A. Appendix

A.1. Kronecker Algebra

We exploit the identity (Steeb & Hardy, 2011):

$$(B^\top \otimes A)v = \text{vec}(AVB) \tag{A22}$$

where $v = \text{vec}(V)$ and the vec operator turns a matrix into a vector by stacking columns vertically. Since a full $n \times n$ matrix is never formed, this approach is very efficient in terms of space and time complexity, relying only on operations with the smaller matrices K_i and the matrix V which only has n entries. We analyzed the complexity in Section 6.5. Another result we use is that given the eigendecompositions of $K_d = Q_d \Lambda_d Q_d^T$, we have:

$$K = (\bigotimes Q_d)(\bigotimes \Lambda_d)(\bigotimes Q_d^T) \tag{A23}$$

A.2. Supplementary Results

q	Weight	Period	Length-scale
1	52.72	10813.9	133280.2
2	5.48	4.0	1.1
3	0.33	52.1	27700.8
4	0.05	22.0	1.6
5	0.02	97.4	7359.1

Table A2. The top five spectral mixture components learned for the temporal kernel in the LGCP fit to 8 years of assault data. The components are visualized in Figure 4 where component q corresponds to the row of the table.

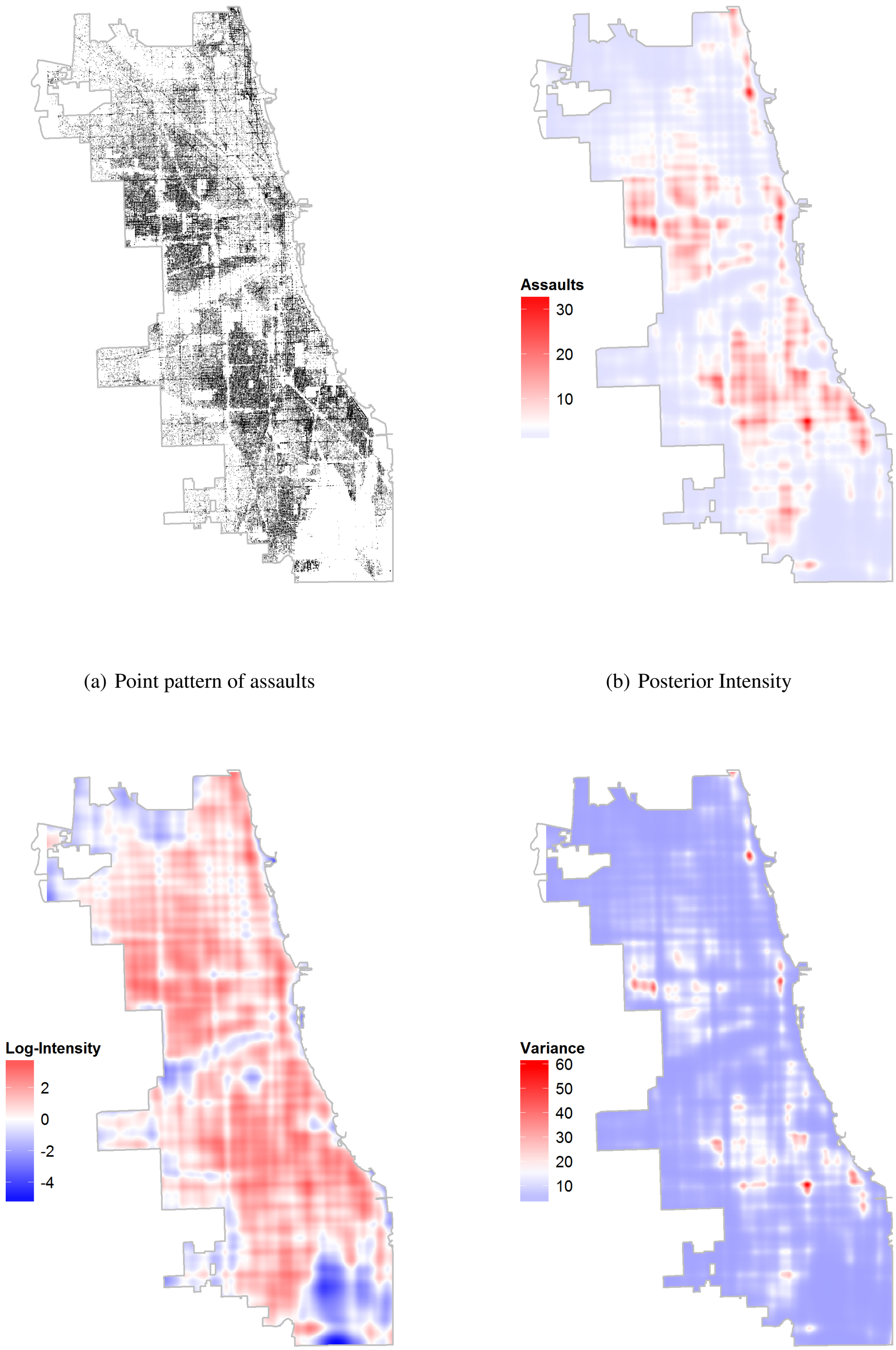
N	Standard	Kronecker	FITC-100
125	-62.12	-61.52	-61.20
343	-157.47	-157.80	-159.21
1000	-445.48	-443.87	-455.84
1728	-739.56	-740.31	-756.95
8000	-3333.10	-3333.66	-3486.20

Table A3. Predictive log-likelihoods are shown corresponding to the experiment in Figure 2. A higher log-likelihood indicates a better fit. The differences between the standard and Kronecker results were not significant but the difference between FITC-100 and the others was significant (two-sample paired t-test, $p \leq .05$) for $n \geq 1000$.

