The Annals of Applied Statistics
2023, Vol. 17, No. 3, 2078–2094
https://doi.org/10.1214/22-AOAS1708

LOG-GAUSSIAN COX PROCESS MODELING OF LARGE SPATIAL LIGHTNING DATA USING SPECTRAL AND LAPLACE APPROXIMATIONS

BY MEGAN L. GELSINGER[1,a], MARYCLARE GRIFFIN[2,d], DAVID MATTESON[1,b] AND JOSEPH GUINNESS[1,c]

[1]*Department of Statistics and Data Science, Cornell University,* [a]*mlg276@cornell.edu,* [b]*matteson@cornell.edu,* [c]*guinness@cornell.edu*

[2]*Department of Mathematics and Statistics, University of Massachusetts at Amherst,* [d]*maryclaregri@umass.edu*

Lightning is a destructive and highly visible product of severe storms, yet there is still much to be learned about the conditions under which lightning is most likely to occur. The GOES-16 and GOES-17 satellites, launched in 2016 and 2018 by NOAA and NASA, collect a wealth of data regarding individual lightning strike occurrence and potentially related atmospheric variables. The acute nature and inherent spatial correlation in lightning data renders standard regression analyses inappropriate. Further, computational considerations are foregrounded by the desire to analyze the immense and rapidly increasing volume of lightning data. We present a new computationally feasible method that combines spectral and Laplace approximations in an EM algorithm, denoted SLEM, to fit the widely popular log-Gaussian Cox process model to large spatial point pattern datasets. In simulations we find SLEM is competitive with contemporary techniques in terms of speed and accuracy. When applied to two lightning datasets, SLEM provides better out-of-sample prediction scores and quicker runtimes, suggesting its particular usefulness for analyzing lightning data which tend to have sparse signals.

1. Introduction. Lightning has great destructive capabilities, and there is growing concern surrounding the relationship between climate change and lightning activity (Clark, Ward and Mahowald (2017), Finney et al. (2018)). In 2016, lightning was added to the Global Climate Observing System's (GCOS) list of Essential Climate Variables, indicators of particular focus for scientists looking to understand and mitigate climate impacts (System (2016)). In October 2017, scientists associated with GCOS and several other meteorological organizations assembled a task force to spearhead a new wave of lightning research (Aich et al. (2018)). Satellite data was identified as a crucial source of information for future lightning study (Aich et al. (2018)). Modern satellite technology is capable of monitoring lightning activity over large, for example, 1000×1000 km, spatial grids. This spatial scale allows researchers to conduct novel studies of macrolevel lightning dynamics but poses a challenge to computational feasibility. Another difficulty lies in the sparsity of lightning count data. At most locations at any given time, there is no lightning which means that the vast majority of recorded counts are zero. These problems motivate our study of computationally feasible statistical methods for satellite lightning data, with the end goal of facilitating novel studies of lightning dynamics that can make use of modern satellite data.

We concentrate on data collected by instruments on the GOES-16 satellite, launched in 2016 by the National Oceanic and Atmospheric Administration (NOAA) and the National Aeronautics and Space Administration (NASA). The first instrument of interest to our study is the Advanced Baseline Imager (ABI), which records images in 16 different spectral bands, corresponding to environmental factors such as water-based cloud coverage and dust, haze, and smoke presence. We focus on the ABI's mesoscale mode of operation, which collects

Received November 2021; revised May 2022.
Key words and phrases. Spatial point pattern, Laplace approximation, expectation-maximization.

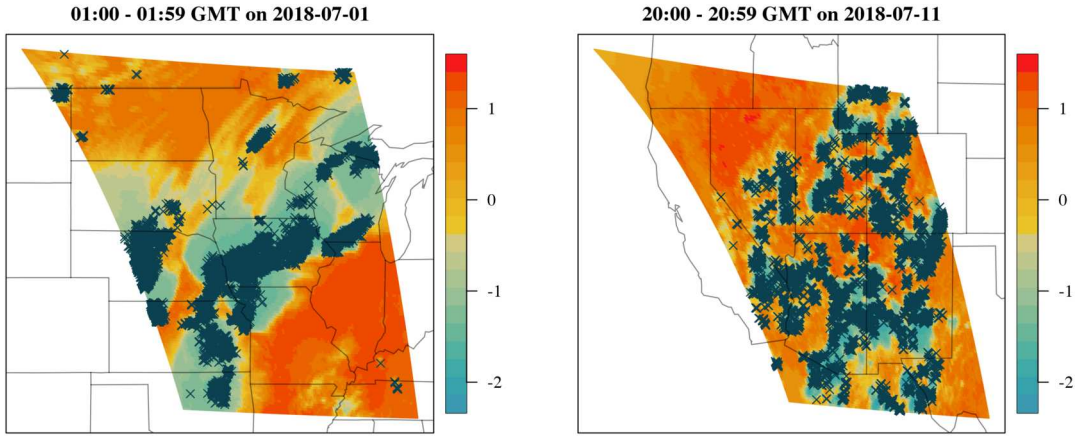

FIG. 1. *Lightning strike locations* (×) *overlaid on proxy data for cloud-top height.*

information over approximately a 1000 × 1000 km field-of-view, often targeted at areas of intense storm activity in North America. This field-of-view can change hourly as the pattern of storms changes. Data is recorded on a minute-by-minute basis, at either two km, one km, or 0.5 km resolution, depending on the spectral band (DOC NOAA and NASA (2017)). The second instrument is the Geostationary Lightning Mapper (GLM), which continuously measures all types of lightning activity at an eight km resolution over the Americas and adjacent oceanic regions. Flashes are detected by their radiance signature—optical pulses which exceed the background instrument threshold (Goodman et al. (2013)). For ease of comparison, we analyze both the ABI and GLM data at an eight km spatial resolution. Figure 1 shows environmental proxies derived from ABI data overlaid with GLM strike data. Construction of these environmental proxies is later described in detail in Section 4.1.

Log-Gaussian Cox process (LGCP) models are commonly used to model spatial point pattern data (Møller, Syversveen and Waagepetersen (1998)). As such, they are natural tools for studying lightning dynamics using strike and environmental proxy data. To define the LGCP model, consider a point pattern whose locations $\mathbf{U} = \{U_1, \ldots, U_\ell\}$ fall within the domain $S \subset \mathbb{R}^2$. As is common when analyzing point patterns, we use a discretization approximation and partition the domain S into an $n_1 \times n_2$ grid with n pixels $\{B_1, \ldots, B_n\}$ and pixel centroids $\{s_1, \ldots, s_n\}$. We transform the observed locations into counts per pixel via $Y_i = \sum_{j=1}^{\ell} \mathbf{1}_{U_j \in B_i}$, for $i = 1, \ldots, n$. We also consider $\mathbf{X}(s_i) = (1, X_1(s_i), \ldots, X_p(s_i))^T$, a $1 \times (p+1)$ row vector of an intercept and covariates considered constant within the ith pixel. An LGCP model for $\mathbf{U}$ implies the following model for $Y_1, \ldots, Y_n$:

$$Y_i | \lambda \overset{\text{indep}}{\sim} \text{Poisson}\left(\int_{B_i} \lambda(s)\, ds\right) = \text{Poisson}\{\Delta_i \lambda(s_i)\}, \tag{1}$$

$$\lambda(s_i) = \exp\{\mathbf{X}(s_i)\boldsymbol{\beta} + Z(s_i)\}, \tag{2}$$

$$Z(s) \sim GP\{0, K(\boldsymbol{\eta})\}, \tag{3}$$

where λ is an intensity function that is constant within each pixel, Δ_i is the area of pixel B_i, $\boldsymbol{\beta}$ is a $(p+1) \times 1$ vector of coefficients, and $K(\boldsymbol{\eta})$ is a covariance function parameterized by $\boldsymbol{\eta}$. Letting $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))$ denote points from the Gaussian field, this is equivalent to assuming that $\mathbf{Z} \sim N(\mathbf{0}, \Sigma_{\boldsymbol{\eta}})$, where $\Sigma_{\boldsymbol{\eta}}$ is the covariance matrix formed by applying $K(\boldsymbol{\eta})$ to locations $s_1, \ldots, s_n$. Even with the discretezation approximation, evaluating the likelihood

remains challenging due to integration over the random effects $\mathbf{Z}$,

$$L(\boldsymbol{\theta}; \mathbf{Y}) = p(\mathbf{Y}|\boldsymbol{\theta}) = \int_{\mathbb{R}^n} p(\mathbf{Y}, \mathbf{Z}|\boldsymbol{\theta})\, d\mathbf{Z} = \int_{\mathbb{R}^n} p(\mathbf{Y}|\mathbf{Z}, \boldsymbol{\theta})\, p(\mathbf{Z}|\boldsymbol{\theta})\, d\mathbf{Z}, \tag{4}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\eta})$.

