Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

USING DISCRETE EVENT SIMULATION TO EVALUATE TIME SERIES FORECASTING METHODS FOR SECURITY APPLICATIONS

Samuel H. Huddleston

Center for Army Analysis
U.S. Army
Fort Belvoir, VA 24061, USA

Donald E. Brown

Department of Systems and Information Engineering
University of Virginia
Charlottesville, VA 22904, USA

ABSTRACT

This paper documents the use of a discrete event simulation model to compare the effectiveness of forecasting systems available to support routine forecasts of criminal events in security applications. Military and police units regularly use forecasts of criminal events to divide limited resources, assign and redeploy special details, and conduct unit performance assessment. We use the simulation model to test the performance of available forecasting methods under a variety of conditions, including the presence of trends, seasonality, and shocks. We find that, in most situations, a simple forecasting method that fuses the outputs of crime hot-spot maps with the outputs of univariate time series methods both significantly reduces modeling workload and provides significant performance improvement over the three currently used methods: naive forecasts, Holt-Winters smoothing, and ARIMA models.

1 INTRODUCTION

Every day, government executives, police officials, and military leaders must decide how to most efficiently and effectively employ their limited resources in an effort to secure the large and diverse populations they are charged to protect. Planning and decision-making processes for these leaders are often oriented around dividing limited resources across subordinate commands and geographic regions. Military and police leaders regularly rely on recurring forecasts of criminal events indexed by geographic region to support these processes. For example, Gorr, Olligshlaeger, and Thompson (2003) note the many applications of these forecasts for police use including: efficient resource allocation, geographic mission assignment planning, and unit performance assessment. Therefore, both military and police units stand to benefit from accurate models for producing regular forecasts of geographic time series.

Geographic time series are counts of events indexed over time by geographic region. Often, geographic time series are very noisy, with the observed counts by region varying considerably from period to period and region to region. This high variance is to be expected for time series of criminal events since these time series are the result of the superposition of many low-intensity point processes. The Palm-Khintchine theorem asserts that the superposition of many low-intensity independent renewal processes behaves asymptotically as a Poisson process (Heyman and Sobel 1982). Thus, when the actions of many criminals acting independently in geography generate a time series, the time series should exhibit the noisy behavior of a Poisson process, in which the variance of event counts in a period is equal to the average count of a period. These noisy geographic time series are very difficult to forecast accurately.

This paper documents the use of a discrete event simulation model to compare the effectiveness of forecasting systems available to support routine forecasts of criminal events in security applications. The simulation model provides the opportunity to compare the performance of the considered forecasting methods under a wide variety of conditions, including the presence of trends, seasonality effects, and shocks. We find that, in most situations, a simple forecasting method that fuses the outputs of crime hot-spot maps with

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2013		**00-00-2013 to 00-00-2013**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Using Discrete Event Simulations to Evaluate Time Series Forecasting Methods for Security Applications	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Center for Army Analysis,Fort Belvoir,VA,24061**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Proceedings of the 2013 Winter Simulation Conference, 8-11 Dec, Washington, DC.

14. ABSTRACT
This paper documents the use of a discrete event simulation model to compare the effectiveness of forecasting systems available to support routine forecasts of criminal events in security applications. Military and police units regularly use forecasts of criminal events to divide limited resources, assign and redeploy special details, and conduct unit performance assessment. We use the simulation model to test the performance of available forecasting methods under a variety of conditions, including the presence of trends, seasonality and shocks. We find that, in most situations, a simple forecasting method that fuses the outputs of crime hot-spot maps with the outputs of univariate time series methods both significantly reduces modeling workload and provides significant performance improvement over the three currently used methods: naive forecasts, Holt-Winters smoothing, and ARIMA models.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**12**	

the outputs of univariate time series methods both significantly reduces modeling workload and provides significant performance improvement over three currently used methods: naive forecasts, Holt-Winters smoothing, and ARIMA models.

2 PROBLEM DEFINITION

The forecasting problem considered in this paper represents the requirement to regularly forecast criminal event counts for subordinate elements within the context of security operations taking place in a geographic region such as a city or province. To formally define the forecasting problem, let Y_t denote the number of events that occur within a domain of interest D during the time period t. The domain of interest could represent a city, a military area of operation (AO), or a police precinct. The domain of interest D contains sub-regions of the domain indexed by j: $\{D_1, D_2, ..., D_J\}$. These regions represent geographic entities such as patrol districts within a police precinct or the military area of operations for subordinate military units (the regions) within a larger military unit's defined AO (the domain). The quantity of interest is Y_{jt}: the event count for each of the J regions during each time period t. The problem considered is the regular production of one-step-ahead forecasts for the noisy geographic time series indexed by Y_{jt}.

3 CRIME FORECASTING METHODS

The most used forecasting method in police applications are the naive approaches of referencing the crime count from the previous time period (i.e. previous week or month) or the crime count from the same period twelve months earlier (Gorr, Olligshlaeger, and Thompson 2003). Gorr, Olligshlaeger, and Thompson (2003) demonstrate that the Holt-Winters exponential smoothing approach can significantly improve upon the naive forecast in crime forecasts, and the crime analysis and forecasting literature contains almost exclusively time series models from the nested exponential smoothing (Holt-Winters) family. Exceptions to this rule include several applications of ARIMA models used to forecast city-level crime rates (Chen, Yuan, and Shu 2008a; Chen, Yuan, and Shu 2008b; Chamlin 1988).

