Modeling travel demand in Portland, Oregon

Katja Ickstadt
Robert Wolpert
Xuedong Lu

ABSTRACT
An important problem in transportation planning is the modeling of patterns of trip-making—especially trips from home (origin) to work (destination) within a fixed geographic area. The standard tool used to study such OD flows is the so-called "gravity model", a Poisson log-linear regression model for studying the number of trips from origins within one element of a fixed partition area (traffic analysis zone, for example) to destinations within another.
We present an alternative: a gridless Bayesian hierarchical Poisson/gamma random field model, allowing us to incorporate spatial correlation explicitly. The models are fitted to a subset of data from the 1994/95 METRO survey of Portland, Oregon, and are internally validated on a reserved portion of the survey data. Bayesian posterior probability distributions are calculated using a Markov chain Monte Carlo integration scheme based on a novel method for simulating samples from gamma random fields.

Some key words: Bayesian mixture models, gamma process, Markov chain Monte Carlo, simulation.

1 Introduction

One of the most important problems in transportation planning is to model the patterns of trip-making from an origin (O) to a destination (D), in order to predict travel demand for such OD flows.

Usually, travel demand studies are based on a system of coarsely partitioned geographical areas, known as traffic analysis zones (TAZs). Interest centers on estimating the expected number of trips T_{ij} from an origin TAZ i to a destination TAZ j. The model most commonly used is the so-called gravity model (Casey (1955), Sen and Smith (1995))

$$T_{ij} = A_i \cdot B_j \cdot \exp\left(-\sum_{n=1}^{N} c_{ij}^n \theta_n\right) ,$$

where the traffic flows T_{ij}, and the costs, travel times or distances c_{ij}^n, $n =$

$1, ..., N$, are observed or known, and the parameters $A_1, ..., A_I$, $B_1, ..., B_J$, and $\theta_1, ..., \theta_N$ are estimated from the data; for $c_{ij}^n = \log|x_i - x_j|$ and $\theta_n = 1$ the equation reduces to the Newtonian potential, hence the name "gravity model". Methods commonly used for analyzing gravity models include fitting marginals in trip-distribution tables, Poisson log-linear regression, and multinomial logit regression.

New legislation, the Clean Air Act Amendments (CAAA) of 1990 and the Intermodal Surface Transportation Efficiency Act (ISTEA) of 1991, requires higher standards of travel demand modeling, such as finer spatial and temporal resolution and incorporating inherent urban structural patterns into the analysis. Although disaggregated forms of the gravity model such as the continuous Poisson log-linear model (Smith (1987)) and the continuous multinomial logit model (Ben-Akiva and Watanatada (1981)) refine the spatial resolution infinitely and do not force the transportation planner to pick an arbitrary partition of the study area before the analysis, they do assume independence between neighboring areas for both origin zones and destination zones. However, neighboring sites often exhibit similar travel behavior, enabling us to learn about travel demands in nearby, structurally similar regions if we model spatial correlation explicitly.

Although little used in transportation science, Bayesian statistical models for possibly correlated, spatial count data are increasingly used in other areas of application. Most commonly applied are Gaussian random field models in which the count data, following a variance-stabilizing transformation, are treated as if normally distributed with a partially unknown covariance structure (this is closely related to the spatial interpolation method of Kriging (Matheron (1963))), and log-linear Poisson/log-normal models, in which the data are assumed to be Poisson distributed with intensities whose logarithms are modeled with a Gaussian random field. Unfortunately Gaussian models (Cressie (1993)) fail to respect the discrete nature of the count data, while log-linear models (Besag et al. (1991); Clayton and Kaldor (1987)) do not scale properly under spatial refinement of partitions (the logarithmic structure leads to products rather than sums for the Poisson means for unions of neighboring partition elements) excluding a continuous, disaggregated version.

Other methods often applied to spatial point count data include auto-Poisson models (Besag (1974)) and Strauss models (Strauss (1975)), which cannot reflect the positive associations that might be expected for nearby reporting units (all count correlations must be negative in those models), and conjugate Poisson/gamma hierarchical models (Clayton and Kaldor (1987)), under which all counts are independent under both prior and posterior distributions. Recently a flexible class of Bayesian hierarchical Poisson/gamma random field models was introduced by Wolpert and Ickstadt (1997) for possibly spatially-dependent count data. These models generalize both conjugate Poisson/gamma models and mixture of Dirichlet process models, and are suitable for spatially continuous applications without aggregation.

In this paper, we apply these geographically continuous hierarchical Poisson/gamma models to analyze part of the 1994/95 Portland Metropolitan Area Household Activity and Travel Survey concentrating on trips from home (origin) to work (destination). We treat all home locations and some of the work locations as known, and forecast the missing work locations. At the lowest stage of the hierarchy the home locations will be modeled as a Poisson random field. At the next level of the hierarchy the intensity measure of this Poisson random field will be modeled as a distance-based kernel mixture of a gamma random field describing the work locations.

In previous applications of this model class the gamma random field was unobserved and latent, and the analysis mainly focussed on smoothing the intensity measure of the Poisson random field (e.g Wolpert and Ickstadt (1997), Ickstadt and Wolpert (1997)). In this application interest will mainly center on estimating the gamma random field (here partially observed and describing the work locations). Using the given home locations and the estimated work locations we obtain the resulting estimated origin/destination pairs. We also estimate the parameters of the kernel function in order to help us understand the influence of travel distance on the home/work travel pattern.

Our analysis is meant as a first step toward modeling origin/destination patterns in Portland, Oregon, intended eventually to enable the prediction of how changes in zoning or road construction work will affect the OD flows.

In section 2 the data are described, in section 3 the class of hierarchical Bayesian Poisson/gamma models is introduced, and its connection to the mixture of Dirichlet processes is pointed out. The Markov chain Monte Carlo sampling scheme is given in section 4. The results of our analysis are presented in section 5, followed by a discussion in section 6.

