Modeling Space-Time Data Using Stochastic Differential Equations

Jason A. Duan
Yale University, New Haven, CT

Alan E. Gelfand
Duke University, Durham, NC

C. F. Sirmans
University of Connecticut, Storrs, CT [1]

[1]Jason A. Duan is a Postdoctoral Associate in the School of Management at Yale University, New Haven, CT, 06520 (email: jd522@som.yale.edu). Alan E. Gelfand is a Professor in the Department of Statistical Science at Duke University, Durham, NC 27708 (email: alan@stat.duke.edu). C. F. Sirmans is a Professor in the Department of Finance at School of Business, University of Connecticut, Storrs, CT 06169 (email: cf.sirmans@business.uconn.edu). The authors thank Thomas Thibodeau for providing the Dallas house construction data.

Abstract

This paper demonstrates the use and value of stochastic differential equations for modeling space-time data in two common settings. The first consists of point-referenced or geostatistical data where observations are collected at fixed locations and times. The second considers random point pattern data where the emergence of locations and times is random. For both cases, we employ stochastic differential equations to the describe a latent process within a hierarchical model for the data. A motivating problem for the second setting is to model urban development through observed locations and times of new home construction; this gives rise to a space-time point pattern. We show that a spatio-temporal Cox process whose intensity is driven by a stochastic logistic equation is a viable mechanistic model that affords meaningful interpretation for the results of statistical inference. Other applications of stochastic logistic differential equations with space-time varying parameters include modeling population growth and product diffusion, which motivate our first, point-referenced data application. We propose a method to discretize both time and space in order to fit the model. We demonstrate the inference for the geostatistical model through a simulated dataset. Then, we fit the Cox process model to a real dataset taken from the greater Dallas metropolitan area.

Key words: geostatistical data; hierarchical model; logistic differential equation; Markov chain Monte Carlo; point patterns; urban development

1 INTRODUCTION

The contribution of this paper is to demonstrate the use and value of stochastic differential equations (SDE) for modeling space-time data. In particular, we consider two common settings. The first consists of point-referenced or geostatistical data where random observations are taken at locations s and times t. The second considers point pattern data where the assumption is that the locations and times themselves are random. In both cases, we assume $s \in D \subset R^2$ and $t \in (0, T]$. Conceptually, space is not gridded and time is not discretized.

Examples of geostatistical data abound in the literature. Diverse but relevant examples appropriate to our objectives include ecological process models such as photosynthesis, transpiration, and soil moisture; diffusion models for populations, products or technologies; financial processes such as house price and/or land values over time. Here, we employ a customary geostatistical modeling specification, i.e.,

$$Y(t, s) = \Lambda(t, s) + \epsilon(t, s) \tag{1}$$

where $\epsilon(t, s)$ is a space time pure error process (contributing the "nugget") and focus is on the modeling for the "process", $\Lambda(t, s)$. Assuming $\Lambda(t, s)$ is random, we describe it as a realization of a space-time process and thus, we focus on this specification, employing stochastic differential equations.

Space-time point patterns also arise in many settings, e.g., ecology where we might seek the evolution of the range of a species over time by observing the locations of its presences; disease incidence examining the pattern of cases over time; urban development explained using say the pattern of single family homes constructed over time. The random locations and times of these events are customarily modeled with an inhomogeneous intensity surface, say, $\Lambda(t, s)$. Here, the theory of spatial point processes provides convenient tools; the most commonly used and easily interpretable

model is the spatial Poisson process: for any region in the area under study, the total number of observed points is a Poisson random variable with mean equal to the integrated intensity over that region. If the points are emerging dynamically and the exact times of their occurrence are viewed as continuous variables, we turn to the spatio-temporal version of the Poisson process.

There is a substantial literature by now on modeling point-referenced space-time data. The most common tool here is the introduction of spatio-temporal random effects described through a Gaussian process, introducing a suitable space-time covariance function (e.g. Brown et al., 2000; Gneiting, 2002; Stein, 2005). If time is discretized, we can employ dynamic models as in Gelfand et al. (2005). If locations are on a lattice (or are projected to a lattice), we can employ Gaussian Markov random fields (Rue and Held, 2005). For general discussion of such space-time modeling see Banerjee et al. (2004).

There is much less statistical literature on space-time point patterns. However, the mathematical theory of point process on a general carrying space is well established (Daley and Vere-Jones, 1988; Karr, 1991). Cressie (1993) and Møller and Waagepetersen (2004) focus primarily on two-dimensional spatial point processes. Recent developments in spatio-temporal point process modeling include Ogata (1998) with application to statistical seismology and Brix and Møller (2001) with application in modeling weeds. Brix and Diggle (2001), in modeling a plant disease, extend the log Gaussian Cox process (Møller et al., 1998) to a space-time version by using a stochastic differential equation model. See Diggle (2005) for a comprehensive review of the literature.

In either case, we propose to work with stochastic differential equation models. That is, we intentionally specify $\Lambda(t, s)$ through a stochastic differential equation rather than a spatio-temporal process (see, e.g. Banerjee et al., 2004 and references

therein). The prototype of our study is the logistic equation

$$\frac{\partial \Lambda(t,s)}{\partial t} = r(t,s)\Lambda(t,s)\left[1 - \frac{\Lambda(t,s)}{K(s)}\right],$$

where $K(s)$ is the "carrying capacity" (assuming it is time-invariant) and $r(t,s)$ is the "growth rate". The logistic equation finds various practical applications, e.g., population growth in ecology (Kot, 2001), product and technology diffusion in economics (Mahajan and Wind, 1986), and urban development (see Section 4.2).

We intentionally seek to introduce a mechanistic modeling component; we are directly interested in the parameters in our differential equation, parameters such growth rates and carrying capacities, parameters which may be best modeled as spatially varying (see below). We recognize the flexibility that comes with a "purely empirical" model such as a realization of a stationary or nonstationary space-time Gaussian process. However, we seek to incorporate features of the complex process we are trying to model into these space-time surfaces. We seek the interpretation that accompanies a particular choice of differential equation. We demonstrate, through a simulation example, that, when such differential equations guide our data (up to noise), the data can inform about the parameters in these equations and that model performance is preferable to that of model performance employing a random process realization.

Indeed, the real issue for us is how to introduce randomness into a selected differential equation specification. Section 2 is devoted to a brief review of available options and their associated properties. Section 3 develops the geostatistical setting and provides the aforementioned simulation illustration.

A motivating problem for our research is to model urban development through new house construction. The exact locations and times of new houses form a point pattern over space and time. Conceptually, a house can be built at any location and time yielding a continuous-time spatio-temporal point process. However, we must

integrate over time and space in order to achieve a positive expectation of *observing* construction. In particular, if we discretize time and presume that there are only a finite number of periods, then in each period there is a finite random set of observed locations, which can be treated as a spatial point process with intensity cumulated over the time period. Mathematically, the conceptual connection between urban house construction and population growth suggests adapting suitable population growth models in mathematical ecology (Kot, 2001) to model the intensity process.

So, Section 4.1 details the modeling for space-time point patterns and addresses formal modeling and computational issues. Section 4.2 provides a careful analysis of the house construction dataset. Finally, in Section 5, we conclude with a summary and some future directions.

2 SDE MODELS FOR SPATIO-TEMPORAL DATA

2.1 A Brief Review of the Non-spatial Case

Ignoring location for the moment, a usual nonlinear (non-autonomous) differential equation subject to the initial condition takes the form

$$d\Lambda(t) = g(\Lambda(t), t, r\,(t))dt \text{ and } \Lambda\,(0) = \Lambda_0 \tag{2}$$

A natural way to add randomness is to make $r\,(t)$ random with a SDE model:

$$dr\,(t) = a(r(t), t, \beta)dt + b\,(r\,(t)\,, t)\,dB\,(t)$$

where $B(t)$ is an independent increments process on R^1.[1]

[1]This is a very general specification. For example, a common SDE model is $d\Lambda(t) = f(\Lambda(t), t)dt + h(\Lambda(t), t)dB(t)$ where $B(t)$ is Brownian motion over R^1 with f and h the "drift" and "volatility" respectively. This model can be considered as model (2) with $g\,(\Lambda\,(t)\,, t, r\,(t)) = f\,(\Lambda\,(t)\,, t) + r\,(t)\,h\,(\Lambda\,(t)\,, t)$ and $dr\,(t) = r\,(t)\,dt = dB\,(t)$, which implies $d\Lambda(t) = f(\Lambda(t), t)dt + h\,(\Lambda\,(t)\,, t)\,dB(t)$.

Analytic solutions of SDE's are rarely available so we usually employ a first order Euler approximation of the form,

$$\begin{aligned} \Lambda(t+\Delta t) &= \Lambda(t) + g(\Lambda(t), t, r(t))\Delta t \\ \text{and } r(t+\Delta t) &= r(t) + a(r(t), t, \beta)\Delta t + b(r(t), t)\left[B(t+\Delta t) - B(t)\right] \end{aligned}$$

where Δt is the interval between time points and $B(t+\Delta t) - B(t) \sim N(0, \Delta t)$. Higher order (Runge-Kutta) approximations can be introduced but these do not seem to be employed in the statistics literature. Rather, recent work introduces latent variables $\Lambda(t')$'s between $\Lambda(t)$ and $\Lambda(t+\Delta t)$. See, e.g., Elerian et al. (2001); Golightly and Wilkonson (2008) and Stramer and Roberts (2007).