Exponential smoothing and ARIMA models are the most popular time series methods in both business and crime forecasting for two reasons. First, both methods provide very flexible modeling frameworks that are robust to many types of time series patterns such as trends, seasonality, or unusual changes in the pattern such as the introduction of shocks (Hyndman and Khandakar 2008). Second, even non-statisticians can easily automate Holt-Winters and ARIMA forecasting models using widely available software. For example, Microsoft Excel easily optimizes the parameters of a Holt-Winters smoothing model using Solver and freely available statistical software such as *R* provides algorithms for automatically fitting ARIMA models (Hyndman and Khandakar 2008). Thus, naive, exponential smoothing, and ARIMA methods provide benchmarks for the performance of any new method for recurring short-term demand forecasts in many applications. In this paper, we automatically optimize the needed Holt-Winters (HW) model parameters using the *stats* package in R software, which uses Nelder-Meade optimization to minimize the average squared prediction errors over the previously observed weeks (Nelder and Mead 1965, Hyndman and Khandakar 2008). We fit the ARIMA forecasts by using the *forecast* package in R software, which applies all appropriate ARIMA models to the training data set (the time series from previous weeks), optimizes the parameters for those models, and selects the best model according to the Akaike Information Criteria (AIC) (Hyndman and Khandakar 2008, Akaike 1974).

Recently, Huddleston, Porter, and Brown (2013) introduced a new method for forecasting geographic time series. The new method, Geographic Probability Forecasting (GPF), combines HW or ARIMA univariate time series methods with predictive crime hot-spot maps, tools commonly used in law enforcement applications for identifying areas with a high probability of criminal activity. This new approach and its motivating principle are briefly outlined here. The first step for the GPF method is to develop a spatial probability model (hot-spot map) for event intensity in a geographic region using kernel density estimation. Let b_i index two-dimensional blocks within a spatial study region $D \subset \Re^2$. These two-dimensional spatial

blocks denote unique locations created by laying a grid at a fine resolution across the study domain: $\{b_1, b_2, ..., b_I\}$. Let s_y denote the location in $\Re^2$ of event y and Y the total number of events occurring within D during the time period used to fit the model. The event intensity, $f(b_i)$, for each location is calculated using the following kernel density function:

$$\hat{f}_h(b_i) = \frac{1}{hY} \sum_{y=1}^{Y} K\left(\frac{\|b_i - s_y\|}{h}\right). \tag{1}$$

In Equation 1, the notation $\|b_i - s_y\|$ denotes the Euclidean norm (distance) between location b_i and event s_y. Model fitting requires the selection of the kernel function K and the bandwidth parameter h. The choice of the kernel function K has relatively little effect on the kernel density model performance, with the default kernel function usually the standard Gaussian probability density function. The bandwidth parameter h does significantly affect model performance. Various statistical procedures automate the selection of the modeling parameter using plug-in estimates (Sheather and Jones 1991, Jones, Marron, and Sheather 1996, Berman and Diggle 1989). The approach demonstrated here selects the plug-in estimate for bandwidth that minimizes the Mean Squared Error (MSE) of the hot-spot map over the previously observed time horizon using the procedure outlined by Berman and Diggle (1989).

The second step of the GPF method is to estimate the risk for each geographic region. The weighting parameter $\hat{w}_j$ represents the spatial probability-weight for geographic region j derived from the kernel density hot-spot map.

$$\hat{w}_j = \frac{\sum_{b_i \in D_j} \hat{f}_h(b_i)}{\sum_{b_i \in D} \hat{f}_h(b_i)}. \tag{2}$$

The weighting factor $\hat{w}_j$ captures the proportion of overall event probability across the domain that falls within the geographic region D_j. Note that this risk weighting both converts the kernel density estimate into a probability estimate over the domain of interest and removes any of the risk probability that falls outside of the considered domain (because the kernel density estimate will map onto a square grid whose boundaries extend beyond the considered domain). GIS systems easily automate the calculation of the weighting factor $\hat{w}_j$ for any subset of the domain (precincts, patrol sectors, etc).

The third step in the GPF methodology is to develop a domain (city-wide) forecast for event counts. The domain level forecast for time period t, F_{Dt}, is fit using either the HW or ARIMA method. The final step of the GPF methodology is to break the domain-level forecast across the geographic regions using the probability weights $\hat{w}_j$. The forecast for region j in the next time period is a function of the domain level forecast for that time period $\hat{F}_{Dt}$ and the estimated probability weight $\hat{w}_j$:

$$F_{jt} = \hat{w}_j \hat{F}_{Dt}. \tag{3}$$

This modeling approach incorporates several modeling assumptions that should be addressed. First, the GPF model assumes that the spatial distribution of the events (as described by the kernel density hot-spot map) remains fixed over the period used to fit the model and forecast. The size of this modeling horizon is flexible, but in this paper we used the entire time series horizon up to the current time period to forecast the next time period.

The second assumption is that any underlying trend or seasonality that affects event counts in one region affects all regions. Thus, any existing trend (or seasonal effect) applies to all regions (i.e., the entire domain) simultaneously. The GPF model therefore models the crime counts as a separable space-time process in which the spatial probability density of events is fixed (or changes slowly) while the distribution of counts (rate) can change rapidly.

The motivating principle for the GPF methodology lies in exploiting the reduction in error variance gained by forecasting at a higher level and dividing this forecast according to a geographic probability. Examining the simple problem of estimating the mean of a process based upon a sample of observed counts

in a geographic time series provides insight into how the GPF methodology improves forecasts. Consider a sequence of observations X_t that come from a Gaussian distribution $N(\mu, \sigma^2)$. The error for the sample mean estimated during time t is bounded for an arbitrary value ε as follows (Shumway and Stoffer 2006):

$$P\{|\bar{x} - \mu| > \varepsilon\} \leq E\left[(\bar{x}_t - \mu)^2\right] = \frac{\sigma^2}{t\varepsilon^2}. \tag{4}$$

Now consider that the sequence of events that make up X can be split between several geographic regions. The count of observations X_t is made up of the sum of the counts in the several regions. For regions indexed by j:

$$X_t = \sum_{j=1}^{J} X_{jt}. \tag{5}$$

If the mean of the distribution for region j is some known fixed percentage w_j of the mean of the distribution for the domain, then one can use the sample mean of the higher distribution to develop an estimate of the mean for the region. For region j, assert that X_{jt} is Gaussian distributed $N(\mu_j, \sigma_j)$. Define that:

$$\mu_j = w_j \mu. \tag{6}$$

Two approaches for estimating the mean for region j now exist. One can use the estimate $\bar{x}_{jt}$ to estimate μ_j during time t or one can weight the domain estimate such that:

$$\hat{\bar{x}}_{jt} = w_j \bar{x}_t. \tag{7}$$

Using Equations 6 and 7, the error bound of the weighted estimate is:

$$P\{|w_j \bar{x}_t - \mu_j| > \varepsilon\} \leq E\left[(w_j \bar{x}_t - \mu_j)^2\right] = w_j^2 \frac{\sigma^2}{t\varepsilon^2}. \tag{8}$$

