geographical analysis

Geographical Analysis (2016) **00,** 00–00

A Spatiotemporal Bayesian Hierarchical Approach to Investigating Patterns of Confidence in the Police at the Neighborhood Level

Dawn Williams [1], James Haworth [1], Marta Blangiardo [2], Tao Cheng [1]

[1] SpaceTimeLab for Big Data Analytics, University College London, London, United Kingdom,
[2] Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom

Public confidence in the police is crucial to effective policing. Improving understanding of public confidence at the local level will better enable the police to conduct proactive confidence interventions to meet the concerns of local communities. Conventional approaches do not consider that public confidence varies across geographic space as well as in time. Neighborhood level approaches to modeling public confidence in the police are hampered by the small number problem and the resulting instability in the estimates and uncertainty in the results. This research illustrates a spatiotemporal Bayesian approach for estimating and forecasting public confidence at the neighborhood level and we use it to examine trends in public confidence in the police in London, UK, for Q2 2006 to Q3 2013. Our approach overcomes the limitations of the small number problem and specifically, we investigate the effect of the spatiotemporal representation structure chosen on the estimates of public confidence produced. We then investigate the use of the model for forecasting by producing one-step ahead forecasts of the final third of the time series. The results are compared with the forecasts from traditional time-series forecasting methods like naïve, exponential smoothing, ARIMA, STARIMA, and others. A model with spatially structured and unstructured random effects as well as a normally distributed spatiotemporal interaction term was the most parsimonious and produced the most realistic estimates. It also provided the best forecasts at the London-wide, Borough, and neighborhood level.

Introduction

Public confidence in the police is a state in which the public regard the police as competent and capable of fulfilling their roles (Hohl, Stanko, and Newburn 2012). This includes engagement with the community, fair treatment, and effectiveness in dealing with crime and antisocial

Correspondence: Dawn Williams, SpaceTimeLab for Big Data Analytics, University College London, London, United Kingdom.
e-mail: dawn.williams.10@ucl.ac.uk

Submitted: June 01, 2017. Revised version accepted: January 20, 2018.

doi: 10.1111/gean.12160

behavior (Hohl, Stanko, and Newburn 2012). Higher public confidence in policing fosters a better relationship between the police and the public, improving the likelihood of members of the public coming forward with tips, obeying the law, or joining the force as volunteers (Tyler and Fagan 2008; Stanko and Bradford 2009; Association of Chief Police Officers 2012). The British model of policing is underpinned by a philosophy of "policing by consent." In this approach, the police are empowered by the common consent of the public (Home Office 2012). The cooperation of the public in observing the law results from their approval, respect, and affection for the police rather than compliance motivated by fear (Home Office 2012). Within this context, public confidence is a vital component of effective policing (Jackson et al. 2013).

Improving public confidence in policing is one of the main concerns of the London metropolitan police service (MPS) and the body which oversees them, the Mayor's office for policing and crime (MOPAC). MOPAC surveys attitudes toward the police in a large scale, rolling survey; the public attitudes survey (PAS). The survey is designed to be representative of the residents of London annually at the borough level (BMG Research 2012) and is too small to allow reliable direct estimation at the neighborhood level. However, neighborhood policing initiatives are central to the strategy for improving public confidence so, in the absence of more data, it is a pressing concern to make the data useable at the small area level (Greenhalgh et al. 2015).

Literature on public confidence in the police is rich. In the UK context, the importance of the role public confidence in the police plays in effective policing is well supported. An evidence-based criminological theory of public confidence was developed using MOPAC PAS survey data (Stanko et al. 2012). This confidence model has been adopted by MOPAC/MPS and informs how confidence is analyzed and reported today. Recently, the focus has shifted toward the effect of the neighborhood context on public confidence in the police. Jackson et al. (2013) investigated the impact of the social neighborhood characteristics (e.g., collective efficacy and fear of crime) and structural neighborhood characteristics (e.g., deprivation and ethnic composition) on public confidence and police legitimacy.

The literature and corresponding theory on predictive policing is recent and growing. Work in this area has largely surrounded predictive hotspot mapping, where maps of areas likely to contain high crime densities are produced (Bowers, Johnson, and Pease 2004; Kulldorff et al. 2005; Chainey, Tompson, and Uhlig 2008; Mohler et al. 2011). Bayesian methods are more frequently applied to estimation of crime patterns rather than forecasting. Space-time Bayesian methods have been applied to estimation of property crime at the local level (Law, Quick, and Chan 2013), burglary risk (Li et al. 2014), and violent crime (Law, Quick, and Chan 2015). However, Aldor-Noiman et al. (2016) successfully used a Bayesian nonparametric approach to forecasting low counts of violent crime in Washington. Comparatively, little work has been done so far in predictive public confidence modeling. This is likely due to the lack of data and less perceived importance to "softer" aspects of policing. In the UK context, (Sindall, Sturgis, and Jennings 2012) have examined the temporal aspect by using the auto-regressive integrated moving average (ARIMA) model put forward by (Box and Jenkins 1976) to analyze public confidence at the country-wide level. However, their approach does not take into account that public confidence varies across space as well as in time as shown in (Williams, Haworth, and Cheng 2015). To the best of our knowledge, there has been no study to date that attempts to model the spatio-temporal variation in public confidence at the small area level.

This article seeks to address three main research questions through its illustration of a Bayesian spatiotemporal approach to investigating trends in public confidence in the police.

First, can the spatial and temporal trends in public confidence in the police be uncovered at the city-wide and neighborhood levels? Second, how does choice of spatiotemporal representation structure affect the trends uncovered? Third, how does the predictive ability of the spatiotemporal Bayesian hierarchical approach compare with traditional time-series forecasting methods?

Spatiotemporal estimation and forecasting for public opinion

Statistical modeling makes inferences about a population from a subsample (Iversen 1984). This study is concerned with the spatiotemporal modeling of discrete survey data that was subsampled in repeated cross-sections, that is, several independent subsamples were taken at different points in time. Since only a small percentage of the population was sampled, this data can be considered sparse, particularly in contrast to the big data generated today. This problem is most acute at the small area level. However, while sample survey data cannot match the sample sizes, collection frequency, and level of detail provided by big data (Whitaker 2014), they are more robust and suitable for use in rigorous statistical analyses (Groves 2015). Furthermore, sample surveys provide rich categorical data and discrete count data that can be used to contextualize public opinion or behavior. Despite this, the development of spatiotemporal models for this non-normally distributed data lags behind that of normally distributed data (Wikle 2015).

Hierarchical modeling

The hierarchical modeling framework has been identified as the ideal vehicle for transferring spatiotemporal statistical methodologies into the realm of social science survey analysis (Wikle, Holan, and Cressie 2013). A Bayesian hierarchical approach is particularly suited to this study as it allows information to be shared across areal units, compensating for unstable/missing data (Gelman and Price 1999). When modeling the risk of occurrence of a phenomenon, prior to data analysis modelers may have probabilities of occurrence based on past occurrences or expert knowledge. These probabilities can be described by a *prior distribution*. Observation of the current data will lead to *likelihoods* of occurrence. This *prior distribution* and *likelihood function* can be combined into the *posterior distribution*, proportional to the product of the *prior distribution* and the *likelihood function*. In this way, the Bayesian approach to inference is characterized by conditioning upon what is known to make probabilistic statements on the unknown interest and is underpinned by two principles: explicit formulation and relevant conditioning (Geweke and Whiteman 2006). The data measurement scale precludes the use of certain likelihood functions. For instance, discrete count data is best described using a Poisson likelihood function. Occam's razor also applies whereby the simplest possible likelihood function should be chosen. Through the use of a prior distribution, the Bayesian inference paradigm provides a formal framework to enable information from various sources to be coherently combined and can be described as formalized subjective judgement (Pole, West, and Harrison 1994). In the absence of prior information on the matter under study, vague priors (characterized by a large variability) are typically used, which result in the posterior distribution being heavily driven by the data.

A generalized linear modeling approach will be adopted to allow the modeling of non-normally distributed data (counts). Ghosh et al. (1998) provide a comprehensive discussion of the use of generalized linear modeling in Bayesian hierarchical small area estimation. Notable examples of a space-time Bayesian hierarchical approach for modeling count data include modeling disease risk (Knorr-Held 2000), election polls (Shor et al. 2007), burglary (Li et al. 2014), cycle accidents (DiMaggio 2015), and road traffic accidents (Boulieri et al. 2016).