Our prototype is the logistic equation

$$d\Lambda(t) = r(t)\Lambda(t)\left[1 - \frac{\Lambda(t)}{K}\right]dt. \tag{3}$$

Let $r(t) = \mu_r + \zeta(t)$ where μ_r is a constant and $\zeta(t)$ is a mean-zero residual process. Suppose we assume $\zeta(t)\,dt = \sigma_\zeta(t)\,dB(t)$ where $B(t)$ is the standard Brownian motion process with variance being one. This specification introduces a logistic SDE model for $\Lambda(t)$:

$$d\Lambda(t) = \mu_r\Lambda(t)\left[1 - \frac{\Lambda(t)}{K}\right]dt + \sigma_\zeta(t)\Lambda(t)\left[1 - \frac{\Lambda(t)}{K}\right]dB(t), \tag{4}$$

which enjoys many desirable stochastic regularity and stability properties (see Golec and Sathananthan, 2003 and Schurz, 2007). However, the white noise specification for $\zeta(t)$ implies that the $r(t)$ vary independently around μ_r. A richer model for $r(t)$ that introduces temporal correlation (and can be easily extended to the spatio-temporal setting) adopts an Ornstein-Uhlenbeck process for $\zeta(t)$, written as the SDE:

$$d\zeta(t) = -\alpha\zeta(t)\,dt + \sigma_\zeta dB(t), \tag{5}$$

yielding a mean-reverting Ornstein-Uhlenbeck process for $r(t)$:

$$dr(t) = -\alpha\left(\mu_r - r(t)\right)dt + \sigma_\zeta dB(t). \tag{6}$$

Hence, $r(t)$ is a stationary Gaussian process with $\text{cov}(r(t), r(t')) = (\sigma_\zeta^2/\alpha)\exp(-\alpha|t-t'|)$.

Equations (3) and (6) define a continuous time dynamic model where we have a nonstationary process $\Lambda(t)$ and stationary $r(t)$. In the next subsection, we offer an extension to an infinite-dimensional SDE for spatio-temporal models.

2.2 Modeling Spatio-Temporal Data

Returning to (1), we model $\Lambda(t,s)$ with an infinite-dimensional evolution equation subject to the initial condition

$$\frac{\partial \Lambda(t,s)}{\partial t} = g(\Lambda(t,s), s, t; r(t,s)) \text{ and } \Lambda(0,s) = \Lambda_0(s) \tag{7}$$

where $r(t,s)$ is a parameter varying over space and time. Evidently, (7) is a direct extension of (2), replacing $\Lambda(t)$ with a random field $\Lambda(t,\cdot)$. We next add randomness to (7) in the form of an infinite dimensional SDE, continuing to illustrate with the logistic equation.

Given a compact region $D \subset R^2$, at time t and at any location s in D, the space-time logistic model can be intuitively written as

$$\frac{\partial \Lambda(t,s)}{\partial t} = r(t,s)\Lambda(t,s)\left[1 - \frac{\Lambda(t,s)}{K(s)}\right]. \tag{8}$$

We shall see how (8) is derived from (3). Cumulating over $s \in D$, define $\Lambda(t,D) = \int_D \Lambda(t,s)\,ds$. Assume, as in (3), that $\Lambda(t,D)$ follows the *non-spatial* logistic equation model

$$\frac{\partial \Lambda(t,D)}{\partial t} = r(t,D)\Lambda(t,D)\left[1 - \frac{\Lambda(t,D)}{K(D)}\right]. \tag{9}$$

Here $r(t,D)$ is the *average* growth rate of $\Lambda(t,D)$ at t, i.e., $r(t,D) = (\int_D r(t,s)\,ds)/|D|$ with $|D|$ being the area of D while $K(D)$ is the *aggregate* carrying capacity, i.e., $K(D) = \int_D K(s)\,ds$.

The model for $\Lambda(t,s)$ at any location s can be considered as the infinitesimal limit

of the model (9) when D is a neighborhood, δ_s of s whose area goes to zero. Then,

$$\begin{aligned}
\lim_{|\delta_s|\to 0} \Lambda(t,\delta_s)/|\delta_s| &= \lim_{|\delta_s|\to 0} (\int_{\delta_s} \Lambda(t,s')\,ds')/|\delta_s| = \Lambda(t,s);\\
\lim_{|\delta_s|\to 0} K(\delta_s)/|\delta_s| &= \lim_{|\delta_s|\to 0} (\int_{\delta_s} K(s')\,ds')/|\delta_s| = K(s);\\
\lim_{|\delta_s|\to 0} r(t,\delta_s) &= \lim_{|\delta_s|\to 0} \int_{\delta_s} r(t,s')\,ds'/|\delta_s| = r(t,s).
\end{aligned}$$

Therefore, the limit of the *global* model is

$$\begin{aligned}
\lim_{|\delta_s|\to 0} \frac{\partial \Lambda(t,\delta_s)/|\delta_s|}{\partial t} &= \lim_{|\delta_s|\to 0} r(t,\delta_s)\frac{\Lambda(t,\delta_s)}{|\delta_s|}\left[1-\frac{\Lambda(t,\delta_s)/|\delta_s|}{K(\delta_s)/|\delta_s|}\right] \Rightarrow\\
\frac{\partial \Lambda(t,s)}{\partial t} &= r(t,s)\Lambda(t,s)\left[1-\frac{\Lambda(t,s)}{K(s)}\right]
\end{aligned}$$

which is exactly our *local* model (8). In other words, $r(t,s)$, $K(s)$ and $\Lambda(t,s)$ provide coherent behavior under integration across space.

Extending our Ornstein-Uhlenbeck process specification in (5) to the case of infinite dimension, we characterize $\zeta(t,s)$ by an infinite-dimensional SDE

$$\frac{\partial \zeta(t,s)}{\partial t} = -L\zeta(t,s) + \frac{\partial B(t,s)}{\partial t} \tag{10}$$

or equivalently, for $r(t,s) = \mu_r(s) + \zeta(t,s)$:

$$\frac{\partial r(t,s)}{\partial t} = L(\mu_r(s) - r(t,s)) + \frac{\partial B(t,s)}{\partial t}.$$

$L(s)$ is a spatial linear operator given by

$$L(s) = a(s) + \sum_{i=1,\ldots,d} b_i(s)\frac{\partial}{\partial x_i} - \frac{1}{2}\sum_{i=1,\ldots,d} c_i(s)\frac{\partial^2}{\partial x_i^2} \tag{11}$$

where $a(s)$, $b_i(s)$ and $c_i(s)$ are positive deterministic functions and $x_1,\ldots,x_d$ are coordinates of s. $B(t,s)$ is a spatially correlated Brownian motion with the covariance operator C_B over space. Equation (10) defines an infinite-dimensional Ornstein-Uhlenbeck process (Markovian in time) for $\zeta(t,s)$. See Itô (1984) and Atherya et al.

(2006) for theoretical discussions. Here, equation (8) and (10) define a spatio-temporal model with non-stationary and non-Gaussian $\Lambda(t,s)$ and the latent stationary Gaussian $r(t,s)$.

In general, the invariant measure for $\zeta(t,s)$ is a stationary Gaussian process with non-separable spatio-temporal covariance $C_\zeta(t,s)$, if $b_i(s)$ and $c_i(s)$ are not zero. Brown et al. (2000) introduced (10) into spatio-temporal modeling as the continuous-time version of their "blur-generated" spatio-temporal processes with Gaussian kernels. A simplified Ornstein-Uhlenbeck process model used in Brix and Diggle (2001) sets b_i and c_i to zero; the resulting covariance is separable in space and time. For example, with the Matérn spatial covariance function, we have

$$\varrho(t,u) = \sigma^2 \exp(-\alpha t)(\phi|u|)^\nu \kappa_\nu(\phi|u|), \tag{12}$$

where $u = s - s'$ and $\kappa_\nu(\cdot)$ is the modified Bessel function of the second kind.

Whittle (1963) and Jones and Zhang (1997) propose other fractional Stochastic partial differential equation models which can be related to the Ornstein-Uhlenbeck process models above, the latter providing a version with a nonseparable covariance function. Other constructions for non-separable spatio-temporal covariance functions include Cressie and Huang (1999), Gneiting (2002) and Stein (2005).

3 GEOSTATISTICAL MODELS USING SDE

3.1 A Discretized Space-Time Model with White Noise

To apply the SDE development of the previous section to the continuous time geostatistical model in (7), we introduce a discretization. We assume time discretized to small, equally spaced intervals of length Δt, indexed as t_j, $j = 0, 1, ..., J$. The data is considered to be $Y(t_j, s_i)$, i.e., an observation at location s and any $t \in (t_j, t_j + \Delta t)$ is labelled as $Y(t_j, s)$. Then, we assume

$$Y\left(t_j, s_i\right) = \Lambda\left(t_j, s_i\right) + \varepsilon\left(t_j, s_i\right).$$

We have created a spatial dynamic model, i.e., a discrete time state space model with transition driven by an infinite dimensional SDE. Using the logistic equation as an example, the dynamics of the discretized $\Lambda\left(t_j, s\right)$ can be derived as a difference equation using Euler's approximation applied to (8):

$$\begin{aligned} \Delta\Lambda\left(t_j, s\right) &= r(t_{j-1}, s)\Lambda\left(t_{j-1}, s\right)\left[1 - \frac{\Lambda\left(t_{j-1}, s\right)}{K\left(s\right)}\right]\Delta t, & (13) \\ \Lambda\left(t_j, s\right) &\approx \Lambda\left(0, s\right) + \sum_{l=1}^{j} \Delta\Lambda\left(t_{j-1}, s\right). & (14) \end{aligned}$$

We do not have to discretize the space-time model for $r\left(t, s\right)$ if the stationary spatio-temporal Gaussian process $\zeta\left(t, s\right)$ allows direct evaluation of its covariance function. For example, the model (10) with constant $L\left(s\right) = \alpha_r$ has the following closed-form separable covariance function

$$\varrho(t_{j_1} - t_{j_2}, s_{i_1} - s_{i2}) = \sigma_r^2 \exp\left(-\alpha_r \left|t_{j_1} - t_{j_2}\right|\right)\left(\phi_r \left|s_{i_1} - s_{i_2}\right|\right)^{\nu} \kappa_\nu\left(\phi_r \left|s_{i_1} - s_{i_2}\right|\right) \quad (15)$$

which can be directly used in modeling and be estimated. Using this form saves one approximation step.

