EPA Public Access

Author manuscript

Stoch Environ Res Risk Assess. Author manuscript; available in PMC 2019 February 01.

About author manuscripts

Submit a manuscript

Published in final edited form as:

Stoch Environ Res Risk Assess. 2018 February ; 32(2): 401–417. doi:10.1007/s00477-017-1417-9.

Alternating Gaussian Process Modulated Renewal Processes for Modeling Threshold Exceedances and Durations

Erin M. Schliep,

University of Missouri 146 Middlebush Columbia, MO 65211, Tel.: 573-882-4455,
schliepe@missouri.edu

Alan E. Gelfand, and

Duke University Durham, NC 27708-0251, alan@stat.duke.edu

David M. Holland

U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC 27711, holland.david@epa.gov

Abstract

It is often of interest to model the incidence and duration of threshold exceedance events for an environmental variable over a set of monitoring locations. Such data arrive over continuous time and can be considered as observations of a two-state process yielding, sequentially, a length of time in the *below threshold* state followed by a length of time in the *above threshold* state, then returning to the below threshold state, etc.

We have a two-state continuous time Markov process, often referred to as an alternating renewal process. The process is observed over a truncated time window and, within this window, time in each state is modeled using a distinct cumulative intensity specification. Initially, we model each intensity over the window using a parametric regression specification. We extend the regression specification adding temporal random effects to enrich the model, using a realization of a log Gaussian process over time. With only one type of renewal, this specification is referred to as a Gaussian process modulated renewal process. Here, we introduce Gaussian process modulation to the intensity for each state.

Model fitting is done within a Bayesian framework. We clarify that fitting with a customary log Gaussian process specification over a lengthy time window is computationally infeasible. The nearest neighbor Gaussian process (NNGP), which supplies sparse covariance structure, is adopted to enable tractable computation. We also propose methods for both generating data under our models and for conducting model comparison. The model is applied to hourly ozone data for four monitoring sites in different locations across the United States for the ozone season of 2014. For each site, we obtain estimated profiles of up-crossing and down-crossing intensity functions through time. In addition, we obtain inference regarding the number of exceedances, the

Disclaimer

The U.S. Environmental Protection Agency through its Office of Research and Development partially collaborated in this research. Although it has been reviewed by the Agency and approved for publication, it does not necessarily reflect the Agency's policies or views.

distribution of the duration of exceedance events, and the proportion of time in the above and below threshold state for any time interval.

Keywords

cumulative risk; hazard; log Gaussian process; Markov chain Monte Carlo; nearest neighbor Gaussian process; representative points; stochastic integration

1. Introduction

Threshold exceedance events for some environmental processes can be infrequent yet have major adverse impacts on humans. For example, high levels of ozone concentration or noise pollution in an urban area may occur only a few times a year yet pose dangerous health risks for sensitive populations (Bell et al., 2004; Bell and Dominici, 2008; U.S. Environmental Protection Agency, 2011, 2013). Accurate inference regarding both the number and duration of exceedance events would provide critically important information in the evaluation of these risks across time and spatial locations.

Exceedance events arise from and are important features of many environmental processes, e.g., temperature, rainfall, and air quality. Suppose the data come from a monitoring station collected over a specified time period. With a spatially referenced set of stations, a direct spatio-temporal approach builds a geostatistical model for the contaminant levels over space and time (e.g., Huerta et al., 2004; Gryparis et al., 2007; Sahu et al., 2007). The geostatistical model enables a predictive distribution for the level of the contaminant at any point in space and time. From the predictive distribution, we can attempt to tease out inference for exceedances by, say, obtaining the probability of exceeding a specified threshold at a given location and time. Here, however, we wish to expand our understanding of exceedance events in terms of quantifying their incidence and duration during a time period of interest, thus, learning more about patterns of exceedances not available from geostatistical models.

A non-homogeneous Poisson process (NHPP) (Cressie, 1993; Waller and Gotway, 2004; Diggle, 2013) provides an approach for explaining what is viewed as a point pattern of exceedance events. The intensity of the process (rate of occurrence) is modeled as a function that varies, say over time and/or space. When the intensity is specified through a regression model with spatially and/or temporally referenced explanatory variables, this process is referred to as a *modulated* Poisson process (Cox, 1972). NHPPs have been used for modeling exceedance events of air quality standards (Achcar et al., 2008, 2011; Rodrigues et al., 2014), drought indices (Achcar et al., 2016), noise thresholds (Guarnaccia et al., 2015), and ground-motion intensities of earthquakes (Lervolino et al., 2014). Various rate functions have been investigated for capturing the spatially and/or temporally varying intensities, such as the Weibull and Goel-Okumoto rate functions. Additionally, NHPPs in the presence of change-points can also be considered (e.g. Achcar et al., 2011, 2016).

The shortcoming of using NHPP models for exceedance events is that they fail to take into account the duration of each event. Threshold exceedance events occur for some positive

duration of time and these durations may be useful when assessing the impact of the events. Moreover, when generating exceedance events from an NHPP, multiple events can be simulated with a short window of time. It is unclear whether these events have the same interpretation as a single event occurring at the time the threshold was first exceeded and lasting for some duration.

Our contribution is to develop a novel stochastic process model to capture the physical processes of threshold exceedances and to generate exceedance events that are consistent with how they occur and are observed. The modeling framework encompasses the information about when exceedance events occur and their durations recognizing that, together, this is a two state process yielding, sequentially, a length of time in the below threshold state followed by a length of time in the above threshold state, then returning to the below threshold state, etc. Such a process is a two-state continuous time Markov process, often referred to as an alternating renewal process (Norris, 1998; Trivedi, 2001). The classic example of a continuous time alternating two-state process arises in reliability analysis where a system is in the operational state (below threshold) until a *failure* (an exceedance event) occurs. Then it is in repair state (above threshold) until it is *repaired* and returns to the below threshold state. An alternative view of the process is in terms of survival analysis. Since the events arrive sequentially, event times can be viewed as alternating survival times for two different survival processes.

The process is observed over a truncated time window and, within this window, time in each state is modeled using a distinct cumulative intensity/hazard specification. Initially, we model the intensity over the window using a parametric regression specification defining a *modulated* renewal process (Cox, 1972; Berman, 1981). We extend the regression specification in a nonparametric fashion adding temporal random effects through a realization of a log Gaussian process over time. When the nonparametric modulation is introduced through a Gaussian process, it has been referred to as a Gaussian process modulated renewal process. See, e.g., Rao and Teh (2011); Lasko (2014). The novelty provided here is the introduction of Gaussian process modulation to the intensity/hazard for each state.

