Chapter 24:

The *CrimeStat* Time Series Forecasting Module

Wilpen L. Gorr
H. John Heinz III College
Carnegie Mellon University
Pittsburgh, Pennsylvania

Andreas M. Olligschlaeger
TruNorth Data Systems
Baden, Pennsylvania

Table of Contents

Chapter 24:
The *CrimeStat* Time Series Forecasting Module

Introduction

The *CrimeStat* Time Series Forecasting module is designed for the forecasting of crime counts (or counts of any type of event) and the early detection of unusual levels of activity in current data. A single run detects and forecasts all districts making up a jurisdiction. The module has a single interface page. It requires the user to specify an input file—either the Primary file or another file, identify variables in the file used for forecasting, select a seasonality adjustment, specify an exponential smoothing model, turn on Trigg tracking signal, use default values or choose Trigg parameter values, and save the output. Included in this chapter is an overview of the module. The theory behind the methods is discussed in Chapter 23.

Rationale of the Module

While time series forecasting is useful to police for estimating future crime levels by extrapolating the current time series trends and seasonal adjustments, the impetus for the Time Series Forecasting Module was to provide a detection mechanism for early warning of large changes in crime patterns, either large increases or decreases. Through experience, police generally know seasonal patterns, such as increases in summer months and decreases in winter months, and know if crime is gradually on the increase or decrease. The time series methods in CrimeStat make objective estimates and forecasts for such trends which can be an aid for decision making.

Likely more valuable is early detection of a crime flare up or evidence of an abrupt crime decrease during a police intervention. Forecasting *large crime changes* requires advanced models not included in this module (e.g., see Gorr, 2009) but early detection (also called "early warning") of large crime changes which have already started is quite feasible (Cohen, Garman, & Gorr, 2009). This module uses the Trigg tracking signal, the best of the simple time series monitoring methods, and requires counterfactual forecasts as inputs to make jurisdiction-wide scans of all subareas (districts) for detection. More sophisticated tracking signals exist, in particular the spatial scan statistic (e.g., Neill, 2009), but the Trigg signal is widely used in the private sector for monitoring product demand and has simplicity as a virtue. Used as inputs to the Trigg tracking signal are one-week-ahead or one-month-ahead forecasts for all districts in a jurisdiction as the basis for judging each district's status as being a departure from the existing time trends or not. These forecasts are extrapolations of past patterns and thus represent what would have happened given no pattern change (i.e., are counterfactual). The most recent crime count of each district is compared to a forecast made from data up to but not including the most recent time period. For districts flagged as having a new pattern or large change, the crime analyst can then drill down to details to diagnose problems and determine a course of recommended action.

The advantage of using CrimeStat for detection is that it is automated and objective. It saves the analyst from visually reviewing time series plots for all district time series and making judgments as to what series have unusual changes in crime levels.

Overview of the Module

This module implements the univariate time series models and methods described in Chapter 23. It takes as input space and time series data for weekly or monthly crime counts of each district of a single police jurisdiction. The crime counts are for a single crime type or crime aggregate (such as vehicle thefts, all serious violent crimes, or all property crimes). The districts can be any partition of areas dividing up a jurisdiction, for example police zones, patrol districts, census tracts, or grid cells. The data are stacked by district with a district name or identification number. See data descriptions later in this chapter and the sample data sets provided: "*Pittsburgh monthly crimes by tract 1990-99.dbf*" and "*Pittsburgh weekly crimes by tract 1990-99.dbf*".

CrimeStat provides two exponential smoothing estimation methods for time series models: 1) Simple Exponential Smoothing which estimates a smoothly-varying time series mean for each district and 2) Holt Exponential Smoothing which estimates a smoothly changing time trend line used in forecasting expected change (growth or decline) for each time series. The module computes the smoothing parameters of exponential smoothing to optimize one-step-ahead forecast accuracy. It does so by individual district (district optimization) to achieve the most tailored and widely-ranging forecasts, but also does so with pooled data by jurisdiction to provide robust, stable and often most accurate forecasts and counterfactual forecasts for detection.

The methods use multiplicative Classical Decomposition for seasonal adjustments in two forms, the first using data from the entire jurisdiction and the second using data for each district. Multiplicative seasonal adjustments (factors) are dimensionless quantities, for example, 1.25 which increases a forecasted crime count by 25 percent relative to the underlying mean estimated by exponential smoothing. Jurisdictional seasonal adjustments, while estimated from all district time series summed to the jurisdiction level, are used for each district's seasonal adjustments. District-specific seasonal adjustments apply to each district separately. All model estimates in the module use seasonal adjustments. There are no model degrees of freedom consumed by estimating seasonal adjustments with Classical Decomposition (the method uses averages and ratios in a simple computational procedure to make estimates). If a time series is non-seasonal, then the adjustments will all be near one in value and have no effect on performance.

The most valuable forecasts for tactical deployment of police resources are ***one step ahead***, either one week ahead or one month ahead. The routine chooses smoothing parameters that minimize the sum of squared one step-ahead forecast errors and provides one-step ahead forecasts for <u>each</u> district in its output. If the user wishes forecasts for additional steps ahead, the necessary parameters are available in the output and can be carried out in Excel.

Finally, the module uses the Trigg tracking signal to provide an assessment of each time period in each district as to whether the number of events is expected ("business-as-usual") or is exceptional (e.g., large change). The basis of the tracking signal for a given time period is a ***counterfactual*** forecast for each district made using exponential smoothing on time series data up to but not including the time period in question. The counterfactual forecast is the expected count given business-as-usual conditions in a district. Then if the actual count of the time period is very different than its counterfactual (expected) value, there is a ***signal trip*** that provides evidence of a change in the structure of the time series.

Data Preparation for Time Series Forecasting

The type of data that is needed is recorded events by individual areas over time. Each record must have the number of events that occurred during a single time period for a single area. The events can be crime events (e.g., burglaries, robberies, total part 1 crimes) or they can be other types of events (e.g., motor vehicle crashes, flu incidents). For example, if there are 100 districts that are being monitored and the data are measured by month over a three year period, then there will be 3,600 records, one for each month for each district.

