Bayesian Analysis (2009) **4**, Number 4, pp. 733–758

# Modeling space-time data using stochastic differential equations

Jason A. Duan*, Alan E. Gelfand† and C. F. Sirmans‡

**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 describe a latent process within a hierarchical model for the data. The intent is to view this latent process mechanistically and endow it with appropriate simple features and interpretable parameters. 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.

**Keywords:** geostatistical data, point pattern, hierarchical model, stochastic logistic equation, Markov chain Monte Carlo, urban development

## 1 Introduction

The contribution of this paper is to demonstrate the use and value of stochastic differential equations (SDE) to yield mechanistic and physically interpretable models for space-time data. We consider two common settings: (i) real-valued and point-referenced geostatistical data where observations are taken at non-random locations $s$ and times $t$ and (ii) spatio-temporal point patterns where the locations and times themselves are random. In either case, we assume $s \in D \subset R^2$ where $D$ is a fixed compact region and $t \in (0, T]$ where $T$ is specified.

Examples of spatio-temporal geostatistical data abound in the literature. Examples appropriate to our objectives include ecological process models such as photosynthesis,

*Department of Marketing, McCombs School of Business,University of Texas, Austin, TX, mailto:duanj@mccombs.utexas.edu

†Department of Statistical Science, Duke University, Durham, NC mailto:alan@stat.duke.edu

‡Department of Risk Management, Insurance, Real Estate and Business Law, Florida State University, Tallahassee, FL mailto:cfsirmans@cob.fsu.edu

 DOI:10.1214/09-BA427

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., noisy space-time data are modeled by

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

where $\epsilon(s,t)$ is a space-time noise/error process (contributing the "nugget") and the process of interest is $\Lambda(s,t)$. For us, $\Lambda(s,t)$ is a realization of a space-time stochastic process generated by a stochastic differential equation.

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 denoted again by $\Lambda(s,t)$. Here, the theory of point processes provides convenient tools; the most commonly used and easily interpretable model is the spatio-temporal Poisson process: for any region in the area under study and any specified time interval, the total number of observed points is a Poisson random variable with mean equal to the integrated intensity over that region and time interval.

There is a substantial literature on modeling point-referenced space-time data. The most common approach is the introduction of spatio-temporal random effects described through a Gaussian process with 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 this literature.

In either of the settings (i) and (ii) above, we propose to work with stochastic differential equation models. That is, we intentionally specify $\Lambda(s,t)$ through a stochastic differential equation rather than a spatio-temporal process (see, e.g., Banerjee et al. 2004 and references therein). We specifically introduce a mechanistic modeling scheme where we are directly interested in the parameters that convey physical meanings in the mechanism described by a stochastic differential equation. For example, the prototype of our study is the logistic equation

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

where $K(s)$ is the "carrying capacity" (assuming it is time-invariant) and $r(s,t)$ is the "growth rate". Spatially and/or temporally varying parameters, such as growth rate and carrying capacity, can be modeled by spatio-temporal processes. In practice, the logistic equation finds various 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 recognize the flexibility that comes with a "purely empirical" model such as a realization of a stationary or nonstationary space-time Gaussian process or smoothing splines and that such specifications can be made to fit a given dataset at least as well as a specification limited to a given class of stochastic differential equations. However, for space-time data collected from physical systems it may be preferable to view them as generated by appropriate simple mechanisms with necessary randomness. That is, we particularly seek to incorporate the features of the mechanistic process into the model for the space-time data enabling interpretation of the spatio-temporal parameter processes that is more natural and intuitive. We also demonstrate, through a simulation example, that, when such differential equations generate the data (up to noise), the data can inform about the parameters in these equations and that model performance is preferable to that of a customary model employing a random process realization.

In this regard, we must resort to discretization of time to work with these models, i.e., we actually fit stochastic finite difference models. In other words, the continuous-time specification describes the process we seek to capture but to fit this specification with observed data, we employ first order (Euler) approximation. Evidently, this raises questions regarding the effect of discretization. Theoretical discussion in terms of weak convergence is presented, for instance, in Kloeden and Platen (1992) while practical refinements through latent variables have been proposed in, e.g., Elerian et al. (2001). In any event, Euler approximation is widely applied in the literature and it is beyond our contribution here to explore its impact. Moreover, the results of our simulation example in Section 3.3 reveal that we recover the true parameters of the stochastic PDE quite well under the discretization.

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.

An early motivating problem for our research was the modeling of new home constructions in a fixed region, e.g., within a city boundary. Mathematically and conceptually, the continuous trend that drives the construction of new houses can be captured by the logistic equation, where there is a rate for the growth of the number of houses and a carrying capacity that limits the total number. When new houses can be built at any locations and times, yielding a space-time point pattern, a spatio-temporal point process governed by a version of the stochastic logistic equation becomes an appealing mechanistic model. However, we also acknowledge its limitations, suggesting possible extensions at the end of the paper.

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

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 model the random parameter $r\,(t)$ with a SDE:

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

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

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

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

where $\Delta t$ is the interval between time points and $B\,(t+\Delta t) - B\,(t) \sim N(0, \Delta t)$ if $B(t)$ is a Brownian motion process. 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 Wilkinson (2008) and Stramer and Roberts (2007).

Our prototype is the logistic equation

$$d\Lambda\,(t) = r(t)\Lambda\,(t)\left[1 - \frac{\Lambda\,(t)}{K}\right]dt \text{ and } \Lambda\,(0) = \Lambda_0. \tag{3}$$

To introduce systematic randomness into this model, we specify a mean-reverting Ornstein-Uhlenbeck process for $r\,(t)$:

$$dr\,(t) = -\alpha\,(\mu_r - r\,(t))\,dt + \sigma_\zeta dB\,(t)\,. \tag{4}$$

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

To extend model (3) to a spatio-temporal setting, we can model $\Lambda\,(s,t)$ at every location $s$ with a SDE

$$\frac{\partial\Lambda\,(s,t)}{\partial t} = r(s,t)\Lambda\,(s,t)\left[1 - \frac{\Lambda\,(s,t)}{K\,(s)}\right] \tag{5}$$

[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 $d\Lambda(t) = f(\Lambda(t), t)dt + h\,(\Lambda\,(t)\,, t)\,dB(t)$.

subject to the initial conditions $\Lambda(s,0) = \Lambda_0(s)$. Expression (5) is derived directly from (3) as follows. Assume model (3) for the *aggregate* $\Lambda(D,t) = \int_D \Lambda(s,t)\,ds$:

$$\frac{\partial \Lambda(D,t)}{\partial t} = r(D,t)\Lambda(D,t)\left[1 - \frac{\Lambda(D,t)}{K(D)}\right]. \tag{6}$$

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

The model for $\Lambda(s,t)$ at any location $s$ can be considered as the infinitesimal limit of the model (6) when $D$ is a neighborhood, $\delta_s$ of $s$ whose area goes to zero. Then,

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

Plugging $\delta_s$ into (6) and passing to the limit, we obtain our local model (5).