Alternating renewal process modeling enables much richer inference for exceedance events than is possible with the foregoing NHPP models. We can explain the duration of exceedance events, assess the probability of being in a particular state at a particular time, and predict incidence of exceedance events (with associated uncertainty) over specified time periods. Fitting is done within a Bayesian framework, using Markov chain Monte Carlo, yielding full posterior inference. We fit both the alternating regression modulated renewal process model and the alternating Gaussian process modulated renewal process model and find that the latter provides much better model fit and better predictive inference for our exceedances application. However, the Gaussian process modulated renewal process is much more demanding to fit due to the stochastic integrations it introduces. We employ a nearest neighbor Gaussian process model (Datta et al., 2016), which provides convenient computational sparsity, for each of the Gaussian process modulations.

For a given threshold and location, over the interval $(0, T]$, the data will be of the form $\{\mathcal{X}, \mathcal{D}, \mathbf{W}\}$. Here, $\mathcal{X} = \{X_1, \dots, X_N\}$ is a set of N event times denoting the end of a period in the below threshold state or the beginning of an above threshold state. After an X_i , we move to the above threshold state and stay in that state for duration D_i . Let $\mathcal{D}$ denote the vector of the durations for each of the exceedances. Therefore, each $Y_i \equiv X_i + D_i$ denotes the end of a period in the above threshold state or the beginning of a below threshold state. At time T , we are either in the below threshold state meaning N renewals of the (below, above) pair have been completed, or in the above threshold state meaning we have completed N below threshold renewals and $N - 1$ above threshold renewals. Lastly, $\mathbf{W}$ denotes time-specific covariates.

In the spirit of an alternating renewal process, we can write the data as an alternating sequence, $0 < X_1 < Y_1 < X_2 < Y_2 < \dots < T$. Renewal processes are discussed in the literature primarily from a probabilistic perspective, investigating long term behavior, without introducing truncation of time to a bounded interval. For example, working with R^+ , using a specified intensity/risk function, the processes are used to investigate expected length of a renewal, expected proportion of time in each state, or the probability of being in state j at time t . See, e.g., Beichelt (2006) and Birolini (2012) and references therein. In this work, we truncate to a finite window of observation and view the challenge as an inference problem, specifying models for the incidence of X s and Y s. Note that the sequence $X_i - Y_{i-1}$, $i = 1, 2, \dots$ (where $Y_0 = 0$) is the sequence of durations of the below threshold state while $Y_i - X_i$, $i = 1, 2, \dots$ is the sequence of durations of the above threshold state. The differences, $Y_i - Y_{i-1}$, $i = 1, 2, \dots$ denote the sequence of durations of an alternating (below, above) pair.

Two noteworthy remarks are as follows. First, we can imagine $\mathcal{X}$ and $\mathcal{Y}$ as realizations of a point pattern model but they can not be modeled using standard NHPP specifications. The imposed alternation of the points along with the fact that the occurrence of N X s in $(0, T]$ implies either N or $N - 1$ Y s precludes an NHPP specification. Additionally, we can not generate the realization by selecting $N = n$ at random and then locating the points, with conditional independence, over $(0, T]$. This alternating process mandates specifying the model sequentially. That is, assuming at time 0 we are waiting for an exceedance event to occur, we model $[X_1 | \mathbf{W}]$. Then, once the first exceedance event occurs, we model the ending time of the exceedance event conditionally as $[Y_1 | X_1, \mathbf{W}]$. Continuing in this fashion over the time interval provides the formal probabilistic specification. This leads to the second remark. We assume renewal times are independent. However, with a time dependent intensity which, as above, is modeled either as a parametric regression in time or as a realization of a log Gaussian process, the renewal times of the $\mathcal{X}$ and $\mathcal{Y}$ are not i.i.d. That is, the clock does not reset after each event time.

It is important to mention that while the term “exceedance” in the literature may suggest *extremes*, here, exceedance events are not being considered as extreme events. Extremes pose a different problem; extreme value theory (Coles et al., 2001) and the peaks over threshold approach (Leadbetter, 1991) assume that the observations are already in the tail of the distribution and, therefore, interest is in the probability of exceeding some level v given that u , where $u < v$, is already exceeded. Here, the threshold of interest need not be an

extreme. For example, it could be an environmental standard for the variable and therefore, exceedance events need not be in the tail of the distribution.

We analyze hourly monitoring station air quality data from U.S. Environmental Protection Agency's (EPA) Air Quality System at four monitoring stations across the United States for the 2014 ozone season which spans from May 1st to September 30th. The data can be obtained through EPA's AirData portal at <https://www3.epa.gov/airquality/airdata/>.

The plan of this paper is as follows. In Section 2 we offer a framework for modeling exceedance events and durations using alternating renewal process, adopting parametric and stochastic process realizations for the intensities. We also make the connection to survival time models. Methods for implementing model-based inference, conducting model selection, and simulating exceedance events and durations under the model are given in Section 3. In Section 4 we give details on the threshold exceedance and duration data for ozone at monitoring stations across the United States. In Section 5, we apply the model first to simulated exceedance and duration data as proof of concept and then to data collected across four urban monitoring stations in the United States. Section 6 offers a discussion and directions for future work.

2. Modeling framework for exceedance events and durations

We begin with a reminder of the basic renewal theory model in the context of an intensity $\lambda(t)$ over a bounded interval, $(0, T]$. On this interval, we have the sequence of observed events $\mathbf{x} = \{x_1, \dots, x_n\}$ such that $0 < x_1 < \dots < x_n \leq T$. We look at the events in $\mathbf{x}$ arriving sequentially and construct the likelihood as

$$\begin{aligned} L(\lambda; \mathbf{x}) = & P(\text{no event in } (0, x_1)) \\ & \times P(\text{event in } (x_1, x_1 + \Delta x_1) \mid \text{no event in } (0, x_1)) \\ & \times P(\text{no event in } (x_1 + \Delta x_1, x_2) \mid \text{event in } (x_1, x_1 + \Delta x_1)) \\ & \times P(\text{event in } (x_2, x_2 + \Delta x_2) \mid \text{no event in } (x_1 + \Delta x_1, x_2)) \\ & \dots \\ & \times P(\text{no event in } (x_n + \Delta x_n, T) \mid \text{event in } (x_n, x_n + \Delta x_n)) \end{aligned} \quad (1)$$

where each Δx_i creates a small interval of time.

We assume renewal times are independent and define an intensity/hazard function $\lambda(x)$ with cumulative hazard $\Lambda(x) = \int_0^x \lambda(u) du$. If $\Lambda(x) \sim \text{Exp}(1)$ ¹, then $1 - F_X(x) = P(X > x) = e^{-\Lambda(x)}$.