Markov chain Monte Carlo (MCMC) methods are popular for fitting Bayesian LGCP models, providing exact inference given infinitely many samples from the posterior. Brix and Diggle (2001) and Diggle, Rowlingson and Su (2005) implement the Metropolis-adjusted Langevin kernel discussed in Møller, Syversveen and Waagepetersen (1998) in an MCMC routine for LGCPs, making use of circulant embedding (Wood and Chan (1994)), which leverages fast Fourier transforms to speed up matrix computations. This method is implemented in the R package `lgcp` (Taylor et al. (2013, 2015)). While the "exactness" of this method is appealing, it is also known to have slow runtime, can mix poorly, and requires specification of user-defined tuning parameters (Taylor and Diggle (2014), Shirota and Gelfand (2016)). To address these issues, Guan and Haran (2018) introduce an approximate method which projects the random effects onto a lower-dimensional subspace. This reduces the dimension of the random effects and alleviates spatial confounding. Likewise, sampling the random effects involves manipulation of a lower dimensional matrix with better mixing properties. Gonçalves and Gamerman (2018) provide an exact MCMC method for spatial-temporal data in the sense that the method has no spatial discretization error.

Maximum likelihood schemes are also popular, but approximations are used due to the intractability of evaluating the likelihood. Guan and Haran (2020) use an expectation-maximization (EM) algorithm in which the E-step is approximated via sampling or Laplace approximation. Park and Haran (2021) use a Monte Carlo likelihood approximation instead, introducing a method for finding a good importance function iteratively. Both Guan and Haran (2020) and Park and Haran (2021) also use similar projection-based approximations to Guan and Haran (2018) to reduce computational burden and address spatial confounding.

An especially well-known approximation method is the integrated nested Laplace approximation (INLA). As the name suggests, the key feature of INLA is its nested approximation of the marginal posterior distribution of the model's hyperparameters, such as $\boldsymbol{\eta}$, via the Laplace approximation (Rue, Martino and Chopin (2009), Illian, Sørbye and Rue (2012)). INLA assumes that the Gaussian process driving the spatial point process is a Gaussian Markov random field and thus has sparse precision matrices (Lindgren, Rue and Lindström (2011)), facilitating faster matrix operations.

In a related alternative, Zilber and Katzfuss (2021) combine the Laplace approximation with a computationally efficient Vecchia approximation to the latent Gaussian process, implemented in the R package GPvecchia (Katzfuss et al. (2021)). Guan and Haran (2020) also propose a variant to their method which leverages the Laplace approximation in the E-step, instead of using Monte Carlo averages. However, scalability to datasets measured on large spatial grids, on the order of tens of thousands as opposed to hundreds of locations, still remains in question, even with the general computational time advantages of these methods compared to MCMC based approches (Taylor and Diggle (2014), Guan and Haran (2020)).

Despite recent advances computational considerations remain critical due to the ever-increasing sizes of modern datasets such as the GOES-16 satellite data. In this work: (i) we introduce an EM algorithm which leverages both the Laplace approximation and fast and scalable FFT algorithms to facilitate matrix computations. While spectral methods are powerful, they do not solve all of the computational challenges within the EM algorithm. (ii) To address these remaining challenges, we also use the Hutchinson trace approximation (Hutchinson (1989)). (iii) Additionally, we craft a local covariance matrix approximation that can be combined with the Laplace approximation to approximate the conditional mean of the Gaussian

field, given the data, after $\boldsymbol{\eta}$ and $\boldsymbol{\beta}$ have been estimated. Combined, these techniques form the proposed Spectral-Laplace-Expectation-Maximization (SLEM) method for efficient estimation of LGCP models from large spatial point pattern data. In simulations SLEM yields sizeable computatational advantages with faster runtimes than the Vecchia-Laplace method. These runtime gains are accompanied by competitive estimation of $\boldsymbol{\beta}$ and slightly less accurate estimation of the latent field. On the lightning datasets we consider, which are sparser than the simulation data, SLEM is both faster and more accurate on an out-of-sample log score prediction metric. Importantly, the speed and accuracy of SLEM on the lightning datasets allows us to draw insights into lightning dynamics from the GOES-16 satellite data that are consistent with the existing literature.

1.1. *Expository analysis of lightning data.* To illustrate the difficulty of estimating LGCPs on GOES lightning data, which have both large spatial scale (125×125 pixel grid) and sparse signals (few pixels with nonzero strikes), we present results from two contemporary techniques—the Vecchia–Laplace (VL) algorithm, as implemented in the R package `GPvecchia` (Katzfuss et al. (2021)), and INLA, as implemented in the `INLA` and `inlabru` R packages (Martins et al. (2013), Bachl et al. (2019)).

We apply both methods to GLM lightning data collected over the upper Midwest region of the United States between 01:00–01:59 GMT on 2018-07-01. We convert the strikes to counts per pixel on an evenly spaced 125×125 grid ($n = 15{,}625$) in order to model the lightning and covariate data on the same spatial scale. We include an intercept and several covariates in the model. The covariates include proxies for cloud growth and cloud-top height which are described in Section 4.1. We also use elevation as a covariate in our model. These covariates are currently believed to be associated with lightning occurrence, making them natural predictors to include in the model (Henderson, Otkin and Mecikalski (2021), Lee, Kummerow and Zupanski (2021), Kilinc and Beringer (2007), Kotroni and Lagouvardos (2008)). All covariates are centered and scaled before including them in the model. Their inclusion means that we interpret the Gaussian field as the effect of environmental factors on lightning intensity after controlling for cloud growth, cloud-top height, and elevation information. Figure 2 provides visuals of the lightning strikes and several covariates.

Figure 3 shows the estimated intensity per one km^2/hour returned by INLA, VL, and SLEM. INLA appears to estimate a finer scale of lightning activity than is actually present in the observed counts. The spurious locations of activity are especially troubling, as they seem associated with relatively large intensity values, as evidenced by the white spaces in Figure 3B which indicate values exceeding the plotting range. Not only does INLA estimate activity where it is not present, it estimates a large amount of activity there. The VL algorithm returns possibly overly smooth estimates of the intensity, washing out isolated areas of activity, and grouping more closely occurring ones together. The proposed SLEM approach provides a middle-ground between INLA and VL. SLEM's estimated intensity captures fine-scale lightning patterns like INLA but without introducing the same spurious activity. SLEM also provides recognition of larger areas of activity like VL but does so with less smoothing. In the sections that follow, we define the SLEM method and describe each of its components in detail. We then perform in-depth studies of SLEM and VL applied to simulated and lightning datasets, focusing on VL as the most competitive method based on preliminary results.

2. Methodology.

2.1. *EM applied to LGCP.* Throughout our analysis and to improve convergence of our algorithm below, we perform a change of variables

