MASTER'S THESIS

Statistical Modelling of Pedestrian Flows

ERIK HÅKANSSON

Thesis for the Degree of Master of Science

Statistical Modelling of Pedestrian Flows

Erik Håkansson

Statistical Modelling of Pedestrian Flows
ERIK HÅKANSSON

Statistical Modelling of Pedestrian Flows

ERIK HÅKANSSON
Department of Mathematical Sciences
University of Gothenburg

Abstract

Pedestrian counts and in particular their relation to the buildings in the vicinity of the street and to the structure of the street network is of central interest in the space syntax field. This report is concerned with using statistical techniques to model pedestrian counts and in particular how these counts vary over the day. Of interest is whether the variation over the day for a street can be predicted based on its *density type*, describing the nearby buildings, and *street type*, describing its role in in the city's overall street network.

Using data from Amsterdam, London and Stockholm the hour-by-hour pedestrian counts are modelled with the so-called *functional ANOVA* method, using the aforementioned types to divide the streets into groups. Additionally, the effect of the presence of schools, stores and public transport stops near the streets on pedestrian counts is considered. The model is fitted in a Bayesian framework using the *integrated nested Laplace approximation* technique. The results indicate that this model works well but that it might be somewhat too rigid to capture all the variability in the data, failing to capture some of the difference between groups and between the cities. Some possible extensions to the model to remedy this are suggested.

Acknowledgements

Det är kallt, nästan höst nu
Jag går runt ditt kvarter
Fjärde gången den här kvällen
men vem räknar på det

Annika Norlin, *Ditt Kvarter*

Contents

Contents

x

Chapter 1

Introduction

In the fields of urban planning and design, pedestrian counts, or *pedestrian flows* as they are often called, are of central interest. Of particular interest is how the pedestrian counts can be explained by the urban area itself, or in more statistical language how they depend on various covariates. Knowledge about this can be of use in the planning and design of new urban developments to adopt the infrastructure to the expected pedestrian flows. This report describes the results of a statistical modelling for such counts and how they vary during the day.

The data used in this report was originally collected in a survey by the Spatial Morphology Group at Chalmers in October of 2017, using mobile phone Wi-Fi signals. This masters project began with analysing this data with a focus on modelling total pedestrian counts over the entire day in. This analysis is discussed very briefly in Appendix A. The results of it are not essential to understanding the current report but give some insight into the choices made when conducting the analysis at hand. The results of this total counts analysis is the focus of the (forthcoming) article [Sta+19].

In this report the main goal is to make a more fine-grained analysis by also considering how the counts vary over time; specifically we will model how the vary hour by hour throughout the day. This makes for a more complicated modelling problem but also serves to give further insight into the data.

The starting point of the modelling is the street typology described in [Ber+17]. In that article, streets are divided into *street types* and *density types*; these (roughly speaking) describe how central streets are to the street network and what kind buildings there are around the street respectively. Since this typology is used as the main explanatory variable this project also serves as an investigation of how well the types describe the time-effects in the data.

To model the time dependence one has to move beyond the most basic statistical methods of ANOVA and regression which were used in the analysis of total counts [Sta+19]. The time effects are therefore handled in a so-called *functional ANOVA* framework [Yue+16]. That is, we consider the counts as functions of time and use the types to give ANOVA categories. In addition to giving a model for the counts this will lead to interpretable effects for individual types, which can be of independent interest, and also show what kinds of behaviour are coming from which parts of the data.

The model fitting is done using the *INLA* method [RMC09] as implemented in the R-INLA package for the R progamming language [R C18].

Lastly we remark that part of the more long-term goal of this project (including both this report and the analysis in [Sta+19]) is to visualise the results of the models along with the data to provide further insight. The total count results have been incorporated into a web based GUI that allows easy overview of these results and of the data [Ber+]. As of writing this has not yet been done for the hour by hour models due to time constraints.

Outline

Chapter 2 describes the data used. It is divided into two Sections: Section 2.1, which gives a general overview of the data, and Section 2.2, which gives more detail on the main covariates used.

Chapter 3 gives a description of the statistical inference, including the model used and how it is fitted. It is divided into a few sections. A short review of the Bayesian statistical paradigm is given in Section 3.1, followed by a description of the statistical model used in Sections 3.2 and 3.3. This is followed by Section 3.4 where the INLA model fitting procedure is described. The last section of this chapter, Section 3.5, introduces the methods for model checking and model comparison that we use.

Chapter 4 shows and discusses the results of the model fitting.

Finally, Chapter 5 summarises the project and gives some suggestions for possible extensions.

Chapter 2

Data

This chapter describes the data used in this report.

2.1 Data Description

Here we give a quick rundown of the data, the variables we use, and what they mean. The data was gathered in three European cities: Amsterdam, London and Stockholm. This was done in October of 2017. The data contains quite a few variables so the description in this section will be limited to the ones that are used in the statistical analysis in this report. Overall, the data contains 10848 observations, corresponding to 678 streets across 46 neighbourhoods.

The location of the measured neighbourhoods are shown in Figure 2.2 for Amsterdam, Figure 2.3 for London and Figure 2.4 for Stockholm.

Each of these observations consist of the pedestrian count during one hour for one street segment with the hours ranging from 6 in the morning to 21 in the evening, so a total of 16 hours for each street segment. Figure 2.1 shows an excerpt from the data split up by the types described below. As mentioned above, the goal of this project is to make a statistical model for these counts over time based on covariates describing the streets (these covariates remain the same over time).

The focus of the analysis is the categorical covariates arising from the typology of [Ber+17], which divides the streets into *density types* (6 levels) and *street types*[1] (4 levels). They are described in more detail below. Each street segment belongs to one density type and one street type. The types come from clustering of continuous variables; these continuous variables are available in the dataset but will not be used by themselves in this project.

There are two major reasons for using these categorical variables instead of the continous variables they are built from. The first is interpretability. The interpretation of the continuous variables is somewhat opaque, but (as we see below) the type variables can be interpreted. This is also useful when planning new urban developments, as it is likely easier to e.g. construct an area of density type 3 than one with a specific value of some continuous measurement. The other reason is ease of statistical modelling. As was mentioned in the introduction, the model considered in this report is based on ANOVA, but

---

[1] Or *centrality types.*
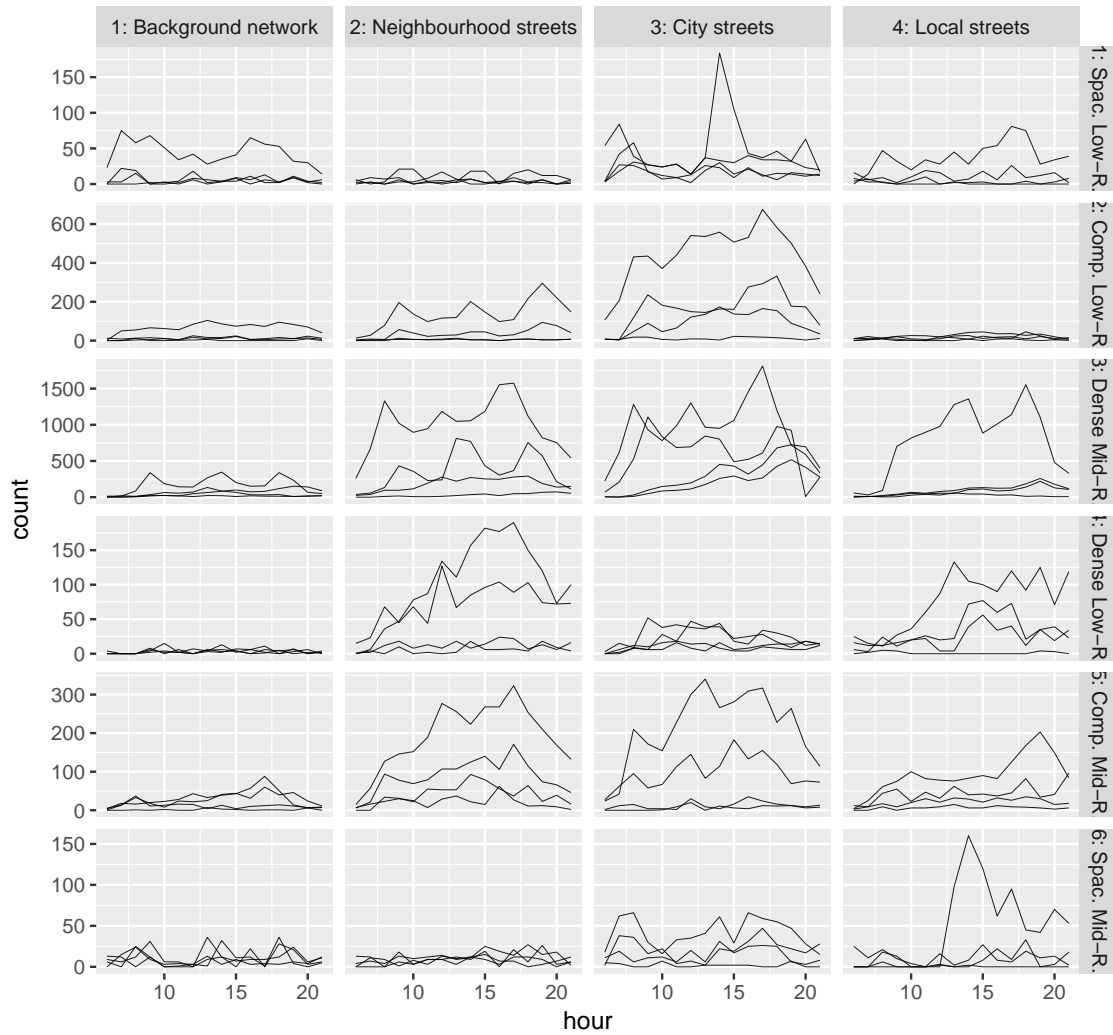
Figure 2.1: Counts for 3 randomly chosen streets in each density-street group.

with a time effect for each group in the ANOVA. This type of model of course requires categorical covariates to divide the data into groups. If we were to use the continuous covariates, we would have to devise a model where the mean value is a function $f(t, s)$ of both time t and continuous covariate values s. This is more complicated than the ANOVA situation where we instead for each group i estimate a function $f_i(t)$ of just one variable.

In addition to the types the analysis will use so-called 'attraction' covariates. These measure the presence of (potentially) pedestrian-attracting institutions in the vicinity of the street segment. The attractions considered are public transport stops, schools, and so-called 'local markets'. Local markets in this case refers to retail stores, cafes, and restaurants. For all three (public transport, schools, local markets) we have data on both the number of institutions on the street segment and the number within a 500 meter walking distance of the street segment.

We note that these values are computed in a somewhat strange manner and that because of this, these attraction covariates are actually *continuous*. The reason for this is as follows. Some street segments considered in the data are divided into shorter paths.
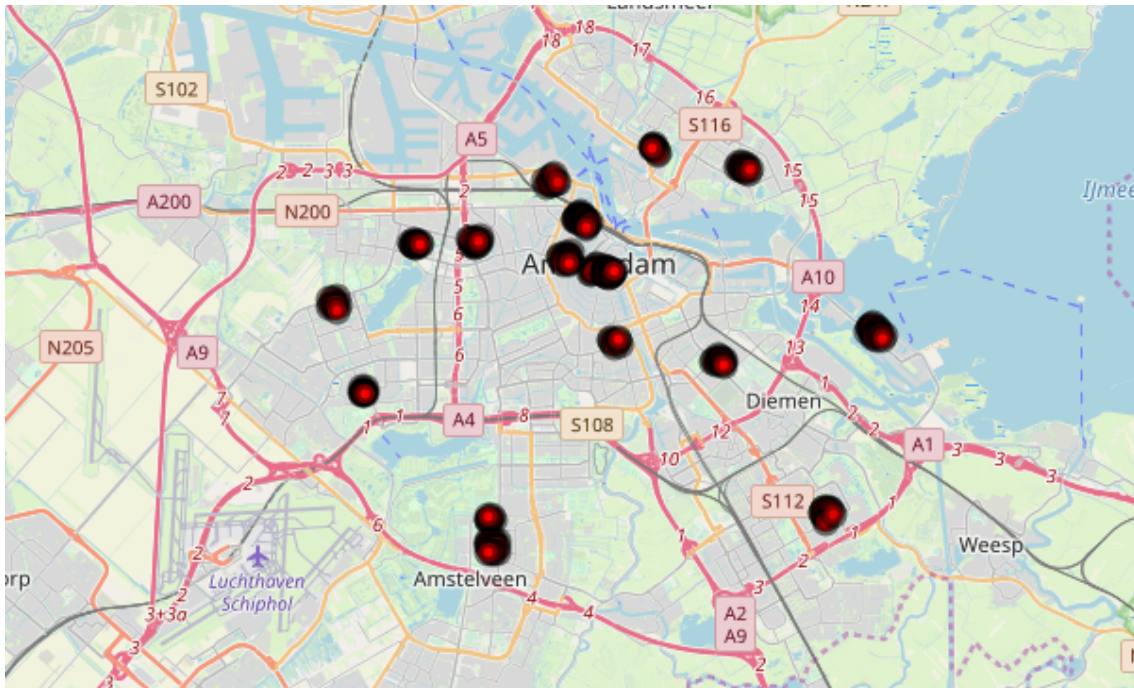
Figure 2.2: Neighbourhood locations in Amsterdam. Image taken from the GUI [Ber+].
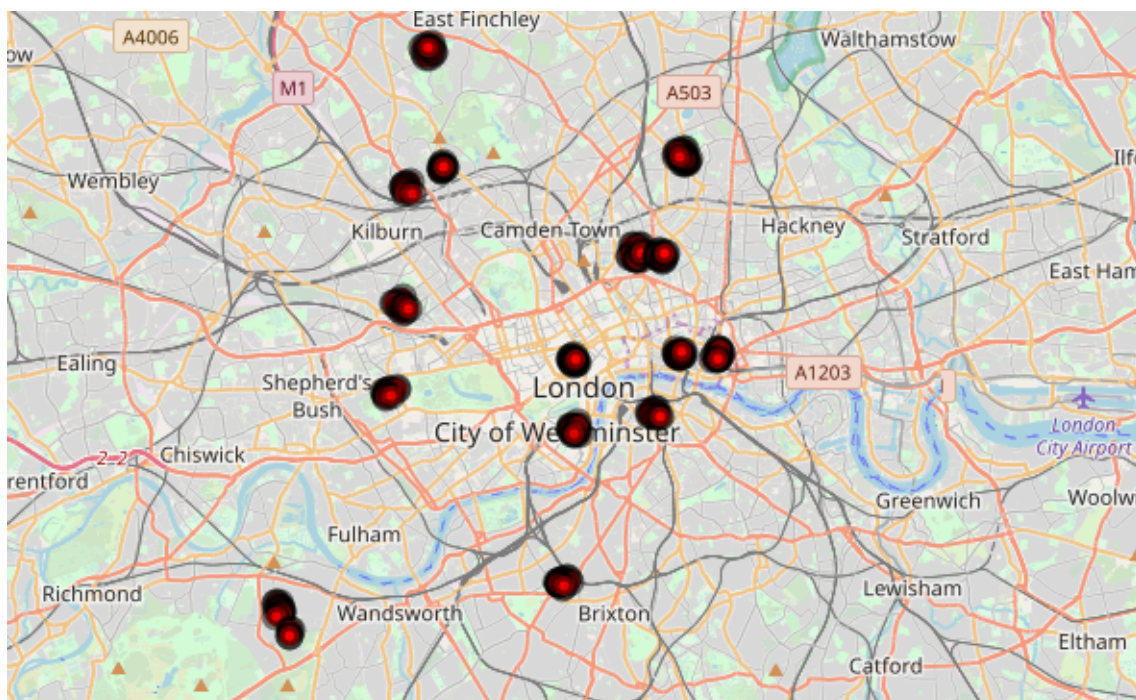
Figure 2.3: Neighbourhood locations in London. Image taken from the GUI [Ber+].
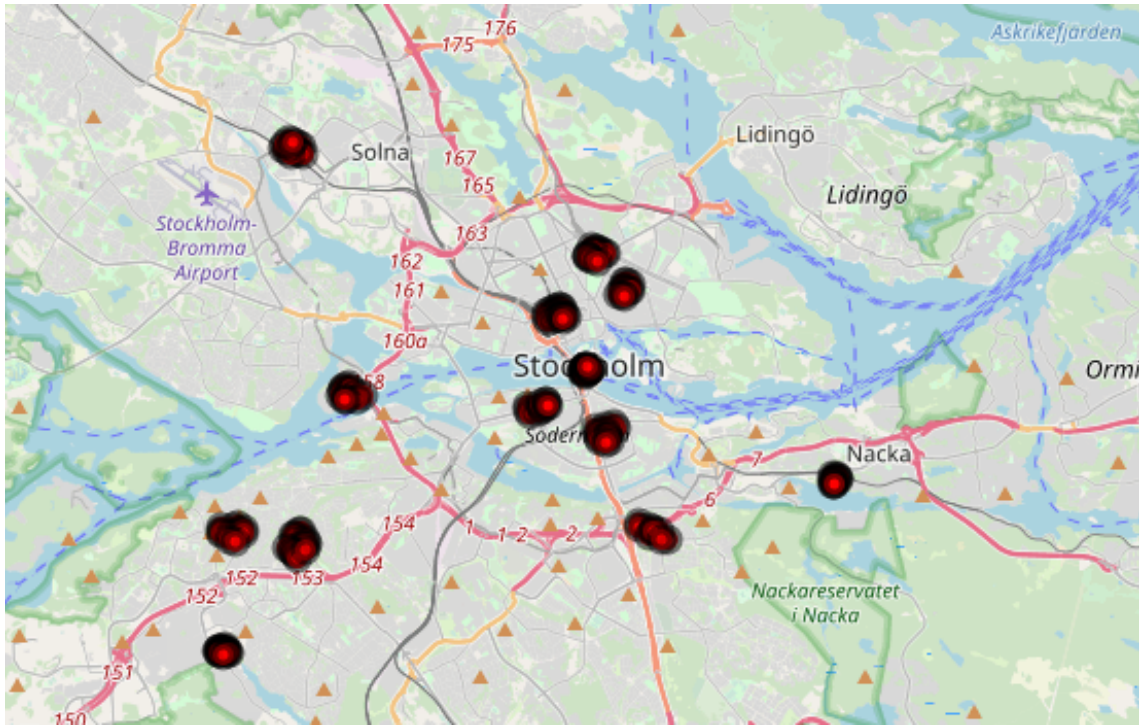
Figure 2.4: Neighbourhood locations in Stockholm. Image taken from the GUI [Ber+].