Space-time hierarchical modeling

For the modeling of phenomena which vary in space-time, methods are required which can account for the underlying spatiotemporal autocorrelation in the data. Nonparametric, additive models of spatial, temporal, and spatiotemporal random errors are the most appropriate for this. Bernardinelli et al. (1995), Waller et al.(1997), Besag, York, and Mollié (1991), Knorr-Held and Besag (1998), and Knorr-Held (2000) have utilized this approach with progressively sophisticated ways of representing space and time. Bernardinelli et al. (1995) presented a parametric space-time model with random effects used to model an area-specific intercept and temporal trend. Waller et al. (1997) extended an existing hierarchical spatial model to allow time varying trends and spatiotemporal interactions. Besag, York, and Mollié's (1991) nonparametric, additive model of structured and unstructured spatial random effects was extended to include a random temporal effect (Knorr-Held and Besag 1998). This was further developed by the inclusion of a spatiotemporal interaction term with a typology of four spatiotemporal interaction structures by Knorr-Held (2000). Boulieri et al. (2016) incorporated these ideas into a multivariate analytical framework. A recent approach by Bauer et al. (2015) replaces spatial, temporal, and spatiotemporal random errors with penalized spline functions.

Forecasting with Bayesian models

Bayesian forecasting is an intuitive extension of the Bayesian approach to inference. A subset of the unknown values to be estimated are taken to be future values of the quantity of interest (Geweke and Whiteman 2006) and obtained from the predictive distribution (de Alba and Mendoza 2007). The latter found a Bayesian approach well suited to overcoming difficulties forecasting short, seasonal time series, that is, 1–2 years of monthly or quarterly data. Using a partial accumulation approach, whereby past proportions of events are used for forecasting, they found a Bayesian approach superior to the standard ARIMA model for forecasting short, seasonal time series, but the ARIMA was found to be superior for forecasting longer seasonal time series (de Alba and Mendoza 2007). Linzer (2013) created a dynamic Bayesian/frequentist forecasting approach to forecasting election results. Baio and Cerina (2015) improved on this approach with a fully Bayesian approach to forecasting election results. Zaman, Fox, and Bradlow (2014) used a Bayesian approach to predict the evolution of retweets of a tweet on the micro-blogging website Twitter. The most frequently used Bayesian time-series forecasting model is the Bayesian vector autoregressive (BVAR) model. The BVAR has been used to forecast city arrivals from Google analytics data (Gunter and Önder 2016). A recent innovation saw the BVAR combined with a principal components analysis-based approach to resulting in the Bayesian factor-augmented vector autoregression (BFAVAR) (Gunter and Önder 2016).

Based on the literature, the spatiotemporal Bayesian hierarchical approach was taken. This approach provides a flexible framework which allows the spatiotemporal variation of the patterns to be fully explored, as well as estimates and forecasts to be produced at the neighborhood level despite the small number problem.

Data and study region

This study considers a real-world application, forecasting quarterly rates of public confidence in the police in London, United Kingdom. London is the capital city of the United Kingdom, located to the south-east of England. The MPS are responsible for policing approximately 8.9 million inhabitants of diverse heritage and cultures (Greater London Authority 2017). For administration, the city is divided into 32 boroughs and 628 census area wards. MOPAC and the MPS have combined the wards into 107 larger units of operational significance called borough neighborhoods that consist of two or three wards (Mayor's Office for Policing and Crime 2016). For simplicity, the term neighborhood will refer to this level of aggregation throughout this article. With respect to temporal aggregation, MOPAC ties the analysis of public confidence in the police to the financial year, with performance figures collated four times a year. The study uses data collected over 36 financial quarters, spanning nine years between April 2006 (Q5) and March 2015 (Q40).

MOPAC and the MPS have used surveys for public consultation since the 1980s (Harrison, Dawson, and Walker 2009). The PAS is a large scale, rolling, population representative survey which collects opinions on policing, has been its current form since 2002 (Harrison, Dawson, and Walker 2009; Mayor's Office for Policing and Crime 2017). It is conducted face-to-face and is designed to be representative of the residents of London annually at the borough level (BMG Research 2012). MOPAC and the MPS use the question "Taking everything into account, how good a job do you think police IN THIS AREA are doing?" as a public confidence indicator. A respondent is confident if they state that the police did an "excellent" or "good" job. Fig. 1 is a plot of the temporal trend of public confidence over the survey period at the London-wide level. While public confidence is on the decline from historically high percentages of 70% in March 2014, it has been greatly improved from April 2006 to present. A spatiotemporal semivariogram (see Appendix) was used to examine the underlying spatiotemporal dependence structure. A Moran's I test further confirmed the presence of spatial autocorrelation with an average value of 0.3 (*P*-value 0.05) over the study period.

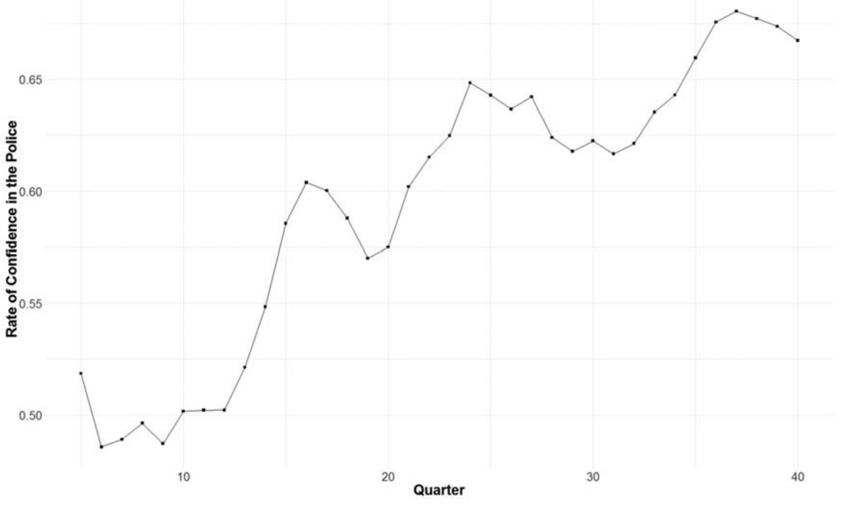

Figure 1. Time-series plot of London-wide rates of public confidence in the police April 2006 to March 2015.

Methodology

To model the PAS data, an extension of the spatial approach for mapping disease risk proposed by Besag, York, and Mollié (1991) is used. In the first level of the model, the binomial likelihood of the data is used to describe the within area variability of the counts conditional on the unknown risk parameters. The second level of the model specifies the space-time structure and parameterizes the unknown risk parameters with a prior distribution. Explanatory variables (covariates) may be added to the model as fixed effects. As the purpose of this study is to demonstrate a novel approach to modeling public confidence in the police rather than to investigate associations with other factors, a purely autoregressive approach was taken. In this case, we specify seven structures with increasingly complex representations of space and time. The third level of the model specifies prior distributions on the hyperparameters, which are the unknown parameters from level 2.

For each neighborhood, the number of confident persons can be modeled using a Binomial distribution as follows

$$Y_{it} \sim \text{Binomial}(n_{it}, \pi_{it}) \quad (1)$$

where i is the neighborhood, t represents the period, n, is the population of interest (population at risk), and π_{it} is the probability that the neighborhood i is considered confident at time t. The binomial model is particularly suited to modeling short-time series such as this one (de Alba and Mendoza 2007). On the second level of the model, the logit transformation π_{it} is taken and expressed it as an additive combination of the overall probability α, overall spatial random effects $(\mu_i + \lambda_i)$, overall temporal random effects, $(\gamma_t + \xi_t)$, and space-time interactions δ_{it} as follows:

$$\text{Model 1 (purely spatial)} \quad \text{logit}(\pi_{it}) = \alpha + (\mu_i + \lambda_i) \quad (2)$$

$$\text{Model 2 (purely temporal)} \quad \text{logit}(\pi_{it}) = \alpha + (\gamma_t + \xi_t) \quad (3)$$

$$\text{Model 3 (spatiotemporal)} \quad \text{logit}(\pi_{it}) = \alpha + (\mu_i + \lambda_i) + (\gamma_t + \xi_t) \quad (4)$$

$$\text{Model 4–7 (spatiotemporal with interactions)} \quad \text{logit}(\pi_{it}) = \alpha + (\mu_i + \lambda_i) + (\gamma_t + \xi_t) + \delta_{it} \quad (5)$$