The error bound for the region sample mean based upon the observations (counts) in that region is:

$$P\{|\bar{x}_{jt} - \mu_j| > \varepsilon\} \leq E\left[(\bar{x}_{jt} - \mu_j)^2\right] = \frac{\sigma_j^2}{t\varepsilon^2}. \tag{9}$$

Therefore:

$$P\{|w_j \bar{x}_t - \mu_j| > \varepsilon\} \leq P\{|\bar{x}_{jt} - \mu_j| > \varepsilon\} \Leftrightarrow w_j^2 \leq \frac{\sigma_j^2}{\sigma^2}. \tag{10}$$

For noisy geographic time series, this inequality holds true. A numerical example illustrates the dramatic reduction in estimation error that this weighting estimate can provide. If the counts for the domain come from a stationary (temporal component) homogenous (geographic component) Poisson process with mean and variance λ, then for a Poisson process with a rate greater than 10, the counts observed during each period are approximately Gaussian distributed as $N(\mu = \lambda, \sigma^2 = \lambda)$. If the domain is divided into four equal geographic regions, then for each region, $\lambda_j = \frac{1}{4}\lambda$. In this example, by substitution into Equations 8 and 9:

$$P\{|w_j \bar{x} - \mu_j| > \varepsilon\} = w_j^2 \frac{\lambda}{t\varepsilon^2} = \frac{1}{16}\frac{\lambda}{t\varepsilon^2} < \frac{1}{4}\frac{\lambda}{t\varepsilon^2} = w_j \frac{\lambda}{t\varepsilon^2} = P\{|\bar{x}_j - \mu_j| > \varepsilon\}. \tag{11}$$

In the example above, both sample estimates of the mean ($\bar{x}$ and $\bar{x}_j$) are converging to their respective means (μ and μ_j). However the probability-weighted estimate $\hat{\bar{x}}_j = w_j \bar{x}$ converges to μ_j much faster than than $\bar{x}_j$ does. This result holds for any stationary Poisson process, regardless of the observed rates, because the relationship described in Equation 10 holds true for any Poisson process in which the domain counts

are made up of the sum of sub-domain counts. The GPF methodology exploits this principle to improve the forecasts for noisy geographic time series by using kernel density estimation to estimate w_k.

Ideally, the relationship between the convergence of a GPF forecast and the convergence of a sub-domain forecast using ARIMA or HW models could be established using estimation theory as demonstrated for estimating the mean above. However, several significant roadblocks prevent the development of closed-form proofs. First, the bandwidth selection problem for kernel density estimation is confounded with the convergence of the kernel density surface, so the use of a plug-in bandwidth selection procedure confounds the assessment of the convergence of the probability weighting parameter (Mason 2003). Second, the HW and ARIMA models are employed algorithmically, using step-wise model selection procedures. Step-wise model selection does not guarantee the selection of the optimal model. Furthermore, the HW or ARIMA models selected by the algorithmic procedure can (and do) vary from step to step and region to region, thus requiring comparisons such as comparing the convergence properties of an ARIMA model with an estimated trend at the domain level to an ARIMA model with no trend estimated for a region. However, simulation modeling provides the opportunity to compare the performance of the different forecasting approaches to assess performance under a variety of conditions.

4 SIMULATING NOISY GEOGRAPHIC TIME SERIES

Using a simulation model to study the properties of these forecasting methods offers three significant benefits. First, with a simulation model, one can easily vary the conditions of the simulation and observe the resulting effects on the performance of the methods. Within the simulation model, not only can one generate noisy geographic time series that include trends, seasonality, and shocks but one can vary the intensity of these effects at will. Second, in a simulation model, a *known* process generates the various time series. So, one can evaluate forecasting methods on how well they model a known process instead of conducting performance comparisons against observed counts in an observational setting for which the true spatial-temporal process is unknown. Removing the random noise from the evaluation measures is especially helpful when evaluating performance against exceptionally noisy processes such as Poisson event counts. Finally, simulation models replicate, repeatedly generating simulated outcomes from the same processes. This replication facilitates the study of the convergence properties of the forecasting methods.

Figure 1 provides a visualization of one of the simulation models developed for this analysis. This graphic illustrates the state of the simulation model in period 1 and after 50 periods of observation. The simulation environment contains a geographic extent (from -100 to 100 in x and y coordinates), a domain of interest (from -60 to 60) and four smaller equal sized geographic regions. The modeling problem is to accurately forecast the observed counts in each of the four spatial regions during each time step. The simulation model provides the opportunity to vary many modeling parameters, including the number of spatial/temporal processes and the location, spatial distribution, and rate for each spatial process. The model in Figure 1 contains five spatial processes, each of which has a unique spatial distribution. Figure 2 shows the known process rate and resulting event counts due to the noisy process in Region 3 over the first 50 time steps of the simulation.

Once the number of processes is defined for a given experiment, the simulation models randomizes the location and dispersion of those spatial processes by uniformly selecting parameters for the Gaussian spatial processes from the following intervals: $\mu_{x,y} = U(-60,60)$, $\sigma_{x,y} = U(5,30)$, $\rho = U(-.5,.5)$. In the notation above, μ denotes the location of the center of the spatial process, σ the spatial variance, ρ the covariance parameter, and $U(-,-)$ selection from the uniform distribution. The notation for the Gaussian spatial process is summarized as $N(\mu_p, \Sigma_p)$, where Σ_p denotes the covariance matrix for spatial process p containing ρ_p, σ_x, and σ_y. Note that given these distributions, some portion of the spatial process may overlap with the domain boundary, so that some of the simulated incidents fall outside of the domain of interest for the forecasting problem. The spatial distribution of the processes remains fixed over the conduct of an experiment (although the simulation model generates many randomized replicates of the geographic time series within each experiment), so the spatial processes do not migrate or shift.