$$W(s) = \mathbf{X}(s)\boldsymbol{\beta} + Z(s) \tag{5}$$

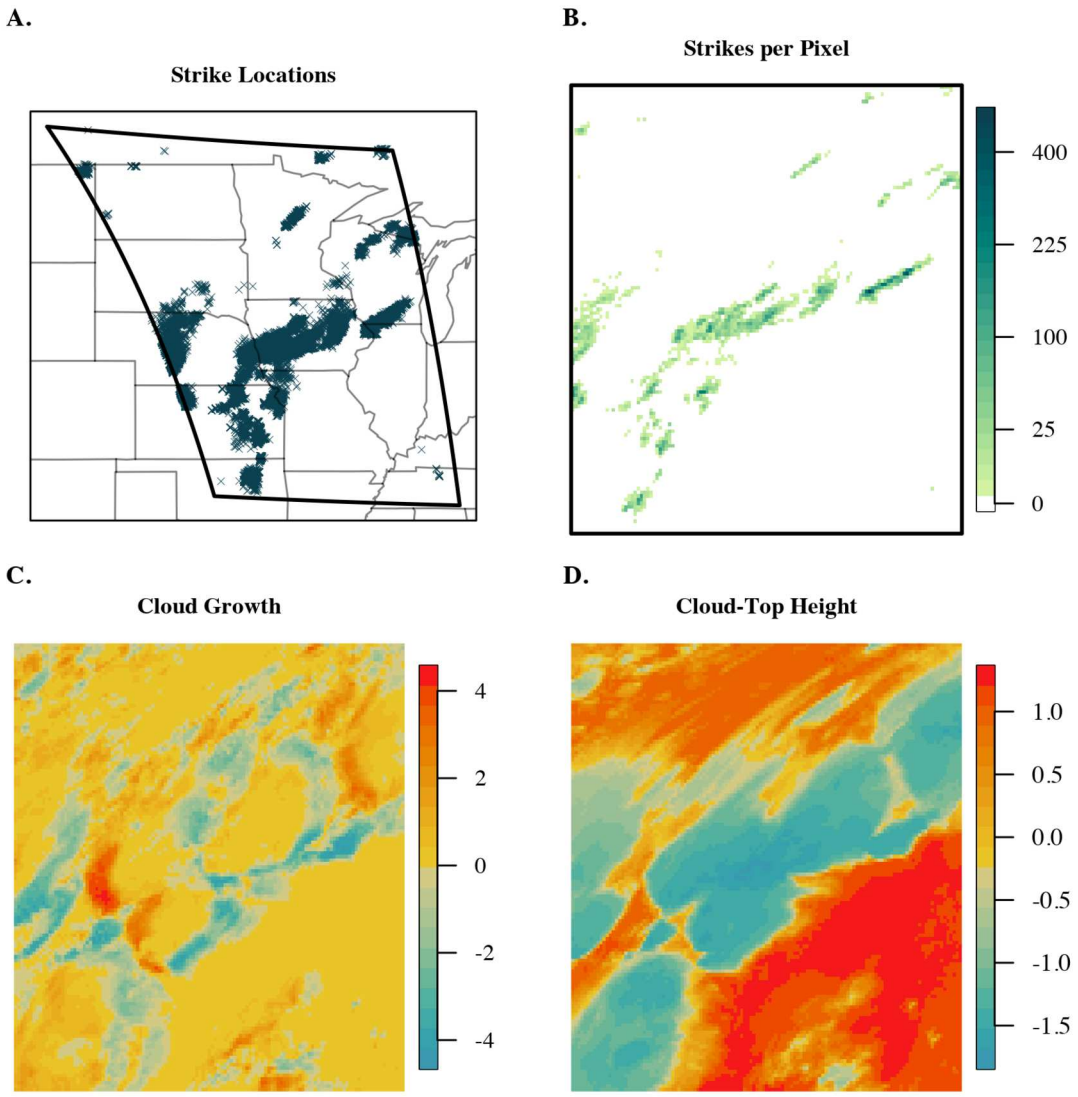

FIG. 2. A. *Individual lightning strikes recorded from* 01:00–01:59 *GMT on* 2018-07-01 *in designated area of North America*. B. *Lightning strikes from A. converted to counts per pixel on* 125 × 125 *grid, colors on square-root scale*. C, D. *Cloud growth and cloud-top height proxy data, respectively, averaged over same time frame as A and grid as B. Covariates are centered and scaled.*

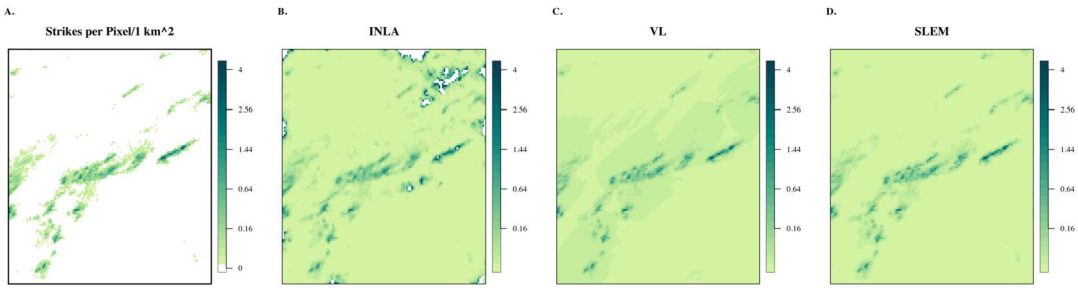

FIG. 3. *Counts and intensity per one km*^{2}*/hour estimates returned by INLA* (153 *min. CPU time*), *VL* (523 *min.*), *and SLEM* (549 *min.*) *for lightning dataset* 1 *shown in Figure* 2. *We employ a square-root transformation for better image definition. White spaces in B and C reflect estimates exceeding indicated range.*

which results in the likelihood function

$$L(\boldsymbol{\theta};\mathbf{Y}) = p(\mathbf{Y}|\boldsymbol{\theta}) = \int p(\mathbf{Y},\mathbf{W}|\boldsymbol{\theta})\,d\mathbf{W} = \int p(\mathbf{Y}|\mathbf{W},\boldsymbol{\theta})p(\mathbf{W}|\boldsymbol{\theta})\,d\mathbf{W}.$$

We obtain a value $\boldsymbol{\theta}$ that approximately maximizes the likelihood function iteratively using an approximate EM algorithm. Let $\boldsymbol{\theta}^{(t)}$ refer to the value of the parameters at iteration t. At iteration $t+1$, the EM algorithm finds a new value $\boldsymbol{\theta}^{(t+1)}$ by increasing the objective function,

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}\big[\log\{p(\mathbf{Y}|\mathbf{W},\boldsymbol{\theta})\}\big] + E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}\big[\log\{p(\mathbf{W}|\boldsymbol{\theta})\}\big] \\ &= E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}\left[\sum_{j=1}^{n} Y_j\big[\log(\Delta_j) + W_j\big] - \Delta_j \exp(W_j) - \log(Y_j!)\right] \\ &\quad + E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}\left[-\frac{1}{2}\big[\log(|\Sigma_{\boldsymbol{\eta}}|) + (\mathbf{W}-X\boldsymbol{\beta})^T\Sigma_{\boldsymbol{\eta}}^{-1}(\mathbf{W}-X\boldsymbol{\beta}) + n\log(2\pi)\big]\right].\end{aligned}$$

In our alternative parameterization, $E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}(W_j)$ and $E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}(\exp(W_j))$ depend on $\boldsymbol{\theta}^{(t)}$ but not $\boldsymbol{\theta}$, so the objective function simplifies to

$$\begin{aligned}Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = &-\frac{1}{2}\big[\log(|\Sigma_{\boldsymbol{\eta}}|) + \big(E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}] - X\boldsymbol{\beta}\big)^T\Sigma_{\boldsymbol{\eta}}^{-1}\big(E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}] - X\boldsymbol{\beta}\big) \\ &+ \operatorname{tr}\big(\Sigma_{\boldsymbol{\eta}}^{-1}E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}\big[\big(\mathbf{W} - E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}]\big)\big(\mathbf{W} - E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}]\big)^T\big]\big)\big] + c,\end{aligned}$$

where c contains terms that do not depend on $\boldsymbol{\theta}$. Evaluating the objective function is computationally challenging and requires a novel approach. The next several subsections detail how we perform computations and approximate this objective function in SLEM.

2.2. *Circulant covariance assumption.* The model includes random effects $\mathbf{Z}$ with spatial covariance matrix $\Sigma_{\boldsymbol{\eta}}$, whose entries are determined by covariance function $K(\boldsymbol{\eta})$. We assume that $K(\boldsymbol{\eta})$ is the circulant version of the quasi-Matérn covariance function with variance and range parameters σ and α, as presented in Guinness and Fuentes (2017). Letting $\boldsymbol{\omega}$ refer to the Fourier frequencies associated with the spatial grid, the quasi-Matérn covariance function is defined as

$$\operatorname{Cov}\big(Z(\boldsymbol{s}), Z(\boldsymbol{s}+\boldsymbol{h})\big) = \frac{1}{n}\sum_{\boldsymbol{\omega}}\left(\sigma^2\left(1+\alpha^2\sin^2\left(\frac{\omega_1}{2}\right)+\alpha^2\sin^2\left(\frac{\omega_2}{2}\right)\right)^{-2}\right)e^{i\boldsymbol{\omega}\cdot\boldsymbol{h}}\,d\boldsymbol{\omega}.$$

Given this spectral representation, $\boldsymbol{\eta} = (\sigma^2,\alpha)$, and $\Sigma_{\boldsymbol{\eta}}$ is block circulant. The circulant assumption on $\Sigma_{\boldsymbol{\eta}}$ simplifies evaluation of $\log(|\Sigma_{\boldsymbol{\eta}}|)$, as the log determinant of a circulant matrix is equal to the sum of the log-spectral density, evaluated at the Fourier frequencies, which can be computed with $O(n)$ memory and in $O(n)$ time for n spatial grid points. The circulant assumption also allows for fast matrix-vector multiplications $\Sigma_{\boldsymbol{\eta}}^{-1}\boldsymbol{v}$ which are leveraged for evaluating other terms in the objective function. These can be computed with $O(n)$ memory and $O(n\log n)$ time via fast Fourier transform algorithms.

2.3. *Laplace approximation.* As described in Section 1, it is common to approximate the distribution $p(\mathbf{W}|\mathbf{Y},\boldsymbol{\theta})$ with a Gaussian distribution. The Laplace approximation is obtained by performing a second-order Taylor series expansion about the mode of $\log\{p(\mathbf{W}|\mathbf{Y},\boldsymbol{\theta})\}$, resulting in a Gaussian approximation with mean equal to the mode and precision matrix equal to the Hessian at the mode. The Supplementary Material (Gelsinger et al. (2023)) provides more details on Laplace approximations.