When this is the case, the attractions are counted separately for each path. The covariate value given to the street segment is then the average of these counts per path. As a simple example, if a street segment is made up of four shorter paths with 1, 2, 3 and 4 schools within a 500 metre distance then the street segment is considered to have $(1 + 2 + 3 + 4)/4 = 2.5$ schools in a 500 metre distance.

The attraction covariates are normalized to have mean 0 and standard deviation 1.

2.2 Density Types and Street Types

As mentioned above two major covariates used are density and street types. Both of these are based on clustering on building density metrics and betweenness at different scales respectively, the details are described in [Ber+17]. Note that the names used in this article differ from the names we use. A summary of the types is shown in Table 2.1.

The clustering for density types is based on two measures of density known as *floor space index*[2] (FSI) and *ground space index* (GSI). Specifically, these measures computed in a 500m radius around the street are used. Roughly speaking FSI is the ratio of total floor area to lot area, and GSI is the ratio of 'footprint size' to lot area. Thus FSI is a measure of the density of total living area, and GSI is a measure of how large part of the land is taken up by buildings. Note that while GSI $\in [0, 1]$ always, FSI can in principle take arbitrary positive values as buildings can have multiple floors and hence more square footage than the lots they are placed on. High FSI and low GSI means that buildings are tall with many floors.

---

[2]Swedish: exploateringstal.

Density type	Description	Street type	Description
1	Spacious low-rise	1	Background network
2	Compact low-rise	2	Neighbourhood streets
3	Dense mid-rise	3	City streets
4	Dense low-rise	4	Local streets
5	Compact mid-rise		
6	Spacious mid-rise		

Table 2.1: Summary of density types and street types.

The street types were created using a clustering on principal components of angular betweenness measures on different scales. This measures (in some sense) how important the street is to the road network on different 'zoom levels'. Streets with high betweenness on large scales are of importance when moving larger distances, while high betweenness on smaller scales means that they are important when moving shorter distances.

The density type is a function of the kind of area in the city a street is located in; the street type rather describes what type of street it is. While there is some relation between these the data was chosen in such a manner that there are streets in the data belonging to each of the 24 density-street groups. To be more precise, Table 2.2 shows the distribution of the type combinations.

Both the density and street types are to some extent interpretable. These interpretations will be helpful when discussing model fitting results, so we present them here. We begin with the density types.

Type 3 (dense mid-rise) is typical in city centres. The area directly surrounding the city centre typically belongs to type 5, which has lower buildings than type 3 but the same compact coverage. Streets in both these types are of course expected to have high pedestrian counts. Types 1,2, and 6 are likely to have lower counts. Type 1 consists to a large extent of villa areas and urban sprawl and type 2 tends to be typical suburban areas. The third of these, type 6, is at least in a Swedish milieu exemplified by typical 'miljonprogrammet' areas, with tall buildings placed somewhat sparsely. The last density type, type 4 or dense low-rise is a (rather odd) combination of two kinds of building development – namely, areas of this type are usually either industrial or historic (i.e. medieval) city centres.

For the street types the names given in Table 2.1 are more descriptive than for the Densities. Type 1, background network, has streets which essentially are of low betweenness at all scales and hence are not so important to the street network. The three remaining categories can be roughly arranged by the scale at which they are most important. First of these is type 4, local streets, which are important when moving within neighbourhoods but not on larger scales than that. Then comes type 2, neighbourhood streets. These are important for moving in between neighbourhoods. Finally, type 3 (city streets) are important on the largest scales, i.e. when moving across the city.

An overview of the spatial distribution of density types is shown in Figure 2.5 and of street types in Figure 2.6. The density types are shown for the city centres, and hence types 3 and 5 are dominant. When it comes to the street types most streets are part of the background network (i.e. type 1), with larger streets in types 2 through 4.

Figure 2.5: Spatial distribution of density types in the three cities and Gothenburg. Taken (with permission) from [Ber+17, Figure 8].
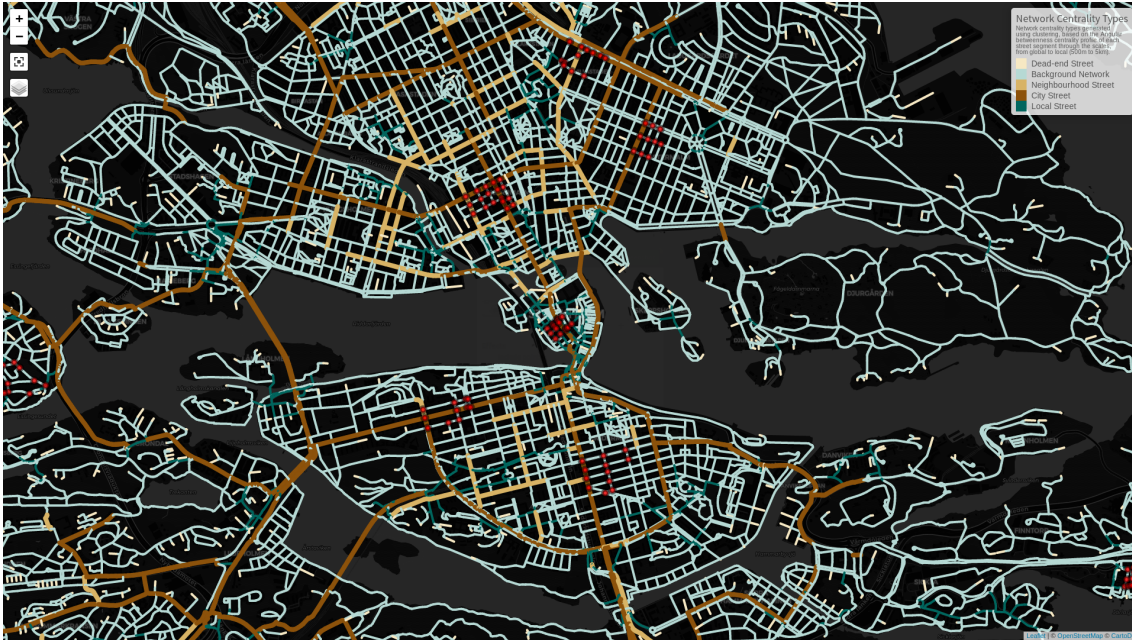
Figure 2.6: Excerpt from the GUI [Ber+], showing the spatial distribution of street types in central Stockholm. The colours show represent the types: light blue – 1 (background network), light brown – 2 (neighbourhood street), dark brown – 3 (city street), darker blue – 4 (local street).

	Street 1	Street 2	Street 3	Street 4	Total
Density 1	0.09	0.02	0.01	0.03	0.15
Density 2	0.05	0.02	0.03	0.05	0.15
Density 3	0.11	0.03	0.03	0.07	0.24
Density 4	0.05	0.04	0.01	0.01	0.12
Density 5	0.13	0.03	0.03	0.02	0.21
Density 6	0.05	0.03	0.01	0.03	0.13
Total	0.48	0.17	0.13	0.22	

Table 2.2: Distribution of density and street types in the data.

Chapter 3

Model and Methods

3.1 A Quick Review of Bayesian Inference

Throughout this thesis we use Bayesian methods for inference. Since this framework is essential to understanding some parts of the model description we will quickly review the basic idea of it in this section. For a thorough account the interested reader is referred to any textbook on the subject, e.g. [Gel+13].

First a small note on notation. Throughout this report we will use the common convention in Bayesian statistics that $\pi(\cdot)$ generically denotes a probability density, and $\pi(\cdot \mid \cdot)$ denotes a conditional density; the exact density that π refers to at the moment should be clear from its arguments.

Suppose we have observed some data y and are interested in doing inference about some parameter θ that controls the distribution of y. For example, y might be the observed weight of some observed subjects and θ might be the average weight in the population. These are related through the *likelihood* $\pi(y \mid \theta)$, i.e. the distribution of the observations for a known θ. This setup is the same as in the frequentist case.

In the Bayesian framework we consider θ to be a random variable. That is, we introduce a *prior distribution* $\pi(\theta)$ for θ. The prior distribution is meant to represent our knowledge of the parameter before we observe any data. Roughly speaking, the variance in the prior reflects the uncertainty in our knowledge of θ – the more sure we are, the lower the variance.

Using the prior and the likelihood we want to compute the *posterior distribution* $\pi(\theta \mid y)$; this represents our knowledge about θ after observing the data y. This computation is done by using Bayes' theorem as

$$\pi(\theta \mid y) = \frac{\pi(y \mid \theta)\,\pi(\theta)}{\pi(y)},$$

where the marginal density can be computed as $\pi(y) = \int \pi(y \mid \theta)\,\pi(\theta)\,d\theta$. Since the denominator does not depend on θ and only serves to normalize the density this formula is usually restated as

$$\pi(\theta \mid y) \propto \pi(y \mid \theta)\,\pi(\theta)$$

where $\propto$ denotes that the left hand side is a constant multiple of the right hand side. We will use this expression in several times throughout this report.

We will in fact use slightly more complicated, hierarchical, models – between the above data and parameters we add a layer of *latent variables x*. The term latent refers to the fact that these variables are not directly observed; we only observe them indirectly through y. For such models the terminology changes slightly: the term *prior* now refers to the conditional distribution $\pi(x \mid \theta)$, and to distinguish them $\pi(\theta)$ is called the *hyper-prior* (hence θ is called *hyperparameters*). The likelihood $\pi(y \mid x, \theta)$, and inference is all about studying the joint posterior

$$\pi(x, \theta \mid y) \propto \pi(y \mid x, \theta)\,\pi(x \mid \theta)\,\pi(\theta),$$

from which one can calculate quantities of interest regarding x and θ. For our models y will be hour-by-hour counts of pedestrians, x will be average hour-by-hour counts, and θ will consist of parameters controlling the variability of x and y.

The principles of Bayesian inference sketched above are rather simple. In practice however some complications appear. One is the precise choice of prior. We mostly use *random walk* priors which are described later. Another is the actual computation. It is often rather easy to write down the posterior, but it can be difficult. What is definitely difficult in most cases is to say anything of interest about the posterior (e.g. compute a mean), and one often has to resort to simulations. In this context Markov Chain Monte Carlo methods such as the Metropolis-Hastings algorithm are used. Such methods are discussed in detail in [Gel+13], and applied in a functional ANOVA in [KS10]. In this thesis we instead use an approximate computation method known as INLA [RMC09] which is discussed below in Section 3.4.

3.2   Model Description

In this section we describe the statsitcal model used. We split this up into separate (but of course connected) discussions of the likelihood, prior, and posterior distributions repsectively.

3.2.1   Data model

In one sentence, we use a *functional ANOVA negative binomial model with logarithmic link*, inspired by the models considered in [Yue+16].

Briefly, the model can be described as follows. We model the count for every street and hour as a negative binomial distribution. The average count changes from hour to hour; how it changes depends on the density type and street type of the street. Each city is given a separate, time-independent intercept term. In addition to this, we include Schools, Local markets, and Public transport as variables that affect the average for the street but not how it changes over time.

In full detail, the model is this: Let $y_k(t)$ denote the pedestrian count on street k at time (hour) t ($t = 6, 7, \ldots, 21$). We assume that this follows a negative binomial distribution:

$$\mathbf{P}\left(y_k(t) = n \mid \eta_k, s\right) = \frac{\Gamma(n+s)}{\Gamma(s)\Gamma(n+1)} \left(\frac{s}{e^{\eta_k(t)} + s}\right)^s \left(\frac{e^{\eta_k(t)}}{e^{\eta_k(t)} + s}\right)^n.$$

This distribution is related to the Poisson distribution but allows for overdispersion, i.e., that

$$\frac{\text{Var}\,(y_k(t))}{\text{E}\,[y_k(t)]} > 1.$$

One way of viewing the negative binomial distribution is that it arises as a Poisson distribution where the intensity λ is itself random and Gamma distributed; this is for example described in Section 17.2 of [Gel+13]. The parameter $s > 0$ is called a *size* parameter, and controls the amount of overdispersion compared to a Poisson distribution. Specifically, $1/s$ is the overdispersion and the weak limit as $s \to \infty$ is the Poisson distribution. While a Poisson distribution is perhaps more natural, the earlier analysis of total counts indicated that the data is overdispersed and we therefore use the negative binomial distribution. Later we will also look at the estimate of s to see whether the negative binomial distribution is necessary.

