Supporting Information

Zammit-Mangion et al. 10.1073/pnas.1203177109

SI Text

Nonparametric Estimation of the Stochastic Integro-Difference Equation (SIDE) from Point Process Observations. Standard nonparametric methods may be easily extended to estimate unknown quantities in the SIDE. First, for $\upsilon = ||\boldsymbol{s} - \boldsymbol{r}||$, it can be shown that

$$\ln g_{k,k+1}(\upsilon) = [k_I * \ln g_{k,k}](\upsilon), \quad \textbf{[S1]}$$

where * is the convolution operator. Nonparametric estimators for the pair auto-correlation function (PACF) and the pair cross-correlation function (PCCF) are well known (*SI Text*), suggesting that the kernel $k_I(\upsilon)$ can be obtained through the *deconvolution* of Eq. **S1**. The problem may be seen as one of image restoration where the task is to recover an original image $k_I(\upsilon)$ from a degraded image $\ln g_{k,k+1}(\upsilon)$, for which standard image processing techniques such as the direct inverse filtering can be used. Further, one can show that

$$k_Q(\upsilon) = (\ln g_{k+1,k+1}(\upsilon) - [k_I * k_I * \ln g_{k,k}](\upsilon)). \quad \textbf{[S2]}$$

The proofs for Eqs. **S1** and **S2** are given below. For ease of the exposition we consider a one-dimensional domain $s, r \in \mathbb{R}$, $\upsilon = |s - r|$ with the intensity defined as $\lambda_k(s) = \exp(z_k(s))$ where $z_k(s)$ is a homogeneous Gaussian process with zero mean and covariance function $\sigma_k^2\psi_k(\upsilon)$. The results hold for a homogeneous and isotropic field in any dimension.

Proof for Eq. S1: Because $z_k(s)$ has zero mean, the following relationships hold:

$$\mathbb{E}[z_k(s) + [k_I * z_k](s + \upsilon)] = 0, \quad \textbf{[S3]}$$

$$\mathbb{E}[z_k(s)^2] = \sigma_k^2, \quad \textbf{[S4]}$$

$$\mathbb{E}[z_k(s)[k_I * z_k](s + \upsilon)] = \sigma_k^2[k_I * \psi_k](\upsilon), \quad \textbf{[S5]}$$

$$\mathbb{E}[[k_I * z_k](s)^2] = \sigma_k^2[k_I * k_I * \psi_k](0). \quad \textbf{[S6]}$$

By Eq. 1 and the assumption that $e_k(s)$ is uncorrelated with $z_k(s)$, the intensity cross second moment is given by

$$\begin{aligned}\lambda^{(2)}_{k,k+1}(\upsilon) &= \mathbb{E}[\lambda_k(s)\lambda_{k+1}(s+\upsilon)] \\ &= \mathbb{E}[\exp(z_k(s) + z_{k+1}(s+\upsilon))] \\ &= \mathbb{E}[\exp(z_k(s) + [k_I * z_k](s+\upsilon) + e_k(s+\upsilon))] \\ &= \exp\left(\frac{\sigma_k^2}{2} + \frac{\sigma_k^2[k_I * k_I * \psi_k](0)}{2} + \sigma_k^2[k_I * \psi_k](\upsilon) + \frac{k_Q(0)}{2}\right).\end{aligned} \quad \textbf{[S7]}$$

Because $\lambda_k = \exp(z_k(\boldsymbol{s}))$, the quantities $\lambda_k^{(1)}$ and $\lambda_{k+1}^{(1)}$ (through Eqs. **S4** and **S6**) are given as

$$\lambda_k^{(1)} = \exp(\sigma_k^2/2), \quad \textbf{[S8]}$$

$$\lambda_{k+1}^{(1)} = \exp\left(\frac{\sigma_k^2[k_I * k_I * \psi_k](0)}{2} + \frac{k_Q(0)}{2}\right), \quad \textbf{[S9]}$$

so that the log PCCF is given by

$$\ln g_{k,k+1}(\upsilon) = \sigma_k^2[k_I * \psi_k](\upsilon). \quad \textbf{[S10]}$$

But by Eq. 4 $\ln g_{k,k}(\upsilon) = \sigma_k^2\psi_k(\upsilon)$ to give Eq. **S1**.

Proof for Eq. S2: To obtain an expression for $k_Q(\upsilon)$, the PACF at subsequent time steps is considered. Again, considering zero mean stationarity of $z_k(s)$

$$\begin{aligned}\lambda^{(2)}_{k+1,k+1}(\upsilon) = \mathbb{E}[\exp([k_I * z_k](s) + [k_I * z_k](s+\upsilon) + e_k(s) \\ + e_k(s+\upsilon))],\end{aligned} \quad \textbf{[S11]}$$

which, after some lengthy algebraic manipulation, can be shown to be

$$\begin{aligned}\lambda^{(2)}_{k+1,k+1}(\upsilon) = \exp(k_Q(0) + k_Q(\upsilon) + \sigma_k^2[k_I * k_I * \psi_k](0) \\ + \sigma_k^2[k_I * k_I * \psi_k](\upsilon)).\end{aligned} \quad \textbf{[S12]}$$

Eq. **S2** subsequently follows from Eqs. 4 and **S9**.

Nonparametric Estimation of First-Order and Second-Order Point Process Statistics. Denote the spatial point process at k as $\mathscr{P}_k$. If $\mathscr{P}_k$ is first-order stationary, then an estimator for $\lambda_k^{(1)}$ is given as (1)

$$\lambda_k^{(1)} = \frac{N_k}{|\mathscr{O}|}, \quad \textbf{[S13]}$$

where N_k is the cardinality of $\mathscr{P}_k$. In some cases, this assumption does not hold and one may instead employ either standard linear regression methods to mark out clear intensity trends (2) or a standard nonparametric kernel estimator (3)

$$\lambda_k^{(1)}(\boldsymbol{s}) = \sum_{\boldsymbol{s}_i \in \mathscr{P}_k} \frac{k_b(||\boldsymbol{s} - \boldsymbol{s}_i||)}{c_{\mathscr{O},b}(\boldsymbol{s}_i)}. \quad \textbf{[S14]}$$

Here, $c_{\mathscr{O},b}(\boldsymbol{s}_i)$ is an edge-correction factor given as $c_{\mathscr{O},b}(\boldsymbol{s}_i) = \int_{\mathscr{O}} k_b(||\boldsymbol{s} - \boldsymbol{s}_i||))d\boldsymbol{s}$ and $k_b(\boldsymbol{s})$ is the *Epanečnikov kernel* which in one dimension is given as

$$k_b(s) = \frac{3}{4b}\left(1 - \frac{s^2}{b^2}\right)\mathbf{1}(|s| \le 1). \quad \textbf{[S15]}$$

A nonparametric estimator for the PACF is given by (2)

$$\hat{g}_{k,k}(\upsilon) = \frac{1}{2\pi\upsilon|\mathscr{O}|}\sum_{\boldsymbol{s}_i,\boldsymbol{s}_j \in \mathscr{P}_k}^{\neq} \frac{k_b(||\boldsymbol{s}_i - \boldsymbol{s}_j|| - \upsilon)}{\lambda_k^{(1)}(\boldsymbol{s}_i)\lambda_k^{(1)}(\boldsymbol{s}_j)w(\boldsymbol{s}_i, \boldsymbol{s}_j)}, \quad \textbf{[S16]}$$

where $w(\boldsymbol{s}_i, \boldsymbol{s}_j)$ is the fraction of the circle (in two dimensions) with center $\boldsymbol{s}_i$ and radius $||\boldsymbol{s}_i - \boldsymbol{s}_j||$ lying in $\mathscr{O}$. Similarly, an estimate of the PCCF is given by

$$\hat{g}_{k,k+1}(v) = \frac{1}{2\pi v|\mathcal{O}|} \sum_{\substack{s_i \in \mathcal{P}_k \\ s_j \in \mathcal{P}_{k+1}}}^{\neq} \frac{k_b(||s_i - s_j|| - v)}{\lambda_k^{(1)}(s_i)\lambda_{k+1}^{(1)}(s_j)w(s_i, s_j)}. \quad \textbf{[S17]}$$

If the processes are taken to be second-order stationary also in time, to smooth out the nonparametric estimates an average over all K time steps may be taken so that

$$\bar{g}_{k,k}(v) = \frac{1}{K}\sum_{k=1}^{K} \hat{g}_{k,k}(v), \quad \textbf{[S18]}$$

$$\bar{g}_{k,k+1}(v) = \frac{1}{K-1}\sum_{k=1}^{K-1} \hat{g}_{k,k+1}(v). \quad \textbf{[S19]}$$

The set of nonparametric estimators Eqs. **S16–S17** and the averaged estimates Eqs. **S18–S19** may be used to estimate Eq. **S2**. Note that if temporally averaged PACF/PCCFs are used, the inverse filter is given as

$$\hat{k}_I(v) = \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\ln \bar{g}_{k,k+1}(v))}{\mathcal{F}(\ln \bar{g}_{k,k}(v))}\right). \quad \textbf{[S20]}$$

Case Studies: Nonparametric Estimation from Point Process Observations. ***High frequency (small-scale) spatial interactions.*** Here we consider the SIDE of Eq. 1 with $f_1(z_k(\boldsymbol{s})) = z_k(\boldsymbol{s})$, $\Delta_t = 1$, $k_I(v) = 0.05\exp(-||v||^2)$ and $k_Q(v) = 0.8\exp(-||v||^2/5)$ on a domain $\boldsymbol{s} = (s_1, s_2) \in \mathcal{O} = [36 \times 36]$. Synthetic data was generated by discretizing Eq. 1 on a 50×50 grid and carrying out the recursion for $K = 100$ time points; this involved the discretization of $k_Q(v)$ to obtain a $2,500 \times 2,500$ covariance matrix which was then sampled from. The initial field $z_0(\boldsymbol{s})$ was assumed to be drawn from the distribution of the disturbance. Point process observations were then generated using the method of thinning (4) from the exponentiated underlying field. The sequence of time frames shown in Fig. S5 is representative of the whole set; here the point process observations are superimposed on the instantiation of the spatiotemporal field which in this case is known, but not used in the analysis.

In the analysis stage, intensities governing the point process at each time instant were first assumed to constitute a series of correlated log-Gaussian Cox processes (LGCPs) with first-order stationarity, the estimate of which is given through Eq. **S13**. Algorithm S1 was then carried out on the dataset to give nonparametric estimates of the mixing kernel $\hat{k}_I(v)$ and the noise kernel $\hat{k}_Q(v)$ using the exact inverse filter of Eq. **S20**. Fig. S6*A* shows that the estimates (red) conform excellently with the true kernels (blue). The figure also compares the bandwidth of the estimated kernels to the true ones and it clearly shows that the spatial dynamics are very localized (high frequency) when compared to the in-time correlations.

Low frequency (large-scale) spatial interactions. A second synthetic dataset was generated under the same conditions as above but with a significantly wider mixing kernel characteristic of a diffusion process, $k_I(v) = 0.01\exp(-||v||^2/15)$. The data was subsequently analyzed with Algorithm S1. As seen in Fig. S6*B*, the kernel estimates are seen to conform well with the true kernels. Note that in this case the dynamics exhibit interactions over a wide range.