That is, $P(\text{no event in } (0, x_1)) = e^{-\int_0^{x_1} \lambda(u) du}$. Differentiating, $f_X(x) = e^{-\Lambda(x)} \lambda(x)$. Therefore,

¹Following Berman (1981) and Lasko (2014), we could adopt a more general $G(a, b)$ model for $\Lambda(x)$. We do not explore this here or in the sequel due to the computational demands imposed under Gaussian process modulation for each of the alternating renewal processes.

$$\lim_{\Delta x_1 \rightarrow 0} \frac{1}{\Delta x_1} P(\text{event in } (x_1, x_1 + \Delta x_1) \mid \text{no event in } (0, x_1)) = \lambda(x_1).$$

Proceeding sequentially, (1) becomes

$$\begin{aligned} L(\lambda; \mathbf{x}) &= e^{-\int_0^{x_1} \lambda(t) dt} \lambda(x_1) e^{-\int_{x_1}^{x_2} \lambda(t) dt} \lambda(x_2) \cdots e^{-\int_{x_n}^T \lambda(t) dt} \\ &= e^{-\int_0^T \lambda(t) dt} \prod_{i=1}^n \lambda(x_i). \end{aligned} \quad (2)$$

We see that this likelihood is exactly that associated with an NHPP having intensity $\lambda(x)$ (Waller and Gotway, 2004; Cressie, 1993; Diggle, 2013; Illian et al., 2008). We have demonstrated a modeling equivalence between the NHPP as a finite point process model over $(0, T]$ and a renewal process truncated to the same interval. This equivalence is obtained only because observations are restricted to a bounded interval. We can not observe either process over R^+ and can not write a likelihood over R^+ . Furthermore, as noted in the Introduction, we can not extend this equivalence to the case of an alternating renewal process.

2.1 Alternating renewal processes

Above and below threshold events are observed as an alternating sequence of event times such that $0 < x_1 < y_1 < x_2 < \cdots < x_n < y_n < T$. Let $\{\mathbf{x}\}$ denote the sequence of start times for threshold exceedances, referred to as up-crossing events, and $\{\mathbf{y}\}$ denote the sequence of end times for threshold exceedances, referred to as down-crossing events. The duration of the i th exceedance is $y_i - x_i$ and the duration between when the i th exceedance ends and the $(i + 1)$ th exceedance begins is $x_{i+1} - y_i$. Assuming at time 0 the threshold is not being exceeded, the below threshold state renewal process, referred to as the up-crossing process, models the time until the first threshold exceedance occurs. Then, the above threshold state renewal process, or the down-crossing process, models the duration of the first exceedance event. We alternate between the two processes for the length of time of the study window with independent event times, hence, independent duration times.

Without loss of generality, we present the likelihood of the alternating renewal processes assuming that at time 0 we are in the below threshold state such that the first event is an up-crossing event. Additionally, we assume that the last event before time T is a down-crossing event such that at time T we are in the below threshold state waiting for the next exceedance event. Under independent renewals of up-crossing and down-crossing events, the likelihood can be written analogous to (1) where now

$$\begin{aligned}
L(\theta; \mathbf{x}, \mathbf{y}) &= P(\text{no up-crossing event in } (0, x_1)) \\
&\times P(\text{up-crossing event in } (x_1, x_1 + \Delta x_1) \mid \text{no up-crossing event in } (0, x_1)) \\
&\times P(\text{no down-crossing event in } (x_1 + \Delta x_1, y_1) \mid \text{up-crossing event in } (x_1, x_1 + \Delta x_1)) \\
&\times P(\text{down-crossing event in } (y_1, y_1 + \Delta y_1) \mid \text{no down-crossing event in } (x_1 + \Delta x_1, y_1)) \\
&\vdots \\
&\times P(\text{no up-crossing event in } (y_n + \Delta y_n, T) \mid \text{down-crossing event in } (y_n, y_n + \Delta y_n)).
\end{aligned}
\tag{3}$$

Letting $\lambda(t)$ and $\gamma(t)$ denote the intensity functions of the up-crossing and down-crossing processes, respectively, analogous to (2), the likelihood in (3) can be written as

$$\begin{aligned}
L(\theta; \mathbf{x}, \mathbf{y}) &= e^{-\int_0^{x_1} \lambda(t) dt} \lambda(x_1) e^{-\int_{x_1}^{y_1} \gamma(t) dt} \gamma(y_1) \\
&\times e^{-\int_{y_1}^{x_2} \lambda(t) dt} \lambda(x_2) e^{-\int_{x_2}^{y_2} \gamma(t) dt} \gamma(y_2) \\
&\times \vdots \\
&\times e^{-\int_{y_{n-1}}^{x_n} \lambda(t) dt} \lambda(x_n) e^{-\int_{x_n}^{y_n} \gamma(t) dt} \gamma(y_n) \\
&\times e^{-\int_{y_n}^T \lambda(t) dt}.
\end{aligned}
\tag{4}$$

If the last event before time T was an up-crossing event, the likelihood is written as

$$\begin{aligned}
L(\theta; \mathbf{x}, \mathbf{y}) &= e^{-\int_0^{x_1} \lambda(t) dt} \lambda(x_1) e^{-\int_{x_1}^{y_1} \gamma(t) dt} \gamma(y_1) \\
&\times e^{-\int_{y_1}^{x_2} \lambda(t) dt} \lambda(x_2) e^{-\int_{x_2}^{y_2} \gamma(t) dt} \gamma(y_2) \\
&\times \vdots \\
&\times e^{-\int_{y_{n-1}}^{x_n} \lambda(t) dt} \lambda(x_n) e^{-\int_{x_n}^T \gamma(t) dt}.
\end{aligned}$$

As an aside, in our application, durations of the below threshold state are long relative to durations of the above threshold state. Also, the latter never occur at the end of the ozone season. So, if in our analysis, T is the end of the ozone season, at time T we are always in the below threshold state.

A parametric model can be developed for the intensity functions, $\lambda(t)$ and $\gamma(t)$, using regression on covariates to help explain heterogeneity in the up-crossing and down-crossing processes. For example, we could set

$$\log(\lambda(t)) = W_1(t)\alpha$$

and

$$\log(\gamma(t)) = W_2(t)\beta.$$

Here, $W_1(t)$ and $W_2(t)$ are vectors of time-specific covariates that need not be the same for the two processes. We refer to these processes as regression modulated renewal processes. Further details regarding the parametrization of the alternating regression modulated renewal model are deferred to Section 3.

2.2 Extension to alternating Gaussian process modulated renewal processes

The modulated renewal process is driven by the time-varying intensity/hazard function, $\lambda(t)$. Following Rao and Teh (2011), we enrich the regression modulated process intensity function $\lambda(t)$ to $\lambda(t) = g(\lambda_0(t), \eta(t))$ where $\lambda_0(t)$ is the parametric form in the previous section while $\eta(t)$ is a Gaussian process over time with stationary covariance function. g is an *interaction* function which we take to be $g(\lambda_0(t), \eta(t)) = \lambda_0(t)e^{\eta(t)}$, i.e., $e^{\eta(t)}$ is a log Gaussian process. We now have a Gaussian process modulated renewal process.