where λ_i represents the unstructured spatial random effects, μ_i represents the structured spatial random effects in an area i, γ_t represents the structured temporal random effects, and ξ_t represents the unstructured temporal random effects at period t. In the Bayesian context, prior distributions are needed on each parameter. Following the Besag, York, Mollie specification (Besag, York, and Mollié 1991) an Intrinsic Conditional Autoregressive model (Besag and Kooperberg 1995) was chosen for μ_i, so that

$$\mu | \mu_{-i} \sim \text{Normal}\left(\frac{1}{N_i}\sum_{j \in \delta} \mu_j, \frac{\sigma_\mu^2}{N_i}\right) \quad (6)$$

where N_i is the number of neighboring areas of i (i.e., sharing boundaries). In practice, the parameter μ_i is normally distributed with a mean equal to the mean of the μ parameters for the set of areas sharing boundaries with i (identified by δ in equation 5). For the area i the variance of μ_i is a global variance σ_μ^2 divided by the number of neighbors, to follow the assumption that an area with many neighbors will have a more precise estimate of μ_i than an area rather isolated. On λ_i and ξ_t, a normal distribution is specified centered on 0 with variance σ_λ^2 and σ_γ^2, respectively; on γ_t a random walk (RW) is used to represent temporal dependence (Blangiardo and Cameletti 2015) which assumes that the parameter for each time point depends on the previous one as follows:

$$\gamma_t | \gamma_{t-1} \sim \text{Normal}(\gamma_{t-1}, \sigma_\gamma^2) \tag{7}$$

and is characterized by a variance (σ_γ^2). The form of the spatiotemporal interaction term varies with differing assumptions per the Knorr-Held framework to give models 4–7 outlined in the Table 1 below.

To complete the model specification, distributions need to be specified on the intercept α and on all the variances. An improper uniform prior on the whole real line was used for the intercept(α). A minimally informative Gamma(a, b) distribution was used as the hyperprior distribution for$\frac{1}{\sigma_\lambda^2}$, $\frac{1}{\sigma_\mu^2}$, $\frac{1}{\sigma_\gamma^2}$, $\frac{1}{\sigma_\xi^2}$, and $\frac{1}{\sigma_\delta^2}$ where a and b equal 1.

Random draws from the product of the likelihood ratio and the prior distribution are summarized to give a posterior distribution (Gelman and Hill 2006). Most commonly, this is done using a Markov Chain Monte Carlo approach using Gibbs sampling (Hahn, 2014). A recent alternative approach called integrated nested Laplace approximation (INLA) was used instead. In this deterministic approach, the posterior distribution is analytically approximated using the Laplace method. An implementation of INLA developed by Rue, Martino, and Chopin (2009) for the *R* statistical programing environment was used.

Effect of the spatiotemporal representation structures

Fig. 2 presents time-series plots of π_{it}, the probability that a neighborhood, i, is confident at time t. The empirical values and estimates produced by the seven models are presented together for 12 neighborhoods across London. Neighborhoods within a Borough are symbolized in shades of the same color. This allows the spatial variation in confidence to be seen, with neighborhoods in different Boroughs having markedly different temporal profiles.

The assumed spatiotemporal representation structures result in varying amounts of smoothing in the estimates. For the spatial benchmark, model 1, a single value is estimated for each neighborhood for the entire test period. For the temporal benchmark, model 2, estimated values are allowed to vary over the test period but not across neighborhoods, resulting in a single profile for all the neighborhoods. Model 3, an additive combination of models 1 and 2, allows each neighborhood to have a different value while retaining a single temporal profile. These representation structures are quite restrictive resulting in over smoothed estimates which do not represent the data well. The inclusion of the spatiotemporal interaction term in models 4–7 allows more flexibility as the temporal trend can deviate across the different neighborhoods. This allows the temporal trend of the empirical data to be more accurately represented, at the same time providing some smoothing in areas of extremely high or low confidence. Models 4 and 6 appear most similar to the empirical data with models 5 and 7 appearing slightly more smoothed.

Model selection and forecast accuracy evaluation

Model fit was investigated using the posterior predictive check proposed by Gelman, Meng, and Stern (1996). This self-consistency check evaluates whether the observed data can reasonably be expected from the posterior predictive distribution. Replicated data simulated from the

Table 1. Knorr-Held's Four Inseparable Spatiotemporal Interaction Structures (Knorr-Held 2000)

Model	Type/Parameters interacting	Interaction term	Description	Prior distribution used
Model 4	Type I λ_i and ξ_t	$\delta_{it} \sim N\left(0, \sigma_{it}^2\right)$	Interactions have no spatial or temporal pattern	Normally distributed
Model 5	Type II λ_i and γ_t	$\delta_{i,1:t} \sim \mathrm{RW}\left(W_t, \sigma_t^2\right)$	Interactions vary in time with trends differing by neighborhood	Random walk
Model 6	Type III ξ_t and μ_i	$\delta_{1:i,t} \sim \mathrm{ICAR}\left(W_i, \sigma_\delta^2\right)$	Interactions have a trend in space. Nearby areas have similar differences from the overall trend at time *t*.	Prior smoothly varying in space (ICAR)
Model 7	Type IV μ_i and γ_t	$\delta_{1:\mu,t} \sim \mathrm{ICAR}\left(W_\mu, \sigma_\delta^2\right) \otimes \mathrm{RW}\left(W_t, \sigma_t^2\right)$	Full spatiotemporal interactions. Nearby areas have similar but different trends.	The Kronecker product of smoothing varying in space (ICAR) and a random walk in time.

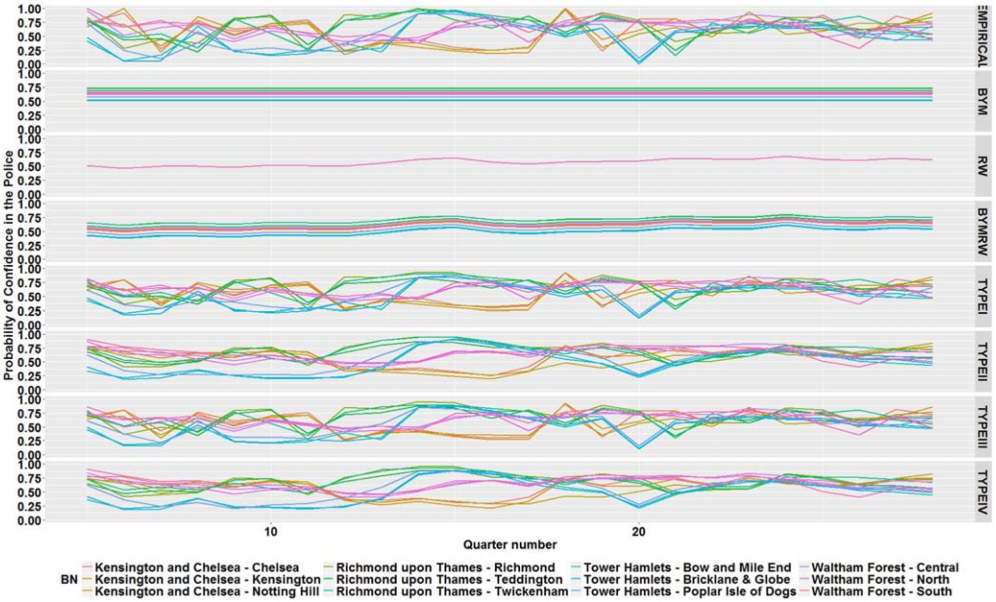

Figure 2. Time series plots displaying the values of π_{it} estimated by the seven models for 12 neighborhoods across London.

joint posterior predictive distribution is compared with the observed data using the posterior predictive P value defined as follows:

$$P_{B_i} = P\left(y_i^{\text{rep}} \leq y_i | y\right)$$

where y is the entire data set, y^{rep} is the replicated value and y_i is the empirical value. Posterior predictive P values near to 0 or 1 indicate an ill-fitting model. The deviance information criterion (DIC), a likelihood based measure of model complexity and fit (Spiegelhalter et al. 2014), was also used to evaluate the parsimony of the model.