We still need to model the initial $\Lambda\left(0, s\right)$ and $K\left(s\right)$.[2] For example, we can model the positive $\Lambda\left(0, s\right)$ and $K\left(s\right)$ as log-Gaussian spatial processes with regression forms for the means,

$$\begin{aligned} \log\Lambda\left(0, s\right) &= \mu_\Lambda\left(X_\Lambda\left(s\right), \beta_\Lambda\right) + \theta_\Lambda\left(s\right), \ \theta_\Lambda\left(s\right) \sim \mathrm{GP}\left(0, \varrho_\Lambda\left(s - s'; \varphi_\Lambda\right)\right); \\ \log K\left(s\right) &= \mu_K\left(X_K\left(s\right), \beta_K\right) + \theta_K\left(s\right), \ \theta_K\left(s\right) \sim \mathrm{GP}\left(0, \varrho_K\left(s - s'; \varphi_K\right)\right). \end{aligned}$$

Similarly, $\mu_r(s)$ below (10) can be modeled as $\mu_r\left(X_r\left(s\right), \beta_r\right)$.

[2] In the simulated example in the next section, we assume $K\left(s\right)$ is known and set to 1. A typical example is the rate of product and technology penetration where 100% is the capacity.

Conditioned on $\Lambda(t_j, s)$, the $Y(t_i, s)$ are mutually independent. With data $Y(t_j, s_i)$ at locations $\{s_i, i = 1, \ldots, n\} \subset D$, we can provide a hierarchical model based on the evolution of $\Lambda(t, s)$ and the space-time parameters. We fit this model within a Bayesian framework so completion of the model specification requires introduction of suitable priors on the hyper-parameters.

For simplicity, we suppress the indices t and s and let our observations at time t_j be $y_j = \{y_{j1}, \ldots, y_{jn}\}$ at the corresponding $s_1, \ldots, s_n$ locations. Accordingly, we let Λ_j, $\Delta\Lambda_j$, r_j, K, $\mu_\Lambda(\beta_\Lambda)$, $\mu_K(\beta_K)$, $\mu_r(\beta_r)$, θ_Λ, θ_K and ζ be the vectors of the corresponding functions and processes in our continuous model evaluated at $s_i \in \{s_1, \ldots, s_n\}$. Note that we begin with the initial observations y_0. The hierarchical model for $y_0, \ldots, y_J$ becomes

$$
\begin{aligned}
& y_j|\Lambda_j \sim N\left(\Lambda_j, \sigma_\varepsilon^2 I_n\right), \;\; j = 0, \ldots, J, \\
& \Delta\Lambda_j = \; r_{j-1}\Lambda_{j-1}\left[1 - \frac{\Lambda_{j-1}}{K}\right]\Delta t, \\
& \Lambda_j = \Lambda_0 + \sum_{l=1}^{j-1} \Delta\Lambda_l, \\
& \log\Lambda_0 = \mu_\Lambda(\beta_\Lambda) + \theta_\Lambda, \;\; \theta_\Lambda \sim N\left(0, C_\Lambda(s - s'; \varphi_\Lambda)\right), \\
& \log K = \mu_K(\beta_K) + \theta_K, \;\; \theta_K \sim N\left(0, C_K(s - s'; \varphi_K)\right), \\
& r = \mu_r(\beta_r) + \zeta, \;\; \zeta \sim N\left(0, C_r(t - t', s - s'; \varphi_r)\right), \\
& \beta_\Lambda, \beta_r, \beta_K, \varphi_\Lambda, \varphi_K, \varphi_r \sim \text{priors},
\end{aligned}
\tag{16}
$$

where $\beta_{(\cdot)}$ are the parameters in the mean surface function; C_Λ, C_K and C_r [3] are the covariance matrices. In this model, Λ_0, r and K are latent variables. Note that the Λ_j's are deterministic functions of Λ_0, r and K. The joint likelihood for the $J + 1$

[3] We will write $\mu_\Lambda(\beta_\Lambda)$, $\mu_K(\beta_K)$, $\left\{\mu_r^{(j)}(\beta_r); j = 1, \ldots, J\right\}$, $C_\Lambda(\varphi_\Lambda)$, $C_K(\varphi_K)$ and $C_r(\varphi_r)$ as μ_Λ, μ_K, μ_r, C_Λ, C_K and C_r when there is no ambiguity.

conditionally independent observations and latent variables is

$$\prod_{j=0}^{J}\left\{N\left(y_j|\Lambda_j\left(\Lambda_{j-1},r_j,K\right),\sigma_\varepsilon^2 I_n\right)\right\}N\left(\log\Lambda_0|\mu_\Lambda,C_\Lambda\right)N\left(\log K|\mu_K,C_K\right)N\left(\log r|\mu_r,C_r\right), \tag{17}$$

where we let $r=\{r_0,\ldots,r_{J-1}\}$.

3.2 Bayesian Inference and Prediction

With regard to inference for the model in (16), there are three latent vectors: r, K and Λ_0. The hyper-parameters in this model include the β_r, β_K and β_Λ in the parametric trend surfaces, the spatial random effects ζ, θ_K, θ_Λ and the hyper-parameters φ_r, φ_K, φ_Λ in the covariance functions.

The priors for the hyper-parameters are assumed to have the form

$$\beta_r,\beta_K,\beta_\Lambda\sim\pi\left(\beta_r\right)\cdot\pi\left(\beta_K\right)\cdot\pi\left(\beta_\Lambda\right);\ \varphi_r,\varphi_K,\varphi_\Lambda\sim\pi\left(\varphi_r\right)\cdot\pi\left(\varphi_K\right)\cdot\pi\left(\varphi_\Lambda\right) \tag{18}$$

where each of β_r, β_K, β_Λ, φ_r, φ_K, φ_Λ may represent multiple parameters; for example, we have $\varphi_r=\{\alpha_r,\phi_r,\sigma_r^2,\nu\}$ in the Matérn class covariance function with for the separable model (15). Exact specifications of the priors for the β's and φ's depends on the particular application. For example, if we take $\mu_r(s;\beta_r)=X\left(s\right)\beta_r$, we adopt a weak normal prior $N\left(0,\Sigma_\beta\right)$ for β_r. The parameter σ^2 receives the usual Inverse-Gamma prior. Note that $\Delta\Lambda_j$ in the likelihood (17) for the discretized model are deterministic functions of r, K and Λ_0 defined by (13) and (14). Therefore the joint posterior is proportional to

$$\prod_{j=0}^{J}\left\{N\left(y_j|\Lambda_j\left(\Lambda_{j-1},r_j,K\right),\sigma_\varepsilon^2 I_n\right)\right\}N\left(\log\Lambda_0|\mu_\Lambda,C_\Lambda\right)N\left(\log K|\mu_K,C_K\right)N\left(\log r|\mu_r,C_r\right)\cdot$$
$$\pi\left(\beta_r\right)\pi\left(\beta_K\right)\pi\left(\beta_\Lambda\right)\pi\left(\varphi_r\right)\pi\left(\varphi_K\right)\pi\left(\varphi_\Lambda\right). \tag{19}$$

We simulate the posterior distributions of the model parameters and latent variables in (19) using a Markov Chain Monte Carlo algorithm. Because the intensities

in the likelihood function are very irregular recursive and nonlinear functions of the model parameters and latent variables, it is very difficult to obtain derivatives for a *directed* MCMC. So, instead we use the random-walk Metropolis-Hastings algorithm in the posterior simulation. Each parameter is updated in turn in every iteration of the simulation.

The prediction problem concerns (i) interpolating the past at new locations and (ii) forecasting the future at current and new locations. Indeed, we can hold out the observed data at new locations or in a future time period to *validate* our model. For the logistic growth function, conditioning on the posterior samples of Λ_0, K, r and β_r, β_K, β_Λ, φ_r, φ_K, φ_Λ, we can use spatio-temporal interpolation and temporal extrapolation to obtain $\Delta\Lambda_{J+\Delta J}(s)$ in period $J+\Delta J$ at any new location $s \in D$ by calculating $\mu_r(s,\beta_r)$, $\mu_K(s,\beta_K)$, $\mu_\Lambda(s,\beta_\Lambda)$ and obtaining $\zeta(t,s), t=1,\ldots,J+\Delta J$, $\theta_K(s)$ and $\theta_\Lambda(s)$ by spatio-temporal prediction, and then using (13) and (14) recursively. Because we can obtain a predictive sample for $\Delta\Lambda_{J+\Delta J}(s)$ from the posterior samples of the model fitting, we can infer on any feature of interest associated with the predictive distribution of $\Delta\Lambda_{J+\Delta J}(s)$. The spatial interpolation of past observations at new locations is demonstrated in the subsection below using a simulated example. We will also demonstrate temporal prediction when we apply a Cox-process version of our model to the house construction data in Section 4.