We introduce Gaussian process modulations for both states of the alternating renewal process. That is, we modulate the intensity functions $\lambda(t)$ and $\gamma(t)$ using log-Gaussian processes. We have

$$\log(\lambda(t)) = W_1(t)\alpha + \eta_\lambda(t)$$

and

$$\log(\gamma(t)) = W_2(t)\beta + \eta_\gamma(t).$$

The η 's can also be viewed as temporal random effects providing local adjustments to the regression modulated intensities. In the point pattern literature, such intensities are associated with so-called log Gaussian Cox processes (Illian et al., 2008). We assume a Matérn family for the covariance functions with smoothness parameter ν fixed. As noted in the Introduction, a significant tradeoff arises between the model flexibility offered by this alternating modulation and the associated computational feasibility in model fitting. The computational challenge is accommodation of all of the integrals which are now stochastic in the likelihood.

3 Model inference and simulation

3.1 Model-based inference

There are challenges in modeling exceedance events and durations in continuous time with the foregoing alternating Gaussian process modulated renewal process specification. Exact fitting is computationally infeasible using the Gaussian process because it requires handling of all of the stochastic integrations in (4), none of which have a closed form. Note from (4), that altogether the stochastic integrations span $(0, T]$. With our application to ozone exceedances, at the hourly scale, the ozone season from May 1st to September 30th comprises 3,672 hours.

If event times are recorded hourly (i.e., x_i and y_j are at hourly scale), then any integral in (4) will be viewed as spanning a number of hours. We will express an integral spanning x hours as a sum of x integrals each spanning one hour. So, we need to approximate $\int_s^{s+1} \lambda(t) dt$ and $\int_s^{s+1} \gamma(t) dt$ where s is any hour during the ozone season. The hours play the role of *representative points* for the stochastic integrations and will introduce 3,672 $\lambda(s)$'s and 3,672 $\gamma(s)$'s as Gaussian random effects into the model. We use a first order approximation to obtain $\int_s^{s+1} \lambda(t) dt = \Lambda(s+1) - \Lambda(s) \approx \lambda(s)$ as a first order approximation. Substituting these approximations into (4) enables approximation of the likelihood. For example, the first line of (4) can be approximated as

$$\begin{aligned} e^{-\int_0^{x_1} \lambda(t) dt} \lambda(x_1) e^{-\int_{x_1}^{y_1} \gamma(t) dt} \gamma(y_1) \\ \approx e^{-\sum_{t=0}^{x_1-1} \lambda(t)} \lambda(x_1) e^{-\sum_{t=x_1}^{y_1-1} \gamma(t)} \gamma(y_1). \end{aligned} \quad (5)$$

We could imagine working at finer temporal resolution for the representative points but, for instance, at the scale of minutes, this would require introducing $60 \times 3,672$ random effects, increasing the computational burden by two orders of magnitude. Moreover, as we show below, employing the hourly scale over 3,672 hours performs well.

3.2 Nearest-neighbor Gaussian process specification

For computational feasibility, we turn to the sparse nearest-neighbor Gaussian process model (NNGP) which offers an approach to model inference by enabling tractable computation for Gaussian processes (Datta et al., 2016). Specifically, NNGPs provide efficient computation for large spatial or temporal datasets. We define NNGPs with neighborhoods of size m to model the Gaussian processes η_λ and η_γ . The benefit of using NNGPs for temporal random effects in $\mathbb{R}^1$ is that there is a natural ordering in time. Let $t_1, t_2, \dots, t_M$ denote a fixed and ordered set of equally space time points. Also, let $\mathcal{N}(t_i)$ denote the nearest neighbors of time point t_i which we assume to be the previous m time points for $i > m$ and the previous $i-1$ time points for $1 < i \leq m$. The nearest neighbor multivariate joint density of $(\eta_\lambda(1), \dots, \eta_\lambda(M))$ can be defined as

$$p(\eta_{\lambda}(1), \dots, \eta_{\lambda}(M)) = p(\eta_{\lambda}(1)) \prod_{k=2}^M p(\eta_{\lambda}(k) | \eta_{\lambda}(N(t_k)))$$

where $p(\eta_{\lambda}(1))$ is a normal distribution and $p(\eta_{\lambda}(k) | \eta_{\lambda}(N(t_k)))$ is a conditional normal distribution for $k = 2, \dots, M$. The density is a proper multivariate joint density and requires manipulations on matrices of at most dimension $m \times m$ making inference computationally feasible. The nearest neighbor joint density for $\eta_{\gamma}(t)$ is defined analogously.

3.3 Prior specifications

Model inference is obtained in a Bayesian framework. Let θ denote the model parameters, i.e., $\theta = (\alpha, \beta, \eta_{\lambda}, \eta_{\gamma}, \phi_{\eta_{\lambda}}, \sigma_{\eta_{\lambda}}^2, \phi_{\eta_{\gamma}}, \sigma_{\eta_{\gamma}}^2)$ and $p(\theta)$ denote the joint prior distribution. We obtain samples from the joint posterior distribution $p(\theta | \mathbf{x}, \mathbf{y}) \propto L(\theta; \mathbf{x}, \mathbf{y})p(\theta)$ using Markov chain Monte Carlo and a hybrid Metropolis within Gibbs sampling algorithm.

Prior distributions are assigned to the model parameters. The coefficient vectors α and β are each assigned multivariate normal prior distributions with mean vector $\mathbf{0}$ and covariance matrix $100^2 \mathbf{I}_p$ where p is the number of covariates and $\mathbf{I}_p$ denotes the identity matrix of dimension $p \times p$. Samples of α and β are each obtained in block using Metropolis-Hastings algorithms. The random effects, η_{λ} and η_{γ} , are modeled using NNGPs with covariance matrices defined above and means $-\sigma_{\eta_{\lambda}}^2 / 2$ and $-\sigma_{\eta_{\gamma}}^2 / 2$ where $\sigma_{\eta_{\lambda}}^2$ and $\sigma_{\eta_{\gamma}}^2$ are temporal variances yielding $E(e^{\eta_{\lambda}(t)}) = 1$ and $E(e^{\eta_{\gamma}(t)}) = 1$. Posterior samples are obtained using an elliptical slice sampling algorithm (Murray et al., 2010). The temporal variance parameters $\sigma_{\eta_{\lambda}}^2$ and $\sigma_{\eta_{\gamma}}^2$ are assigned inverse-gamma prior distributions. Lastly, the temporal decay parameters, $\phi_{\eta_{\lambda}}$ and $\phi_{\eta_{\gamma}}$ are assigned uniform prior distributions with lower and upper bounds set to control the range of temporal correlation. Posterior samples of the range parameters are obtained using a Metropolis-Hastings algorithm.

3.4 Model comparison

For full development of the alternating process models introduced in Section 2, methods for model comparison and variable selection are needed. The likelihood functions for both the alternating regression modulated and Gaussian process modulated models are constructed as products of sequential renewal processes and can be approximated by extending (5) to each element of (4). Therefore, we are able to approximate the likelihood functions of both models and can compute the log-likelihood ratio for competing models. Such ratios can be used to conduct variable selection as well as to compare the regression modulated intensity model to the Gaussian process modulated intensity model. A posterior distribution for the log-likelihood ratio can be computed for competing models using parameter values sampled from the posterior distributions under each of the models. This model comparison approach is demonstrated in Section 5 and Figure 1.