The other parameter of the negative binomial distribution, η_k, is a (time- and observation-dependent) *linear predictor*. It is related to the mean of y_k through

$$\eta_k(t) = \log \text{E}\,[y_k(t)].$$

Finally, the linear predictor is related to the covariates through

$$\eta_k(t) = X_k \phi + \tilde{\mu}(t) + \tilde{\alpha}_{i[k]}(t) + \tilde{\beta}_{j[k]}(t).$$

Here X_k is a row-vector of time-independent covariates: Schools, Local markets, public transport and intercept terms for each city, and ϕ are fixed effects coefficients. The covariates are normalized by centering and dividing by the standard deviation, producing values with mean 0 and variance 1. We also have time-dependent, or *functional*, effects $\tilde{\mu}, \tilde{\alpha}_i, \tilde{\beta}_i$. The first of these, $\tilde{\mu}$, is a baseline time effect, which is the same for all streets. The second $\tilde{\alpha}_i$, $i = 1, 2, 3, 4, 5, 6$ depends on which density type $i[k]$ that street k has. The third and final one, $\tilde{\beta}_j$, $j = 1, 2, 3, 4$ is similar but instead depends on the street type $j[k]$ of street k. To ensure identifiability we require that $\tilde{\alpha}_1 = \tilde{\beta}_1 = 0$ everywhere. We will in the sequel split up the time effects as

$$\tilde{\mu}(t) = m + \mu(t)$$
$$\tilde{\alpha}_i(t) = a_i + \alpha_i(t)$$
$$\tilde{\beta}_j(t) = b_j + \beta_j(t)$$

where m, a_i, b_j are time independent mean levels and μ, α_i, β_j are constrained to sum to zero:

$$\sum_t \mu(t) = \sum_t \alpha_i(t) = \sum_t \beta_j(t) = 0$$

for all i, j. Note that the previous identifiability constraint means that $a_1 = b_1 = 0$ and that $\alpha_1 = \beta_1 = 0$. This split is done to more clearly separate changes over time from mean levels in groups.

Note that while this model is similar to the model described in equation (2) of [Yue+16], we do not include interactions effects γ_{ij}. As Table 2.2 shows that there are rather few streets in some of the i-j-groups making fitting a separate effect rather difficult. This also makes the model more rigid, providing a guard against overfitting.

We will later also experiment with a model that includes one further random effect, changing the linear predictor to

$$\eta_k(t) = X_k\phi + \mu(t) + \alpha_{i[k]}(t) + \beta_{j[k]}(t) + \gamma_k$$

where the γ_k is a street level random intercept.

3.2.2 Prior distributions

To fully describe our model we also need to specify priors for ϕ, μ, α, β, and hyperprios for s and the other hyperparameters. The hyperpriors and the fixed effect priors are the default ones in R-INLA and the priors for the time effect are second-order random walk models. These priors are based on the discussion in [Yue+16], and the hyperpriors are the default priors in R-INLA, and are generally not so informative. The intercept is given a flat prior $\pi(\phi_0) \propto 1$ and the other fixed effects $\phi_i, i \neq 0$ are given vague normal priors $\phi_i \sim \text{Normal}(0, 1000)$ (that is, variance 1000). This also includes the fixed mean effects m, a_i, b_j for each density and street type. The usage of the same default priors for all variables is reasonable as the (continuous) covariates are normalized.

All the functional effects are given second order random walk (RW2) priors. These are described in Section 3.3; briefly the linear combinations $\mu(t_i) - 2\mu(t_{i+1}) + \mu(t_{i+2})$ are taken to be independent and normally distributed with some precision κ_μ for all i, and similarly for α and β, with each effect having its own precision parameter. The precision parameters $\kappa_\mu, \kappa_\alpha, \kappa_\beta$ are given independent and rather flat Gamma priors with shape parameter $a = 1$ and rate parameter $b = 5 \cdot 10^{-5}$, i.e.,

$$\pi(\kappa) = \frac{b^a}{\Gamma(a)} \kappa^{a-1} e^{-b\kappa},$$

for $\kappa = \kappa_\mu, \kappa_\alpha, \kappa_\beta$.

Lastly, there is the prior on s. Instead of defining a prior on s directly we let $\theta = \log s$ and give θ the prior

$$\pi(\theta) = \frac{7}{\theta^2} \frac{\left|\psi'(\theta^{-1}) - \theta\right|}{\sqrt{2\log(\theta^{-1}) - 2\psi(\theta^{-1})}} \exp\left(-7\sqrt{2\log(\theta^{-1}) - 2\psi(\theta^{-1})}\right)$$

where ψ is the digamma function; this is described in [Sim+17b]. This prior seems strange at first glance but it arises rather naturally as a *penalized complexity prior*, a type of prior introduced in [Sim+17a]. Briefly, s (and hence θ) can be thought of as a parameter that controls the model complexity compared to a Poisson distribution (which corresponds to the limit as $s \to +\infty$). The prior is constructed by requiring that increased complexity is penalized in a certain uniform manner, the details of which are described in [Sim+17a].

3.2.3 Posterior

In the language of Section 3.1 we have $y = y$, $x = (X_k\phi, \mu, \alpha, \beta)$ and $\theta = (s, \kappa_\mu, \kappa_\alpha, \kappa_\beta, \kappa_\gamma)$.

Let $\kappa = (\kappa_\mu, \kappa_\alpha, \kappa_\beta, \kappa_\gamma)$. The likelihood is

$$\pi(y \mid \eta, s) = \prod_k \pi(y_k \mid \eta_k, s) = \prod_k \prod_{t=6}^{21} \pi(y_k(t) \mid \eta_k, s).$$

The posterior is

$$\pi\left(\phi,\mu,\alpha,\beta,\gamma,s,\kappa\mid y\right) \propto \prod_k \pi\left(y_k\mid\eta_k,s\right)\cdot\pi\left(\mu\,\middle|\,\kappa_\mu\right)\pi\left(\alpha\mid\kappa_\alpha\right)\pi\left(\beta\,\middle|\,\kappa_\beta\right)\pi\left(\phi\right)\pi\left(s\right)\pi\left(\kappa\right)$$

where $\pi\left(\kappa\right) = \pi\left(\kappa_\mu\right)\pi\left(\kappa_\alpha\right)\pi\left(\kappa_\beta\right)$.

3.3 Random walk Priors

The goal of this section is to give some brief comments on the RW2 priors used for the time effects in the model. Informally, these can be seen as a higher-dimensional analogue to 'flat priors' for mean parameters in one dimension. The RW2 model is an example of an *intrinsic gaussian Markov random field*, the theory of which are detailed in [RH05, Ch.3]. A vector $x = (x_1, \ldots, x_n)$ is a *second order random walk model* with precision κ if the second order differences

$$(x_i - x_{i+1}) - (x_{i+1} - x_{i+2}) = x_i - 2x_{i+1} + x_{i+2}$$

are independent and distributed as

$$x_i - 2x_{i+1} + x_{i+2}\mid\kappa \sim \text{Normal}\left(0, \kappa^{-1}\right). \tag{3.1}$$

There are $n - 2$ such increments, giving the density

$$\pi\left(x\mid\kappa\right) \propto \kappa^{(n-2)/2}\exp\left(-\frac{\kappa}{2}\sum_{i=1}^{n-2}(x_i - 2x_{i+1} + x_{i+2})^2\right)$$

$$\propto \kappa^{(n-2)/2}\exp\left(-\frac{\kappa}{2}x^t Q x\right)$$

where matrix Q has the form

$$Q = \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & 1 \\ & & & & 1 & -2 & 1 \end{pmatrix}.$$

Note that this distribution is improper (it does not have a finite integral), as Q is rank deficient. In fact the density does not change if we add a linear trend to x. To be precise,

$$\pi\left(x_1, x_2, \ldots, x_n\mid\kappa\right) = \pi\left(x_1 + a + 1b, x_2 + a + 2b, \ldots, x_n + a + nb\mid\kappa\right)$$

for any $a, b \in \mathbb{R}$. Since it might be that there is a linear trend in pedestrian counts this type of model is appropriate for the time effects considered in this project.

An interpretation of the RW2 model can be found by considering (3.1). The normal distribution has most of its mass concentrated around 0, so this equation says that most
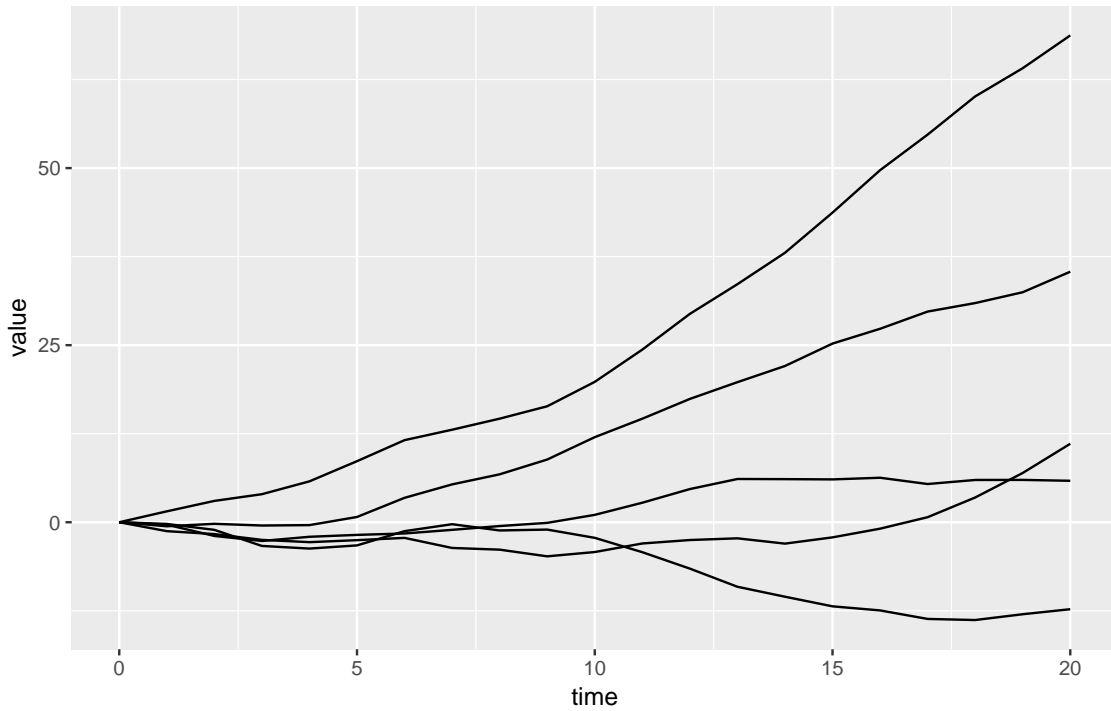
Figure 3.1: Simulated trajectories from a RW2 model with $\kappa = 1$. The RW2 model here has the constraints $x_0 = 0$ and $\mathrm{E}\left[x_{i+1} - x_i\right] = 0$ to get a proper distribution.

of the time the second difference $(x_i - x_{i+1}) - (x_{i+1} - x_{i+2})$ will be small (how small is of course controlled by κ). Since $x_i - x_{i+1}$ gives the 'trend' at i, the RW2 model says that the change in the trend is small. It should be noted that the trend itself can be arbitrarily large as long as it does not change quickly. This is also reflected by the fact that density is maximized for any perfectly linear x. The typical realization of an RW2 process therefore tends to look rather smooth.

Figure 3.1 shows simulated trajectories from a RW2 model with $\kappa = 1$. Since the RW2 model is improper the trajectories in this plot are subject to the additional constraints $x_0 = 0$ and $\mathrm{E}\left[x_{i+1} - x_i\right] = 0$, making the distribution proper. We see quite clearly their typical 'smooth-esque' behaviour.

3.4 Integrated Nested Laplace Approximation

To fit the models to data we use the integrated nested Laplace approximation (INLA) method of Rue, Martino and Chopin [RMC09]. INLA is based around approximating densities with Gaussians and using numerical integration. If $\pi(x) = C \exp(h(x))$ is a (sufficiently smooth) density with mode x^* then we can approximate it by the Gaussian density

$$\pi_G(x) = C \exp(h(x^*)) \exp\left(-\frac{1}{2}(x - x^*)^t H(x - x^*)\right)$$

where $H = -\left(\frac{\partial^2}{\partial x_i \partial x_j} h\right)(x^*)$ is the negative of the hessian matrix of h evaluated in x^*. The approximation comes from a second order Taylor expansion of h around x^* and is

exact if π is a Gaussian density.

The INLA method is applicable for *latent gaussian models*. These are hierarchical models that can be schematically written as

$$\theta \longrightarrow x = (x_1, \ldots, x_n)$$

$$x_1 \qquad x_2 \qquad \ldots \qquad x_n$$

$$y_1 \qquad y_2 \qquad \ldots \qquad y_n$$
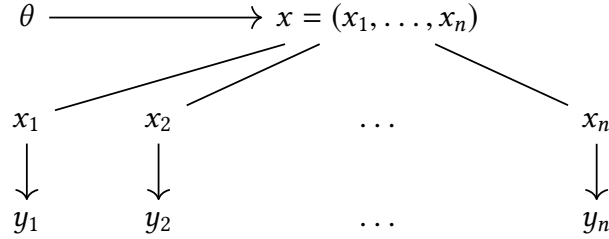
where θ is a vector of hyperparameters, x is a Gaussian vector and y_i are observations. The x are refered to as *latent variables* (hence the name *latent Gaussian models*), and x is assumed to be a GMRF. Finally we assume that observations are conditionally independent given the latent field and hyperparameters:

$$\pi\left(y \mid x, \theta\right) = \prod_i \pi\left(y_i \mid x_i, \theta\right).$$

This last assumption means that the y_i do not affect each other directly; if they affect each other they only do so through the parameters θ and latent field x.

The goal when doing inference is (as discussed in Section 3.1) to figure out the posterior distribution of θ and x. This is formally given by

$$\pi\left(x, \theta \mid y\right) \propto \pi\left(y \mid x, \theta\right) \pi\left(x \mid \theta\right) \pi\left(\theta\right)$$

but this expression can be hard to do computations with in general. What INLA does is to provide approximations of the *marginal* posterior distributions $\pi\left(\theta \mid y\right)$, $\pi\left(\theta_j \mid y\right)$ and $\pi\left(x_i \mid y\right)$ that are reasonable to compute. Note that we can write the marginals as integrals

$$\pi\left(x_i \mid y\right) = \int \pi\left(x_i \mid \theta, y\right) \pi\left(\theta \mid y\right) d\theta \tag{3.2}$$

$$\pi\left(\theta_j \mid y\right) = \int \pi\left(\theta \mid y\right) d\theta_{-j}, \tag{3.3}$$

where as usual θ_{-j} denotes all the values of θ *except* for θ_j. These integrals are however not trivial to compute. The next step is therefore to replace the densities π with more tractable approximations $\tilde{\pi}$. First we approximate the marginal posterior $\pi\left(\theta \mid y\right)$. This density can be written as

$$\pi\left(\theta \mid y\right) \propto \pi\left(\theta, y\right) = \frac{\pi\left(x, \theta, y\right) \pi\left(\theta, y\right)}{\pi\left(x, \theta, y\right)} = \frac{\pi\left(x, \theta, y\right)}{\pi\left(x \mid \theta, y\right)}$$

for any choice of x, where the proportionality is due to the missing normalizing constant $\pi\left(y\right)$. We can replace the denominator with its gaussian approximation $\pi_G\left(x \mid \theta, y\right)$ and

evaluate the expression at the mode $x^* = x^*(\theta)$ of $\pi(x \mid \theta, y)$ to get the *Laplace approximation*

$$\tilde{\pi}(\theta \mid y) \propto \left. \frac{\pi(x, \theta, y)}{\pi_G(x \mid \theta, y)} \right|_{x=x^*}$$

The reason for using this approximation instead of the direct Gaussian approximation $\pi_G(\theta \mid y)$ is that the latter works very badly when $\pi(\theta \mid y)$ is skewed and this tends to occur in practice.

With this approximation we can use (3.3) to approximate the marginal posterior for each θ_j as

$$\pi(\theta_j \mid y) \approx \tilde{\pi}(\theta_j \mid y) = \int \tilde{\pi}(\theta \mid y) \, d\theta_{-j},$$

where the integral can be computed numerically.