A.3. A two-dimensional LGCP

We used a product of Matérn-$5/2$ kernels: $k_x(d)$ with length-scale λ_x and variance σ^2 and $k_y(d)$ with length-scale λ_y and variance fixed at 1: $k((x,y),(x',y')) = k_x(|x-x'|)k_y(|y-y'|)$.

We discretized our data into a 288×446 grid for a total of 128,448 observations. Locations outside of the boundaries of Chicago – about 56% of the full grid—were treated as missing. In Figure A5 we show the location of assaults represented by dots, along with a map of our posterior intensity, log-intensity, and variance of the number of assaults. It is clear that our approach is smoothing the data. The hyperparameters that we learn are $\sigma^2 = 5.34$, $\lambda_x = 2.23$, and $\lambda_y = 2.24$, i.e., length-scales for moving north-south and east-west were found to be nearly identical for these data; by assuming Kronecker structure our learning happens in a fashion analogous to Automatic Relevance Determination (Neal, 1996).



(a) Point pattern of assaults

(b) Posterior Intensity

(c) Posterior Latent Log-Intensity

(d) Posterior Variance

Figure A5. We fit a log Gaussian Cox Process to the point pattern of reported incidents of assault in Chicago (a) and made posterior estimates of the intensity surface (b). The latent log-intensity surface is visualized in (c) and the posterior variance is visualized in (d).

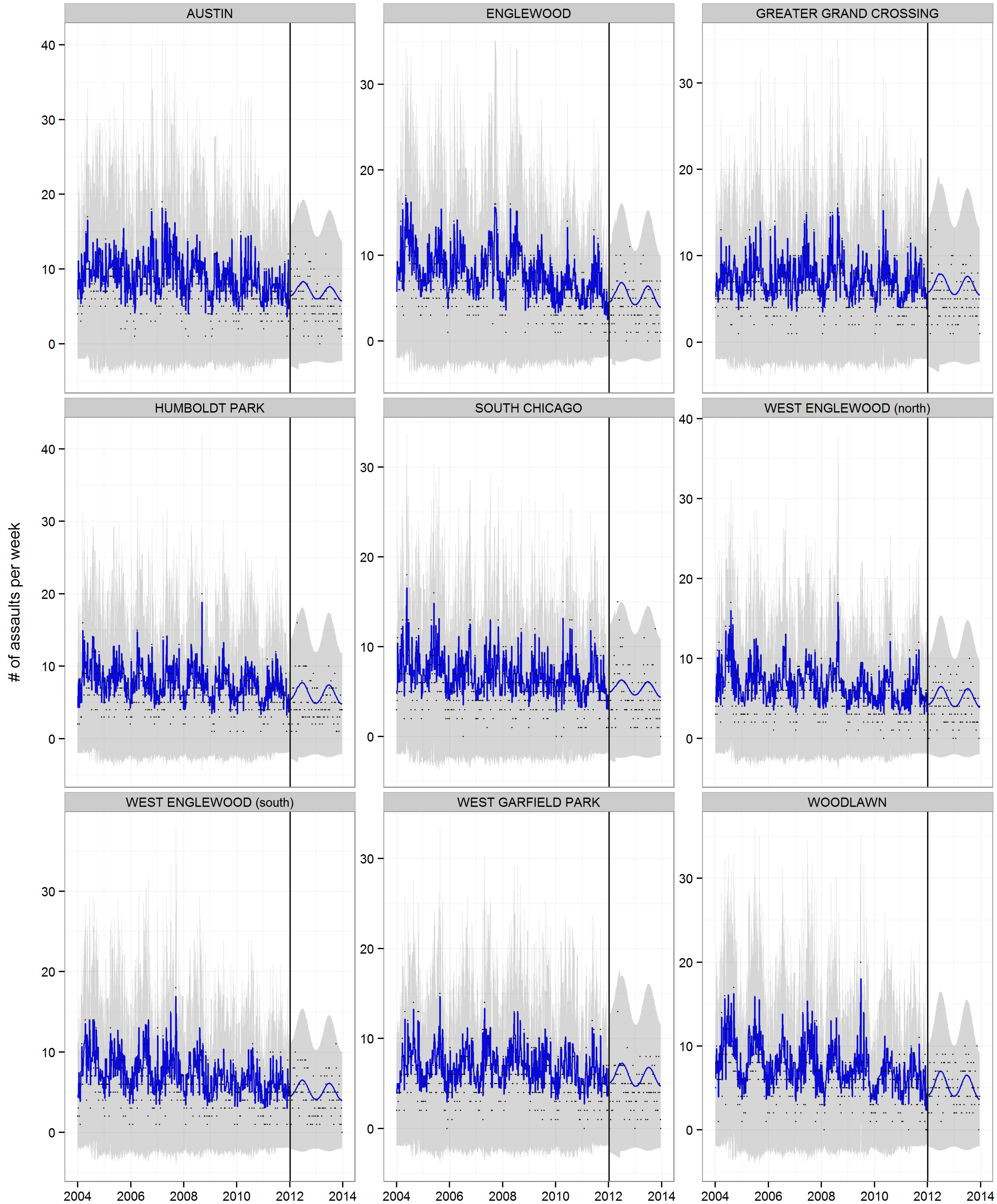


Figure A6. We show the time series of weekly assaults in the nine neighborhoods with the most assaults in Chicago. The blue line shows our posterior prediction (training data, first 8 years of data) and forecast (out-of-sample, last 2 years of data, to the right of the vertical bar). Observed counts are shown as dots. 95% posterior intervals are shown in gray.