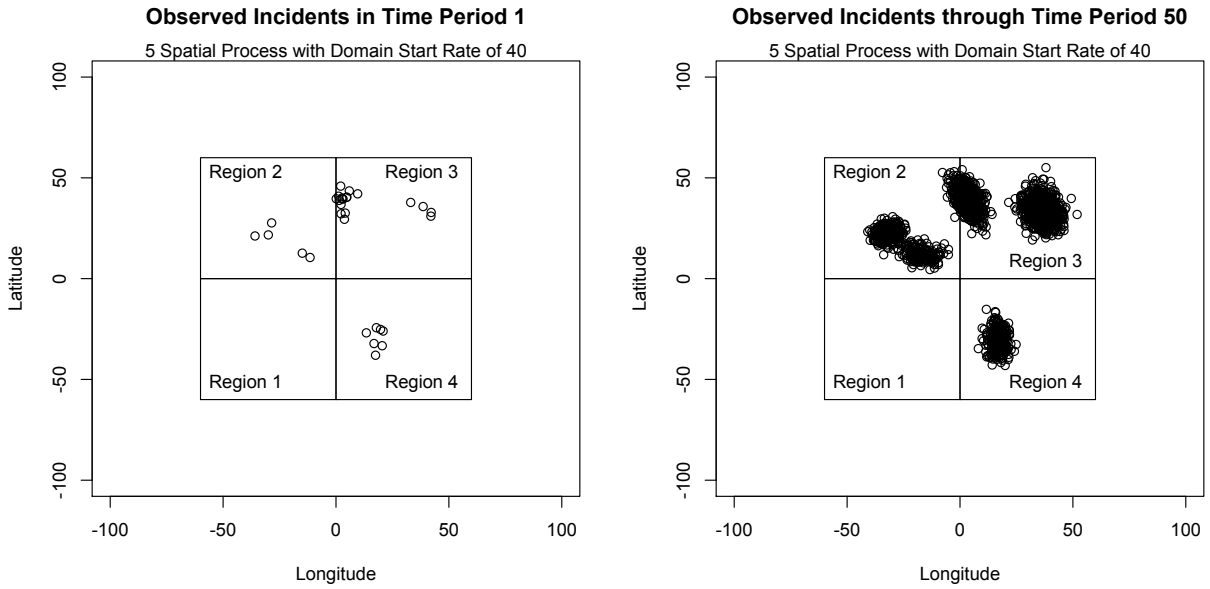

Figure 1: Graphical illustration of a non-homogenous point process model during time step 1 (left panel) and time step 50 (right panel).

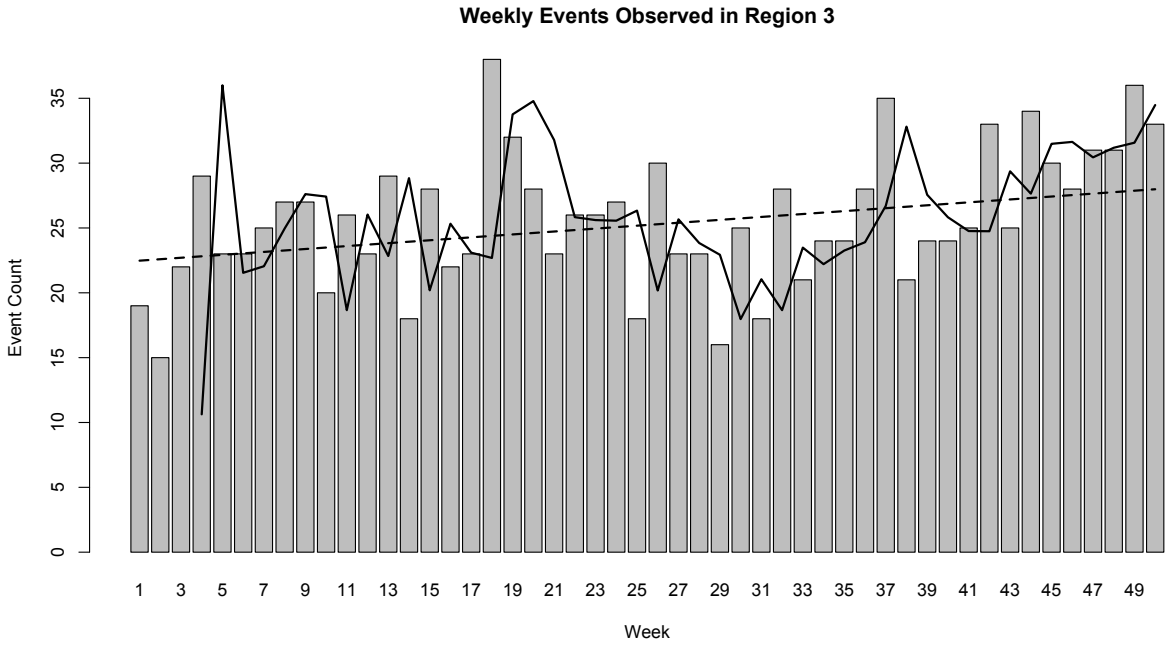

Figure 2: Observed event counts (barplot), known process mean (dashed line), and Holt-Winters forecast (solid line) for Region 3 for 1 replicate of the simulation illustrated in Figure 1. The observed error is the difference between the forecast (solid line) and the observed event count (barplot). The process error is the difference between the forecast and the known process mean (dashed line).