There must be a minimum of three years worth of data for the module to work. The reason is that seasonality (and other parameters) must be estimated with sufficient precision. For example, because a season is defined by the time period (month or week), with three years worth of data each season (month or week) only occurs three times. Clearly, the variability of an estimate based on only a sample size of 3 is very high. That is why having more years worth of data provides more reliable estimates. Three years is the minimum: if there is less than three years, the module will stop and output an error message.

Data for input to the Time Series Forecasting Module need to be prepared using a GIS package or a database package. For example, a discussion and methods are found in chapter 9 of *Preparing Incident Data for Mapping* by Gorr and Kurland (2012). In the module, time series data can be input in two different ways. First, time series data can be input as spatial data as the Primary file. The requirement for either of these is that the X and Y coordinates (centroids) of each study district be listed on each of the records. For example, if there are 100 districts and 52 weeks per area with three years worth of data, then there will be 15,600 records each with an X/Y coordinate listed (100 x 52 x 3).

Second, many time series data will not have spatial coordinates assigned. Consequently, the data can be input as a non-spatial file. In the input file dialogue on the Time Series Forecasting page, the user will choose 'Other' for the input file.

Required Fields

Whether the data is spatial or not and whether there are other data elements listed, each record must incorporate the following four data elements (with names that can be different than those listed):[1]

- **Areal unit** —the name or identifier for the district of the incident
- ***Year***—the year (e.g., 2012)
- ***Season number***—either the month number (1–12) or the week number (1–52)
- ***Event Count***—the count of crimes or other types of events

All districts need data starting at the same period (Year and Week or Year and Month) through to the end period. Note that there can be **no** missing records or data values in records. If a district has zero crimes in a given period, the period for the district must be included with the value zero for the event count.

Note that data aggregation algorithms generally leave out records with zero frequency. One way to add records that have zero frequency is to create a table with all possible districts and time periods[2] and then to left join that table with the aggregated data, forcing all rows of the new table to be in the join. The district and month rows with zero frequency will have the null value and the analyst must replace null values with zeros.

The finished data must be sorted by District, Year, and Season number (Week or Month) in that order.

1 Individual incidents can be aggregated into time periods by district in one of several different types of programs – GIS, database packages such as Microsoft Access, or spreadsheet programs like Excel. The result is that each record includes a count of the number of events that occurred in a particular district for a particular time period. See chapter 9 in Gorr and Kurland (2012) for examples on how to do this.

2 An easy way to create the table with all possible districts and time periods in Microsoft Access is to (1) make a table with all district names or identifiers, (2) make a table with all years; (3) make a table with all time periods (e.g., 1, 2, …, 12 for months; 1,2, ….52 for weeks), and (4) use a make table query that includes data elements from all three tables but has no joins. This is called a "Cartesian product" and has all possible combinations of districts, years, and time periods.

See the sample data sets, "*Pittsburgh monthly crimes by tract 1990-99.dbf* "and "*Pittsburgh weekly crimes by tract 1990-99.dbf*", that are available on the download site. These data meet these requirements. The examples below use the Pittsburgh monthly data.

The week function in Excel creates week 53 of one or two day's length. The analyst must delete points for week 53 because the module only accepts weeks 1–52.

Fields to be Defined

Figure 24.1 shows the interface for the Time Series Forecasting module for a run with monthly data, jurisdiction-wide seasonality, Simple Exponential Smoothing, and with Trigg tracking. There are 10 fields that must be selected for the module to work. These include:

Input file.

This is the file with the data for the Time Series Forecasting module. It can be the Primary file or another file. If it is the Primary file, then it must be a spatial data file with and X and Y coordinates defined on each record. If it is another file, select Other and then identify the file.

Areal unit

This is the variable name of each district being forecasted. For example, in Figure 24.1 the Pittsburgh monthly data set is shown and the areal unit is TRACT, the census tract ID.

Year

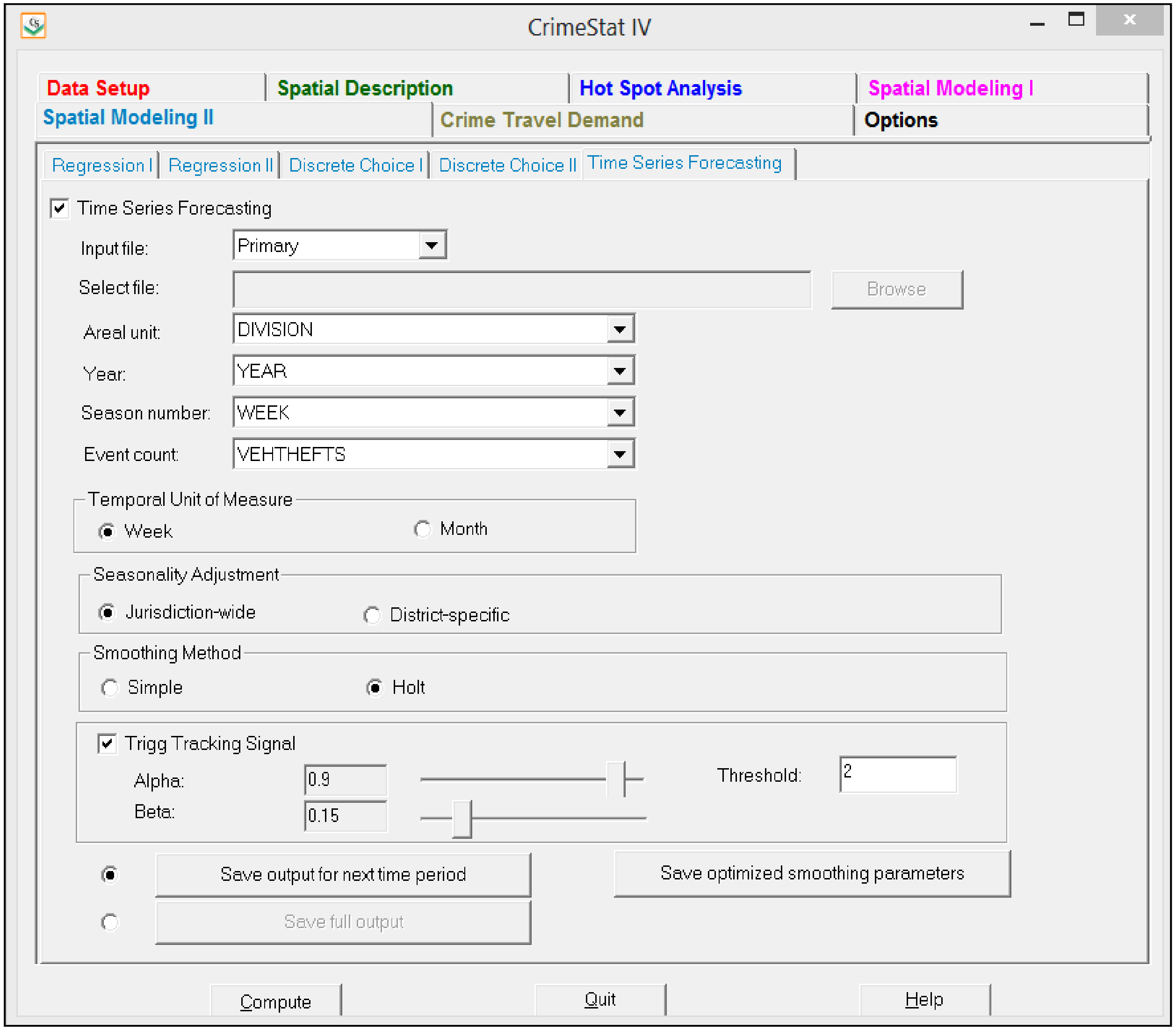
The year is the calendar year such as 2012 of each data record. This must be recorded. As mentioned above, there must be at least three years of data.

