Working paper

2021-10

Statistics and Econometrics
ISSN 2387-0303

A Bayesian Spatio-temporal model for predicting passengers' occupancy at Beijing Metro

Stefano Cabras, Flor Sunhe

uc3m | Universidad Carlos III de Madrid

A Bayesian Spatio-temporal model for predicting passengers' occupancy at Beijing Metro

Stefano Cabras,
Department of Statistics, University Carlos III of Madrid
and
Sun He,
Beijing Metro Group Ltd (Beijing, China)

December 15, 2021

Abstract

This work focuses on predicting metro passenger flow at Beijing Metro stations and assessing uncertainty using a Bayesian Spatio-temporal model. Forecasting is essential for Metro operation management, such as automatically adjusting train operation diagrams or crowd regulation planning measures. Different from another approach, the proposed model can provide prediction uncertainty conditionally on available data, a critical feature that makes this algorithm different from usual machine learning prediction algorithms. The Bayesian Spatio-temporal model for areal Poisson counts includes random effects for stations and days. The fitted model on a test set provides a prediction accuracy that meets the standards of the Beijing Metro enterprise.

Keywords: Bayesian modelling; Integrated Nested Laplace Approximation; Spatio-temporal modelling; Poisson Counts

1 Introduction

For some Metro operation companies and the government department of transportation, it is imperative about knowing the accurate metro passenger flow daily and hourly. Almost all metropolises have permanent and persistent traffic problems as they spend much fiscal revenue on the construction of transportation infrastructure while only getting back a few effects. On the other hand, traffic issues are the most dangerous problems happening in parts of each city as this causes accidents. Developing a traffic diversion plan also depends on the prediction of daily passenger flow and passenger range. Therefore, the observed past data is essential, but the future data affects whether a particular traffic plan would be suitable or not. So getting an accurate prediction of each station each day is necessary for the metropolis's governments. Traffic plans are also adopted based on the uncertainty around their effects. This uncertainty is intrinsic related to that predicting traffic flow. It will be possible to program train traffic and the needs of new metro stations only given the uncertainty of investments returns. Such uncertainty is, in turn, a function of traffic flow prediction uncertainty. A good forecast process is such if it can reasonably provide the uncertainty about the forecast. The uncertainty, conditional to the sample, is available only under a Bayesian modelling approach.

The real problem at hand requires models for count series. Davis et al. (2021) contains a review of the current literature on models for count series, including the Bayesian one. Specifically, the proposed model belongs to the class of level correlated models (see Ma et al. (2008); Chib & Winkelmann (2001); Aitchison & Ho (1989)) illustrated in Section 6.2.1. in Davis et al. (2021). To the best of our knowledge, we do not know if there is any literature regarding using these models for estimating passenger train traffic in a Metro station network.

The core of this approach is the predictive posterior distribution of passenger flow and model parameters that explain the traffic flow. Specifically, Spatio-temporal Bayesian models require calculating the posterior distribution of several scalar parameters, which challenges the use of Markov Chain Monte Carlo Methods. This problem calls for alternative posterior estimation methods, such as the Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009).

Given a specific station, its passenger occupancy is related to its historical occupation and depends on the occupation pattern of other stations. These spatial correlations are nonlinear and complex. The distance between stations gathers spatial correlations over the network. According to Tobler's first law of geography, the correlation between nearby things is stronger. The station's impact on its neighbours may be more significant than neighbours far away in the subway system. Also, those stations on the same line, or those in the same geographical zone, will have a similar passenger trend. The character similarity between stations produces another kind of spatial correlation. For example, stations with similar functional characteristics (e.g. suburbs and centres, education/Commercial/residential areas) may have similar mobility patterns to some other stations. Meanwhile, the day of a week will also affect the passenger flow, not only the labour days but also the traffic limitations of each day. Furthermore, the daily weather, the visitor flow of some attractions are also the possible impact. Although we could enumerate other possible influencing factors, these are absent from the available dataset.

The organization of the exposition is standard: available data is described in Section 2, the Bayesian Spatio-temporal model in Section 3, results and model interpretation in Section 4. Section 5 contain a discussion of the results and comparison with other methods.

2 Data

Station and day is the statistical unit that reports the number of passengers at station i and day t, $Y_{it} \in \mathbb{N}_0$. Every station has at least one entrance. So when one passenger uses his/her Identity Card (IC) or QR-code to enter one station, it counts as a passenger enter. So for one station, the passenger flow includes the whole day passengers who enter this station no matter how they enter or which entrance they go through.

Observed days are from September 1st 2020, to October 30th 2020. The used period coincides with when Beijing started to open after the epidemic of covid-19. Therefore, the observed passenger flow is most likely that of the past years.

Let $i = 1, \ldots, I$ in day $t = 1, \ldots, T, T+1, \ldots$, where the total number of stations is $I = 268$ and the total number of days used to estimate the model is $T = 51$ (September 1st to October 21st) and days $t > T$ (say $t = 52, \ldots, 61$) had been used for testing the model. Thus the first day in the dataset is Tuesday, and the last is Saturday. In total, we should have $I \times T = 13668$ counts of passengers that entered into the metro system on the observed period. Because of 28 missing data (all in the training set), we have 13640 counts. 28 missing counts are not a problem for the estimation (11) and (10). The test set consists of 10 days and thus 2680 counts that the model predicts. For overall stations and days, the mean occupancy is 13,974 passengers per station and day, ranging from a minimum of 563 to a maximum of 150,474 passengers. From October 1st to October 10th are national-day holidays in China, which induced a change in traffic flow, exhibiting heavier passenger density in some stations.

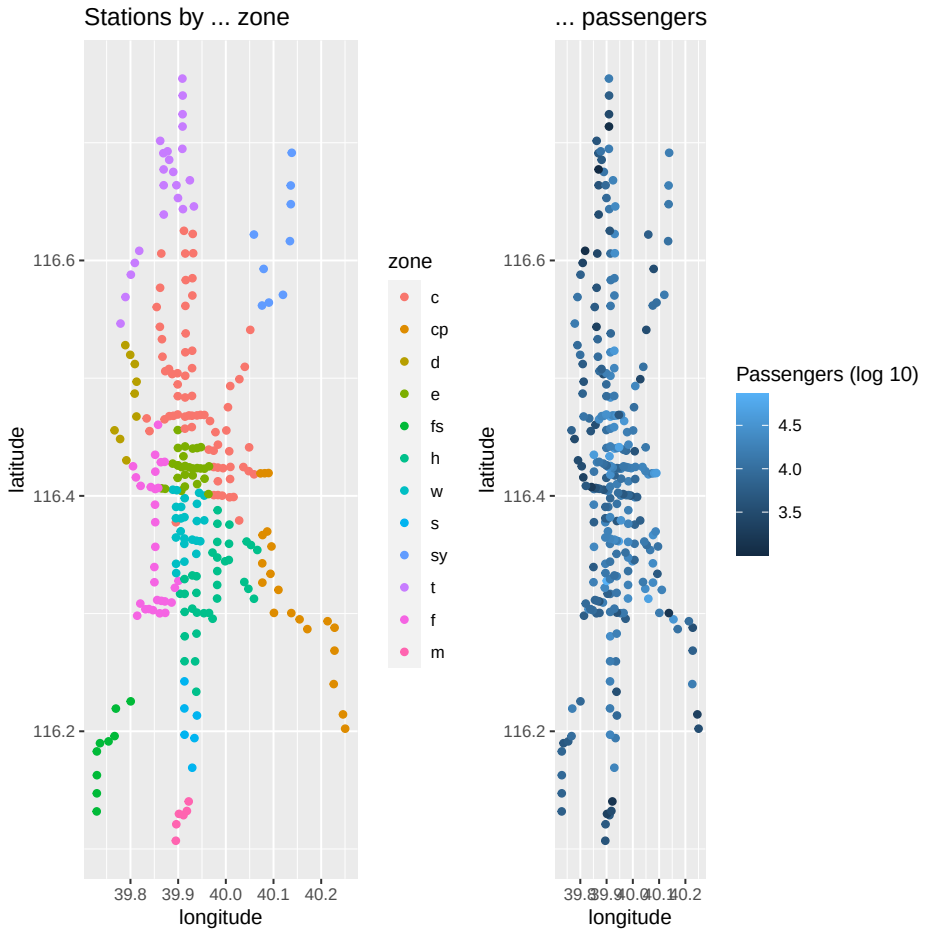

Figure 1: Map of the stations and metro line zones (left) and by mean number of passengers in $\log_{10}$ scale (right).