The rate of the spatial processes can be controlled dynamically, introducing trends, seasonality, or shocks into the process by adjusting the intensity (rate) parameter λ_p over time. Note Figure 2 shows a positive trend, although the actual observed counts fluctuate wildly, reflecting the noise of the Poisson process. During each model step t, the simulation model randomly draws an event count for each process from the Poisson distribution defined for that process by the rate parameter $\lambda_p(t)$ and randomly places each of those events within the model geography in accordance with the spatial distribution defined by a the individual processes' spatial distribution model. These spatial distribution models are Gaussian: $N(\mu_p, \Sigma_p)$. As observed in Figures 1 and 2, these processes are noisy.

The use of square regions and known process parameters facilitates direct calculation of the expected counts for each region D_j for each time step. For example, with P spatial processes taking place in the domain during time t, and each of these spatial processes Gaussian distributed in space with a Poisson arrival rate λ_p, the expected count for a given geographic region during time period t is:

$$E[Y_{jt}] = \sum_P \lambda_p(t) \int_{D_j} N(\mu_p, \Sigma_p). \tag{12}$$

Table 1 contains the Design of Experiments (DOE) for considered cases. This table reports the number of experiments conducted at each unique setting depicted in the table as well as the number of replicates that make up each experiment. As the DOE table depicts, experimental blocks include three different domain rates (20, 40, and 80 events per period) and five different numbers of spatial processes (5, 20, 40, 80, and 100). This blocking enables performance comparison under a wide spectrum of conditions, from those in which region counts are high and processes are intense (i.e., a domain rate of 80 and number of processes is 5) to those when region counts are low and each process is very intermittent. For each experiment, the simulation randomly samples for the spatial location (spatial mean), spatial distribution, and domain rate for the defined number of spatial processes, scaling the rates of the individual processes so that they add up to the rate defined for the experimental block. The simulation then produces many geographic time series (replicates) for each experimental setting over 100 time periods, which provides 97 time periods per replicate over which to evaluate one-step-ahead forecasts.

The first six scenarios presented in Table 1 represent situations in which the assumptions of the GPF method apply (i.e. any existing trend or seasonality effects are global effects that impact all regions). These scenarios include stationary time series, the introduction of trends, the introduction of seasonality effects, and scenarios that include both trends and seasonal effects. Positive trends are scaled so that the expected event count of the process increases by 100% (i.e., the event count in period 100 is double that during period 1). Negative trends are scaled such that the expected event count of the process decreases by 50% over the time horizon, so the expected count in period 100 is $\frac{1}{2}$ that expected in period 1.

The seasonality effects are sinusoidal functions, with two different amplitudes and cycle frequencies considered. The high-amplitude setting [identified using the notation (H) in the DOE table] uses an amplitude equivalent to $\frac{1}{2}$ of the process mean while the low-amplitude setting (L) uses an amplitude of $\frac{1}{4}$ of the process mean. The $\frac{1}{50}$ cycle frequency corresponds to completing one cycle every 50 periods. This mimics the annual temperature cycle which has been shown to affect crime rates, with two full cycles over the evaluation period. The $\frac{1}{25}$ cycle completes four cycles over the evaluation period. Simulation experiments containing seasonality are evaluated from time steps 151 through 250 (100 periods) to provide enough observations for estimating seasonality effects using the HW method.

The last three scenarios recorded in Table 1 consider cases in which the GPF modeling assumptions do not apply. The experiments investigating the impact of shocks contain a stationary homogenous process throughout the domain with a Poisson rate of 40 events per time period throughout the domain. At time period 50, a new non-homogenous spatial process is introduced in the domain centered in Region 3 and scaled so that the new spatial process only affects that region. Shock process rates vary from 5% of the domain rate (i.e., a rate of 2 events per time period) to 30% of the domain rate (i.e., the rate within Region 3 more than doubles from 10 to 22). Note that with the introduction of the shock process, both of the

Table 1: Design of Experiments (DOE) for simulation study cases showing the number of experiments and replicates per experiment (in parentheses) for all studied cases. Each experimental setting listed in the table columns is repeated for each of the domain rates or strength of trend settings listed in the left-most column.

Stationary Homogenous Point Process

Domain Rate	20	40	80	100
Experiments (Replicates)	1 (100)	1(100)	1(100)	1(100)

Stationary Non-Homogenous Point Processes

	Number of Spatial Processes			
Domain Rate	5	20	40	100
20, 40, 80	10 (50)	10 (50)	10 (50)	10 (50)

Non-Stationary Non-Homogenous Point Processes with Trend

	Positive Trend (100%)		Negative Trend (50%)	
Domain Start Rate	Number of Spatial Processes			
	20	80	40	100
20, 40, 80	10 (50)	10 (50)	10 (50)	10 (50)

Non-Stat. Non-Homog. Point Processes with Seasonality

	Cycle Freq. = 1/50		Cycle Freq. = 1/25	
Domain Start Rate	Number of Spatial Processes (Amplitude)			
	20 (L)	80 (H)	40 (H)	100 (L)
20, 40, 80	1 (50)	1 (50)	1 (50)	1 (50)

Non-Stat. Non-Homog. with Positive Trend and Seasonality

	Cycle Freq. = 1/50		Cycle Freq. = 1/25	
Domain Start Rate	Number of Spatial Processes (Amplitude)			
	20 (L)	80 (H)	40 (H)	100 (L)
20, 40, 80	1 (50)	1 (50)	1 (50)	1 (50)

Non-Stat. Non-Homog. with Negative Trend and Seasonality

	Cycle Freq. = 1/25		Cycle Freq. = 1/50	
Domain Start Rate	Number of Spatial Processes (Amplitude)			
	20 (H)	80 (L)	40 (L)	100 (H)
20, 40, 80	1 (50)	1 (50)	1 (50)	1 (50)

Shocks

Size of Shock	5%	10%	20%	30%
Experiments (Replicates)	1 (50)	1 (50)	1 (50)	1 (50)

Competing Trends

	Unique Process Rate as % of the Global Rate			
Strength of	10		25	
Negative Trend	Strength of Positive Trend			
	100%	200%	100%	200%
25%, 50%	10 (50)	10 (50)	10 (50)	10 (50)

Random Trends

	Number of Spatial Processes			
Domain Rate	20	40	80	100
20, 40, 80	10 (50)	10 (50)	10 (50)	10 (50)

model assumptions of the GPF method are violated: the spatial distribution of crimes is no longer fixed over the time horizon used to fit and forecast the model and a phase change has occurred in one geographic region that does not affect the other regions.