Model (5) specifies an infinite-dimensional SDE model for the random field $\Lambda(s,t)$, $s \in D$. Similar to (4), we can add randomness to (5) by extending the Ornstein-Uhlenbeck process to the case of infinite dimension,

$$\frac{\partial r(s,t)}{\partial t} = L(\mu_r(s) - r(s,t)) + \frac{\partial B(s,t)}{\partial t}, \tag{7}$$

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

$$L(s) = a(s) + \sum_{l=1}^{2} b_l(s)\frac{\partial}{\partial s_l} - \frac{1}{2}\sum_{l=1}^{2} c_l(s)\frac{\partial^2}{\partial s_l^2} \tag{8}$$

where $a(s)$, $b_l(s)$ and $c_l(s)$ are positive deterministic functions with $s_1$ and $s_2$ the coordinates of location $s$. $B(s,t)$ is a spatially correlated Brownian motion. Here, equation (5) and (7) define a spatio-temporal model with a nonstationary and non-Gaussian $\Lambda(s,t)$ and a latent stationary Gaussian $r(s,t)$. Note that a well-specified $C_B$ will guarantee mean-square continuity and differentiability of $r(s,t)$, $s \in D$. Because the logistic equation is Lipschitz, $\Lambda(s,t)$ will also be mean-square continuous and differentiable.

The simplest Ornstein-Uhlenbeck process model for $r(s,t)$ sets $b_l$ and $c_l$ to zero (see Brix and Diggle 2001); the resulting covariance is separable in space and time. For example, with the Matérn spatial covariance function, we have

$$C_r(s-s',t-t') = \sigma^2 \exp(-a|t-t'|)\,(\phi_\zeta|s-s'|)^\nu\,\kappa_\nu(\phi_\zeta|s-s'|), \tag{9}$$

where $\kappa_\nu(\cdot)$ is the modified Bessel function of the second kind. When $b_l$ and $c_l$ are not equal to zero, $r(s,t)$ defined by equation (7) is a blur-generated process in Brown et al. (2000) with a nonseparable non-explicit spatio-temporal covariance function. Whittle (1963) and Jones and Zhang (1997) propose other stochastic partial differential equation models, which are shown by Brown et al. (2000) to be special examples of the Ornstein-Uhlenbeck process model above.

Returning to the discussion in the Introduction, conceptually, $\Lambda(s,t_j)$ is generated by the continuous-time process defined by (5) and (7). However, the exact solution of this infinite-dimensional SDE and the transition probability of the resulting Markov process for $\Lambda(s,t_j)$ are not generally known in closed-form given the error process $B(s,t)$. Hence, for handling these models, time-discretization is usually required to compute $\Lambda(s,t)$ in simulation and estimation. So, we use Euler approximation to discretize the SDE model for $\Lambda(s,t_j)$. The Euler scheme is broadly employed in simulating SDE's because it the simplest method that is proven to generate a stochastic process that has both strong convergence of order 1/2 and weak convergence of order 1 (see Kloeden and Platen 1992 for theoretical discussions). That is the stochastic *difference* equation resulting from the Euler discretization scheme generates a process that will converge to the process defined by the stochastic differential equation when the length of time steps goes to zero.

# 3 Geostatistical models using SDE

## 3.1 A discretized space-time model with white noise

We assume time is discretized to small, equally spaced intervals of length $\Delta t$, indexed as $t_j$, $j = 0, 1, ..., J$. The data is considered to be $Y(s_i, t_j)$, i.e., an observation at location $s$ and any $t \in (t_j, t_j + \Delta t)$ is labeled as $Y(s,t_j)$. Then, we assume

$$Y(s,t_j) = \Lambda(s,t_j) + \varepsilon(s,t_j),$$

where $\varepsilon(s,t_j)$ models either sampling or measure errors because a researcher cannot directly observe $\Lambda(s,t_j)$.

The dynamics of the discretized $\Lambda(s,t_j)$ is therefore modeled by a difference equation using Euler's approximation applied to (5):

$$\Delta\Lambda(s,t_j) = r(s,t_{j-1})\Lambda(s,t_{j-1})\left[1 - \frac{\Lambda(s,t_{j-1})}{K(s)}\right]\Delta t, \tag{10}$$

$$\Lambda(s,t_j) \approx \Lambda(s,0) + \sum_{l=1}^{j} \Delta\Lambda(s,t_{j-1}). \tag{11}$$

We do not have to discretize the space-time model for $r(s,t)$ if the stationary spatio-temporal Gaussian process $\zeta(s,t)$ allows direct evaluation of its covariance function. For example, the model (7) with constant $L(s) = a_r$ has the closed-form separable

covariance function given in (9) which can be directly used in modeling and can be estimated. Using this form saves one approximation step.

We still need to model the initial $\Lambda(s,0)$ and $K(s)$ if they are not known. For example, because $\Lambda(s,0)$ and $K(s)$ are positive in the logistic equation, we can model them by the log-Gaussian spatial processes with regression forms for the means,

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

Similarly, $\mu_r(s)$ below (7) can be modeled as $\mu_r(X_r(s),\beta_r)$.

Conditioned on $\Lambda(s,t_j)$, the $Y(s,t_j)$ are mutually independent. With data $Y(s_i,t_j)$ at locations $\{s_i, i=1,\ldots,n\}\subset D$, we can provide a hierarchical model based on the evolution of $\Lambda(s,t)$ 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 $\Lambda_j$, $\Delta\Lambda_j$, $r_j$, $K$, $\mu_\Lambda(\beta_\Lambda)$, $\mu_K(\beta_K)$, $\mu_r(\beta_r)$, $\theta_\Lambda$, $\theta_K$ and $\zeta$ 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

$$\begin{aligned}
&y_j|\Lambda_j\sim N\left(\Lambda_j,\sigma_\varepsilon^2 I_n\right),\ \ j=0,\ldots,J,\\
&\Delta\Lambda_j = r_{j-1}\Lambda_{j-1}\left[1-\frac{\Lambda_{j-1}}{K}\right]\Delta t,\\
&\Lambda_j=\Lambda_0+\sum_{l=1}^{j-1}\Delta\Lambda_l,\\
&\log\Lambda_0=\mu_\Lambda(\beta_\Lambda)+\theta_\Lambda,\ \theta_\Lambda\sim N(0,C_\Lambda(s-s';\varphi_\Lambda)),\\
&\log K=\mu_K(\beta_K)+\theta_K,\ \theta_K\sim N(0,C_K(s-s';\varphi_K)),\\
&r=\mu_r(\beta_r)+\zeta,\ \zeta\sim N(0,C_r(s-s',t-t';\varphi_r)),\\
&\beta_\Lambda,\beta_r,\beta_K,\varphi_\Lambda,\varphi_K,\varphi_r\sim \text{priors},
\end{aligned}\tag{12}$$