2 The Data

We use part of the Portland Metropolitan Area Household Activity and Travel Survey conducted in Spring 1994 and Fall/Winter 1994/95 (available on-line at `http://www.webconcepts.com/m2lib/pdxsvy.htm`). The sample includes 4,400 households, selected by random dialing, with about 10,000 people altogether. For each respondent the activity and travel behavior for a 48 hours period was reported. For every change in location the exact (x, y) coordinates and the activity were recorded for the destination. The pairs of days were staggered across the seven days of the week. In addition, household characteristics, such as household size, income, type of dwelling, and number of vehicles, and personal socio-demographics, such as age, gender, employment status, and occupation, were reported.

For our analysis we are particularly interested in the adults' regular travel behavior from home to work. Thus, we take only data entries of employed individuals of age 16 years and older, who made home to work trips on the survey day; we restrict ourselves to the first survey day to avoid correlation; we pool observations from Monday to Friday; and we use only the spring sample.

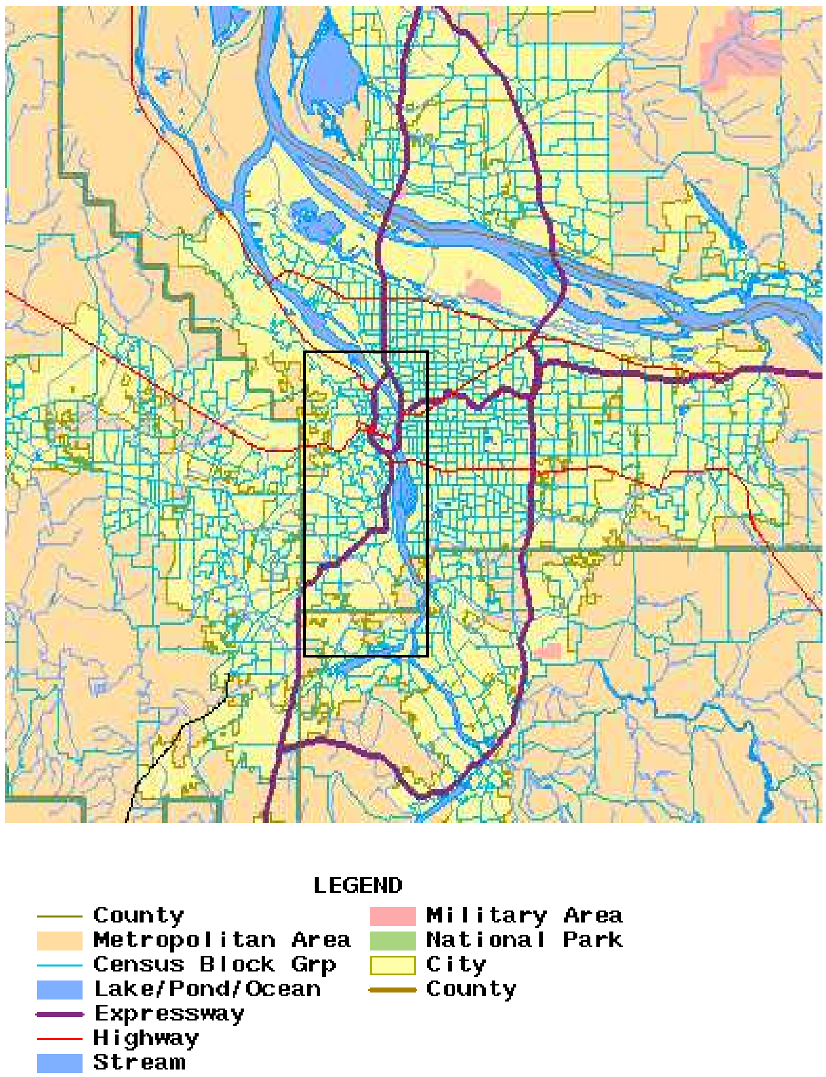
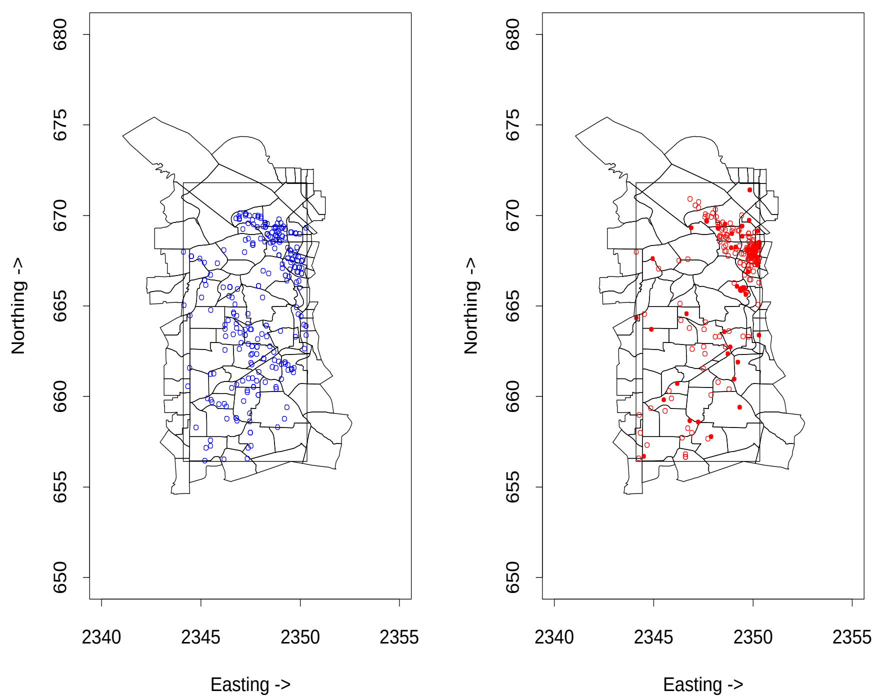


FIGURE 1. Metro and downtown Portland. The box indicates the study region.