Season number

The season number is a unique temporal identifier. With this module, only months or weeks are allowed. Thus, the season number is 1 (January) through 12 (December) for months

Figure 24.1:

Time Series Forecasting



and 1 (first week of the year) through 52 (last week of the year) for weeks. Note that there cannot be partial weeks.

Event count

This is the count of the number of events for a given areal unit, year, and time period.

Temporal Unit of Measure

This defines the type of season used, either week or month.

Seasonality Adjustment

The seasonality adjustment is the adjustment made for each time observation for seasonal patterns such as when, for example, crime is low in February and high in July relative to the time series trend line. The routine uses:

1. **Jurisdiction-wide**. Either data from the entire jurisdiction (e.g., the entire city) and applies this to each district or

2. **District-specific**. Individual data from each district so that each gets its own unique seasonal pattern.

In Figure 24.1, the choice is jurisdiction-wide, which generally provides more accurate forecasts overall because district-specific seasonal factors are overly influenced by individual data points.

Smoothing Method

There are two alternative smoothing models, simple smoothing or Holt exponential smoothing:

1. **Simple smoothing** assumes that there is no trend and that future values will follow the mean of recent past values. Estimated means for data are weighted with weights summing to one for all data points but falling of exponentially with the age of data points.

2. **Holt smoothing** adds a smoothed time trend line into the expected number of future events. The models have smoothing parameters which determine how quickly

exponential smoothing "forgets" the past. The larger a smoothing parameter, the more quickly weights fall off with data point age.

The routine automatically chooses smoothing parameters by minimizing one-step-ahead forecast errors. It uses jurisdiction-level optimization of simple exponential smoothing's smoothing parameter for a single time series for the entire jurisdiction. The result is a smoothing parameter that is relatively small and does a good job of ignoring large changes and, therefore, yields good counterfactual forecasts. If there is a strong time trend (increasing or decreasing) in time series, then Holt is the better choice. Also, given that the option of optimizing smoothing parameters by individual district (instead of by the entire jurisdiction) is only available for Holt in CrimeStat, Holt is the better choice for estimating widely-differing time series patterns across districts and highly dynamic time series patterns within districts. For detection, Holt is likely more conservative because it captures more change in model estimates and, therefore, issues fewer signal trips than Simple Exponential Smoothing.

Trigg Tracking Signal

The Trigg Tracking Signal provides a test statistic for unusual activity in the number of events. If the absolute value of the signal exceeds a pre-specified threshold value, then there is a "signal trip" meaning that it is likely that there is an unusual change in events. The signal has three parameters with default values provided, alpha, beta and the threshold value.

1. Alpha is a smoothing parameter that varies between 0 and 1. An alpha of 0.9 (the default value) makes the tracking signal very reactive to current data on the anticipation of changes in a time series pattern. Note that "Alpha" is the same name for a parameter as used in Simple Exponential Smoothing for forecasting, but here is used to smooth the Trigg tracking signal instead of crime counts.

2. Beta is a smoothing parameter that varies between 0 and 1. A value of beta of 0.15 (the default value) smooths the measure of spread used to standardize the Trigg signal and retains a good amount of history while allowing estimates to drift and follow changing spread in the data.

3. The threshold is the value of the Trigg Tracking Signal that indicates whether the expected number of events will be greater than what is normally expected ("business-as-usual"). The default threshold of 1.5 is somewhat liberal in the sense that it will signal more periods of unusual activity. However, most police organizations would rather respond to more expected events even if the increased activity does not materialize (i.e., are false positives) than not respond and have events blow up. To use more conservative values, try 1.75 or 2.0 to get fewer signal trips.

Cohen, Garman, and Gorr (2009) found that the default values in the CrimeStat module are the best-performing parameter values. However, the user can experiment. Making alpha smaller than 0.9 will reduce the importance of recent events and give more sensitivity to increases building over several time periods. Similarly, increasing beta above 0.15 will smooth the data less and make the Trigg more reactive to changing variability in time series.

Running the Time Series Forecasting Module

With variations in seasonal adjustment and smoothing method, there are 8 possible models that can be run: four for weekly data and four for monthly data (Table 24.1).