We obtain a Laplace approximation of the form,

$$\mathbf{W}|\mathbf{Y}, \boldsymbol{\theta}^{(t)} \sim N\big(\mathbf{W}^{(t)}, (\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C_{\mathbf{W}^{(t)}})^{-1}\big),$$

where $\mathbf{W}^{(t)}$ is the mode of $p(\mathbf{W}|\mathbf{Y}, \boldsymbol{\theta}^{(t)})$, $C_{\mathbf{W}^{(t)}}$ is a diagonal matrix with diagonal entries $\boldsymbol{\Delta} \circ \exp(\mathbf{W}^{(t)})$, $\boldsymbol{\Delta}$ represents the $n \times 1$ vector of pixel areas, and $\circ$ is elementwise multiplication. To obtain $\mathbf{W}^{(t)}$, we select a starting value $\mathbf{W}_0^{(t)} = \mathbf{W}^{(t-1)}$ and then iterate as

$$\mathbf{W}_{\ell+1}^{(t)} = \mathbf{W}_{\ell}^{(t)} + (\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C_{\mathbf{W}_{\ell}^{(t)}})^{-1}(\mathbf{Y} - \boldsymbol{\Delta} \circ \exp(\mathbf{W}_{\ell}^{(t)}) - \Sigma_{\boldsymbol{\eta}^{(t)}}^{-1}(\mathbf{W}_{\ell}^{(t)} - X\boldsymbol{\beta}^{(t)})) \tag{6}$$

which corresponds to performing Newton–Raphson updates. Convergence is determined by $n^{-1/2}\|\mathbf{W}_{\ell+1}^{(t)} - \mathbf{W}_{\ell}^{(t)}\| < \epsilon$, where we use $\epsilon = 1 \times 10^{-3}$. We set $\mathbf{W}^{(t)}$ equal to the converged value.

We use preconditioned conjugate gradient (PCG) to solve the system of equations required to evaluate (6) (Hestenes, Stiefel et al. (1952)). PCG is a standard algorithm for solving positive definite systems. In this case PCG relies on successive matrix-vector multiplication involving $\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C_{\mathbf{W}_{\ell}^{(t)}}$ and a preconditioning matrix. This is computationally efficient because $\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1}$ is block circulant, so the multiplication can be done with FFTs, and $C_{\mathbf{W}_{\ell}^{(t)}}$ and our preconditioning matrix are diagonal. The PCG algorithm is included in Algorithm 1 of the Supplementary Material. Having obtained the Laplace approximation, we approximate the posterior mean $E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}]$ with $\mathbf{W}^{(t)}$ and the posterior variance $E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[(\mathbf{W} - E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}])(\mathbf{W} - E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}])^T]$ with $(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C_{\mathbf{W}^{(t)}})^{-1}$.

2.4. *Hutchinson trace approximation.* The Laplace approximation allows us to replace $E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[(\mathbf{W} - E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}])(\mathbf{W} - E_{\mathbf{W}|\mathbf{Y},\boldsymbol{\theta}^{(t)}}[\mathbf{W}])^T]$ with $(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)})^{-1}$. Because $\operatorname{tr}(\Sigma_{\boldsymbol{\eta}}^{-1}(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)})^{-1})$ remains challenging to evaluate, we use the Hutchinson trace approximation (HTA) (Hutchinson (1989)). HTA is a technique for calculating the trace when a matrix A is too hard to compute, but performing matrix-vector multiplication, $\mathbf{v}^T A\mathbf{v}$, is feasible. In this context, $A = \Sigma_{\boldsymbol{\eta}}^{-1}(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)})^{-1}$. Feasibility of evaluating $\mathbf{v}^T A\mathbf{v}$ is a consequence of the assumed circulant structure of $\Sigma_{\boldsymbol{\eta}}^{-1}$ and the relationship between evaluation of $(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)})^{-1}\mathbf{v}$ and evaluation of the Newton–Raphson updates used to compute $\mathbf{W}^{(t)}$. For $M \geq 1$ random vectors $\mathbf{v}_i$ with independent, identically distributed Rademacher distributed elements, HTA approximates $\operatorname{tr}(\Sigma_{\boldsymbol{\eta}}^{-1}(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)})^{-1})$ with

$$\frac{1}{M}\sum_{i=1}^{M} \mathbf{v}_i^T\big(\Sigma_{\boldsymbol{\eta}}^{-1}(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)})^{-1}\big)\mathbf{v}_i.$$

For each $\mathbf{v}_i$ we use the PCG algorithm described in Section 2.3 to quickly solve $(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)})\mathbf{r}_i = \mathbf{v}_i$. We then leverage FFTs to efficiently evaluate $\frac{1}{M}\sum_{i=1}^{M} \mathbf{v}_i^T \Sigma_{\boldsymbol{\eta}}^{-1}\mathbf{r}_i$. The choice of M controls the tradeoff between speed and accuracy.

2.5. *Defining and increasing the approximate objective function.* We define an approximate objective function $\widetilde{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}; M)$, via combining the Laplace and Hutchinson trace approximations,

$$\begin{aligned}\widetilde{Q}(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}; M) = -\frac{1}{2}\Bigg[&\log(|\Sigma_{\boldsymbol{\eta}}|) + (\mathbf{W}^{(t)} - X\boldsymbol{\beta})^T \Sigma_{\boldsymbol{\eta}}^{-1}(\mathbf{W}^{(t)} - X\boldsymbol{\beta}) \\ &+ \frac{1}{M}\sum_{i=1}^{M} \mathbf{v}_i^T\big(\Sigma_{\boldsymbol{\eta}}^{-1}(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)})^{-1}\big)\mathbf{v}_i\Bigg].\end{aligned} \tag{7}$$

We consider the problem of finding a new value $\boldsymbol{\theta}^{(t+1)}$ that satisfies $\widetilde{Q}(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}; M) > \widetilde{Q}(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}; M)$. First, we set

$$\boldsymbol{\beta}^{(t+1)} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} (\mathbf{W}^{(t)} - X\boldsymbol{\beta})^T \Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} (\mathbf{W}^{(t)} - X\boldsymbol{\beta}).$$

This is equivalent to computing the regression coefficients for a regression of the mode $\mathbf{W}^{(t)}$ on the predictors with error covariance $\Sigma_{\boldsymbol{\eta}^{(t)}}$.

Next, we set

$$\boldsymbol{\eta}^{(t+1)} = \underset{\boldsymbol{\eta}}{\operatorname{argmax}} -\frac{1}{2}\Bigg[\log(|\Sigma_{\boldsymbol{\eta}}|) + (\mathbf{W}^{(t)} - X\boldsymbol{\beta}^{(t+1)})^T \Sigma_{\boldsymbol{\eta}}^{-1} (\mathbf{W}^{(t)} - X\boldsymbol{\beta}^{(t+1)}) + \frac{1}{M}\sum_{i=1}^{M} \mathbf{v}_i^T \big(\Sigma_{\boldsymbol{\eta}}^{-1} \big(\Sigma_{\boldsymbol{\eta}^{(t)}}^{-1} + C^{(t)}\big)^{-1}\big)\mathbf{v}_i\Bigg].$$

This yields a new value $\boldsymbol{\theta}^{(t+1)} = (\boldsymbol{\beta}^{(t+1)}, \boldsymbol{\eta}^{(t+1)})$ that satisfies $\widetilde{Q}(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}; M) \geq \widetilde{Q}(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}; M)$.

2.6. *Practical implementation details*. We iterate between the E- and M-steps until we reach convergence. Our convergence criterion is

$$\sqrt{\frac{1}{p+3}\sum_{i=1}^{p+3} \big(\boldsymbol{\theta}_i^{(t+1)} - \boldsymbol{\theta}_i^{(t)}\big)^2} < \epsilon,$$

where $p+3$ corresponds to the length of $\boldsymbol{\theta}$ and $\epsilon = 10^{-5}$. We measure convergence in terms of absolute changes in $\boldsymbol{\theta}$; if one were interested in relative changes, $\boldsymbol{\theta}_i$ could be replaced with $\log \boldsymbol{\theta}_i$ in the convergence criterion. Given the approximations throughout this method, convergence is not guaranteed. However, our studies suggest that iterating through about 100 EM steps provides reasonable results; aside from one of the fits to the lightning data—which converged in 109 iterations—all of the simulation and lightning fits converged in fewer than 100 iterations.

Implementation of this EM algorithm requires specification of starting values $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)}, \boldsymbol{\eta}^{(0)})$. We recommend setting $\boldsymbol{\beta}^{(0)} = \mathbf{0}$ and $\boldsymbol{\eta}^{(0)} = \widetilde{\boldsymbol{\eta}}^*$, where $\widetilde{\boldsymbol{\eta}}^*$ refers to the EM estimate of variance parameters based on assuming an LGCP model with no predictors for the same data. For EM estimation of the variance parameters under an LGCP model with no predictors, we recommend the initial value $\widetilde{\boldsymbol{\theta}}^{(0)} = (\widetilde{\boldsymbol{\eta}}^{(0)}) = (\bar{Y}, n_1/4)$.