It remains to approximate $\pi(x_i \mid y)$. To do this we need to somehow compute $\pi(x_i \mid \theta, y)$ in the integral above in (3.2). The authors of [RMC09] describe three different approaches to approximating this, all of which are based around using Gaussian approximations, either directly or in Laplace-style quotients. These methods present a choice between accuracy and computational cost. The method used in this report is the default approach in the R-INLA library which is called *simplified Laplace approximation*. This method takes as its starting point the Laplace approximation

$$\tilde{\pi}(x_i \mid \theta, y) \propto \left. \frac{\pi(x, \theta, y)}{\pi_G(x_{-i} \mid x_i, \theta, y)} \right|_{x_{-i}=x_{-i}^*}$$

which is in turn made more computationally feasible by using series expansion and disregarding the effect of x_{-i}-elements which correspond to locations that are in some sense far from the location of x_i. The marginal $\pi(x_i \mid y)$ can then be approximated using (3.2) as

$$\pi(x_i \mid y) \approx \tilde{\pi}(x_i \mid y) = \int \tilde{\pi}(x_i \mid \theta, y) \, \tilde{\pi}(\theta \mid y) \, d\theta,$$

after which we are done with our presentation of how the INLA method works.

We conclude this section with some remarks on the properties and pros and cons of INLA compared to MCMC. The big advantage of INLA compared to an MCMC approach is speed. In addition to being fast it also has the advantage of being deterministic. This makes it easier to reason about the computation time. It also means that there is no need to carfully check for convergence of chains and to do multiple runs as is usually necessary when performing MCMC.

There are also some drawbacks to INLA. One of these is that it can only fit latent gaussian models. In this project this restriction does not cause any problems as the models considered are of that type. Another, which has more effect on this project, is that INLA only gives posterior marginals and not joint distributions. This for example makes the computation of posterior predictive distributions difficult as this really needs the joint distribution of s and μ_k which is not available. However, this problem is for the most part alleviated by the fact that R-INLA can generate samples from the (joint) posterior $\pi(x, \theta \mid y)$ from which such distributions can be approximated.

3.5 Comparing and Checking Models

The main model used in this project was described above in Section 3.2. In addition to this model a few other models were also tried for comparison. This section aims to briefly describe these models and what methods will be used for comparing the models. The goal of trying other models is twofold. First of all it makes sense to investigate whether there is some strictly better model than the main model – if there is, it makes sense to change the main model to get better predictions. Second, trying other models will give some insight into the robustness of the estimates. For example, we can see if removing the fixed effects will have any major impact on the estimates of the time effects.

We begin by describing the other models considered. With one exception, the other models can be seen as simplifications to the main where some parts of it have been removed. These simplified models are as follows:

- No Fixed Effects: same as the Main model, but without the fixed effects for schools, markets and public transport.
- No Street type effect: Does not use the street type, only the density type.
- Only mean effect: Uses neither street nor density, there is only a single time-dependent mean effect μ.
- No time effects: Does not include any of the time effects μ, α, β. The fixed type effects m, a, b are still used however Note that as the time is ignored, this is just a regular negative binomial regression.

The last additional model (random street intercepts) was alluded to in the model description, it takes the main model and then adds a (time-independent) random intercept γ_k to each street k.

To compare the models we will use so-called *proper scoring rules* [GR07]. These are a type of predictive performance measures which have the useful property that they are always minimised for the true model[1]. Scoring rules also take prediction uncertainty into account. This contrasts them with direct error measures such as RMSE or MAE that only measure the error in pointwise predictions when compared to observed values. Since we use Bayesian methods, it is natural to in this way account for more of the posterior predictive distribution and not only their means.

Unlike model performance measures such as DIC or WAIC [Gel+13, Ch. 7], scoring rules do not penalise model complexity. To avoid overfitting it is therefore paramount to use separate training and test data. We achieve this by using 10 random splits of the streets into 90% training data and 10% testing data and comparing by the average score on the test datasets for each model. The same train-test-splits where used for each model to ensure fairness.

As for the specific scores we use, the main ones are *log score* and *quasi-ignorance score* (QI score). The log score of an observation y_k^{test} is the negative of the log posterior predicitve distribution, i.e.

$$\text{logscore}_k = -\log \pi \left(y_k^{\text{test}} \mid y^{\text{train}} \right).$$

The QI score is essentially the same log score, but assuming that the posterior pre-

---

[1] Note that [GR07] uses the opposite sign convention and hence seeks to maximize the score instead.

dictive distribution is normal. It is given by

$$QI_k = \frac{\left(\mu - y^{\text{test}}\right)^2}{\sigma^2} + \log \sigma$$

where μ, σ are the mean and standard deviation of the posterior predictive distribution. This is no longer a *strictly* proper scoring rule as it gives the same score to all distributions with the same mean and variance as the true one. It is still proper however. To get a single value from the scoring rules we take the average of the scores across all of the training data.

Both of these scores are computed as MC averages of samples from the posterior distribution. To give a little more detail, the posterior predictive distribution is given by an expectation, namely

$$\pi \left(y_k^{\text{test}} \mid y^{\text{train}}\right) = \int \pi \left(y_k^{\text{test}} \mid \eta_k, s\right) \pi \left(\eta_k, s \mid y^{\text{train}}\right) d(\eta_k, s)$$
$$= \mathbf{E}_{\eta_k, s \sim \pi\left(\eta_k, s \mid y^{\text{train}}\right)} \left[\pi \left(y_k^{\text{test}} \mid \eta_k, s\right)\right].$$

This expectation can then be approximated as

$$\mathbf{E}_{\eta_k, s \sim \pi\left(\eta_k, s \mid y^{\text{train}}\right)} \left[\pi \left(y_k^{\text{test}} \mid \eta_k, s\right)\right] \approx \frac{1}{I} \sum_{i=1}^{I} \pi \left(y_k^{\text{test}} \mid \eta_k^{(i)}, s^{(i)}\right)$$

where $\eta_k^{(i)}, s^{(i)}$ are samples from the posterior $\pi \left(\eta_k, s \mid y^{\text{train}}\right)$. For this report, $I = 1000$ was used. By taking negative logarithms we get an approximation of the logscore. Similar calculations can be performed to approximate the mean and standard deviation of the predictive distribution, and hence the QI score.

We follow this with a minor comment on the choice of scoring rule. Initially the idea was to mainly use the *continuous ranked probability score* (CRPS). This is defined as

$$\text{CRPS} = \mathbf{E} \left| Y - y^{\text{test}} \right| - \frac{1}{2} \mathbf{E} \left| Y - Y' \right|$$

where Y, Y' are independent and distributed accoring to the posterior predictive distribution for y^{test}. However, this score suffers from giving more weight to parts of the predictive distribution where there is more variability. Since the variability of the negative binomial distribution scales with the mean this will give more weight to streets with large counts which is not appropriate as the goal is to get good results for all kinds of streets. Hence the more scale-invariant log score and QI score were chosen.

In addition to these scores we will also compute 95% posterior predictive intervals for the test data and check their coverage. This helps ensure that the models estimate their own uncertainty properly – too low coverage is an indication that the model underestimates its uncertainty.

Finally, we will use the *probability integral transform*, or PIT as a model check. It serves as an alternative to the normal QQ-plot in ordinary linear regression. Such QQ-plots are not directly applicable here as the theoretical posterior is not known and in all likelihood not normal. The PIT is based on the following idea from probability theory:

if X is a (continuous) random variable with cdf F, then $F(X)$ is uniformly distributed on $[0, 1]$. For an observation k the PIT is defined as

$$\mathrm{PIT}_k = \mathbf{P}\left(Y_k \leq y_k \mid y_{-k}\right)$$

where Y_k is distributed according to $\pi\left(y_k \mid y_{-k}\right)$, i.e. the predictive distribution of observation k given all other observations. The PIT values are computed as a part of model fitting in R-INLA. If the model is appropriate for the data then this predictive distribution is really the 'true' distribution for y_k, and hence PIT_k should be uniformly distributed. One can therefore asses the model by for example plotting sorted PIT values against the corresponding uniform quantiles.

There are two concerns with this. The first is that the PIT values are subject to random error so it might therefore by good to try to figure out how much they are expected to vary. The second is that the uniformity in some sense relies on having a continuous distribution – the data considered here is discrete, so it might not be that the distribution of the transformed values is uniform in this case. To handle these to issues we will also create simulated PIT values, and compare these to the observed ones. This is done by sampling new data from the posterior, refitting the model to this new data and computing the PIT values.

Chapter 4

Results and Discussion

In this section we describe and discuss the results from the modelling. We divide the results section as follows:

1. Results for the main model,
2. comparisons to other models,
3. looking at cities separately.

4.1 Main model results

We begin by showing the results for the main model. When looking at the estimates from the model, there are essentially three things to consider: the overall mean levels (i.e. the fixed effects), the structure over time (i.e. the functional effects) and the hyperparameters. We will get to each of these in order. Then we look at more 'overall model diagnostics', e.g. PIT, and checking dispersion.

4.1.1 Fixed effects

We begin by looking at the fixed effects estimates, shown in Figure 4.1. These are of less interest than the time effects but there are a couple things to consider. We see that there is a rather clear difference between the different types and Cities. The mean value for Amsterdam is largest, followed by Stockholm with London last. It is worth stating that this might just be an effect of the specific streets that where measured in each city. Regarding the 'attraction covariates' the schools seem to have very small effect, while the markets and public transport stops have larger effects. The street-level local markets effect is estimated to zero, so is probably not worth including at all. The mean effect for density type 6 also has an interval that covers zero, but since it is really part of an overall density effect it would make little sense to drop this effect. Note also that the fact that these effects might be zero should not have any bearing on the predictive ability of a model including them compared to one leaving them out, and since the goal is not to model these fixed effects it is perhaps not worth spending so much time on optimizing them. However, it is interesting to see what fixed effects affects pedestrian flow.

Overall the density fixed effects estimates seem natural considering what the types represent; it is no surprise that the typical city centre types 3 and 5 have clearly larger estimates than the other types. Type 6 (Spacious Mid-rise) has the lowest pedestrian
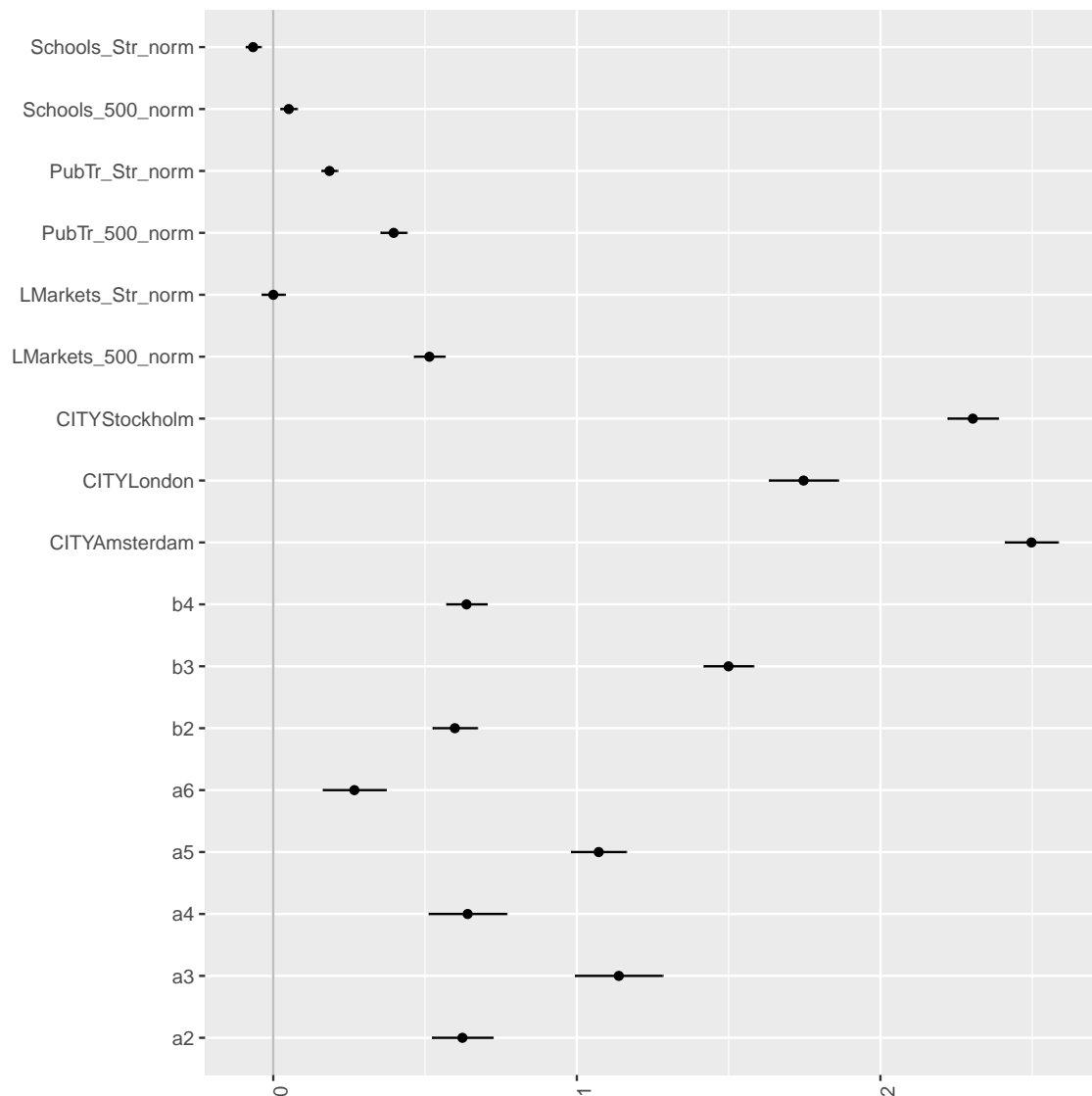
Figure 4.1: Fixed effects estimates for the main model. Points show posterior means and lines 95% credible intervals.

counts, followed by types 2 and 4 which are pretty much even. Looking back at the descriptions of the types this is natural, with these types roughly corresponding to suburban and industrial areas where not so much pedestrian traffic is to be expected. It is worth noting that for all these density types (2 through 6) the effect estimates are positive. Due to the way the model is specified type 1 is taken as a baseline, with $a_1 = 0$ by definition. This type has villa areas which probably have very low pedestrian counts and the all-positive estimates simply mean that all other types of areas have larger pedestrian counts than the villa areas.

Similarly, street type 1 which consists of background streets is taken as a baseline so that b_1 is fixed to 0, and thus the positive estimates for b_2, b_3, b_4 mean that background streets are the least travelled by pedestrians.
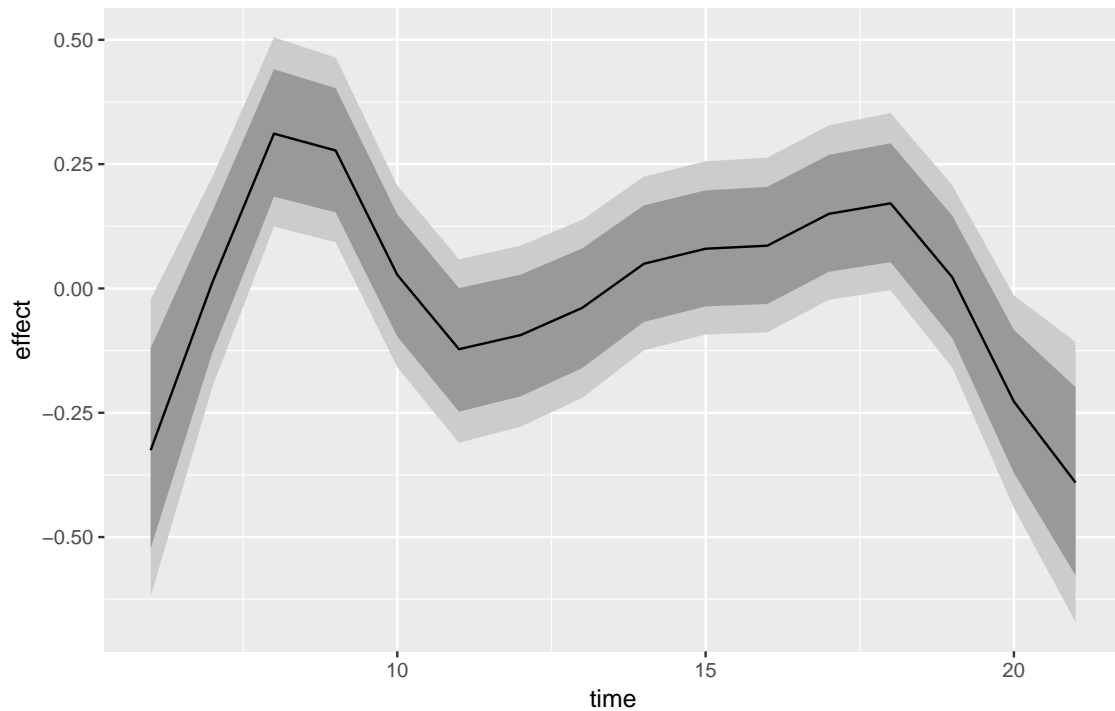
Figure 4.2: Baseline effect over time (black) with 95% pointwise (darker) and simultaneous (lighter) credible intervals.

