Spatio-Temporal Forecasting of Crime:

Application of Classical and Neural Network Methods

by

Andreas Olligschlaeger

Research Associate

H. John Heinz III School of Public Policy and Management

Carnegie Mellon University

Wilpen Gorr

Professor of Public Policy and Management Information Systems

H. John Heinz III School of Public Policy and Management

Carnegie Mellon University

June 7, 1997

1. Introduction

The problem of forecasting space-time phenomena is one that has occupied geographers for decades. Complicating the search for accurate and robust forecasting models is the fact that space-time phenomena more often than not exhibit behavior inconsistent with that assumed by modern statistical theory. For example, the spatio-temporal distributions of some types of criminal activity appearing chaotic or almost random in nature are often non-linear and discontinuous across space and time. Model heterogeneity is especially true for criminal activity such as street level drug dealing, which is difficult to model at the microlevel of census tracts or smaller units.

Geographers and Regional Scientists have long realized that local context and spatial heterogeneity are extremely important when forecasting space-time phenomena (see for example Anselin, 1988), and have consequently devised a number of ways of addressing heterogeneity (see for example Cliff et al, 1975). Some of these methodologies have been applied to modeling and forecasting spatial patterns of criminal behavior (see for example Dunn, 1980; Gorr and Olligschlaeger, 1994), albeit with varying degrees of success. Nevertheless, difficulties in obtaining high quality localized data and building successful models have yielded only few examples in the literature that employ spatio-temporal forecasting techniques to crime patterns.

Over the past five years, however, the use of geographic information systems (GIS) in police departments across the country has increased considerably (Maltz, 1993 and McEwen and Taxman, 1994). In addition, some police agencies have begun to integrate GIS with other sources of data such as 911 calls for service and police records management systems. This has resulted in the availability of high quality data at the address, or point level. Consequently data can be aggregated according to any desired spatial unit or grid system while still preserving all other attributes related to that point, thus producing very accurate spatio-temporal data sets.

This paper introduces a new spatio-temporal forecasting methodology that combines artificial neural networks and cellular automata with GIS-based data. The technique, which we refer to as chaotic cellular forecasting (CCF) is similar to spatial adaptive filtering due to Foster and Gorr (1986) and weighted spatial adaptive filtering due to Gorr and Olligschlaeger (1994) in that it uses contiguity relationships and the geographer's assumption that influence between data points decays with distance. As with spatial adaptive filtering the methodology uses an iterative process to arrive at a solution. Unlike spatial adaptive filtering, however, chaotic forecasting uses a gradient descent method rather than a grid search to find the optimal set of parameters (or, in the case of artificial neural networks, weights). In addition, and most importantly, CCF has the nonlinear and multi-model functional form commonly used in neural net modeling, allowing for increased pattern recognition and accommodation of spatio-temporal heterogeneity. The result is a robust spatio-temporal forecasting method that requires very little model specification, is self - adaptive and performs very well on data sets that exhibit non-traditional statistical behavior.

The second section in this paper briefly reviews the spatially varying parameter model and the literature on spatio-temporal forecasting. The third section outlines the chaotic cellular forecasting method and illustrates how it was derived from cellular automata and a particular type of artificial neural network: a multilayer feedforward network with backpropagation. The fourth section describes a comparative study of chaotic cellular forecasting with a number of traditional spatial forecasting techniques using GIS-based data derived from the Pittsburgh Drug Market Analysis Program (DMAP). Holdout samples in a rolling forecast design over 12 months are used to determine the out-of-sample forecast accuracy of each method. The fifth section reviews the results and last section provides a brief summary and conclusion.

2. Spatial Forecasting Methods and Models

According to Anselin (1988) there is much evidence of the model heterogeneity over spatial context. He argues that modeling strategies should take into account features unique to each location or spatial unit. In the absence of a completely specified model over space and time, researchers must resort to a model that accommodates spatio-temporal heterogeneity using varying parameters:

$$Y_i = \sum_{k=0}^{p} \boldsymbol{b}_{ki} x_{ki} + \boldsymbol{e}_i, \; i \in C \tag{1}$$

where C is an index for the space-time context of parameter variation, i is an index in C denoting place and time, Y_i is the independent variable at observation i, β_{ik} is the parameter for the k_{th} independent variable for observation i (k = 0,1,2...p), and ε_i is the error term for observation i. Implementing this model in unconstrained form is impossible because the number of parameters is p+1 times the number of observations. Consequently researchers have devised a number of strategies to counter this problem by constraining variation of the β_{ik} over C (e.g., Anselin, 1988; Gorr and Olligschlaeger, 1994).

Classical econometric fixed and random effects models (see Cliff et al, 1975) contextuate via dummy variables, assuming that effects of space and/or time are fixed. These types of models are most often used in situations where the number of observations is limited and spatio-temporal heterogeneity is viewed as a nuisance in estimating the non-varying parameters that hold across space and time. Those methods have the goal of estimating behavior which is constant over space and time while controlling or screening out varying behavior.

Two methods which explicitly model heterogeneity for use in exploration and forecasting are locally weighted regression (see for example Cassetti, 1982, and Cleveland and Devlin, 1988) and Kriging (David, 1977; Haining, 1990). Locally weighted regression techniques require that the weights determining spatial variation be specified a priori via trial and error. Kriging, or in the multivariate case, co-kriging, use an empirically estimated function, called the variogram, to determine the spatial weighting of data observations. Both methods assume that the influence of other observations declines with distance from the current observation.

Model (1) is also often used in exploratory data analysis (see for example Bretschneider and Gorr, 1983). Maps of residuals can show undetected heterogeneity suggesting additional theory or model structure, and maps of estimated spatially varying parameters are useful in determining the functional form of parameter variation. A similar type of exploratory data analysis is expansion modeling, e.g. step-wise regression models using polynomial or other functions of time and space coordinates which interact with an initial model's variables (Cassetti, 1986 and Cassetti and Jones, 1992).

A non-causal modeling or "naïve" approach to accommodating heterogeneity uses spatial and temporal lags of variables or model residuals. A univariate example is the Space-Time Autoregressive model, or STAR (Tobler, 1969) which is an extension of the purely temporal autoregressive model due to Box and Jenkins (1970) and assumes that the influence of neighboring observations declines with distance from the current observation according to a set of predefined spatial weights. Other examples include the Space-Time Autoregressive Integrated Moving Average (STARIMA) model which incorporates repeated differencing for trend elimination and the exponential smoothing model (Cliff et al, 1975).

Still another approach uses pattern recognition with the geographer's assumption of distance decay for influence, and shares a heredity that also led to neural network models. Two such methods are spatial adaptive filtering (SAF) and weighted spatial adaptive filtering (WSAF), both of which are based on adaptive filtering due to Widrow and Hoff (1960). Neural network model estimation also draws upon adaptive filtering. Foster and Gorr (1986) introduced SAF as an extension of multivariate damped negative feedback estimation for pattern recognition by using a grid search approach to optimizing individual damping factors for each β_k in model (1). WSAF was introduced by Gorr and Olligschlaeger (1994) as an extension of SAF. It incorporates an additional pattern recognizer component into SAF that reduces an inherent bias due to applying equal weights to feedback signals from neighboring observations and allows for self-modeling of spatial discontinuities. Based on the magnitude of forecast errors using the β_k of neighboring observations WSAF automatically assigns appropriate weights to feedback signals: those observations with small forecast errors receive relatively large weights, whereas those with larger errors receive smaller weights. The resulting weighting scheme is similar to those used in time-series combination forecasting (see for example Bates and Granger, 1969, and Clemen, 1989).

The idea of using artificial neural networks (ANNs) for forecasting is relatively new. While ANNs have been the subject of intense research efforts over the past decade, yielding many technological advances in areas such as robotics, speech recognition and the detection of explosives, it was not until White (1988) used ANNs to predict IBM daily stock returns that they were first used to forecast temporal patterns of economic behavior. Nevertheless there have been no examples in the domain of spatio-temporal forecasting (see Sharda and Patil, 1990, and Tang et al, 1990, for one of the few exceptions). To the best of our knowledge no ANN-based spatio-temporal forecasting models have been developed to date.