The survey data covers the entire Multinomah county, as well as parts of Clackamas county, Washington county, and Clark county. Figure 1, obtained from `http://tiger.census.gov/cgi-bin/mapbrowse-tbl`, shows large parts of Multinomah county, including the Metro Portland and downtown Portland areas. For illustration purposes we base our analysis only on the 290 home to work trips originating and ending in a $6.1km$ by $15.2km$ rectangular study region in downtown Portland, Multinomah County, indicated by a box in figure 1.

Figure 2 displays the 290 OD data pairs in the study region box; overlaid are the TAZs. In figure 2a) the home locations are shown as circles, in figure 2b) the work locations are shown. Here, filled and unfilled circles represent the unknown and known work locations, respectively. In our analysis in section 5, we will treat different fractions of work locations as known.



(a) Home (b) Work

FIGURE 2. (a) Home locations and (b) work locations in the study region.

3 Poisson/Gamma Random Field Models

In this section we describe the models in somewhat more generality than needed for the present application, both to clarify their structure and to simplify their use in other applications; for example, we maintain the distinction between the spaces $\mathcal{X}$ and $\mathcal{S}$ of trip origins and destinations, even though in our application both represent the study region.

3.1 The Model Class and its Interpretation

At the first stage of the hierarchy we will model the point pattern of the origin locations as a Poisson random field in which the number of points $N(A)$ in each set A of the origin space $\mathcal{X}$, part of the plane, is a Poisson random variable with mean $\Lambda(A)$. The representations of the Poisson random field are (almost surely) discrete, consisting of a number of jump locations σ_m, all with unit mass (jump height), so that

$$N(A) = \sum_{\sigma_m \in A} 1 \, .$$

In our application the jump locations represent the house locations in the study region; equal jump height for all locations means that all households are regarded equally influential for the OD behavior. Since the mean intensity $\Lambda(A)$ of the Poisson random field is unobserved, we regard it as a non-negative random variable and thus use a doubly stochastic Poisson model.

Doubly stochastic models that assign independent "increments" $\Lambda(A_i)$ to disjoint sets A_i, such as the gamma and stable random fields, would necessarily feature independent counts $N(A_i)$ as well, and thus could not directly be used to study possible spatial dependence. Instead we construct the intensity measure at the second stage of the hierarchy as a kernel mixture $\Lambda(A) \equiv \int_{\mathcal{S}} \int_A k(x,s) dx \Gamma(ds)$ for an independent-increment infinitely divisible random measure $\Gamma(ds)$ on an auxiliary space $\mathcal{S}$, usually a subset of the plane or a discrete set.

In our application the random measure $\Gamma(ds)$ describes the destinations, i.e. the work locations. Work places with many employees will be more influential for the OD behavior than places with a small number of employees. Thus, we model the destination locations by an independent-increment infinitely divisible random field that features a range of possible jump heights at different locations, larger jumps representing a larger number of employees, such as the gamma or stable random fields (Wolpert and Ickstadt (1998)). The entire set of possible workplaces is infinite (and dense) in such a model, but for each bounded subregion $A \subset \mathcal{S}$ and number $\varepsilon > 0$ the number of work places in A with employment intensity greater than ε will be finite, again with a Poisson distribution. In this paper we illustrate the ideas taking $\Gamma(ds)$ to have the gamma random field distribution $\Gamma(ds) \sim \mathrm{Ga}\big(\alpha(ds), \beta(s)^{-1}\big)$ with shape measure $\alpha(ds)$ and inverse scale function $\beta(s) > 0$.

At the third stage of the hierarchy we introduce a prior probability distribution $\pi(d\theta)$ on a parameter space Θ in order to express uncertainty about the kernel, $k(x,s) = k^\theta(x,s)$, and some aspects of the $\Gamma(ds)$ random field, $\alpha(ds) = \alpha^\theta(ds)$ and $\beta(s) = \beta^\theta(s)$. Thus, our class of doubly stochastic Poisson random field models is given hierarchically by:

$$\begin{aligned} \theta &\sim \pi(d\theta) \\ \Gamma(ds)\big|_\theta &\sim \mathrm{Ga}(\alpha^\theta(ds), \beta^\theta(s)^{-1}) \\ N(dx)\big|_{\theta,\Gamma} &\sim \mathrm{Po}\big(\Lambda(dx) \equiv \int_{\mathcal{S}} k^\theta(x,s)\, dx \Gamma(ds)\big). \end{aligned}$$

Figure 3 presents the hierarchical Poisson/gamma random field model in pictures. On the lowest level the home locations (dots) are shown; on the middle level a contour plot of the estimated mean intensity $\Lambda(dx)$ is given; and on the highest level a realization of the gamma random field (jump locations are work locations, jump heights are proportional to the number of employees) is presented. The dashed lines indicate that the gamma random field jumps are averaged (weighted by the kernel) to obtain an estimate for $\Lambda(dx)$ at a certain home location x.

Conditional on the parameter values θ the means and covariances of the Poisson random field are given by:

$$\begin{aligned} \mathrm{E}^\theta(N(A)) &= \int_{\mathcal{S}} k^\theta(A,s)\beta^\theta(s)^{-1}\alpha^\theta(ds) \\ \mathrm{Cov}^\theta(N(A), N(B)) &= \int_{\mathcal{S}} k^\theta(A \cap B, s)\beta^\theta(s)^{-1}\alpha^\theta(ds) \\ &\quad + \int_{\mathcal{S}} k^\theta(A,s) k^\theta(B,s)\beta^\theta(s)^{-2}\alpha^\theta(ds), \end{aligned}$$

where $k^\theta(A,s) \equiv \int_A k^\theta(x,s)\, dx$. For further details on this model class see Wolpert and Ickstadt (1997).

