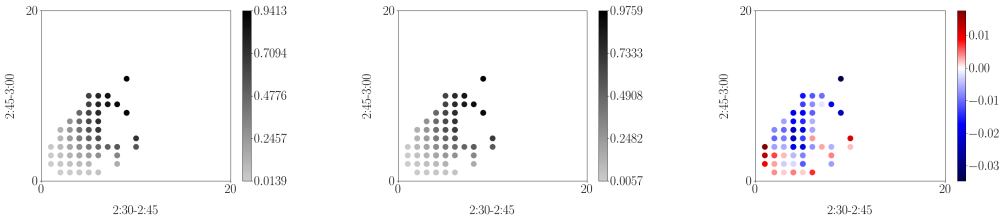
(a) Estimate (b) Target (c) Estimate-Target

Figure G.2 Average CDF estimate between 5:30h and 21:00h, Mondays 2016 (1S). (a) estimated CDF, (b) target CDF and (c) difference between estimated and target CDF. The average estimate in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons is given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.



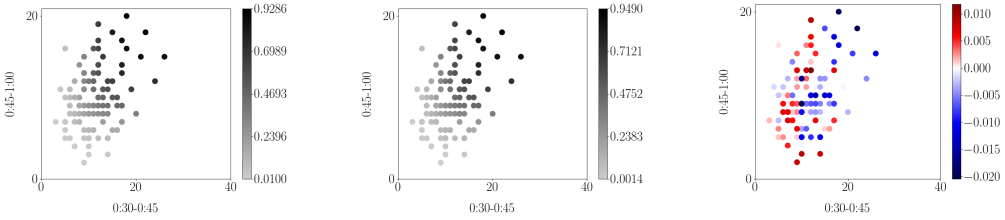
(a) Estimate (b) Target (c) Estimate-Target

Figure G.3 Worst CDF estimate between 5:30h and 21:00h, Mondays 2016 (1S). (a) estimated CDF, (b) target CDF and (c) difference between estimated and target CDF. The worst estimate in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons is given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.



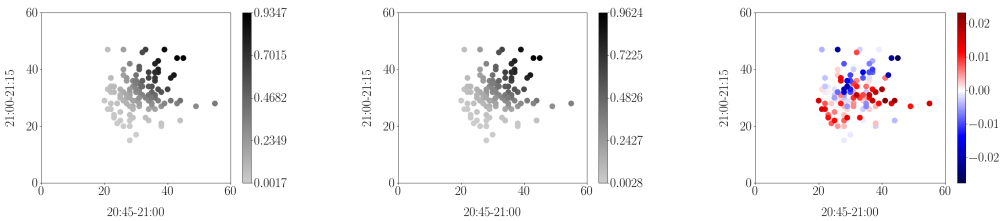
(a) Estimate (b) Target (c) Estimate-Target

Figure G.4 Best CDF estimate between 21:00h and 5:30h, Mondays 2016 (1S). (a) estimated CDF, (b) target CDF and (c) difference between estimated and target CDF. The best estimate in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons is given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.



(a) Estimate (b) Target (c) Estimate-Target

Figure G.5 Average CDF estimate between 21:00h and 5:30h, Mondays 2016 (1S). (a) estimated CDF, (b) target CDF and (c) difference between estimated and target CDF. The average estimate in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons is given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.



(a) Estimate (b) Target (c) Estimate-Target

Figure G.6 Worst CDF estimate between 21:00h and 5:30h, Mondays 2016 (1S). (a) estimated CDF, (b) target CDF and (c) difference between estimated and target CDF. The worst estimate in terms of MSE loss values out of 96 PMFs using 1 hidden layer and 25 neurons is given for the test sample. The x-axis and y-axis labels refer to the number of vehicle counts per minute during the time slot of the label.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information