Table 24.1:
Time Series Forecast Combinations

Season Number	Seasonality	Smoothing Method	Optimization
Weekly data	Jurisdiction-wide	Simple smoothing	Jurisdiction
Weekly data	District-specific	Simple smoothing	Jurisdiction
Weekly data	Jurisdiction-wide	Holt smoothing	District
Weekly data	District-specific	Holt smoothing	District
Monthly data	Jurisdiction-wide	Simple smoothing	Jurisdiction
Monthly data	District-specific	Simple smoothing	Jurisdiction
Monthly data	Jurisdiction-wide	Holt smoothing	District
Monthly data	District-specific	Holt smoothing	District

Output

There are three types of output: full, one-step ahead, and the optimized smoothing parameters. The first two outputs produce the following calculated values:

1. DE_SEASON is the number of events per period (EVENTCOUNT) divided by the seasonal factor for the current observation's season (December) and, thus, is a de-seasonalized count of events. To calculate the seasonal factor for each record divide EVENTCOUNT by DE_SEASON.

2. SMTH_LEVEL is the smoothed estimate for the current observation (e.g., December 2012).

3. When using the Holt smoothing method, there is one additional estimated parameter. SMTH_SLOPE is the change in estimated crime for each step ahead. If, for example, you need the forecasts for February 2013 and your current time period is December 2012, you add two times SMTH_SLOPE to SMTH_LEVEL because February 2013 is two steps ahead of December 2012.

4. SQ_ERROR is the squared forecast error of the current observation from the forecast made for it from the previous period (e.g., November 2012 if the current period is December 2012).

5. TRIGG is the value of the Trigg Tracking Signal for the current observation.

6. SIGNALTRIP indicates whether the Trigg level was higher than the threshold. If it was, this field will have a **1** to indicate that the Trigg value was greater than or equal to the threshold selected and the detected change is an <u>increase</u>, a **-1** if the Trigg value is greater than or equal to the threshold but the detected change was a <u>decrease</u>, and a **0** otherwise.

7. FORECAST is the one-step-ahead forecast, for the next observation in time (e.g., January 2013 if the current period is December 2012). For a January 2013 forecast and simple exponential smoothing it is SMTH_LEVEL for December 2012 multiplied by the seasonal factor for January 2013. For January 2013 and Holt smoothing it is the sum of SMTH_LEVEL and SMTH_SLOPE times the seasonal factor for January 2013.

Full Output

First, the full output includes all the input fields plus the calculated values. If the user clicks the option button for Save full output button and then clicks the Save full output button, a save output window opens (Figure 24.2). Select dBase 'DBF' for the Save output to field, browse to the folder of your choice, and type a file name (Run99.dbf in Figure 24.2). Both the input data and the one-step ahead forecast are output to the screen and to a 'dbf' file. The file will be saved with a "TS_F" prefix before the defined file name, Run99.dbf, resulting in TS_FRun99.dbf.

Figure 24.3 is an example of the full output from the Pittsburgh monthly data given the selections made in Figure 24.1. This output is useful for not only seeing current conditions but also the history of a district. For example, census tract 20300 has had a lot of unusual activity according to the Trigg signal. It had an unexpected decline in August 1998, two exceptionally high values in November and December 1998 and another two in April and May 1999. So if

Figure 24.2:

Defining Full File Output

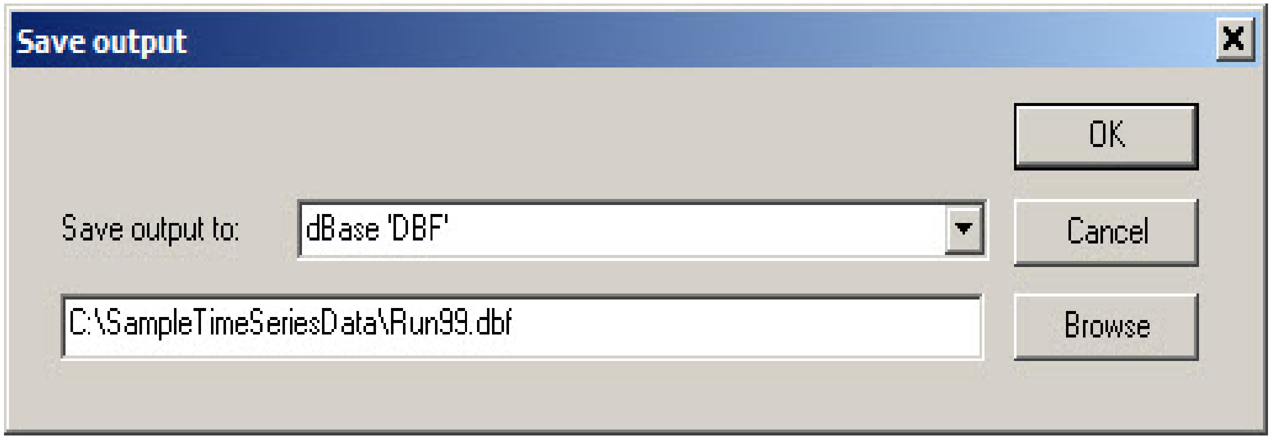


Figure 24.3:

Example Full Output from Pittsburgh Monthly Data

	A	B	C	D	E	F	G	H	I	J
1	AREA	YEAR	SEASON	EVENTCOUNT	DE_SEASON	SMTH_LEVEL	SQ_ERROR	TRIGG	SIGNALTRIP	FORECAST
341	20300	1998	4	2	2.11	1.85	0.09	0.37	0	1.86
342	20300	1998	5	1	1.00	1.75	0.74	0.76	0	1.62
343	20300	1998	6	4	4.34	2.06	6.68	1.76	1	2.23
344	20300	1998	7	2	1.85	2.04	0.05	0.00	0	2.13
345	20300	1998	8	0	0.00	1.79	4.14	1.60	-1	1.71
346	20300	1998	9	3	3.15	1.95	1.84	0.80	0	2.08
347	20300	1998	10	1	0.94	1.83	1.03	0.73	0	1.81
348	20300	1998	11	5	5.07	2.22	10.50	1.86	1	2.47
349	20300	1998	12	6	5.39	2.60	10.03	1.92	1	2.39
350	20300	1999	1	2	2.18	2.55	0.18	0.00	0	2.45
351	20300	1999	2	3	3.12	2.62	0.32	0.34	0	2.61
352	20300	1999	3	2	2.00	2.54	0.38	0.38	0	2.41
353	20300	1999	4	6	6.34	3.00	14.41	1.93	1	3.01
354	20300	1999	5	6	5.98	3.36	8.85	1.62	1	3.10
355	20300	1999	6	2	2.17	3.21	1.42	0.40	0	3.48
356	20300	1999	7	2	1.85	3.05	1.87	0.82	0	3.20

census tract 20300 had a positive signal trip in the last time period in the series, most likely it would have been a true positive needing police intervention, given its history of past flare-ups.