3.3 A Simulated Data Example

In order to see how well we can learn about the true process, we first illustrate the fitting of the models in (16) with a simulated data set. In a study region D of 10×10 square miles shown as the block in Figure 1, we simulate 44 locations at which spatial observations are collected over 30 periods. Therefore our observed spatio-temporal data that constitute a 44×30 matrix. The data are sampled using (16) where we fix

the carrying capacity to be one at all locations.[4] The initial condition Λ_0 is simulated as a log-Gaussian process with a constant mean surface μ_Λ and the Matérn class covariance whose smoothness parameter ν is set to be 3/2. The spatio-temporal growth rate r is simulated using a constant mean μ_r and the separable covariance function (15), where the Matérn smoothness parameter ν is also set to be 3/2. This separable model induces a convenient covariance matrix as the Kronecker product of the temporal and spatial correlation matrices: $\sigma_r^2 \Sigma_t \otimes \Sigma_s$. The values of the fixed parameters in our data simulation are presented in Table 1.

We use the simulated r and Λ_0 and the transition equation (13) recursively to obtain $\Delta\Lambda_j$ and Λ_j for each of the 30 periods. The observed data y_j are sampled as mutually independent given Λ_j with the random noise ε_j. The data at four selected locations (marked as 1, 2, 3, and 4 in Figure 1) are shown as small circles in Figure 2. We leave out the data at four randomly chosen locations (shown in diamond shape and marked as A, B, C and D in Figure 1) for spatial prediction and out-of-sample validation for our model.

We fit the same model (16) to the data at the remaining 40 locations (hence a 40×30 spatio-temporal data set). We use very vague priors for the constant means: $\pi(\mu_\Lambda) \sim N(0, 10^8)$ and $\pi(\mu_r) \sim N(0, 10^8)$. We use natural conjugate priors for the precision parameters (inverse of variances) of r and Λ_0: $\pi(1/\sigma_r^2) \sim Gamma(1,1)$ and $\pi(1/\sigma_\Lambda^2) \sim Gamma(1,1)$. The positive parameter for the temporal correlation of r also has a vague log-normal prior: $\pi(\alpha_r) \sim log\text{-}N(0, 10^8)$. Because the spatial range parameters ϕ_r and ϕ_Λ are only weakly identified (Zhang, 2004), we only use informative and discrete prior for them. Indeed we have chosen 20 values (from 0.1 to 2.0) and assume uniform priors on them for both ϕ_r and ϕ_Λ.

[4] We may view the data as the household adoption rates for a certain durable product (e.g. air conditioners, motorcycles) in 44 cities over 30 months. The capacity being one means 100% adoption. Household adoption rates are collected by surveys with measurement errors.

We use the random-walk Metropolis-Hastings algorithm to simulate posterior samples of r and Λ_0. We draw the entire vector of Λ_0 for all forty locations as a single block in every iteration. Because r is very high-dimensional (r being a 40×30 matrix), we cannot draw the entire matrix of r as one block and have satisfactory acceptance rate (between 20% to 40%). Our algorithm divides r into 40 row blocks (location-wise) in every odd-numbered iteration and 30 column blocks (period-wise) in every even numbered iteration. Each block is drawn in one Metropolis step. We find the posterior samples start to converge after about 30,000 iterations. Given the sampled r and Λ_0, the mean parameters μ_r, μ_Λ and the precision parameters $1/\sigma_r^2$ and $1/\sigma_\Lambda^2$ all have conjugate priors, and therefore their posterior samples are drawn with Gibbs samplers. ϕ_r and ϕ_Λ have discrete priors and therefore discrete Gibbs samplers too. We also use the random-walk Metropolis-Hastings algorithm to draw α_r.

We obtain 200,000 samples from the algorithm and discard the first 100,000 as burn-in. For the posterior inference, we use 4,000 subsamples from the remaining 100,000 samples, with a thinning equal to 25. It takes about 15 hours to finish the computation using the R statistical software on an Intel Pentium 4 3.4GHz computer with 2GB of memory. The posterior mean and 95% equal-tail quantile for the model parameters are presented in Table 1. Evidently we are recovering the true parameter values very well. Figure 2 displays the posterior mean of the growth curves and 95% Bayesian predictive bound for the four locations (1, 2, 3 and 4), compared with the actual latent growth curve $\Lambda(t, s)$ and observed data. Up to the uncertainty in the model we approximate the actual curves very well. The fitted mean growth curves almost perfectly overlap with the actual simulated growth curves. The empirical coverage rate of the Bayesian predictive bounds is 93.4%.

We use the Bayesian spatial interpolation in Section 3.2 to obtain the predictive growth curve for four new locations (A, B, C and D). In Figure 3 we display the means

of the predicted curves and 95% Bayesian predictive bounds, together with the hold-out data. We can see the spatial prediction captures the patterns of the hold-out data very well. The predicted mean growth curves overlap with the actual simulated growth curves very well except for location D, because location D is rather far from all the observed locations. The empirical coverage rate of the Bayesian predictive bounds is 95.8%.

We also fit the following customary process realization model with space-tiem random effects to the simulated data set

$$y_j = \mu + \xi_j + \varepsilon_j;\ \varepsilon_j \sim N\left(0, \sigma_\varepsilon^2 I_n\right),\ \ j = 0, \ldots, J \tag{20}$$

where the random effects $\xi = [\xi_0, \ldots, \xi_j]$ come from a Gaussian process with a separable spatio-temporal correlation of the form:

$$C_\xi(t - t', s - s') = \sigma_\xi^2 \exp\left(-\alpha_\xi \left|t - t'\right|\right) \left(\phi_\xi \left|s - s'\right|\right)^\nu \kappa_\nu \left(\phi_\xi \left|s - s'\right|\right),\ \nu = \frac{3}{2}\ . \tag{21}$$

Comparison of model performance between our model in (16) and the model in (20) is conducted using spatial prediction at the 4 new locations in Figure 1. The computational cost of the model in (20) is, of course, much lower; this model can be fitted with a Gibbs sampler and requires one hour for 100,000 iterations. After we discard 20,000 as burn-in and thin the remaining samples to 4,000, we conduct the prediction on the four new sites (A, B, C and D). In Figure 4 we display the means of the predicted curves and 95% Bayesian predictive bounds, together with the hold-out data. For the four hold-out sites, the average mean square error of the model (16) is 1.75×10^{-3} versus 3.34×10^{-3} of the model (20); the average length of the 95% predictive bounds for the model (16) is 0.29 versus 0.72 for the model (20). It is evident that the prediction results under the benchmark model are substantially worse than those under our model (16); the mean growth curves are less accurate and less smooth. The 95% Bayesian predictive bounds are also much wider.

4 SPACE-TIME COX PROCESS MODELS USING SDE

4.1 The Model

Here, we turn to the use of a SDE to provide a Cox process model for space-time point patterns. Let D again be a fixed region and let X_T denote an observed space-time point pattern within D over the time interval $[0, T]$. The Cox process model assumes a positive space-time intensity that is a realization of a stochastic process. Denote the stochastic intensity by $\Omega(t, s), s \in D, t \in [0, T]$. In practice, we may only know the spatial coordinates of all the points whereas the time coordinates are only known to be in the time interval $[0, T]$.[5] The integrated process $\Lambda(T, s) = \int_0^T \Omega(t, s)\, dt$, provided that $\Omega(t, s)$ is integrable over $[0, T]$, is the intensity for this kind of point patterns. We may also know multiple subintervals of $[0, T]$: $[t_1 = 0, t_2), \ldots, [t_{J-1}, t_J = T]$, and observe a point pattern in each subinterval. These data constitute a series of discrete-time spatio-temporal point patterns, which are denote by $X_{[t_1=0,t_2)}, \ldots, X_{[t_{j-1},t_N=T]}$. The integrated process also provides stochastic intensities for these point patterns

$$\Delta\Lambda_j(s) = \Lambda(t_j, s) - \Lambda(t_{j-1}, s) = \int_{t_{j-1}}^{t_j} \Omega(t, s)\, dt.$$

In this paper, we will model the dynamics of these point patterns by an infinite dimensional SDE subject to the initial condition for $\Lambda(t, s)$. Note an equivalent infinite dimensional SDE for $\Omega(t, s)$ can also be derived from the equation for $\Lambda(t, s)$.

If we observed the complete space-time data $X_T(t, s)$, we can still approximate temporally dependent $X_{[t_1=0,t_2)}, \ldots, X_{[t_{j-1},t_N=T]}$ will provide a good approximation to $X_T(t, s)$, when the time intervals are sufficiently small (Brix and Diggle, 2001).

[5] For example, in our house construction data, for Irving, TX, we only have the geo-coded locations of the newly constructed houses within a year. The exact time when the construction of a new house starts is not available.

Moreover, this will also facilitate the use of the approximated intensity

$$\Delta\Lambda_j(s) = \Lambda(t_j, s) - \Lambda(t_{j-1}, s) = \int_{t_{j-1}}^{t_j} \Omega(\tau, s)\, d\tau \approx \Omega(t_{j-1}, s)(t_j - t_{j-1}).$$

As a concrete example, return to the house construction dataset mentioned in Section 1. Let $X_j = X_{[t_{j-1},t_j)} = x_j$ be the observed set of locations of new houses built in region D and period j=$[t_{j-1}, t_j)$. We can apply the Cox process model to X_j and assume that the stochastic intensity $\Lambda(t, s)$ follows the logistic equation model (8). We can also apply the discretized version (13) to $\Delta\Lambda_j(s)$.