4.1.2 Functional effects

The next point to discuss is the time effects. We will begin by looking at the effects for each density and street type one by one and then combine the effects and compare this to the data.

We begin by looking at the effects one by one. INLA directly gives pointwise (i.e. hour-wise) credible intervals; the simultaneous credible intervals were computed using the excursion method of [BL15]. Since density/street type 1 is taken as baseline, i.e. $\alpha_1 = \beta_1 = 0$, we only plot levels 2 and up of these effects.

Figure 4.2 shows the baseline effect μ. For this effect we see that there are two peaks, one in the morning and one in the evening. We also see that the effect is clearly nonzero – the simultaneous credible interval does not contain the zero function.

Figure 4.3 shows the time effects α_i for each density type. With the exception of type 6, these are all clearly negative in the early morning. Since type 1 is taken as baseline, this means that the other types will have lower average pedestrian counts in the morning compared to type 1. Looking at the credible intervals, only the one for type 6 contains the zero function, and hence the density type seems to have a real effect on the counts over time.

The time effects β_j of each street type are shown in Figure 4.4. Unlike the baseline and density type effects, these have credible intervals containing zero, meaning that there might plausibly be no real time effect of the street type. It might therefore not really be worth it to include a time effect for the street types. However, Figure 4.1 shows that the b_j are clearly nonzero so it is still helpful to include a fixed effect for the street types. This is discussed further below when comparing the models.
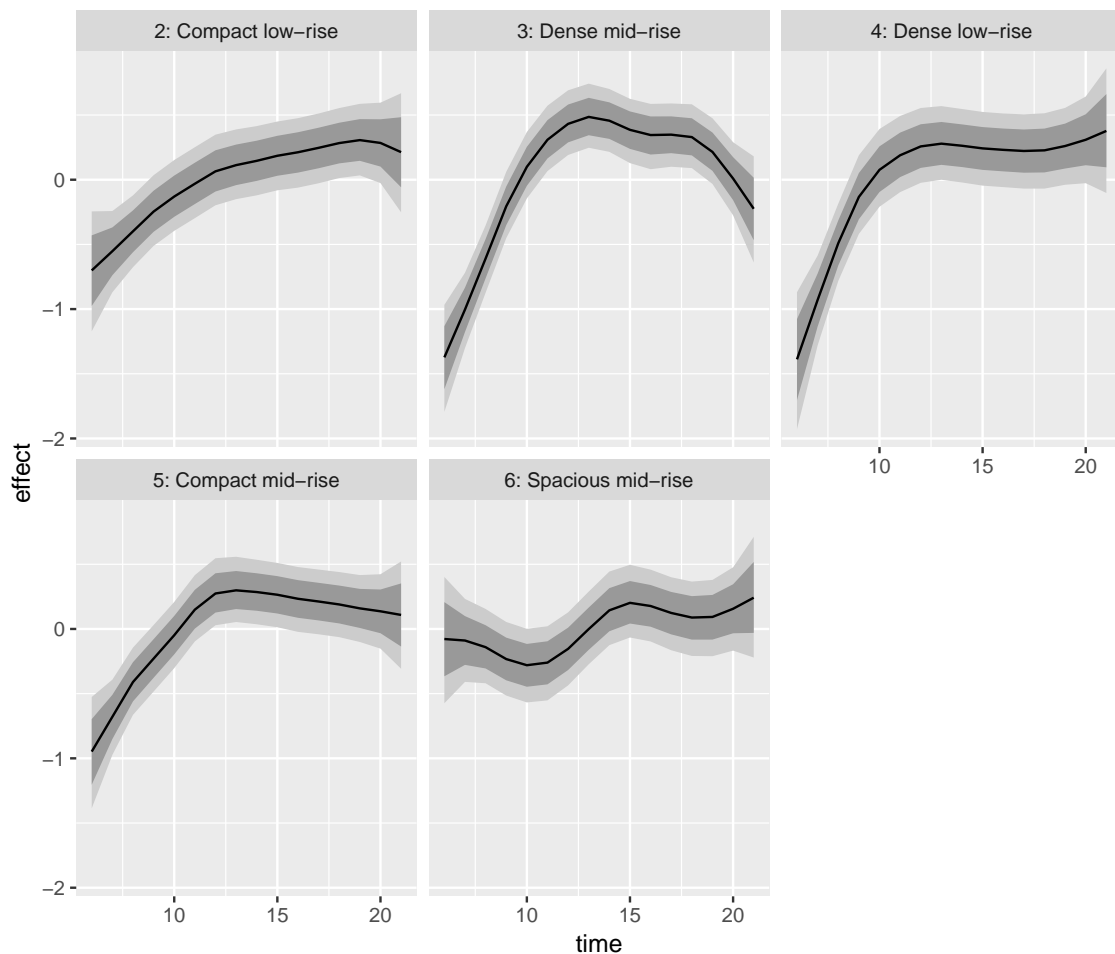
Figure 4.3: Time functions for different density types. Solid lines are posterior means, shaded regions 95% pointwise (darker) and simultaneous (lighter) intervals. Note that effects are normalized so they sum to one.

One overall trend is that the street type effects are smaller in magnitude than the density type effects and vary less. This can also be seen by looking at the precision parameters $\kappa_\mu, \kappa_\alpha, \kappa_\beta$. Figure 4.5 shows the posterior distribution of the corresponding standard deviations $1/\sqrt{\kappa_\mu}, 1/\sqrt{\kappa_\alpha}, 1/\sqrt{\kappa_\beta}$. We see directly that the standard deviation for β is smaller than the others, indicating less variability.

We next move to the main point of interest which is the actual time effect curves we get for each type combination when we combine the μ-, α-, and β-effects, and include the inverse link function. What is of course of interest is to see whether these agree with the actual data. This is investigated in Figure 4.6. This plot compares the average observed counts with the effects for each type combination. We are mainly interested in the shape of these curves as the mean values are mainly controlled by the fixed effects. To be able to focus on the shapes we normalize both the estimated and observed curves. This means that the numbers in the plot are not (directly) interpretable – only the overall shapes of the curves are. This plot also ignores the fixed effects and the random intercepts entirely,
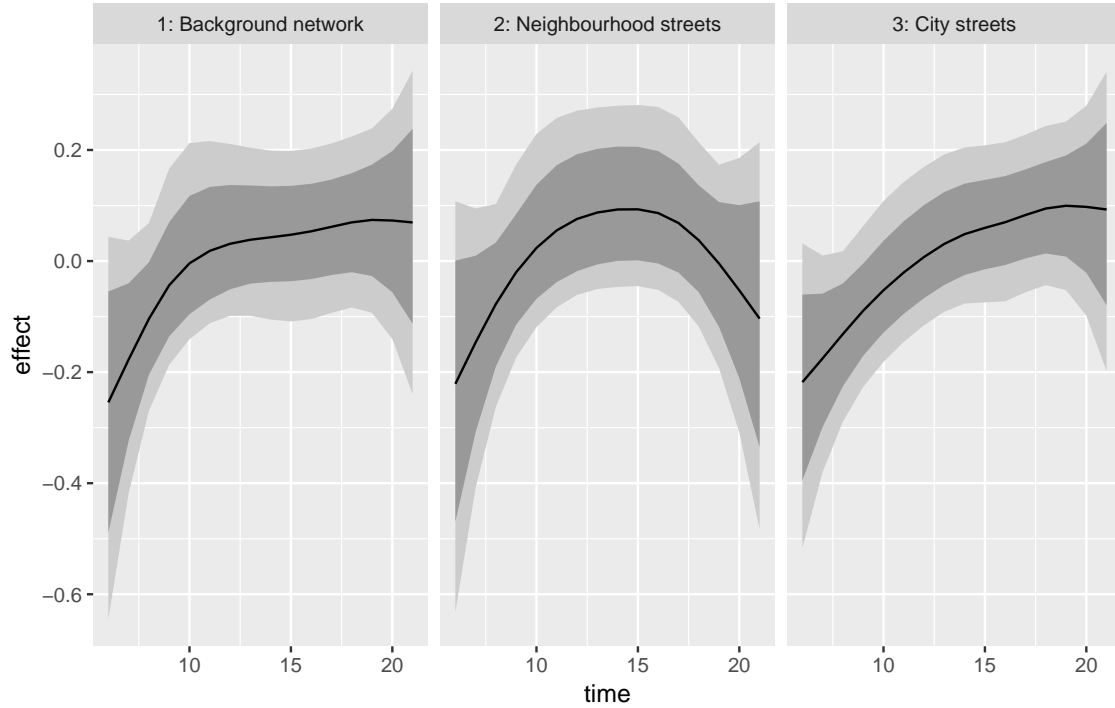
Figure 4.4: Time effects for different street types. Solid lines are posterior means, shaded regions 95% pointwise (darker) and simultaneous (lighter) credible intervals.

as these would only appear as a scaling factor,

$$\mu_k(t) = \exp\left(\eta_k(t)\right) = \exp\left(X_k\phi + \mu(t) + \alpha_{i[k]}(t) + \beta_{j[k]}(t) + \gamma_k\right)$$
$$= \exp\left(X_k\phi + \gamma_k\right)\exp\left(\mu(t) + \alpha_{i[k]}(t) + \beta_{j[k]}(t)\right),$$

and when this is normalized the corresponding factor $\exp\left(X_k\phi + \gamma_k\right)$ will disappear.

Looking at the fit of the curves we see that it is rather good overall, with some exceptions here and there. In particular, the model does seem to have some trouble capturing the more rapid changes in the data. This is consistent with the RW2 priors tendency towards smooth curves, making very rapid turns unlikely. Of course, the fitted curves are rather rigid compared to the possible variability in the data so a perfect fit is not to be expected anywhere.

We note also that this is probably not just because of the smoothness from the RW2 model. Changing to a RW1 model affected the fitted curves only very slightly. It thus seems that some of the rigidity is due to the small number of time effects we have allowed ourselves to use when compared to the possible variability of the observed data.

We can see that the behaviour in the data differs between types. Two types of patterns are worth noting. The first concerns the length of trends over time. The counts in some groups seem to be on average the same throughout the day; variation in these groups is mostly on a shorter time scale. This happens pretty much across every street type in density type 6 (spacious mid-rise). In contrast we have behaviour like in type 3 where there is an overall almost linear increase in pedestrian counts throughout the day until they drop off again quickly in the evening. Here there are trends over longer intervals of time.
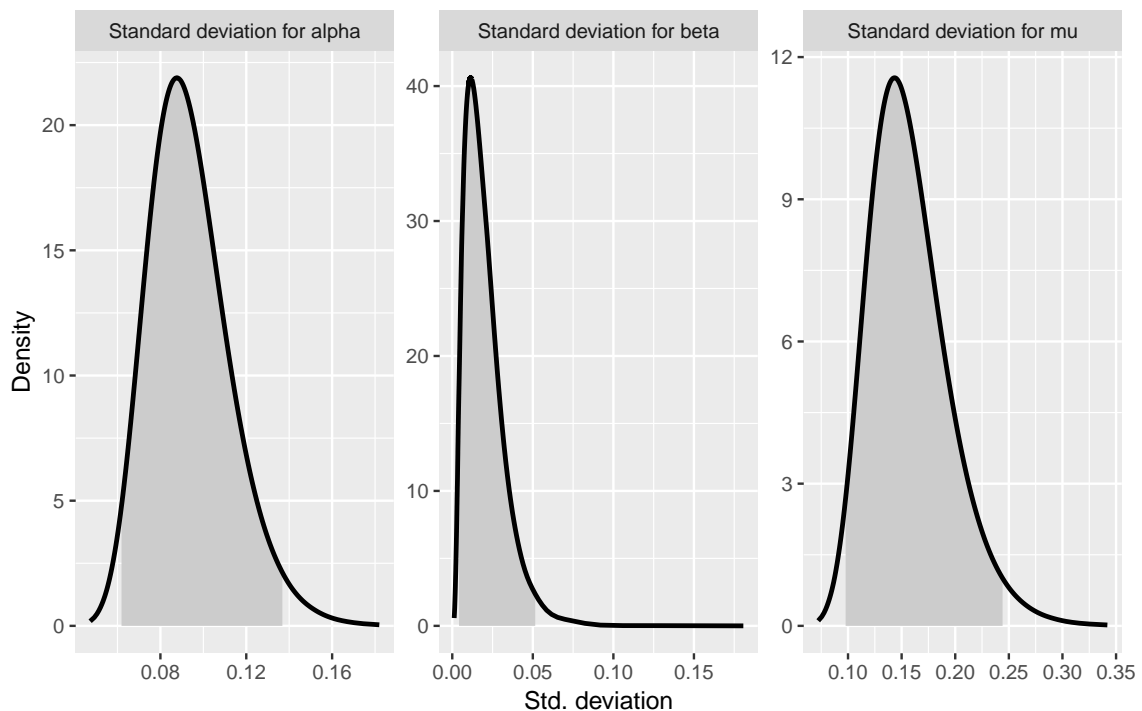
Figure 4.5: Posterior standard deviations. Shaded regions are 95% credible intervals.

The other interesting pattern is the variability in the number of peaks of the curve. Some groups, e.g. density 4 - street 4 do not really have peaks at all. Some, e.g. density 1 - street 4 has peaks in the morning and evening and dips in between. Finally some groups (for example 2-2) have three clear peaks: morning, evening and a middle peak which happen slightly *after* lunchtime.

It is possible to give reasoning behind some of these behaviours. For example, take the clear two-peak group 4-1. Streets from this group are often in suburban villa areas. In these areas people live but they (usually) do not work there. Thus, we get a peak in traffic in the morning and evening when people travel to and from work, but in between it is quiet. In more central areas, i.e. density type 3, there is often a 'lunch peak' as well – people go out to eat.

One curious phenomenon is that the model seems to have trouble capturing the morning peak at about 8-9 in street type 3, regardless of the density type. Maybe this is because the peaks really fall at slightly different times but seems to be consistent under-estimation going on. Looking back at the distribution of types in Table 2.2 street type 3 is the least common, but this should not necessarily give bad estimates across this type, especially not in such a consistent manner.