Next Time Period Output

Second, the next time period output includes only the calculated fields for both the screen and saved file. The word "next" refers to the forecast made for the next time period, while the Trigg tracking signal evaluates the current period. Again, in the dialog for saving the output file, type the .dbf extension in the chosen file name. The file is saved as a 'dbf' file with a "TS_C" (for 'current') prefix, resulting in TS_CRun2.dbf if Run2.dbf were entered by the user. The same field definitions as used in full output apply.

Figure 24.4 shows a sample output, which provides a scan of the entire jurisdiction for the current time period. The last time period in the corresponding input data set was December 1999, so this was taken as the current time period. Here you can see that the first 11 areas shown appear to have ordinary crime levels for December 1999 but the last four areas have unusual activity, three large increases and one large decrease. The forecasts of this output, just as for the full output, are "business-as-usual" forecasts for the next time period, January 2000.

> Note that the user must **choose** between full output and next time period output. Only one of these can be output for a single run.

Optimized Smoothing Parameters Output

The third type of output shows the results of the optimization process for exponential smoothing. This provides information on the parameters used to optimize the smoothing for each district. Define the file name and it will be saved as an ASCII text file with a 'txt' extension. Figure 24.5 illustrates the major output of the optimized smoothing parameters file, in this case for all selections in Figure 24.1 except the smoothing method which is now Holt.

1. Optimum Alpha is the smoothing parameter value for *level* of a time series that minimizes the one-step-ahead forecast sum of squared errors.

2. Optimum Gamma is the smoothing parameter value for *time trend slope* of a time series that minimizes the one-step-ahead forecast sum of squared errors.

3. SSE is the resulting optimal sum of squared errors for the time series.

Figure 24.4:

Example Output for Next Period Forecast

	A	B	C	D	E	F	G	H	I	J
1	AREA	YEAR	SEASON	EVENTCOUNT	DE_SEASON	SMTH_LEVEL	SQ_ERROR	TRIGG	SIGNALTRIP	FORECAST
17	51100	1999	12	1	0.90	1.57	0.58	0.81	0	1.44
18	60300	1999	12	2	1.80	2.05	0.08	0.05	0	1.88
19	60500	1999	12	1	0.90	0.54	0.16	0.77	0	0.50
20	70300	1999	12	0	0.00	0.73	0.69	1.37	0	0.67
21	70500	1999	12	0	0.00	0.80	0.83	1.35	0	0.73
22	70600	1999	12	0	0.00	0.21	0.06	0.64	0	0.19
23	70800	1999	12	0	0.00	0.65	0.54	1.21	0	0.59
24	70900	1999	12	1	0.90	1.13	0.07	0.24	0	1.03
25	80200	1999	12	0	0.00	0.98	1.23	0.80	0	0.90
26	80400	1999	12	3	2.69	1.52	1.79	1.39	0	1.39
27	80600	1999	12	1	0.90	0.51	0.20	0.71	0	0.47
28	80700	1999	12	2	1.80	0.71	1.53	1.72	1	0.65
29	80900	1999	12	0	0.00	2.38	7.32	1.64	-1	2.18
30	90100	1999	12	2	1.80	0.83	1.20	1.66	1	0.76
31	90200	1999	12	4	3.59	1.74	4.43	2.02	1	1.59

Figure 24.5:

Example Optimized Smoothing Parameters Output

Pittsburgh Monthly Data

Holt Exponential Smoothing Results

Data File:
Output File:
Deseasonalization Level: District
Optimum Sum of Squared Errors: 38140.61
Optimum Mean Square Error: 2.27
Trigg Alpha : 0.90
Trigg Beta: 0.15
Trigg Threshold: 1.50
Results by District:
District: 10300 Optimum Alpha: 0.03 Optimum Gamma: 0.01 SSE: 771.71
District: 20100 Optimum Alpha: 0.27 Optimum Gamma: 0.04 SSE: 2800.01
District: 20300 Optimum Alpha: 0.03 Optimum Gamma: 0.28 SSE: 230.15
District: 30500 Optimum Alpha: 0.07 Optimum Gamma: 0.14 SSE: 713.50
District: 40200 Optimum Alpha: 0.18 Optimum Gamma: 0.04 SSE: 485.49
District: 40300 Optimum Alpha: 0.11 Optimum Gamma: 0.01 SSE: 63.12
District: 40400 Optimum Alpha: 0.04 Optimum Gamma: 0.22 SSE: 328.64
District: 40500 Optimum Alpha: 0.26 Optimum Gamma: 0.04 SSE: 621.25
District: 40600 Optimum Alpha: 0.03 Optimum Gamma: 0.29 SSE: 126.85
District: 40900 Optimum Alpha: 0.01 Optimum Gamma: 0.03 SSE: 201.86
District: 50100 Optimum Alpha: 0.17 Optimum Gamma: 0.10 SSE: 986.29
District: 50600 Optimum Alpha: 0.16 Optimum Gamma: 0.13 SSE: 176.69
District: 50700 Optimum Alpha: 0.10 Optimum Gamma: 0.01 SSE: 288.04
District: 50900 Optimum Alpha: 0.01 Optimum Gamma: 0.39 SSE: 368.07

It is valuable to review the optimal parameters to see which areas have stable versus dynamic time series. Note that for the Trigg calculation, we want a large alpha to detect large changes in the number of recent events. That is why the default value of alpha is 0.9. However, for forecasting, we want a low alpha in order to smooth the data to produce a stable forecast.