where $\beta_{(\cdot)}$ are the parameters in the mean surface function; $C_\Lambda$, $C_K$ and $C_r$ [2] are the covariance matrices. In this model, $\Lambda_0$, $r$ and $K$ are latent variables. Note that the $\Lambda_j$'s are deterministic functions of $\Lambda_0$, $r$ and $K$. The joint likelihood for the $J+1$ conditionally independent observations and latent variables is

$$\prod_{j=0}^{J}\left\{N\left(y_j|\Lambda_j(\Lambda_{j-1},r_j,K),\sigma_\varepsilon^2 I_n\right)\right\}N(\log\Lambda_0|\mu_\Lambda,C_\Lambda)\,N(\log K|\mu_K,C_K)\,N(r|\mu_r,C_r),\tag{13}$$

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

[2] 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 $\mu_\Lambda$, $\mu_K$, $\mu_r$, $C_\Lambda$, $C_K$ and $C_r$ when there is no ambiguity.

## 3.2 Bayesian inference and prediction

With regard to inference for the model in (12), there are three latent vectors: $r$, $K$ and $\Lambda_0$. The hyper-parameters in this model include the $\beta_r$, $\beta_K$ and $\beta_\Lambda$ in the parametric trend surfaces, the spatial random effects $\zeta$, $\theta_K$, $\theta_\Lambda$ and the hyper-parameters $\varphi_r$, $\varphi_K$, $\varphi_\Lambda$ in the covariance functions.

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

$$\beta_r, \beta_K, \beta_\Lambda \sim \pi(\beta_r) \cdot \pi(\beta_K) \cdot \pi(\beta_\Lambda); \ \varphi_r, \varphi_K, \varphi_\Lambda \sim \pi(\varphi_r) \cdot \pi(\varphi_K) \cdot \pi(\varphi_\Lambda) \tag{14}$$

where each of $\beta_r$, $\beta_K$, $\beta_\Lambda$, $\varphi_r$, $\varphi_K$, $\varphi_\Lambda$ may represent multiple parameters. For example, we have $\varphi_r = \{\alpha_r, \phi_r, \sigma_r^2, \nu\}$ in the Matérn class covariance function for the separable model (9). Exact specifications of the priors for the $\beta$'s and $\varphi$'s depend on the particular application. For example, if we take $\mu_r(s;\beta_r) = X(s)\beta_r$, we adopt a weak normal prior $N(0, \Sigma_\beta)$ for $\beta_r$. The parameter $\sigma_r^2$ receives the usual Inverse-Gamma prior. Note that $\Delta\Lambda_j$ in the likelihood (13) for the discretized model are deterministic functions of $r$, $K$ and $\Lambda_0$ defined by (10) and (11). Therefore the joint posterior is proportional to

$$\prod_{j=0}^{J} \left\{ N\left(y_j | \Lambda_j(\Lambda_{j-1}, r_j, K), \sigma_\varepsilon^2 I_n\right) \right\} N(\log \Lambda_0 | \mu_\Lambda, C_\Lambda) N(\log K | \mu_K, C_K) N(r | \mu_r, C_r) \cdot$$
$$\pi(\beta_r) \pi(\beta_K) \pi(\beta_\Lambda) \pi(\varphi_r) \pi(\varphi_K) \pi(\varphi_\Lambda). \tag{15}$$

We simulate the posterior distributions of the model parameters and latent variables in (15) using a Markov Chain Monte Carlo algorithm. Because the intensities in the likelihood function are irregular recursive and nonlinear functions of the model parameters and latent variables, it is very difficult to obtain derivatives for an MCMC with directional moves, such as the Langevin method. So, instead we use a 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 $\Lambda_0$, $K$, $r$ and $\beta_r$, $\beta_K$, $\beta_\Lambda$, $\varphi_r$, $\varphi_K$, $\varphi_\Lambda$, 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 (10) and (11) 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 illustrate the fitting of the models in (12) with a simulated data set. In a study region $D$ of 10×10 square units 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 (12) where we fix the carrying capacity to be one at all locations. We may envision that the data simulate the household adoption rates for a certain durable product (e.g., air conditioners, motorcycles) in 44 cities over 30 months. A capacity of one means 100% adoption. Household adoption rates are collected by surveys with measurement errors. The initial condition $\Lambda_0$ is simulated as a log-Gaussian process with a constant mean surface $\mu_\Lambda$ and the Matérn class covariance whose smoothness parameter $\nu$ is set to be 3/2. The spatio-temporal growth rate $r$ is simulated using a constant mean $\mu_r$ and the separable covariance function (9), where the Matérn smoothness parameter $\nu$ 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.

| Model Parameters | True Value | Posterior Mean | 95% Equal-tail Interval |
|---|---|---|---|
| $\mu_\Lambda$ | -4.2 | -4.14 | (-4.88, -3.33) |
| $\sigma_\Lambda$ | 1.0 | 0.91 | (0.62, 1.46) |
| $\phi_\Lambda$ | 0.7 | 0.77 | (0.50, 1.20) |
| $\sigma_\varepsilon$ | 0.05 | 0.049 | (0.047, 0.052) |
| $\mu_r$ | 0.24 | 0.24 | (0.22, 0.26) |
| $\sigma_r$ | 0.08 | 0.088 | (0.077, 0.097) |
| $\phi_r$ | 0.7 | 0.78 | (0.60, 1.10) |
| $\alpha_r$ | 0.6 | 0.64 | (0.51, 0.98) |

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

We use the simulated $r$ and $\Lambda_0$ and the transition equation (10) recursively to obtain $\Delta\Lambda_j$ and $\Lambda_j$ for each of the 30 periods. The observed data $y_j$ are sampled as mutually independent given $\Lambda_j$ with the random noise $\varepsilon_j$. 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 (12) 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: $\pi(\mu_\Lambda) \sim N\left(0, 10^8\right)$ and $\pi\left(\mu_r\right) \sim N\left(0, 10^8\right)$. We use natural conjugate priors for the precision parameters (inverse of variances) of $r$ and $\Lambda_0$: $\pi\left(1/\sigma_r^2\right) \sim Gamma(1,1)$ and $\pi\left(1/\sigma_\Lambda^2\right) \sim Gamma(1,1)$. The positive parameter for the temporal correlation of $r$ also has a vague log-normal prior: $\pi\left(\alpha_r\right) \sim log\text{-}N\left(0, 10^8\right)$. Because the spatial range parameters $\phi_r$ and $\phi_\Lambda$ 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 $\phi_r$ and $\phi_\Lambda$.