The best fitting model was then used to produce one-step ahead, quarterly forecasts of public confidence. We have chosen to evaluate the model forecast for one-step ahead as one financial quarter is a meaningful forecast unit for the end user with public confidence planning and evaluation meetings taking place quarterly. The performance of the ST-BHM approach to forecasting was assessed by comparing its accuracy to eight more commonly used time-series forecasting methods: historical mean, naïve method, seasonal naïve method, naïve model with drift, simple exponential smoothing (Gardner 1985), Holt-Winters seasonal exponential smoothing (Holt 2004), ARIMA (Box and Jenkins 1976), STARIMA (Pfeifer and Deutsch 1980; Pfeifer and Deutsch 1981), and the hierarchical time-series approach with optimal combination. These are the most widely applied time-series forecasting methods and are useful from a benchmarking perspective. The performance rankings of forecast models used can vary with the evaluation statistics used (Makridakis and Hibon 2000). For this reason, a variety of statistical metrics were employed. Three frequently used forecast accuracy evaluation metrics were used: the mean absolute error

(MAE), the relative mean absolute error (RelMAE), and the mean absolute scaled error (MASE). The MAE is

$$\mathrm{MAE}=1/n\sum_{t=1}^{n}e_t$$

where the forecast error at time t, $e_t=y_i-y_{i+h}$ and y_{i+h} is a forecast of y_i, h steps ahead. The MAE, is given in measurement units making it scale dependent (Hyndman and Koehler 2006). Scale dependent measures are suitable for comparing performances of forecasts with identical measurement units made at the same scale. The use of a relative error measure is particularly appropriate for evaluating forecast performance on the same series where scale issues are not a factor. The RelMAE is given by

$$\mathrm{RelMAE}=\mathrm{MAE}/\mathrm{MAE}_{\mathrm{benchmark}}$$

As is customary, we have chosen the naïve method, that is, the random walk without drift as our benchmark model. As such, the RelMAE is equivalent to Thiel's coefficient of inequality (Theil 1966; Bliemel 1973) as our forecasts are one step ahead. Finally, the MASE was applied as it has been described as the "best available measure of forecast accuracy" provides better, more robust accuracy estimates less sensitive to scale errors (Hyndman and Koehler 2006). The MASE is the best metric to compare forecast performances across various levels in this case, the neighborhood level, the Borough level, and the London wide level.

$$\mathrm{MASE}=1/n\sum_{t=1}^{n}\left(\frac{e_t}{\frac{1}{n-1}\sum_{i=2}^{n}|Y_i-Y_{i-1}|}\right)$$

which is the mean of the forecast error, e_t, divided by the average forecast error of the naïve method. A value less than 1 signifies a better performance than the naïve model and a value greater than 1 signifies the opposite. As originally proposed by Hyndman and Koehler (2006), the MASE is an internal measure as the comparison is made with performance of the naïve benchmark during the fit period (Clements and Hendry 2005). However, some authors, Huddleston, Porter, and Brown (2015) for example, have adapted it to an external measure by choosing to scale by the performance of the naïve benchmark during the forecast period instead of during the training period.

Results and discussion

The data was separated by a 2:1 ratio into two subsets: the training set (April 2006 to March 2012; 24 quarters) for model calibration and the testing set (April 2012 to March 2015; 12 quarters) for out of sample forecast evaluation.

Model selection

Analyzing histograms of the posterior predictive P values in Fig. 3 allows us to assess the suitability of the models given the data. A bell-shaped histogram indicates superior model fit. Models 1–3 were poorly fitted with pronounced u-shaped histograms. Comparing models 1–3 with models 4–7, the fit was greatly improved with the inclusion of the spacetime interaction term.

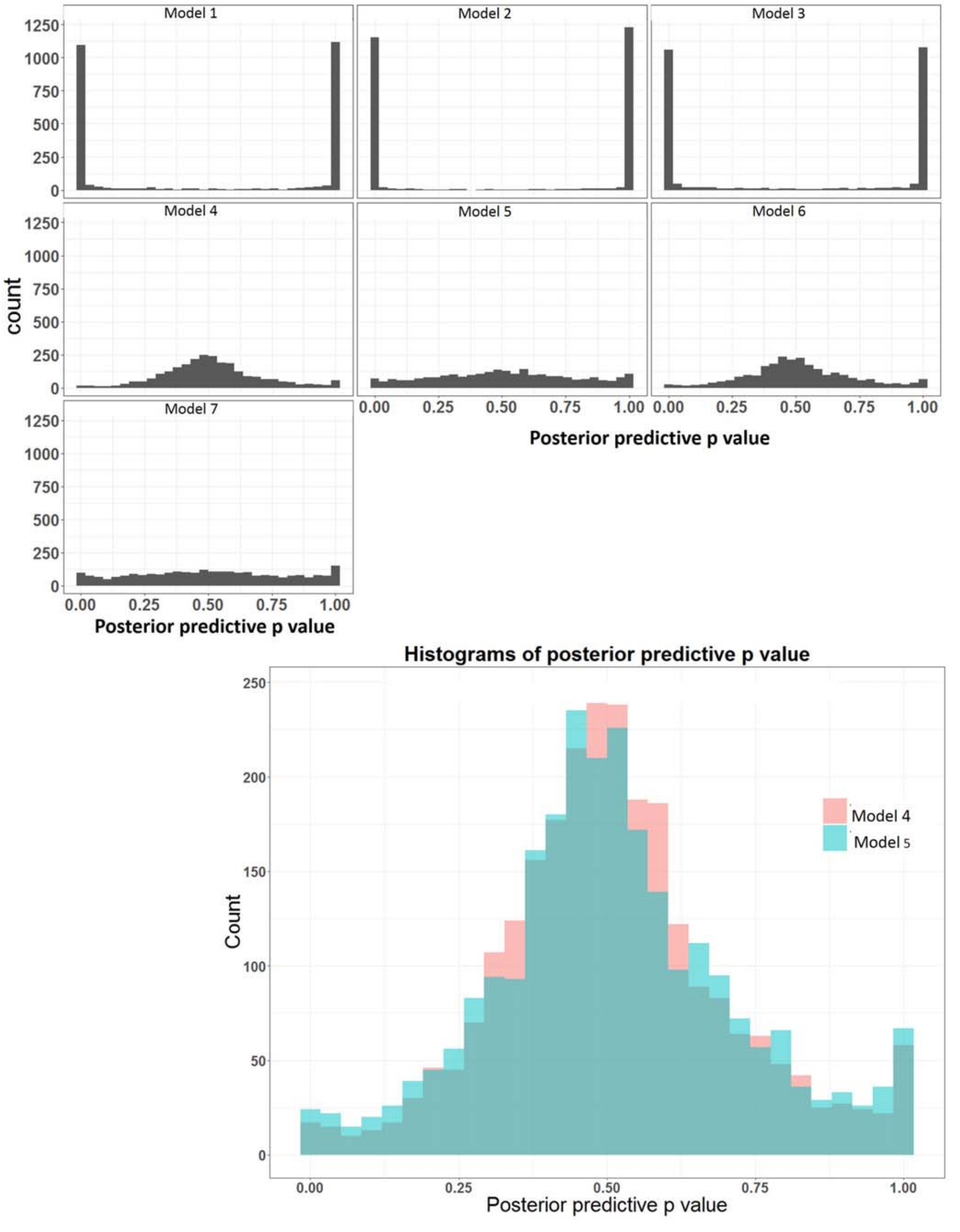

Figure 3. Histograms of the posterior predictive *P* values for all models (above) and models 4 and 6 (below).

This result means that the inclusion of the normally distributed spatiotemporal interaction term is very useful to the model, describing some of the variability of public confidence in the police in London. Model 4, with a normally distributed spatiotemporal interaction structure, and model 6, with a spatially structured interaction term, had the best fit overall. Upon closer inspection of the histogram (Fig. 3, below), model 4 (in pink) had a slightly better fit with lower

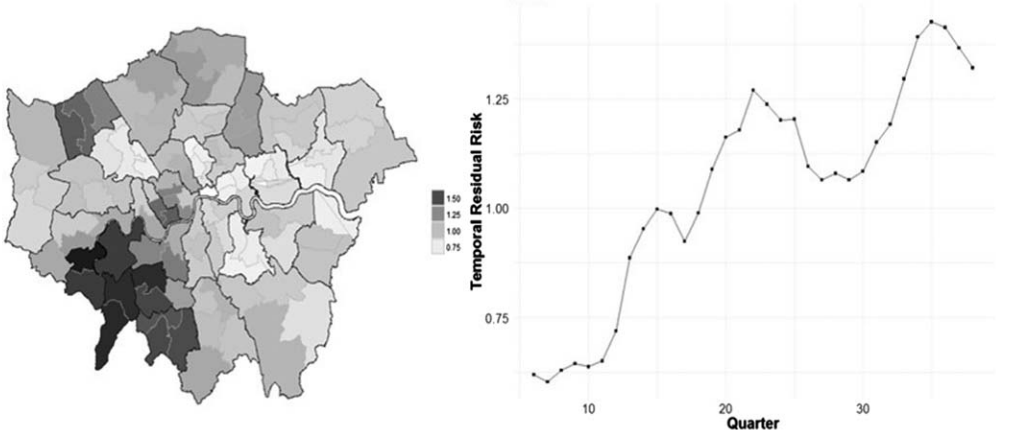

Figure 4. Estimated structured spatial (left) and temporal (right) random effects.

values at the tails of the distribution and higher values at the center. Furthermore, model 4 had the lowest DIC value, indicating the best fit and parsimony. Thus model 4 was chosen for forecasting.