Let our initial point pattern be x_0 and the intensity be $\Lambda_0(s) = \Lambda(0, s) = \int_{-\infty}^{0} \Omega(\tau, s)\, d\tau$. The hierarchical model for the space-time point patterns is merely the model (16) with the first stage of the hierarchy replaced by the following

$$\begin{aligned} x_j|\Delta\Lambda_j &\sim \text{Poisson Process}(D, \Delta\Lambda_j), \;\; j = 1, \ldots, J \\ x_0|\Lambda_0 &\sim \text{Poisson Process}(D, \Lambda_0), \end{aligned} \tag{22}$$

where we suppress the indices t and s again for the periods $t_1, \ldots, t_J$. Note that, unlike in (16), the intensity $\Delta\Lambda_j$ for x_j must be positive. Therefore, we model the growth rate r as a log-process, that is

$$\log r(t, s) = \mu_r(s; \beta_r) + \zeta(t, s), \;\; \zeta \sim \text{GP}(0, \varrho(t - t', s - s'; \varphi_r)). \tag{23}$$

The J spatial point patterns are conditionally independent given the space-time intensity so the likelihood is

$$\prod_{j=1}^{J} \left\{ \exp\left(-\int_D \Delta\Lambda_j(s)\, ds \right) \prod_{i=1}^{n_j} \Delta\Lambda_j(x_{ji}) \right\} \cdot \exp\left(-\int_D \Lambda_0(s)\, ds \right) \prod_{i=1}^{n_0} \Lambda_0(x_{0i}). \tag{24}$$

This likelihood is more difficult to work with than that in (17). There is a stochastic integral in (24), $\int_D \Delta\Lambda_j(s)\, ds$, which must be approximated in model fitting by a Riemann sum. To do this, we divide the geographical region D into M cells and assume the intensity is homogeneous within each cell. Let $\Delta\Lambda_j(m)$ and $\Lambda_0(m)$ denote

this average intensity in cell m. Let the area of cell m be $A(m)$. Then, the likelihood becomes

$$\prod_{j=1}^{J}\left[\exp\left(-\sum_{m=1}^{M}\Delta\Lambda_j(m)A(m)\right)\prod_{m=1}^{M}\Delta\Lambda_j(m)^{n_{jm}}\right] \cdot\exp\left(-\sum_{m=1}^{M}\Lambda_0(m)A(m)\right)\prod_{m=1}^{M}\Lambda_0(m)^{n_{0m}} \tag{25}$$

where n_{jm} is the number of point in cell m in period j. Our parameter processes $r(t_j, s)$ and $K(s)$ are also approximated accordingly as $r(t_j, m)$ and $K(m)$, which are homogeneous in each cell m.

4.2 Modeling House Construction Data for Irving, TX

Our real house construction dataset consists of the geo-coded locations and years of the newly constructed residential houses in Irving, TX from 1901 to 2002. Figure 5 shows how the city grows from the early 1950's to the late 1960's. Irving started to develop in the early 1950's and the outline of the city was already in its current shape by the late 1960's. The city became almost fully developed by the early 1970's with much fewer new constructions after that era. Therefore, for our data analysis, we select the period from 1951 through 1969 when there was rapid urban development. In our analysis, we use the data from year 1951–1966 to fit our model and hold out the last three years (1967, 1968 and 1969) for prediction and model validation.

As shown in the central block of Figure 6, our study region D in this example is a square of 5.6×5.6 square miles with Irving, TX in the middle. This region is geographically disconnected from other major urban areas in Dallas County, which enables us to isolate Irving for analysis. We divide the region into 100 (10×10) equally spaced grid cells shown in Figure 6. Within each cell, we model the point pattern with a homogeneous Poisson process given $\Delta\Lambda_j(m)$. The corresponding $\Lambda_0(m)$, $K(m)$ and

$r(j,m)$ are collected into vectors Λ_0, K, and r which are modeled as follows.

$$\begin{aligned}
\log \Lambda_0 &= \mu_\Lambda + \theta_\Lambda,\ \theta_\Lambda \sim N(0, C_\Lambda) \\
\log K &= \mu_K + \theta_K,\ \theta_K \sim N(0, C_K) \\
\log r &= \mu_r + \zeta,\ \zeta \sim N(0, C_r)
\end{aligned}$$

where the spatial covariance matrix C_Λ and C_K are constructed using the Matérn class covariance function with distances between the centroids of the cells. The smoothness parameter ν is set to be 3/2. The variances σ^2_Λ, σ^2_K and range parameters ϕ_Λ and ϕ_K are to be estimated. The spatio-temporal log growth rate r is assumed to have a separable covariance matrix $C_r = \sigma^2_r \Sigma_t \otimes \Sigma_s$, where the spatial correlation Σ_s is also constructed as a Matérn class function of the distances between cell centroids with smoothness parameter ν being set to 3/2. The temporal correlation Σ_t is of exponential form as in (15). The variance σ^2_r, spatial and temporal correlation parameters ϕ_r and α_r are to be estimated.

We use very vague priors for the parameters in the mean function: $\pi(\mu_\Lambda)$, $\pi(\mu_K)$, $\pi(\mu_r) \overset{ind}{\sim} N(0, 10^8)$. We use natural conjugate priors for the precision parameters (inverse of variances) of r and Λ_0: $\pi(1/\sigma^2_\Lambda)$, $\pi(1/\sigma^2_K)$, $\pi(1/\sigma^2_r) \overset{ind}{\sim} Gamma(1,1)$. The temporal correlation parameter of r also has a vague log-normal prior: $\pi(\alpha_r) \sim$ *log-*$N(0, 10^8)$. Again, the spatial range parameters ϕ_Λ ϕ_K and ϕ_r are only weakly identified (Zhang, 2004), so we use informative, discrete priors for them. Indeed we have chosen 40 values (from 1.1 to 5.0) and assume uniform priors on them for ϕ_Λ ϕ_K and ϕ_r.

We use the same random-walk Metropolis-Hastings algorithm as in the simulation example to simulate posterior samples with the same tuning of acceptance rates. As a *production run* we obtain 200,000 samples from the algorithm and discard the first 100,000 as burn-in. For the posterior inference, we use 4,000 subsamples from the

remaining 100,000 samples, with a thinning equal to 25. The posterior mean and 95% equal-tail quantile for the model parameters are presented in Table 2.

Figure 7 shows the posterior mean growth curves and 95% Bayesian predictive bound for the intensity in the four blocks (marked as block 1, 2, 3 and 4) in Figure 6. Comparing with the observed number of houses in the four blocks from 1951 to 1966, we can see the estimated curves fit the data very well.[6]

In Figure 8 we display the posterior mean intensity surface for year 1966 and the predictive mean intensity surfaces for years 1967, 1968 and 1969. We also overlay the actual point patterns of the new homes construct in those four years on the intensity surfaces. Figure 8 shows that our model can forecast the major areas of high intensity, hence high growth very well. For example, in the upper left corner, the intensity continues rising from 1966 to 1968 and starts to wane in 1969. We can see increasing numbers of houses are built from 1966 to 1968 and much fewer are built in 1969. In the lower left part of the plots near the bottom, we can see areas of high intensity gradually shift down to the south and the house construction pattern confirms this trend too.

5 DISCUSSION

We have illustrated the use of stochastic differential equation models to handle both geostatistical and point pattern space-time data. We have illustrated with the nonlinear logistic specification but can handle more general models as given in (8) with (10) and (11). We have argued that a mechanistic explanation, up to error, can be more attractive in explaining process behavior than a customary Gaussian process

[6]The growth curves for the house construction data are much smoother than those in our simulated data example in Section 3.3. Although our fitted mean growth curves seem to match the data too perfectly, we do not think we overfit because our hold-out prediction results are sufficiently accurate too.

realization. In fact, we have shown that if the data is driven by the SDE model, we achieve improved inference using the former relative to the latter.

Our application to the house construction data is really only a first attempt to incorporate a structured growth model into a spatio-temporal point process to afford insight into the mechanism of urban development. However, if it is plausible to assume that the damping effect of growth is controlled by the carrying capacity of a logistic model, then it is not unreasonable to assume the growth rate is mean-reverting. Of course, we can envision several ways to make the model more realistic and these suggest directions for future work. We might have additional information at the locations to enable a so-called marked point process. For instance, we might assign the house to a group according to its size. Fitting the resultant multivariate Cox process can clarify the intensity of development. We could also have useful covariate information on zoning or introduction of roads which could be incorporated into the modeling for the rates. We can expect "holes" in the region - parks, lakes, etc. - where no construction can occur. For locations in these regions, we should impose zero growth. Finally, it may be that growth triggers more growth so that so-called self exciting process specifications might be worth exploring.

References

Atherya, S., Bass, R., Gordina, M., and Perkins, E. (2006), "Infinite-dimensional stochastic differential equations of Ornstein-Uhlenbeck type," *Stochastic Processes and Their Applications*, 116, 381–406.

Banerjee, S., Carlin, B., and Gelfand, A. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC.

Brix, A. and Diggle, P. (2001), "Spatiotemporal prediction for log-Gaussian Cox processes," *Journal of the Royal Statistical Society: Series B*, 63, 823–841.

Brix, A. and Møller, J. (2001), "Space-time multi type log Gaussian Cox Processes with a View to Modelling Weeds." *Scandinavian Journal of Statistics*, 28, 471–488.

Brown, P., Karesen, K., Roberts, G., and Tonellato, S. (2000), "Blur-generated non-separable space-time models," *Journal of the Royal Statistical Society: Series B*, 62, 847–860.

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley, 2nd ed.

Cressie, N. and Huang, H. (1999), "Classes of nonseparable, spatio-temporal stationary covariance functions," *Journal of the American Statistical Association*, 94, 1330–1340.

Daley, D. and Vere-Jones, D. (1988), *Introduction to the Theory of Point Processes*, New York: Springer Verlag.

Diggle, P. (2005), "Spatio-temporal point processes: methods and applications," Working Paper, Johns Hopkins University.

Elerian, O., Chib, S., and Shephard, N. (2001), "Likelihood inference for discretely observed non-linear diffusions," *Econometrica*, 69, 959–993.

Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005), "Spatial process modelling for univariate and multivariate dynamic spatial data," *Environmetrics*, 16, 465–479.