There are many different types of artificial neural networks. Since even a cursory introduction is far beyond the scope of this paper the interested reader is referred to Rumelhart and McClelland (1988) or Kroese and Van der Smagt (1993) for an introduction. The most widely used and studied type of ANN is the multi-layer feedforward network with backpropagation. This type of ANN is a nonlinear extension of Minsky and Papert’s (1969) Perceptrons and the same type of network used by White (1988) in temporal forecasting and in CCF introduced in this paper. ANN’s have features which make them attractive as candidates for spatio-temporal forecasting. Most important for modeling spatial heterogeneity is that the logistic activation function and hidden-layer model components of common ANN’s provide a multi-model functional form. Such ANN’s can automatically identify and estimate multi-models, plus switching rules for their deployment (Gorr, 1994). Thus, if there are patterns in spatio-temporal data that discriminate areas and times as having different multivariate models, the ANN model developed in this paper can automatically identify, estimate, and deploy multi-models. Furthermore, Hornik et al (1989) proved that certain ANN’s with logistic activation functions and hidden layer architectures form a class of universal approximators, i.e., they can estimate any functional form to any desired degree of accuracy provided sufficiently many hidden units are available. In addition, Lapedes and Farber (1987) have provided evidence that for chaotic time series artificial neural networks can forecast more accurately than conventional methods.

3. Chaotic Cellular Forecasting

Chaos theory provides a framework for modeling complex spatio-temporal phenomena (for a detailed introduction to chaos theory see Schroeder, 1991). A classical example of a chaotic system is the Mandelbrot set which is frequently used in artwork and as an illustration of how to graphically display complex mathematical functions. Some real world examples include weather patterns, stock market activity and brain activity. One of the fundamental tenets of chaos theory is that although chaotic systems seem to display totally random and unpredictable behavior they actually follow strict mathematical rules which can be derived and studied (Pickover, 1990). For the purpose of forecasting spatio-temporal phenomena, such rules are based on spatially and temporally lagged variables that form the basis of the forecast model.

Cellular automata are a specific type of chaotic system. They differ from other chaotic systems in that they act on discrete space, or grids rather than a continuous medium such as a surface. In cellular automata machines, each time frame is represented by a layer of cells (representing the "population" of cells). The next time frame is derived by replacing the previous layer according to a set of rules (Toffoli and Margolus, 1987). A key determinant of cellular automata rules is how each cell is influenced by neighboring cells.

Figure 1 shows three epochs (time frames) in the life of a population of cells occupying a nine by nine grid. An empty square represents a dead cell, whereas a dotted square represents a live cell. Upon initial examination of the three epochs it appears that cells are born and die randomly. However, this is not the case. The behavior of the cell population from one epoch to the other actually follows a very simple set of rules introduced by mathematician John Conway in his game of life (Gardner, 1970). Assuming that the neighborhood of a cell consists of all immediately adjacent or touching cells (Queen's case) the rules are as follows:

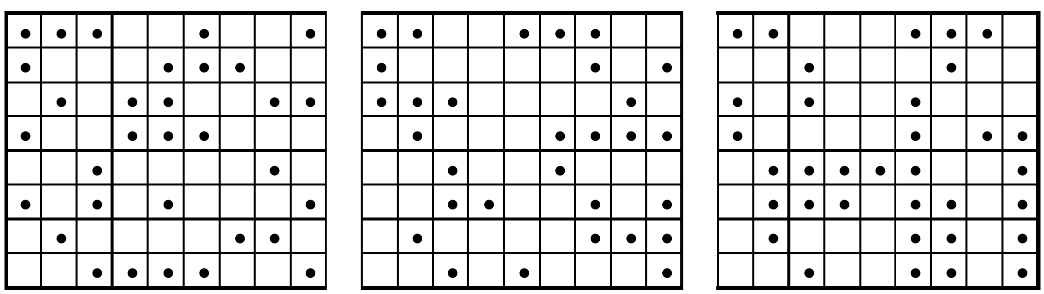
1. A live cell remains alive if and only if it has exactly two or three living neighbors. Otherwise it dies of either "overcrowding" or "loneliness".

2. A dead cell will come to life if and only if it has exactly three living neighbors. Otherwise it remains dead.

Note that the rules for the game of life do not vary spatially, i.e., the same rules apply to each cell in the population regardless of its location. A complexity of real-world chaotic systems is that they likely have spatially-varying rules. Nevertheless, spatially constant rules result in various spatial patterns. A further important observation in this regard is that the Game of Life rules produce one-step-ahead transitions, or forecasts without error, of cell populations based on lagged variables.

Spatial phenomena often exhibit the same kind of apparently random and chaotic behavior as the cells in the game of life described above. For example, if we were to divide space into cells and

Figure 1: A Game of Life



Epoch 0 Epoch 1 Epoch 2

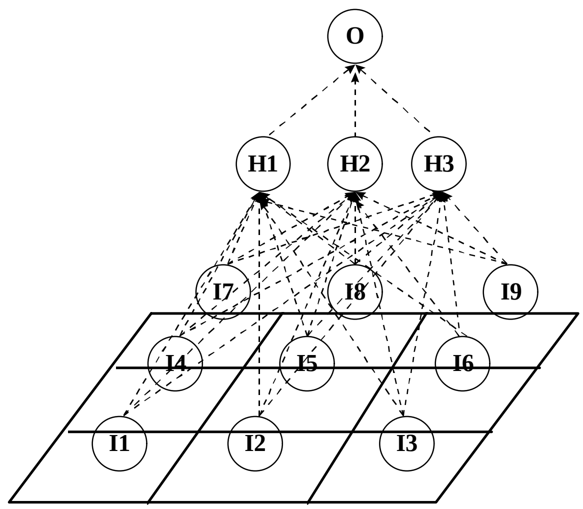
count the number of infected individuals per cell per time frame, then it is quite possible that over time cells would display similar seemingly unpredictable behavior as in the Game of Life, even though the rules of infection are quite simple. In other words it is conceivable that some spatio-temporal phenomena are actually the result of chaotic systems. Even minute changes in the parameters of cellular automata machines can result in a changeover from long term stability to short term chaotic behavior within the cell population (Schroeder, 1991).

This could help explain, for example, why spatial patterns of criminal activity are difficult to model. For example, it is possible to draw some analogies between the rules of the game of life and the behavior of street level drug markets. A drug market in a cell could “die” if there is overcrowding in its neighborhood, i.e., there are too many drug dealers, which either invites police intervention or leads to turf wars. On the other hand, one active drug market cell in an area where there are otherwise none would result in its being too inconspicuous. Finally, for an area to become a place of drug dealing certain conditions must be met. The analogy between rule number two could be that there must be displacement or spillover from neighboring drug markets.

Figure 2 shows how CCF was used to estimate the rules of the game of life. The network consists of three layers: the input layer, three hidden units and the output layer. The number of input units in CCF is determined by the neighborhood of a cell as well as the number of signals (or inputs) each cell sends to the hidden layer (this is analogous to the number of independent variables in regression). Since the game of life only considers whether a cell is dead or alive, each cell has only one input unit. The value of the signal is 1 if the cell is alive, and 0 otherwise. The neighborhood in this case consists of nine cells (the Queen’s case), consisting of the current observation and the

The rules of the game of life are very simple. Rules governing real world chaotic systems are far more complex, however. The problem is that they are only rarely, if ever, known a priori. If for the moment we assume that spatio-temporal patterns are indeed the manifestations of chaotic systems, then many conventional space-time models can be viewed as suggestions for rules. These suggestions are then tested against consecutive populations of cells, census tracts or any other areal unit. Often suggested models, or rules work quite well, sometimes they do not. Chaotic cellular forecasting (CCF) is essentially a cellular automata machine that works in reverse: instead of applying the rules or testing suggested rules it attempts to learn them. The mechanism used for learning the rules are feedforward neural networks with backpropagation. The rest of this section describes chaotic cellular forecasting in more detail using the game of life as an example.

Figure 2: Game of Life Neural Network



eight cells that it touches upon. We therefore have a total of nine input units. The intermediate, or hidden layer consists of three units. Each of the three hidden units is connected to all input units. Finally, the output unit receives its signal from the hidden layer via three connections. Each processing unit (or neuron) in a backpropagation network receives input from the units to which it is connected in the layer below.

As one of the first tests of the algorithm CCF was used to estimate the rules of the game of life using the network shown in Figure 2. The neighborhood used to produce one-step-ahead forecasts in CCF consists of the current observation (cell) and the eight surrounding cells. As in Figure 2 each hidden unit is connected to each grid cell in the neighborhood. The difference is there are multiple connections to each grid cell because multiple signals (independent variables) are processed. This produces spatially and, because the network produces one-step-ahead forecasts, also temporally lagged data points.

The network was able to learn the rules of the Game of Life perfectly, i.e., without error. Moreover, the network performed flawlessly even when shown patterns that were not used during training: different random initializations of the grid cells had no effect on the performance of the network. In addition, it was able to predict successive generations of cells ad infinitum, requiring only the first epoch of randomly initialized cells to do so. This indicates that the network is robust and able to generalize well, at least for this particular problem.