Comparing the estimates to the distribution of types also gives some overall results. The estimates pretty much agree with data in all the groups with 5% or above of the data (all of street type 1, and density types 2,3 in street type 4), perhaps with the exception of density 4-street 1 where the estimate has a spurious morning peak.
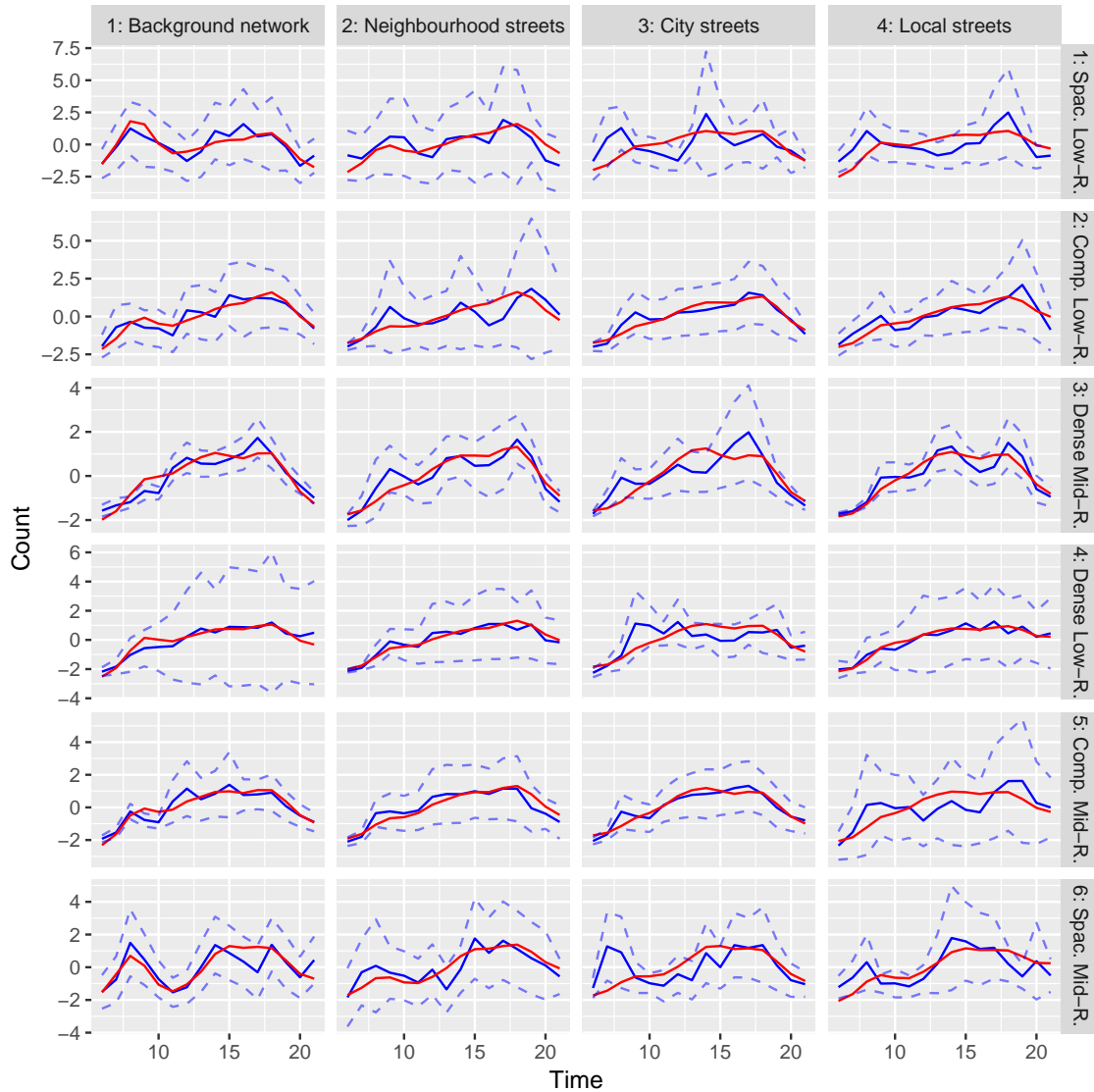
Figure 4.6: Normalized true counts (blue, solid) and pointwise 95% CIs (blue, dashed) vs normalized estimated counts (Red), split by density type (rows) and street type (columns). Data from all cities. The normalization consists of subtracting the mean and dividing by the standard deviation. This is done separately for the mean counts and the mean effects. The CI width is divided by the same standard deviation as the mean estimate.

4.1.3 Model checks

In Figure 4.7 we see the posterior overdispersion (that is, the posterior of $1/s$ where s is the size parameter). The overdispersion is with high likelihood (about 0.95) between 1.51 and 1.6, so it is clearly nonzero. Thus the negative binomial distribution (or some other overdispersion adjustment) is thus necessary, and Poisson regression is not appropriate for this data. Hence we can not simplify the model and use a Poisson regression without undermining the integrity of the model.
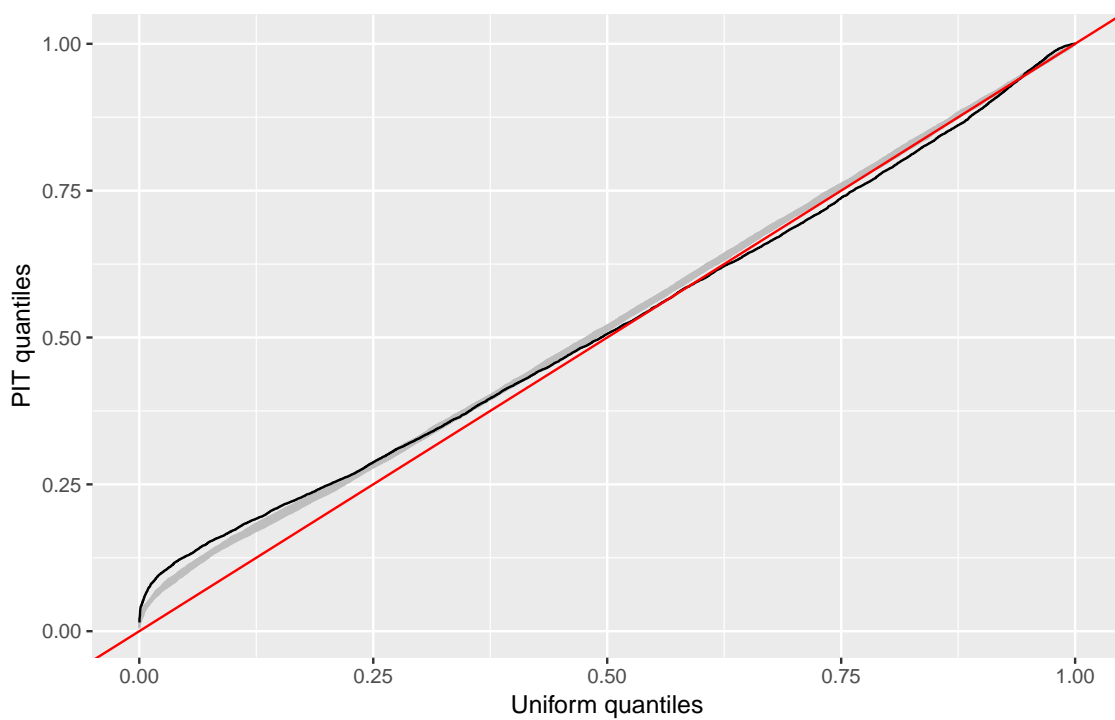
Figure 4.8 shows sorted PIT values compared to uniform quantiles, with PIT on 20 simulated datasets in grey.

The PIT values of the fit do not seem to be uniformly distributed. However, the values on the data do not look so different from the simulated values, and these do not look uniform either. Thus, the conclusion from this plot is twofold. The first is that (as was suspected) we do not actually get uniformly distributed PIT values in this case. The second is that the model seems appropriate for the data – the behaviour of the PIT values on the observed data is similar to the values for simulated data with the assumed distribution.
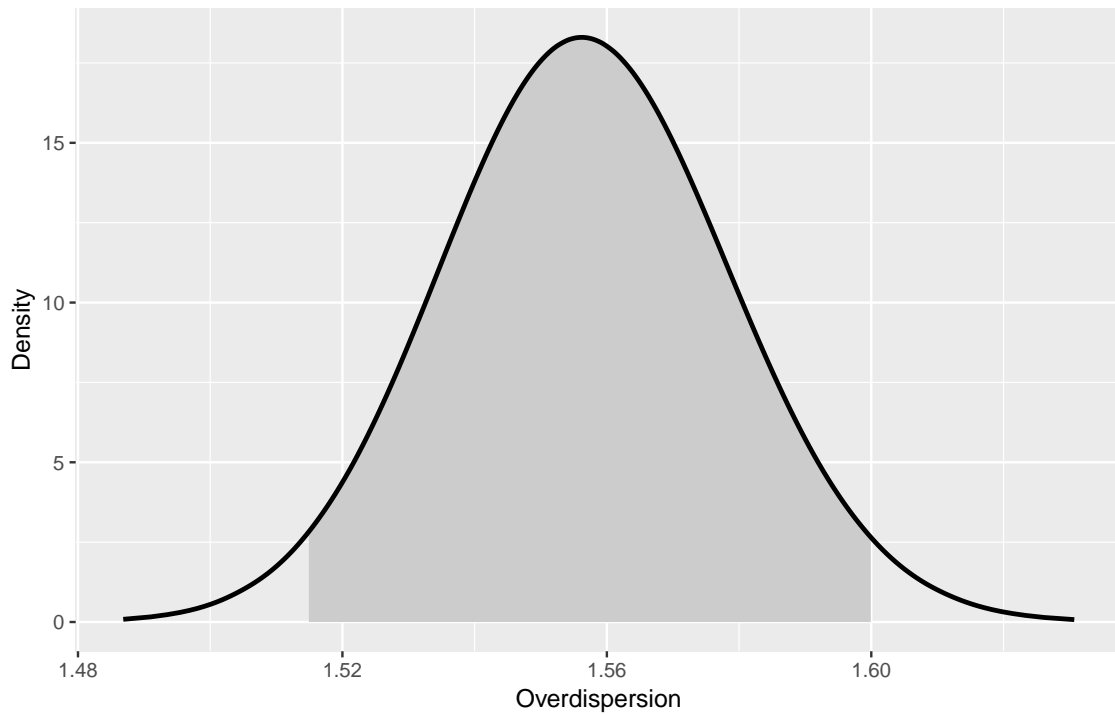
Figure 4.7: Posterior distribution of overdispersion. The shaded region shows a 95% credible interval.

Figure 4.8: PIT-QQ-plot: Sorted PIT values versus uniform quantiles (black); 20 simulated runs in grey.

4.2 Model Comparisons

We next compare different models. The chief model comparison results are shown in Table 4.1.

This table displays the test data scores of the models considered, and some further summary statistics (i.e. predictive interval coverage, RMSE and MAE). What can be seen from this is that the best models seem to be the main model and the model with no street type time effects. These models perform similarly with respect to both scores used (logarithmic and QI) and outperform the others, with the exception of the random intercept model. The random intercept model actually has the best performance with respect to QI score, and logarithmic score performance that is not not far behind that of the main model and the model without street type effects. It does however have markedly worse performance with respect to CRPS. Thur, the 'safer bet' is probably to go with the main model, but models with some form of random intercept can be worth investigating further.

Thus, the conclusions regarding model selection is that the main model is a reasonable choice but that the street type time effects are perhaps not strictly needed, as was alluded to above when discussing their estimates. The scores indicate that the other parts of the model (the fixed effects, the other time effects etc.) do improve the accuracy and should not be dropped without good reason.

The predictive interval coverage for all models is close to 95% so it seems that the uncertainty is estimated correctly in all cases.

Finally, the time effect estimate comparisons of Figures 4.9, 4.11, and 4.10 show that the model is reasonably robust. The removal of fixed effects and the inclusion of random intercepts do not change the time effects greatly, perhaps changing the street type effects a bit more than the density type effects. This indicates that the exact choice of model is not very important if the main interest is in time effect estimates in the vein of Figure 4.6. We can therefore quite safely reason about these based only on the results of the main model.

| Model	Log Score	QI Score	CRPS	Coverage (%)	MAE	RMSE
Main	**4.43**	4.22	6.74	95.9	**13.89**	124.1
No fixed effects	4.48	4.62	7.01	96.1	16.89	123.1
No street type	**4.43**	4.24	6.71	96	14.01	124.2
Only mean effect	4.44	4.37	**6.67**	95.7	14.66	**120.6**
No time effects	4.48	4.34	7.14	95.9	14.78	126.5
Street random intercept	4.45	**4.16**	7.42	95.6	14.09	133