For example in Figure 24.5 district 40900 has a very stable time series with low alpha and gamma while district 20100 has a dynamic level with large alpha and district 40600 has a dynamic time trend with high gamma. Therefore, the forecast for district 40900 is liable to be more accurate than for district 20100 but district 20100 is liable to have more accurate detections of large crime increases.

Guidelines for Running Forecast Models

Each model that is run for a fixed set of data will produce slightly different output. Based on our experience, we have found that there is not a single model type that will cover all data sets. Indeed, much of the forecasting literature for the last 35 years has been attempting to design and verify rules on which forecast model to use in different situations and for different data. Each jurisdiction will have its own unique characteristics in its time series data and districts within the same jurisdiction will often differ in their characteristics. Consequently, an analyst must experiment with this framework to identify which parameter choices produce the best overall fit for that jurisdiction and for the individual districts.

However, some guidelines can be provided based on our studies with this methodology.

1. Detection does not need very accurate forecasts, but just reasonable values because large changes in a time series are fairly easily detected. Detection does not attempt numerical accuracy but just a binary categorization, either a crime count is exceptional or not. That reduces the need for forecast accuracy of the counterfactuals. We recommend using Simple Exponential Smoothing with jurisdiction seasonality and optimization for detection.

2. Gorr, Olligschlaeger, and Thompson (2003) provide the guideline from empirical testing, that to get good forecast accuracy (20% mean absolute forecast error or less) crime time series should average at least 25 crimes or so per time period (month or week). It is hard to get that volume of crime unless using crime aggregates (such as all serious property crimes) and fairly large districts (tracts or police divisions). Very large cities such as New York, Los Angeles, Chicago, Houston, Philadelphia or Phoenix may have sufficient crime levels in some areas to provide good accuracy by week and for some crime types (e.g., vehicle theft, burglaries). It is the high-crime areas that are the most important and need the best forecasting accuracy as well as good detection accuracy for tactical

deployment of police. The low-crime areas do not need as good forecast accuracy for resource allocation as much as they need good detection for large crime increases, which is easily obtainable.

3. Seasonal factors are the least accurately estimated parameters in univariate time series models because a season, such as month (e.g., July, 7) or a week (e.g., 23) is only observed once a year. If an analyst has five years of data, then there are only five observations per seasonal adjustment. Gorr, Olligschlaeger, and Thomson (2003) found that jurisdiction-wide seasonality is about 10% more accurate for forecasting than district-level seasonality (estimated separately for each district using district data).

4. Different districts, however, have different seasonal patterns, some widely different from each other. For example, motor vehicle thefts peak in Pittsburgh in summer months in most areas but the major university area has a trough in summer because students (and their cars) are mostly gone then. For some applications, then, it may be valuable to give up some forecast accuracy to gain information (albeit noisy) on seasonal patterns by district. To obtain the most detailed forecasts tailored to each district, we recommend district seasonality with the district optimization of smoothing parameters available with the Holt method.

5. There is an objective way to choose forecast models in Table 24.1. The CrimeStat output includes a sum of squared errors term for each district from minimizing one-step-ahead forecasts in selecting smoothing parameters. Average these for a random sample of districts a across alternative models and select the method with the minimum sum of squared errors.

Overall, for detection/early warning in small-to medium-sized cities, the authors recommend using monthly data, simple exponential smoothing, jurisdiction-wide seasonality and smoothing parameter optimization. For the most informative and tailored one-step-ahead forecasts, we recommend using monthly data, Holt Smoothing, district seasonality and optimization of smoothing parameters. For large metropolises, however, it may be possible to run weekly forecasts as long as the average number of events per week is 25 or more in high crime areas.

Current research under way by one of the authors suggests that once detected, crime flare ups in areas tend to continue for some stretch of time with some periods on and an others off. Figure 24.3 showed an example of this behavior. This research suggests that police can adopt decision rules to maximize their effectiveness such as “Each time a crime flare up is detected, maintain extra police resources in the area for a fixed number additional time periods.” For example, the research has experimented with one through three period stretches of time for prevention efforts, each time a signal is tripped. Such rules can expose a relatively large number of crimes to prevention while maintaining reasonable workloads for police officers on patrol (Gorr & Lee, 2013).

Counterfactual Detection v. Forecasting

The time series forecasting module has two main purposes: 1) as a signal for early **detection of large changes** in time series patterns; and 2) as a **forecast** for the next time period into the future to aid resource allocation decisions.

Signal detection needs a forecast of the level of a time series under "business-as-usual" conditions for the most recent data point to answer the question "Does this data point seem unusual?" The counterfactual forecast applies exponential smoothing to all data before the most recent time period in the time series to forecast its data point. Exponential smoothing is ideal for making counterfactual forecasts because by definition it largely ignores recent large changes by smoothing them with a relatively low weight, even for the last data point used for estimation. Then if the difference between the forecast of what was expected versus the actual value is large, we have a signal trip.

Time series forecasting for the next or future time periods, on the other hand, is different in that when we make the forecast we do not already have the future actual values. As time passes after a forecast is made, eventually the future is realized and we get the actual values. Then we can calculate forecast errors "after the fact". The module also uses the same exponential smoothing routines for forecasting as it does for making counterfactual but for different purposes. Smoothing for detection tunes the signal for how reactive the signal is to current large changes as opposed to a series of smaller but accumulating changes while smoothing for forecasting estimates modules that "drift" with the data, albeit with a lag, to self-adapt to changing conditions.

However, we believe that major value of exponential smoothing for police is in providing counterfactual forecasts for detection while forecasts of the future necessarily are limited to extrapolations of known patterns under the existing conditions (i.e., with no surprises). The user should be very cautious with forecasting and not assume that the forecast will necessarily come to pass. The forecasts are useful for extrapolating what is to be expected *if* current conditions persist.