4. CCF Model

Each processing unit (or neuron) in a backpropagation network receives input from the units to which it is connected in the layer below. Figure 3 shows how these input signals are processed. The total (net) input of a unit is the sum of all signals multiplied by their connection strengths (weights) plus a bias. More formally:

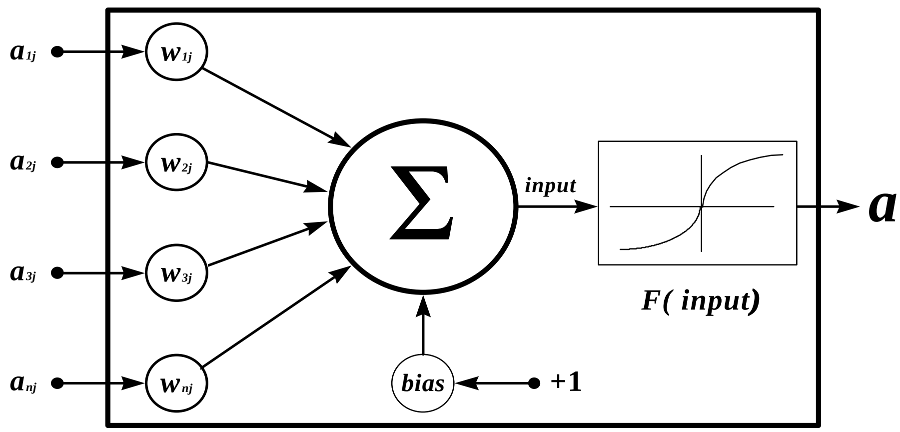
$$net_{pi} = \sum_j w_{ij} a_{pj} + \theta_{pi} \tag{2}$$

where net_i is the net input for processing unit i for input pattern p, w_{ij} is the weight of the connection between processing unit i and processing unit j in the previous layer, a_{pj} is the activation of unit j in the previous layer and θ_{pi} is the bias associated with unit i. The activation function used to process the net input varies, but generally takes on a sigmoid functional form. The most commonly used function is the logistic function. It yields values in the range [0,1] and determines the activation of the processing unit at the next step as follows:

$$a_{ip} = \frac{1}{1+e^{-net_{ip}}} \tag{3}$$

where a_{ip} is the activation of unit i for input pattern p. For input units the activation is usually equal to the net input since only one signal needs to be processed.

Figure 3: A Neural Processing Unit



The goal in feedforward networks is to map the input units to a desired output similar to the way in which the dependent variable is a function of the independent variables in regression analysis. The difference is that regression uses direct linear models whereas multi-layer feedforward networks use indirect nonlinear models. The hidden layer creates an internal representation of the patterns to be modeled. This internal representation is then mapped to the output unit. It is the hidden layer, along with the use of a nonlinear activation function that allows multi-layer networks to map far more complex functions than simple direct input to output unit mappings.

Feedforward networks with backpropagation “learn” to map the input units to the output units by adjusting the weights on the connections in response to error signals transmitted back through the network. This is the ANN equivalent of model estimation. During training the network is presented with each input pattern and computes the activation of the output unit(s) using the current network weight structure (the weights are initialized randomly prior to training). The difference between the output of the network and the target mapping constitutes the error signal. This signal is then propagated back through the network via the processing units and their connections and the weights are updated. The goal is to continually update the weights until the sum of all error signals is minimized.

In backpropagation networks the weights are estimated using the generalized delta rule derived by Rumelhart et al (1988) from the Perceptron convergence procedure due to Minsky and Papert (1969), which in turn is a variation of the delta rule proposed by Widrow and Hoff (1960). The generalized delta rule can be summarized in three equations (for a formal derivation of the generalized delta rule see Kroese and Van der Smagt, 1993 or Rumelhart et al., 1988). The first specifies that the weight change should be proportional to the product of the error signal sent to a receiving unit along a connection and the activation of the sending unit. More formally,

$$\Delta_p w_{ij} = \eta \delta_{pj} a_{pi} \tag{4}$$

where $\Delta_p w_{ij}$ is weight change for training pattern p and the connection between processing units i and j, δ_{pj} is the error signal sent to unit j, a_{pi} is the activation of unit i for input pattern p and η is the "learning rate". The learning rate is usually a small number less than 1 and close to zero. The definition of the error signal differs between output units and hidden units. For output units using a logistic activation function it is defined as

$$\delta_{pj} = (t_{pj} - a_{pj}) a_{pj} (1 - a_{pj}) \tag{5}$$

where t_{pj} is the target activation for the output unit. For hidden units the error signal is given by

$$\delta_{pj} = a_{pj} (1 - a_{pj}) \sum_i \delta_{pi} w_{ij} \tag{6}$$

The generalized delta rule implements a gradient descent in the error term. Training of the network proceeds by repeatedly presenting all input patterns and adjusting the weights until the sum of all errors is minimized, i.e. the network converges to a solution. In this respect it is crucial to select a learning rate that during training will allow the network to iterate towards a true global minimum rather than getting stuck in local minima. Learning rates that are too large can lead to oscillations between local minima, whereas small learning rates can require hundreds of thousands of iterations to converge. While it is theoretically possible that even with small learning rates the network will converge to a local minimum, empirical evidence suggests that this is rarely the case (Weiss and Kulikowski, 1991). One way to detect and avoid convergence to local optima is to train the network several times with different random initializations of the weights and to compare the results (Rumelhart et al, 1988).

CCF uses batch processing to adjust connection weights. Rather than adjusting the weights after each input pattern (or observation and time period), weight changes are summed over all input patterns (observations and time periods). After all observations have been processed, the sum of changes is divided by the number of observations multiplied by the number of time periods (which is equal to the number of input to output mappings) to arrive at a “smoothed” weight change for each connection. Experimentation has shown that updating weights in this manner yields better results and leads to faster convergence (Rumelhart et al, 1988).

The data for CCF is obtained by superimposing a regular grid on the study area and aggregating data values for the independent variables for each time period and grid cell. In the game of life the only data point per grid cell and time period consists of whether the cell is dead or alive. In CCF an example of a data point might be the number of burglary arrests per time period in a grid cell. A further example would be the proportion of housing units in the cell that are abandoned or tax delinquent during a time period.

The neighborhood used to produce one-step-ahead forecasts in CCF consists of the current observation (cell) and the eight surrounding cells. As in Figure 2 each hidden unit is connected to each grid cell in the neighborhood. The difference is there are multiple connections to each grid cell because multiple signals (independent variables) are processed. This produces spatially and, because the network produces one-step-ahead forecasts, also temporally lagged data points.

The algorithm for training the CCF network using spatially varying weights for all connections and a single hidden layer follows. Each observation has its own set of input to hidden unit connections and associated weights for each cell in the neighborhood. However, the hidden to output unit weights are spatially constant. This produces a hybrid model that has the advantage of some spatial variation in the weight structure but not too much to cause overfitting. However, using a varying weight structure, even in a hybrid model as described above, requires a very large number of observations.

The algorithm for this hybrid CCF model, which assumes temporally constant but spatially varying input to hidden unit weights and a single layer of hidden units is therefore:

For each grid cell:
  For each input to hidden unit connection:
    - randomly initialize the weight
For each hidden to output unit connection:
    - randomly initialize the weight

Begin Iteration

For each grid cell:
  For each input to hidden unit connection:
    - set the sum of weight changes to zero
For each hidden to output unit connection:
    - set the sum of weight changes to zero

1. Produce One-Step-Ahead Forecasts for all Spatial Units:
  For each time period in the data set:
    For each grid cell:
      1. Forward Pass:
        - for each hidden unit:
- calculate the net input using signals (data) from the current grid cell as well as those in the neighborhood (spatially lagged variables) and input to hidden layer weights associated with the current grid cell
- calculate the hidden unit activation

        - calculate the output of the network; this is the estimated forecast
      2. Backward Pass:
        - calculate the forecast error (actual - forecast)

- for each connection (input to hidden and hidden to output) calculate the weight change and add it to the sum of weight changes for that connection and grid cell over all forecasts
- calculate the squared forecast error and add it to the sum of squared forecast errors for the current grid cell and connection

2. Batch Update of Network Weights.

For each grid cell:

For each input to hidden unit connection:

- divide the total weight change (sum of all weight adjustments) by the number of time periods; this is the average weight change
- adjust the weight by the average weight change

For each hidden to output unit connection:

- divide the total weight change (sum of all weight adjustments) by the number of observations multiplied by the number of time periods; this is the average weight change
- adjust the weight by the average weight change

If the total sum of squared forecast errors in the current iteration is greater than or equal to that of the previous iteration, stop. Otherwise, continue iterating.