3.5 Simulating exceedance events and durations

One of the benefits of the alternating process models is that they offer a data generating mechanism that can be used for both simulating and predicting threshold exceedance events and durations that are consistent with how the events occur. To illustrate, first consider a single renewal process model for exceedance events. A convenient approach for generating a point pattern of exceedance events from a renewal process is to take advantage of the equivalence to the NHPP over a finite interval. That is, compute the maximum intensity over the temporal domain $(0, T]$, denoted $\lambda_{\max}$ (see Banerjee et al., 2014, for more details). Then, sample $n \sim \text{Poisson}(\lambda_{\max} T)$ and locate the n events uniformly over the interval $(0, T]$. Finally, thin the samples using rejection sampling where event t_i is retained with probability $\lambda(t_i)/\lambda_{\max}$ for i in $1, \dots, n$. The number and set of retained t s gives a realization of the number of exceedance events and set of times of when they occur.

Simulating or predicting exceedance events and durations from the alternating process model needs to be done sequentially. Again, let $\lambda(t)$ and $\gamma(t)$ be the intensities of the alternating renewal processes. Without loss of generality, assume that at time $t = 0$ we are in the below threshold state waiting for the first up-crossing event and that the alternating point patterns are defined over $(0, T]$. To simulate the first event, first draw $z_1 \sim \text{Uniform}(0, 1)$. Then, find x_1 such that

$$F(x_1) = 1 - e^{-\int_0^{x_1} \lambda(t) dt} = z_1.$$

When time is discretized to the hourly scale, we can approximate the integral and obtain the time of the first exceedance event as the x_1 that satisfies

$$-\log(1 - z_1) = \int_0^{x_1} \lambda(t) dt \approx \sum_{t=0}^{x_1-1} \lambda(t). \quad (6)$$

If $1 - e^{-\int_0^T \lambda(t) dt} < z_1$, then no up-crossing events occurred in $(0, T]$. Assuming

$1 - e^{-\int_0^T \lambda(t) dt} > z_1$, we obtain our first up-crossing event at time x_1 . We then simulate the first down-crossing event by again drawing $z_2 \sim \text{Uniform}(0, 1)$ and finding y_1 such that

$F(y_1 | x_1) = 1 - e^{-\int_{x_1}^{y_1} \gamma(t) dt} = z_2$. Here, on the discretized time scale, y_1 is the value that satisfies

$$-\log(1 - z_2) = \int_{x_1}^{y_1} \gamma(t) dt \approx \sum_{t=x_1}^{y_1-1} \gamma(t). \quad (7)$$

Following this alternating pattern, we continue simulation until either $1 - e^{-\int_{x_k}^T \gamma(t) dt} < z_{2k}$ indicating that at time T we are waiting for the k th down-crossing event to occur or $1 - e^{-\int_{y_k}^T \lambda(t) dt} < z_{2k+1}$ indicating that we are waiting for the $k+1$ th up-crossing to occur at time T .

By simulating realizations of the alternating renewal process during the time interval $(0, T]$ using (6) and (7), we realize exceedance events that can be used to obtain predictions of the number of events, the times in which the events occur, and their durations. Since each predicted exceedance event has to end before the next exceedance event begins, the model-generated predictions align with how exceedance events occur.

4 Hourly ozone data in the United States

Hourly ozone data from EPA Air Quality System² (AQS) were obtained for the 2014 ozone season spanning from May 1st to September 30th. The observed hourly ozone data for the week July 1 - July 7, 2014 for four monitoring stations located in Los Angeles, Phoenix, Colorado Springs, and Oklahoma City are shown in Figure A.1 Horizontal lines are drawn at two thresholds, 50 parts per billion (ppb) and 70 ppb, to identify below and above threshold states. Climate data, such as average hourly temperature, are also available at each of the monitoring stations. As seen in Figure A.1, ozone levels tend to follow a diurnal pattern where ozone peaks during the daytime hours and drops during the overnight hours (Bloomer et al., 2009; Jacob and Winner, 2009). Therefore, when an up-crossing event occurs at a time when the temperature trend is positive, the duration of the event may be longer than if it had occurred later in the day when the temperature trend was negative.

For each monitoring station and threshold, we obtained the set of up-crossing and down-crossing events for the 2014 ozone season. Observed up-crossing events were the hours in which ozone concentration exceeded the threshold such that the ozone concentration during the previous hour was below the threshold. Down-crossing events were the hours in which ozone concentration fell below the threshold such that the ozone concentration was above the threshold during the previous hour. From the hourly discretization, the minimum duration of an exceedance event is 1 hour.

The two thresholds, 50 ppb and 70 ppb, were chosen based on the possible future and current standards for ozone in the U.S. The National Ambient Air Quality Standard (NAAQS) for ground-level ozone is an ambient air regulation established by the U.S. EPA under the Clean Air Act. The primary (health based) NAAQS³ was lowered in October 2015 to 70 ppb and is based on the 4th highest daily maximum (using 8-hour averages) averaged over 3 consecutive years. All four monitoring sites considered here for the 2014 ozone season are out of the compliance with the new ozone NAAQS.

²Data are available for download at http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html

³<https://www.epa.gov/criteria-air-pollutants/naaqs-table#3>

A summary of the number of exceedance events of the two thresholds for each of the four stations and their durations are given in Table 1. For the Los Angeles location, we also computed the 8-hour average ozone concentration throughout the ozone season and obtained the resulting up-crossing and down-crossing events to more closely align with the standard. The number of exceedance events ranged from 1 to 175. Los Angeles had more exceedances of the 70 ppb threshold than of the 50 ppb threshold. This is the result of multiple up-crossing and down-crossing pairs of events of the higher threshold occurring during one longer duration in the above 50 ppb threshold state. The number of exceedances using the hourly ozone data was greater than the number using the 8-hour average ozone data, however, the durations were on average much shorter. Hourly ozone data for Phoenix and Oklahoma City resulted in more exceedances of the 50 ppb threshold than the 70 ppb threshold and Colorado Springs had the same number of exceedance events for the two thresholds. Fewer exceedance events of the higher threshold indicate that the exceedance events are less severe. The median duration of an exceedance event for the 50 ppb threshold ranged from 6 to 8 hours for the four stations. For the 70 ppb threshold, the median duration was between 1 and 3 hours for the hourly ozone data and was 7 hours for the 8-hour average ozone data. The maximum durations of exceedance events of the 50 and 70 ppb threshold were 63 and 13 hours, respectively, both of which occurred in Los Angeles.

5 Alternating modulated renewal process analyses: simulation and real data

We first illustrate the performance of the alternating renewal process model through a simulation study and then apply the model to exceedance data from four cities in the U.S.