Finite-Dimensional Reduction of the SIDE. For conciseness, define $\mathscr{A}z_k(\boldsymbol{s}) = \int k_I(\boldsymbol{s}, \boldsymbol{r})z_k(\boldsymbol{r})\mathrm{d}\boldsymbol{r}$. We now employ a standard Galerkin-type finite-dimensional reduction on $z_k(\boldsymbol{s})$ by first expanding $z_k(\boldsymbol{s}) \approx \sum_{i=1}^{n} \phi_{z_i}(\boldsymbol{s})x_k = \boldsymbol{\phi}_z^T\boldsymbol{x}_k$, $\boldsymbol{\phi}_z(\boldsymbol{s}) \in \mathbb{R}^{n_z}$ to obtain

$$\boldsymbol{\phi}_z(\boldsymbol{s})^T\boldsymbol{x}_{k+1} = \mathscr{A}\boldsymbol{\phi}_z(\boldsymbol{s})^T\boldsymbol{x}_k + e_k(\boldsymbol{s}), \quad \textbf{[S21]}$$

and, subsequently projecting through the inner product with $\phi_z(\boldsymbol{s})$, $\langle \boldsymbol{\phi}_z(\boldsymbol{s}), \cdot \rangle$, to obtain

$$\langle \boldsymbol{\phi}_z(\boldsymbol{s}), \boldsymbol{\phi}_z(\boldsymbol{s})^T\rangle\boldsymbol{x}_{k+1} = \langle \boldsymbol{\phi}_z(\boldsymbol{s}), \mathscr{A}\boldsymbol{\phi}_z(\boldsymbol{s})^T\rangle\boldsymbol{x}_k + \langle \boldsymbol{\phi}_z(\boldsymbol{s}), e_k(\boldsymbol{s})\rangle. \quad \textbf{[S22]}$$

Let $\boldsymbol{\Psi}_{\boldsymbol{x}} = \langle \boldsymbol{\phi}_z(\boldsymbol{s}), \boldsymbol{\phi}_z(\boldsymbol{s})^T\rangle$ and $\boldsymbol{\Psi}_{\mathscr{A}} = \langle \boldsymbol{\phi}_z(\boldsymbol{s}), \mathscr{A}\boldsymbol{\phi}_z(\boldsymbol{s})^T\rangle$. Then

$$\boldsymbol{x}_{k+1} = \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\boldsymbol{\Psi}_{\mathscr{A}}\boldsymbol{x}_k + \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\langle \boldsymbol{\phi}_z(\boldsymbol{s}), e_k(\boldsymbol{s})\rangle. \quad \textbf{[S23]}$$

We next decompose the heterogeneous effects in $\mu_Q(\boldsymbol{s}), k_I(\boldsymbol{s}, \boldsymbol{r})$ and $k_Q(\boldsymbol{s}, \boldsymbol{r})$ using vectors of basis functions $\boldsymbol{\phi}_{\mu_Q} \in \mathbb{R}^{n_{\mu_Q}}$, $\boldsymbol{\phi}_{k_I} \in \mathbb{R}^{n_{k_I}}$ and $\boldsymbol{\phi}_{k_Q} \in \mathbb{R}^{n_{k_Q}}$ respectively to obtain

$$\mu_Q(\boldsymbol{s}) = \boldsymbol{\phi}_{\mu_Q}(\boldsymbol{s})^T\boldsymbol{\vartheta}, \quad \textbf{[S24]}$$

$$k_I(\boldsymbol{s}, \boldsymbol{r}) = \boldsymbol{\phi}_{k_I}(\boldsymbol{s})^T\boldsymbol{\Sigma}_I\boldsymbol{\phi}_{k_I}(\boldsymbol{r}), \quad \textbf{[S25]}$$

$$k_Q(\boldsymbol{s}, \boldsymbol{r}) = \boldsymbol{\phi}_{k_Q}(\boldsymbol{s})^T\boldsymbol{\Sigma}_Q\boldsymbol{\phi}_{k_Q}(\boldsymbol{r}), \quad \textbf{[S26]}$$

where $\boldsymbol{\vartheta} \in \mathbb{R}^{n_{\mu_Q}}$, $\boldsymbol{\Sigma}_I \in \mathbb{R}^{n_{k_I} \times n_{k_I}}$ and $\boldsymbol{\Sigma}_Q \in \mathbb{R}^{n_{k_Q} \times n_{k_Q}}$ are unknown and, in addition to each $\boldsymbol{x}_k$, need to be estimated from the data. Under this decomposition, $\boldsymbol{\Psi}_{\mathscr{A}}$ is given by

$$\boldsymbol{\Psi}_{\mathscr{A}} = \langle \boldsymbol{\phi}_z(\boldsymbol{s}), \mathscr{A}\boldsymbol{\phi}_z(\boldsymbol{s})^T\rangle \quad \textbf{[S27]}$$

$$= \langle \boldsymbol{\phi}_z(\boldsymbol{s}), \int_{\mathcal{O}} \boldsymbol{\phi}_{k_I}(\boldsymbol{s})^T\boldsymbol{\Sigma}_I\boldsymbol{\phi}_{k_I}(\boldsymbol{r})\boldsymbol{\phi}_z(\boldsymbol{r})^T\mathrm{d}\boldsymbol{r}\rangle \quad \textbf{[S28]}$$

$$= \iint_{\mathcal{O}} \boldsymbol{\phi}_z(\boldsymbol{s})\boldsymbol{\phi}_{k_I}(\boldsymbol{s})^T\boldsymbol{\Sigma}_I\boldsymbol{\phi}_{k_I}(\boldsymbol{r})\boldsymbol{\phi}_z(\boldsymbol{r})^T\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{s} \quad \textbf{[29]}$$

$$= \langle \boldsymbol{\phi}_z(\boldsymbol{s}), \boldsymbol{\phi}_{k_I}(\boldsymbol{s})^T\rangle\boldsymbol{\Sigma}_I\langle \boldsymbol{\phi}_{k_I}(\boldsymbol{s}), \boldsymbol{\phi}_z(\boldsymbol{s})^T\rangle, \quad \textbf{[S30]}$$

to give

$$\boldsymbol{\Psi}_{\mathscr{A}} = \boldsymbol{\Phi}_{z,k_I}\boldsymbol{\Sigma}_I\boldsymbol{\Phi}_{z,k_I}^T, \quad \textbf{[S31]}$$

where $\boldsymbol{\Phi}_{z,k_I} = \langle \boldsymbol{\phi}_z(\boldsymbol{s}), \boldsymbol{\phi}_{k_I}(\boldsymbol{s})^T\rangle$. Similarly, the mean of the added disturbance $\boldsymbol{w}_k = \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\langle \boldsymbol{\phi}_z(\boldsymbol{s}), e_k(\boldsymbol{s})\rangle$ is given by

$$\mathbb{E}[\boldsymbol{w}_k] = \mathbb{E}[\boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\langle \boldsymbol{\phi}_z(\boldsymbol{s}), e_k(\boldsymbol{s})\rangle] \quad \textbf{[S32]}$$

$$= \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\langle \boldsymbol{\phi}_z(\boldsymbol{s}), \boldsymbol{\phi}_{\mu_Q}(\boldsymbol{s})^T\rangle\boldsymbol{\vartheta}, \quad \textbf{[S33]}$$

to give

$$\mathbb{E}[\boldsymbol{w}_k] = \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\boldsymbol{\Phi}_{z,\mu_Q}\boldsymbol{\vartheta}. \quad \textbf{[S34]}$$

Further, the second moment is given by

$$\mathbb{E}[\boldsymbol{w}_k\boldsymbol{w}_k^T] = \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\mathbb{E}[\langle\boldsymbol{\phi}_z(\boldsymbol{s}), e_k(\boldsymbol{s})\rangle\langle e_k(\boldsymbol{r}), \boldsymbol{\phi}_z(\boldsymbol{r})^T\rangle]\boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}$$
$$= \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\iint_{\mathcal{O}}\boldsymbol{\phi}_z(\boldsymbol{s})\mathbb{E}[e_k(\boldsymbol{s})e_k(\boldsymbol{r})]\boldsymbol{\phi}_z(\boldsymbol{r})^T\mathrm{d}\boldsymbol{s}\mathrm{d}\boldsymbol{r}\boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}$$
$$= \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}[\iint_{\mathcal{O}}\boldsymbol{\phi}_z(\boldsymbol{s})\boldsymbol{\phi}_{\mu_Q}(\boldsymbol{s})^T\boldsymbol{\vartheta}\boldsymbol{\vartheta}^T\boldsymbol{\phi}_{\mu_Q}(\boldsymbol{r})\boldsymbol{\phi}_z(\boldsymbol{r})^T\mathrm{d}\boldsymbol{s}\mathrm{d}\boldsymbol{r}$$
$$+ \iint_{\mathcal{O}}\boldsymbol{\phi}_z(\boldsymbol{s})\boldsymbol{\phi}_{k_Q}(\boldsymbol{s})^T\boldsymbol{\Sigma}_Q\boldsymbol{\phi}_{k_Q}(\boldsymbol{r})\boldsymbol{\phi}_z(\boldsymbol{r})^T\mathrm{d}\boldsymbol{s}\mathrm{d}\boldsymbol{r}]\boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}, \quad \text{[S35]}$$

so that

$$\mathrm{cov}[\boldsymbol{w}_k] = \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\langle\boldsymbol{\phi}_z(\boldsymbol{s}), \boldsymbol{\phi}_{k_Q}(\boldsymbol{s})^T\rangle\boldsymbol{\Sigma}_Q\langle\boldsymbol{\phi}_{k_Q}(\boldsymbol{s}), \boldsymbol{\phi}_z(\boldsymbol{s})^T\rangle\boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}, \quad \text{[S36]}$$

to give

$$\mathrm{cov}[\boldsymbol{w}_k] = \boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}\boldsymbol{\Phi}_{z,k_Q}\boldsymbol{\Sigma}_Q\boldsymbol{\Phi}_{z,k_Q}^T\boldsymbol{\Psi}_{\boldsymbol{x}}^{-1}, \quad \text{[S37]}$$

where the matrices $\boldsymbol{\Phi}_{z,\mu_Q} = \langle\boldsymbol{\phi}_z(\boldsymbol{s}), \boldsymbol{\phi}_{\mu_Q}(\boldsymbol{s})^T\rangle$ and $\boldsymbol{\Phi}_{z,k_Q} = \langle\boldsymbol{\phi}_z(\boldsymbol{s}), \boldsymbol{\phi}_{k_Q}(\boldsymbol{s})^T\rangle$.

As shown in the main text considerable simplifications are obtained when choosing the basis $\boldsymbol{\phi}_z(\boldsymbol{s}) = \boldsymbol{\phi}_{\mu_Q}(\boldsymbol{s}) = \boldsymbol{\phi}_{k_I}(\boldsymbol{s}) = \boldsymbol{\phi}_{k_Q}(\boldsymbol{s})$.

Consistency in Event Rate Between the Afghan War Diary (AWD), Afghanistan NGO Safety Office (ANSO) Armed Opposition Groups (AOG) Reports, Armed Conflict Location and Event Dataset (ACLED), and Global Terrorism Database (GTD) Datasets. The AWD is a large but highly heterogeneous collection of event logs but the reluctance of government officials to confirm the dataset as being factually accurate and a complete portrayal of the Afghan war (5) has led a few to doubt its use as a reflection of the ground truth. As a result, substantial verification efforts were carried out by several researchers and organisations alike. The New York Times Company, for instance, cross-validated a number of logged entries with its own media reports*. O'Loughlin et al. on the other hand compared the spatial and temporal distribution of violent events in the AWD to that in the ACLED, finding significant correlation in support of consistency (6).