End Iteration.

Further experimentation revealed that adding direct input to output connections to the network architecture kept overfitting in check, i.e., the network would generalize better. In effect adding direct input to output connections results in a hybrid Backpropagation/Perceptron architecture, with both linear direct and nonlinear indirect input to output mappings. In addition, it was found that averaging spatially-lagged variables from the neighborhood of an observation rather than connecting each independent variable in each neighboring grid to each hidden unit (as is the case in Figure 2) reduced the number of connections, and, by extension, also the amount of overfitting. In other words, instead of nine sets of inputs, one from each observation in the neighborhood, the architecture only had two sets: one from the current observation, and one for the averaged spatially lagged independent variables of the eight neighboring grid cells. It was also found that this CCF architecture converged must faster than any other previously tested, on the average four times faster.

The following section tests the capabilities of CCF with a model that forecasts drug calls for service using data derived from a geographic information system developed for the City of Pittsburgh police department.

5. Forecasting Drug Calls for Service: A Comparison Between Classical Forecasting Methods and Chaotic Cellular Forecasting

Like many other cities of its size Pittsburgh experienced a marked increase in street level drug dealing during the late 1980's and early 1990's as a result of the crack cocaine epidemic. Prior to the appearance of crack cocaine street level drug dealing in Pittsburgh was largely confined to two areas which specialized primarily in heroin and marijuana. Other sporadic areas of open air drug dealing did exist, but were mainly limited to the sale of prescription drugs such as painkillers and the "Yuppie" drug powder cocaine.

In the summer of 1991 Pittsburgh also experienced a surge in gang-related violence. While initially most gangs in Pittsburgh were merely loosely organized groups of adolescents, experienced gang members from larger cities quickly attempted to gain a foothold in what they perceived as "virgin territory" for crack cocaine sales. Street-level drug markets in other major cities were already saturated by dealers and there was little opportunity for entry into a market tightly controlled by gangs. Pittsburgh, on the other hand, was still a "free-for-all": demand was greater than supply. Thus at least a part of the increase in violence can be attributed to street gangs setting up and defending "turfs" within which they plied their illicit drug trade.

In reacting to the increase in street level drug dealing the Pittsburgh Bureau of Police was using disruptive enforcement strategies which were proven highly successful in other cities: reverse stings, undercover buys, on-sight arrests of drug dealers after having observed illicit transactions and placing community oriented police officers in plain view of established drug hot spots. These strategies were used because street-level drug dealing is widely regarded as a weak link in the chain: once a street market has been disrupted it is very difficult for dealers to relocate (Cohen et al., 1993). They are unable to advertise their new location and are severely restricted in establishing new ones because they might infringe upon turfs already claimed by other drug dealers. However, there were a few instances where new hot spots did eventually surface.

GIS perform quite well when tracking the geographic displacement of drug dealers due to their ability to plot the locations and frequencies of the number of drug calls for service and drug arrests. But they do not perform as well at identifying emerging drug markets. The reason for this is twofold: first, police officers rarely make arrests in areas of which they are unaware that street level drug dealing is going on unless they happen to stumble upon a transaction. Street sweeps tend to concentrate on known drug markets. Second, residents of areas in which street level drug dealing is a relatively new phenomenon frequently are unaware of what is going on. They initially do not perceive the activity as drug dealing (Olligschlaeger, 1997). Rather, they tend to notice an increase in the level of crimes associated with street level drug dealing such as robberies, burglaries and assaults. In addition, violence associated with drug trafficking increases. Thus there is a lag between the time a drug market has established itself and when residents begin to make drug related calls for service.

A forecasting model for the activities of street-level drug markets must therefore also be able to predict drugs calls for service based on factors other than simply the level of drug calls for service in previous time periods. Based on the results of previous work and the availability of data from the Pittsburgh Drug Market Analysis Program (DMAP) it was decided to use three types of calls for service as indicators of emerging drug activity: weapon related calls (shots fired, person shot, man with a gun, etc.), robbery calls and assaults. In addition, Cohen et al (1993) showed that ecological factors such as the proportion of commercial properties in an area are important contributors to the level of drug calls for service. Commercial areas (especially older, run-down ones) lend themselves more to open air drug dealing because of factors such as the relative lack of population outside of regular business hours (there are fewer residents to observe drug dealing).

On the other hand, residential areas tend to reduce the amount of street level drug dealing due to increased guardianship of the neighborhood at night and after working hours. Thus the proportion of residential and commercial properties were included as indicator variables. Finally, open air drug dealing is a seasonal phenomenon. In the winter months drug dealers tend to stay inside not only because it is cold, but also because fewer people in general are on the streets and they become more visible. A seasonality index was therefore also included.

One additional independent variable that would be desirable to include as a predictor for the number of drug related calls for service is a measure of the impact of policing. In the absence of daily manpower allocation data the only available candidate for such a measure would be the number of drug arrests in an area. Arrest data, however, especially for drug related offenses, are misleading as an indicator of police activity. The reason for this is that many, if not most, drug arrests occur in an area other than the one in which the offense occurred. For example, arrests are often made in residential areas at the homes of suspected drug dealers where there is no other apparent street level drug activity. With the exception of public housing projects, drug dealers rarely ply their trade in the same area where they live. Police officers do not always arrest suspected drug dealers on sight, i.e., as soon as they observe what they believe to be an illegal drug transaction. This is especially true for undercover drug buys. It may take three or more buys until an arrest is made. Even if only one undercover buy is made, there can be a considerable time delay between the drug transaction and an arrest simply because of the time it takes for the crime lab to process the purchased substance and confirm whether it is indeed an illicit drug.

Certainly not all drug arrests are made at the homes of offenders. The point, however, is that many times simply arresting a person on sight can be counterproductive because potentially useful information about the suppliers and coconspirators of the offender can be lost. For that reason police officers often conduct an investigation, the result of which are arrests which may or may not be in the same area as where the offense occurred. Since the data do not provide any insight as to where the original offense occurred in relation to the location of the arrest, it was decided not to include drug arrests as an explanatory variable.

Three forecasting models were estimated using four different methodologies. The first model is a simple regression model using only the current observations' independent variables, i.e., no spatially lagged variables, and produces one-step-ahead (i.e., one month) forecasts for the number of drug related calls for service:

$$Y_i^F(t+1) = \sum_{k=1}^{p} \hat{\boldsymbol{b}}_k x_{ik}(t) + \boldsymbol{e} \qquad (7)$$

where i is the current observation, Y_i^F is the forecast for observation i, p is the number of independent variables, the $\boldsymbol{b}_k$ are the estimated regression parameters, $x_{ik}(t)$ are the independent variables for observation i at time t, and $\boldsymbol{e}$ is the error term.

The second model adds spatially lagged variables to model (7) above as follows:

$$Y_i^F(t+1) = \sum_{k=1}^{p} \hat{\boldsymbol{b}}_k x_{ik}(t) + \sum_{l=1}^{p-1} \hat{\boldsymbol{b}}_l \frac{1}{n} \sum_{j=1}^{n} x_{jl}(t) + \boldsymbol{e} \qquad (8)$$

where p is the number of independent variables (in this case it is seven), j represents each observation in the neighborhood, and n is the number of observations in the neighborhood (which is equal to 8 using the Queen's case except in the case of boundary grid cells). The x_{ik} represent temporally lagged variables whereas the x_j represent spatially and temporally lagged variables. Note that the second part of the equation represents the spatial average of observations in the neighborhood for all variables except the seasonality index. Model (8) assumes that the parameters are spatially constant. For the spatially varying parameter case an additional model would be:

$$Y_i^F(t+1) = \sum_{k=1}^{p} \hat{\boldsymbol{b}}_{ik} x_{ik}(t) + \sum_{l=1}^{p-1} \hat{\boldsymbol{b}}_{ml} \frac{1}{n} \sum_{j=1}^{n} x_{jl}(t) + \boldsymbol{e}; i, m \in C \qquad (9)$$

where i and m are indexes in C, the context of spatial parameter variation. The only difference between (8) and (9) is therefore that in model (9) each observation has its own set of parameters. However, this means that model (9) cannot be estimated due to the number of parameters. Note that models (8) and (9) also assume that the dependent variable is a linear function of the independent variables. Feedforward networks with backpropagation do not require this assumption since functional dependencies do not need to be specified a priori.