We use the random-walk Metropolis-Hastings algorithm to simulate posterior samples of $r$ and $\Lambda_0$. We draw the entire vector of $\Lambda_0$ for all forty locations as a single block in every iteration. Because $r$ is very high-dimensional ($r$ being a $40\times30$ 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 $\Lambda_0$, the mean parameters $\mu_r$, $\mu_\Lambda$ 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. $\phi_r$ and $\phi_\Lambda$ have discrete priors and therefore discrete Gibbs samplers too. We also use the random-walk Metropolis-Hastings algorithm to draw $\alpha_r$.

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 means and 95% equal-tail Bayesian posterior predictive intervals 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 intervals 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 intervals, 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 intervals is 95.8%.

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

$$y_j = \mu + \xi_j + \varepsilon_j;\ \varepsilon_j \sim N\left(0, \sigma_\varepsilon^2 I_n\right),\ \ j = 0, \ldots, J \tag{16}$$

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

$$C_\xi(t-t', s-s') = \sigma_\xi^2 \exp\left(-\alpha_\xi |t-t'|\right) \left(\phi_\xi |s-s'|\right)^\nu \kappa_\nu \left(\phi_\xi |s-s'|\right),\ \nu = \frac{3}{2}\,. \tag{17}$$

Comparison of model performance between our model in (12) and the model in (16) is conducted using spatial prediction at the 4 new locations in Figure 1. The computational cost of the model in (16) 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 intervals, together with the hold-out data. For the four hold-out sites, the average mean square error of the model (12) is $1.75\times10^{-3}$ versus $3.34\times10^{-3}$ of the model (16); the average length of the 95% predictive intervals for the model (12) is 0.29 versus 0.72 for the model (16). It is evident that the prediction results under the benchmark model are substantially worse than those under our model (12); the mean growth curves are less accurate and less smooth, and the 95% predictive intervals are 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 $\Omega(s,t), s \in D, t \in [0,T]$. 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 $[0,T]$. 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. The integrated process $\Lambda(s,T) = \int_0^T \Omega(s,t)\,dt$, provided that $\Omega(s,t)$ 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

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

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

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

the use of the approximated intensity

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

As a concrete example, we 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 $\Lambda(s,t)$ follows the logistic equation model (5). We can also apply the discretized version (10) to $\Delta\Lambda_j(s)$.

Let our initial point pattern be $x_0$ and the intensity be $\Lambda_0(s) = \int_{-\infty}^{0} \Omega(s,\tau)\,d\tau$. The hierarchical model for the space-time point patterns is merely the model (12) with the first stage of the hierarchy replaced by the following

$$\begin{aligned} x_j|\Delta\Lambda_j &\sim \text{Poisson Process}(D,\Delta\Lambda_j), \quad j=1,\ldots,J \\ x_0|\Lambda_0 &\sim \text{Poisson Process}(D,\Lambda_0), \end{aligned} \tag{18}$$

where we suppress the indices $t$ and $s$ again for the periods $t_1,\ldots,t_J$. Note that, unlike in (12), the intensity $\Delta\Lambda_j$ for $x_j$ must be positive. Therefore, we model the log growth rate, that is

$$\log r(s,t) = \mu_r(s;\beta_r) + \zeta(s,t), \ \zeta \sim \text{GP}(0, C_\zeta(s-s',t-t';\varphi_r)). \tag{19}$$

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

$$\prod_{j=1}^{J}\left\{\exp\left(-\int_D \Delta\Lambda_j(s)\,ds\right)\prod_{i=1}^{n_j}\Delta\Lambda_j(x_{ji})\right\}\cdot\exp\left(-\int_D \Lambda_0(s)\,ds\right)\prod_{i=1}^{n_0}\Lambda_0(x_{0i}). \tag{20}$$

This likelihood is more difficult to work with than that in (13). There is a stochastic integral in (20), $\int_D \Delta\Lambda_j(s)\,ds$, 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

$$\begin{aligned} &\prod_{j=1}^{J}\left[\exp\left(-\sum_{m=1}^{M}\Delta\Lambda_j(m)A(m)\right)\prod_{m=1}^{M}\Delta\Lambda_j(m)^{n_{jm}}\right] \\ &\cdot\exp\left(-\sum_{m=1}^{M}\Lambda_0(m)A(m)\right)\prod_{m=1}^{M}\Lambda_0(m)^{n_{0m}} \end{aligned} \tag{21}$$

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

## 4.2 Modeling house construction data for Irving, TX

Our 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 $\Lambda_0, K$, 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 $C_\Lambda$ and $C_K$ are constructed using the Matérn class covariance function with distances between the centroids of the cells. The smoothness parameter $\nu$ is set to be 3/2. The variances $\sigma^2_\Lambda$, $\sigma^2_K$ and range parameters $\phi_\Lambda$ and $\phi_K$ 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 $\Sigma_s$ is also constructed as a Matérn class function of the distances between cell centroids with smoothness parameter $\nu$ being set to 3/2. The temporal correlation $\Sigma_t$ is of exponential form as in (9). The variance $\sigma^2_r$, spatial and temporal correlation parameters $\phi_r$ and $\alpha_r$ 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 $\Lambda_0$: $\pi(1/\sigma^2_\Lambda)$, $\pi(1/\sigma^2_K)$, $\pi(1/\sigma^2_r) \overset{ind}{\sim} Gamma(1,1)$. The temporal correlation parameter of $r$ also has a vague log-normal prior: $\pi(\alpha_r) \sim log\text{-}N(0, 10^8)$. Again, the spatial range parameters $\phi_\Lambda$ $\phi_K$ and $\phi_r$ 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 $\phi_\Lambda$ $\phi_K$ and $\phi_r$.