3.2 The Specific Model

In our application we represent uncertain model features with $\theta \in \mathbb{R}^3$ as follows. We take $\Gamma(ds)$ to be a gamma random field with a uniform shape measure $\alpha^\theta(ds) = (\theta_1/\theta_2)\, ds$, and constant scale $\beta^\theta(s)^{-1} = \theta_2$, hence mean $\beta^\theta(s)^{-1}\alpha(ds) = \theta_1\, ds$.

We choose a gravity model kernel with a single cost function expressing the undesirability of commuting from x to s as squared Euclidean distance, leading to the Gaussian kernel density

$$k^\theta(x,s) \equiv \frac{1}{2\pi\theta_3^2} e^{-|x-s|^2/2\theta_3^2}$$

that allows for a range of distances over which spatial interaction between the home and work locations might extend.

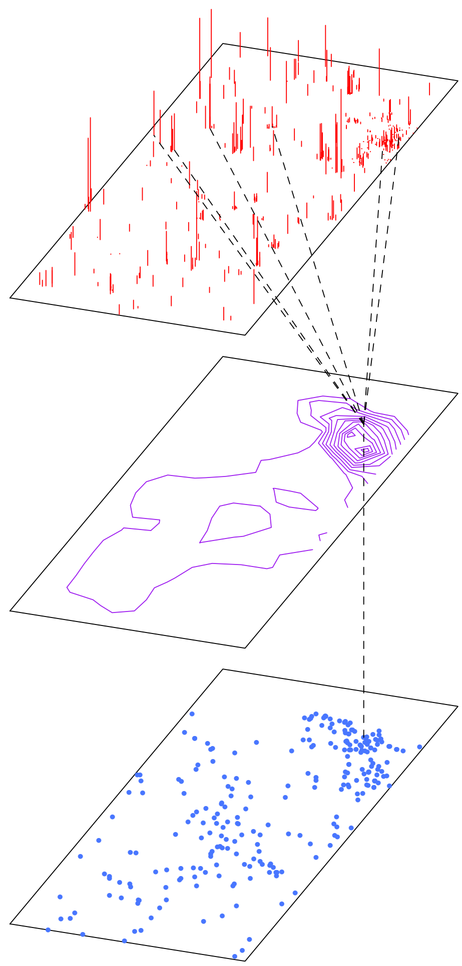


FIGURE 3. Graphical representation of the hierarchical Poisson/gamma random field model.

For sets A of large diameter the above parameterization leads to the following approximate means and variances (see Wolpert and Ickstadt (1997)):

$$\mathrm{E}^{\theta}(N(A)) \approx \theta_1 |A|, \qquad \mathrm{Var}^{\theta}(N(A)) \approx \theta_1 |A| (1+\theta_2).$$

Evidently, θ_1 is the home location density, in counts per unit area (hectare (ha), in our application), and θ_2 is the fraction of the extra-Poisson variation or large-scale overdispersion.

3.3 Connection to Mixture of Dirichlet Processes

If we specifically set $\beta^{\theta}(s) \equiv \beta$ for a constant β, and scale the gamma random field to total mass one, $\mathcal{D}(ds) = \Gamma(ds)/\Gamma(\mathcal{S})$, our model can be viewed as a mixture of Dirichlet processes model (Antoniak (1974), Escobar and West (1998), MacEachern (1998)) for estimating the Poisson intensity measure $\Lambda(dx)$. In our hierarchical Poisson/gamma model class nonconstant functions $\beta^{\theta}(s)$ are allowed reflecting our prior opinion about the space $\mathcal{S}$ to be more informed in some places than in others; the kernel function $k^{\theta}(x, s)$ need not be a Gaussian density; and the gamma random field can be replaced by another Lévy random field (Wolpert and Ickstadt (1998)) to model a different tail behavior for the distributions on the hierarchical (gamma) level.

4 The Computational Scheme

Here we describe the Markov chain Monte Carlo (MCMC) computational scheme needed for the posterior analysis presented in section 5. While the likelihood function for θ from our Poisson/gamma model is unavailable analytically, we can study the intractable posterior distributions of the uncertain quantities θ and the gamma random field $\Gamma = \Gamma(ds)$ by simulating steps (θ^t, Γ^t) from an ergodic Markov chain whose stationary distribution is the posterior distribution of θ and Γ given the observed home locations $N(dx)$ (Gelfand and Smith (1990); Gilks et al. (1996); Tierney (1994)).

The simulation approach requires sampling from the complete conditional distributions of each uncertain quantity $(\theta^t, \Gamma^t, etc.)$ given all the others; see Wolpert and Ickstadt (1997). The simulation is simplified dramatically by the introduction of data augmentation, in this application selecting an augmented work location $s_n \in \mathcal{S}$ for each of the $N = 290$ survey participants residing at $x_n \in \mathcal{X}$. With this augmentation the data become conjugate for the gamma random field, enabling us to simulate the conditional distribution of Γ given θ and the augmented points $\{s_n\}_{n \le N}$ using the methods of Wolpert and Ickstadt (1998). The conditional distribution of the augmentation points given θ and Γ is discrete and easy to sample from. The conditional distribution of θ given Γ and the augmentation points $\{s_n\}_{n\le N}$ depends on the arbitrary prior distribution $\pi(d\theta)$. It has a density function if $\pi(d\theta)$ does, and can be sampled from by the Metropolis/Hastings step given below.