The available covariates are the other information available for making predictions:

1. Metro line j, denoted with $l_{ij} \in \{0, 1\}$ and $l_{ij} = 1$ if metro station i belongs to metro line j. If metro station $i$is an intersection then $\sum_{j=1}^{J} l_{ij} > 1$, where $J = 22$ is the total number of metro lines in the city, denoted in the graphs with the following labels: BT, CP, DX, FS, L1, L10, L13, L14D, L14X, L15, L2, L4, L5, L6, L7, L8, L8N, L9, LAirport, LAirportDX, S1 and YZ.;

zone $z_{ik} \in \{0, 1\}$ and $z_{ik} = 1$ if metro station i is located at zone k, $k = 1, \ldots, K$. We have $K = 12$ zones denoted in the graphs with labels: c, cp, d, e, f, fs, h, m, s, sy, t, w (see Chinese translation appendix for their real names);

day of the week $w_t \in \{$Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday$\}$ is the day of the week for day t.

For each station, we also have the GPS coordinates of the station's entry points to which we derived a central GPS coordinate of the station (just averaging latitudes and longitudes separately). Figure 1 reports the stations' map along with the metro line zone.

Based on those central points we derived the euclidean distance among stations and build the $I \times I$ adjacency matrix among stations, C_s, whose generic element $c_{i,i'} = 1$ if station i' is one of the 10 most nearest stations to station i, otherwise $c_{i,i'} = 0$. The actual adjacency matrix among metro stations is reported in Figure 2.

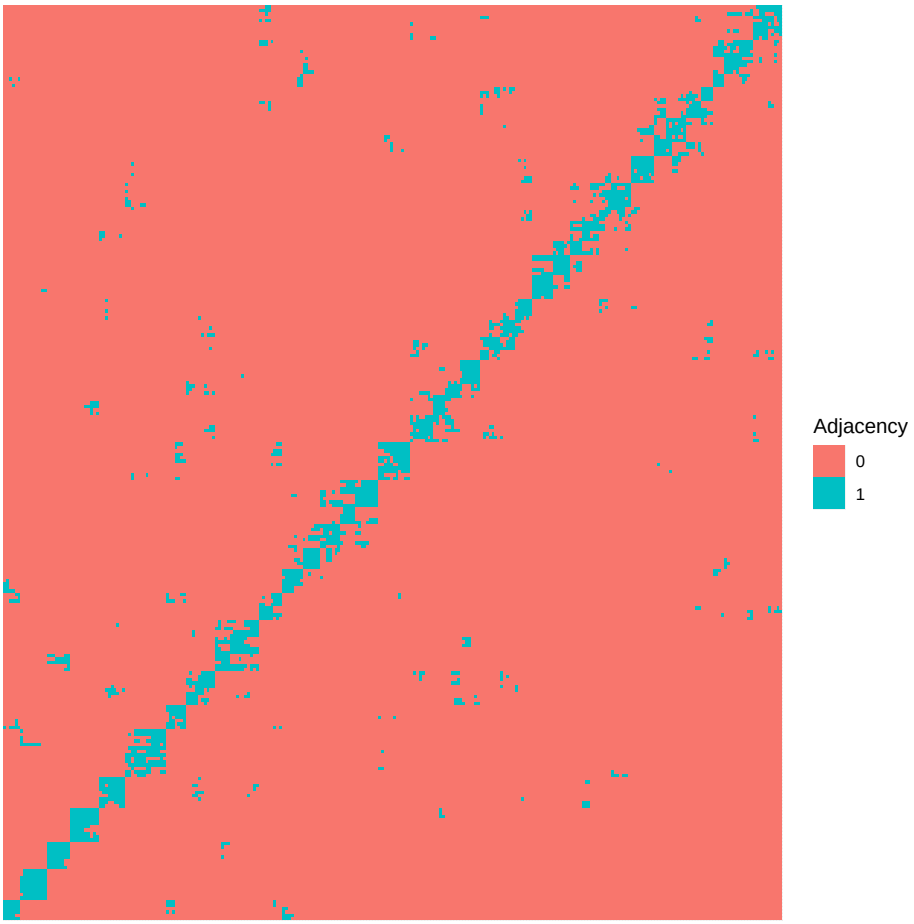

Figure 2: The adjacency matrix C_s among metro stations.

The Manhattan distances instead of the Euclidean one could have been another choice. However, measuring Manhattan distances requires data that are not available (i.e. google walking or car distances from one station to another). Furthermore, as we are modelling passenger traffic that partially moves along metro lines, any means of transport (including walking) is compatible with metro usage. So, the Manhattan distances do not necessarily represent actual distances along metro lines. Therefore, we must assume that the distance relevant for measuring an eventual spatial effect is proportional to the Euclidean distance among stations. A number different than ten stations adjacent to station i induces a decrease of spatial effects. That is why the ten nearest stations have been considered the most appropriate for capturing general citizens traffic around a station. We assume that a passenger generally uses the metro line to cross around ten stations and that this behaviour generates traffic relations among stations. However, because of the randomness nature of the covariance matrix specified in (4) the assumption of 10 stations and the Euclidean distance is questioned *a priori*.

The time correlations include that among days with an Autoregressive random effect of order one and a random effect of the days of the week. Days of the week correlate among the corresponding before and after in a cycle that links Mondays with Sundays. The 7×7 matrix C_w implements this correlation structure among days of the week, as its generic element is one if two days are one after the other and vice versa (e.g., Mondays and Sundays). Similarly, as done for the station, *a priori*, we will put in doubt the usefulness of C_w to capture the cycling behaviour of the traffic during weeks.

3 Statistical Model

The posited Bayesian hierarchical model has the following stochastic representation:

$$
\begin{aligned}
Y_{it} \mid \theta_{it} \sim & \quad Poisson(\theta_{it}),\ E(Y_{it}) = \theta_{it} && \text{(Observable)} && (1)\\
\log(\theta_{it}) = & \quad \alpha + \boldsymbol{\gamma}_l^T \mathbf{l}_i + \boldsymbol{\gamma}_z^T \mathbf{z}_i + s_i + w_t + d_t && \text{(Linear predictor)} && (2)\\
\alpha, \boldsymbol{\gamma} \sim & \quad \pi(x) \propto 1\ x \in \mathbb{R} && \text{(Fixed effects)} && (3)\\
\mathbf{s} \sim & \quad N(0, \Sigma_s),\ \Sigma_s = \tfrac{1}{\tau_s}\left(\mathbf{1} - \tfrac{\beta_s}{\lambda_s} C_s\right)^{-1} && \text{(Station effect)} && (4)\\
\mathbf{w} \sim & \quad N(0, \Sigma_w),\ \Sigma_w = \tfrac{1}{\tau_w}\left(\mathbf{1} - \tfrac{\beta_w}{\lambda_w} C_w\right)^{-1} && \text{(Day of week eff.)} && (5)\\
d_1 \sim & \quad N\left(0, \tfrac{1}{\tau_d(1-\rho^2)}\right) && \text{(First day effect)} &&\\
d_t \mid d_{t-1} \sim & \quad N\left(\rho d_{t-1}, \tfrac{1}{\tau_d}\right),\ t = 2, \ldots, T && \text{(Day AR(1) eff.)} && (6)\\
\log\left(\frac{1+\rho}{1-\rho}\right) \sim & \quad N(0, 6.\overline{6}) && \text{(AR(1) correlation)} && (7)\\
\log(\tau_s), \log(\tau_w), \log(\tau_d(1-\rho)) \sim & \quad \text{log-Gamma}(1, 0.0005) && \text{(Precisions)} && (8)\\
\log\left(\frac{\beta_s}{1-\beta_s}\right), \log\left(\frac{\beta_w}{1-\beta_w}\right) \sim & \quad N(0, 100) && (C_s,\ C_w \text{ corr.}), && (9)
\end{aligned}
$$

where

- $\boldsymbol{\gamma}$ is the vector of fixed effect coefficients for lines, $\boldsymbol{\gamma}_l$ and zones $\boldsymbol{\gamma}_z$. Jointly with α, they represent the fixed effects in the model;
- $\mathbf{s}$ and $\mathbf{w}$ are vectors of size I and 7 respectively;
- $N(\cdot, \cdot)$ is the usual univariate or multivariate normal distribution parametrized in terms of mean and variance (or covariance if multivariate);

- log-gamma$(\cdot,\cdot)$ the log-gamma distribution parametrized in terms of shape and rate;
- λ_s and λ_w are the maximum eigen value of the matrix C_s and C_w, respectively;
- $\mathbf{1}$ is the identity matrix of appropriate size.