Gneiting, T. (2002), "Nonseparable, stationary covariance functions for space-time data," *Journal of the American Statistical Association*, 97, 590–600.

Golec, J. and Sathananthan, S. (2003), "Stability analysis of a stochastic logistic model," *Mathematical and Computer Modelling*, 38, 585–593.

Golightly, A. and Wilkonson, D. (2008), "Bayesian inference for nonlinear multivariate diffusion models observed with error," *Computational Statistics and Data Analysis*, 52, 1674–1693.

Itô, K. (1984), *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces*, Philadelphia: Society for Industrial and Applied Mathematics.

Jones, R. and Zhang, Y. (1997), "Models for continuous stationary space-time processes," in *Modelling Longitudinal and Spatially Correlated Data*, eds. Gregoire, G., Brillinger, D., Diggle, P., Russek-Cohen, E., Warren, W., and Wolfinge, R., Springer, New York, pp. 289–298.

Karr, A. (1991), *Point Processes and Their Statistical Inference*, New York: Marcel Dekker, 2nd ed.

Kot, M. (2001), *Elements of Mathematical Ecology*, Cambridge Press.

Mahajan, V. and Wind, Y. (1986), *Innovation Diffusion Models of New Product Acceptance*, Harper Business.

Møller, J., Syversveen, A., and Waagepetersen, R. (1998), "Log Gaussian Cox processes," *Scandanavian Journal of Statistics*, 25, 451–482.

Møller, J. and Waagepetersen, R. (2004), *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC Press.

Ogata, Y. (1998), "Space-time point-process models for earthquake occurrences," *Annals of the Institute for Statistical Mathematics*, 50, 379–402.

Rue, H. and Held, L. (2005), *Gaussian Markov random fields: theory and applications*, Chapman & Hall/CRC.

Schurz, H. (2007), "Modeling, analysis and discretization of stochastic logistic equations," *International Journal of Numerical Analysis and Modeling*, 4, 178–197.

Stein, M. L. (2005), "Space-time covariance functions," *Journal of the American Statistical Association*, 100, 310–321.

Stramer, O. and Roberts, G. (2007), "On Bayesian analysis of nonlinear continuous-time autoregression models," *Journal of Time Series Analysis*, 28, 744–762.

Whittle, P. (1963), "Stochastic processes in several dimensions," *Bulletin of the International Statistical Institute*, 40, 974–994.

Zhang, H. (2004), "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics," *Journal of the American Statistical Association*, 99, 250–261.

Table 1: Parameters and their posterior inference for the simulated example

Model Parameters	True Value	Posterior Mean	95% Equal-tail Interval
μ_Λ	-4.2	-4.14	(-4.88, -3.33)
σ_Λ	1.0	0.91	(0.62, 1.46)
ϕ_Λ	0.7	0.77	(0.50, 1.20)
σ_ε	0.05	0.049	(0.047, 0.052)
μ_r	0.24	0.24	(0.22, 0.26)
σ_r	0.08	0.088	(0.077, 0.097)
ϕ_r	0.7	0.78	(0.60, 1.10)
α_r	0.6	0.64	(0.51, 0.98)

Table 2: Posterior inference for the house construction data in Section 4.2

Model Parameters	Posterior Mean	95% Equal-tail Interval
μ_Λ	2.78	(2.15, 3.40)
σ_Λ	1.77	(1.49, 2.11)
ϕ_Λ	3.03	(2.70, 3.20)
μ_r	-2.76	(-3.24, -2.29)
σ_r	2.48	(2.32, 2.68)
ϕ_r	4.09	(3.70, 4.30)
α_r	0.52	(0.43, 0.62)
μ_K	6.49	(5.93, 7.01)
σ_K	1.17	(1.02, 1.44)
ϕ_K	1.91	(1.60, 2.20)

Figure 1: Locations for the simulated data example in Section 3.3.

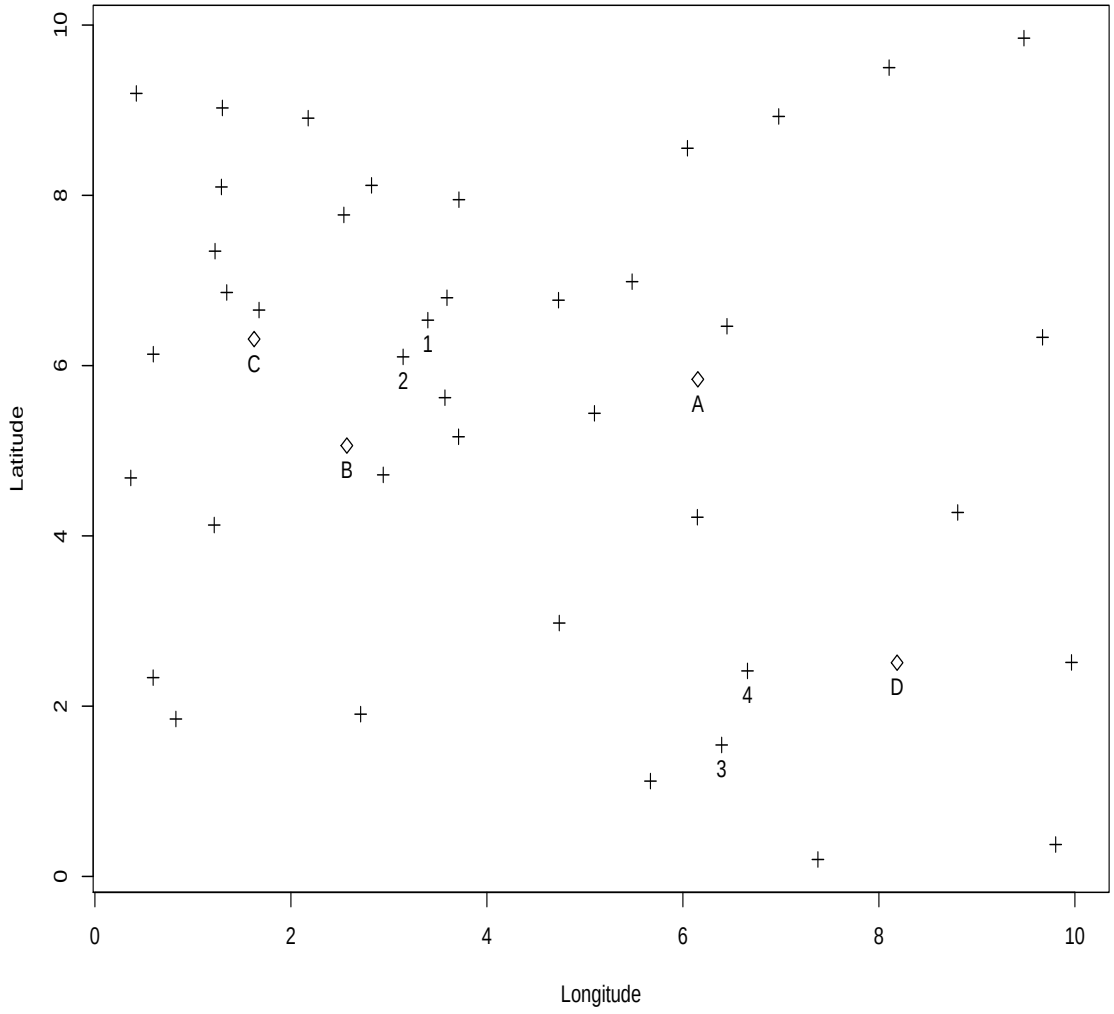

Figure 2: Observed space-time geostatistical data at 4 locations, actual (dashed line) and fitted mean growth curves (solid line), and 95% predictive intervals (dotted line) by our model (16) for the simulated data example.