The chaotic cellular forecasting model can be rewritten in algebraic terms. For the network architecture outlined in the previous section we can write:

$$Y_i(t+1) = \sum_{n=1}^{p} w_{no} I_{nit} + \sum_{j=1}^{h} w_{jo} \frac{1}{1+e^{net_{jt}}} \qquad (10)$$

where Y_i is the target output of the network for observation i, t is the time period, h is the number of hidden units, p is the number of input units (including the averaged spatially lagged variables), w_{no} is the weight along the direct input to output unit connection between input unit, n and the output unit, o, I_{nit} is the input of unit n for observation i at time t, w_{jo} is the weight along the connection between the output unit, o and the hidden unit, j, and net_{jt} is the net input for hidden unit j at time t which is calculated as follows:

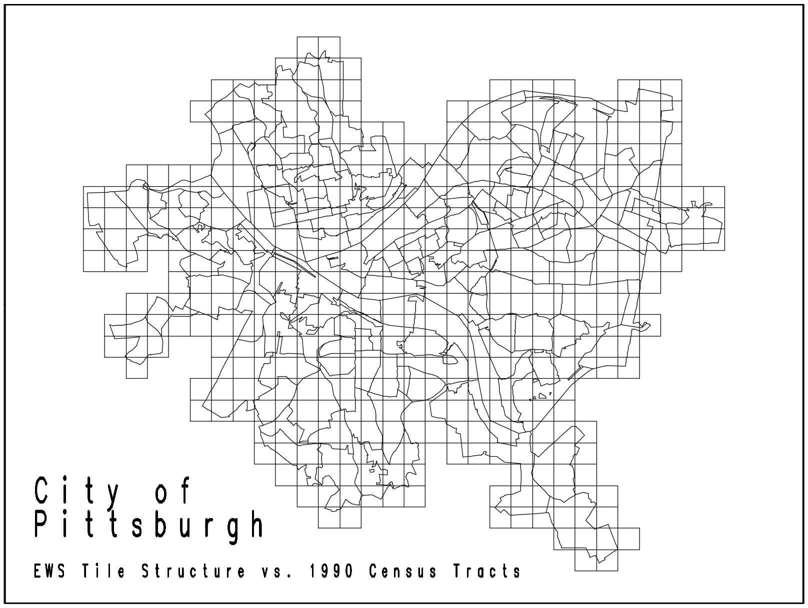
$$net_{jt} = \sum_{l=1}^{p} w_{jlm} I_{jlt} + \boldsymbol{q}_j; m \in C \qquad (11)$$

where p is the number of input units per cell in the neighborhood (this is analogous to the number of independent variables) plus the number of averaged spatially lagged variables, I_{jlt} is the input from unit l at time t for hidden unit j, w_{jlm} is the weight along the connection between hidden unit j and the lth input unit, and θ_j is the bias for hidden unit j, and m is an index in C, the context of spatial weight variation. In equations (10) and (11) the weights are the parameters to be estimated.

Note that the output of the network is simply the net input to the output unit. The number of hidden units in this instance of CCF is nine. This number was arrived at using informal experimentation on the data set and does not imply that there should always be nine hidden units in CCF models.

The data for the early warning system were obtained by superimposing a grid on the area of the city of Pittsburgh (see Figure 4) and aggregating data for each grid cell and time period (in this case one month). The cells are 2150 feet square (which represents approximately two city blocks), resulting in a total number of 445 cells. In selecting the size of the cells it was important not to make them too small because otherwise only few cells would have more than one or two calls for service. Too large cells would have resulted in too few data points for neural net modeling. Another criterion used in selecting the size of the grid cells was that pin maps of historical drug activity show that the typical drug hot spot area in Pittsburgh is about two city blocks square.

Figure 3



Call for service data were obtained by counting the number of calls per month within each cell using the (x,y) coordinates of the geocoded locations. The data spanned the years 1990 to 1992 resulting in 35 months worth of data (December of 1992 could not be used since there was no value for the number of drug related calls for service in January 1993). Future testing and additional modeling will utilize later data. With 445 cells the total number of data points was therefore 15,575. The relative proportions of commercial and residential properties were arrived at by relating property ownership information to parcel polygons via the lot and block number. The (x,y) coordinates of the geographic center of a parcel were then used to determine which grid cell a particular property falls into. The zoning classification for each property provided the basis for the relative frequencies. Finally, the seasonal index was arrived at by assigning values between 0.1 and 0.9 in equal increments to each month, where a value of 0.9 was assigned to June and July and 0.1 to December and January. Since backpropagation networks require a signal from the input units in order for weight adjustment to occur, all variables with a value of zero were adjusted to 0.1. This ensured that connection weights were not adjusted only in the case of non-zero inputs.

Table 1 shows the total number of calls for service, by year, for each of the four nature codes used in this study. Notice the remarkable increase in the number of weapon related calls for service of almost 100% over three years. The number of drug and robbery related calls for service also increased. Only assaults showed a decline.

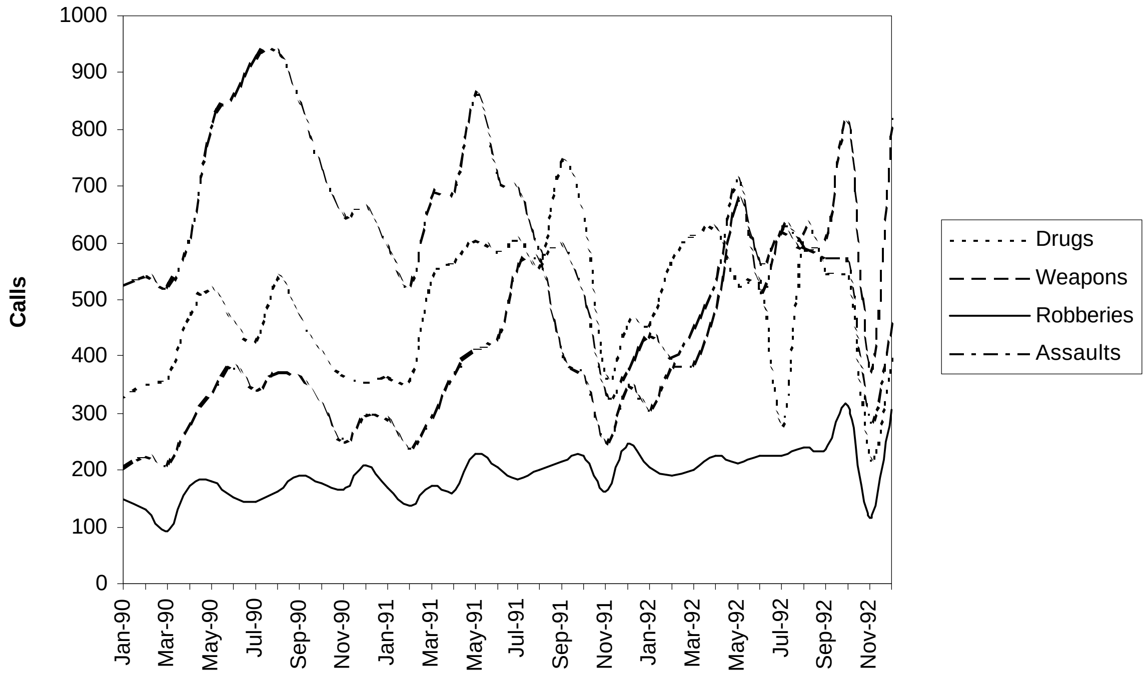
Table 1: Annual Total Number of Calls for Service

	Drugs	Weapons	Robberies	Assaults
1990	5053	3580	1922	8618
1991	6397	4523	2309	7154
1992	6223	6622	2699	6147

Figure 6 shows the number of calls for each nature code broken down by month. The seasonal variation is quite noticeable, especially for drugs, weapons and assaults. The figure also shows how, with the exception of assaults, most of the increase in the number of calls for service is accounted for in the summer months.

Models (7) and (8) were estimated using three different methodologies resulting in six different model estimates. The first methodology is simple regression and the second and third are Poisson regression and Tobit regression, respectively. The Tobit regression model was left censored at 0.1. Equation (9) could not be estimated since there are too many parameters for the results to be of any significance. Model (10) was estimated using chaotic cellular forecasting. In addition, all results were compared to a simple random walk, i.e., a model where the assumption is that the number of drug calls in the next time period are the same as those in the current period. With the exception of CCF and the random walk all models were estimated using the Stata statistical software package.

Figure 5: Monthly Calls for Service



The CCF model was estimated using a learning rate of 0.001 and nine hidden units. Again these figures were arrived at via experimentation on the data. Other learning rates and a different number of hidden units may be optimal for different data sets. The connection weights were randomly initialized to values in the range [-0.1, +0.1].

Both the regression and chaotic cellular forecasting models were estimated using the first two years' worth of data. The last year was used as a holdout sample to test the robustness of the estimated parameters. Tables 2 through 7 report the results of the simple (OLS) regression, Poisson regression and Tobit regression estimates of model (7) and (8) on the training data set for one-step-ahead forecasts.