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 means and 95%

equal-tail posterior intervals for the model parameters are presented in Table 2.

| Model Parameters | Posterior Mean | 95% Equal-tail Interval |
|---|---|---|
| $\mu_\Lambda$ | 2.78 | (2.15, 3.40) |
| $\sigma_\Lambda$ | 1.77 | (1.49, 2.11) |
| $\phi_\Lambda$ | 3.03 | (2.70, 3.20) |
| $\mu_r$ | -2.76 | (-3.24, -2.29) |
| $\sigma_r$ | 2.48 | (2.32, 2.68) |
| $\phi_r$ | 4.09 | (3.70, 4.30) |
| $\alpha_r$ | 0.52 | (0.43, 0.62) |
| $\mu_K$ | 6.49 | (5.93, 7.01) |
| $\sigma_K$ | 1.17 | (1.02, 1.44) |
| $\phi_K$ | 1.91 | (1.60, 2.20) |

**Table 2:** Posterior inference for the house construction data

Figure 7 shows the posterior mean growth curves and 95% Bayesian predictive intervals 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.[3]

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 equations to model both geostatistical and point pattern space-time data. Our examples demonstrate that the proposed hierarchical modeling can accommodate the complicated model structure and achieve good estimation and prediction. The major challenges in fitting our proposed models are: (i) the evaluation of a likelihood that involves discretization of the SDE in time and stochastic integrals and (ii) a likelihood that does not allow an easy formulation of an efficient Metropolis-Hastings algorithm. In dealing with the first challenge, we utilize the Euler approximation and the space discretization method in Benes et al. (2002). Though the simulation results are encouraging, further investigation of these approximations or

[3]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 very accurate as well.

alternatives would be of interest. For the second, we apply the random-walk Metropolis algorithm to the posterior simulation, which is liable to create large auto-correlation in the sampling chain. The nonlinear and recursive structure of our likelihood makes most of the current Metropolis methods inapplicable, encouraging future research for a more efficient Metropolis-Hastings algorithm for this class of problems.

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

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC. 734

Benes, V., Bodlak, K., Møller, J., and Waagepetersen, R. (2002). "Bayesian Analysis of Log Gaussian Cox Processes for Disease Mapping." Technical Report R-02-2001, Department of Mathematical Sciences, Aalborg University. 746

Brix, A. and Diggle, P. (2001). "Spatiotemporal prediction for log-Gaussian Cox processes." *Journal of the Royal Statistical Society: Series B*, 63: 823–841. 734, 737, 743

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. 734

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. 734, 738

Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley, 2nd edition. 734

Daley, D. and Vere-Jones, D. (1988). *Introduction to the Theory of Point Processes*. New York: Springer Verlag. 734

Diggle, P. (2005). "Spatio-temporal point processes: methods and applications." Dept. of Biostatistics Working Paper 78, Johns Hopkins University. 734

Elerian, O., Chib, S., and Shephard, N. (2001). "Likelihood inference for discretely observed non-linear diffusions." *Econometrica*, 69: 959–993. 735, 736

Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005). "Spatial process modelling for univariate and multivariate dynamic spatial data." *Environmetrics*, 16: 465–479. 734

Gneiting, T. (2002). "Nonseparable, stationary covariance functions for space-time data." *Journal of the American Statistical Association*, 97: 590–600. 734

Golightly, A. and Wilkinson, D. (2008). "Bayesian inference for nonlinear multivariate diffusion models observed with error." *Computational Statistics and Data Analysis*, 52: 1674–1693. 736

Jones, R. and Zhang, Y. (1997). "Models for continuous stationary space-time processes." In Gregoire, G., Brillinger, D., Diggle, P., Russek-Cohen, E., Warren, W., and Wolfinge, R. (eds.), *Modelling Longitudinal and Spatially Correlated Data*, 289–298. Springer, New York. 738

Karr, A. (1991). *Point Processes and Their Statistical Inference*. New York: Marcel Dekker, 2nd edition. 734

Kloeden, P. and Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Springer. 735, 738

Kot, M. (2001). *Elements of Mathematical Ecology*. Cambridge Press. 735

Mahajan, V. and Wind, Y. (1986). *Innovation Diffusion Models of New Product Acceptance*. Harper Business. 735

Møller, J., Syversveen, A., and Waagepetersen, R. (1998). "Log Gaussian Cox processes." *Scandanavian Journal of Statistics*, 25: 451–482. 734

Møller, J. and Waagepetersen, R. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC Press. 734

Ogata, Y. (1998). "Space-time point-process models for earthquake occurrences." *Annals of the Institute for Statistical Mathematics*, 50: 379–402. 734

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. Chapman & Hall/CRC. 734

Stein, M. L. (2005). "Space-time covariance functions." *Journal of the American Statistical Association*, 100: 310–321. 734

Stramer, O. and Roberts, G. (2007). "On Bayesian analysis of nonlinear continuous-time autoregression models." *Journal of Time Series Analysis*, 28: 744–762. 736

Whittle, P. (1963). "Stochastic processes in several dimensions." *Bulletin of the International Statistical Institute*, 40: 974–994. 738

Zhang, H. (2004). "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics." *Journal of the American Statistical Association*, 99: 250–261. 741, 745

**Acknowledgments**

The authors thank Thomas Thibodeau for providing the Dallas house construction data.