Table 4.1: Average scores and other summary statistics on test data across 10 random train-test splits. The same splits were used for all models. MAE and RMSE are computed using posterior means as predictions.
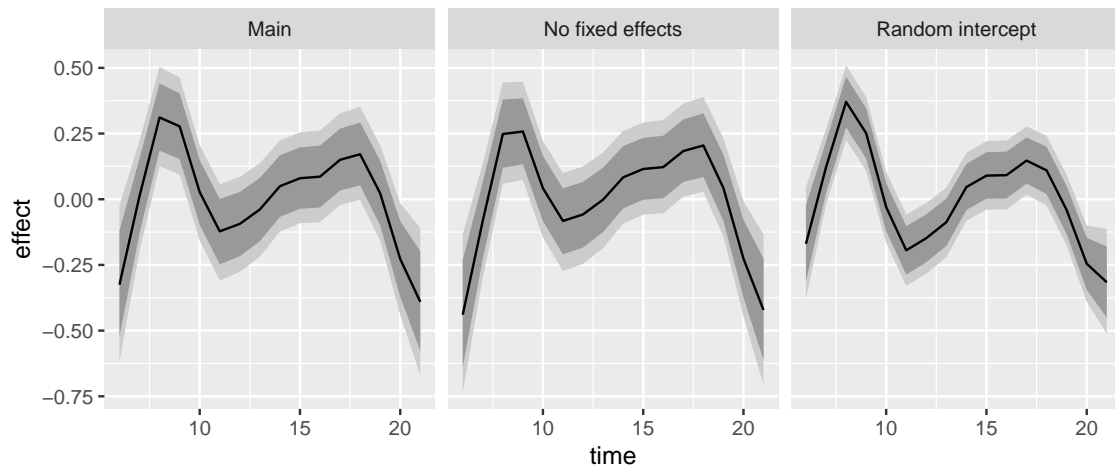
Figure 4.9: Estimated mean effect μ for different models. The figure shows the posterior mean (solid line), and 95% pointwise (dark grey) and simultaneous (light grey) credible intervals.
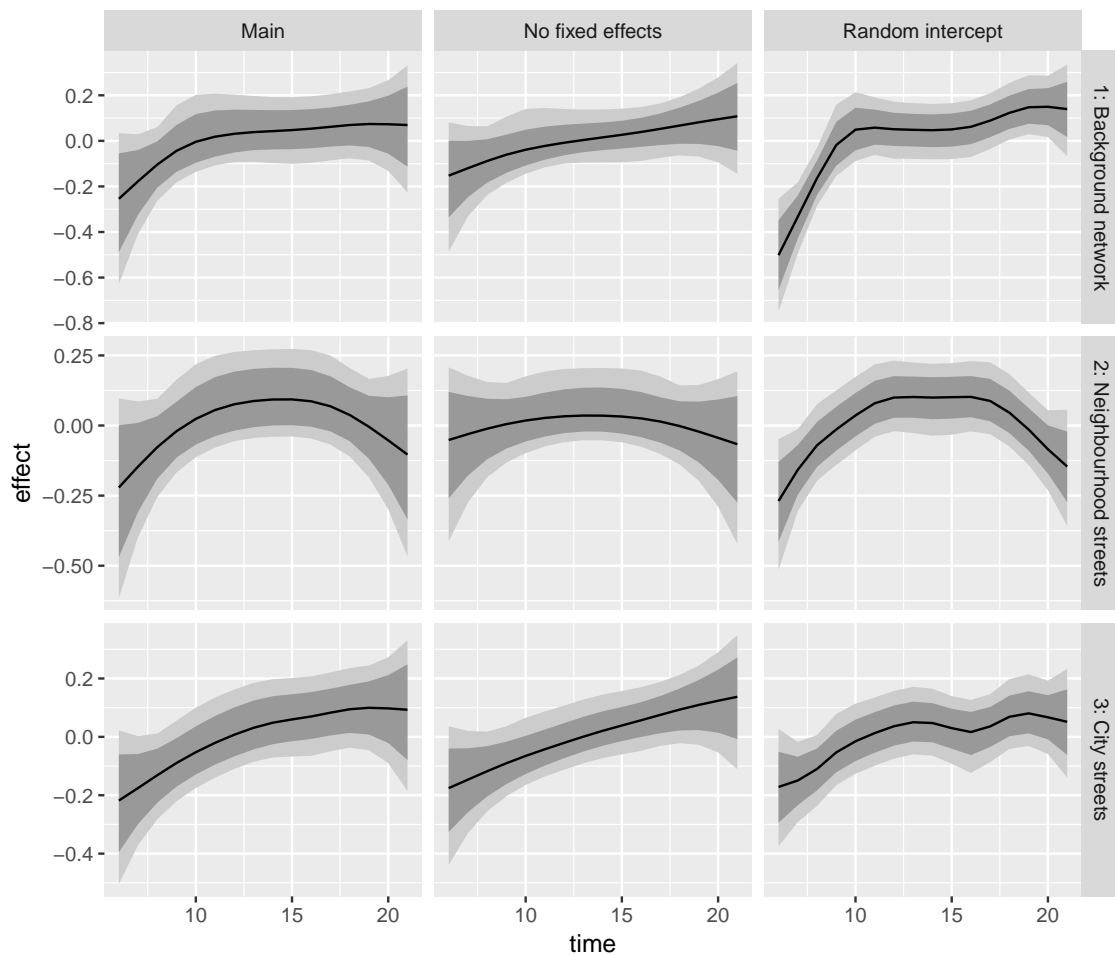
Figure 4.10: Estimated Street type effects β_j for different models. The figure shows the posterior mean (solid line), and 95% pointwise (dark grey) and simultaneous (light grey) credible intervals.
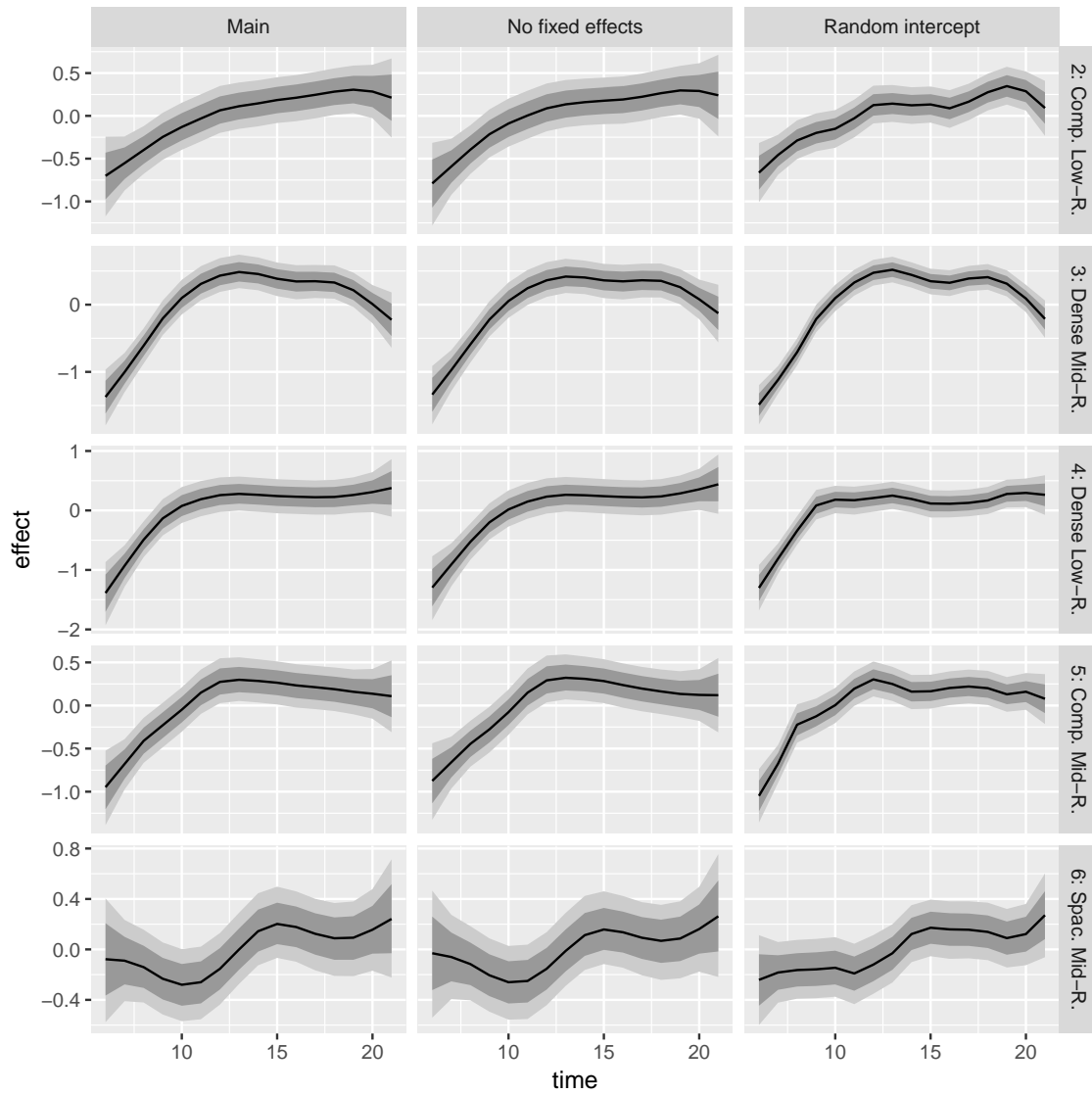
Figure 4.11: Estimated density type effects α_i for different models. The figure shows the posterior mean (solid line), and 95% pointwise (dark grey) and simultaneous (light grey) credible intervals.

4.3 Profiles for individual cities

As a last result we fit the main model to one city at a time and plot the resulting time profiles. These are shown for Stockholm in Figure 4.12, Amsterdam in Figure 4.13, and London in Figure 4.14. Due to the smaller datasets some groups are missing completely, and some have only one street in them.

The peaks are less obvious in Amsterdam, and in some cases (density type 3) more clear in London. Stockholm has no streets in density type 4, and the model does not seem to work so well for density type 1 in this city. The overall fit is not so good in Stockholm but is arguably better in Amsterdam and London than the fit to all cities simultaneously. This is natural if the shape over time of the data differs between cities which (considering the data curves in the figures) seems to be the case. When fitting the model to one city at a time there is also less data for the model to adapt to, while the 'degrees of freedom' remain the same. So in this sense it is also easier for the model to fit to one city at a time.

We can also compare the type effect estimates when fitting the model to cities one at a time. Figure 4.15 compares the density type effects α_i between using all cities, just Amsterdam and just London (Stockholm is not included due to the lack of density type 4 streets). We can see that there is a rather clear difference between the effect for Amsterdam and London by looking at types 4 and 5. The Amsterdam and London estimates also differ from the estimates when using all cities, both in uncertainty (which is to be expected) and shape. Similar behaviour happens for the baseline and street effects, though we do not show them here. This is an additional indication that there is a difference between the cities that is naturally not captured well by the main model.

In the light of these remarks a natural extension of the model to investigate is one where we let the time effects vary by city as well. However, the inclusion of city-by-city variation in this manner impedes the generalisability of the model to cities not in the dataset so if this is to be done one has to make sure that this is not at odds with the modelling goals.
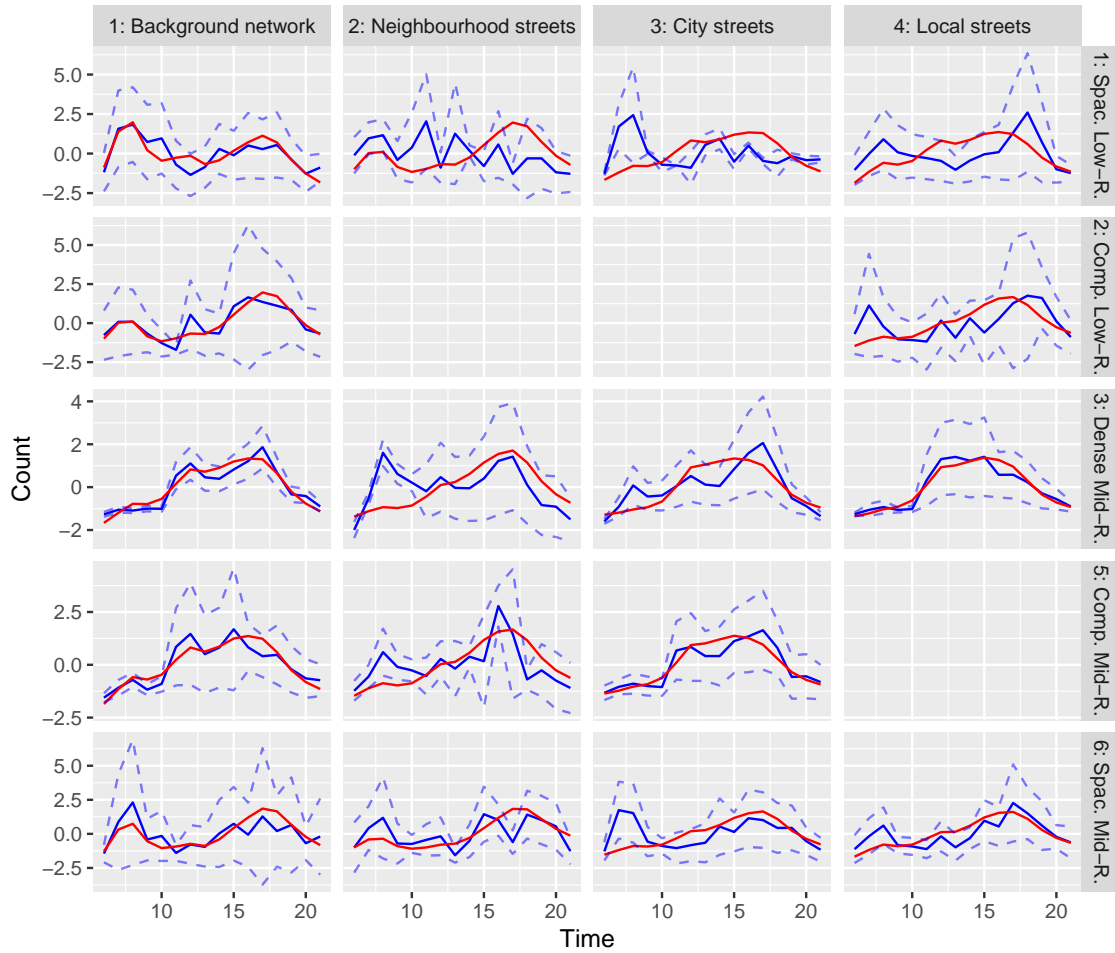
Figure 4.12: Normalized true counts (blue, solid) and pointwise 95% CIs (blue, dashed) vs normalized estimated counts (Red), split by density type (rows) and street type (columns). Data from Stockholm. The normalization is as in Figure 4.6. Note that there are no streets in density type 4.
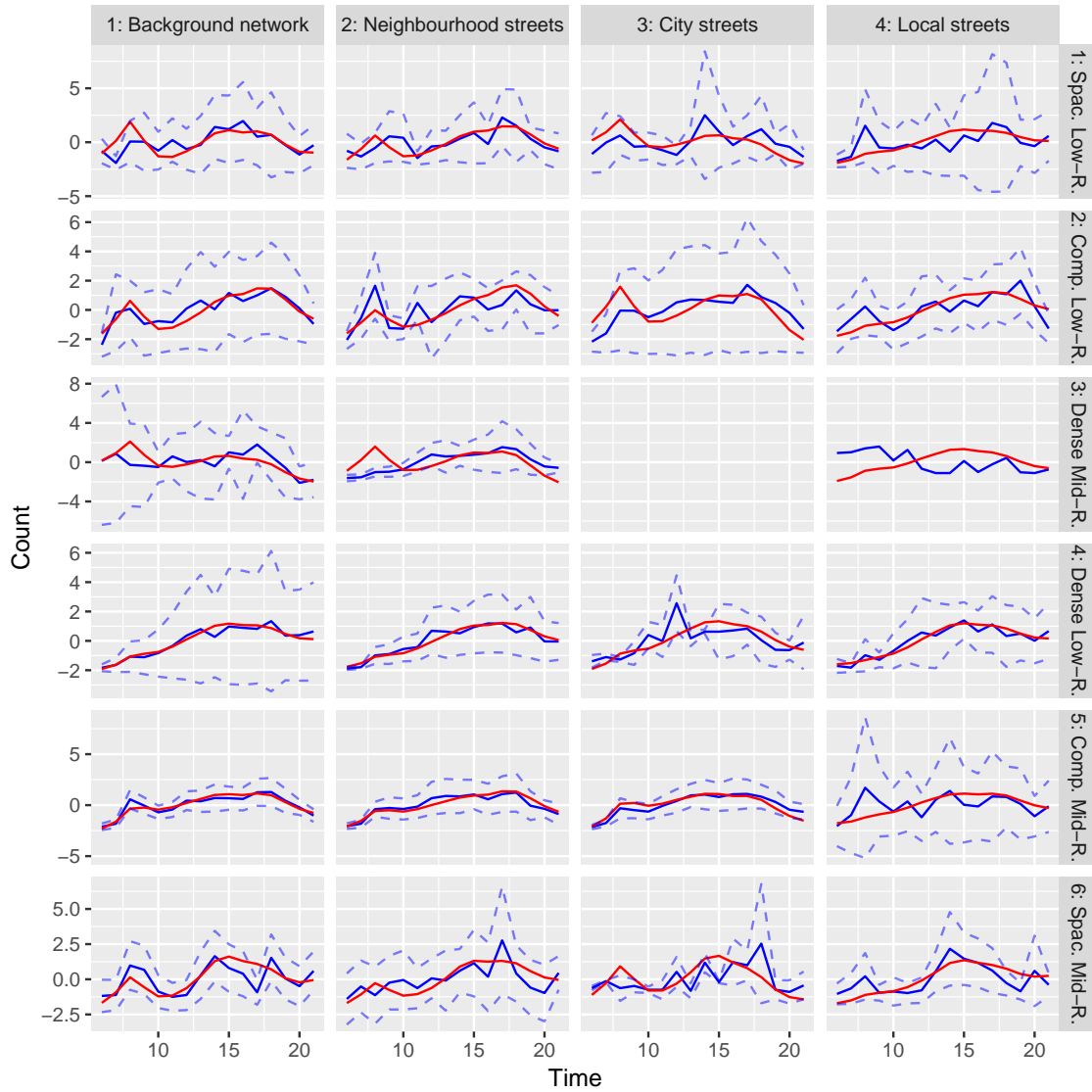
Figure 4.13: Normalized true counts (blue, solid) and pointwise 95% CIs (blue, dashed) vs normalized estimated counts (Red), split by density type (rows) and street type (columns). Data from Amsterdam. The normalization is as in Figure 4.6
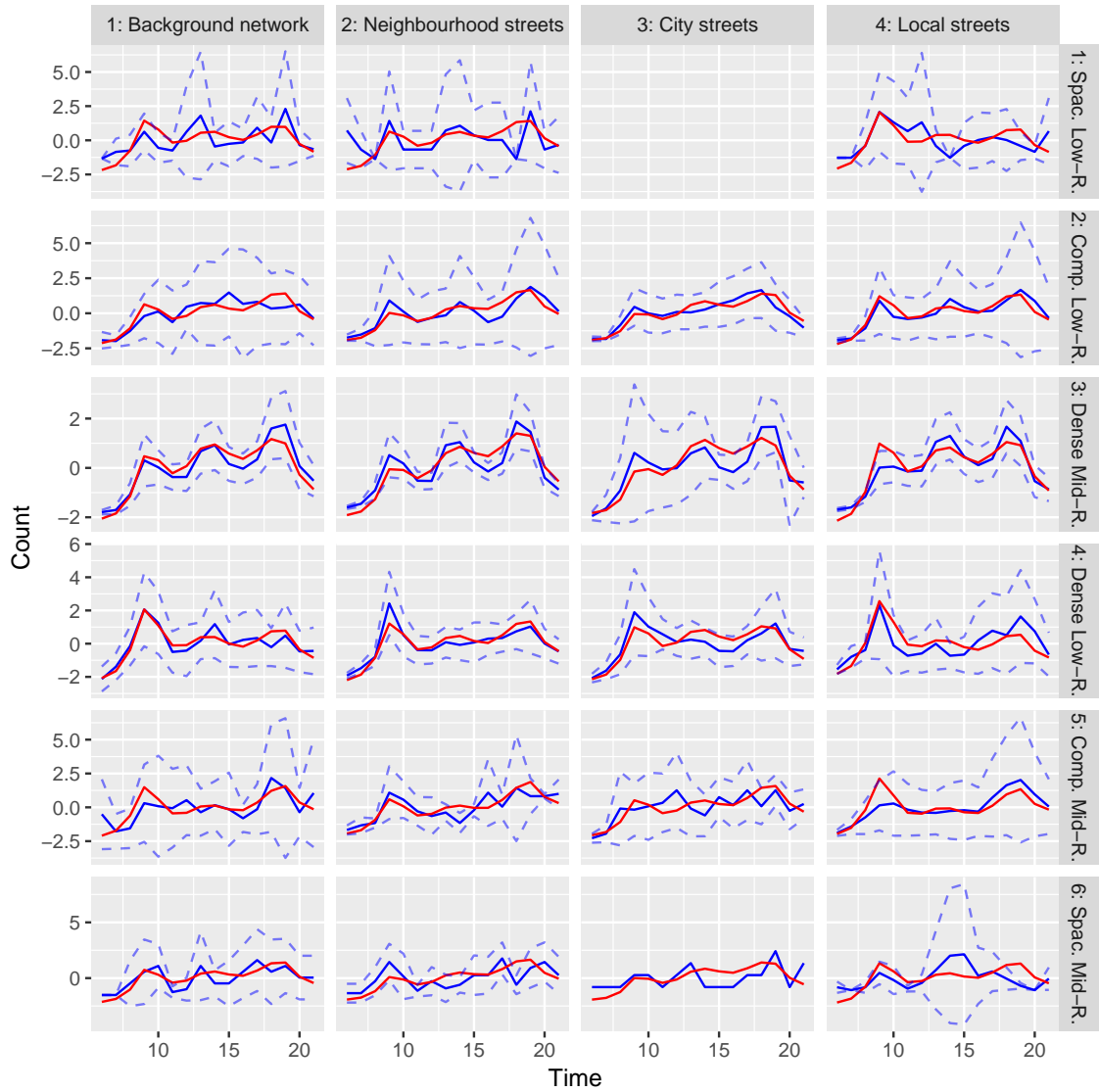
Figure 4.14: Normalized true counts (blue, solid) and pointwise 95% CIs (blue, dashed) vs normalized estimated counts (Red), split by density type (rows) and street type (columns). Data from London. The normalization is as in Figure 4.6
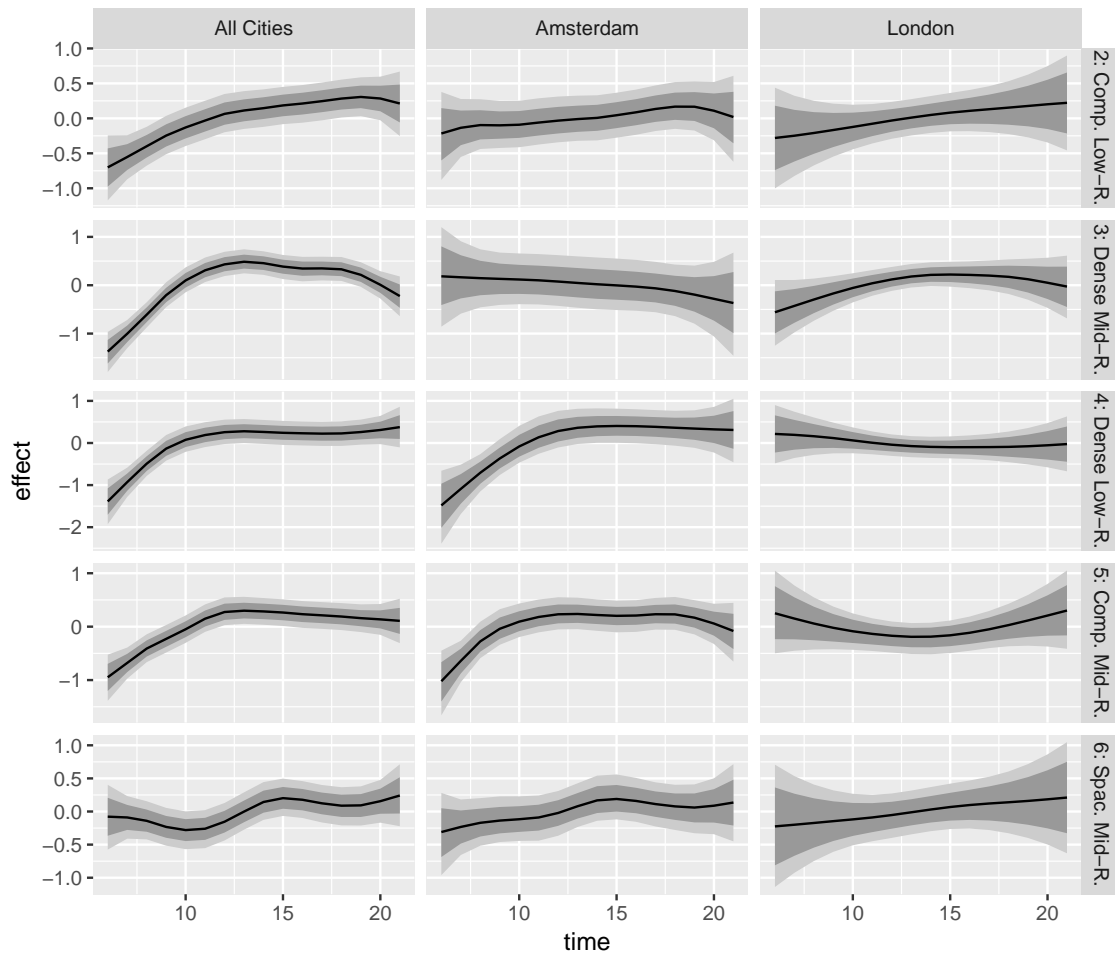
Figure 4.15: Density type effects (α_i) for the main model compared to models fitted separately to Amsterdam and London. Lines are posterior means, shaded regions are 95% pointwise (dark grey) and simultaneous (light grey) credible intervals.