As pointed in Davis et al. (2021), the proposed model do not imply that the marginal distribution of Y_{it} is Poisson as it can be far from Poisson. However, the Poisson conditional assumption allows the straightforward derivation of a conditional likelihood for observations (1).

The goal here pursued in solving this problem is double.

1. Firstly, we have to predict unobserved future passengers occupancy $\hat{Y}_{it}$ for $t > T$, without knowing the exact value of all unknown parameters $\boldsymbol{\theta} \in \Theta$ using past observed occupancy $\mathcal{D} = \{y_{it}, i = 1, \ldots, I, t = 1, \ldots, T\}$. This can be done by looking at summaries of the posterior predictive distribution (or summaries of it) of counts at station i for a future day t
$$\pi(\hat{Y}_{it} \mid \mathcal{D}) = \int_{\boldsymbol{\theta}\in\Theta} \Pr(\hat{Y}_{it} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathcal{D})d\boldsymbol{\theta}, \tag{10}$$
where $\Pr(\hat{Y}_{it} \mid \boldsymbol{\theta})$ is the Poisson distribution given in (1). Point predictions are the median of (10) and its 90% Highest Posterior Density (HPD) interval provides the uncertainty around such a point prediction.

2. Secondly, we have to interpret the model in order to be confident on model prediction. This is done by looking at the posterior distribution (or summaries of it) of the parameters
$$\pi(\boldsymbol{\theta} \mid \mathcal{D}) = \int_{\boldsymbol{\theta}\in\Theta} \Pr(\mathcal{D} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}, \tag{11}$$
where $\Pr(\mathcal{D} \mid \boldsymbol{\theta})$ is the Poisson likelihood induced by equations (1) (2) and $\pi(\boldsymbol{\theta})$ is the prior distribution induced by equations (3) - (7).

Posterior (11) has been derived using Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009) instead of Markov Chain Montecarlo Methods (MCMC) (Robert, 2001). Luckily, Spatio-temporal for areal data that fits into the INLA framework (Bivand et al., 2014) and the dimension of $\boldsymbol{\theta}$ challenges the employment of MCMC methods to derive the posterior distribution. INLA is nowadays a standard approximation method based on iterated Laplace approximations of the Gaussian Random Field. This field is induced by the prior distributions on d_t, s_i and w_t given the values of the hyperparameters. Monte Carlo samples approximate their posterior distribution. The appendix provides the INLA formula to estimate the model given the abovementioned variables' symbol in the data set.

4 Results for Beijing Metro occupancy

The accuracy of the approximation with INLA seems acceptable as it is zero the Kullback-Leibler divergence between the simplified Gaussian approximation and complete Laplace approximations for the marginal posteriors. These are two approximation strategies implemented in INLA, and their convergence to the same marginal density indicates that the overall posterior approximation converged.

Before considering the results from the fitted model, it is necessary to assess the model Goodness of Fit (GOF). GOF is here understood as part of the model building process, in that if the model were not compatible with the data, we should elaborate a better one. GOF

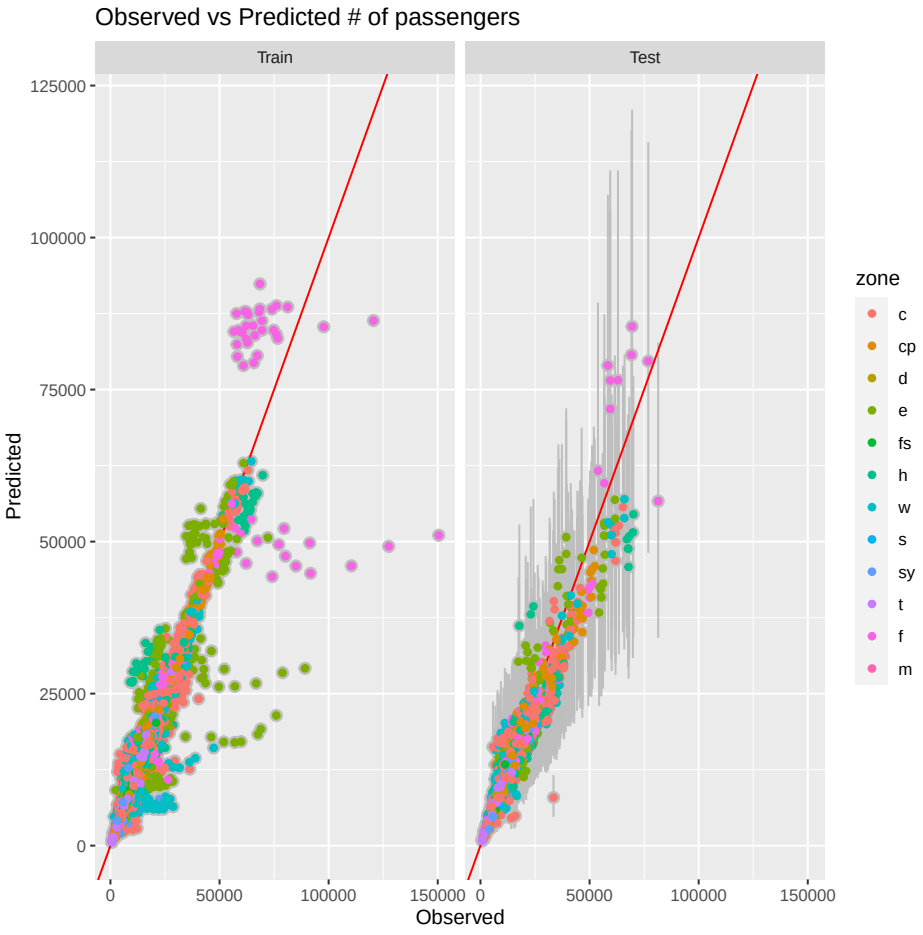

Figure 3: The R^2 statistics of the predicted median number of passengers in both train and test is around 92%, the median absolute prediction error is around 1,800. For the test set the HPD 90% intervals are larger, leading to cover 97% of all passengers count on the test set with a median wide of the intervals around 9,300 passengers.

assessment consists in comparing the median of (10) with the number of passengers observed on the test set, the remaining ten days on the data set as done in Figure 3.

Comparing point predictions, we can see that the model performs equally on the train and test set as R^2 in both cases is 92%. Therefore, the R^2 calculated on the test set is an adequate measure of model GOF, allowing us to consider the model as compatible with the observed data.

4.1 Prediction

To achieve the first goal mentioned in Section 3 we have to study the point prediction errors and the coverage and range of HPD intervals.

The median prediction error on the test set is around 1,800 passengers above or below the proposed median posterior prediction, and 90% of predictions have an absolute error of fewer than 6,000 passengers. Thus, the HPD intervals cover 97% of passengers counts on the test set, and the intervals have a median of 9,300 passengers. Overall, this is in line with the Beijing Metro Group Ltd. standard. Interestingly, the unconditional (over the test set) coverage of the 90% HPD intervals is more extensive than the actual one. This finding could suggest the enterprise can safely decrease the nominal 90% to obtain narrower HPD intervals.