The following version of the MCMC algorithm merges steps 1 (updating Γ) and 3 (updating θ) into a single block, improving the efficiency of the algorithm given in Wolpert and Ickstadt (1997). The algorithm is also more efficient if the sampling distribution $\Pi(ds)$ can be chosen to be nearly proportional to the gamma shape measure $\alpha^\theta(ds)$, so that the density function $\alpha^\theta(s) = \alpha^\theta(ds)/\Pi(ds)$ is nearly constant; in this application we address this by sampling heavily from the downtown region, where most employees work.

For the properties and computational algorithms for evaluating the exponential integral function $E_1(x) \equiv \int_x^\infty u^{-1}e^{-u}\,du$ see (Abramowitz and Stegun, 1964, p. 228). We denote the integral of the kernel by $k^\theta(\mathcal{X}, s) \equiv \int_{\mathcal{X}} k^\theta(x,s)\,dx$. Note that gamma random field realizations are almost-surely discrete, so admit a representation $\Gamma(ds) \equiv \sum_{m<\infty} \upsilon_m \delta_{\sigma_m}(ds)$; in the algorithm we can draw only a large finite number M of the mass points, of course.

Given a prior density function $\pi(\theta)$ on the parameter space Θ, a probability measure $\Pi(ds)$ on $\mathcal{S}$, a kernel function $k^\theta(x,s)$, a gamma random field shape density $\alpha^\theta(s)$ and inverse scale function $\beta^\theta(s)$, a Markov transition density $Q(\theta,\theta^*)$ on Θ, an integer $M \gg N$ and initial points $\theta^0 \in \Theta$ and $\{s_n^0\}_{n\le N}$, successive points can be generated by the following hybrid Gibbs/Metropolis scheme starting at iteration $t = 1$.

1. Gibbs step to update the gamma random field:

Given θ^{t-1} and $S^{t-1} = \{s_n^{t-1}\}_{n\le N}$,

a) Set $\sigma_m^t \equiv s_m^{t-1}$, $1 \le m \le N$, and generate independent $\sigma_m^t \sim \Pi(ds)$, $N < m \le M$;

b) Set $\alpha_m^t \equiv \alpha^{\theta^{t-1}}(\sigma_m^t)$, $\beta_m^t \equiv \beta^{\theta^{t-1}}(\sigma_m^t) + k^{\theta^{t-1}}(\mathcal{X}, \sigma_m^t)$, and $i_m^t \equiv 0$ if $s_n^{t-1} = \sigma_m^t$ for some $n < m$, otherwise $i_m^t \equiv 1$;

c) Generate successive jumps $\{\tau_m\}_{m\le M}$ of a standard Poisson process;

d) Set $\upsilon_m^t \equiv (\tau_m - \tau_{m-1})/\beta_m^t$, $1 \le m \le N$, and $\upsilon_m^t \equiv E_1^{-1}\big((\tau_m - \tau_N)/\alpha_m^t\big)$, $N < m \le M$;

e) Set $\Gamma^t(ds) \equiv \sum_{m\le M} \upsilon_m^t \delta_{\sigma_m^t}(ds)$.

2. Gibbs step to update the augmentation points:

Given θ^{t-1} and Γ^t,

a) Generate independent $S^t = \{s_n^t\}_{n\le N}$ with $\mathrm{pr}(S_n^t = \sigma_m^t) \propto \upsilon_m^t k^{\theta^{t-1}}(x_n, \sigma_m^t)$;

3. Metropolis/Hastings step to update the parameter θ:

Given θ^{t-1} and $S^t = \{s_n^t\}_{n\le N}$,

a) Set $\theta^- = \theta^{t-1}$ and generate a new candidate $\theta^+ \sim Q(\theta^-, \theta^+)$;

b) Set $k_n^- \equiv k^{\theta^-}(x_n, s_n^t), \quad k_n^+ \equiv k^{\theta^+}(x_n, s_n^t)$;

c) Set $\alpha_n^- \equiv \alpha^{\theta^-}(\sigma_n^t), \quad \alpha_n^+ \equiv \alpha^{\theta^+}(\sigma_n^t)$;

d) Set $\beta_n^- \equiv \beta^{\theta^-}(\sigma_n^t) + k^{\theta^-}(\mathcal{X}, \sigma_n^t), \quad \beta_n^+ \equiv \beta^{\theta^+}(\sigma_n^t) + k^{\theta^+}(\mathcal{X}, \sigma_n^t)$;

e) Calculate the Metropolis/Hastings acceptance probability

$$
\begin{aligned}
P^t \equiv\ & \{\pi(\theta^+)/\pi(\theta^-)\}\{Q(\theta^+,\theta^-)/Q(\theta^-,\theta^+)\} \prod_{n\le N} (k_n^+/k_n^-) \\
\times\ & \exp\Big(\sum_{n\le N} \big(i_n^t \log(\alpha_n^+/\alpha_n^-) - \upsilon_n^t(\beta_n^+/\beta_n^-)\big)\Big) \\
\times\ & \exp\Big(-\int_{\mathcal{S}} \log(1 + k^{\theta^+}(\mathcal{X}, s)/\beta^{\theta^+}(s))\, \alpha^{\theta^+}(ds) \\
& \quad + \int_{\mathcal{S}} \log(1 + k^{\theta^-}(\mathcal{X}, s)/\beta^{\theta^-}(s))\, \alpha^{\theta^-}(ds) \Big)\ ;
\end{aligned}
$$

f) Set $\theta^t \equiv \begin{cases} \theta^+ & \text{with probability } \min(1, P^t) \\ \theta^- = \theta^{t-1} & \text{otherwise.} \end{cases}$

4. Increment $t \leftarrow t+1$ and return to step **1**.

Software (in ISO C) is available from the authors on request.

5 Posterior Analysis

Underlying our analysis is the specific Poisson/gamma random field model introduced in section 3.2. For the three parameters we use vague but proper independent exponential prior distributions.