Example with Pittsburgh Monthly Crime Data

Using the sample data for Pittsburgh monthly part 1 crime counts, the monthly data was run with jurisdiction-wide seasonality and simple smoothing. Figure 24.6 shows a choropleth map of Pittsburgh of serious violent crimes for December 1999 along with tracts that have signal trips. Out of the 140 tracts, there are 19 (or 13.5%) that have signal trips for increases and 8 (or 5.7%) that have signal trips for decreases. Clearly, police are more interested in the areas where crime increased than in areas where crime decreased. Nineteen seems manageable for investigation by crime analysts and targeted patrol or other police interventions.

Figure 24.7 shows the map zoomed into tract 261400 which has a signal trip for an increase and a current crime count of 6. There are three serious violent crimes recorded at the same street intersection (at different dates) which display as one point in the map, two crimes adjacent to each other, plus a crime near the border of the tract. Research on crime hot spots (e.g., Weisburd, Bushway, Lum, & Yang, 2004; see Chapters 7, 8 and 9 of the *CrimeStat* manual) and current research on crime flare ups show that they tend to occur in very small areas ("micro areas" on the order of blocks) so when detected at the tract level, the crime analyst can zoom in and specify small subareas for targeted patrol.

Conclusion

CrimeStat's Time Series Forecasting Module provides a crime early warning system for police that is comprehensive, uses simple but proven methods, and is easy to use. A crime analyst can scan all the districts or areas of an entire jurisdiction in a single run of the module and see which ones are starting large crime increases (or decreases). Once detected, police have the option and ability to intervene with directed patrol and other means to prevent additional crimes in affected districts. The Time Series Forecasting Module also provides extrapolative forecasts of expected crimes for the next week or month by district, which can aid resource allocations by police.

A major limitation of any approach to working with crime time series data for tactical deployment of police resources (e.g., "Where should we target patrols this week?") is that the size of area units needs to be small, certainly patrol districts and smaller, but then the associated time series data has relatively low crime counts and any estimated models have sizable estimation and prediction errors. In such a situation it is better to use simple models, if for no other reason than there is very little else to get out of the limited data than what simple models can find. In our academic research we have used many complex models on this kind of data, from spatial multiple regression, to Bayesian versions of spatial regression models, to neural networks, and to spatial scan statistics without much more benefit than is available now in this chapter. Nevertheless, time series analysis of crime data does add value to crime analysis. We recommend using the automated detection methods of this chapter for early warning of crime increases in conjunction with crime mapping and other sources of information to diagnose and respond to emerging crime-area problems.

Figure 24.6:

Pittsburgh Violent Crimes in December 1999 with Signal Trips

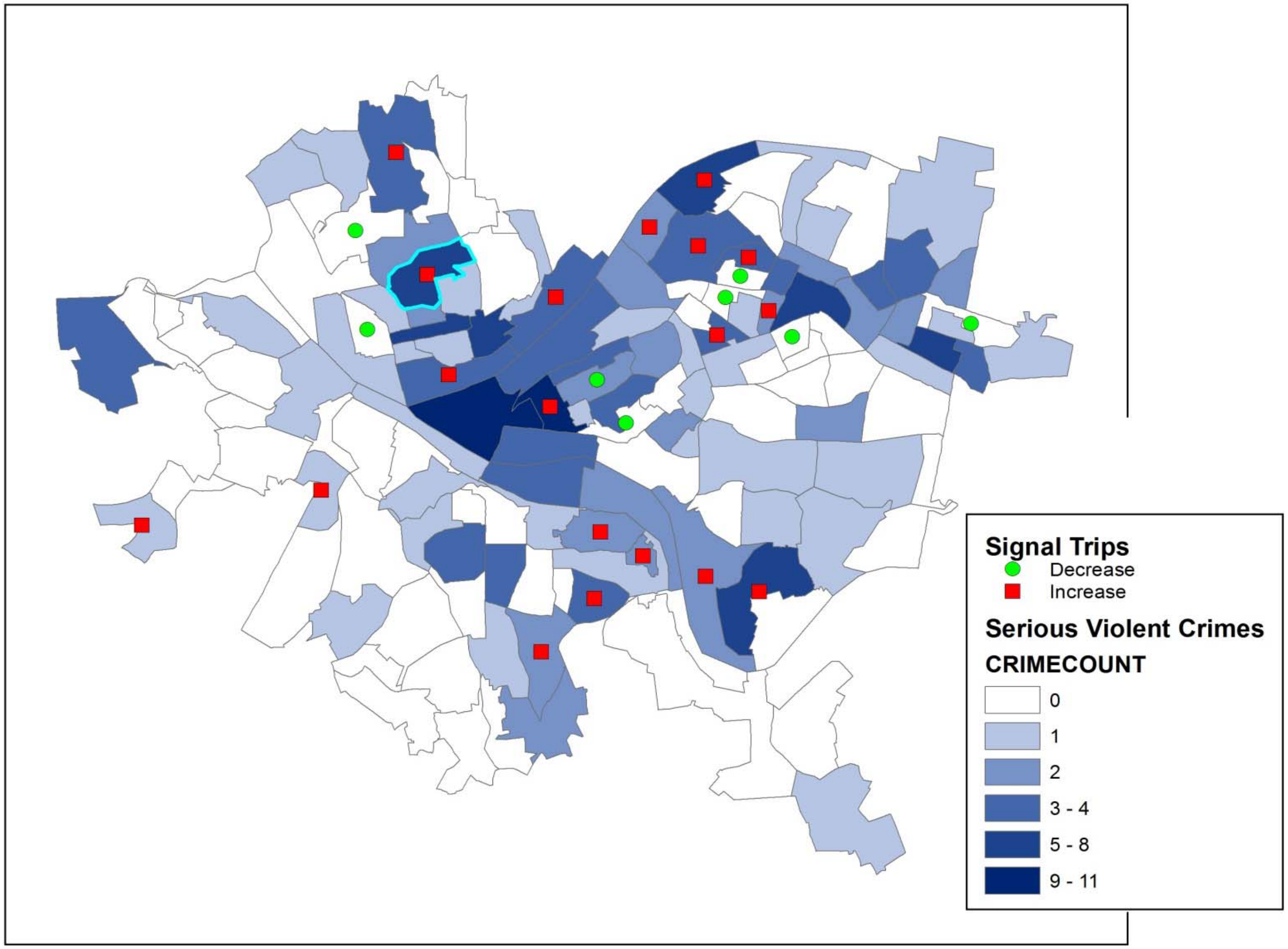
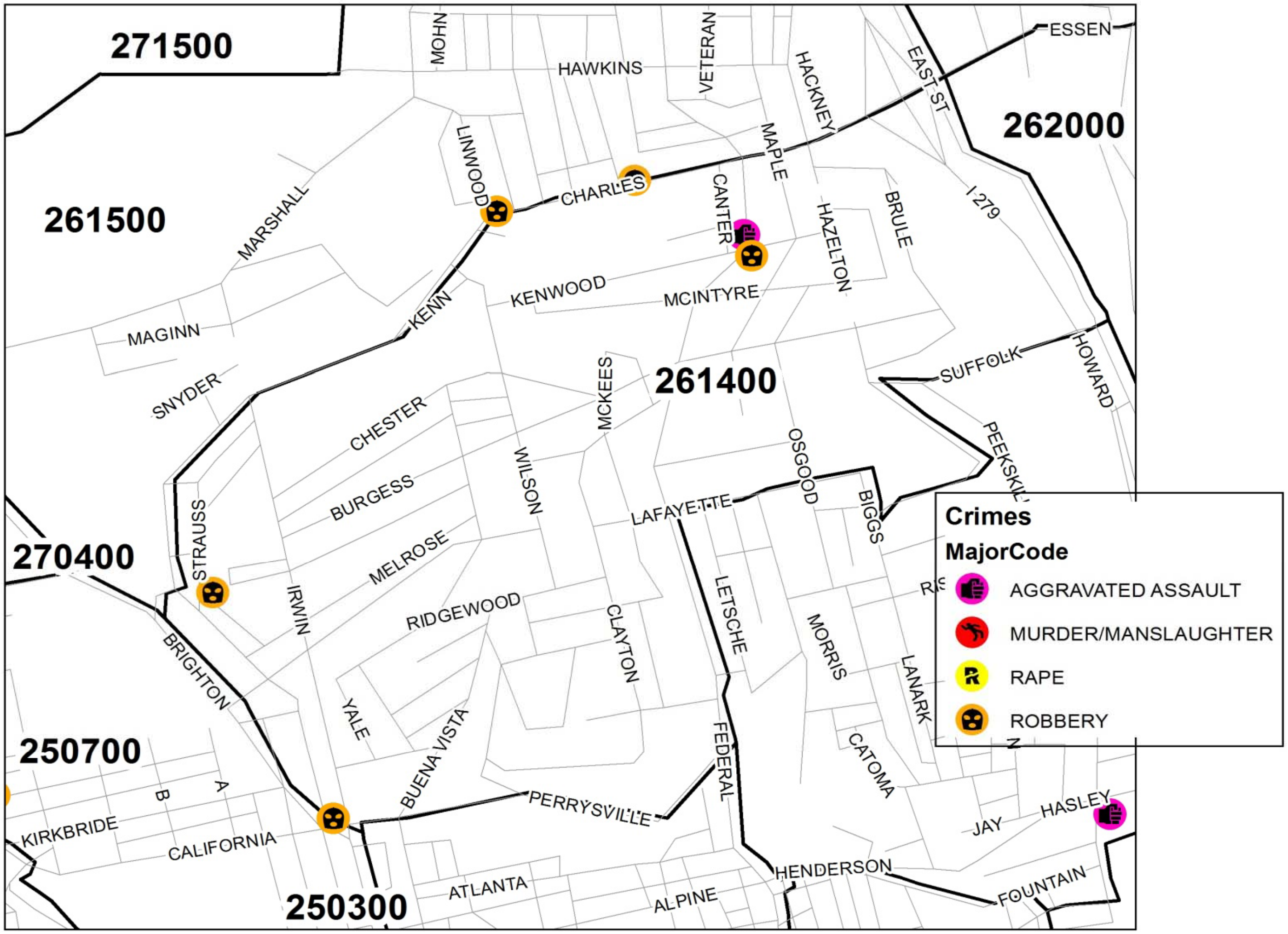


Figure 24.7:

Tract 261400 Showing Crimes Causing Signal Trip



References

Cohen, J., Garman, S. & Gorr, W. L. (2009). Empirical calibration of time series monitoring methods using receiver operating characteristic curves, *International Journal of Forecasting*, 2009, 25(3), 484–497.

Gorr, W. L. (2009). Forecast accuracy measures for exception reporting using receiver operating characteristic curves, *International Journal of Forecasting*, 2009, Vol. 25(1), 48–61.

Gorr, W. L. & Kurland, K. S. (2012). *GIS Tutorial for Crime Analysis*, Esri Press, Redlands. See chapter 9, Preparing incident data for mapping, especially Tutorials 10-1 and 10-2 of Chapter 10 for automating those steps).

Gorr, W.L. & Lee, Y. J. (2013). Early warning system for crime hot spots, Heinz College, Carnegie Mellon University Working Paper Series (http://www.heinz.cmu.edu/faculty-and-research/research/research-details/index.aspx?rid=482).

Gorr, W. L., Olligschlaeger, O. M. & Thompson, Y. (2003). Short-term forecasting of crime, *International Journal of Forecasting, Special Section on Crime Forecasting*, 19(4), 579–594.

Neill, D. B. (2009). Expectation-based scan statistics for monitoring spatial time series data. *International Journal of Forecasting* 25: 498–517.

Weisburd, D. L., Bushway, S., Lum, C., Yang, S. (2004). Trajectories of crime at places: a longitudinal study of street segments in the city of Seattle. *Criminology* 42:283−321