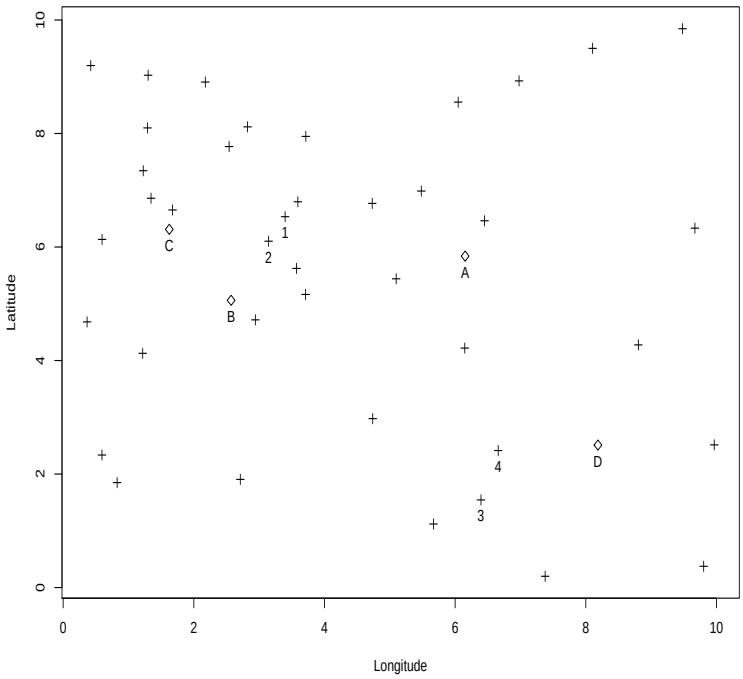


**Figure 1:** Locations for the simulated data example in Section 3.3.

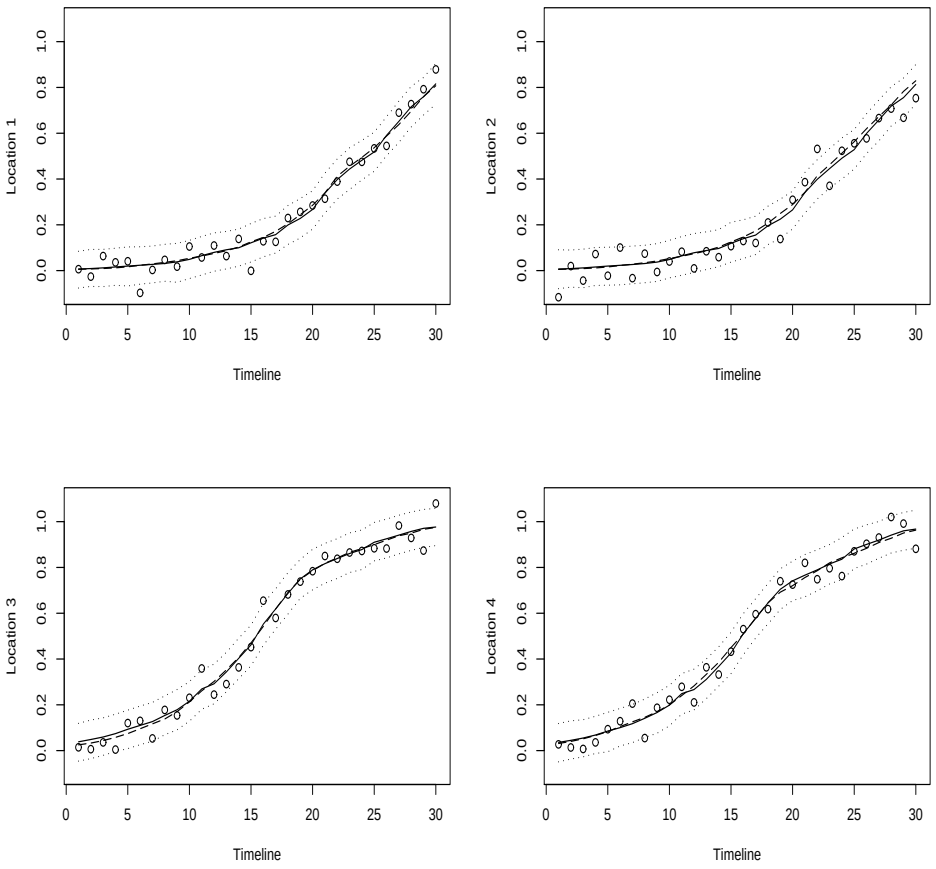


**Figure 2:** Observed space-time geostatistical data at 4 locations, actual (dashed line) and fitted mean growth curves (solid line), and 95% predictive intervals (dotted line) by our model (12) for the simulated data example.

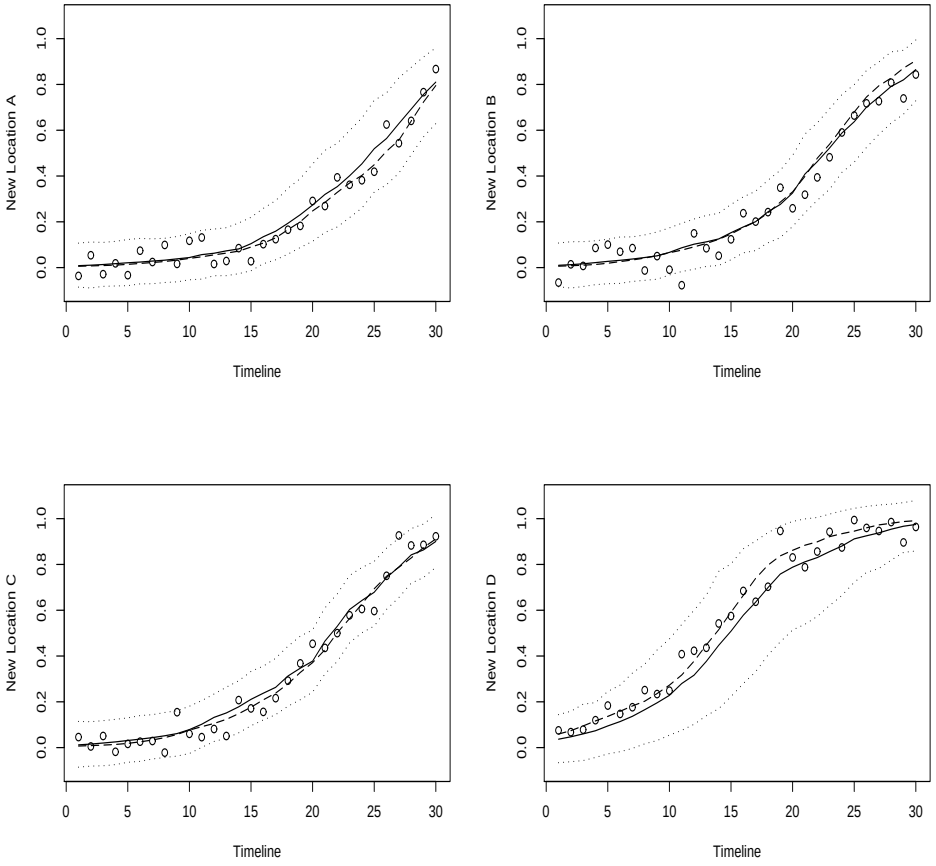


**Figure 3:** Hold-out space-time geostatistical data at 4 locations, actual (dashed line) and predicted mean growth curves (solid line) and 95% predictive intervals (dotted line) by our model (12) for the simulated data example.