Table 2 Estimates on Training Data Set (No Spatially Lagged Averages) - Simple Regression

Source		*Sum of Squared Errors*		*Degrees of Freedom*		*Mean Squared Error*	
Model		162212		7		23173.26	
Residual		45890		10672		4.30	
Total		208103		10679		19.49	
Adj. R^2	0.7793	*F-Value*	7541.32***	*Root MSE*	2.074	*# of Obs.*	10680
Variable		*Coefficient*		*Standard Error*		*t-Value*	
Constant		0.0070305		0.0580121		0.121	
Drugs		0.8224579		0.0057901		142.045***	
Weapons		0.1488242		0.128355		11.595***	
Robberies		0.0125878		0.0211918		0.594	
Assaults		0.062614		0.0085916		7.288***	
% Residential		0.0099798		0.0586685		0.170	
% Commercial		0.4258341		0.1881995		2.263**	
Season		-0.0853638		0.0738103		-1.127	
Significance Levels: * = 0.1, ** = 0.05, *** = 0.01							

The adjusted R-squared for the simple regression model without spatial lags was 0.7793. The F-statistic, significant at the 0.01 level, was 7541.32. However, only half of the estimated coefficients were statistically significant. Somewhat surprisingly the coefficient for the seasonality index was negative. Given that the number of drug calls for service tends to increase during the summer months one would have expected a positive coefficient. Not surprisingly, the number of drug calls for service in the previous time period had the largest coefficient, followed by the % of properties that are zoned commercial.

Table 3 Estimates on Training Data Set (No Spatially Lagged Averages) - Poisson Regression

Goodness-of-Fit Chi^2	*Log Likelihood*	*Pseudo R^2*	*Model Chi^2*
20145.380***	-15697.372	0.4649****	27276.032***
Variable	*Coefficient*	*Standard Error*	*z-Score*
Constant	-1.128023	0.0377612	-29.873***
Drugs	0.0395313	0.0005303	74.551***
Weapons	0.1004241	0.0017777	56.492***
Robberies	0.0437552	0.0040948	10.686***
Assaults	0.0643317	0.0015185	42.365***
% Residential	0.9181392	0.0404966	22.672***
% Commercial	1.054886	0.0757209	13.931***
Season	-0.2396972	0.0334107	-7.174***
Significance Levels: * = 0.1, ** = 0.05, *** = 0.01, ****as reported by Stata			

Table 3 shows the results of the Poisson regression using model (7). The pseudo R-squared reported by Stata is 0.4649, which is significant at the 0.01 level. Unlike the OLS results, all of the coefficients are significant. The % of properties that are commercial and residential have the largest coefficients.

Table 4 Estimates on Training Data Set (No Spatially Lagged Averages) - Tobit Regression

	Log Likelihood	*Pseudo R^2*	*Model Chi^2*
	-8909.5345	0.2564****	6142.94***
Variable	*Coefficient*	*Standard Error*	*t-Value*
Constant	-8.3880	0.2810	-29.847***
Drugs	0.8700	0.0137	63.402***
Weapons	0.4748	0.0318	14.912***
Robberies	0.1067	0.0536	1.990**
Assaults	0.3587	0.0222	16.188***
% Residential	2.8595	0.2617	10.927***
% Commercial	4.8561	0.5767	8.420***
Season	-0.2321	0.2510	-0.902
Significance Levels: * = 0.1, ** = 0.05, *** = 0.01, ****as reported by Stata			
Note: 8406 left censored observations at < 0.1, 2274 uncensored observations			

The Tobit regression results of model (7) shown in Table 4 indicate a pseudo R-squared of 0.2654, which is significant at the 0.01 level. Again, the most dominant coefficients are those of the ecological variables. 8406 observations were left censored.

Table 5 Estimates on Training Data Set (Spatially Lagged Averages) - Simple Regression

Source		*Sum of Squared Errors*		*Degrees of Freedom*		*Mean Squared Error*	
Model		162378		13		12490.63	
Residual		45725		10666		4.29	
Total		208103		10679		19.49	
Adj. R^2	0.7800	*F-Value*	2913.61***	*Root MSE*	2.0705	*# of Obs.*	10680
Variable		*Coefficient*		*Standard Error*		*t-Value*	
Constant		0.0323524		0.0712593		0.454	
Drugs		0.816167		0.0059132		138.025***	
Weapons		0.1283622		0.0135344		9.484***	
Robberies		0.0120371		0.0214847		0.560	
Assaults		0.0498935		0.0091128		5.475***	
% Residential		-0.007064		0.0736159		-0.096	
% Commercial		0.678445		0.2186717		3.103***	
Season		-0.1403245		0.0750493		-1.870*	
Drugs (Spat. Ave.)		0.0112826		0.0164524		0.686	
Weapons (Spat. Ave.)		0.0619071		0.0317452		1.950**	
Robberies (Spat. Ave.)		0.0316789		0.0571666		0.554	
Assaults (Spat. Ave.)		0.0487492		0.0206204		2.364**	
% Residential (Spat. Ave.)		-0.0766389		0.1161294		-0.660	
% Comm. (Spat. Ave.)		-1.367454		0.485068		-2.819***	
Significance Levels: * = 0.1, ** = 0.05, *** = 0.01							

Table 5 reports the (OLS) results of model 8. The adjusted R-squared is 0.7800, which is statistically significant at the 0.01 level. Only approximately half of the coefficients were statistically significant, as in model (7). Interestingly, the coefficient of the spatially lagged average of the proportion of commercial properties had a strong negative influence.

Table 6 Estimates on Training Data Set (No Spatially Lagged Averages) - Poisson Regression

Goodness-of-Fit Chi²	*Log Likelihood*	*Pseudo R²*	*Model Chi²*
17386.336***	-14317.850	0.5119****	30035.076***
Variable	*Coefficient*	*Standard Error*	*z-Score*
Constant	-1.89738	0.0528566	-35.897***
Drugs	0.0333808	0.0006144	54.329***
Weapons	0.0761946	0.0022289	34.185***
Robberies	0.0146796	0.0048114	3.051***
Assaults	0.043634	0.0018067	24.151***
% Residential	0.5973702	0.049468	12.076***
% Commercial	0.8313732	0.0845584	9.832***
Season	-0.3538514	0.0364302	-9.713***
Drugs (Spat. Ave.)	0.0775448	0.0035547	21.815***
Weapons (Spat. Ave.)	-0.0106852	0.0066702	-1.602*
Robberies (Spat. Ave.)	0.2112664	0.0163722	12.904***
Assaults (Spat. Ave.)	0.0474487	0.0056467	8.403***
% Residential (Spat. Ave.)	1.113307	0.0699516	15.915***
% Comm. (Spat. Ave.)	0.7582296	0.1932213	3.924***

Significance Levels: * = 0.1, ** = 0.05, *** = 0.01, ****as reported by Stata

Table 6 reports the Poisson regression results on model (8). The pseudo R-squared, as reported by Stata, is 0.5119, which is statistically significant at the 0.01 level. As with model (7), all coefficients are statistically significant, and the ecological variables have the strongest influence.

Table 7 Estimates on Training Data Set (No Spatially Lagged Averages) - Tobit Regression

	Log Likelihood	*Pseudo R²*	*Model Chi²*
	-8741.0613	0.2704****	6479.89***
Variable	*Coefficient*	*Standard Error*	*t-Value*
Constant	-9.3351	0.3455	-27.018***
Drugs	0.8230	0.0139	57.171***
Weapons	0.3501	0.0332	10.542***
Robberies	0.0438	0.0537	0.816
Assaults	0.2572	0.0231	11.156***
% Residential	2.0991	0.3133	6.700***
% Commercial	4.4018	0.7038	6.254***
Season	-0.7324	0.2660	-2.754***
Drugs (Spat. Ave.)	0.1426	0.0422	3.375***
Weapons (Spat. Ave.)	0.2461	0.0848	2.903***
Robberies (Spat. Ave.)	0.2972	0.1624	1.831*
Assaults (Spat. Ave.)	0.3609	0.0577	6.260***
% Residential (Spat. Ave.)	1.5087	0.4468	3.376***
% Comm. (Spat. Ave.)	-0.8892	1.6004	-0.556

Significance Levels: * = 0.1, ** = 0.05, *** = 0.01, ****as reported by Stata
Note: 8406 left censored observations at < 0.1, 2274 uncensored observations

The Tobit regression results of model (8) reported in Table 7 indicate a pseudo R-squared of 0.2704, which is significant at the 0.01 level. Again, the ecological variables appear to have the strongest influence, as do the number of drug calls for service in the previous time period for the current observation (not spatially lagged). With two exceptions all coefficients are statistically significant. As in model (7), 8406 observations were left censored at 0.1.