The next scenario investigates competing trends in the various regions. These experiments contain a non-stationary homogenous process in the domain (with a Poisson rate of 40 events per time period at the model start time). This Poisson process has a negative trend that reduces the Poisson rate by either 25% or 50% over the study time horizon. At the same time, a non-homogenous Poisson process with a positive trend is introduced into Region 3. Thus, Region 3 has a trend that is moving in the opposite direction of the rest of the domain. Several different rates and strengths of trend for the process in the unique region are studied, as depicted in the table.

The final scenario randomly assigns trends to every unique process in the domain. The number of spatial processes varies from 20 to 100 and the overall domain rates (at the start of the simulation) vary from 20 to 80. The simulation model randomly selects a unique trend for each spatial process from $U(-50\%, 100\%)$. Thus, a spatial process is equally likely to double in intensity or halve in intensity over the simulation time horizon. Each spatial process is therefore a non-stationary, non-homogenous Poisson process, with the spatial location and spatial distribution parameters randomly chosen for each spatial process at the beginning of the experiment. Thus, each region also has a unique trend, and the overall trend in the domain can be positive or negative but on average should be slightly positive.

5 RESULTS

Results are presented here using the Mean Absolute Scaled Error (MASE) as the performance assessment statistic. Hyndman and Koehler (2006) propose the MASE as a "generally applicable measurement of forecast accuracy without the problems seen in the other measurements." MASE provides the ideal statistic for assessing forecasting performance in this problem due to several properties. First, MASE is a scale-free

Table 2: MASE performance summary for all experimental conditions (less is better).

Scenario	Naive	HW	GPF-HW	ARIMA	GPF-ARIMA
Stationary Homogenous	1.00	0.80	0.42	0.25	0.18
Stationary Non-Homogenous	1.00	0.79	0.44	0.25	0.20
Non-Homogenous Trend	1.00	0.80	0.43	0.42	0.29
Non-Homogenous Seasonality	1.00	0.67	0.35	0.44	0.26
Pos. Trend & Seasonality	1.00	0.63	0.32	0.40	0.28
Neg. Trend & Seasonality	1.00	0.77	0.40	0.52	0.35
Competing Trends	1.00	0.82	0.58	0.41	0.46
Random Trends	1.00	0.78	0.45	0.35	0.28
Shocks	1.00	0.81	0.60	0.31	0.42

measure of forecast error, which allows forecast accuracy comparisons between series. This property is important in this case because event counts occur at different scales depending on the domain rate used and the spatial distribution of event processes (which are randomized for each experiment). Using the scale-free MASE statistic allows us to aggregate performance by averaging MASE performance over the different geographic regions and over experimental runs, which is not possible using scale-dependent error measures such as Root Mean Squared Error (RMSE) or Mean Squared ERROR (MSE). Secondly, MASE can be used on intermittent time series, which contain observed event counts of 0 in at least one time period. This occurs frequently when the domain event rates are low. Finally, MASE scales the error by the Naive forecast. A MASE score less than one indicates a model that has smaller average error than the Naive method, with the reverse true when $MASE > 1$.

Each forecasting method is assessed on how well the method performs at modeling the known process taking place in each region over many replicates for a given simulation experiment. Figure 2 provides an illustration of the difference between the process error (the difference between the known process rate and the forecast) and the observed error (the difference between the observed count and the forest). The MASE statistic for one of the geographic regions averaged over all replicates (indexed by R) and all time periods (indexed by T) is:

$$MASE_j = \frac{1}{RT}\sum_{r=1}^{R}\sum_{t=1}^{T}\left(\frac{|E[Y_{jt}] - F_{jrt}|}{|E[Y_{jt}] - Y_{jr(t-1)}|}\right). \tag{13}$$

Table 2 summarizes the performance of all of the considered forecasting methods using the MASE statistic. Several important conclusions can be drawn from the results in this table. First, all of the univariate forecasting methods improve upon the Naive method, implying that police and military units can improve their forecasting over this popular approach. Second, the ARIMA method uniformly improves upon the HW method. The performance improvement the ARIMA method provides over the HW method in these tables does not reflect that the ARIMA class is a better method than the HW method in general [as Hyndman and Khandakar (2008) notes, they are overlapping model classes]. Rather, the observed performance improvement is due to the more complex model fitting procedure employed to fit the ARIMA models in this study.

When the GPF modeling assumptions apply, the GPF method improves forecasting performance when applied to both HW and ARIMA models. Additionally, when seasonality effects are present, the GPF-HW method improves upon the ARIMA method. This improvement is largely to the improved ability to accurately model seasonality effects at the domain level. Figure 3 illustrates the performance improvement that the GPF method provides when the GPF modeling assumptions apply. Figure 3 provides time series plots for the forecasts generated using ARIMA (at the domain and region level) and GPF-ARIMA for a stationary homogenous point process. As can be seen in the figure, the GPF-ARIMA forecast at the region level is a scaled version of the forecast at the domain level (in this case $w_j = \frac{1}{4}$). The variance of the

GPF-ARIMA forecast at the region level is less than that for the ARIMA forecast. Note that both forecasts are converging to the true process rate, but the GPF-ARIMA method is converging faster.