The next Figure 4 provides more insights into the model prediction performances and its reliability in using point predictions.

The with of HPD intervals measure such reliability. The wider the interval, the more is the prediction uncertainty around the provided point estimation of passenger occupancy at that day and station. From the top of Figure 4, we can see that more significant prediction errors generally correspond to wider HPD intervals which is what one would expect from a Bayesian model. The provided predictions are always conditional to the observed variability in the sample and on model parameters. From the bottom of Figure 4 we can see that the posited model can reliably predict some zones of the city as zone "m" (门头沟) which is an outer suburb. Beijing government just made city planning about it in 2019, so this area generally lacks public transportation. The number of citizens in the "m" zone is much less than in other areas. However, almost all passengers in this area are resident citizens, which means the ratio of students or tourists among passengers is very low. At the same time, the citizens in this area have a more vital willingness to choose public transportation. They travel more regularly. So we have less noise than in other areas. Traffic on others zones is weaker predicted as in zone "h" (海淀), or they have more considerable prediction uncertainty as zones "c", "cp", "e" or "w" (朝阳、昌平、东城、西城). There is a slight overestimation in the number of passengers during labour days (Mondays to Fridays), while on weekends, there is a slight underestimation. The underestimation during weekends occurs because more visitors and more students (during weekends) do not have regular actions in those areas. Finally, the bottom right of Figure 4 illustrates the one-day ahead prediction error. We can see that the absolute error tends to increase for future labour days, but this increment seems not constant. This finding indicates that the model can be reliably used to predict at least ten days of metro traffic over the Beijing cities. The model worked for the considered period without unexpected relevant events that can drastically change day-by-day traffic behaviour (data AFTER Covid-19).

We can study the frequentist coverage of model predictions by combining the prediction error and the range of HPD intervals. As Bayesian models provide inference conditionally to the sample, the frequentist coverage is out of the scope of the conditional inference in a Bayesian analysis. However, looking at the agreement with frequentist principles, Bayesian procedures should also have good frequentist coverage. This agreement is known to be valid for the model parameters and a general class of prior distribution (Peers, 1968).

The marginal coverage is 97%, and Figure 5 reports the conditional coverages to zone, days

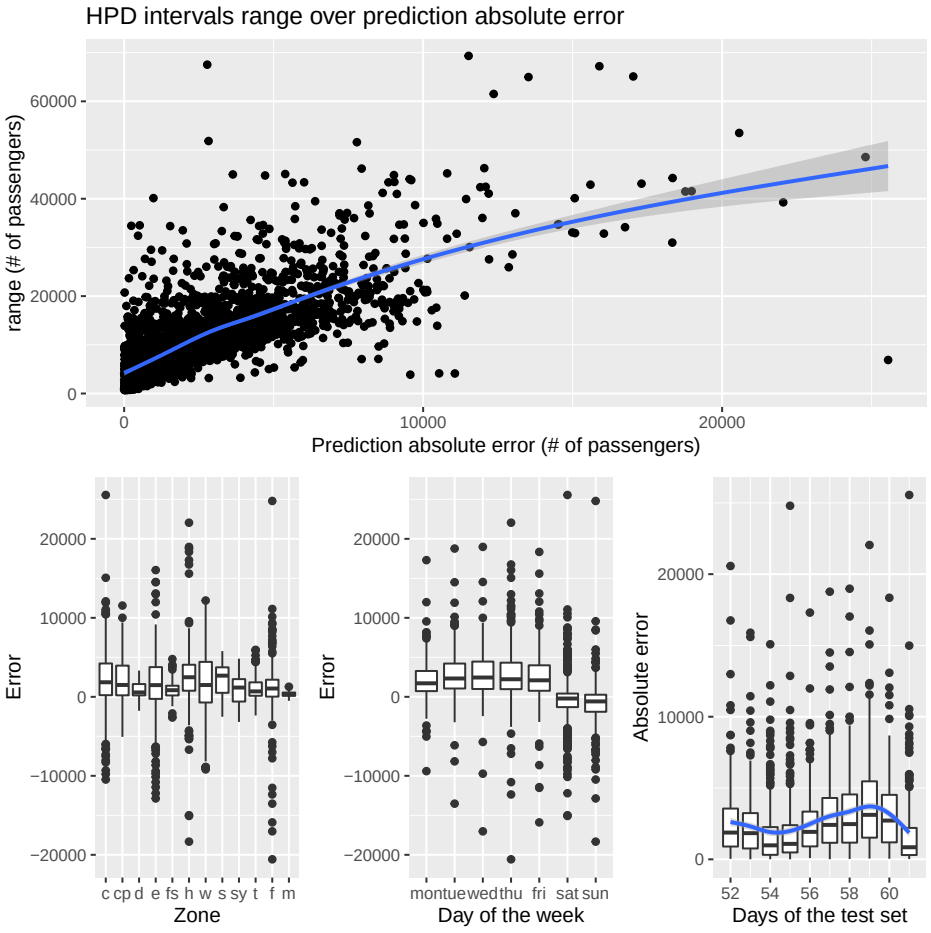

Figure 4: (Top) Absolute prediction error over the range of 90% HPD intervals. (Bottom) Errors of point estimation (median of (10)) by: city zones (left), days of the week (center) and days of the test set (right) where day 52 is one day head predicted and day 61 is ten days ahead predicted.

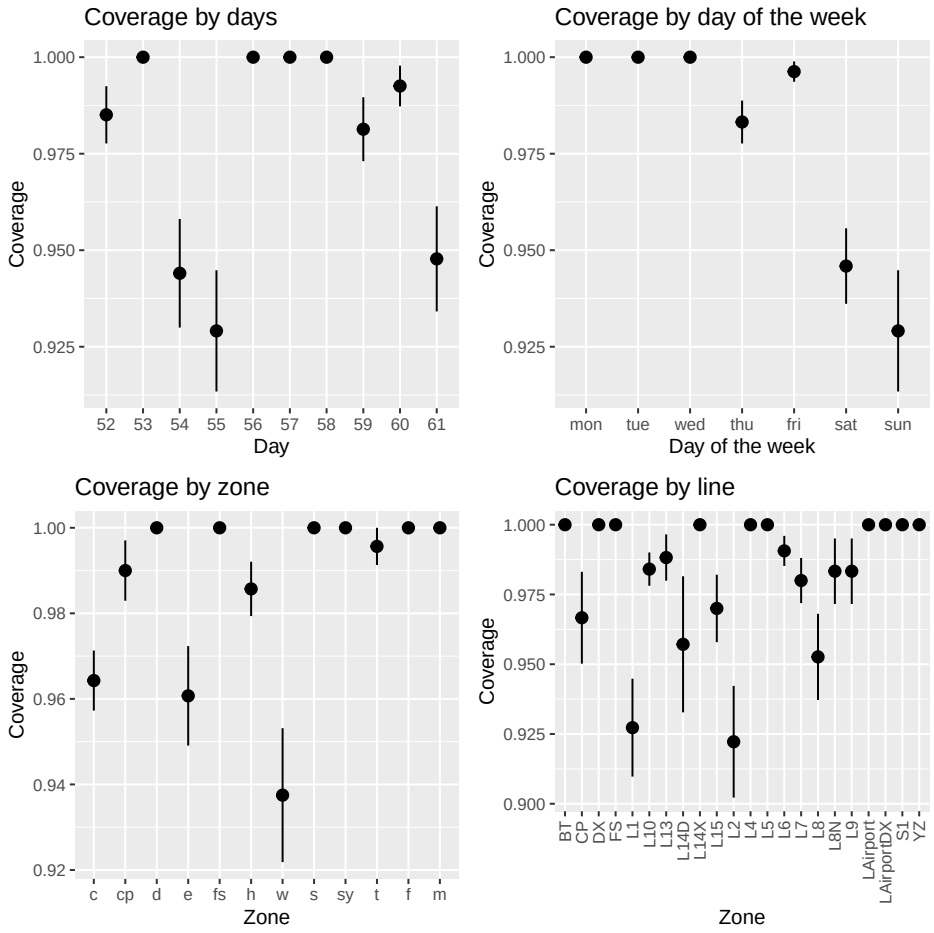

Figure 5: Conditional coverages of 90% HPD intervals.

of the week, metro lines and the number of ahead days predictions (52 is the one day ahead, while 61 is ten days ahead).