Whilst it is is beyond the scope of this article to verify the AWD as a representatory dataset, here we show some of our own correlation tests between the AWD and ACLED (7), the ANSO Q4 reports on AOG initiated attacks†, and the GTD‡ in the hope that our results help to further evidence consistency across different databases. Throughout, our basic assumption is that the incidence rate in the AWD, both violent and nonviolent, should corroborate with that in other datasets, both geographically and temporally. We note here that the corroborating datasets are not free from sampling biases themselves: for instance, NGO reports from extremely dangerous regions such as Helmand may underestimate the event counts simply because NGOs may have a lesser presence there due to security reasons. We stress therefore that the level of corroboration we expect to find needs not be perfect, but simply a confirmation of general trends.

ACLED. The ACLED dataset is an extensive one, constructed from a variety of local and international media sources and NGO reports. The majority of logs in the ACLED dataset denote instances of political violence, where force is exercised by one or more actors (governments, militias, rebel groups) for a political end. As such one would expect the incidence rate of ACLED events to corroborate with violent events in the AWD and indeed (6) found a significant correlation in the respective geographical distributions in 2008/2009. Here we replicated the study with the difference that we included all of the AWD in the analysis, and not just violent events. A geographical assessment revealed significant correlations in both years, with a Pearson's correlation coefficient of $r = 0.88$ in 2008 and $r = 0.92$ in 2009. See Fig. S7 *A* and *B*.

ANSO AOG reports. ANSO monitors the activity of AOG nationwide and its Q4 reports provide detailed overviews of the frequency of AOG incidents on a provincial basis. A similar geographical assessment to that carried out with ACLED revealed significant (although weaker) correlations between AOG activity and event log incidence in the AWD, with a Pearson's correlation coefficient of $r = 0.73$ in 2008 and $r = 0.59$ in 2009. Omission of Helmand, a clear outlier in this analysis, improved r to 0.84 and 0.89 respectively. See Fig. S7 *C* and *D*.

GTD. The GTD is a collection of international terrorist incidents, 1,783 of which are located in Afghanistan between 2004 and 2009. Most terrorist attacks are reported as being perpetrated by the Taliban and list a number of targets including private citzens, the telecommunication infrastructure, government buildings and personnel and NGOs. As with the ANSO reports we thus expected the prevalence of logs in this database to corroborate with the AWD and a correlation test across all years strongly confirmed this with a coefficient of $r = 0.93$. See Fig. S7*E*. Because the geographical location in the GTD is coded through city names, generally inconsistent with those in standard shapefiles, a comparison on a provincial level was omitted. String similarity checks with the use of, for instance, the Levenshtein distance, did not produce reliable merges in this case.

Basis Function Selection. For basis selection we adopt a frequency-based approach. First the Fourier transform of the average PACF is found, from which a cutoff frequency of ν_c cycles/unit is selected. Second, localized reconstruction kernels are placed at regular intervals throughout the spatial domain. The resolution of the lattice on which the functions are placed has to be small enough to avoid aliasing by satisfying Shannon's sampling criterion. In particular, if the centers of the basis functions, denoted $\{\boldsymbol{\zeta}_i\}_{i=1}^n$, are set equal to the sequence of vectors describing a regular lattice of edge length Δ_s in $\mathcal{O}$, then it is required that

$$\Delta_s < \frac{1}{2\nu_c} = \frac{1}{2\alpha_0\nu_c}, \quad \text{[S38]}$$

where α_0 is an oversampling parameter. Lastly, the 'width' of the local functions, which determines the range of frequencies they can represent, is set to cater for the frequency content of the spatial field at each time point.

To demonstrate the last step, consider the case when the basis functions are Gaussian radial basis functions (GRBFs) with functional form $\phi(s) = \exp(-s^2/2\sigma_b^2)$. The Fourier transform of the GRBF is yet another GRBF (in the frequency domain) given as

$$\phi(\nu) = \mathcal{F}\{\phi(s)\} = \sqrt{2\pi\sigma_b^2}\exp(-2\pi^2\sigma_b^2\nu^2), \quad \text{[S39]}$$

so that the variances in the spatial and frequency domain are related through the mappings (8)

$$\sigma_\nu^2 \leftarrow \frac{1}{4\pi^2\sigma_b^2}, \sigma_b^2 \leftarrow \frac{1}{4\pi^2\sigma_\nu^2}. \quad \text{[S40]}$$

The range of frequencies which can be represented by the basis functions has to exceed that of the field for adequate reconstruction. To this end Sanner and Slotine (8) suggest the following relationship

*http://www.nytimes.com/2010/07/26/world/26editors-note.html?_r=1

†Reports are freely available from the official ANSO website http://www.afgnso.org.

‡National Consortium for the Study of Terrorism and Responses to Terrorism (2011) Retrieved from http://www.start.umd.edu/gtd

$$\sigma_\nu = \frac{1}{\sqrt{2}}\nu_c. \quad \textbf{[S41]}$$

Given σ_ν, **[S40]** can then be used to find the width of the desired GRBF in $\mathcal{O}$. By substitution of Eq. **S41** in **[S40]**, $\sigma_b = (2\nu_c^2\pi^2)^{-1/2}$. The resulting basis is hence a set of GRBFs with parameter σ_b placed in the spatial domain centered on the coordinates $\{\boldsymbol{\zeta}_i\}_{i=1}^n$.

Because GRBFs are not of compact support (and hence have a diminishing yet far-reaching effect), we instead use a compact radial basis function, termed the *compact GRBF* (CGRBF), of the form

$$\phi(\boldsymbol{s}) = \begin{cases} \frac{(2\pi - ||\tau\boldsymbol{s}||)(1+(\cos\tau||\boldsymbol{s}||/2)+\frac{3}{2}\sin(\tau||\boldsymbol{s}||)}{3\pi}, & ||\tau\boldsymbol{s}|| < 2\pi, \\ 0, & \text{otherwise}, \end{cases} \quad \textbf{[S42]}$$

for $\tau > 0$ and where $||\cdot||$ denotes the usual Euclidean distance on $\mathcal{O}$. The CGRBF closely resembles the usual GRBF with $\phi(\boldsymbol{s}) = \exp(-\tau^2||\boldsymbol{s}||^2/2\pi)$, however it is of compact support. For a given GRBF parameter σ_b, or a cutoff frequency ν_c, the CGRBF parameter is then given by

$$\tau = \sqrt{\pi}/\sigma_b = \sqrt{2\nu_c^2\pi^3}. \quad \textbf{[S43]}$$

For the AWD a cutoff frequency of $\nu_c = 0.2$ cycles/unit was selected from the average PACF shown in Fig. S8*A*, thus corresponding to the basis parameter $\tau \approx 0.9\sqrt{\pi}$. CGRBFs were then equally spaced within the entire domain with an oversampling parameter of $\alpha_0 = 1.3$. A cross-section of the chosen basis function is shown in Fig. S8*A*.

Initially basis functions were placed on a 16×16 grid with an intercentre spacing of $\Delta_s = 1.9$ and covering the whole of Afghanistan. Many of these basis functions were however considered redundant, representing areas exhibiting no logged events, or a very small number of events. To avoid problems of identifiability in these regions (see also study in *SI Text*), a constant background intensity baseline b_1 was used to represent activity in these areas. Each basis function was analyzed separately; if a basis function had its center more than 0.4 spatial units outside Afghanistan or had on average less than eight logged events per year within 1.3 standard deviations from its center (corresponding to a background event rate $b_1 = -3.5$) it was omitted. The final arrangement of the basis functions together and the spatial distribution of all the logged events of the AWD in Afghanistan are shown in Fig. S8 *B* and *C* respectively. Note that basis functions are omitted in 'quiet' areas. To cater for the inclusion of the background event rate, we set the first covariate, $d_1(\boldsymbol{s}) = 1$ so that

$$\lambda_k(\boldsymbol{s}) = \exp(b_1 + b_2 d_2(\boldsymbol{s}) + b_3 d_3(\boldsymbol{s}) + \ldots + z_k(\boldsymbol{s})). \quad \textbf{[S44]}$$

As can be seen from Eq. **S44**, in the absence of other covariates, the intensity where there are no basis functions simply reduces to $\exp(b_1)$.

Whilst placing basis functions only in regions which are highly represented by the data is not new both in spatial (9) and spatiotemporal (10) systems, care should be exercised if extending the problem to an online setting. Basis function omission induces a strong prior on the model, and unexpected changes in unrepresented regions would not be detected with the current setup (11); a realistic trade-off, in this case, needs to be found between the computational complexity and level of representation. We note that this issue is virtually inexistent in offline studies (such as this one), where it is very common to employ a basis which is not amenable to substantial changes in temporal behavior (12, 13).

Controlling for Purely Spatial Variation. As is typical in spatiotemporal point process applications, we considered the addition of deterministic components to the intensity model to control for certain demographical and topological features:

Population density. It is generally an accepted notion in conflict that more populous regions witness, on average, more conflict events than rural regions (14). This association was confirmed by our study which compared extrapolations from the 1979 census and a precensus survey in 2003/2004 to the spatial intensity of the AWD, see Fig. S10*A*. Data, available from the Central Statistics Organization, http://cod.humanitarianresponse.info/country-region/afghanistan, is aggregated at the district level.

Distance to major city. Similar to population density, locations far away from any large concentrations of population are likely to witness less conflict events. This association was confirmed by our study which compared a digital map of distances to the 33 major cities (as from 1994) in Afghanistan to the intensity map, see Fig. S10*B*. Data for settlement location is available from the Afghanistan Information Management Service, http://www.aims.org.af/services/mapping/shape_files/afghanistan/point/.

Elevation and terrain type. In (6) it was seen that most violent events occurred in flat terrain. This observation does not imply that flat terrain is more susceptible to conflict and in fact we found no evidence that this need be the case, see Fig. S10*C*. We also found no simple relationship between elevation and overall event intensity, see Fig. S10*D*. Topology data was obtained from the GTOPO30 dataset (U.S. Geological Survey 2007, 30-arc second resolution).

Distance to Pakistan border. It is well known that the proximity to borders increases the prospect of conflict. Afghanistan may be no exception; insurgents are known to use regions across the Pakistan border for refuge. However, an analysis showed that there was no direct association between the distance to the Pakistan border and the conflict intensity map. The lack of association is most probably due to the high intensities in the central border area (e.g., Pakitka, Nangarhar, and Kunar) being offshot by the relatively quiet regions in the south near Nimroz and the north in Badakhshan, see Fig. S10*E*.

As a result of this study the intensity model was augmented so that $\boldsymbol{d}(\boldsymbol{s}) = (1, d_1(\boldsymbol{s}), d_2(\boldsymbol{s}))$ where the first element corresponds to a background intensity rate (established in *SI Text*), $d_1(\boldsymbol{s})$ the population density and $d_2(\boldsymbol{s})$ the distance to the closest major city, the latter two variables deemed as having strong associations with the overall spatial intensity of the AWD.