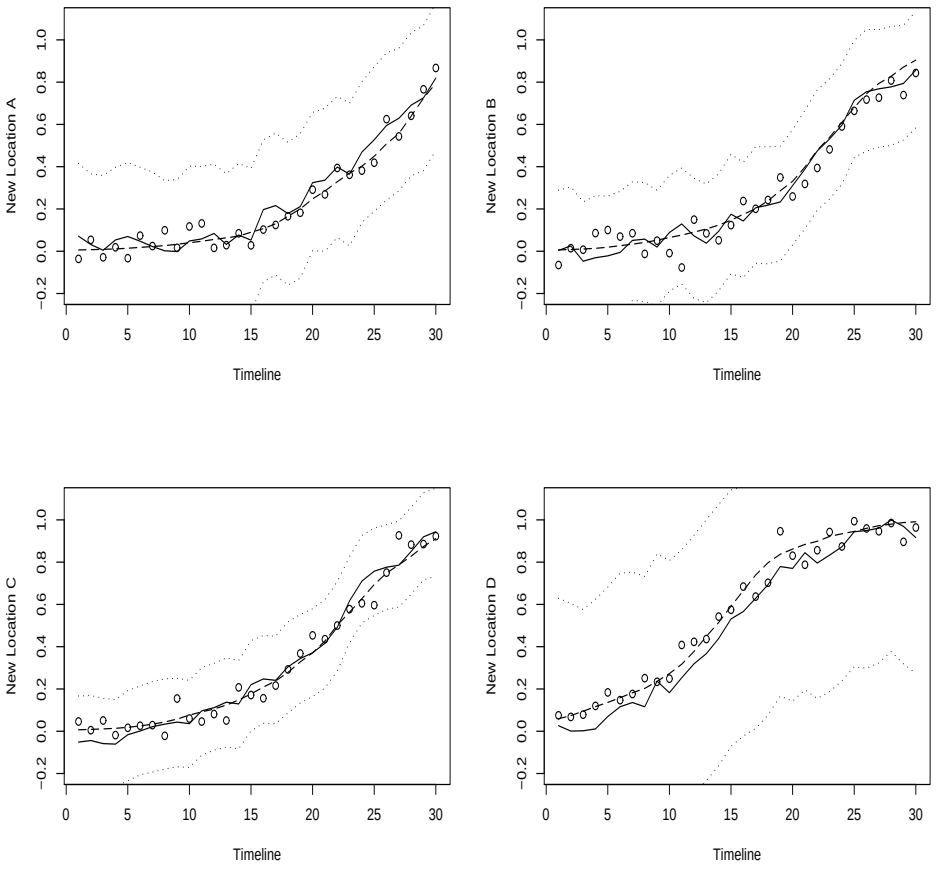


**Figure 4:** Hold-out space-time geostatistical data at 4 locations, actual (dashed line) and predicted mean growth curves (solid line) and 95% predictive intervals (dotted line) by the benchmark model (16) for the simulated data example.