Spatiotemporal trends in public confidence

Examination of the estimates of spatial and temporal random effects provides useful information for practitioners (see Fig. 4). On the left, the map of the spatial random effects captures the persistent spatial trend in public confidence, revealing the areas that are most likely to be confident in the police over the entire test period. Historically, there has been a gap in prospects between east and west Londoners, with the west generally being more prosperous and the east exhibiting more urban deprivation (MacRury 2016). In particular, neighborhoods within the south-west are the least deprived. Previous research has found that persons experiencing higher levels of deprivation are less likely to be confident in the police (Reisig 2007). Policy makers have explored the effect of this "east/west" deprivation divide on public confidence in the police in London confirming that this relationship generally holds true (Mayor's Office for Policing and Crime 2014). Our findings support this, with neighborhoods in the south-west being the most likely to be confident in the police.

On the right, the empirical temporal trend (see Fig. 1) is well captured by the temporal random effects. Examining the structured temporal component at the neighborhood level also provides insights. Stanko and Bradford (2009) used confirmatory factor analysis to establish a link between effective policing of major effects in London and confidence in the policing in local areas. The London 2012 Summer Olympics games, held in the third quarter of 2012 (Q30), are the biggest major event in recent history, with an estimated 90% of the British population tuning into the coverage on the BBC (Plunkett, 2012). Our examination of the empirical temporal trend suggests a lagged increased in confidence after the successful policing of the 2012 London Olympics. Fig. 5 shows the empirical values and the estimates of spatial, structured temporal and spatiotemporal interaction for Waltham Forest—South, one of the host neighborhoods of the Olympics. The increase over the fourth quarter of 2012 (Q31) and the first quarter of 2013 (Q32) can be clearly seen. Examining the temporal trend we may also extend the

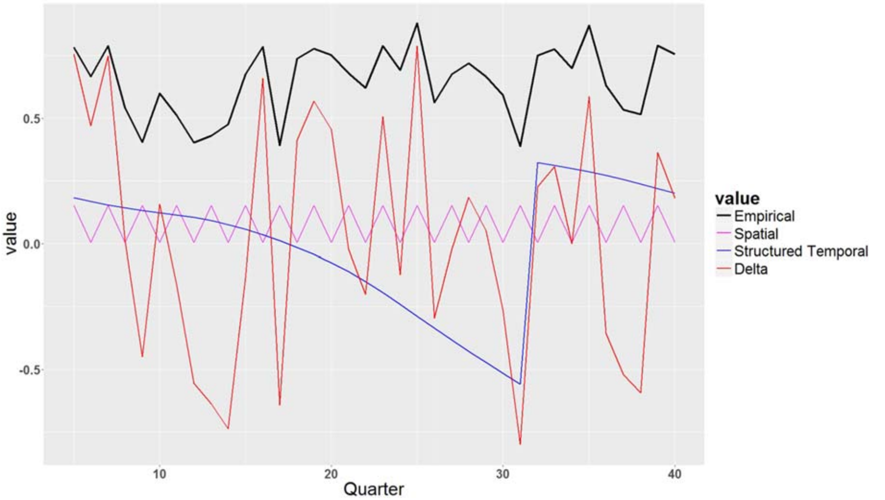

Figure 5. Time-series plot of selected model components for the Waltham Forest South neighborhood.

findings of a net-neutral effect of the 2011 London riots (Q26) on policing, due to the split in public opinion, (Hohl, Stanko, and Newburn 2012) to the neighborhood level.

Examining the spatiotemporal interaction term can provide further insight into the dynamics of confidence at the neighborhood level. This is particularly relevant in the case of public confidence modeling as patterns are very likely to be impacted by factors beyond the control of the police such as social cohesion, and the physical condition of the neighborhood (Mayor's Office for Policing and Crime 2014). The flexibility of the exchangeable prior used allows the underlying spatiotemporal dynamics of these background features which cannot be explicitly modeled to be incorporated in a way that is useful for analysis, rather than being included in a noise term. This approach allows us to look under the hood of the neighborhood variability, particularly when explanatory covariates are included, to examine whether the disparities in confidence are driven in large part by factors under the control of the police or other ancillary considerations. A further exploration of these spatiotemporal patterns including explanatory covariates will be the subject of future work.

Forecasting public confidence

Out of sample forecast, accuracy was measured on a rolling quarterly basis, with all previously observed data used to forecast one quarter into the future for the period April 2012 to March 2015. This allowed methods requiring longer initialization periods, such as ARIMA and STARIMA to be evaluated equally. Previous study was conducted by Williams, Haworth, and Cheng (2015) into the spatiotemporal structure of public confidence in the police, confirming a heterogeneous and nonstationary nature. This violates the assumptions of spatial homogeneity and stationarity required by traditional time-series forecasting methods (ARIMA, STARIMA, etc.). In this case, the data was transformed by differencing prior to modeling.

Visual comparison of the observed and forecast series is an intuitive first step in forecast performance evaluation. Fig. 6 shows the forecasted rates of public confidence on a rolling quarterly basis for the test period at the London wide level (1) and the local level (2). At first

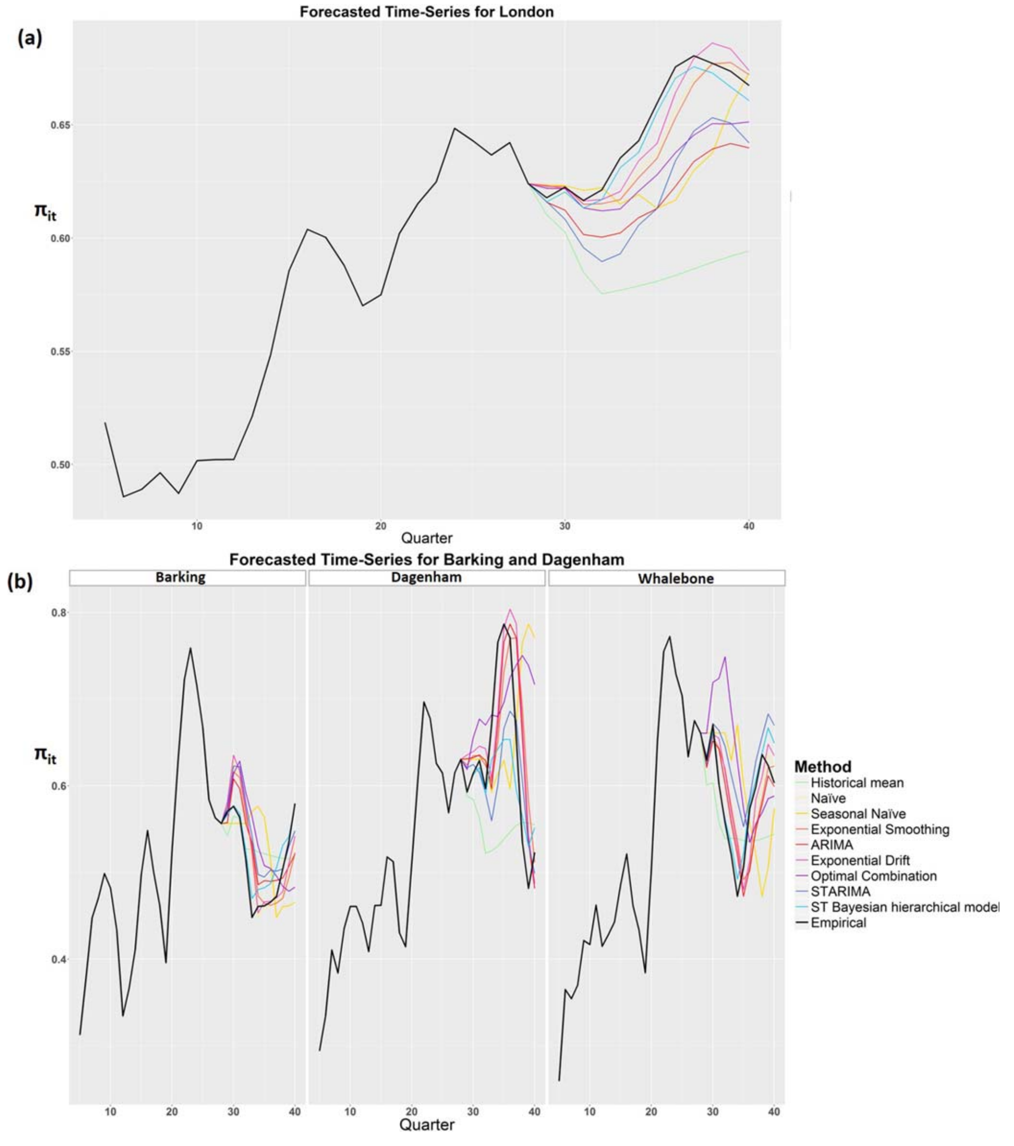

Figure 6. Trend line of the rate of confidence values per rolling quarter for the neighborhoods in the Borough of Barking and Dagenham from October 2006 to March 2015 with forecasted values from April 2012.