Variational Bayes (VB) Update Equations. By finding a lower bound on the marginal likelihood it can be shown that the required VB marginals for the unknown states $\mathcal{X}_K$ and parameters $\boldsymbol{\theta} = (\boldsymbol{\vartheta}, \boldsymbol{\Sigma}_Q^{-1})$ and $\boldsymbol{b} = [b_1, b_2, \ldots, b_d]$ (d denoting the number of covariates) are given as (19)

$$\tilde{p}(\mathcal{X}_K) \propto \exp(\mathbb{E}_{\tilde{p}(\boldsymbol{\theta})\tilde{p}(\boldsymbol{b})}[\ln p(\mathcal{Y}_K, \mathcal{X}_K, \boldsymbol{\theta}, \boldsymbol{b})]), \quad \textbf{[S45]}$$

$$\tilde{p}(\boldsymbol{\vartheta}) \propto \exp(\mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{\theta}^{/\vartheta})\tilde{p}(\boldsymbol{b})}[\ln p(\mathcal{Y}_K, \mathcal{X}_K, \boldsymbol{\theta}, \boldsymbol{b})]), \quad \textbf{[S46]}$$

$$\tilde{p}(\boldsymbol{\Sigma}_Q^{-1}) \propto \exp(\mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{\theta}^{/\boldsymbol{\Sigma}_Q^{-1}})\tilde{p}(\boldsymbol{b})}[\ln p(\mathcal{Y}_K, \mathcal{X}_K, \boldsymbol{\theta}, \boldsymbol{b})]), \quad \textbf{[S47]}$$

$$\tilde{p}(b_i) \propto \exp(\mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{\theta})\tilde{p}(\boldsymbol{b}^{/b_i})})[\ln p(\mathcal{Y}_K, \mathcal{X}_K, \boldsymbol{\theta}, \boldsymbol{b})]), i = 1 \ldots d, \quad \textbf{[S48]}$$

where $\boldsymbol{\theta}^{/\theta}$ denotes the set of variables $\boldsymbol{\theta}$ without θ and $\mathbb{E}_{\tilde{p}(\cdot)}[\cdot]$ is used to render specific the distribution relative to which we are

taking expectations. Next, we formulate the algorithm used to infer the unknown quantities. Throughout, the notation $i|j$ refers to the estimate at time i conditioned on data up to time j. For ease of the exposition the model is rewritten as

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{\vartheta} + \tilde{\boldsymbol{w}}_k, \qquad \text{[S49]}$$

where $\tilde{\boldsymbol{w}}_k$ is now with 0 mean.

State inference. For the computation of [S45] we employ an approximate variational Kalman smoother (15). Let $\boldsymbol{x}_0 \sim \mathcal{N}_{x0}(\mu_0, \boldsymbol{\Sigma}_0)$. and denote the variational forward message as $\tilde{\alpha}(\boldsymbol{x}_k) = \tilde{p}(\boldsymbol{x}_k|\boldsymbol{y}_{1:k})$. This latter quantity is further approximated through the Laplace method as

$$\tilde{\alpha}(\boldsymbol{x}_k) \propto \int \tilde{\alpha}(\boldsymbol{x}_{k-1}) \exp(\mathbb{E}_{\tilde{p}(\boldsymbol{\theta})\tilde{p}(\boldsymbol{b})}[\ln p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{\theta}) p(\boldsymbol{y}_k|\boldsymbol{x}_k, \boldsymbol{b})]) \mathrm{d}\boldsymbol{x}_{k-1}$$
$$\overset{\text{Laplace}}{\rightarrow} \mathcal{N}_{\boldsymbol{x}_k}(\hat{\boldsymbol{x}}_{k|k}, \boldsymbol{\Sigma}_{k|k}), \qquad \text{[S50]}$$

where $\boldsymbol{y}_k$ is conditionally independent of $\boldsymbol{\theta}$. Similarly the backward message $\tilde{\beta}(\boldsymbol{x}_k) = \tilde{p}(\boldsymbol{y}_{k+1:K}|\boldsymbol{x}_k)$ is given by

$$\tilde{\beta}(\boldsymbol{x}_k) = \int \tilde{\beta}(\boldsymbol{x}_{k+1}) \exp(\mathbb{E}_{\tilde{p}(\boldsymbol{\theta})\tilde{p}(\boldsymbol{b})}[\ln p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k, \boldsymbol{\theta}) p(\boldsymbol{y}_{k+1}|\boldsymbol{x}_{k+1}, \boldsymbol{b})]) \mathrm{d}\boldsymbol{x}_{k+1}$$
$$\overset{\text{Laplace}}{\rightarrow} \mathcal{N}_{\boldsymbol{x}_k}(\hat{\boldsymbol{x}}_{k|k+1:K}, \boldsymbol{\Sigma}_{k|k+1:K}). \qquad \text{[S51]}$$

The two messages are then combined to give the smoothed estimate:

$$\tilde{p}(\boldsymbol{x}_k|\boldsymbol{y}_{1:K}) \propto \tilde{p}(\boldsymbol{x}_k|\boldsymbol{y}_{1:k}) \tilde{p}(\boldsymbol{y}_{k+1}|\boldsymbol{x}_k) = \tilde{\alpha}(\boldsymbol{x}_k) \tilde{\beta}(\boldsymbol{x}_k)$$
$$= \mathcal{N}_{\boldsymbol{x}_k}(\hat{\boldsymbol{x}}_{k|K}, \boldsymbol{\Sigma}_{k|K}). \qquad \text{[S52]}$$

The resulting equations of the forward-backward smoother are quite involved and given in Algorithm S2.

Escalation inference. Let the prior $p(\boldsymbol{\vartheta}) = \mathcal{N}_{\boldsymbol{\vartheta}}(\hat{\boldsymbol{\vartheta}}_p, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta},p})$. Then, the posterior $\tilde{p}(\boldsymbol{\vartheta})$ in [S46] is given by

$$\tilde{p}(\boldsymbol{\vartheta}) \propto p(\boldsymbol{\vartheta}) \exp\left(-\frac{1}{2} \mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{\Sigma}_Q^{-1})} \left[\sum_{k=0}^{K-1} (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k - \boldsymbol{\vartheta})^T \times \boldsymbol{\Sigma}_Q^{-1} (\boldsymbol{x}_{k+1} - \boldsymbol{x}_k - \boldsymbol{\vartheta})\right]\right), \qquad \text{[S53]}$$

so that $\boldsymbol{\vartheta} \sim \mathcal{N}_{\boldsymbol{\vartheta}}(\hat{\boldsymbol{\vartheta}}, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}})$ where

$$\hat{\boldsymbol{\vartheta}} = \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}} \left[\boldsymbol{\Sigma}_{\boldsymbol{\vartheta},p}^{-1} \hat{\boldsymbol{\vartheta}}_p + \mathbb{E}_{\tilde{p}(\boldsymbol{\Sigma}_Q^{-1})}[\boldsymbol{\Sigma}_Q^{-1}] \sum_{k=0}^{K-1} (\mathbb{E}_{\tilde{p}(\mathcal{X}_K)^{(i+1)}}[\boldsymbol{x}_{k+1} - \boldsymbol{x}_k])\right], \qquad \text{[S54]}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}} = (\boldsymbol{\Sigma}_{\boldsymbol{\vartheta},p}^{-1} + K \mathbb{E}_{\tilde{p}(\boldsymbol{\Sigma}_Q^{-1})}[\boldsymbol{\Sigma}_Q^{-1}])^{-1}. \qquad \text{[S55]}$$

Volatility inference. Let the prior $p(\boldsymbol{\Sigma}_Q^{-1}) = \mathcal{W}i_{\boldsymbol{\Sigma}_Q^{-1}}(\boldsymbol{V}_p, d_p)$ where $\mathcal{W}i_{\boldsymbol{\Sigma}_Q^{-1}}(\boldsymbol{V}, d)$ denotes a Wishart distribution with $\boldsymbol{V}$ a positive-definite, symmetric scale matrix and d degrees of freedom. The variational posterior of [S46] is given by

$$\tilde{p}(\boldsymbol{\Sigma}_Q^{-1}) \propto p(\boldsymbol{\Sigma}_Q^{-1}) \exp\left(\frac{K}{2} \ln|\boldsymbol{\Sigma}_Q^{-1}| - \frac{1}{2} tr(\boldsymbol{\Gamma} \boldsymbol{\Sigma}_Q^{-1})\right), \qquad \text{[S56]}$$

where

$$\boldsymbol{\Gamma} = \sum_{k=1}^{K} \mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{\theta})}[(\boldsymbol{x}_k - \boldsymbol{x}_{k-1} - \boldsymbol{\vartheta})(\boldsymbol{x}_k - \boldsymbol{x}_{k-1} - \boldsymbol{\vartheta})^T]. \qquad \text{[S57]}$$

It can then be easily shown that $\tilde{p}(\boldsymbol{\Sigma}_Q^{-1}) = \mathcal{W}i_{\boldsymbol{\Sigma}_Q^{-1}}(\hat{\boldsymbol{V}}, \hat{d})$ where

$$\hat{\boldsymbol{V}} = (\boldsymbol{V}_p^{-1} + \boldsymbol{\Gamma})^{-1}, \qquad \text{[S58]}$$

$$\hat{d} = d_p + K. \qquad \text{[S59]}$$

Evaluation of $\boldsymbol{\Gamma}$ requires evaluation of the cross-covariance matrix in addition to the usual posterior covariance matrices. The computation of the cross-covariance, also requiring Laplace approximations (see also ref. 16), is given in the last for loop of Algorithm S2.

Regression parameters. Under VB we let $\tilde{p}(\boldsymbol{b}) = \prod_{i=1}^{d} \tilde{p}(b_i)$. Let the prior $p(b_i) = \mathcal{N}_{b_i}(\hat{b}_{i.p}, \sigma_{\hat{b}_{i.p}}^2)$. Then the variational posterior $\tilde{p}(b_i)$ of Eq. **S48**, under a Laplace approximation, is given by

$$\tilde{p}(b_i) \propto p(b_i) \prod_{k \in \mathcal{K}} \left\{ \left[\prod_{\boldsymbol{s}_j \in \boldsymbol{y}_k} \exp(\mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{b}^{/b_i})}[\boldsymbol{b}^T \boldsymbol{d}(\boldsymbol{s}_j) + \boldsymbol{\phi}(\boldsymbol{s}_j)^T \boldsymbol{x}_k])\right] \exp\left(\mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{b}^{/b_i})}\left[-\int_{\mathcal{O}} \exp(\boldsymbol{b}^T \boldsymbol{d}(\boldsymbol{s}) + \boldsymbol{\phi}^T(\boldsymbol{s}) \boldsymbol{x}_k)\right]\right) \mathrm{d}\boldsymbol{s}\right\} \overset{\text{Laplace}}{\rightarrow} \mathcal{N}_{b_i}(\hat{b}_i, \sigma_{\hat{b}_i}^2),\ i = 1 \ldots d, \qquad \text{[S60]}$$

where it can be easily shown that

$$\hat{b}_i = \hat{b}_{i.p} + \sigma_{\hat{b}_i.p}^2 \left(\sum_{\substack{k \in \mathcal{K} \\ \boldsymbol{s}_j \in \boldsymbol{y}_k}} d_i(\boldsymbol{s}_j) - \sum_{k \in \mathcal{K}} \mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{b}^{/b_i})} \left[\int_{\mathcal{O}} d_i(\boldsymbol{s}) \exp(\boldsymbol{b}^T \boldsymbol{d}(\boldsymbol{s}) + \boldsymbol{\phi}^T(\boldsymbol{s}) \boldsymbol{x}_k) \mathrm{d}\boldsymbol{s}\right]\right), \qquad \text{[S61]}$$

$$\sigma_{\hat{b}_i}^2 = \left(\sigma_{\hat{b}_i.p}^{-2} + \sum_{k \in \mathcal{K}} \mathbb{E}_{\tilde{p}(\mathcal{X}_K)\tilde{p}(\boldsymbol{b}^{/b_i})} \left[\int_{\mathcal{O}} d_i^2(\boldsymbol{s}) \exp(\boldsymbol{b}^T \boldsymbol{d}(\boldsymbol{s}) + \boldsymbol{\phi}^T(\boldsymbol{s}) \boldsymbol{x}_k) \mathrm{d}\boldsymbol{s}\right]\right)^{-1},\ i = 1 \ldots d. \qquad \text{[S62]}$$

Inference for the AWD was completed in less than an hour on a standard PC: this included the approximation of integrals within the optimization routines for variational-Laplace on a 100×100 grid, the use of a relatively low tolerance value for terminating the optimisation routine (0.1% change in sequential function evaluations) and six state-parameter iterations for convergence.