We refer to the implementation scheme described above as the "joint" implementation because we update both $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ during each M-step. We also consider the alternative method of fixing $\boldsymbol{\beta}$ at the generalized least squares estimate and updating only $\boldsymbol{\eta}$ at each M-step. We will refer to our implementation of this alternative scheme as the "fixed" case.

All computations are done in R on a machine with eight cores (Intel Xeon W-2145 CPU @ 3.70 GHz) and 16 GB RAM.

2.7. *Recovery of the residual latent field*. Having obtained an optimal value of the parameters $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^*, \boldsymbol{\eta}^*)$, we can recover $\mathbf{Z}^*$, the posterior mode of the latent field $\mathbf{Z}$, as defined in the original stochastic representation of the LGCP model, from $\mathbf{W}^*$, the posterior mode of $\mathbf{W}$ at $\boldsymbol{\theta}^*$, by setting $\mathbf{Z}^* = \mathbf{W}^* - X\boldsymbol{\beta}^*$. A detailed derivation is provided in the Supplementary Material.

In practice, it can also be of interest to approximate the posterior mean of the latent field on the intensity scale, $E[\exp(\mathbf{Z})|\mathbf{Y}, \boldsymbol{\theta}^*]$, as opposed to the log scale. Again, we use a Laplace

approximation to the posterior distribution to approximate this expectation. The Laplace approximation to the posterior distribution of $\mathbf{Z}$, given $\mathbf{Y}$ and $\boldsymbol{\theta}^*$, is

$$\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^* \sim N(\mathbf{Z}^*, \boldsymbol{\Psi}^{-1}),$$

where $\boldsymbol{\Psi} = \Sigma_{\boldsymbol{\eta}^*}^{-1} + \operatorname{diag}(\boldsymbol{\Delta} \circ \exp(X\boldsymbol{\beta}^* + \mathbf{Z}^*))$. Given the Laplace approximation for $\mathbf{Z}|\mathbf{Y}, \boldsymbol{\theta}^*$, it follows that $\exp(\mathbf{Z})|\mathbf{Y}, \boldsymbol{\theta}^*$ has a multivariate log-normal distribution with

$$E\big[\exp(Z_j)|\mathbf{Y}, \boldsymbol{\theta}^*\big] \approx \exp\left(Z_j^* + \frac{1}{2}\boldsymbol{\Psi}_{jj}^{-1}\right).$$

It is too computationally expensive to invert the dense $n \times n$ matrix $\boldsymbol{\Psi}$ and extract the diagonal elements. Instead, we propose a local approximation. For each j we extract the entries of $\boldsymbol{\Psi}$ corresponding to the $k \times k$ square neighborhood of pixels surrounding pixel j. We then invert the submatrix containing these entries and extract the diagonal entry of the inverse corresponding to pixel j as our approximation to $\boldsymbol{\Psi}_{jj}^{-1}$. Note that the circulant covariance structure has an implied assumption that $\mathbf{Z}$ is dependent across opposite boundaries of the domain. Likewise, constructing neighborhood submatrices for pixels along one edge of the spatial domain involves incorporating pixels from the other edge of the spatial domain. Like the value of M used to construct the HTA, smaller values of k yield faster but less accurate, approximations to $E[\exp(\mathbf{Z})|\mathbf{Y}, \boldsymbol{\theta}^*]$.

3. Simulation study. Because our initial exploration of the data in Section 1.1 suggested that VL is the most competitive alternative method, we focus on SLEM and VL in simulations. For each method we implement both the fixed and joint implementations suggested in Section 2.6. For SLEM we vary M, the number of vectors in the HTA. We compare average runtime, estimates of $\boldsymbol{\beta}$, and average root-mean-square-error of the log-intensity $\log(\boldsymbol{\lambda})$ across 100 simulation replicates of $\mathbf{Y}$, corresponding to a 70×70 grid.

Each simulation replicate uses the same $\mathbf{Z}$ and thus the same intensity. We simulate $\mathbf{Z}$ from a multivariate normal distribution with zero mean and Matérn covariance with variance $\sigma^2 = 2$, range $\alpha = 18$, and smoothness $\nu = 1$. We define $\boldsymbol{X}(s_i)$ to include an intercept and several covariates and consider two different settings for covariate construction. The first covariate setting produces a noisy true intensity using three covariates, where two covariates are simulated from a standard normal distribution and another is constructed from raw Channel 5 ABI data. The second covariate setting produces a smooth true intensity using two covariates, where the two covariates are constructed from raw Channel 5 and Channel 8 ABI data, respectively. We refer to the first noisy true intensity setting as "Setting 1" and the second smooth true intensity setting as "Setting 2."

The results for Setting 1 are summarized in Table 1 and Figure 4. In SLEM we see that increasing M affects runtime but not estimation of $\boldsymbol{\beta}$ or the log-intensity. In general, the joint settings are slower but more accurate in terms of RMSE than their fixed counterparts. SLEM-joint is more accurate than VL-joint for estimating β_0, but VL-joint is more accurate than SLEM-joint for estimating β_3 and the log-intensity. However, VL-joint takes three times as long as SLEM-joint with $M = 1$ to run. For SLEM fixed, all fits converged in roughly 40 iterations. All SLEM joint fits converged in about 60 iterations. A more detailed description of the number of iterations needed for convergence in simulations is provided in the Supplementary Material. An additional single simulation for a larger 200×200 grid, described in the Supplementary Material, suggests that the speed advantage of SLEM over VL persists for larger grids.

Figure 4 shows that all of the methods produce visually similar intensity estimates. SLEM-joint sacrifices some accuracy along the boundary of the domain which is not surprising since

TABLE 1

Results of applying SLEM and VL to Setting 1. *True* $\boldsymbol{\beta} = 1, 0.85, 0.6, 0.95$. *Parameter estimates reported as Mean* (*RMSE*), *where RMSE stands for "root-mean-square-error" throughout. Runtime and iterations to convergence presented as average over* 100 *trials. RMSE*($\log \lambda$) *reported as "full grid|interior points," where the interior point calculation is restricted to to pixels* $3 : (n_1 - 2) \times 3 : (n_1 - 2)$, *that is, two pixels in from the edge of the spatial domain*

Method	Update	M	Time (min.)	Iter	β_0	β_1	β_2	β_3	RMSE(log(λ))
SLEM	fixed	1	3.71	39	1.98 (0.98)	0.8 (0.05)	0.6 (<0.01)	0.68 (0.27)	0.279|0.19
SLEM	fixed	10	21.24	40	1.98 (0.98)	0.8 (0.05)	0.6 (<0.01)	0.68 (0.27)	0.279|0.19
SLEM	joint	1	14.21	59	1.05 (0.05)	0.85 (<0.01)	0.6 (<0.01)	0.88 (0.07)	0.269|0.171
SLEM	joint	10	71.43	59	1.05 (0.05)	0.85 (<0.01)	0.6 (<0.01)	0.88 (0.07)	0.269|0.171
VL	fixed	–	5.26	–	1.98 (0.98)	0.8 (0.05)	0.6 (<0.01)	0.68 (0.27)	0.178|0.178
VL	joint	–	49.46	–	1.13 (0.15)	0.85 (0.01)	0.6 (<0.01)	0.94 (0.01)	0.137|0.136

circulant covariance methods are known to suffer from edge effects. This is also reflected in Table 1 which includes RMSE for the log-intensity restricted to interior points. The accuracy advantage of VL for estimating the log-intensity shrinks when only interior points are considered, although the VL-joint estimates remain superior. The same conclusions are echoed in the analysis of Setting 2 which has a smoother intensity field. Full details are described in the Supplementary Material. The SLEM fits took more iterations to converge in Setting 2, between 80 and 90 iterations for each fit. A more detailed description of the number of iterations needed for convergence in simulations is provided in the Supplementary Material. The same conclusions are echoed once more in an additional set of simulations, allowing for heavier-than-normal tailed $\mathbf{Z}$, which are summarized in the Supplementary Material.