glance, it is apparent that the performance of the models varies considerably at the London-wide level, with the average/historical mean and seasonal naïve methods performing the worst overall. The forecasts for the naïve, simple exponential smoothing, ARIMA, and ST-BHM appear to follow the trajectory of the empirical data reasonably well. However, the ST-BHM appears the most responsive and best at capturing both the trajectory and overall level of the rates.

Table 2 presents the average errors for the forecast models at the London-wide level for the entire forecast window as defined previously. Per the most robust statistic, the MASE, all

Table 2. Forecast Accuracy Statistics

Method	MAE	RelMAE	MASE
Average/historical mean	0.08	2	0.45
Naïve (Random walk without drift)	0.04	1	0.2
Seasonal naïve	0.08	2	0.4
Simple exponential smoothing	0.04	1	0.21
Holt-Winters seasonal exponential smoothing	0.04	1	0.2
ARIMA	0.06	1.5	0.32
STARIMA (1,0,0)	0.06	1.5	0.33
Optimal combination	0.06	1.5	0.29
ST Bayesian hierarchical modeling	**0.02**	**0.5**	**0.13**

the forecast models performed better than the naïve method during the training window with MASE values less than 1. All three statistics used confirmed that the ST-BHM was the best performing model. Using the external measures (MAE, RelMAE) which compared the performance of the naïve benchmark and the ST-BHM within the forecast window, the ST-BHM approach was twice as accurate. Using an internal measure of forecast accuracy (MASE) which compared the performance of the naïve benchmark and the ST-BHM to the performance of the naïve benchmark within the training window, the ST-BHM approach was 1.5 times as accurate the benchmark. It confirms that the average/historical mean and the seasonal naïve models were the worst performing, being at least twice as bad as the benchmark per the internal and external metrics. A spatiotemporal variogram of the forecast errors, e_t, (see Appendix) further confirmed the performance of the ST-BHM model.

Maps of the forecast errors, e_t, of model 4, the spatiotemporal Bayesian hierarchical model with type I interactions, were produced for each quarter in the forecast period, April 2012 (Q29) to March 2015 (Q40), Fig. 7. Areas that were over forecast are symbolized in blue, while under forecasted areas are symbolized in red. While some clustering was found, particularly in Q33, generally the residuals had no obvious spatial patterns or strong forecast bias with an average Moran's Index score of 0.12 (*P*-value 0.04).

Conclusions and further work

Public confidence in the police is a key component of effective policing that varies across geographical space and over time. This study illustrates a Bayesian spatiotemporal approach to investigating and predicting trends in public confidence in the police. The effect of spatiotemporal representation structures on the estimates produced was investigated. A model with spatial and temporal random effects as well as a normally distributed spatiotemporal interaction term was the best fitting and produced the most realistic estimates and forecasts.

Stable spatial and temporal trends in public confidence in the police were uncovered at the city-wide and neighborhood levels. Our examination of the spatial trends further supports the finding that confidence in the local police is related to neighborhood deprivation levels. Our examination of temporal trends suggests a possible lagged "Olympic effect" to be seen, whereby Londoners were more confident in the police after the successful policing of the 2012

Figure 7. Maps of the ST-BHM forecast errors for the test period.

London Olympic games Hohl, Stanko, and Newburn (2012). This aligns with the findings of Stanko and further supports the MET confidence model. Additionally, we have extended the findings of a net neutral effect of the 2011 London riots on confidence in policing, due to the split in public opinion, (Hohl, Stanko, and Newburn 2012) to the neighborhood level. The inclusion of a spatiotemporal interaction term improved the fit of the model and allowed us to further examine the neighborhood variability. The predictive ability of the spatiotemporal Bayesian hierarchical approach was validated by comparison with traditional time-series forecasting methods.

The modeling was motivated by a desire to understand the variations in public confidence in space-time rather than investigate relationships between associated factors. For this reason, a purely autoregressive approach was taken investigating the effect of seven different underlying spatio-temporal dependence structures without the use of explanatory covariates. In future, covariates such as geodemographic classifier, land use, and neighborhood cohesion could be incorporated as fixed effects. This would make the output more useful to the policy maker. Furthermore, a mixture modeling approach could be used to classify the spatiotemporal interactions (Abellan, Richardson, and Best 2008). This will allow the full spatiotemporal profile of public confidence at the neighborhood level to be examined and unusual patterns of public confidence to be better identified. Additionally, research into the impact of high profile policing incidents on public confidence particularly for Black and Minority Ethnic Londoners has also been identified as future research after (Desmond, Papachristos, and Kirk 2016).

It should be noted that forecast performance varies with the forecast horizon (Makridakis and Hibon 2000). We have chosen to evaluate the model forecast for one-step ahead as one financial quarter is a meaningful forecast unit for the end user MOPAC/MET, with Borough level public confidence planning and evaluation meetings taking place quarterly. Future work, will see forecasts evaluated for longer horizons. Additionally, the accuracy of forecasts produced from a combination of models has been shown to be better than the forecasts of the individual models on their own (Makridakis and Hibon 2000). This has also been earmarked for future work.

Acknowledgments

This work is part of the project – Crime, Policing and Citizenship (CPC): Space-Time Interactions of Dynamic Networks (“http://www.ucl.ac.uk/cpc” www.ucl.ac.uk/cpc), supported by the UK Engineering and Physical Sciences Research Council (EP/J004197/1). The authors would like to acknowledge the Metropolitan Police Service (MPS) and Mayor’s Office for Policing and Crime (MOPAC) for provision of the data. They are also grateful to Trevor Adams for many valuable discussions about the manuscript and related work. The results presented and views expressed in this manuscript are the responsibility of the authors alone and do not represent the views of Trevor Adams, the MPS or MOPAC. The authors would like to thank the editor Dr. Rachel Franklin, and three anonymous referees for their valuable comments and suggestions.

Appendix

Table A1. Deviance Information Criterion Statistics for Models

Model	Effective No. of Parameters	Mean Deviance	DIC
Model 1: BYM	102.1834	34,015.67	34,117.85
Model 2: RW	134.9651	31,757.98	31,892.95
Model 3: BYM + RW	134.9651	31,757.98	31,892.95
Model 4: Type I	**2,872.997**	**18,118.93**	**20,991.93**
Model 5: Type II	2,148.502	19,868.31	22,016.82
Model 6: Type III	2,747.565	18,262.22	21,009.79
Model 7: Type IV	1,967.077	20,139.36	22,106.43

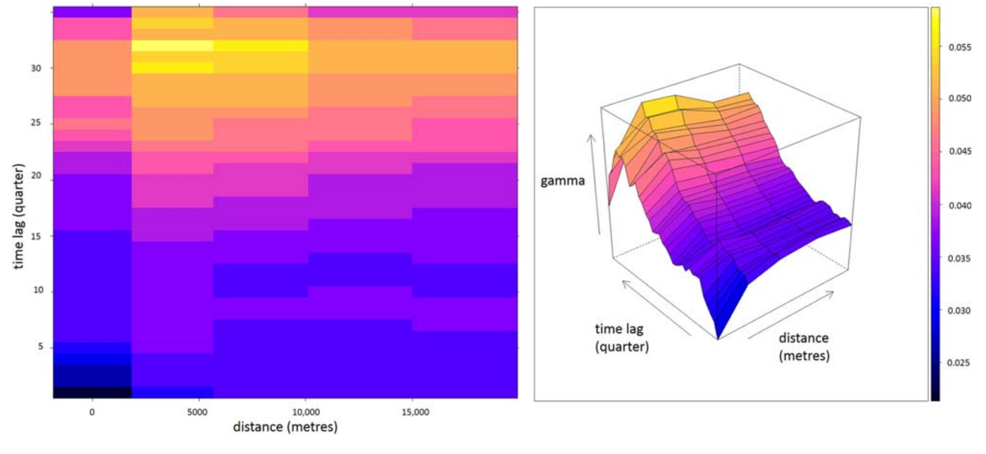

Figure A1. Level plot (left) and 3D (right) representations of the spatiotemporal variogram of public confidence in the police April 2006 to March 2015.