From Figure 5 we can see that the conditional coverages are not very different from the marginal one and almost well above the nominal 90% except for Zone "w" and Sunday. Thus, there is no clear degrading pattern on the HPD coverage by the number of days ahead. This last finding indicates the general reliability of passenger traffic prediction and the reliability of the provided uncertainty.

4.2 Interpretation

To achieve the interpretation goal mentioned in Section 3, we have to study the posterior distribution (11). For this purpose, we have to provide parameter interpretation and their corresponding marginal posterior distribution or some summary of it (e.g., posterior means and 90% HPD intervals).

To have a practical understanding of the magnitude of the effects, we have to report that the posterior mean of the intercept α that includes zone "c" and Line 4 is around 8.9 (8.6 - 9.2 90% HPD), which means that every effect in the scale of passengers is equal to the exponential of the effect multiplied by $\exp(8.9) = 7300$ passengers. If an effect, either random or fixed, were equal to 2 on the exponential scale, then this would correspond to an effect of around 14,600 passengers. Keep in mind that the observed counts have a sample mean of 13,974 passengers, ranging from a minimum of 563 to a maximum of 150,474 passengers. Those extremes are respectively 25 times less and 11 times more than the sample mean, so we expect effects to be between 11 and 25.

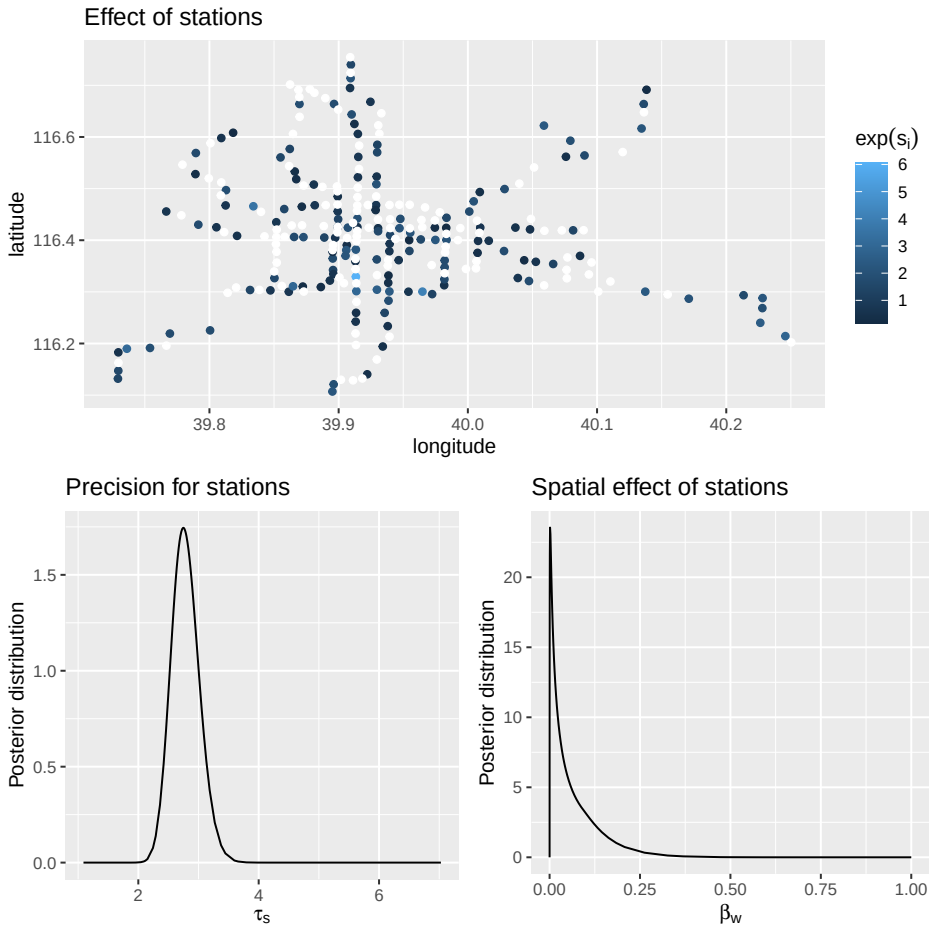

Figure 6: (top) Posterior median of: $\exp(s_i)$ with white points being stations with no significative effects (i.e. $\Pr(s_i < 1|\mathcal{D}) > 0.9$ or $\Pr(s_i > 1|\mathcal{D}) > 0.9$). (bottom-left) Posterior distribution of station random effect precision τ_s. (bottom-right) Posterior distribution of the the random effect, β_s, explained by adjacency matrix C_s.

Appendix A reports the numerical summaries of the marginal posterior distributions for some parameters, while the others are reported graphically into the following subsections.

Station random effects

An interesting future for the enterprise is to understand the most input and output stations in the network. There are stations in which the number of passengers is high and should be under monitoring. The random effect for the station, s_i, provides this information. The station effect joint prior distributions is in (4) and their marginal posterior distributions are reported in Figure 6.

Looking into the map in Figure 6 we can see that there are some stations in the city centre that exhibit a high random effect as well as some stations at the border of the city. Stations under-occupied (in respect of the value of $\exp(\alpha)$), the black points are usually located nearest to other overcrowded ones. The top of Figure 6 can be helpful to understand which stations have to be monitored for their traffic by starting from those with more significant random effects. These stations can also suggest where to eventually build new stations to diminish the traffic at nearby stations.

Looking at the role of $\beta_s \in [0, 1]$ and $\tau_s > 0$ in (4) we can see that we allowed for stations to either exhibits a random effect which is a weighted mixture of a pure random effect ($\beta_s \to 0 \implies \Sigma_s = \tau_s^{-1}\mathbf{1}$) and pure spatial effect ($\beta_s \to 1 \implies \Sigma_s = \tau_s^{-1}C_s$) with weights $1-\beta_s$