In our survey data set home and work locations are both known. Home locations for every Portland resident are available from census data, but the corresponding work locations are only known for respondents of the survey. Motivated by this practical problem we treat a portion of the survey work locations as if unknown. In order to see how well the model is able to replicate these "unknown" work locations, we omit a range (10%, 33%, 50%, 66%, and 90%) of the work locations.

For our analysis 6,000 iterations of the MCMC scheme introduced in section 4 (after a burn-in period of 2,000 steps) turned out to be more than adequate for convergence. For the different work location percentages treated as unknown, all three Markov chains passed various statistical tests such as Raftery and Lewis, Heidelberger and Welch, and Gelman and Rubin as implemented in CODA (Best et al. (1995)). We used a symmetric

Gaussian random walk for the Metropolis step (3a), with appropriate step sizes to ensure acceptance probabilities of about 40%.

An important goal of our analysis is to obtain estimates for all work locations, both those we treat as known and those we treat as unknown. Figure 4 shows contour plots of these mean work location densities versus the left-out work locations (given as dots), for the different proportions of work locations assumed to be known. It can be seen that the model replicates the work location density quite well; with only 10% of the data treated as known, the high destination density area of downtown Portland (in the upper right corner) is identified. Assuming higher percentages of the work places to be known leads to less smoothed estimates.

From the posterior estimates for all work locations and the given home locations, we can derive the trip OD tables, the traditional objects of central interest for the transportation planning engineer. Since our analysis is independent of any arbitrary grid structure, we can decide a posteriori how to group the OD flows in order to display them. In our example we partitioned the study region into a regular 3 by 3 grid (numbered from 1 to 9, starting at the south-west corner of the study region, assigning numbers 1, 2, 3 moving north in the most western column, 4, 5, 6 moving north in the middle column, and 7, 8, 9 moving north in the most eastern column) and aggregated home locations and mean work location estimates on that grid. Figure 5 shows the resulting OD tables for the different proportions of work locations assumed to be known. It can be seen that destination grid zone 9 is most popular, followed by zones 6 and 8. The diagonal entries include people working at their home locations.

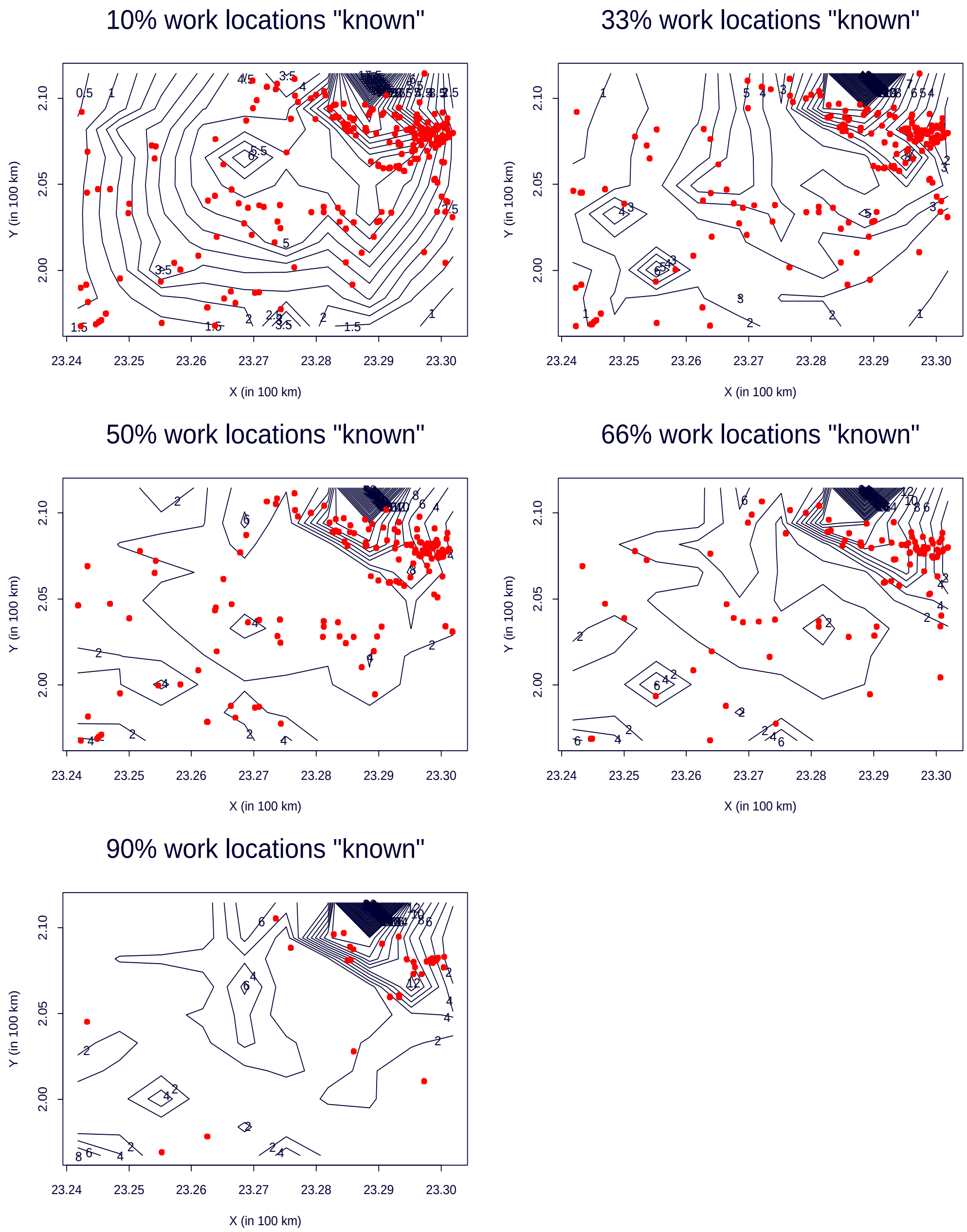


FIGURE 4. Estimated mean posterior work locations.