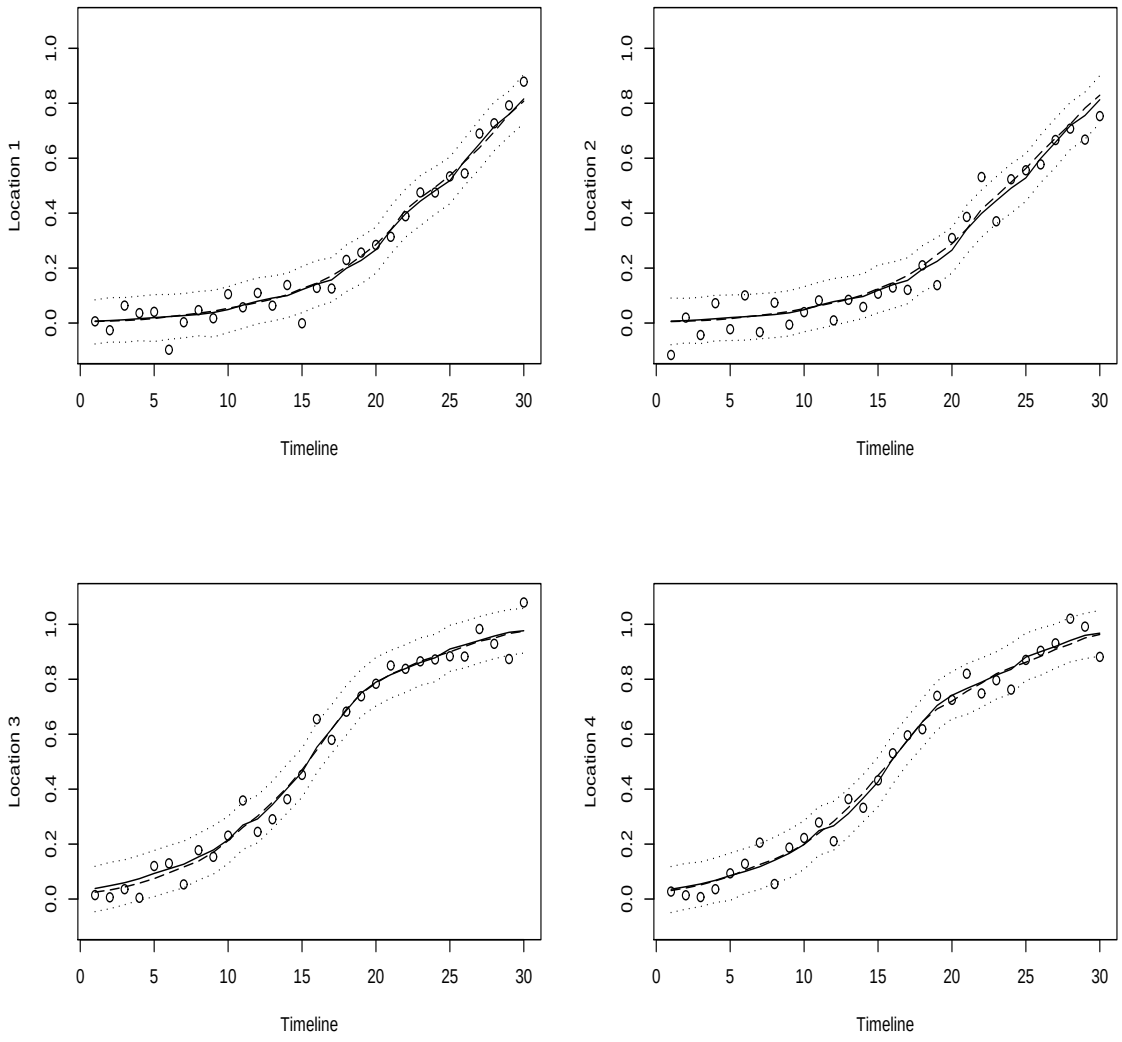

Figure 3: Hold-out space-time geostatistical data at 4 locations, actual (dashed line) and predicted mean growth curves (solid line) and 95% predictive intervals (dotted line) by our model (16) for the simulated data example.

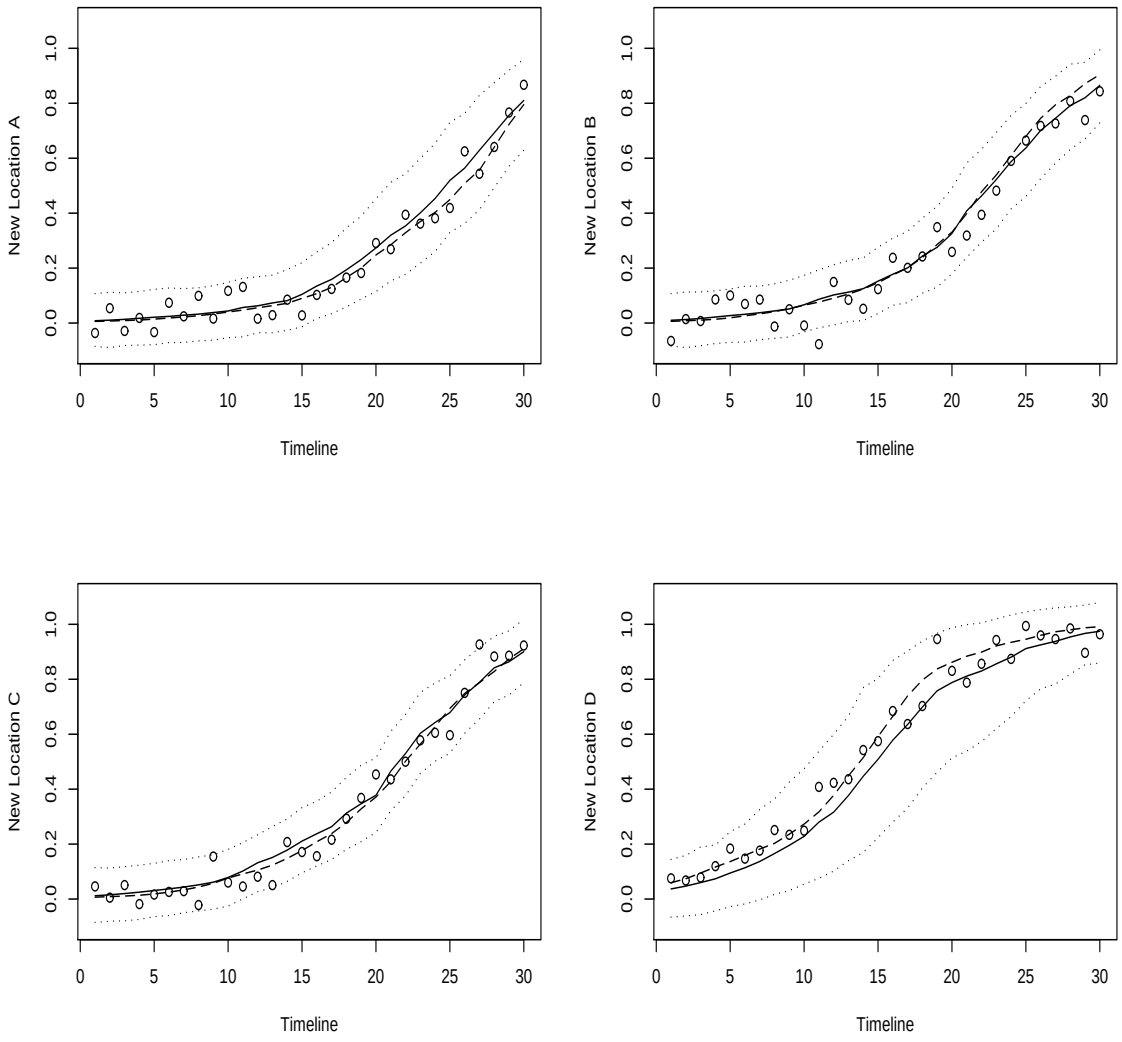

Figure 4: Hold-out space-time geostatistical data at 4 locations, actual (dashed line) and predicted mean growth curves (solid line) and 95% predictive intervals (dotted line) by the benchmark model (20) for the simulated data example.

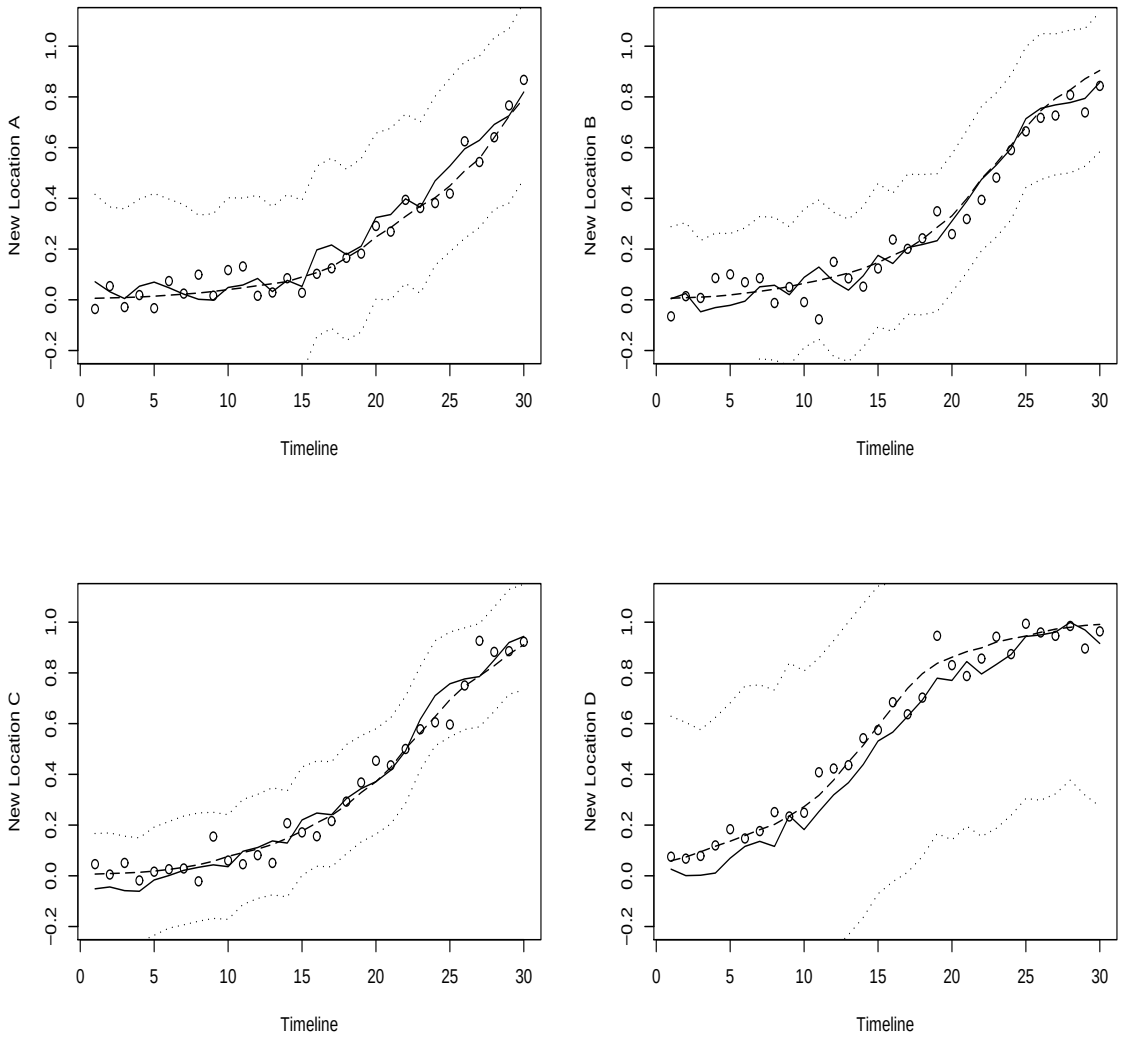

Figure 5: Growth of residential houses in Irving, TX.