Configuration Notes. ***State inference:*** The initial state $\hat{\boldsymbol{x}}_{0|0}$ was set by first carrying out nonparametric estimation of the field in the first week ($k = 1$) using conventional methods (17) and then regressing this onto the chosen basis $\boldsymbol{\phi}$ using ordinary least squares. $\boldsymbol{\Sigma}_{0|0}$ was set to $30\boldsymbol{I}$. In Algorithm S2 the prior from a Kalman filter running in parallel and assuming point estimates was used as an initial condition in the first for loop. In the second for loop, the mean of the forward message was used as initialization. In

both cases, gradient descent was halted after a change of less than a 0.1% in sequential function evaluations (typically 20–30 function evaluations were required). The integrals in Algorithm S2 were approximated on a 100×100 discrete grid using numerical quadrature.

Parameter inference: The parameter priors were configured as follows (recall that b_1 was fixed a priori):

$$\hat{b}_{i,p} = 0, \qquad \sigma^2_{b_i,p} = 10, \qquad i = 2, 3, \qquad \textbf{[S63]}$$

$$\hat{\boldsymbol{\vartheta}}_p = \mathbf{0}, \qquad \boldsymbol{\Sigma}_{\boldsymbol{\vartheta},p} = 1000\boldsymbol{I}, \qquad \textbf{[S64]}$$

$$d_p = 1000, \qquad \boldsymbol{V}_p = 0.025\boldsymbol{I}. \qquad \textbf{[S65]}$$

The prior scale matrix $\boldsymbol{V}_p$ was chosen such that its mean is $25\boldsymbol{I}$, where $\sigma^{-2} = 25$ is equal to the squared reciprocal of the standard deviation of the logged increments in 2006, the largest of the four years 2006–2009 for which homoskedasticity was met.

Stopping conditions: The VB algorithm was assumed to have converged when the change in $\hat{\boldsymbol{\vartheta}}$ and $\hat{\boldsymbol{b}}_i, i = 2, 3$ in subsequent iterations was less than 0.005, and when all diagonal elements in $\mathbb{E}[\boldsymbol{\Sigma}_Q^{-1}] = \hat{d}\,\hat{V}$ changed by less than 1%.

Case Study: VB Estimation from Point Process Observations. ***Simulation setup.*** Here we consider the SIDE of Eq. 1 with $f_1(z_k(\boldsymbol{s})) = \gamma z_k(\boldsymbol{s})$, $\gamma = 0.1$, $\Delta_t = 1$, $k_I(v) = \delta(v)$ and $k_Q(\boldsymbol{s}, \boldsymbol{r}) = \sigma(\boldsymbol{s})\sigma(\boldsymbol{r}) \exp(-||\boldsymbol{s} - \boldsymbol{r}||^2/3)$ on a domain $\boldsymbol{s} = (s_1, s_2) \in \mathcal{O} = [0\ 18] \times [0\ 18]$. Spatially varying volatility is modeled through $\sigma(\boldsymbol{s}) = 0.5 + 2(\exp(-(s_1 - 5)^2/6 - (s_2 - 2.5)^2/12) + \exp(-(s_1 - 10)^2/6 - (s_2 - 13)^2/12))$ and the nonstationary mean $\mu_Q(\boldsymbol{s})$ is itself generated from a Gaussian process (GP) $\mu_Q \sim \mathcal{GP}(0, k_{\mu_Q})$ where $k_{\mu_Q}(\boldsymbol{s}, \boldsymbol{r}) = 3.2 \exp(-||\boldsymbol{s} - \boldsymbol{r}||^2/3)$. The intensity function was modeled as in Eq. **S44** with $b_1 = -2$ ($d_1 = 1$) and $b_2 = 0, b_3 = 0, \ldots$. The simulation configuration reflects one which typically gives a total event count of the same order of magnitude as that present in the AWD.

Synthetic data was generated by discretising the SIDE on a 25×25 grid and carrying out the recursion for $K = 300$ time points. Point processes observations were once again simulated using the method of thinning (4) from the true intensity with linear interpolation used for evaluating the latent data in between grid points. The complete set of points is shown in Fig. S9*A*, where the heterogeneity in the dynamics of the governing intensity is immediately apparent. The simulated growth map and the volatility surface are shown in Fig. S9 *B* and *C* respectively. For low-rank representation, a set of 8×8 grid of CGRBFs of the form Eq. **S42** were placed on a regular grid inside the domain and on the boundary, truncated were appropriate, see Fig. S9*D*. At this stage basis function omission may be carried out to selectively omit functions in highly unrepresented regions; however here they will be kept to explicitly show problems of nonidentifiability (of the parameters) in these regions.

As assumed in the AWD, the background rate b_1 and autoregressive parameter γ were considered known in this study. We then considered two cases, (VB1) one where $\boldsymbol{\Sigma}_Q^{-1}$ is set to the identity matrix (which seemed a reasonable fit on visualization of the true precision matrix) and (VB2) one where $\boldsymbol{\Sigma}_Q^{-1}$ is assumed to be fully unknown, with the aim of showing that estimating the spatially varying volatility may indeed lead to better a-posteriori inference of the intensity function. In both cases both the latent field and the growth (through $\boldsymbol{\vartheta}$) were considered unknown, the latter being equipped with a Gaussian prior with parameters

$$\hat{\boldsymbol{\vartheta}}_p = \mathbf{0}, \qquad \boldsymbol{\Sigma}_{\boldsymbol{\vartheta},p} = \boldsymbol{I}. \qquad \textbf{[S66]}$$

In VB2 the precision matrix was assigned a Wishart prior with degrees of freedom and scale matrix given by

$$d_p = 10, \qquad \boldsymbol{V}_p = 0.2\boldsymbol{I}. \qquad \textbf{[S67]}$$

For the purpose of this study both priors can be considered largely uninformative. The VB expectation maximization (VBEM) algorithm was run for 20 iterations until deemed to have converged. In addition to the VBEM algorithm we also implemented a standard kernel estimator (KE) of the intensity function using homogeneous, but anisotropic spatiotemporal kernels of the form $k(s_1, s_2, t) = \exp(-(s_1 - \mu_1)^2/2\sigma_s^2 - (s_2 - \mu_2)^2/2\sigma_s^2 - (t - \mu_3)^2/2\sigma_t^2)$ with $\sigma_s \in [0.1, 0.2, 0.3, 0.5, 0.8]$ and $\sigma_t \in [1, 1.5, 2, 2.5]$. Volumetric correction was carried out when $\mu_1 \pm 3\sigma_s$, $\mu_2 \pm 3\sigma_s$ or/and $\mu_3 \pm 3\sigma_t$ lay outside the spatiotemporal domain $\mathcal{O}_{\mathcal{ST}} = [0\ 18] \times [0\ 18] \times [0\ 300]$.

SI results. Estimation of $\mu_Q(\boldsymbol{s})$ and the volatility map $\sigma(\boldsymbol{s})$ by VB2 are shown in Fig. S9 *E* and *F* respectively showing a relatively good agreement with Fig. S9 *B* and *C* respectively. Crucially, regions of high volatility are detected independently from the overall growth rate in the region.

A comment is due on regions where the algorithm does not perform so well, in particular in areas which exhibit a considerable low event count (this can be seen by comparing to S9*A*). In these regions, quantifying parameters such as growth and volatility clearly becomes a futile task; even more so given the exponential form of the intensity function where low intensity fields correspond to arbitrarily negative latent fields. Clearly spatially selective priors can be introduced to remedy this problem. However, in line with existing spatiotemporal literature; e.g., (9, 10), with the AWD we have opted to simply concentrate modelling effort in represented areas. This approach also results in obvious computational savings.

A natural question to ask is whether estimating the spatially varying volatility contributes to an increase in accuracy to the latent field and growth variational densities obtained on convergence. To answer this question we computed (*i*) the mean square error (MSE) between the true intensity and the a-posteriori median intensity given by

$$\text{MSE}_\lambda = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{J^2}\sum_{\boldsymbol{s}\in\mathcal{O}_J}(\lambda_k^*(\boldsymbol{s}) - \exp(\boldsymbol{\phi}(\boldsymbol{s})^T\hat{\boldsymbol{x}}_{k|K}))\right), \qquad \textbf{[S68]}$$

where $*$ is used to denote the true value, J^2 is the number of spatial points over which the error is computed (10,000) and $\mathcal{O}_J$ a correspondingly gridded spatial domain with side length J, (*ii*) the MSE between the exponentiated growth field and the a-posteriori median exponentiated growth (with the exponential applied in this case so as not to penalize for excessively negative growth rates in regions witnessing scarce events)

$$\text{MSE}_{\mu_Q} = \frac{1}{J^2}\sum_{\boldsymbol{s}\in\mathcal{O}_J}(\mu_Q^*(\boldsymbol{s}) - \exp(\boldsymbol{\phi}(\boldsymbol{s})^T\hat{\boldsymbol{\vartheta}})), \qquad \textbf{[S69]}$$

and finally (*iii*) the bias in the estimated probabilities of the intensity function computed as

$$\text{bias}_q = q - \frac{1}{KJ^2}\sum_{k=1}^{K}\sum_{\boldsymbol{s}\in\mathcal{O}_J}\mathbb{I}[z_k^*(\boldsymbol{s}) \le z_k^q(\boldsymbol{s})], \qquad \textbf{[S70]}$$

where $\mathbb{I}$ is the indicator function, q is a quantile and $\tilde{p}(\boldsymbol{\phi}(\boldsymbol{s})^T\boldsymbol{x}_k \le z_k^q(\boldsymbol{s})) = q$ so that for each $k, \boldsymbol{s}$ $z_k^q(\boldsymbol{s})$ denotes the

q^{th} quantile as supplied by the variational density (for complete details refer to ref. 20). For a comparison between VB1 and VB2, the mean of bias_q evaluated for selected values of $q(q \in [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99])$ will be used.

Results of the analysis shown in Table S1 where it is immediately evident that estimation of the precision matrix contributes to a considerable reduction in overall intensity MSE, growth MSE, and bias. Further, inspection on a frame-by-frame basis showed that in an overwhelming 93% of the times the intensity MSE given by VB2 was less than that of VB1, thus firmly establishing the importance of correct variance estimation in the methodology. Finally, the nonparametric estimator is seen to give reasonable estimates (only the lowest MSE from all attempted kernels is shown in Table S1, corresponding to $\sigma_s = 0.2$, $\sigma_t = 1$). However, and although it is possible that here the MSE can be further reduced using adaptive methods (18), the key limiation of the KE method remains its inability to (*i*) capture uncertainty in a systematic way and (*ii*) to provide a mechanistic description of the underlying process, both limitations which modern spatiotemporal methods such as that presented here overcome.