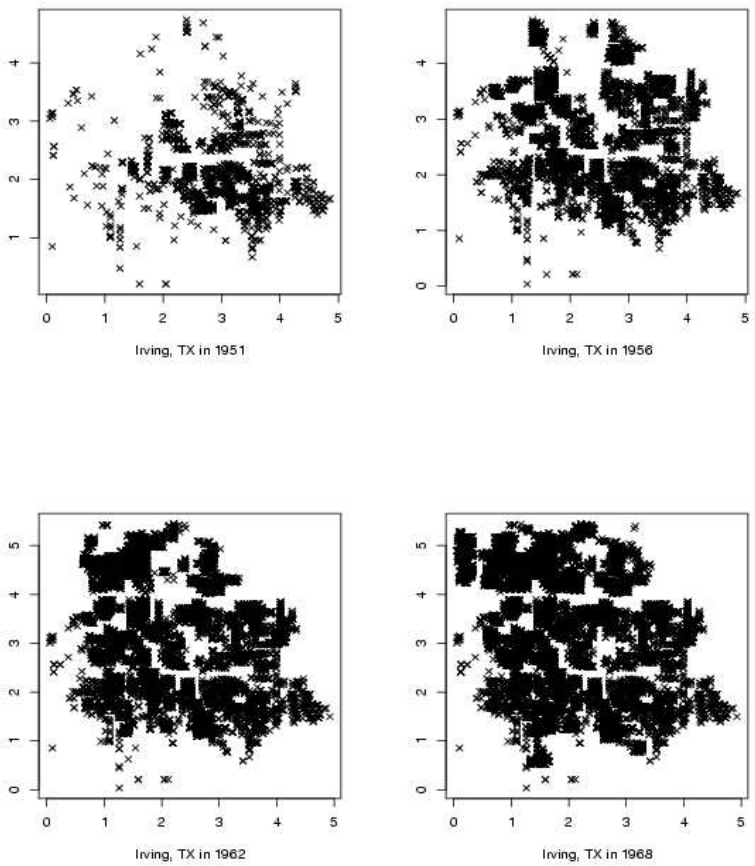


**Figure 5:** Growth of residential houses in Irving, TX.

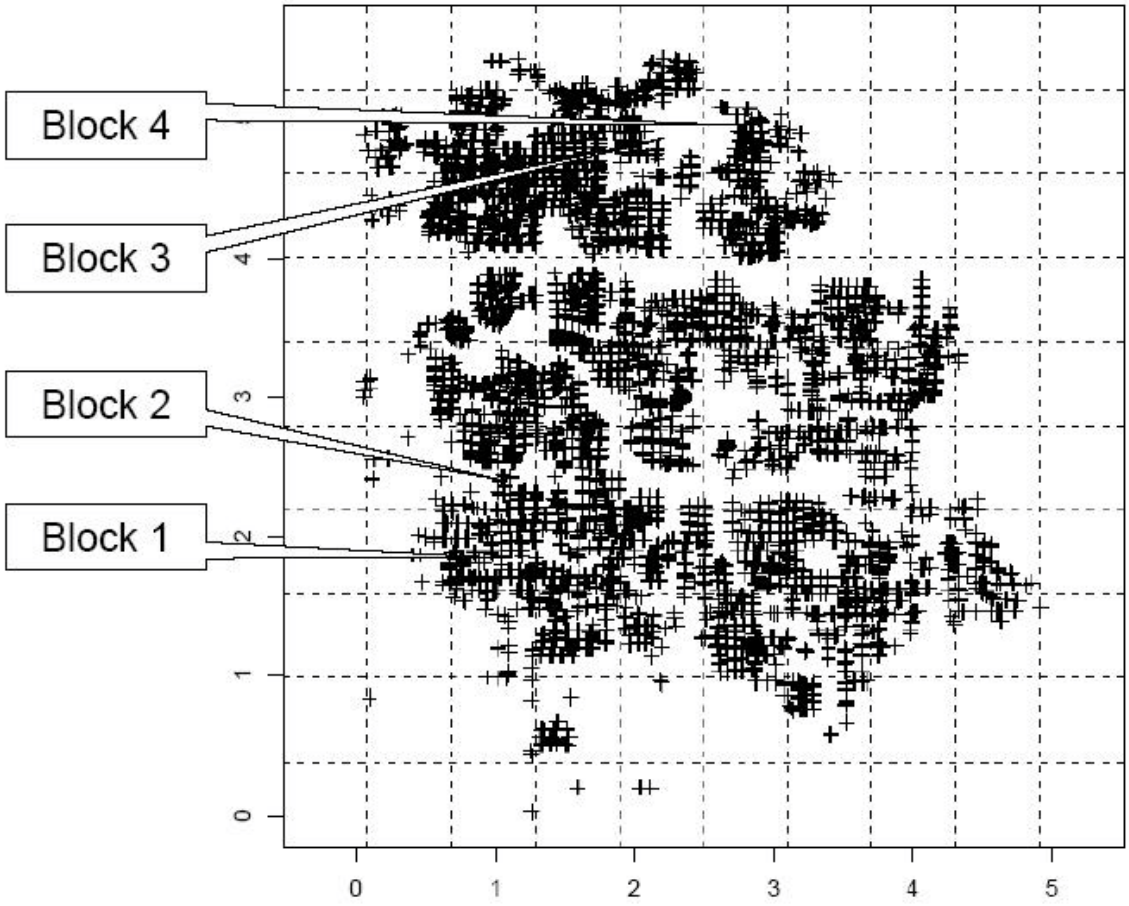


**Figure 6:** The gridded study region encompassing Irving, TX.

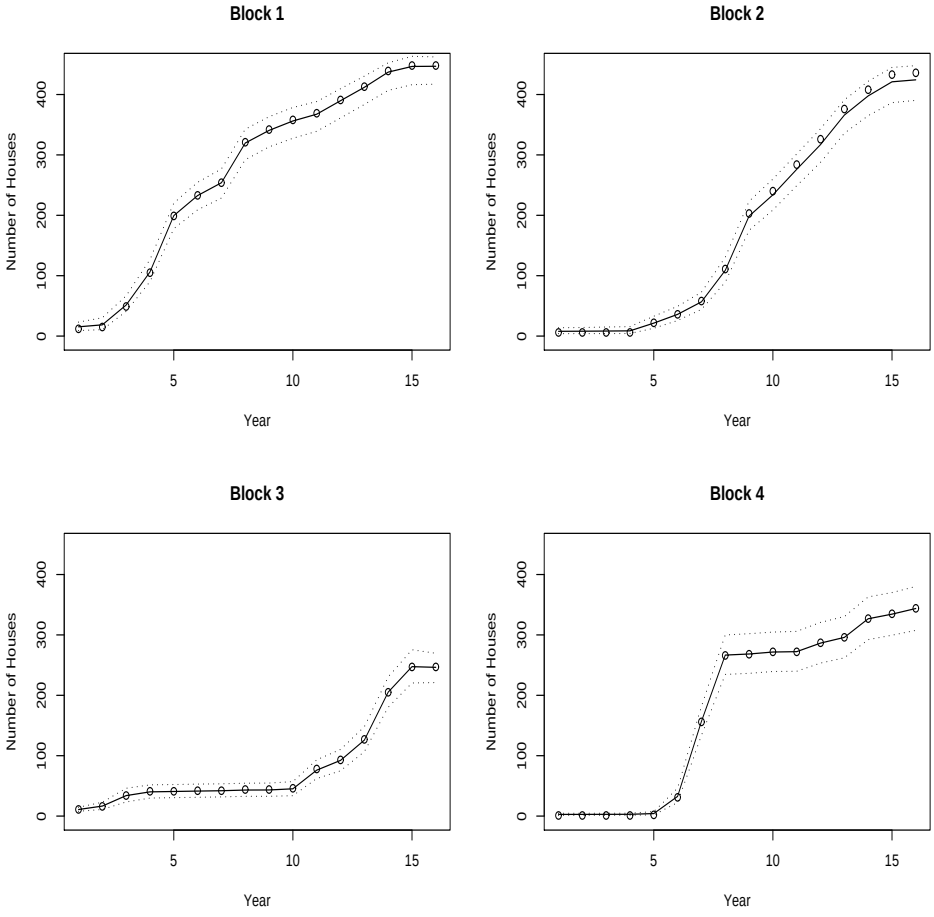


**Figure 7:** Mean growth curves (solid line) and their corresponding 95% predictive intervals (dotted lines) for the intensity for the four blocks marked in Figure 6.

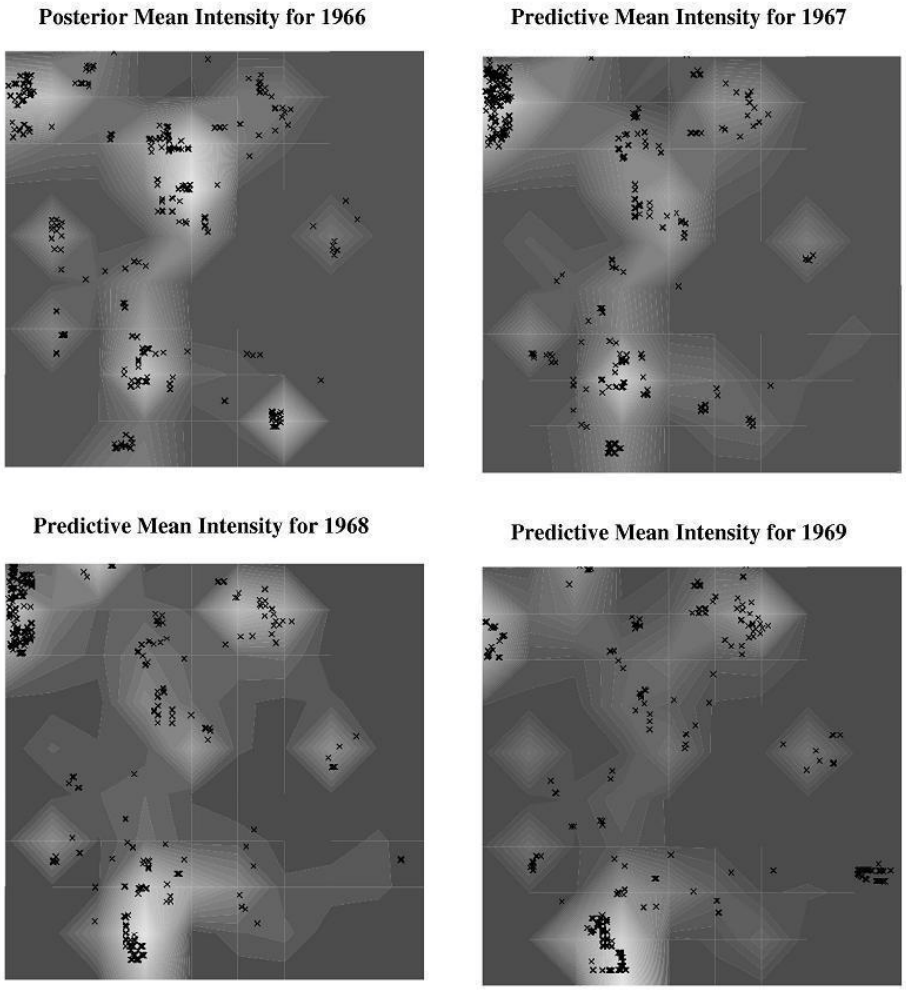


**Figure 8:** Posterior and predictive mean intensity surfaces for the years 1966, 1967, 1968 and 1969