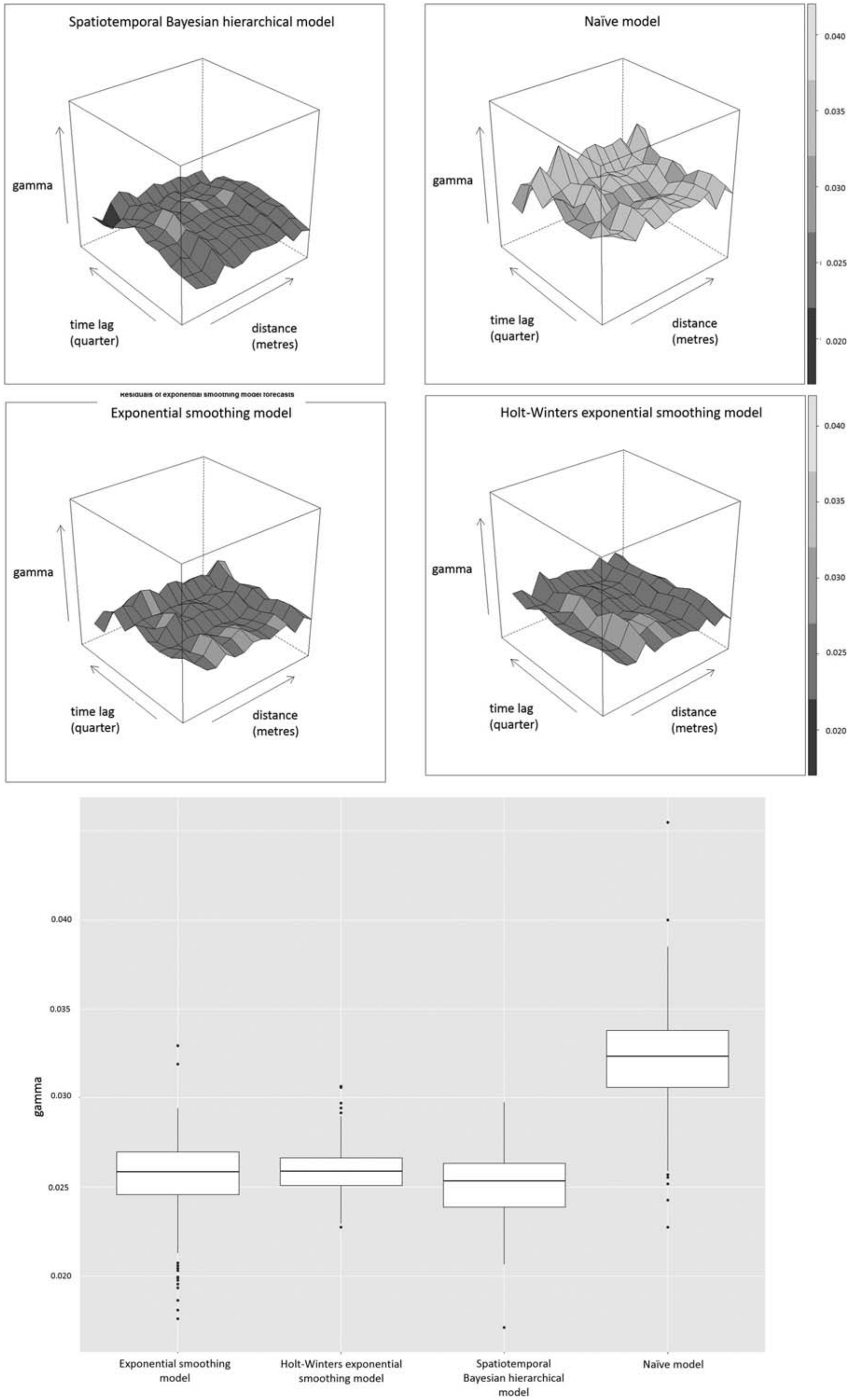

Figure A2. Spatiotemporal variogram of the forecast errors for the four most accurate models (above) and a box and whisker plot of the spatiotemporal variogram values.

References

Abellan, J. J., S. Richardson, and N. Best. (2008). "Use of Space–Time Models to Investigate the Stability of Patterns of Disease." *Environmental Health Perspectives* 116, 1111–9.

Aldor-Noiman, S., L. D. Brown, E. B. Fox, and R. A. Stine. (2016). "Spatio-Temporal Low Count Processes with Application to Violent Crime Events." *Statistica Sinica* 26(4), 1587–10.

Association of Chief Police Officers. (2012). "Policing in the UK: A Brief Guide." Association of Chief Police Officers.

Baio, G., and R. Cerina. (2015). "A Predictable Outcome." *Significance* 12, 11–3.

Bauer, C., J. Wakefield, H. Rue, S. Self, Z. Feng, and Y. Wang. (2015). "Bayesian Penalized Spline Models for the Analysis of Spatio-Temporal Count Data." *Statistics in Medicine* 35(11), 1848–65.

Bernardinelli, L., D. Clayton, C. Pascutto, C. Montomoli, M. Ghislandi, and M. Songini. (1995). "Bayesian Analysis of Space—Time Variation in Disease Risk." *Statistics in Medicine* 14, 2433–43.

Besag, J., and C. Kooperberg. (1995). "On Conditional and Intrinsic Autoregressions." *Biometrika* 82(4), 733–46.

Besag, J., J. York, and A. Mollié. (1991). "Bayesian Image Restoration, with Two Applications in Spatial Statistics." *Annals of the Institute of Statistical Mathematics* 43(1), 1–20.

Blangiardo, M., and M. Cameletti. (2015). *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. Hoboken, NJ: Wiley.

Bliemel, F. (1973). "Theil's Forecast Accuracy Coefficient: A Clarification." *Journal of Marketing Research* 10(4), 444–6.

BMG Research. (2012). "Public Attitude Survey 2011–12." Technical Report Q28, Bostock Marketing Group Ltd.

Boulieri, A., S. Liverani, K. de Hoogh, and M. Blangiardo. (2016). "A Space–Time Multivariate Bayesian Model to Analyse Road Traffic Accidents by Severity." *Journal of the Royal Statistical Society: Series A Statistics in Society* 180(1), 119–39.

Bowers, K. J., S. D. Johnson, and K. Pease. (2004). "Prospective Hot-Spotting: The Future of Crime Mapping?" *The British Journal of Criminology* 44(5), 641–58.

Box, G. E. P., and G. M. Jenkins. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco; London.

Chainey, S., L. Tompson, and S. Uhlig. (2008). "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime." *Security Journal* 21(1–2), 4–28.

Clements, M. P., and D. F. Hendry. (2005). "Evaluating a Model by Forecast Performance*." *Oxford Bulletin of Economics and Statistics* 67, 931–56.

de Alba, E., and M. Mendoza. (2007). "Bayesian Forecasting Methods for Short Time Series." *Foresight. The International Journal of Applied Forecasting* Spring 2007, 41–4.

Desmond, M., A. V. Papachristos, and D. S. Kirk. (2016). "Police Violence and Citizen Crime Reporting in the Black Community." *American Sociological Review* 81(5), 857–76.

DiMaggio, Cs. (2015). "Small-Area Spatiotemporal Analysis of Pedestrian and Bicyclist Injuries in New York City." *Epidemiology* 26(2), 247–54.

Gardner, E. S.(1985). "Exponential Smoothing: The State of the Art." *Journal of Forecasting* 4(1), 1–28.

Gelman, A., and J. Hill. (2006). *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge, UK: Cambridge University Press.