1. Stoyan D, Stoyan H (1994) *in Fractals, Random Shapes, and Point Fields: Methods of Geometrical Statistics* (Wiley, New York), pp 275–305.
2. Brix A, Moeller J (2001) Space-time multi type log Gaussian Cox processes with a view to modelling weeds. *Scand Stat Theory Appl* 28:471–488.
3. Moeller J, Waagepetersen R (2004) *in Statistical Inference and Simulation for Spatial Point Processes* (CRC Press, Boca Raton), pp 29–55.
4. Lewis PAW, Shedler GS (1979) Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly* 26:403–413.
5. Shachtman N (2010) What I saw at Moba Khan: the military reports highlighted by WikiLeaks don't provide a full picture of the war. *Wall Street Journal*, http://online.wsj.com/article/SB10001424052748703977004575393523349648264.html [online: last accessed 23 June 2012].
6. O'Loughlin J, Witmer FDW, Linke AM, Thorwardson N (2010) Peering into the fog of war: the geography of the WikiLeaks Afghanistan war logs, 2004–2009. *Eurasian Geogr Econ* 51:472–495.
7. Raleigh C, Linke A, Hegre H, Karlsen J (2010) Introducing ACLED: an armed conflict location and event dataset. *J Peace Res* 47:651–660.
8. Sanner R, Slotine J (1992) Gaussian networks for direct adaptive control. *IEEE Trans Neural Netw* 3:837–863.
9. Crainiceanu CM, Diggle P, Rowlingson B (2008) Bivariate binomial spatial modeling of Loa loa prevalence in tropical Africa. *Amer Statist Assoc* 103:21–37.
10. Stroud JR, Mueller P, Sanso B (2001) Dynamic models for spatiotemporal data. *J R Stat Soc Series B Stat Methodol* 63:673–689.
11. Rodrigues A, Diggle P (2010) A class of convolution-based models for spatiotemporal processes with nonseparable covariance structure. *Scand Stat Theory Appl* 37:553–567.
12. Wikle CK, Cressie NAC (1999) A dimension-reduced approach to space-time Kalman filtering. *Biometrika* 86:815–829.
13. Berliner LM, Wikle CK, Cressie NAC (2000) Long-lead prediction of Pacific SSTs via Bayesian dynamic modeling. *Journal of Climate* 13:3953–3968.
14. Weidmann NB, Ward MD (2010) Predicting conflict in space and time. *J Conflict Resolut* 54:883–901.
15. Beal MJ (2003) *in Variational Algorithms for Approximate Bayesian Inference*. Ph.D. thesis (Gatsby Computational Neuroscience Unit, University College London, United Kingdom), pp 159–205.
16. Zammit Mangion A, Yuan K, Kadirkamanathan V, Niranjan M, Sanguinetti G (2011) Online variational inference for state-space models with point-process observations. *Neural Comput* 23:1967–1999.
17. Diggle P (1985) A kernel method for smoothing point process data. *Appl Stat* 34:138–147.
18. Diggle P, Rowlingson B, Su T (2005) Point process methodology for on-line spatio-temporal disease surveillance. *Environmetrics* 16:423–434.
19. Smidl V, Quinn A (2005) *The Variational Bayes Method in Signal Processing* (Springer Verlag, New York).
20. Taylor BM, Diggle PJ (2012) INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. http://arxiv.org/pdf/1202.1738v2.pdf [online: last accessed 12 April 2012].

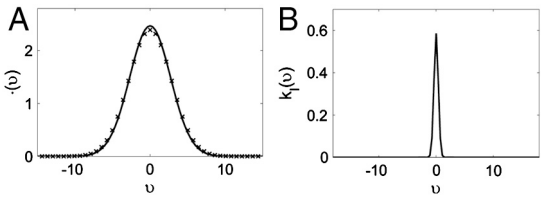


Fig. S1. (*A*) Average log PACF $\ln \bar{g}_{k,k}(v)$ (–) and average log PCCF $\ln \bar{g}_{k,k+1}(v)$ (×). (*B*) $\hat{k}_l$ as computed from Eq. **S20**.

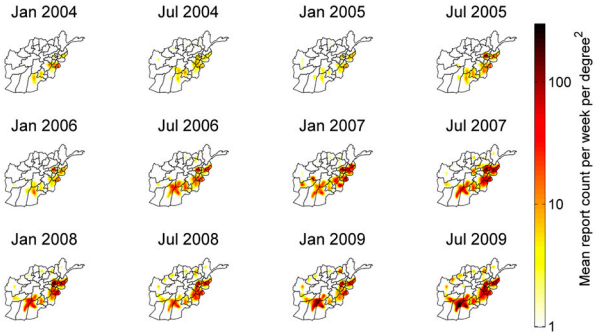


Fig. S2. Estimated mean intensity $\mathbb{E}[\lambda_k(\boldsymbol{s})]$ on the first week of the month and respective year.

Fig. S3. Provincial map of Afghanistan (accurate as from 2010).

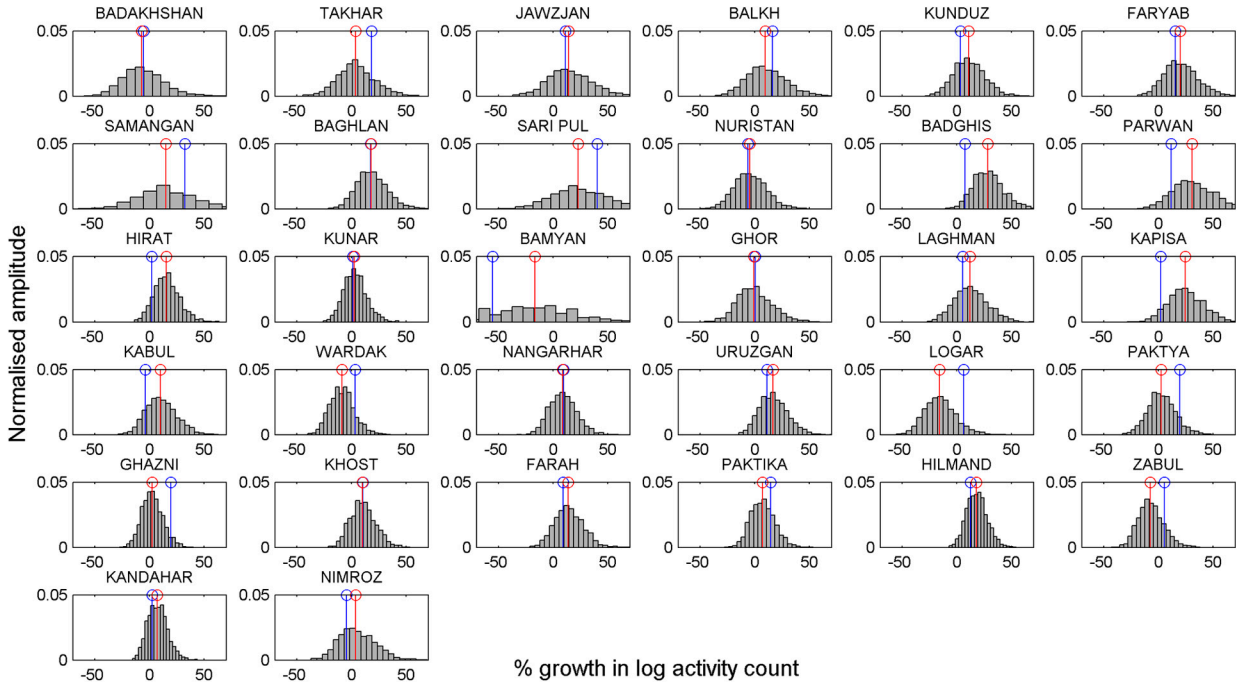


Fig. S4. Normalized histograms of log AOG activity count in 2010 per province as obtained from MC simulations, shown together with true growth (circle, blue) and sample median (circle, red). The two stems overlap considerably in Baghlan, Kunar, Ghor, Nangarhar, and Khost.

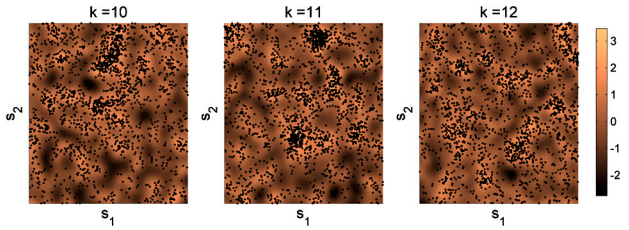


Fig. S5. Three time frames from a single realization of a spatiotemporal point process, where the intensity function exponentiates a field evolving according to the SIDE of Eq. **S1** with $k_I(\upsilon) = 0.05\exp(-||\upsilon||^2)$ and $k_Q(\upsilon) = 0.8\exp(-||\upsilon||^2/5)$.

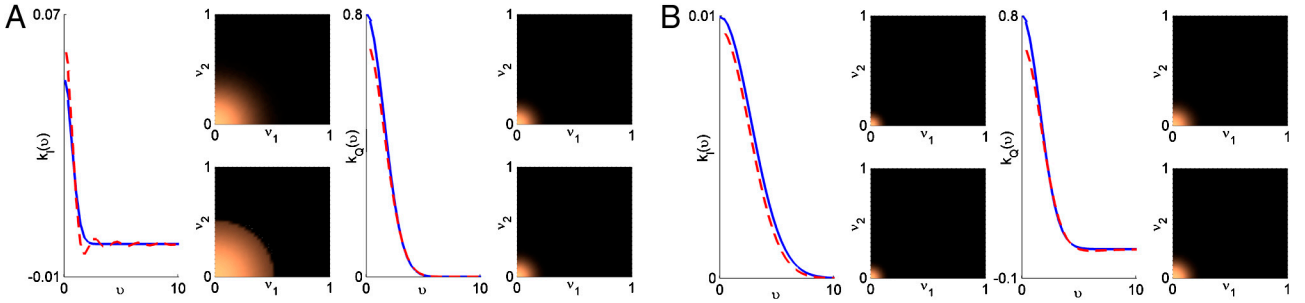


Fig. S6. Line plots: true (–) and nonparametric estimates (- -) of $k_I(\upsilon)$ and $k_Q(\upsilon)$ from point process observations with (*A*) $k_I(\upsilon) = 0.05\exp(-||\upsilon||^2)$, $k_Q(\upsilon) = 0.8\exp(-||\upsilon||^2/5)$ and (*B*) $k_I(\upsilon) = 0.01\exp(-||\upsilon||^2/15)$, $k_Q(\upsilon) = 0.8\exp(-||\upsilon||^2/5)$. Upper surface plots: spatial frequency response of the true kernels. Lower surface plots: spatial frequency response of the estimated kernels.