Table 8. Results of CCF Estimated Forecasts Compared to Regression Estimated Forecasts

	Models Without Spatially Lagged Averages		
Model	*Training Data Set*	*Holdout Sample*	*Iterations*
Simple Regression			
R^2	0.7795	0.7644	-
SSE	45890	22455	
Mean % Error	-128.16	-149.37	
Poisson Regression			
R^2	0.4079	0.4302	-
SSE	197447	88787	
Mean % Error	633.99	516.91	
Tobit Regression			
R^2	0.7108	0.6894	-
SSE	425369	198034	
Mean % Error	4784.72	4664.81	
Models With Spatially Lagged Averages			
Simple Regression			
R^2	0.7802	0.7644	-
SSE	45725	22449	
Mean % Error	-120.74	-149.23	
Poisson Regression			
R^2	0.3383	0.3666	-
SSE	199444	89861	
Mean % Error	702.73	703.03	
Tobit Regression			
R^2	0.6894	0.1887	-
SSE	457224	320908	
Mean % Error	4998.77	2904.63	
Random Walk			
Random Walk			
R^2	0.7725	0.7620	-
SSE	49988	24165	
Mean % Error	-93.30	-118.06	
CCF			
CCF			
R^2	0.8188	0.7800	4500
SSE	37740	20926	

* Note: The R-squared reported for CCF, Tobit, and Poisson, as well as the fit of the holdout sample using the regression coefficients is a simple squared correlation coefficient only, and can thus not be directly compared to the adjusted R-squared of the regression results on the training data set or the pseudo R-squared reported by Stata.

Table 8 compares the results of CCF to the other models, including the random walk. The results show that, overall only CCF and OLS could beat the random walk on both the training data set and the holdout sample. CCF still overfits somewhat, however, as evidenced by the drop in the fit of the holdout sample compared to the training data set, although the decline is still within the range of acceptability. Overall CCF reduces the sum of squared errors by about 17% on the training data set and by about 7% on the holdout sample compared to the closest competitor (OLS). CCF took only 4500 iterations to converge, which is a considerable improvement over early efforts (Olligschlaeger, 1997). The improvement in fit over the closest competitor is evidence that CCF was able to pick up nonlinearities in the data.

Table 9 Comparison of Models by Zero and Non-Zero Cells - Sum of Squared Errors

Model	Training Data Set Zero Cells (MPE)	Holdout Sample Zero Cells (MPE)	Training Data Set Non-Zero Cells (MPE)	Holdout Sample Non-Zero Cells (MPE)
Models Without Spatially Lagged Averages				
Simple Regression	2199 (-163.59)	1762 (-149.37)	43691 (2.79)	20693 (-5.56)
Poisson Regression	4786 (633.99)	2248 (642.19)	192691 (101.77)	86540 (99.01)
Tobit Regression	325666 (6009.70)	147515 (5992.37)	99703 (256.50)	50519 (236.48)
Random Walk	2000 (-120.64)	1577 (-155.34)	47987 (12.45)	22588 (6.31)
Models With Spatially Lagged Averages				
Simple Regression	2364 (-153.56)	1874 (-191.30)	43361 (0.60)	20575 (-8.89)
Poisson Regression	9210 (865.43)	4297 (884.56)	190233 (101.31)	85564 (97.51)
Tobit Regression	363951 (6286.25)	169038 (3802.81)	93274 (239.53)	151060 (-92.59)
CCF				
CCF	1179 (-150.87)	1463 (-205.59)	36560 (10.64)	19463 (6.86)

Table 9 further breaks down the comparison between those cells that are zero, i.e., had no drug related calls for service during a month's period, and those that had a least one call for service. This was done because crime data aggregated by areal units tend to have a Poisson distribution: most values are zero, and hot spots form a "spike" at the tail end of the distribution. When forecasting crime data by areal unit it is desirable to accurately capture the hot spots, rather than just correctly estimate that most areal units have in fact zero drug calls for service. With respect to multiple regression this would imply regression to the mean of zero with the tail end of the distribution being treated as outliers. For day to day police operations, however, it is important that police administrators have some means of predicting which areas are going to "flare" up.

It is clear both from the training data set results and the holdout sample reported in Table 9 that CCF consistently outperforms all of the other models. Indeed, CCF was the only model that was able to beat the random walk model in all categories. The only other model that came close to doing so were the OLS regression models. The uniformity of the results suggests that CCF is more accurately able to estimate the true functional form of the model. Most likely this is because it is able to incorporate nonlinearities in the data.

Table 10 Comparison of Models for Cells Where the Number of Drug Calls for Service is Zero at Time T and Non-Zero at Time T+1 - Sum of Squared Errors (Mean Percent Error in Parentheses)

Models Without Spatially Lagged Averages		
Model	*Training Data Set*	*Holdout Sample*
a - Simple Regression	1846 (72.36)	1071 (71.29)
b - Poisson Regression	3145 (123.94)	1734 (121.38)
c - Tobit Regression	32309 (521.27)	16113 (503.77)
d - Random Walk	2336 (91.37)	1313 (91.60)
Models With Spatially Lagged Averages		
e - Simple Regression	1767 (69.69)	1038 (67.65)
f - Poisson Regression	3206 (125.92)	1850 (124.32)
g - Tobit Regression	30203 (498.77)	23072 (71.11)
CCF		
CCF	1411 (65.04 - a,b,c,d,e,f,g)	1005 (67.49 - b,c,d,f,g)

Note: letters in parentheses indicate a statistically significant difference at the 0.01 level using a Wilcoxon signed rank test

Table 10 compares the relative performance of the models for those grid cells that displayed zero drug activity at time t and greater than zero drug activity in the next time period. In effect Table 10 gauges the performance of the models on emerging drug markets, or their usefulness as an early warning system for previously unknown drug hot spots. Again, CCF outperforms the other models, although the margin of improvement over OLS is not as great as that of the previous tables. However, with the exception of the two OLS models on the holdout sample, the decrease in the sum of squared errors is statistically significant at the 0.01 level using a Wilcoxon signed rank test.

6. Conclusion

This paper has introduced a new spatio-temporal forecasting method - chaotic cellular forecasting - that is derived from cellular automata (a class of chaotic systems) and artificial neural networks. One implementation of CCF, a hybrid model of spatially varying input to hidden units weights and constant hidden to output unit weights and direct input to output unit connections was tested using GIS based data. Specifically, CCF was used to produce one-step-ahead forecasts of drug calls for service based on crime and ecological variables. The results were then compared to six other forecasting models as well as a random walk model and tested for robustness on a holdout data sample. The performance of CCF is quite promising. It outperformed all of the other models, albeit with varying degrees of success, and was the only model to consistently outperform the random walk model in all categories examined. One disadvantage of CCF is that, like all neural networks, there are currently no tests of statistical significance for the estimated weight structures. However, since in forecasting the main goal is to provide accurate and robust forecasts rather than to analyze relationships between dependent and independent variables this should not be an issue.

Although these early results of chaotic cellular forecasting are very encouraging, much work remains to be done. Within the realm of artificial neural networks the algorithm is a fairly simple one. There have been many recent advances in the type of networks used by CCF.

Future work will include incorporating some of the advances in artificial neural network technology mentioned above. For example, there are many ways in which backpropagation networks can be modified so that they converge more quickly to a solution. An additional improvement would be to employ genetic algorithms to develop self-optimizing network architectures. Also, this paper only used a single hidden layer. The literature suggests that for very complex input to output mappings, such as is possibly the case for spatio-temporal data, two or more layers may be more appropriate in that they increase the capability of the network to capture nonlinearities in the relationship between input and outputs units.

Literature Cited

Anselin, Luc (1988). *Spatial Econometrics: Methods and Models*, Dordrecht, Netherlands: Kluwer.

Bates, J. M., and Granger, C. W. J. (1969). “The Combination of Forecasts.” *Operations Research Quarterly*, 20, 451-68.

Belsley, D., Kuh, E., and Welsch, R. (1980). *Regression Diagnostics*, New York: Wiley.

Blumstein, A., Cohen, J., and Farrington, D. P. (1987). “Characterizing Criminal Careers.” *Science,* 238, 985-991.

Blumstein, A., Cohen, J., and Farrington, D. P. (1988). “Criminal Career Research: Its Value for Criminology.” *Criminology,* 26: 1-35.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

Brantingham, P. J. and Brantingham, P. L. (1984). *Patterns in Crime*. Macmillan.