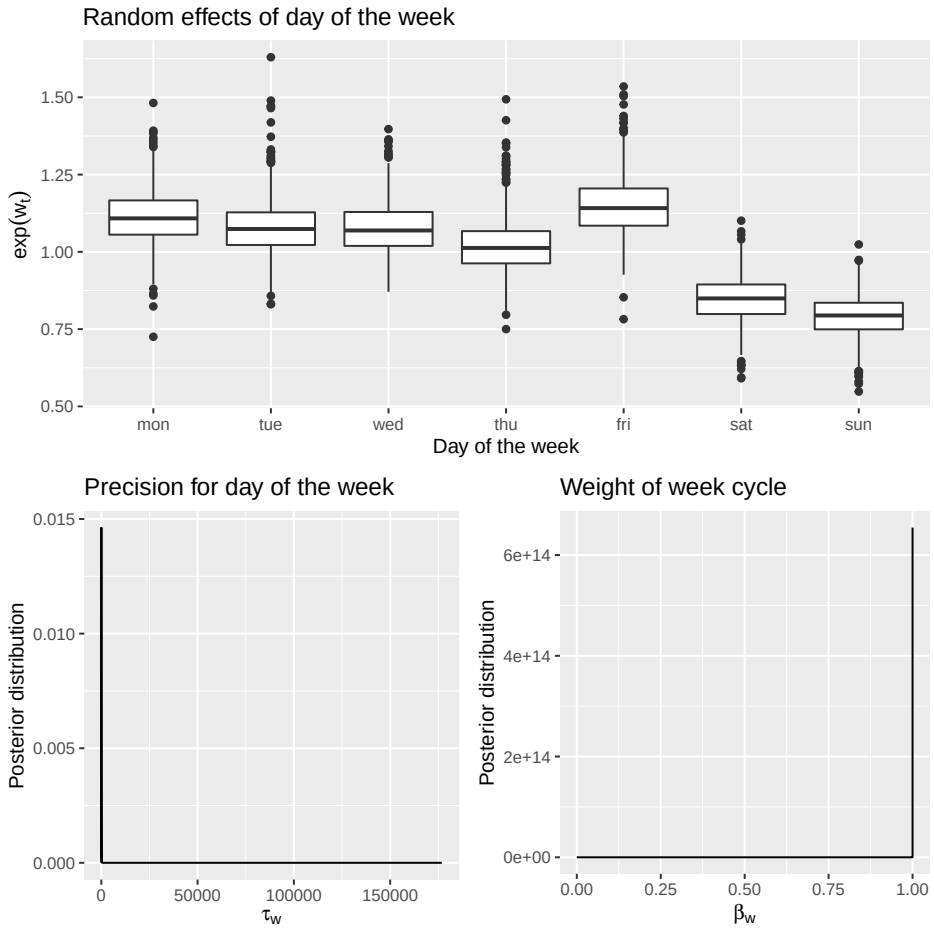

Figure 7: (top) Posterior distribution of $\exp(w_t)$. (bottom-left) Posterior of day of the week's precision τ_w. (bottom-right) Posterior distribution of the random effect, β_w, explained by the week cycle matrix C_w.

and β_s, respectively. This is a flexible way that we employed to account for the uncertainty on the posited spatial effect encoded into matrix C_s as mentioned above. For this we allowed β_s to be random with prior specified in (9), which is a U-shaped prior, centered at $\beta_s = 0.5$, symmetric and with substantial mass at $\beta_s = 0$ and $\beta_s = 1$, in a similar fashion to the Jeffreys's prior for the probability of success in a Bernoulli model. The posterior of β_s reported in Figure 6 (right) indicates that spatial component has a large uncertainty as the 90% HPD of $\pi(\beta_s|\mathcal{D})$ (Figure (6) (bottom-right)) is between 0.01 and 0.29 with a median of 0.13. This indicates that at least 70% of the effect of stations is purely random. This means that either there is not strong spatial effect or that this spatial effect is not properly accounted by the posited matrix C_s.

week cycle random effects

The week cycle encoded into matrix C_w plays a non-negligible role in predicting the number of passengers. From the bottom-left of Figure 7, we can see that this is due to the very high precision of the random effect of the day of the week w_t in (5).

The week cycle matrix C_w plays an important role into the stochastic relations among the days of the week as $\pi(\beta_w|\mathcal{D})$ has a median of 0.46 with a 90% HPD interval of 0.10-0.82 (bottom-right of Figure 7). As expected, the top of Figure 7 indicates that during weekends the number of passengers is below the meanwhile the toughest day is Friday as the traffic reach its highest level.

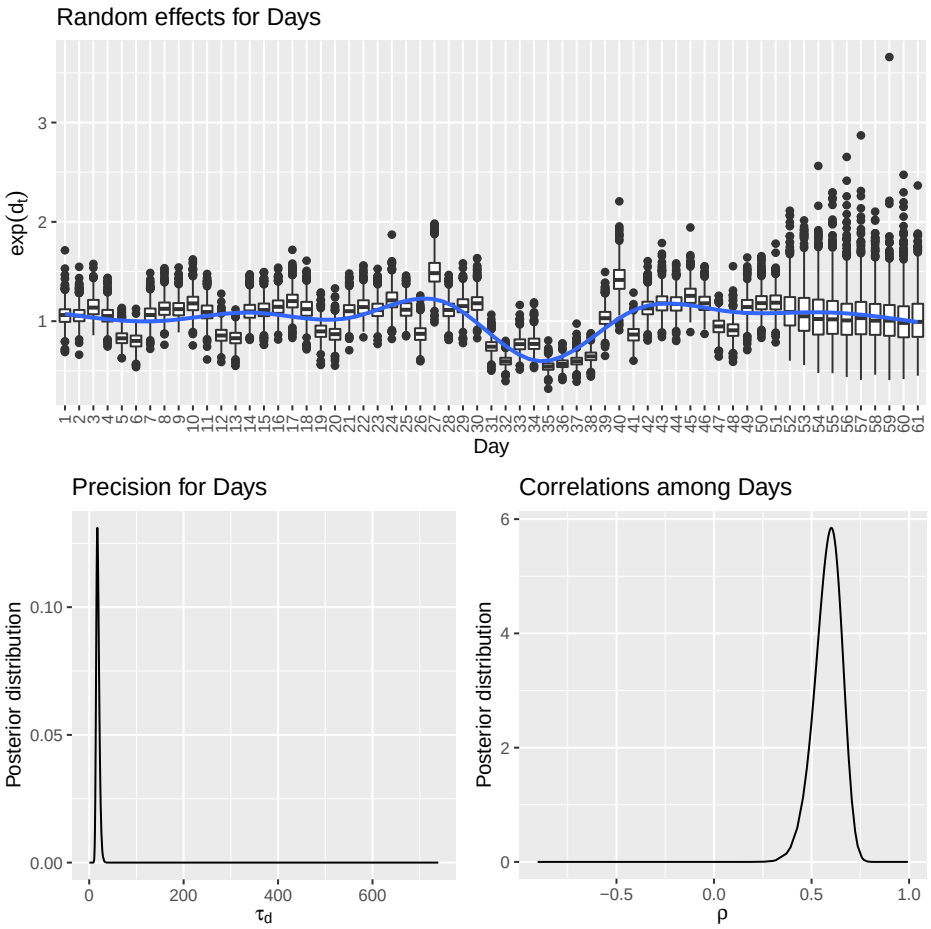

Figure 8: (top) Posterior distribution of $\exp(d_t)$. (bottom-left) Posterior of day precision τ_d. (bottom-right) Posterior distribution of days correlation, ρ.

Time (day) random effects

Days random effects, posed in (6), are capturing the non-cycling day relations. Conditionally one to each other, these random effects are autoregressive (AR) of order one. Days correlation is $\rho \in [-1, 1]$ and τ_d is the conditional (to ρ) variance and the marginal one is $\tau_d(1-\rho)$. The prior on ρ in (7) is U-shaped density, centered at 0, symmetric and with relevant mass at $\rho = 1$ and $\rho = -1$. For the marginal variance, prior (8), is substantially uninformative as all priors on precision parameters used in the model. Figure 8 reports posterior distributions of days random effects as well as those of conditional τ_d and correlation ρ.

The ρ has a posterior median of 0.6 with a 90% HPD interval of (0.42 - 0.77). These values indicate a strong correlation among days and the AR(1) process jointly with the cycle matrix C_w model the temporal part of the traffic process. In particular, we can see from the top of Figure 8 that during the national holidays, the metro traffic decreased, and this behaviour has been taken into account by the AR(1) process (and not by the days of the week effect). Finally, the marginal precision of the day random effect has a median of $18 \times (1 - 0.6) = 7.2$, indicating that its final contribution to predicting the traffic is larger than that of the station random effect and, in any case, smaller than the week cycle effect.