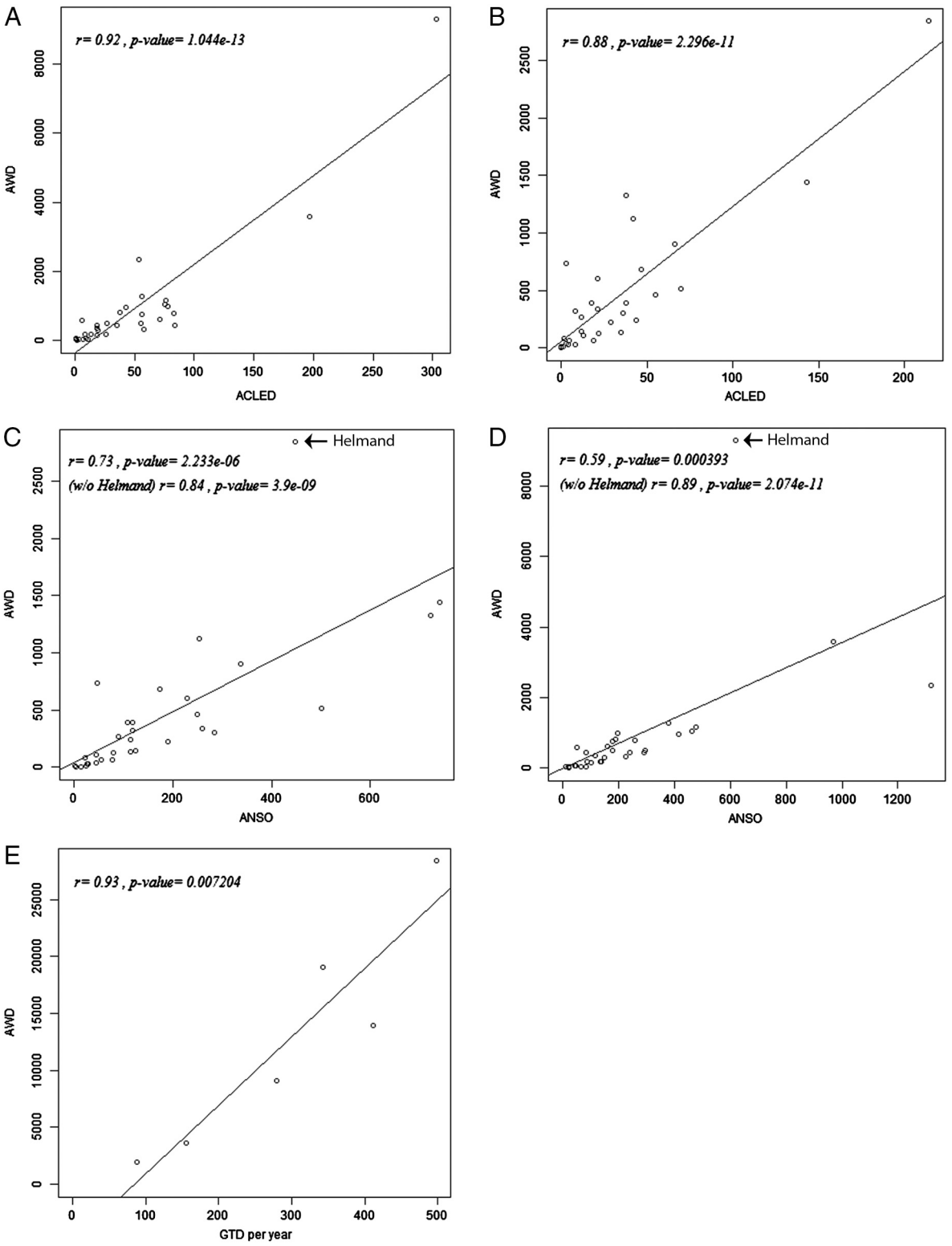


Fig. S7. Correlation between the event incidence in the AWD and other datasets. (*A*) ACLED 2008 on a province-by-province basis. (*B*) ACLED 2009 on a province-by-province basis. (*C*) ANSO AOG initiated attacks 2008 on a province-by-province basis. (*D*) ANSO AOG initiated attacks 2009 on a province-by-province basis. (*E*) GTD 2004–2009 on a year-by-year basis.

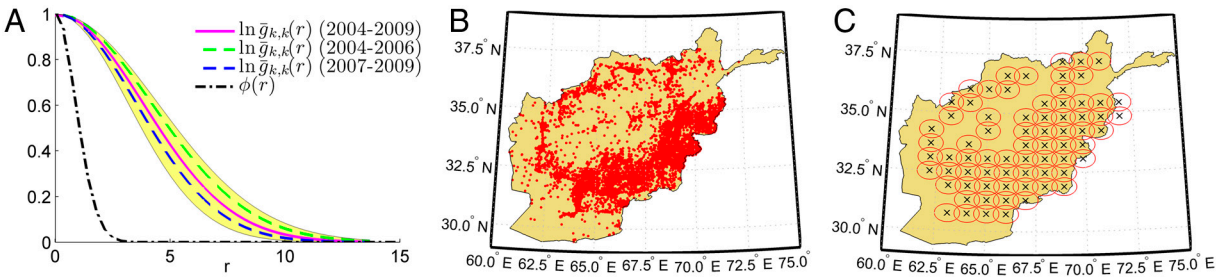


Fig. S8. (*A*) Average log PACF $\bar{g}_{k,k}(r)$ as a function of radial distance r between 2004 and 2009 and cross-section of the isotropic basis function employed in study. Here one unit in r corresponds to approximately 0.4 degrees (latitude/longitude). (*B*) Spatial location of of all logged events between 2004 and 2009. Of the roughly 77,000 logs constituting the AWD, the 75,676 located within Afghanistan's borders were considered in the analysis. (*C*) Basis function placement in spatial domain with the red contours denoting the 1σ mark. Functions were omitted in regions within the country (such as in the extreme North East and South West) which contain few, sparse events. These events which are few and far between were instead captured with the use of a background activity baseline

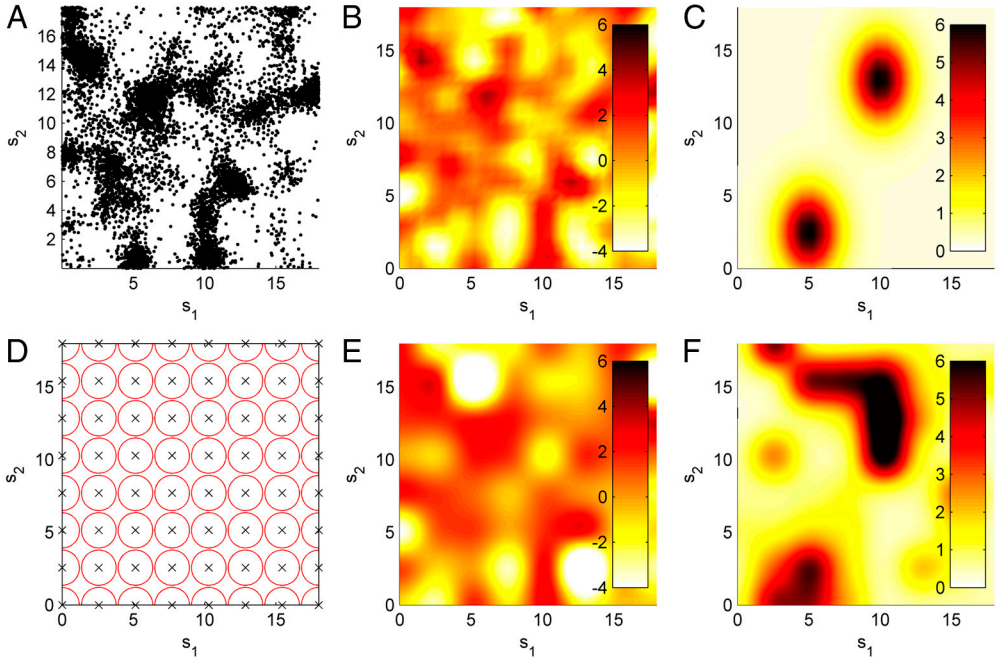


Fig. S9. (*A*) Superposition of the first 50 spatial point patterns in the simulation study. (*B*) True growth function $\mu_Q(\boldsymbol{s})$. (*C*) True variance map $\sigma(\boldsymbol{s})^2$. (*D*) Basis function placement in simulation study with the red contours denoting the 1σ mark. (*E*) Mean a-posteriori estimate of $\mu_Q(\boldsymbol{s})$. (*F*) Mean a-posteriori estimate of $\sigma(\boldsymbol{s})^2$. Note how mismatches occur in regions of very low event count, as expected.

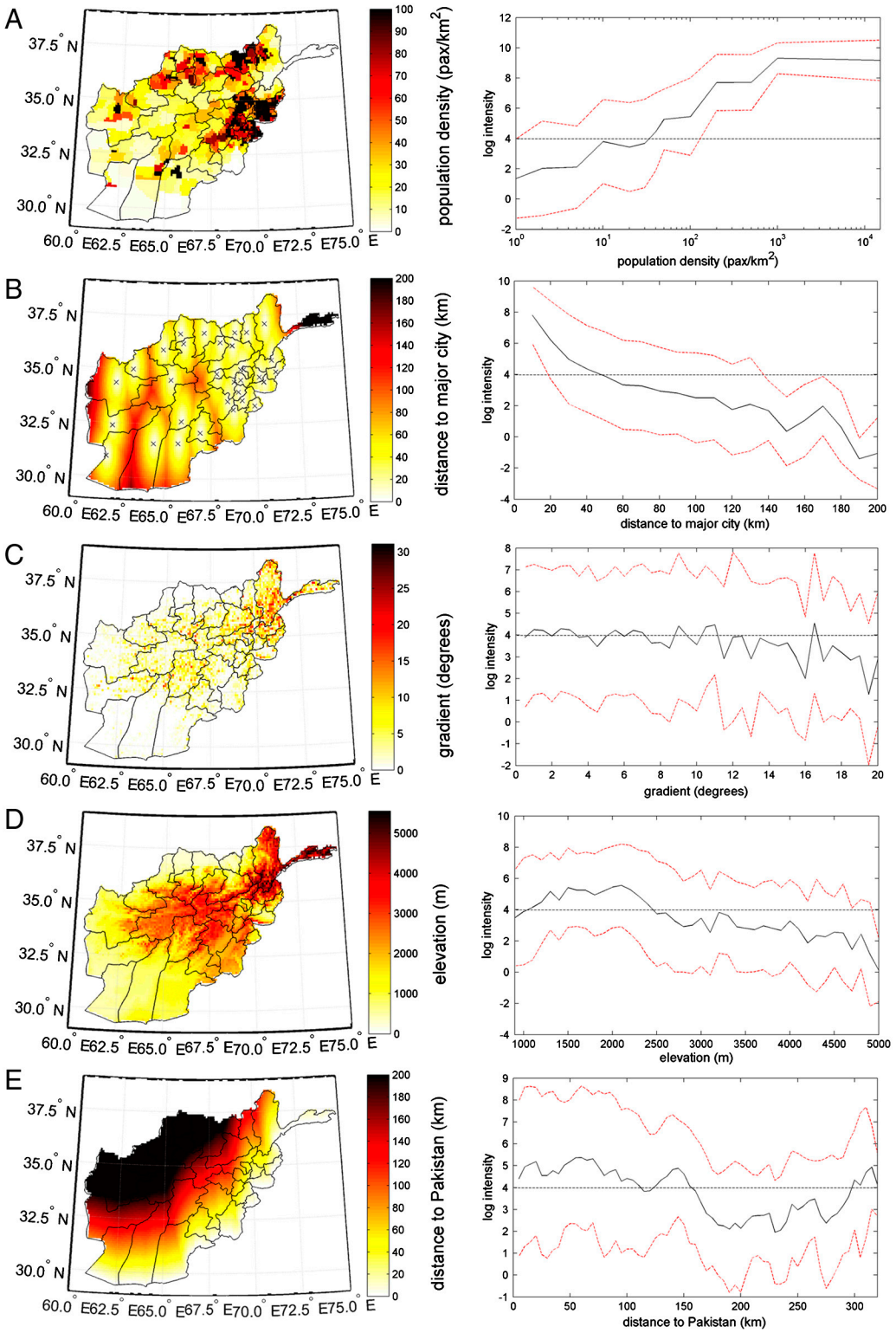
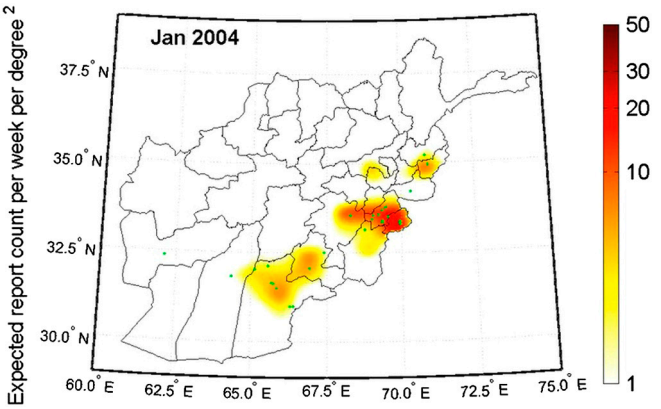
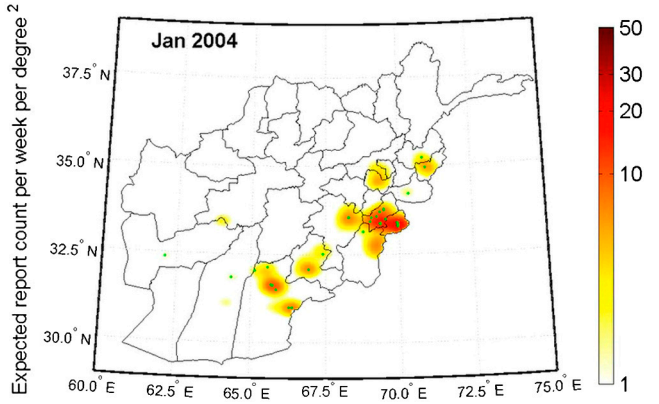