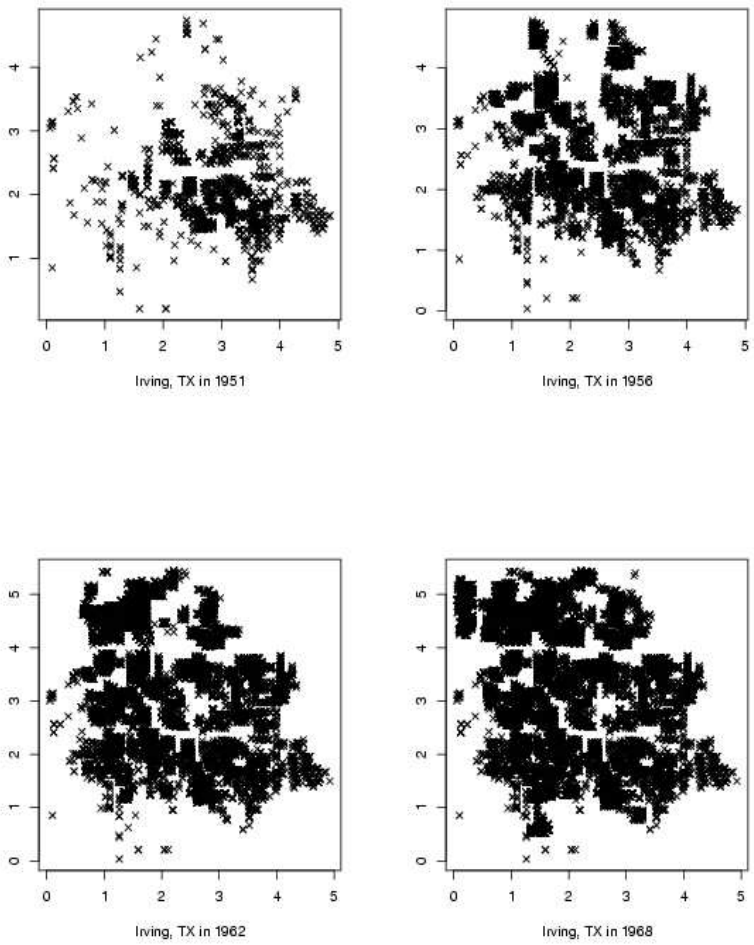

Figure 6: The gridded study region encompassing Irving, TX.

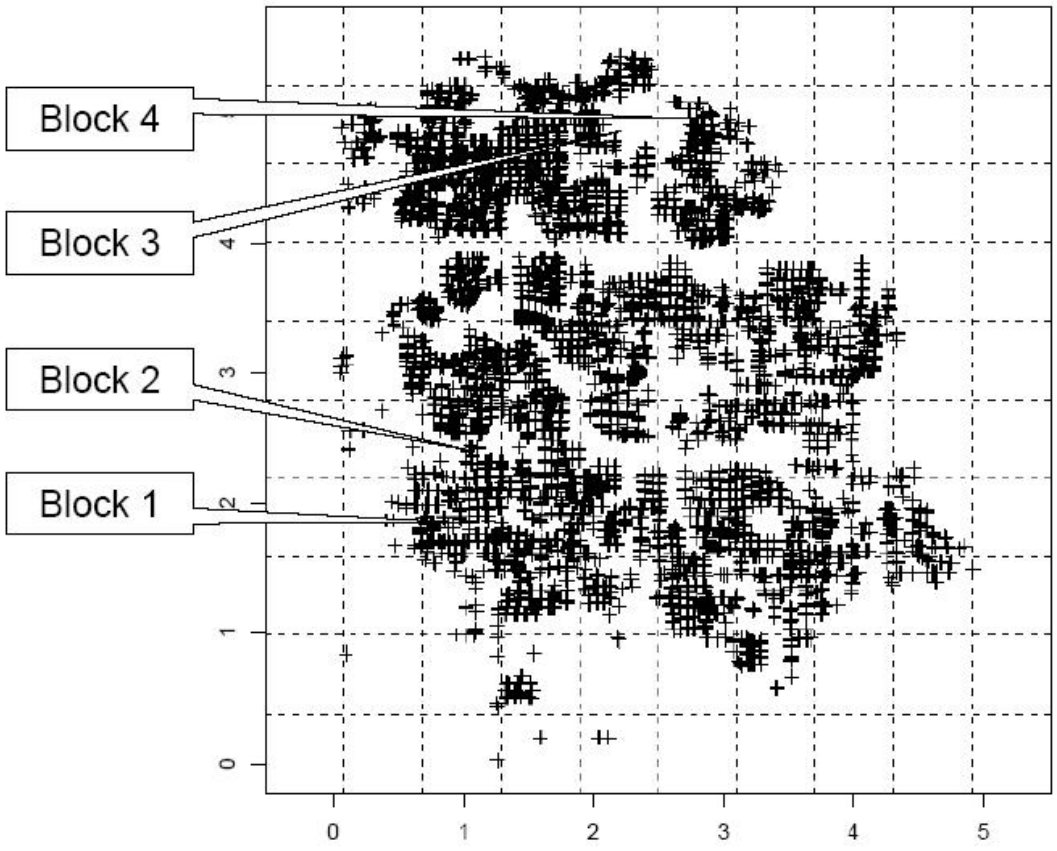

Figure 7: Mean growth curves (solid line) and their corresponding 95% predictive intervals (dotted lines) for the intensity for the four blocks marked in Figure 6.

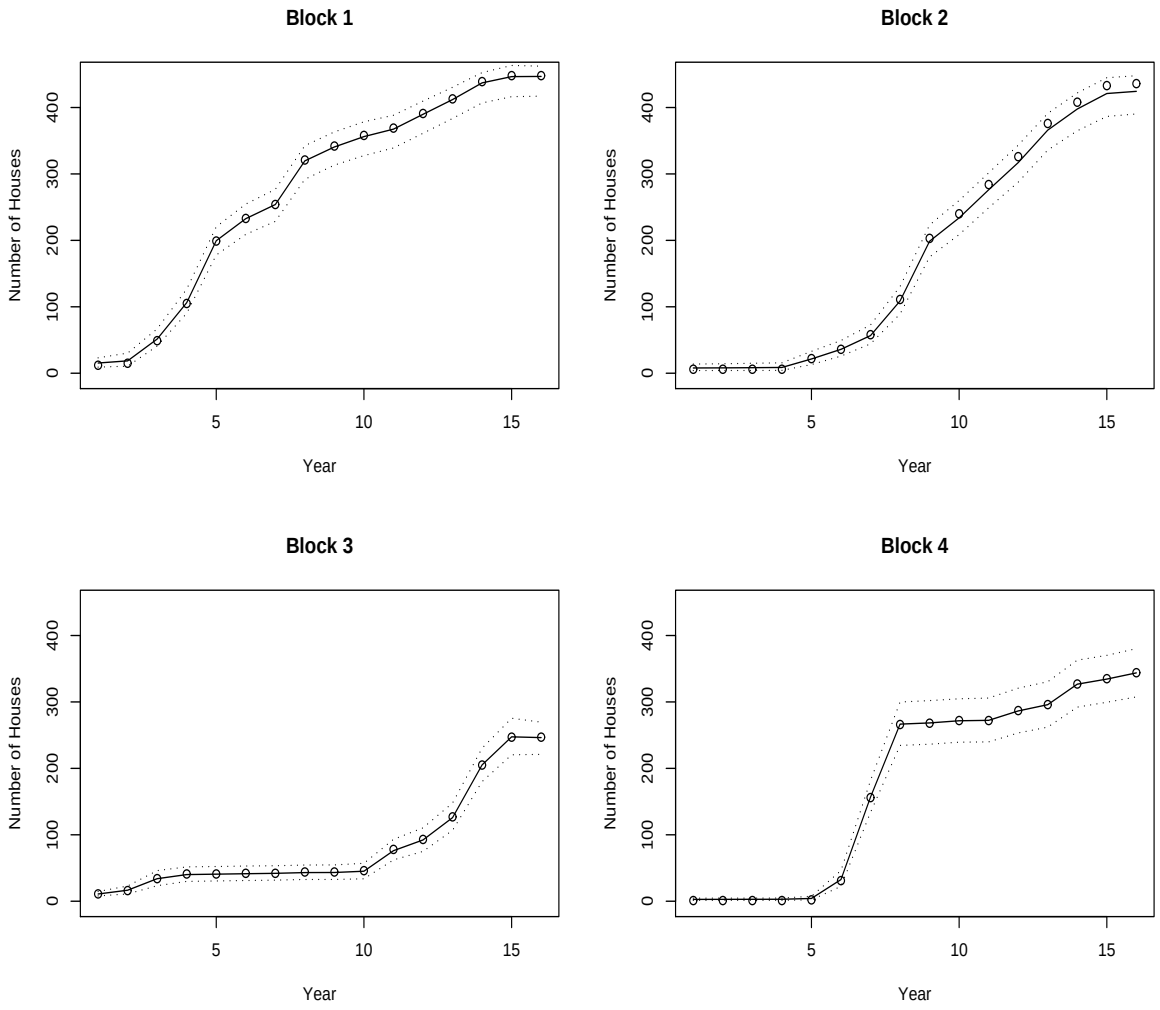

Figure 8: Posterior and predictive mean intensity surfaces for the years 1966, 1967, 1968 and 1969