Gelman, A., X.-L. Meng, and H. Stern. (1996). "Posterior Predictive Assessment of Model Fitness via Realized Discrepancies." *Statistica Sinica* 6, 733–60.

Gelman, A., and P. N. Price. (1999). "All Maps of Parameter Estimates are Misleading." *Statistics in Medicine* 18, 3221–34.

Geweke, J., and C. Whiteman. (2006). "Bayesian Forecasting." In *Handbook of Economic Forecasting*, 3–80, edited by G. Elliott, C. W. J. Granger, and A. Timmermann. North Holland: Elsevier.

Ghosh, M., K. Natarajan, T. W. F. Stroud, and B. P. Carlin. (1998). "Generalized Linear Models for Small-Area Estimation." *Journal of the American Statistical Association* 93, 273–82.

Greater London Authority. (2017). "GLA Population and Household Projections." Greater London Authority.

Greenhalgh, S., F. Boardman, J. Glanz, S. O'Connell, K. Prince, L. Duncan, H. Bailey, R. Lawrence, H. King, M. Chishty, and S. Otter. (2015). *MOPAC Challenge: Performance and Confidence*. City Hall, London: Mayor's Office for Policing and Crime.

Groves, R. R. (2015). "Moving from the Sample Survey Paradigm to a Blended World with High-Dimensional Data." Workshop on Big Data for Social Policy, The Fields Institute for Research in Mathematical Sciences, Toronto, Canada.

Gunter, U., and I. Önder. (2016). "Forecasting City Arrivals with Google Analytics." *Annals of Tourism Research* 61, 199–12.

Hahn, E. D. (2014). *Bayesian Methods for Management and Business: pragmatic Solutions for Real Problems*. Hoboken, NJ: Wiley.

Harrison, A., P. Dawson, and A. Walker. (2009). "Surveys in the MPS: A Journey of Change in Understanding and Practice." *Policing* 3(4), 340–46.

Hohl, K., B. Stanko, and T. Newburn. (2012). "The Effect of the 2011 London Disorder on Public Opinion of Police and Attitudes Towards Crime, Disorder, and Sentencing." *Policing* 7(1), 12–20.

Holt, C. C. (2004). "Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages." *International Journal of Forecasting* 20, 5–10.

Home Office. (2012). "Definition of Policing by Consent." Freedom of Information Release No: 25060, *Home Office*, London, UK.

Huddleston, S. H., J. H. Porter, and D. E. Brown. (2015). "Improving Forecasts for Noisy Geographic Time Series." *Journal of Business Research, Special Issue on Simple versus Complex Forecasting* 68, 1810–8.

Hyndman, R. J., and A. B. Koehler. (2006). "Another Look at Measures of Forecast Accuracy." *International Journal of Forecasting* 22, 679–88.

Iversen, G. (1984). *Bayesian Statistical Inference*. Thousand Oaks, CA: SAGE Publications.

Jackson, J., B. Bradford, E. A. Stanko, and K. Hohl. (2013). *Just Authority: Trust in the Police in England and Wales*, 1st ed. New York, NY: Routledge.

Knorr-Held, L. (2000). "Bayesian Modelling of Inseparable Space-Time Variation in Disease Risk." *Statistics in Medicine* 19, 2555–67.

Knorr-Held, L., and J. Besag. (1998). "Modelling Risk from a Disease in Time and Space." *Statistics in Medicine* 17, 2045–60.

Kulldorff, M., R. Heffernan, J. Hartman, R. Assunção, and F. Mostashari. (2005). "A Space–Time Permutation Scan Statistic for Disease Outbreak Detection." *PLOS Medicine* 2(3), e59.

Law, J., M. Quick, and P. Chan. (2013). "Bayesian Spatio-Temporal Modeling for Analysing Local Patterns of Crime Over Time at the Small-Area Level." *Journal of Quantitative Criminology* 30, 57–78.

Law, J., M. Quick, and P. W. Chan. (2015). "Analyzing Hotspots of Crime Using a Bayesian Spatiotemporal Modeling Approach: A Case Study of Violent Crime in the Greater Toronto Area." *Geographical Analysis* 47, 1–19.

Li, G., R. Haining, S. Richardson, and N. Best. (2014). "Space–Time Variability in Burglary Risk: A Bayesian Spatio-Temporal Modelling Approach." *Spatial Statistics* 9, 180–91.

Linzer, D. A. (2013). "Dynamic Bayesian Forecasting of Presidential Elections in the States." *Journal of the American Statistical Association* 108, 124–34.

MacRury, I. (2016). *London After Recession: A Fictitious Capital*? Farnham: Ashgate.

Makridakis, S., and M. Hibon. (2000). "The M3-Competition: results, conclusions and Implications." *International Journal of Forecasting* 16 (4), 451–76.

Mayor's Office for Policing and Crime. (2014). "MOPAC Challenge Board: Confidence Transcript." Transcript. Mayor's Office for Policing and Crime, London, UK.

Mayor's Office for Policing and Crime. (2017). *Metropolitan Police Public Attitudes Surveys, 2000–2017*. [Data Collection]. 6th ed. UK Data Service.

Mohler, G. O., M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita. (2011). "Self-Exciting Point Process Modeling of Crime." *Journal of the American Statistical Association* 106(493), 100–8.

Pfeifer, P. E., and S. J. Deutsch. (1980). "A Three-Stage Iterative Procedure for Space-Time Modeling." *Technometrics* 22, 35–47.

Pfeifer, P. E., and S. J. Deutsch. (1981). "Seasonal Space-Time ARIMA Modeling." *Geographical Analysis* 13, 117–33.

Plunkett, J. (2012). *BBC Olympics Coverage Watched by 90% of UK Population*. London, UK: The Guardian.

Pole, A., M. West, and J. Harrison. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. Boca Raton, FL: CRC Press.

Reisig, M. D. (2007). "Continuing the Discussion on Community Policing, Issue 2 Procedural Justice and Community Policing—What Shapes Residents' Willingness to Participate in Crime Prevention Programs?" *Policing: A Journal of Policy and Practice* 1(3), 356–69.

Rue, H., S. Martino, and N. Chopin. (2009). "Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations." *Journal of the Royal Statistical Society: Series B Statistical Methodology* 71, 319–92.

Shor, B., J. Bafumi, L. Keele, and David Park. (2007). "A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data." *Political Analysis* 15(2), 165–81.

Sindall, K., P. Sturgis, and W. Jennings. (2012). "Public Confidence in the Police a Time-Series Analysis." *The British Journal of Criminology* 52, 744–64.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. (2014). "The Deviance Information Criterion: 12 Years On." *Journal of the Royal Statistical Society: Series B Statistical Methodology* 76, 485–93.

Stanko, B., J. Jackson, B. Bradford, and K. Hohl. (2012). "A Golden Thread, a Presence Amongst Uniforms, and a Good Deal of Data: Studying Public Confidence in the London Metropolitan Police." *Policing and Society* 22, 317–31.

Stanko, E. A., and B. Bradford. (2009). "Beyond Measuring 'How Good a Job' Police are Doing: The MPS Model of Confidence in Policing." *Policing* 3(4), 322–30.

Theil, H. (1966). *Applied Economic Forecasting*. Amsterdam: North-Holland Publishing Company.

Tyler, T. R., and J. Fagan. (2008). "Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities." *Ohio State Journal of Criminal Law* 6, 231.

Waller, L. A., B. P. Carlin, H. Xia, and A. E. Gelfand. (1997). "Hierarchical Spatio-Temporal Mapping of Disease Rates." *Journal of the American Statistical Association* 92, 607–17.

Whitaker, S. (2014). "Big Data versus a Survey." FRB of Cleveland Working Paper No. 14–40.

Wikle, C. K. (2015). "Modern Perspectives on Statistics for Spatio-Temporal Data." *Wiley Interdisciplinary Reviews: Computational Statistics* 7, 86–98.

Wikle, C. K., S. H. Holan, and N. A. C. Cressie. (2013). "Hierarchical Spatio-Temporal Models and Survey Research." *Statistics Views*.

Williams, D., J. Haworth, and T. Cheng. (2015). "Exploratory Spatiotemporal Data Analysis and Modelling of Public Confidence in the Police in Central London." Presented at the *23rd GIS Research UK (GISRUK)*, The University of Leeds, Leeds, UK.

Zaman, T., E. B. Fox, and E. T. Bradlow. (2014). "A Bayesian Approach for Predicting the Popularity of Tweets." *The Annals of Applied Statistics* 8(3), 1583–11.