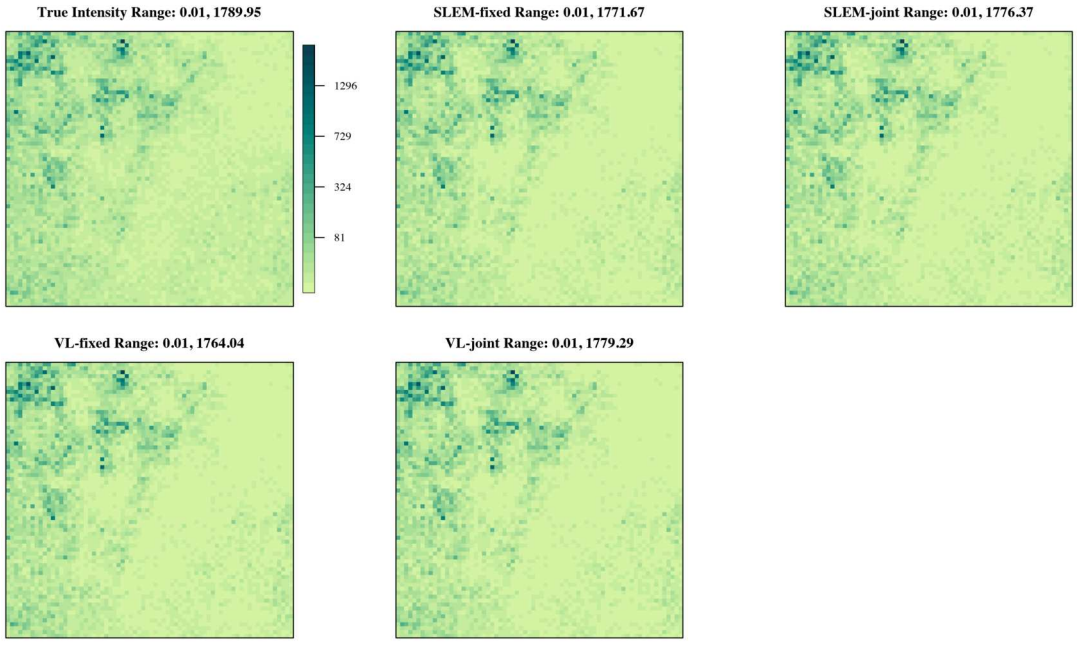

FIG. 4. *Examples of intensities resulting from applying SLEM with* $M = 1$ *and VL to Setting* 1.

4. Lightning data.

4.1. *Covariate construction and selection.* For the lightning data, we define $\boldsymbol{X}(s_i)$ to include an intercept, two covariates derived from ABI data, and a third covariate measuring elevation. All covariates are centered and scaled to facilitate comparison of parameter estimates. The first two covariates (X_1 and X_2) are constructed from brightness temperature from ABI Channel 13 which serves as a proxy for cloud-top temperature (Henderson, Otkin and Mecikalski (2021)). Channel 13 measures light at an infrared wavelength (10.3 μm) which ensures continual measures throughout day and night. Meteorologists use cloud-top temperature, which is inversely related with cloud-top height, to monitor updrafts in severe convective storms (Mecikalski and Bedka (2006)). Some care must be taken to construct covariates capable of connecting the minute-by-minute ABI data to lightning counts, which we have aggregated over one-hour periods. Simple averaging of the Channel 13 data over one hour may not capture the dynamic and transitory nature of clouds in severe storms. Lee, Kummerow and Zupanski (2021) suggested constructing variables based on 10-minute intervals of data. Moreover, both the absolute cloud heights and sharp changes in cloud heights are important factors impacting the severity of storms.

For these reasons we construct X_1 and X_2 as follows. For pixel i and minute t, let $A_{i,t}$ be the Channel 13 brightness temperature. Letting $k = 1, \ldots, 6$, we define six 10-minute proxies for differenced and absolute cloud top temperatures for one hour of data as follows:

$$X_{1,k} = A_{i,10k} - A_{i,10k-9} \quad \text{(10-minute differences)},$$

$$X_{2,k} = \frac{1}{10}\sum_{j=1}^{10} A_{i,10k+j} \quad \text{(10-minute averages)}.$$

The differences are designed to capture changes in cloud top height, whereas the averages measure absolute cloud top height. We next consider several functions of these six values: the average, minimum, maximum, and range. For both differences and averages, the best function of the six values is determined by selecting the function which produces the highest log-likelihood of a simple Poisson regression with a single covariate. Through this process we select the average of the 10-minute differences and the minimum of the 10-minute averages as our covariates,

$$X_1 = \frac{1}{6}\sum_{k=1}^{6} X_{1,k},$$

$$X_2 = \min\{X_{2,1}, \ldots, X_{2,6}\}.$$

Note that the inverse relationship between cloud-top temperature and cloud-top height allows us to interpret the minimum of Channel 13 brightness temperatures (X_2) as a proxy for the maximum cloud-top height. We interpret the average of Channel 13 brightness temperature differences (X_1) as a proxy for cloud growth (Henderson, Otkin and Mecikalski (2021)).

We also include elevation (X_3) as an environmental factor in our model, because a connection between lightning and elevation has been hypothesized in other parts of the world (Kilinc and Beringer (2007), Kotroni and Lagouvardos (2008)). Elevation data is available from the ETOPO5 data repository which contains land and sea-floor elevation at an approximate eight km resolution over the United States (NOAA). We assemble the 125×125 grid of elevation data in each hour by matching each pixel in the 125×125 grid used for the lightning and ABI data to the locations in the ETOPO5 dataset.

TABLE 2
Results for Lightning Dataset 1

Method	Update	Time (min.)	Iter	Log Score	β_0	β_1 (avg diff)	β_2 (min avg)	β_3 (elev)
SLEM	fixed	225	109	−1852	−7.72	−0.29	−3.39	0.10
SLEM	joint	549	71	−1853	−10.06	−0.12	−1.13	0.40
VL	fixed	154	–	−2456	−7.72	−0.29	−3.39	0.10
VL	joint	523	–	−3305	−6.59	−0.10	−0.92	0.00

4.2. *Analysis*. We now analyze the two lightning datasets shown in Figure 1 that motivated this work. Both datasets are from July 2018, but they differ in several ways. The first is an early-evening storm over the upper midwest U.S. (01:00 GMT is 17:00 CST). The second is an early afternoon storm over the southwest U.S. (20:00 GMT is 12:00 PST). The midwest clouds are larger and have lightning strikes that are more clustered, whereas the strike activity in the southwest is more scattered. The terrain of these regions also differs in that the southwest is generally drier and more mountainous. For these reasons we fit separate LGCP models to the two datasets, since we expect the lightning dynamics to differ with respect to both the random term and with respect to the covariates. We fit both the fixed and joint implementations. For VL we used a Matérn covariance with smoothness $\nu = 1$, which outperformed $\nu = 1/2$ and $\nu = 3/2$, as shown in the Supplementary Material.

To facilitate out-of-sample comparisons, we fit each model to a random subset of 90% of the strikes and test the resultant model on the remaining 10% of the strikes. Note that we subsample the strikes, not the pixels, so we always work on a full grid with no missing values. We record the log-score or log-likelihood value associated with the testing data, given a 10/90 scaling of the fitted intensity. This corresponds to evaluating

$$\text{log-score} = \sum_{i=1}^{n} Y_i * \big(\log \Delta_i + \log \lambda_i + \log(1/9)\big) - \big((1/9) * \Delta_i * \lambda_i\big) - \log Y_i!$$

on the testing data values Y_i and estimated intensities λ_i from the training data. We also offer visuals of the estimated intensity functions for qualitative comparisons.

As seen in Tables 2 and 3, SLEM has larger out-of-sample log-scores on test data, indicating a superior model fit. VL's lower log-scores are likely a result of the overly smooth estimates of the intensities, as seen in Figures 5 and 6, which fail to accurately reflect the isolated regions of lightning activity present in the data. In contrast, SLEM appears, both quantitatively and qualitatively, to capture both the isolated regions of lightning activity and those larger areas with sufficient detail, especially along the interior points which are not affected by edge effects. On Dataset 1, SLEM and VL take roughly the same amount of time; however, the SLEM fit is far superior in terms of log score. On Dataset 2, SLEM joint runs about four times faster than VL fixed and about 10 times faster than VL joint, in addition to

TABLE 3
Results for Lightning Dataset 2

Method	Update	Time (min.)	Iter	Log Score	β_0	β_1 (avg diff)	β_2 (min avg)	β_3 (elev)
SLEM	fixed	111	84	−1993	−5.58	−0.15	−1.01	0.29
SLEM	joint	35	53	−2004	−8.46	0.10	−0.75	0.55
VL	fixed	155	–	−2682	−5.58	−0.15	−1.01	0.29
VL	joint	372	–	−2844	−5.21	0.08	−0.41	0.36