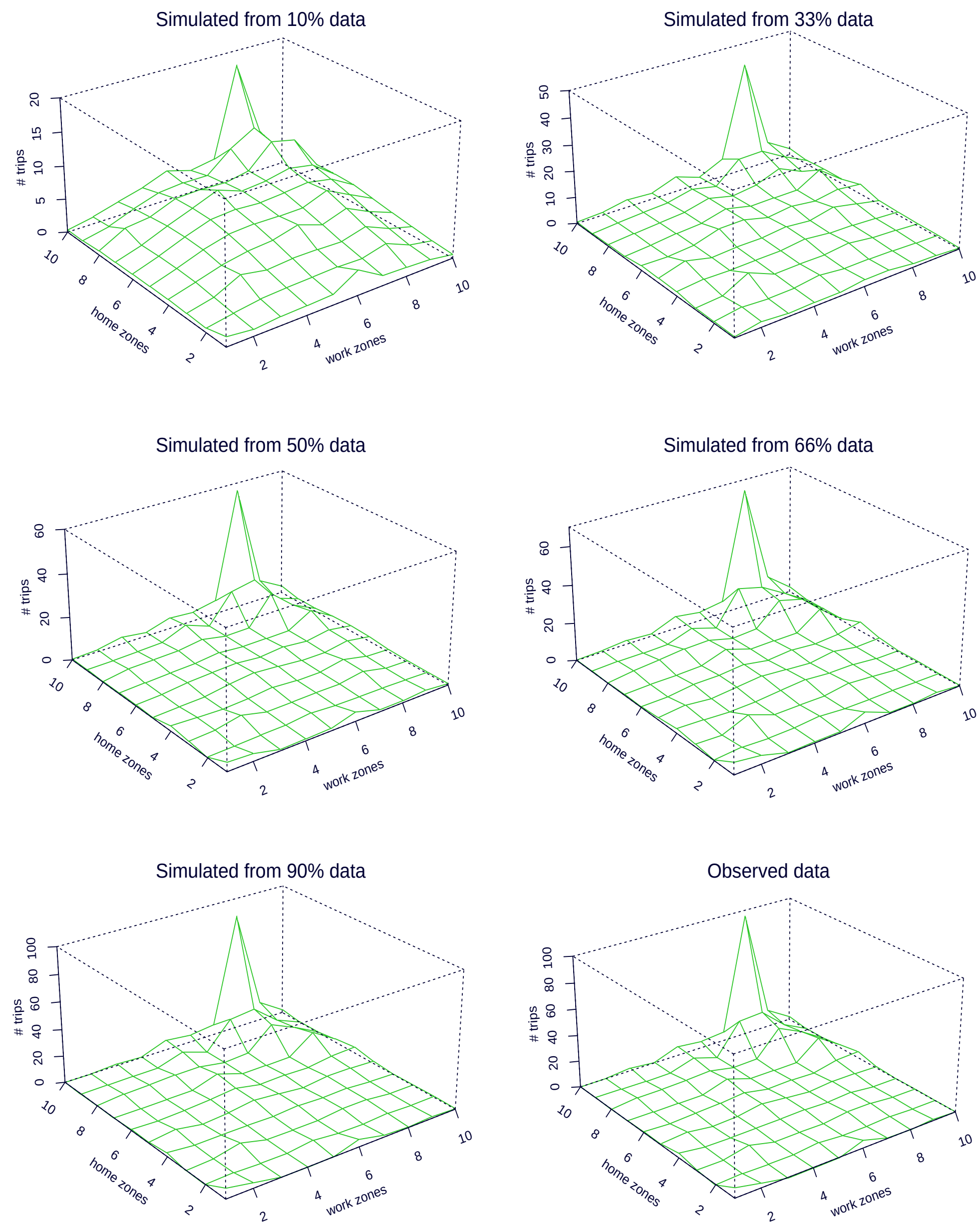


FIGURE 5. Estimated trip OD tables.

Figure 6 shows the prior (curve) and the estimated posterior (histogram) density functions for the home location density θ_1 (in trips/ha), the large-scale extra-Poisson variation θ_2, and the interaction distance θ_3 (in km) for 50% of the work locations assumed to be unknown. The posterior means and standard deviations are summarized in table 1.

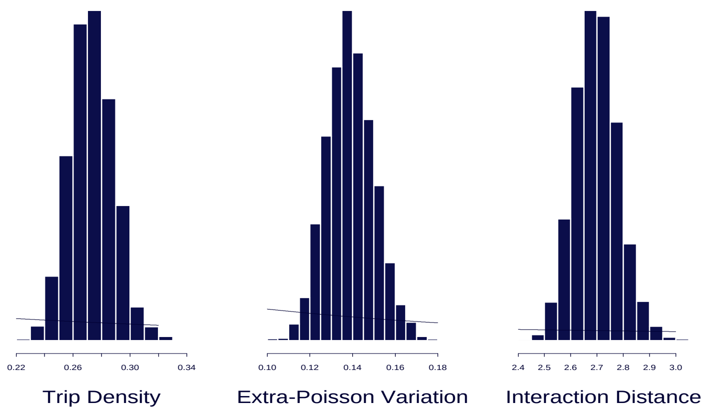


FIGURE 6. Prior densities and posterior histograms.

	Units	Par.	Mean	Std. Dev.
Daily Trip Density	trips/ha	θ_1	0.274	0.015
Extra-Poisson Variation		θ_2	0.139	0.011
Interaction Distance	km	θ_3	2.70	0.083

TABLE 1: Posterior summary statistics.