Chapter 5

Conclusions and Possible Extensions

In this report we have modelled pedestrian counts over time in three European cities. We have seen that functional ANOVA based on the types described in [Ber+17] is rather successful as a method for this, giving estimates in good agreement with the data. This rather simple model manages to capture even the time dependence rather well; especially considering how simple the categorical covariates are. This bodes well for the earlier mentioned usage of these models in the planning of urban developments – the types have rather simple interpretations in terms of what types of buildings they correspond to, and thus predicting hour-by-hour pedestrian flows based on what is being built should be possible. There are also plans for future projects incorporating this type of analysis in actual planning of urband developments.

Throughout this project there has been many wishes, thoughts, and ideas that never made it into this report. This has been for different reasons – lack of data, lack of time, or simply because they seemed to show little promise after doing some initial experiments. While the last category is perhaps not so important, some ideas from the first two can be worth divulging.

The obvious one is the eternal wish in all statistical projects: that there would be *more* data available, both in terms of sheer amount and in terms of the number of explanatory variables. For an example of the latter, it might be interesting to look at weather data in conjunction with the data already used. It is quite natural that the number of pedestrians is correlated with weather conditions such as rain. It would also be interesting to have data for more hours of the day, i.e. data through the night and not just from 6 to 21. With such data it could also be possible to leverage the cyclic structure (i.e. 24=00) to get estimates that unlike ours do not have increasing uncertainty towards the edges.

Another idea is to look into using the spatial structure in the data. The random intercept model (with its inconclusive results) is a first primitive foray into this direction but more complex ideas are possible. It did not make it into this report but models using Matérn random fields based on the distance between streets were tried in this project, with computations using the SPDE methods of [LRL11]. They did however not seem to give much better results than the plain random intercepts and they greatly increased computation time. This might be an indication that a spatial model needs to work directly with the road network topology rather than distances to work well. It might be possible to leverage 'Kirchoff's law' (the number of people that enter an intersection must equal the number that exit it) to get better fitting models. This is however com-

plicated by the fact that the data features isolated clusters of streets, so pedestrians can enter and exit the data, giving leakage. This ties back to the wish for more data: if there was data on every single street this would not happen. Overall this seems to be a non-trivial problem that can yield interesting results if investigated further.

As a final 'idea' it would be nice to work more on visualising the results. In the introduction we mentioned a web GUI where the time-independent results have been visualised, and also that there has not yet been time to incorporate the results of this thesis into the GUI. This was originally part of the plan for this project but the focus ended up shifting more towards the modelling itself. Clear visualisation is often very helpful when developing statistical models and the modelling here is unlikely to be an exception. In particular, the GUI, which includes visualisation on maps of the cities, might be very helpful if one wants to develop proper spatial models for this and related data. Some difficulty arises from trying to simultaneously visualise the data and model over both time and space, but this difficulty might be worth dealing with to get a better understanding of the data.

However, we have reached the end of this masters thesis, and so all these ideas are perhaps best left for other theses.

Bibliography

[Ber+]   Meta Berghauser Pont, David Bolin, Erik Håkansson, Oscar Ivarsson, Gi-
            anna Stavroulaki, and Vilhelm Verendel. *STEPFLOW Web Interface*. URL:
            `http://129.16.20.138:3838/stepflow/stepflow/`.

[Ber+17]   Meta Berghauser Pont, Gianna Stavroulaki, Jorge Gil, Lars Marcus, Miguel
            Serra, Birgit Hausleitner, Jesper Olsson, Ehsan Abshirini, and Ashley Dhanani.
            "Quantitative Comparison Of Cities: Distribution of street and building types
            based on density and centrality measures". In: *11th International Space Syn-
            tax Symposium*. Instituto Superior Técnico, Departamento de Engenharia
            Civil, Arquitetura e Georrecursos, Portugal, 2017.

[BL15]   David Bolin and Finn Lindgren. "Excursion and contour uncertainty regions
            for latent Gaussian models". In: *Journal of the Royal Statistical Society, Series
            B (Statistical Methodology)* 77.1 (Jan. 2015), pp. 85–106. ISSN: 1369-7412. DOI:
            `10.1111/rssb.12055`.

[Gel+13]   Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Ve-
            htari, and Donald B. Rubin. *Bayesian Data analysis*. 3rd ed. Chapman and
            Hall/CRC, 2013. ISBN: 9781439840955.

[GR07]   Tilmann Gneiting and Adrian E Raftery. "Strictly Proper Scoring Rules, Pre-
            diction, and Estimation". In: *Journal of the American Statistical Association*
            102.477 (2007), pp. 359–378. DOI: `10.1198/016214506000001437`.

[KS10]   Cari G. Kaufman and Stephan R. Sain. "Bayesian functional ANOVA model-
            ing using Gaussian process prior distributions". In: *Bayesian Anal.* 5.1 (Mar.
            2010), pp. 123–149. DOI: `10.1214/10-BA505`.

[LRL11]   Finn Lindgren, Havard Rue, and Johan Lindström. "An explicit link between
            Gaussian fields and Gaussian Markov random fields: the stochastic partial
            differential equation approach". In: Journal of the Royal Statistical Society.
            Series B: Statistical Methodology 73.4 (2011), pp. 423–498. ISSN: 1369-7412.
            DOI: `10.1111/j.1467-9868.2011.00777.x`.

[R C18]   R Core Team. *R: A Language and Environment for Statistical Computing*. R
            Foundation for Statistical Computing. Vienna, Austria, 2018. URL: `https:
            //www.R-project.org/`.

[RH05]   Håvard Rue and Leonhard Held. *Gaussian Markov Random Fields*. Chapman
            and Hall/CRC, 2005.

[RMC09]   Håvard Rue, Sara Martino, and Nicolas Chopin. "Approximate Bayesian in-
            ference for latent Gaussian models by using integrated nested Laplace ap-
            proximations". In: *Journal of the Royal Statistical Society. Series B (Statistical
            Methodology)* 71.2 (2009), pp. 319–392.

[Sim+17a]   Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors". In: *Statist. Sci.* 32.1 (Feb. 2017), pp. 1–28. DOI: 10.1214/16-STS576.

[Sim+17b]   Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. "You Just Keep on Pushing My Love over the Borderline: A Rejoinder". In: *Statist. Sci.* 32.1 (Feb. 2017), pp. 44–46. DOI: 10.1214/17-STS576REJ.

[Sta+19]   Gianna Stavroulaki, David Bolin, Meta Berghauser Pont, Lars Marcus, and Erik Håkansson. "Statistical Modelling and Analysis of Big Data on Pedestrian Movement". In: *12th International Space Syntax Symposium.* To appear. 2019.

[Yue+16]   Yu (Ryan) Yue, David Bolin, Håvard Rue, and Xiao-Feng Wang. "Bayesian Generalized Two-way ANOVA Modeling for Functional Data Using INLA". In: *Statistica Sinica* (2016). DOI: 10.5705/ss.202016.0055.

Appendix A

Some notes on the total count analysis

In this Appendix we give some notes on the total counts results from the analysis performed in [Sta+19] and preliminary analyses leading up to that. We do not give any detailed results in this section; we instead mention a few insights from the earlier analyses that carried over to the time analyses used here.

The main distinguishing element in these analyses compared to the current report is the outcome variable. In the current report, the focus is on the hour-by-hour pedestrian flow intensity while in previous analyses only aimed at predicting total pedestrian counts throughout the entire day. Adding up the values over the entire day has the effect of increasing the values, which makes a continuous approximation more reasonable. The analysis in [Sta+19] shows some success transforming the counts as $\log(1 + y)$ and then applying ordinary linear regression, achieving comparable performance to negative binomial regression. With the lower counts we get wehn we split them up by hour such a model is not really reasonable however.

There are also differences in the covariates used. In the functional ANOVA method used for the time analysis the most important covariates are the density and street types, which are categorical. As was mentioned above, these are based on clustering of continuous variables. In the total counts analysis the types are not used. Instead (some of) the clustered continuous variables are used as covariates themselves. This makes the model a more typical GLM regression even in the negative binomial case. It was noted that the variables used when creating the density types were significant in the model. Working with these continuous covariates also showed that it was necessary to normalize them, especially so as the same priors are used for all fixed effects.

The final normal and negative binomial models used in the total counts analysis were preceded by attempts at using Poisson. The Poisson models showed indications of rather strong overdispersion which lead to the negative binomial models. This then carried over to the analysis performed in this report. One might have thought of retrying a Poisson model, but as the posterior overdispersion estimates (Figure 4.7) show the negative binomial is also justified by the data for the time-based models considered here.

Finally, the idea of using PIT as a major model checking tool arose in connection with the total counts analysis. For the log-transformed normal models, normal QQ-plots were useful as a model checking tool. They were however not directly applicable

to the negative binomial models as the distribution of a residual from such a model is not known. This lead to the idea of using PIT as an alternative and thus get something with a known (uniform) distribution.