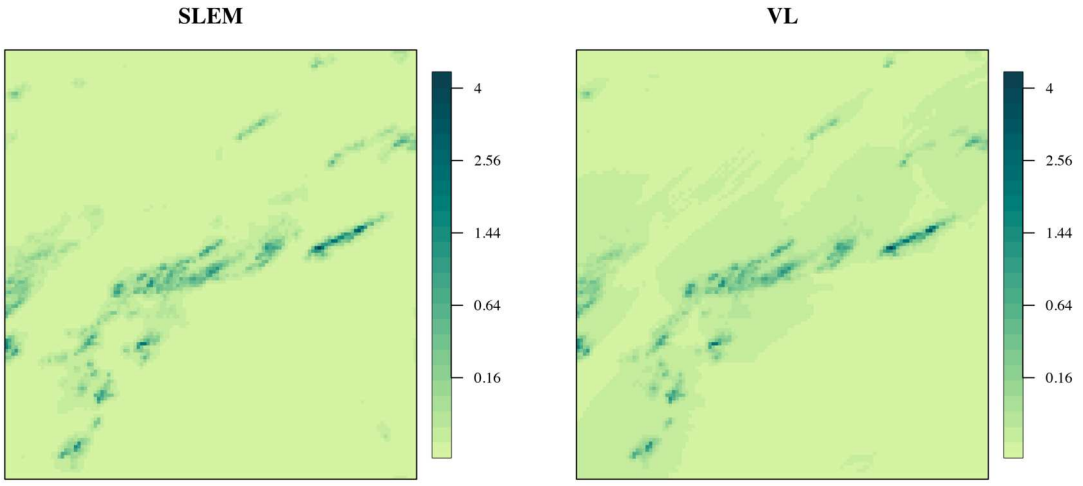

FIG. 5. *Estimated intensities per one km^2 per hour resulting from the SLEM and VL algorithms applied to lightning Dataset* 1.

having a superior log score. We also see that SLEM converges in 71 iterations for Dataset 1 and 53 iterations for Dataset 2. Though not definitive proof, these results, in combination with those from the simulation study, suggest that stronger spatial correlations, represented by more strongly clustered lightning strikes, leads to slower convergence.

Turning our attention to the estimated regression coefficients, we see that the minimum of the 10-minute averages of brightness temperatures has the strongest effect (recall that the covariates are normalized). The estimates are similar for SLEM and VL joint and are negative for both datasets, which is expected, since the minimum of the average brightness temperatures is negatively associated with the maximum of the average cloud-top heights, and high cloud tops indicate severe weather (Mecikalski and Bedka (2006)). Elevation has a smaller but positive effect that is consistent with the relevant literature (Kilinc and Beringer (2007), Kotroni and Lagouvardos (2008)). The weakest effect is the average of the 10-minute differences, which mostly have small negative effects. Negative effects are also expected, since negative differences correspond to cloud growth (Mecikalski and Bedka (2006)). Plots of es-

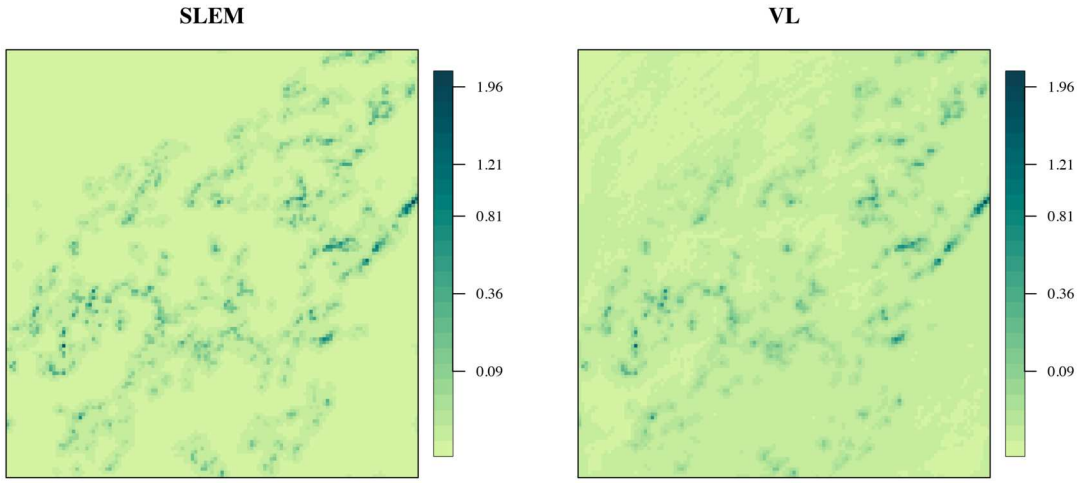

FIG. 6. *Estimated intensities per km^2 per hour resulting from the SLEM and VL algorithms applied to lightning Dataset* 2.

timated fixed effects $\mathbf{X}\boldsymbol{\beta}$ for Lightning Datasets 1 and 2 are provided in the Supplementary Material.

Lastly and importantly, we highlight the impact of jointly estimating the random effects and the regression coefficients simultaneously. In the "fixed" implementation, the regression coefficient estimates are equivalent to those from a generalized linear model with independent responses. In the "joint" implementation, we consider the spatially varying random effect when estimating the regression coefficients. In both datasets this has the impact of lessening the effect of the minimum average brightness temperature. This is perhaps not surprising because the joint implementation accounts for spatially continuous effects that are not attributable to the covariates, whereas the fixed implementation attributes as much variation as possible to the observed covariates. We give more credence to the "joint" estimates which attempt to account for unobserved effects and are more conservative.

5. Discussion. In this paper we introduced SLEM, a new approximate method for fitting LGCP models to large spatial point pattern datasets. This method leveraged spectral, Laplace, and Hutchinson trace approximations to make computational gains without sacrificing much accuracy. We verified this in a simulation study where SLEM ran much faster than VL and competitively estimated $\boldsymbol{\beta}$. Via simulations, we also showed SLEM is robust to the choice of M, the number of vectors in the HTA. When applied to the lightning data, SLEM was superior to its chief competitor, running as fast or faster and with more accuracy in both examples. Our analysis of the lightning data produces findings that are consistent with the meteorology literature. Specifically, we found that high and rapidly increasing cloud tops over high elevation regions are associated with more lightning strikes.

Although we developed SLEM in the context of the lightning data described in this paper and focus on lightning data throughout, SLEM could be applied more generally to large spatial point pattern datasets which are ubiquitous in a variety of fields including epidemiology and finance. SLEM acts as a new tool not only for researchers seeking to understand lightning dynamics but also for those looking to model and investigate large spatial point pattern datasets, in general.

We conclude by describing several potential extensions to our work. First, SLEM's computational advantages could be especially valuable for spatiotemporal data due to the typically large sizes of these datasets and complex relationships between space, time, and the covariates. Second, existing methods for reducing edge effects could be incorporated into SLEM. In particular, methods that embed the circulant covariance on a larger spatial domain could be adapted for use with SLEM in order to gain an edge-effect-free approximation of those covariances within the spatial domain of interest (Guinness and Fuentes (2017)). Additionally, existing methods for accelerating EM algorithms, for example, Varadhan and Roland (2008), Zhou, Alexander and Lange (2011), or Henderson and Varadhan (2019), could be incorporated into SLEM to improve computational efficiency. Furthermore, extensions of the LGCP model may provide better fits to GOES lightning data. We used separate models for the midwest and southwest datasets, but one might consider extending SLEM to fit a spatially varying coefficient (SVC) model which generalizes the LGCP model and allows all regression coefficients to vary within a dataset (Gelfand et al. (2003), Banerjee, Carlin and Gelfand (2014)). Letting $\gamma(s_i) = (\gamma_1(s_i), \ldots, \gamma_{p+1}(s_i))$ refer to a $(p+1) \times 1$ vector of spatially varying regression coefficients, the SVC model assumes

$$
\begin{aligned}
Y_i|\lambda &\overset{\text{indep}}{\sim} \text{Poisson}\left(\int_{B_i} \lambda(s)\,ds\right) = \text{Poisson}\{\Delta_i \lambda(s_i)\},\\
\lambda(s_i) &= \exp\{\mathbf{X}(s_i)\boldsymbol{\beta} + \mathbf{X}(s_i)\boldsymbol{\gamma}(s_i)\},\\
\gamma_j(s) &\sim GP\{0, K(\boldsymbol{\eta}_j)\} \quad \text{for } j = 1, \ldots, p+1.
\end{aligned}
$$

The LGCP model is obtained by defining $Z(s_i) = \gamma_1(s_i)$ and $\gamma_j(s_i) = 0$ for $j = 2, \ldots, p+1$. In this paper we focused on the LGCP model because of the already large computational challenges associated with fitting the simpler LGCP model.

Lastly, another natural extension would relax the assumption that intensities in the LGCP model follow a log-Gaussian distribution. The log transformation could be replaced with a different suitable transformation to allow for intensities with different distributions. Some modifications would have to be made to SLEM, for example the form of Q in the EM algorithm, and the entries of $C_{\mathbf{W}}$.

Acknowledgments. The authors would like to thank Finn Lindgren and Matthias Katzfuss for their aid in implementing INLA and VL, respectively.