5.1 Simulation study

We simulated alternating up-crossing and down-crossing event pairs under the regression modulated renewal process model of Section 2.1 using (6) and (7) for the ozone season spanning 153 days. Time was discretized to the hourly time scale resulting in 3,672 time points. To capture the diurnal pattern of ozone, two climate variables, average hourly temperature and the rate of change in average hourly temperature, computed as the difference in average hourly temperature, were included in the regression equations of the up-crossing and down-crossing process intensity functions. The coefficient vectors $\alpha = (-3.00, 1.00, 0.25)$ and $\beta = (-2.00, -0.50, -0.25)$ were used for the simulation along with the observed climate variables from the Colorado Springs location. These coefficient values are similar to the posterior mean estimates of the coefficients obtained when fitting the model to the observed data (see Section 5.2) and result in exceedance events that resemble those observed for this location both in terms of the number and the distribution of durations. The simulated number of event pairs was 126 and the percentage of hours in the above threshold state was 23.01%. The duration of the exceedances ranged from 1 to 22 hours with a median of 5.5 hours.

Markov chain Monte Carlo model fitting was run for 20,000 iterations. The first 10,000 iterations were disregarded as burn-in and the remainder were used for inference. The posterior mean and 95% credible intervals for the coefficient parameters are given in Table

2. The parameter estimates indicate that the model is able to accurately recover the true values of the parameters.

To assess the variability in the estimates of the number of exceedance events and durations of total exceedances during the ozone season, we simulated up-crossing and down-crossing event pairs according to Section 3.5 using the posterior distribution of the model parameters. We simulated 1,000 realizations from the alternating regression modulated renewal processes and obtained 1,000 sequences of alternating up-crossing and down-crossing events. The mean number of event pairs from the 1,000 simulations was 128 with lower and upper limits of 110 and 148, respectively. The mean percentage of hours during the time period spent in the above threshold state was 23.60% with lower and upper limits of 19.74 and 27.67%. Both estimates are similar to those of the true simulated data (126 event pairs and 23.01% above threshold). For the month of July, the true number of event pairs was 34 and the estimated number of event pairs ranged between 20 and 43 with a mean of 32. In the true simulated data, an exceedance event occurred on July 1st at 3pm and had a duration of 3 hours. Using the generated exceedance events and durations, we computed the probability of being in the above threshold state for three different times on July 1st. The probability of being in the above threshold state at 12pm, 5pm, and 10pm was 0.19, 0.36, and 0.15, respectively.

5.2 Ozone exceedance events and durations data in the United States

We now apply the alternating regression modulated and Gaussian process modulated renewal process models to exceedance data obtained for the two different thresholds and four urban monitoring stations. The two covariates, average hourly temperature and the rate of change in average hourly temperature, are obtained at each of the monitoring stations and included in the regression equations of the up-crossing and down-crossing process intensity functions. For the Gaussian process modulated process model, the nearest-neighbor Gaussian process is employed with neighborhoods of size $m = 10$. Datta et al. (2016) experimented with different values of m and found $m = 10$ to be adequate with little sensitivity. The two temporal covariances are defined independently using Matérn covariance functions, with $\nu = 3/2$. The temporal variance parameters $\sigma_{\eta_\lambda}^2$ and $\sigma_{\eta_\gamma}^2$ are assigned IG(10,1/2) prior distributions penalizing large variances. We assign Uniform(0.4,1) prior distributions to both ϕ_{η_λ} and ϕ_{η_γ} resulting in an effective range between approximately 5 and 12 hours.

When fitting the model to the 70 ppb threshold exceedance data, the temporal random effect $\eta_\gamma(t)$ in the down-crossing intensity $\gamma(t)$ is omitted. That is, the up-crossing process is a Gaussian process modulated renewal process while the down-crossing process is a regression modulated renewal process. This simplification is due to fewer exceedance events and shorter durations in the above 70 ppb threshold state during the ozone season and thus, less information in the data to estimate the parameters of the down-crossing process. For the 70 ppb threshold, the model is fitted to both the hourly and 8-hour average exceedance data for the Los Angeles location. We did not fit the model to the 70 ppb threshold exceedance data for the Colorado Springs location since there was only one event for the 2014 ozone season.

Inference for each of the models is obtained using MCMC. For each threshold and location combination, the chain for the regression modulated renewal process model is run for 20,000 iterations and the chain for the Gaussian process modulated renewal process model is run for 100,000 iterations. The first half of each chain is disregarded as burn-in.

Boxplots of the posterior distributions of the log-likelihood ratio of the two alternating process models are shown in Figure 1 for each location and threshold. The log-likelihood ratios favor the Gaussian process modulated renewal process model over the regression modulated renewal process model for all locations and thresholds.

The parameter estimates for the alternating Gaussian process modulated renewal process model using exceedances of the 50 ppb threshold are given in Table 3. For each location, the parameter estimates for α_1 and α_2 indicate that the intensity functions of the up-crossing processes increase with hourly average temperature and with increases in the rate of change of hourly average temperature. The intensity functions of the down-crossing processes decrease with both hourly average temperature and the rate of change of hourly average temperature. This indicates that when temperature is high or is increasing, the average rate of occurrence of a down-crossing event decreases, resulting in longer durations of exceedance events. Conversely, when temperature is low or is decreasing, the rate of the down-crossing process increases resulting in shorter durations of threshold exceedances. The estimates of the effective range of the temporal random effects of both the up-crossing process and down-crossing process for the 50 ppb threshold are also given in Table 3. For the four locations the effective range was slightly larger for the up-crossing process than the down-crossing process.

Figure 2 shows the posterior median estimates of the intensity functions, $\lambda(t)$ and $\gamma(t)$, for the 50 ppb threshold and each of the four locations for the week July 1 - July 7. The black curve denotes the up-crossing process intensity function, $\lambda(t)$, during the time spent in the below threshold state (waiting for an up-crossing event) and the red curve denotes the down-crossing process intensity function, $\gamma(t)$, during the time spent in the above threshold state (waiting for a down-crossing event). The two intensity functions are shown alternating between the below threshold state and above threshold state. The figures indicate that the intensity functions for each of the processes increase from the time of the previous event until the next event occurs. Also, when in the below threshold state, the up-crossing intensity function $\lambda(t)$ increases during the day time hours and decreases during the night time hours. Similarly, Figure 3 shows the posterior median estimates of the intensity functions for the models fitted to the 70 ppb threshold exceedance data for the week August 29 - September 4. We notice that on August 29th and 30th, there were two up-crossing events for hourly ozone data for Los Angeles but only one up-crossing event for 8-hour average ozone concentration data, indicating that the second event was less severe than the first. The exceedance events on September 1 and 3 were also less severe as they did not result in 8-hour average ozone concentration levels above the 70 ppb threshold.