The right panel of Figure 3 plots ratios of the ARIMA and GPF-ARIMA Mean Squared Error (MSE) to the domain-level ARIMA model MSE. This is the relationship explored in Equations 10 and 11 and the experimental results depicted in Figures 3 are for the example given in Equation 11. With a smoothly converging estimator (such as the sample mean) and a perfect estimate for w_k, we would expect these ratios to conform closely to the dashed lines on the graph. As can be seen, the MSE ratio for the ARIMA method noisily oscillates around the expected value, while the GPF-ARIMA MSE ratio does not fully achieve the performance gains expected given Equation 11 due to the modeling error of the kernel density estimate. In spite of this, the improvement provided by the GPF-ARIMA method is significant: in this case the GPF-ARIMA sub-region MSE averages $\frac{1}{10}$ of the domain ARIMA MSE while the ARIMA sub-region MSE averages $\frac{1}{4}$ of the domain ARIMA MSE. This pattern is repeated for all of the experiments depicted in Table 1: the GPF method MSE ratio is a value slightly higher than (sometimes within one percent of) w_k^2 while the ARIMA or HW MSE ratio noisily oscillates around and averages w_k. The left panel of Figure 4 provides an illustration of how the GPF method improves forecasting in a scenario containing both trend and seasonality effects.

When the GPF modeling assumptions do not apply, the GPF method still improves forecasting when applied to HW models. However, when shocks or competing trends exist, the ARIMA method provides better performance than the GPF-ARIMA method. The right panel of Figure 4 illustrates why the GPF method fails in these cases. In the figure, when the shock is introduced at time period 50 (which changes the process mean event count from 10 to 22), the weighting factor w_k becomes very biased. This bias is slowly reduced over time, but the effect of the 50 initial time periods on the estimate for w_k is very strong. This suggests that using rolling time horizons to estimate w_k might provide some performance improvement and insulation against violations of the GPF modeling assumptions.

6 CONCLUSIONS

The use of a simulation model for this problem provides an invaluable contribution to studying the properties of these forecasting methods. As discussed, closed form estimation theory proofs for performance comparison are not possible. Collecting real-world data sets for the many different scenarios (non-homogenous, homogenous, stationary, trends, seasonality, shocks etc.) investigated in this paper would be very difficult. The simulation model provides the opportunity to identify the scenarios in which each of the methods provides the best performance.

The results of this simulation study indicate that, when the modeling assumptions of the GPF method apply, the GPF method significantly improves forecasting performance for both ARIMA and HW smoothing models. The simulation study confirms that this performance improvement mimics the performance improvement suggested by the motivating principle of exploiting the reduction in error variance to be gained by geographically weighting a domain level forecast, with some loss in performance due to the kernel density estimate of the weighting factor. An additional significant advantage of the GPF method is the significant reduction in modeling workload. The GPF method requires fitting only one forecasting model and one kernel density estimate, while the ARIMA and HW methods require generating unique forecasts for every geographic region. In this simulation model this effect is small, but in applications such as forecasting crime counts for patrol sectors in a large city, this can reduce the modeling workload by up to 50 unique models (Huddleston, Porter, and Brown 2013).

The results of the simulation study also indicate that, although the GPF method improves forecasting performance for the HW method under all studied scenarios, the ARIMA method provides better performance than the GPF-ARIMA method when shocks or competing trends occur. As demonstrated, this is due to the bias introduced to the geographic weighting factor when a shock occurs or over time as the competing trends in different regions diverge. Future work includes investigation of the performance of the GPF

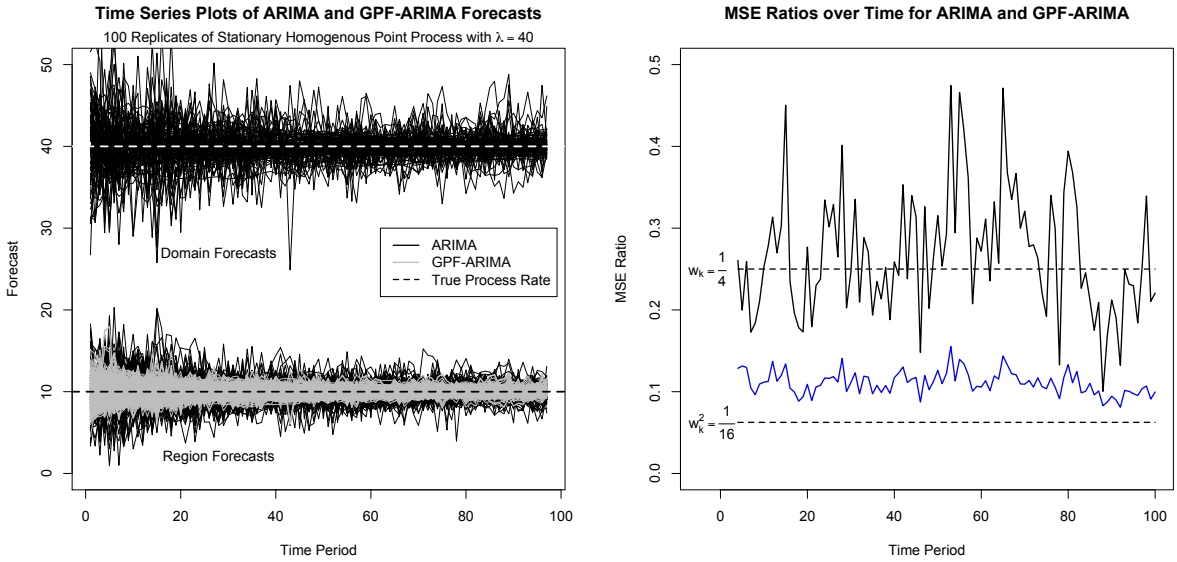

Figure 3: Left Panel: Time series plot comparisons of domain (top of the graphic) and region (bottom of the graphic) forecasts using ARIMA (in black) and GPF-ARIMA (in gray) models. Right Panel: Time series plots of the resulting mean squared error (MSE) ratio for the Sub-Region MSE divided by the Domain MSE for the ARIMA method (top of the graph) and the GPF-ARIMA method (bottom of the graph) during each time step using the 50 experiment replicates. The dashed lines indicate the expected MSE ratios given Equation 11.