Fig. S10. Spatial maps (left) and empirical relationships (right) between the independent variable and the log spatial intensity of the AWD (–) together with the 1σ interval (red lines, - -) and global mean (- -). (*A*) Population density. (*B*) Distance to closest major city. (*C*) Terrain slope. (*D*) Elevation. (*E*) Distance to Pakistan border. All studies were carried out on a 200 × 200 spatial grid.



Movie. S1. A-posteriori intensity of the AWD between 2004 and 2009 at regular intervals of one week. The animation shows strong evidence for a consistent trend of increasing activity in Afghanistan, with particular growth in the Southern provinces from 2006 onwards (MPG; 13.7 MB)

Movie S1 (MPG)



Movie S2. Kernel intensity estimation of the AWD between 2004 and 2009, showed at regular intervals of one week. For this study, the homogeneous kernel chosen was of the form $\exp(-s_1^2/0.5 - s_2^2/0.5 - t^2/2)$ (MPG; 14.7 MB)

Movie S2 (MPG)

Table S1. Results of simulation study of SI Text S10

	MSE_{λ}	MSE_{μ_Q}	mean($bias_q$)
VB1	1.939	0.176	−0.099
VB2	1.699	0.161	−0.063
KE	2.324	N.A.	N.A.

Algorithm S1: Analysis for dynamic, homogeneous, isotropic spatiotmporal point processes

1. Estimate $\lambda_k^{(1)}(s)\forall k$ using Eq. **S13** for stationary systems or simple regression where clear trends (linear or otherwise) are evident.
2. Estimate $\hat{g}_{k,k}(v)$, $\hat{g}_{k,k+1}(v)\forall k$ from Eq. **S16** and Eq. **S17** respectively.
3. Estimate $\hat{k}_I(v)$ from Eq. **S20** using $\bar{g}_{k,k}(v)$ and $\bar{g}_{k,k+1}(v)$ from Eqs. **S18–S19**.
4. Estimate $\hat{k}_Q(v)$ from Eq. **S2** using $\hat{k}_I$ and $\bar{g}_{k,k}(v)$ from Eq. **S18**.

Algorithm S2: VB-Laplace smoother for the AWD model (Note: Integrations are carried out by numerical quadrature).

Time interval $\Delta_t = 1$ is assumed throughout. Expectations are
taken with respect to the relevant distributions.)
Input: Data set $\mathcal{Y}_K$, parameters $b, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0$ and parameter distributions $\tilde{p}(\vartheta)$, $\tilde{p}(\boldsymbol{\Sigma}_Q^{-1}) = \tilde{p}(\boldsymbol{Q})$.
Forward message
Set $\hat{\boldsymbol{x}}_{0|0} = \boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_{0|0} = \boldsymbol{\Sigma}_0$.
for $k = 1$ **to** K
$\boldsymbol{\Sigma}_{k-1}^{*} = (\boldsymbol{\Sigma}_{k-1|k-1}^{-1} + \mathbb{E}[\boldsymbol{Q}])^{-1}$
$\tilde{\boldsymbol{\Sigma}}_k = (\mathbb{E}[\boldsymbol{Q}] - \mathbb{E}[\boldsymbol{Q}]\boldsymbol{\Sigma}_{k-1}^{*}\mathbb{E}[\boldsymbol{Q}])^{-1}$
$\tilde{\boldsymbol{x}}_k = \tilde{\boldsymbol{\Sigma}}_k\left[\mathbb{E}[\boldsymbol{Q}]\boldsymbol{\Sigma}_{k-1}^{*}(\boldsymbol{\Sigma}_{k-1|k-1}^{-1}\hat{\boldsymbol{x}}_{k-1|k-1} - \mathbb{E}[\boldsymbol{Q}]\mathbb{E}[\vartheta]) + \mathbb{E}[\boldsymbol{Q}]\mathbb{E}[\vartheta]\right]$.
$\hat{\boldsymbol{x}}_{k|k} = \arg\max_{\boldsymbol{x}_k} \sum_{\boldsymbol{s}_j \in \boldsymbol{y}_k}(\mathbb{E}[b^T d(\boldsymbol{s}_j)] + \boldsymbol{\phi}(\boldsymbol{s}_j)^T x_k) - \int_{\mathcal{O}}\mathbb{E}[\exp(b^T d(\boldsymbol{s}))]\exp(\boldsymbol{\phi}^T(\boldsymbol{s})x_k)\mathrm{d}\boldsymbol{s} - \frac{1}{2}(x_k - \tilde{x}_k)^T\tilde{\boldsymbol{\Sigma}}^{-1}(x_k - \tilde{x}_k)$
$\boldsymbol{\Sigma}_{k|k} = (\tilde{\boldsymbol{\Sigma}}_k^{-1} + \int_{\mathcal{O}}\boldsymbol{\phi}(\boldsymbol{s})\boldsymbol{\phi}(\boldsymbol{s})^T\exp(\boldsymbol{\phi}(\boldsymbol{s})^T x_{k|k})\mathbb{E}[\exp(b^T\mathrm{d}(\boldsymbol{s}))]d\boldsymbol{s})^{-1}$
end for
Backward message
Set $\boldsymbol{\Sigma}_{K|K+1:K}^{-1} = \boldsymbol{0}$ (ignore estimate of end condition)
for $k = (K-1)$ **down to** 0
$x'_{k+1} = \arg\max_{\boldsymbol{x}_{k+1}} \sum_{\boldsymbol{s}_j \in \boldsymbol{y}_{k+1}}(\mathbb{E}[b^T d(\boldsymbol{s}_j)] + \boldsymbol{\phi}(\boldsymbol{s}_j)^T x_{k+1}) - \int_{\mathcal{O}}\mathbb{E}[\exp(b^T d(\boldsymbol{s}))]\exp(\phi^T(s)x_{k+1})\mathrm{d}\boldsymbol{s} - \frac{1}{2}(x_{k+1} - \hat{x}_{k+1|k+2:K})^T\tilde{\boldsymbol{\Sigma}}^{-1}(x_{k+1} - \hat{x}_{k+1|k+2:K})$
$\boldsymbol{\Sigma}'_{k+1} = \left(\boldsymbol{\Sigma}_{k+1|k+2:K}^{-1} + \int_{\mathcal{O}}\boldsymbol{\phi}(\boldsymbol{s})\boldsymbol{\phi}(\boldsymbol{s})^T\exp(\boldsymbol{\phi}(\boldsymbol{s})^T x'_{k+1})\mathbb{E}[\exp(b^T\mathrm{d}(\boldsymbol{s}))]d\boldsymbol{s}\right)^{-1}$
$\boldsymbol{\Sigma}_{k|k+1:K} = (\mathbb{E}[\boldsymbol{Q}] - \mathbb{E}[\boldsymbol{Q}](\boldsymbol{\Sigma}'^{-1}_{k+1} + \mathbb{E}[\boldsymbol{Q}])^{-1}\mathbb{E}[\boldsymbol{Q}])^{-1}$
$x_{k|k+1:K} = \boldsymbol{\Sigma}_{k|k+1:K}(-\mathbb{E}[\boldsymbol{Q}]\mathbb{E}[\vartheta] + \mathbb{E}[\boldsymbol{Q}](\boldsymbol{\Sigma}'^{-1}_{k+1} + \mathbb{E}[\boldsymbol{Q}])^{-1}(\boldsymbol{\Sigma}'^{-1}_{k+1}\boldsymbol{x}'_{k+1} + \mathbb{E}[\boldsymbol{Q}]\mathbb{E}[\boldsymbol{\vartheta}]))$
end for
Smoothed estimate
for $k = 0$ **to** K
$\boldsymbol{\Sigma}_{k|K} = (\boldsymbol{\Sigma}_{k|k}^{-1} + \boldsymbol{\Sigma}_{k|k+1:K}^{-1})^{-1}$
$\hat{x}_{k|K} = \boldsymbol{\Sigma}_{k|K}[\boldsymbol{\Sigma}_{k|k}^{-1}\hat{x}_{k|k} + \boldsymbol{\Sigma}_{k|k+1:K}^{-1}\hat{x}_{k|k+1:K}]$
end for
Computation of cross-covariance $\{\boldsymbol{M}_k\}_{k=1}^{K}$
for $k = K$ **down to** 1
$M_{k|K} = \boldsymbol{\Sigma}_{k-1}^{*}\mathbb{E}[Q][\boldsymbol{\Sigma}_{k|k+1:K}^{-1} + \mathbb{E}[Q] + \int_{\mathcal{O}}\boldsymbol{\phi}(\boldsymbol{s})\boldsymbol{\phi}(\boldsymbol{s})^T\exp(\boldsymbol{\phi}(\boldsymbol{s})^T\hat{x_{k|K}})\mathbb{E}[\exp(b^T d(\boldsymbol{s}))]\mathrm{d}\boldsymbol{s} - \mathbb{E}[Q]\boldsymbol{\Sigma}_{k-1}^{*}\mathbb{E}[Q]^{-1}$
end for
Output: $\{\hat{x}_{k|K}, \boldsymbol{\Sigma}_{k|K}\}_{k=0}^{K}$, $\{M_{k|K}\}_{k=1}^{K}$..

Other Supporting Information

Dataset S1 (CSV)