Metro lines

In (2) we assume that metro lines have a fixed effect on the mean number of daily passengers. The summaries of $\pi(\boldsymbol{\gamma}_l|\mathcal{D})$ are given in Appendix A and Figure 9 reports the posterior

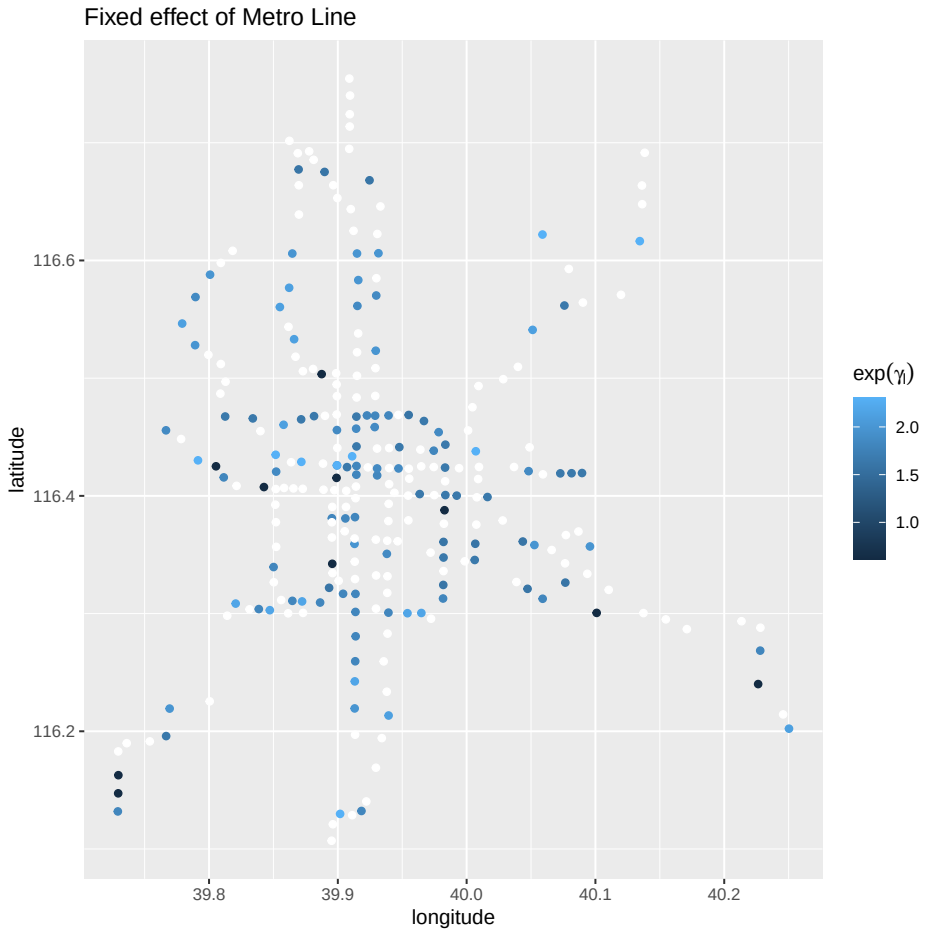

Figure 9: (top) Median of posterior distribution of the fixed effects for lines $\exp(\boldsymbol{\gamma}_l)$. White points are not significative or they are interchange stations.

median of the fixed effect $\boldsymbol{\gamma}_l$ for those metro lines l that have a significant effect, i.e. the corresponding 90% HPD interval is either positive or negative.

From Figure 9 we can see that the most crowded lines are at the border of the city as well as those lines that are more connected. The line effect is essential to estimate the traffic at interchange stations. At these stations, the estimated traffic gathers information by the metro lines present at the interchange.

City zones

In (2) we assume that metro zones also affect the mean number of daily passengers. The summaries of $\pi(\boldsymbol{\gamma}_z|\mathcal{D})$ are given in Appendix A and Figure 10 reports the posterior median zones which have a significant effect according to their corresponding 90% HPD interval.

The lower and higher significative random effects are at zone "m" in the south and zone "t" in the north. Figure 10 provides a clear indication of which zone should be eventually interested in building new stations, expanding an existing one. Figure 10 could suggest a program of daily train traffic indicating, for example, that trains from zone "m" could be subtracted from that zone to serve zone "t".

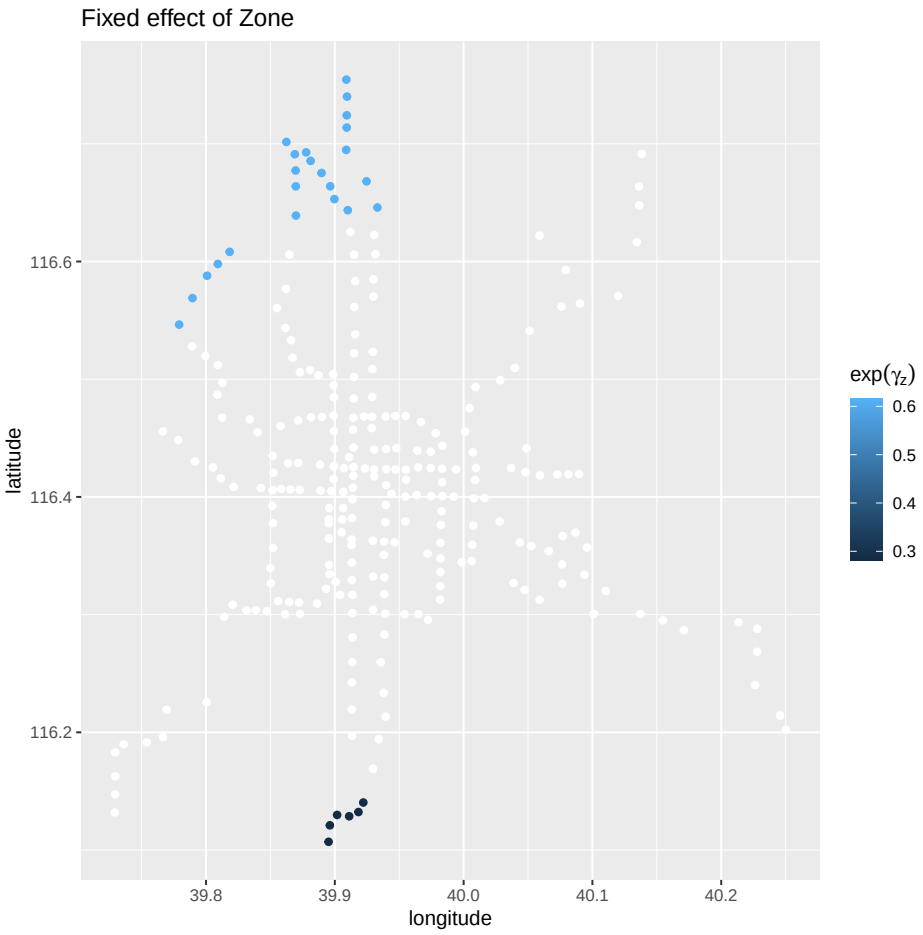

Figure 10: (top) Median of posterior distribution of the fixed effects for zones $\exp(\boldsymbol{\gamma}_l z)$. White points are not significative.