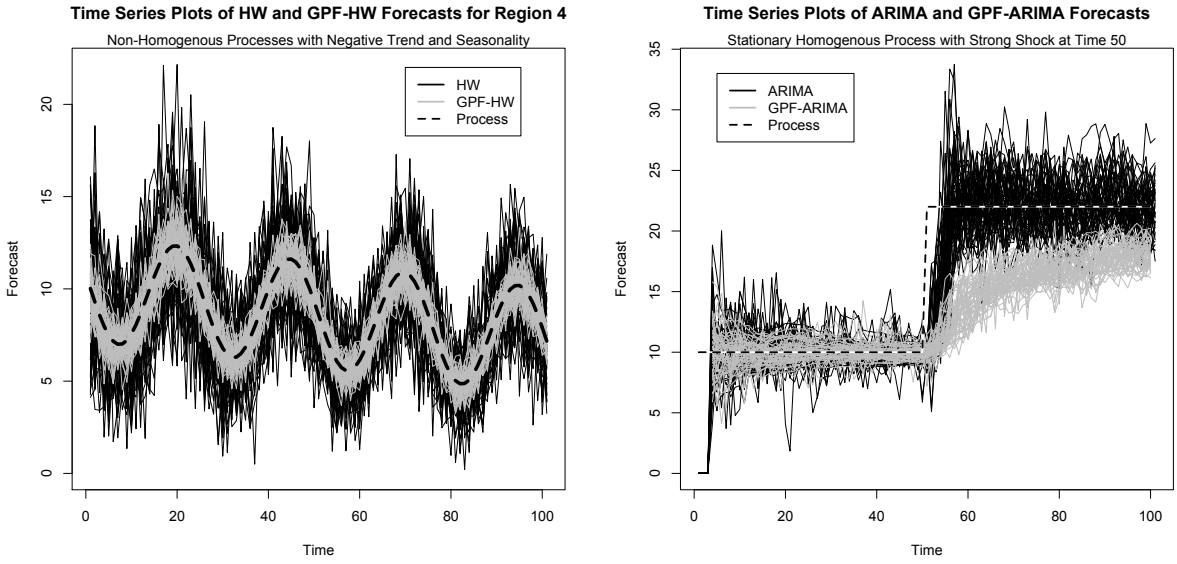

Figure 4: Time series plots of HW and GPF-HW forecasts in a scenario with negative trend and seasonality effects (left panel) and ARIMA and GPF-ARIMA forecasts in a scenario including the introduction of a strong shock process at time period 50 (right panel). As can be seen in the figure, when the GPF assumptions apply (left panel), the GPF method significantly reduces the model error. When the GPF assumptions are violated, the bias in the weighting factor $\hat{w}_j$ has a very negative effect on the GPF methods forecasting performance.

method using rolling horizons to fit the kernel density estimate in an effort to mitigate the bias resulting from shocks or competing trends.

REFERENCES

Akaike, H. 1974. "A new look at the statistical model identification". *IEEE Transactions on Automatic Control* 19 (6): 716–723.

Berman, M., and P. Diggle. 1989. "Estimating Weighted Integrals of the Second-order Intensity of a Spatial Point Process". *Journal of the Royal Statistical Society. Series B (Methodological)* 51 (1): 81–92.

Chamlin, M. B. 1988. "Crime and Arrests: An AutoRegressive Integrated Moving Average (ARIMA) Approach". *Journal of Quantitative Criminology* 4 (3): 247–258.

Chen, P., H. Yuan, and X. Shu. 2008a. "Forecasting Crime Using the ARIMA Model". In *Proceedings of the Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, 627–630: IEEE Computer Society.

Chen, P., H. Y. Yuan, and X. M. Shu. 2008b. "Making Weekly/Daily Forecasting of Property Crime Using ARIMA Model". *Progress in Safety Science and Technology Vol Vii Pts a and B* 7:531–535.

Gorr, W., A. Olligshlaeger, and Y. Thompson. 2003. "Short-Term Forecasting of Crime". *International Journal of Forecasting* 19:579–594.

Heyman, D. P., and M. J. Sobel. 1982. *Stochastic Models in Operations Research: Volume I*. New York: McGraw-Hill.

Huddleston, S. H., J. Porter, and D. E. Brown. Under Review: 2013. "Improving Forecasts for Noisy Geographic Time Series". *The Journal of Business Research*.

Hyndman, R. J., and Y. Khandakar. 2008. "Automatic Time Series Forecasting: The forecast Package for R". *Journal of Statistical Software* 27 (3): 1–22.

Hyndman, R. J., and A. B. Koehler. 2006. "Another look at measures of forecast accuracy". *International Journal of Forecasting* 22 (4): 679 – 688.

Jones, M., J. Marron, and S. Sheather. 1996. "A Brief Survey of Bandwidth Selection for Density Estimation". *Journal of the American Statistical Association* 91 (433): 401–407.

Mason, D. M. 2003. "Representations for Integral Functionals of Kernel Density Estimators". *Austrian Journal of Statistics* 32:131–142.

Nelder, J. A., and R. Mead. 1965. "A Simplex Method for Function Minimization". *Computer Journal* 7:308–313.

Sheather, S., and M. Jones. 1991. "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation". *Journal of the Royal Statistical Society. Series B (Methodological)* 53 (3): 683–690.

Shumway, R. H., and D. S. Stoffer. 2006. *Time Series Analysis and Its Applications*. New York: Springer.

AUTHOR BIOGRAPHIES

SAMUEL H. HUDDLESTON is an Operations Research and Systems Analyst (ORSA) in the US Army. He holds a Ph.D. from the University of Virginia in Systems and Information Engineering. His email address is shh4m@virginia.edu.

DONALD E. BROWN is a Professor in the Department of Systems and Information Engineering, University of Virginia, USA. He serves on the National Research Council Committee on Transportation Security. He is a past member of the Joint Directors of Laboratories Group on Data Fusion and a former Fellow at the National Institute of Justice Crime Mapping Research Center. Dr. Brown is a Fellow of the IEEE and a past President of the IEEE Systems, Man, and Cybernetics Society. He holds a Ph.D. from the University of Michigan. His email address is deb@virginia.edu.