We simulated event pairs according to Section 3.5 using the posterior distribution of the model parameters for each location and threshold. We generated 1,000 realizations from the Gaussian process modulated renewal process model and obtained 1,000 alternating

sequences of up-crossing and down-crossing events. Figure 4 shows boxplots of the distribution of the predicted number of exceedance events, the distribution of durations of exceedance events, and the proportion of hours during the ozone season above the 50 ppb threshold for the four locations. The $\times$ s on each boxplot denote the observed value for the location. The red boxplots on the middle panel show the distributions of the observed durations of exceedance events. These figures illustrate that from the fitted Gaussian process modulated renewal process model, we are able to simulate realizations of exceedance events and durations that effectively capture the physical processes of threshold exceedances. Moreover, these realizations are consistent with how the exceedance events occur and are observed. Similar results are seen in Figure 5 for exceedance events and durations of the 70 ppb threshold.

Using the simulated realizations of event pairs for the ozone season and the 50 ppb threshold, we estimate the probability of being in the above threshold state for each hour of the day. The probability estimate was computed as the proportion of alternating two state process simulations that were in the above threshold state waiting for a down-crossing event each hour. The posterior probabilities are given for each of the four locations for the days May 26th, July 4th, and September 1st in Figures 6-8 respectively. The $\times$ s are the hours of the observed above threshold state for that day and location. In general, the probabilities increase during the daytime hours and during the hours of observed above threshold states. Interestingly, Figure 6 indicates that in Los Angeles, an exceedance event from the previous day had not yet ended by the early morning hours of May 26th. This highlights the benefit of our alternating renewal process model in capturing *both* exceedance events and durations. On September 1st, neither Phoenix nor Oklahoma City experienced a threshold exceedance event and the probabilities estimated from the model are lower than the other two locations that did have threshold exceedance events. These results emphasize the difference in exceedance patterns in terms of incidence and duration across locations as well as between hourly and 8-hour average ozone, thus, increasing our understanding of exceedance events.

6 Discussion and Future work

We developed modeling for alternating point patterns of exceedance events that incorporates both the time in which exceedance events occur and their durations. The alternating processes are defined with parametric intensities and extended to log Gaussian process driven intensities in order to provide temporal dependence and local refinement. The alternating sequence of up-crossing and down-crossing events reflects how exceedance events occur. Therefore, the model can be used to simulate or predict exceedance events and durations given intensity functions of the up-crossing and down-crossing processes.

The generative alternating point pattern model provides a framework for several directions of future work. In particular, we are interested in extending the log Gaussian processes to account for spatial dependence within the intensity functions for the alternating renewal processes across locations. This will enable prediction of exceedance events and durations at unobserved locations. Extending the log Gaussian processes to spatio-temporal intensity functions will require spatio-temporal NNGPs to enable computation. Additionally, models for multivariate exceedance events would be attractive. For example, we might be interested

in jointly modeling exceedances for ozone concentration and for particulate matter, e.g. $PM_{2.5}$. Novel modeling challenges will arise.

A Appendix

Observed hourly ozone for the week July 1 - July 7, 2014 for the four locations, Los Angeles, Phoenix, Colorado Springs, and Oklahoma City.

Fig. A.1.
Observed hourly ozone for the week July 1 - July 7, 2014 for the four locations, Los Angeles, Phoenix, Colorado Springs, and Oklahoma City.

References

- Achcar JA, Coelho-Barros EA, and de Souza RM (2016). Use of non-homogeneous Poisson process (NHPP) in presence of change-points to analyze drought periods: a case study in Brazil. *Environmental and Ecological Statistics*, pages 1–15.
- Achcar JA, Fernández-Bremauntz AA, Rodrigues ER, and Tzintzun G (2008). Estimating the number of ozone peaks in Mexico City using a non-homogeneous Poisson model. *Environmetrics*, 19(5): 469–485.
- Achcar JA, Rodrigues ER, and Tzintzun G (2011). Using non-homogeneous poisson models with multiple change-points to estimate the number of ozone exceedances in mexico city. *Environmetrics*, 22(1): 1–12.
- Banerjee S, Carlin BP, and Gelfand AE (2014). *Hierarchical modeling and analysis for spatial data*. CRC Press.
- Beichelt F (2006). *Stochastic processes in science, engineering and finance*. CRC Press.
- Bell ML and Dominici F (2008). Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities. *American Journal of Epidemiology*, 167(8):986–997. [PubMed: 18303005]
- Bell ML, McDermott A, Zeger SL, Samet JM, and Dominici F (2004). Ozone and short-term mortality in 95 US urban communities, 1987–2000. *Journal of the American Medical Association*, 292(19): 2372–2378. [PubMed: 15547165]
- Berman M (1981). Inhomogeneous and modulated gamma processes. *Biometrika*, 68(1):143–152.
- Birolini A (2012). *On the use of stochastic processes in modeling reliability problems*, volume 252 Springer Science & Business Media.
- Bloomer BJ, Stehr JW, Piety CA, Salawitch RJ, and Dickerson RR (2009). Observed relationships of ozone air pollution with temperature and emissions. *Geophysical Research Letters*, 36(9).
- Coles S, Bawa J, Trenner L, and Dorazio P (2001). *An introduction to statistical modeling of extreme values*, volume 208. Springer.
- Cox D (1972). The statistical analysis of dependencies in point processes. *Stochastic point processes*, pages 55–66.
- Cressie N (1993). *Statistics for spatial data*.
- Datta A, Banerjee S, Finley AO, and Gelfand AE (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association, Series B (Statistical Methodology)*.
- Diggle PJ (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press.
- Gryparis A, Coull BA, Schwartz J, and Suh HH (2007). Semiparametric latent variable regression models for spatiotemporal modelling of mobile source particles in the greater Boston area. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(2):183–209.
- Guarnaccia C, Quartieri J, Tepedino C, and Rodrigues ER (2015). An analysis of airport noise data using a non-homogeneous poisson model with a change-point. *Applied Acoustics*, 91:33–39.
- Huerta G, Sansó B, and Stroud JR (2004). A spatiotemporal model for Mexico City ozone levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(2):231–248.
- Illian J, Penttinen A, Stoyan H, and Stoyan D (2008). *Statistical analysis and modelling of spatial point patterns*, volume 70 John Wiley & Sons.
- Jacob DJ and Winner DA (2009). Effect of climate change on air quality. *Atmospheric Environment*, 43(1):51–63.
- Lasko TA (2014). Efficient inference of Gaussian-process-modulated renewal processes with application to medical event data. In *Uncertainty in Artificial Intelligence (UAI)*, volume 2014, pages 469–476. NIH Public Access.
- Leadbetter MR (1991). On a basis for Peaks over Threshold? modeling. *Statistics & Probability Letters*, 12(4):357–362.
- Lervolino I, Giorgio M, and Polidoro B (2014). Sequence-based probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America*, 104(2):1006–1012.