Funding. The authors gratefully acknowledge financial support from the National Science Foundation 1455172, 1916208, 1934985, 1940124, 1940276, 1953088, and 2114143, USAID 7200AA18CA00014, National Institutes of Health R01ES027892, and Cornell Atkinson Center for Sustainability.

SUPPLEMENTARY MATERIAL

Supplementary details, tables, and figures (DOI: 10.1214/22-AOAS1708SUPP; .pdf). Additional computational details, results of additional simulation studies, further information about fitted models to lightning data.

REFERENCES

AICH, V., HOLZWORTH, R., GOODMAN, S., KULESHOV, Y., PRICE, C. and WILLIAMS, E. (2018). Lightning: A new essential climate variable. *Eos* **99**.

BACHL, F. E., LINDGREN, F., BORCHERS, D. L. and ILLIAN, J. B. (2019). Inlabru: An R package for Bayesian spatial modelling from ecological survey data. *Methods Ecol. Evol.* **10** 760–766.

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, Boca Raton.

BRIX, A. and DIGGLE, P. J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 823–841. MR1872069 https://doi.org/10.1111/1467-9868.00315

CLARK, S. K., WARD, D. S. and MAHOWALD, N. M. (2017). Parameterization-based uncertainty in future lightning flash density. *Geophys. Res. Lett.* **44** 2893–2901.

DIGGLE, P., ROWLINGSON, B. and SU, T. (2005). Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics* **16** 423–434. MR2147534 https://doi.org/10.1002/env.712

DOC NOAA, N. and NASA (2017). GOES R series product definition and users' guide. *Atmos. Res.* **3**.

FINNEY, D. L., DOHERTY, R. M., WILD, O., STEVENSON, D. S., MACKENZIE, I. A. and BLYTH, A. M. (2018). A projected decrease in lightning under climate change. *Nat. Clim. Change* **8** 210–213.

GELFAND, A. E., KIM, H.-J., SIRMANS, C. F. and BANERJEE, S. (2003). Spatial modeling with spatially varying coefficient processes. *J. Amer. Statist. Assoc.* **98** 387–396. MR1995715 https://doi.org/10.1198/016214503000170

GELSINGER, M. L, GRIFFIN, M., MATTESON, D. and GUINNESS, J. (2023). Supplement to "Log-Gaussian cox process modeling of large spatial lightning data using spectral and laplace approximations." https://doi.org/10.1214/22-AOAS1708SUPP

GONÇALVES, F. B. and GAMERMAN, D. (2018). Exact Bayesian inference in spatiotemporal Cox processes driven by multivariate Gaussian processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 157–175. MR3744716 https://doi.org/10.1111/rssb.12237

GOODMAN, S. J., BLAKESLEE, R. J., KOSHAK, W. J., MACH, D., BAILEY, J., BUECHLER, D., CAREY, L., SCHULTZ, C., BATEMAN, M. et al. (2013). The GOES-R geostationary lightning mapper (GLM). *Atmos. Res.* **125** 34–49.

GUAN, Y. and HARAN, M. (2018). A computationally efficient projection-based approach for spatial generalized linear mixed models. *J. Comput. Graph. Statist.* **27** 701–714. MR3890863 https://doi.org/10.1080/10618600.2018.1425625

GUAN, Y. and HARAN, M. (2020). Fast expectation-maximization algorithms for spatial generalized linear mixed models.

GUINNESS, J. and FUENTES, M. (2017). Circulant embedding of approximate covariances for inference from Gaussian data on large lattices. *J. Comput. Graph. Statist.* **26** 88–97. MR3610410 https://doi.org/10.1080/10618600.2016.1164534

HENDERSON, D. S., OTKIN, J. A. and MECIKALSKI, J. R. (2021). Evaluating convective initiation in high-resolution numerical weather prediction models using GOES-16 infrared brightness temperatures. *Mon. Weather Rev.* **149** 1153–1172.

HENDERSON, N. C. and VARADHAN, R. (2019). Damped Anderson acceleration with restarts and monotonicity control for accelerating EM and EM-like algorithms. *J. Comput. Graph. Statist.* **28** 834–846. MR4045852 https://doi.org/10.1080/10618600.2019.1594835

HESTENES, M. R., STIEFEL, E. et al. (1952). *Methods of Conjugate Gradients for Solving Linear Systems* **49**. NBS, Washington, DC.

HUTCHINSON, M. F. (1989). A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Comm. Statist. Simulation Comput.* **18** 1059–1076. MR1031840 https://doi.org/10.1080/03610918908812806

ILLIAN, J. B., SØRBYE, S. H. and RUE, H. (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Ann. Appl. Stat.* **6** 1499–1530. MR3058673 https://doi.org/10.1214/11-AOAS530

KATZFUSS, M., JUREK, M., ZILBER, D., GONG, W., GUINNESS, J., ZHANG, J. and SCHAEFER, F. (2021). GPvecchia: Scalable Gaussian-Process Computing. R package version 0.1.3.

KILINC, M. and BERINGER, J. (2007). The spatial and temporal distribution of lightning strikes and their relationship with vegetation type, elevation, and fire scars in the northern territory. *J. Climate* **20** 1161–1173.

KOTRONI, V. and LAGOUVARDOS, K. (2008). Lightning occurrence in relation with elevation, terrain slope, and vegetation cover in the Mediterranean. *J. Geophys. Res., Atmos.* **113**.

LEE, Y., KUMMEROW, C. D. and ZUPANSKI, M. (2021). A simplified method for the detection of convection using high-resolution imagery from GOES-16. *Atmos. Meas. Tech.* **14** 3755–3771.

LINDGREN, F., RUE, H. and LINDSTRÖM, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 423–498. MR2853727 https://doi.org/10.1111/j.1467-9868.2011.00777.x

MARTINS, T. G., SIMPSON, D., LINDGREN, F. and RUE, H. (2013). Bayesian computing with INLA: New features. *Comput. Statist. Data Anal.* **67** 68–83. MR3079584 https://doi.org/10.1016/j.csda.2013.04.014

MECIKALSKI, J. R. and BEDKA, K. M. (2006). Forecasting convective initiation by monitoring the evolution of moving cumulus in daytime GOES imagery. *Mon. Weather Rev.* **134** 49–78.

MØLLER, J., SYVERSVEEN, A. R. and WAAGEPETERSEN, R. P. (1998). Log Gaussian Cox processes. *Scand. J. Stat.* **25** 451–482. MR1650019 https://doi.org/10.1111/1467-9469.00115

NOAA, N. G. D. C. Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth.

PARK, J. and HARAN, M. (2021). Reduced-dimensional Monte Carlo maximum likelihood for latent Gaussian random field models. *J. Comput. Graph. Statist.* **30** 269–283. MR4270503 https://doi.org/10.1080/10618600.2020.1811106

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

SHIROTA, S. and GELFAND, A. E. (2016). Inference for log Gaussian Cox processes using an approximate marginal posterior. ArXiv preprint. Available at arXiv:1611.10359.

SYSTEM, G. C. O. (2016). The global observing system for climate: Implementation needs Technical Report No. 200 World Meteorological Organization Geneva, Switzerland.

TAYLOR, B. M. and DIGGLE, P. J. (2014). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *J. Stat. Comput. Simul.* **84** 2266–2284. MR3223624 https://doi.org/10.1080/00949655.2013.788653

TAYLOR, B. M., DAVIES, T. M., ROWLINGSON, B. S., DIGGLE, P. J. et al. (2013). lgcp: An R package for inference with spatial and spatio-temporal log-Gaussian Cox processes. *J. Stat. Softw.* **52** 1–40.

TAYLOR, B., DAVIES, T., ROWLINGSON, B. and DIGGLE, P. (2015). Bayesian inference and data augmentation schemes for spatial, spatiotemporal and multivariate log-Gaussian Cox processes in R. *J. Stat. Softw.* **63** 1–48.

VARADHAN, R. and ROLAND, C. (2008). Simple and globally convergent methods for accelerating the convergence of any EM algorithm. *Scand. J. Stat.* **35** 335–353. MR2418745 https://doi.org/10.1111/j.1467-9469.2007.00585.x

WOOD, A. T. A. and CHAN, G. (1994). Simulation of stationary Gaussian processes in $[0,1]^d$. *J. Comput. Graph. Statist.* **3** 409–432. MR1323050 https://doi.org/10.2307/1390903

ZHOU, H., ALEXANDER, D. and LANGE, K. (2011). A quasi-Newton acceleration for high-dimensional optimization algorithms. *Stat. Comput.* **21** 261–273. MR2774856 https://doi.org/10.1007/s11222-009-9166-3

ZILBER, D. and KATZFUSS, M. (2021). Vecchia–Laplace approximations of generalized Gaussian processes for big non-Gaussian spatial data. *Comput. Statist. Data Anal.* **153** 107081. MR4146817 https://doi.org/10.1016/j.csda.2020.107081