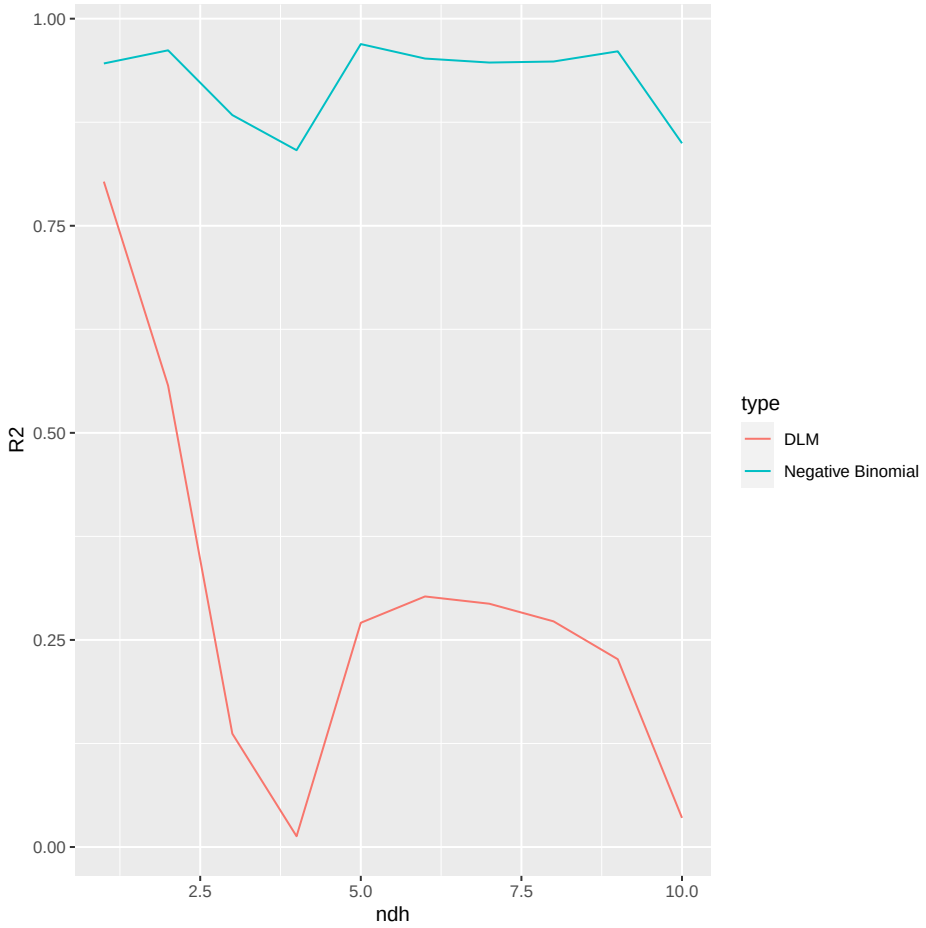

Figure 11: R^2 statistics for point prediction on the number of days ahead predicted. These are for the Negative Binomial and for the Dynamic Linear Model.

5 Discussion

We compare the proposed approach with two different models using how these are precise in predicting Y_{it} and the coverage of the reported uncertainty around them. First, we consider the Negative Binomial regression model. This model is estimated using station effect, a linear trend on the days, the days of the week, the metro line and the city zone. Predictions are accurate with a R^2 around 90%, that it generally degrades with the number of ahead predicted as shown in Figure 11.

However, the coverages of the 90% *confidence* intervals for predictions fail to cover the observed counts on the test set. This coverage is around 38%. Essentially, intervals' wides mean 1,249 passengers, and they do not adequately account that some stations are more variable than others. Thus, this model, estimated under a frequentist setting, fails to manage the uncertainty around prediction.

Finally, another model for comparison purposes would be a multivariate Dynamic Linear Model (DLM) (Petris, 2010; Petris et al., 2009) on the $\log(1 + Y_{it})$. Given the number of involved parameters (grows with I), the joint prediction for all stations is unfeasible. Therefore the DLM model has been fitted for each station separately. Concretely, we used a polynomial DLM of order ten. This model implies that the latent vector has a size of 10 scalars. Again the R^2 of the expected posterior counts is much less than 92% and also decreases with the number of days ahead of predicted (see Figure 11). The corresponding 90% credible intervals coverage is around 70%, and intervals have a mean wide of 1000 passengers.

The comparison for the frequentist model shows that a Bayesian approach is needed;

otherwise, the uncertainty is not correctly estimated. Even under a Bayesian setting, as for the DLM model, the dependency structure among passengers counts is far more complicated than an autoregressive structure on the latent space.

A Appendix - Summaries of posterior distributions

We report in the following Table 1 the summary statistics of the posterior distributions of the main parameters.

B Appendix - INLA Code

The INLA formula for the fitted model is:

```
total ~ f(t,model="ar1")+
f(i, model = "generic1",
Cmatrix=Diagonal(1, n = nrow(Cs))-Cs)+  # Local spatial effects
f(as.numeric(w), model = "generic1",
Cmatrix=Diagonal(1, n = nrow(Cw))-Cw)+ # Week cycle effect
z+BT+CP+FS+L1+L10+L13+L14D+L14X+L15+L2+
L5+L6+L7+L8+L8N+L9+LAirport+LAirportDX+S1+YZ
```

Where symbols coincides with the model notation and symbols `BT, CP, FS, L1, L10, L13, L14D, L14X, L15, L2, L5, L6, L7, L8, L8N, L9, LAirport, LAirportDX, S1, YZ` are dummy variables indicating the metro line.

C Appendix - Label translation

The following Latin character is the labels of zones reported in graphs and tables. Table 2 reports their corresponding English and Chinese names.

References

Aitchison, J & Ho, C (1989), 'The multivariate poisson-log normal distribution,' *Biometrika*, **76**(4), pp. 643–653.

Bivand, RS, Gómez-Rubio, V & Rue, H (2014), 'Approximate bayesian inference for spatial econometrics models,' *Spatial Statistics*, **9**, pp. 146–165.

Chib, S & Winkelmann, R (2001), 'Markov chain monte carlo analysis of correlated count data,' *Journal of Business & Economic Statistics*, **19**(4), pp. 428–435.

Davis, RA, Fokianos, K, Holan, SH, Joe, H, Livsey, J, Lund, R, Pipiras, V & Ravishanker, N (2021), 'Count time series: A methodological review,' *Journal of the American Statistical Association*, pp. 1–15.

Ma, J, Kockelman, KM & Damien, P (2008), 'A multivariate poisson-lognormal regression model for prediction of crash counts by severity, using bayesian methods,' *Accident Analysis & Prevention*, **40**(3), pp. 964–975.

Table 1: Summary of some posterior distribution.

Parameter	**Median**	**90% HPD interval**
α	8.9	(8.59 - 9.16)
τ_s	2.74	(2.34 - 3.19)
β_s	0.08	(0.00 - 0.26)
τ_w	47	(7 - 139)
β_w	0.46	(0.10 - 0.82)
τ_d	18	(10 - 25)
ρ	0.60	(0.42 - 0.77)
	City Zones	
cp	0.30	(-0.07 - 0.67)
d	-0.14	(-0.59 - 0.31)
e	-0.07	(-0.34 - 0.2)
fs	0.18	(-0.64 - 1)
h	-0.07	(-0.3 - 0.15)
w	0.02	(-0.26 - 0.31)
s	0.09	(-0.41 - 0.58)
sy	0.18	(-0.26 - 0.61)
t	-0.49	(-0.77 - -0.22)
f	-0.27	(-0.56 - 0.02)
m	-1.26	(-2.46 - -0.05)
	Metro Line	
BT	0.46	(0.11 - 0.81)
CP	-0.18	(-0.62 - 0.26)
FS	-0.16	(-0.91 - 0.59)
L1	0.68	(0.4 - 0.96)
L10	0.59	(0.35 - 0.83)
L13	0.83	(0.54 - 1.12)
L14D	0.33	(-0.06 - 0.72)
L14X	-0.79	(-1.53 - -0.05)
L15	0.17	(-0.15 - 0.5)
L2	0.75	(0.42 - 1.08)
L5	0.62	(0.35 - 0.89)
L6	0.50	(0.23 - 0.76)
L7	-0.02	(-0.29 - 0.25)
L8	0.15	(-0.14 - 0.44)
L8N	-0.48	(-0.86 - -0.1)
L9	0.78	(0.4 - 1.15)
LAirport	-0.04	(-0.6 - 0.52)
LAirportDX	0.75	(-0.28 - 1.77)
S1	0.41	(-0.69 - 1.51)
YZ	0.13	(-0.25 - 0.52)

Table 2: Chinese translations of zones.

Label	**English**	**Chinese**
c	chaoyang	朝阳
cp	changping	昌平
d	daxing	大兴
e	dongcheng	东城
f	fengtai	丰台
fs	fangshan	房山
h	haidian	海淀
m	mentougou	门头沟
s	shijingshan	石景山
sy	shunyi	顺义
t	tongzhou	通州
w	xicheng	西城

Peers, HW (1968), 'Confidence properties of bayesian interval estimates,' *Journal of the Royal Statistical Society. Series B (Methodological)*, **30**(3), pp. 535–544.

Petris, G (2010), 'An r package for dynamic linear models,' *Journal of Statistical Software*, **36**(1), pp. 1–16.

Petris, G, Petrone, S & Campagnoli, P (2009), 'Dynamic linear models,' in *Dynamic Linear Models with R*, Springer, pp. 31–84.

Robert, CP (2001), *The Bayesian Choice*, Springer.

Rue, H, Martino, S & Chopin, N (2009), 'Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations,' *Journal of the royal statistical society: Series b (statistical methodology)*, **71**(2), pp. 319–392.