- Murray I, Adams RP, and MacKay DJ (2010). Elliptical slice sampling. *Journal of Machine Learning Research: Workshop and Conference Proceedings (AISTATS)*, 9:541?548.
- Norris JR (1998). *Markov chains*. Number 2008. Cambridge University Press.
- Rao V and Teh YW (2011). Gaussian process modulated renewal processes. In *Advances in Neural Information Processing Systems*, pages 2474–2482.
- Rodrigues ER, Gamerman D, Tarumoto MH, and Tzintzun G (2014). A non-homogeneous poisson model with spatial anisotropy applied to ozone data from Mexico City. *Environmental and Ecological Statistics*, 22(2):393–422.
- Sahu SK, Gelfand AE, and Holland DM (2007). High-resolution space–time ozone modeling for assessing trends. *Journal of the American Statistical Association*, 102(480):1221–1234. [PubMed: 19759840]
- Trivedi KS (2001). *Probability & statistics with reliability, queuing and computer science applications*. John Wiley & Sons.
- U.S. Environmental Protection Agency (2011). *Ozone health risk assessment for selected urban areas* EPA Office of Air Quality Planning and Standards, Research Triangle Park, NC, EPA/452/R-07-009.
- U.S. Environmental Protection Agency (2013). *2013 Final Report: Integrated science assessment of ozone and related photochemical oxidants* U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-10/076F.
- Waller LA and Gotway CA (2004). *Applied spatial statistics for public health data*, volume 368 John Wiley & Sons.

Fig. 1.

Log-likelihood ratio of the Gaussian process modulated renewal process model over the regression modulated renewal process model for the four locations and the 50 ppb threshold (left) and 70 ppb threshold (right).

Fig. 2. Intensity functions for the 50 ppb threshold exceedance data for July 1 - July 7, 2014 for the four locations, Los Angeles, Phoenix, Colorado Springs, and Oklahoma City.

Fig. 3. Intensity functions for the 70 ppb threshold exceedance data for August 29 - September 4, 2014 for Los Angeles (hourly and 8-hour average), Phoenix, and Oklahoma City.

Fig. 4.

Boxplots of the number of predicted number of exceedance events, the distribution of durations of exceedance events, and the proportion of hours during the ozone season above the 50 ppb threshold for the four locations, Los Angeles, Phoenix, Colorado Springs, and Oklahoma City.

Fig. 5.

Boxplots of the number of predicted number of exceedance events, distribution of durations of exceedance events, and the proportion of hours during the ozone season above the 70 ppb threshold for Los Angeles (hourly and 8-hour average), Phoenix, and Oklahoma City.

Fig. 6.
Probability of being above the 50 ppb threshold for each hour on May 26, 2014 for the four locations, Los Angeles, Phoenix, Colorado Springs, and Oklahoma City.

Fig. 7.
Probability of being above the 50 ppb threshold for each hour on July 4, 2014 for the four locations, Los Angeles, Phoenix, Colorado Springs, and Oklahoma City.

Fig. 8.
Probability of being above the 50 ppb threshold for each hour on September 1, 2014 for the four locations, Los Angeles, Phoenix, Colorado Springs, and Oklahoma City.

Table 1

Summary of exceedance data for the four monitoring stations and two thresholds. The median (min, max) durations are given in hours.

Threshold = 50 ppb			
Location	# of exceedances	Durations	Exceedance percentage
Los Angeles	175	8 (1, 63)	47.47
Phoenix	148	6 (1, 12)	23.72
Colorado Springs	156	6 (1, 40)	27.40
Oklahoma City	105	6 (1, 20)	18.16
Threshold = 70 ppb			
Location	# of exceedances	Durations	Exceedance percentage
Los Angeles	75	3 (1, 12)	7.73
Los Angeles (8-Hour Average)	32	7 (2, 13)	5.95
Phoenix	32	2 (1, 6)	2.21
Colorado Springs	1	1	0.02
Oklahoma City	12	3 (1, 5)	0.84

Table 2

True values along with posterior mean estimates and 95% credible intervals for the coefficient parameters in the simulation.

Parameter	Truth	Mean	Credible Interval
<i>Up-crossing process</i>			
α_0	-3.00	-3.06	(-3.29, -2.83)
α_1	1.00	1.15	(0.92, 1.37)
α_2	0.25	0.26	(0.20, 0.32)
<i>Down-crossing process</i>			
β_0	-2.00	-1.98	(-2.20, -1.79)
β_1	-0.50	-0.62	(-0.84, 0.39)
β_2	-0.25	-0.24	(-0.30, -0.19)

Table 3

Posterior mean estimates and 95% credible intervals for the coefficient parameters in the Gaussian process modulated renewal process model for each location and 50 ppb threshold.

	Los Angeles	Phoenix	Colorado Springs	Oklahoma City
Parameter	Mean (CI)	Mean (CI)	Mean (CI)	Mean (CI)
<i>Up-crossing process</i>				
a_0	-2.70 (-2.91, -2.50)	-3.39 (-3.65, -3.13)	-3.20 (-3.42, -2.99)	-4.06 (-4.38, -3.76)
a_1	0.80 (0.62, 0.99)	0.99 (0.78, 1.20)	1.11 (0.91, 1.31)	0.93 (0.68, 1.17)
a_2	0.37 (0.31, 0.43)	0.33 (0.27, 0.40)	0.23 (0.18, 0.27)	0.51 (0.41, 0.61)
$\sigma_{\eta_\lambda}^2$	0.08 (0.04, 0.13)	0.06 (0.03, 0.11)	0.05 (0.03, 0.09)	0.05 (0.03, 0.09)
φ_{η_λ}	0.75 (0.46, 0.99)	0.70 (0.41, 0.99)	0.59 (0.41, 0.97)	0.63 (0.41, 0.96)
effective range	6.72 (4.73, 10.32)	7.30 (4.78, 11.45)	9.90 (6.19, 11.72)	8.05 (4.97, 11.64)
<i>Down-crossing process</i>				
B_0	-2.49 (-2.71, -2.28)	-1.92 (-2.19, -1.68)	-1.78 (-2.00, -1.58)	-1.68 (-1.95, -1.44)
B^1	-0.61 (-0.80, -0.43)	-0.18 (-0.37, -0.03)	-0.42 (-0.61, -0.23)	-0.44 (-0.68, -0.20)
B^2	-0.40 (-0.49, -0.31)	-0.58 (-0.67, -0.50)	-0.28 (-0.33, -0.23)	-0.31 (-0.39, -0.24)
$\sigma_{\eta_\gamma}^2$	0.06 (0.03, 0.12)	0.09 (0.04, 0.20)	0.05 (0.03, 0.10)	0.09 (0.03, 0.17)
φ_{η_γ}	0.77 (0.47, 0.99)	0.82 (0.93, 0.99)	0.73 (0.49, 0.98)	0.83 (0.60, 0.99)
effective range	6.44 (4.78, 10.76)	5.89 (4.78, 8.01)	5.65 (4.78, 6.86)	5.86 (4.73, 8.02)