The interaction (i.e. commuting) distance has a posterior mean of 2.7 km, suggesting that employment opportunities at distances of up to a few times this distance (five or six km) have a strong effect on the choice of home location.

Results similar to those of table 1 are obtained for 10%, 33%, 66%, and 90% of the work locations treated as known. The posterior means of θ_1 and θ_2 remain unchanged, while the mean of the interaction distance θ_3 increases with increasing percentages of data points treated as known, and turns out to be 1.84 km, 2.26 km, 2.80 km, 2.91 km, respectively.

6 Discussion

The hierarchical Poisson gamma random field models underlying the present analysis are a useful tool for estimating a continuous work location surface, enabling us in turn to look at OD flows on any arbitrary discretization of both origin and destination space. We model possible spatial correlation explicitly, and allow for incorporating observed data on our hierarchical gamma level.

The model used here can be extended to include additional covariate information. Individual covariates (e.g. age, gender) can be incorporated using a marked Poisson random field on the lowest level of the hierarchy; spatial covariate information (e.g. employment density per TAZ, land use, GIS information about rivers and parks) can be reflected in a spatially dependent measure $\alpha(ds)$ for the gamma random field and might lead to less smoothing for the work location densities. We are also exploring kernels based on different cost functions, e.g. the travel time between home location $x \in \mathcal{X}$ and work location $s \in \mathcal{S}$.

The class of hierarchical Poisson/gamma models introduced in section 3 is applicable in a wide range of problems featuring correlated count data including disease mapping, bioabundance and forest growth studies, image analysis and survival analysis.

Acknowledgments

This work was supported by U.S. N.S.F. grant DMS-9313013 to the National Institute of Statistical Sciences, U.S. E.P.A. grant CR822047-01-0, U.S. N.S.F. grant DMS-9626829, and the Deutsche Forschungsgemeinschaft. The authors would like to thank Keith Lawton and his staff at METRO Portland for providing the data, and the U.S. Census Bureau for offering the TIGER Mapping Service. This work was inspired by and would have been impossible without the ideas, advice and encouragement of Eric Pas, who passed away while the work was in progress. We will all miss him deeply.

Bibliography

Abramowitz, M. and Stegun, I. A., eds., (1964), *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, volume 55 of *Applied Mathematics Series*, Washington, D.C.: National Bureau of Standards.

Antoniak, C. E., (1974), "Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems," *Annals of Statistics*, 2, 1152–1174.

Ben-Akiva, M. and Watanatada, T., (1981), "Application of a Continuous Spatial Choice Logit Model," in *Structural Analysis of Discrete Data with Econometric Applications*, eds. C. F. Manski and D. McFadden, 320–343, Cambridge, Mass.: MIT Press.

Besag, J., (1974), "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B (Methodological)*, 35, 192–236, (With discussion).

Besag, J., York, J., and Mollié, A., (1991), "Bayesian image restoration, with two applications in spatial statistics," *Annals of the Institute of Statistical Mathematics*, 43, 1–59, (With discussion).

Best, N. G., Cowles, M. K., and Vines, S. K., (1995), *Convergence diagnosis and output analysis software for Gibbs sampling output, Version 0.3*, Cambridge, UK: MRC Biostatistics Unit.

Casey, H. J., (1955), "Applications to traffic engineering of the law of retail gravitation," *Traffic Quarterly*, 100, 23–35.

Clayton, D. G. and Kaldor, J., (1987), "Empirical Bayes estimates of age–standardized relative risks for use in disease mapping," *Biometrics*, 43, 671–681.

Cressie, N. A. C., (1993), *Statistics for Spatial Data*, New York, NY, USA: John Wiley & Sons.

Escobar, M. D. and West, M., (1998), "Computing nonparametric hierarchical models," in Müller et al. (1998), 0–0.

Gelfand, A. E. and Smith, A. F. M., (1990), "Sampling-based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398–409.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., eds., (1996), *Markov Chain Monte Carlo in Practice*, New York, NY, USA: Chapman & Hall.

Ickstadt, K. and Wolpert, R. L., (1997), "Multiresolution Assessment of Forest Inhomogeneity," in *Case Studies in Bayesian Statistics, Volume III*, eds. C. Gatsonis, J. S. Hodges, R. E. Kass, R. E. McCulloch, P. Rossi, and N. D. Singpurwalla, volume 121 of *Lecture Notes in Statistics*, 371–386, New York, NY, USA: Springer-Verlag.

MacEachern, S. N., (1998), "Computational Methods for Mixture of Dirichlet Process Models," in Müller et al. (1998), 0–0.

Matheron, G., (1963), "Principles of geostatistics," *Economic Geology*, 58, 1246–1266.

Müller, P., Dey, D., and Sinha, D., eds., (1998), *Practical Nonparametric and Semiparametric Bayesian Statistics*, Lecture Notes in Statistics, New York, NY, USA: Springer-Verlag.

Sen, A. and Smith, T. E., (1995), *Gravity Models of Spatial Interaction Behavior*, Berlin: Springer-Verlag.

Smith, T. E., (1987), "Poisson Gravity Models of Spatial Flows," *Journal of Regional Science*, 27, 315–340.

Strauss, D. J., (1975), "A Model for Clustering," *Biometrika*, 62, 467–476.

Tierney, L., (1994), "Markov chains for exploring posterior distributions," *Annals of Statistics*, 22, 1701–1762, (With discussion).

Wolpert, R. L. and Ickstadt, K., (1997), "Poisson/gamma random field models for spatial statistics," Discussion Paper 95-43, Duke University ISDS, USA, to appear in *Biometrika*.

—, (1998), "Simulation of Lévy Random Fields," in Müller et al. (1998), 0–0.