Bretschneider, S. L. and Gorr, W. L. (1983). “Ad Hoc Model Building Using Time Varying Parameter Models.” *Decision Sciences* 14, 221-39.

Brown, R. G. (1962). *Smoothing, Forecasting, and Prediction of Discrete Time Series*, Englewood Cliffs, NJ: Prentice-Hall

Carbone, R. and Longini, R. (1977). “A Feedback Approach for Automated Real Estate Assessment.” *Atmospheric Environment*, 12: 241-48

Casetti, E. (1972). “Generating Models by the Expansion Method: Applications to Geographic Research.” *Geographical Analysis* 4: 81-91

Casetti, E. (1973). “Testing for Spatial-Temporal Trends: An Application of Urban Population Density Trends Using the Expansion Method.” *Canadian Geographer* 17: 127-36.

Cassetti, E. (1982). “Drift Analysis of Regression Parameters: An Application to the Investigation of Fertility Development Relations.” *Modeling and Simulation 13*, 961-66.

Casetti, E. (1986). “The Dual Expansion Method: An Application to Evaluating the Effects of Population Growth on Development.” *IEEE Transactions on Systems, Man and Cybernetics* SMC-16, 29-39.

Casetti, E. and Fan, C. (1991). “The Spatial Spread of the AIDS Epidemic in Ohio: Empirical Analyses Using the Expansion Method.” *Environmental Planning A* 23: 1-23.

Cassetti, Emilio and Jones, J.P. eds. (1992). *Applications of the Expansion Method*. London: Routledge.

Clemen, R. T. (1989). “Combining Forecasts: An Annotated Bibliography.” *International Journal of Forecasting*, 5, 559-83.

Cleveland, W. S. (1979). “Robust Local Weighted Regression and Smoothing Scatterplots.” *Journal of the American Statistical Association* 74: 829-36.

Cleveland, W. S. and Devlin, S. J. (1988). “Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting”. *Journal of the American Statistical Association 83*, 596-610.

Cliff, A.D., Haggett, P., Ord, J.K., Bassett, K. and Davies, R. (1975). *Elements of Spatial Structure*. Cambridge, England: Cambridge University Press.

Cohen, L.E. and Felson, M. (1979). “Social Change and Crime Rate Trends: A Routine Activity Approach.” *American Sociological Review* 44: 588-607.

David, M. (1977). *Geostatistical Ore Reserve Estimation*, Amsterdam: Elsevier.

Dunn, C. S. (1980). “Crime Area Research.” In *Crime: A Spatial Perspective*, edited by D. E. Georges-Abeyie and K. D. Harries. New York: Columbia University Press.

Felson, M. (1987). “Routine Activities and Crime Prevention in the Developing Metropolis.” *Criminology* 25:911-31.

Foster, S. A. and Gorr, W. L. (1986). “An Adaptive Filter for Estimating Spatially Varying Parameters: Application to Modeling Police Hours Spent in Response to Calls for Service.” *Management Science*, 32: 878-89.

Foster, S. A. W., Gorr, W. L. and Wimberly, F. C. (1992). “Comparison of Functional and Drift Analysis Approaches to Expansion Modeling: Evaluation of Federal Policies on the Supply of Physicians.” In Casetti, E. and Jones, J. P. III (eds.): *Applications of the Expansion Method*. London: Routledge.

Getis, A. (1994). “Spatial Filtering in a Regression Framework: Experiments on Regional Inequality, Government Expenditures and Urban Crime.” In Anselin, L. and Florax, R. (eds.): *New Directions in Spatial Econometrics*. Amsterdam: North Holland.

Gardner, M. (1970). “The Fantastic Combinations of John Conway’s New Solitaire Game ‘Life’”. *Scientific American*, 223:4, pp. 120-23.

Gorr, W.L., (1994). "Research Prospective on Neural Network Forecasting," *International Journal of Forecasting*, Vol. 10, pp. 1-4.

Gorr, W.L. and Olligschlaeger, A.M. (1994). “Weighted Spatial Adaptive Filtering: Monte Carlo Studies and Application to Illicit Drug Market Modeling.” *Geographical Analysis*, 26: 67-87.

Granger, C. W. J. (1969). “Investigating Causal Relations in Econometric Models and Cross-Spectral Models.” *Econometrica*, 37: 424-438.

Haining, T. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge, England: Cambridge University Press.

Hornik, K., Stinchcombe, M. and White, H. (1989). “Multilayer Feedforward Networks are Universal Approximators.” In: *Neural Networks*, Vol. 2, Pergamon Press.

Kabel, J. R. (1992). “A Geographic Perspective on AIDS in the United States: Past, Present and Future.” Ph.D. Dissertation, Geography Department, The Pennsylvania State University, State College, PA.

Kroese, B. J. and Van der Smagt, P. P. (1993). *An Introduction to Neural Networks*, lecture notes, University of Amsterdam, Netherlands.

Lander, B. (1954). *Toward an Understanding of Juvenile Delinquency*. Columbia University Press.

Lapedes, A. and Farber, R. (1987). *Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling*. Technical Report LA-UR-87-2662, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM.

Makridakis, S. and Wheelwright, S. C. (1977). “Adaptive Filtering: An Integrated Autoregressive/Moving Average Filter for Time Series Forecasting.” *Operations Research Quarterly*, 28: 425-37.

Maltz, M. D. (1993). “Crime Mapping and the Drug Market Analysis Program (DMAP).” In *Workshop on Crime Analysis Through Computer Mapping Proceedings: 1993*, edited by C. R. Block and M. Dabdoub, Illinois Criminal Justice Authority.

McEwen, J.T. and Taxman, F. S. (1994). “Applications of Computerized Mapping to Police Operations.” In *Crime, Place and Police*, edited by J. E. Eck and D. Weisburd, Crime Prevention Series, Vol. 4.

Minsky, M. and Papert, S. (1969). *Perceptrons*. Cambridge, PA: MIT Press.

Newbold, P. and Granger, C. W. J. (1974). “Experience with Forecasting Univariate Time Series and the Combination of Forecasts.” *Journal of the Royal Statistical Society* 137: 131-46.

Olligschlaeger, A. M. (1997). “Artificial Neural Networks and Crime Mapping.” In McEwen, T. and Weisburd, D. (eds.): *Computerized Crime Mapping*, Crime Prevention Series, Rutgers University Press.

Pickover, C. A. (1990). *Computers, Pattern, Chaos and Beauty*. New York: St. Martin’s Press.

Poli, I. And Jones, R. D. (1994). “A Neural Net Model for Prediction.” *Journal of the American Statistical Association*, 89:117-21.

Robbins, H. and Monro, S. (1951). “A Stochastic Approximation Method.” *Annals of Mathematical Statistics*, 22: 400-407

Roncek, D. W. and Baier, P. A. (1991). “Bars, Blocks and Crimes Revisited: Linking the Theory of Routine Activities to the Empiricism of ‘Hot Spots’.” *Criminology* 29: 725-53.

Rossmo, K. (1995). “Place, Space and Police Investigations: Hunting Serial Violent Criminals.” In Eck, J. and Weisburd, D. (eds.) *Crime and Place*. Monsey, NY: Criminal Justice Press.

Rumelhart, D. E. and McClelland, J. L. (1988). *Parallel Distributed Processing*, Vols. 1 and 2, Cambridge, MA: MIT Press

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1988). “Learning Internal Representations by Error Propagation.” In Rumelhart, D. E. and McClelland, J. L. (eds.) *Parallel Distributed Processing*, Vol. 1, Cambridge, MA: MIT Press

Schmid, C. F. (1960). “Urban Crime Areas: Part II.” *American Sociological Review* 25: 655-78.

Schroeder, M. (1991). *Fractals, Chaos and Power Laws*. New York, New York: W.H. Freeman.

Sharda, R. and Patil, R. B. (1990). “Neural Networks as Forecasting Experts: An Empirical Test.” In: *IEEE INNS Joint Conference on Neural Networks*, II, 491-494, Washington, DC: January, 1990.

Shaw, C. R. (1929). *Delinquency Areas*. University of Chicago Press.

Sherman, L. W., Gartin, P. R. and Buerger, M. E. (1989). “Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place.” *Criminology* 27: 27-55.

Skogan, W. (1986). “Fear of Crime and Neighborhood Change.” In A. J. Reiss and M. Tonry, *Communities and Crime 8*, University of Chicago Press.

Spring, J. W. and Block, C. R. (1988). “Finding Crime Hot Spots: Experiments in the Identification of High Crime Areas. “ Paper presented at the 1988 annual meeting of the Midwest Sociological Society, Minneapolis, MN

Tang, Z., Almeida, C. and Fishwick, P. A. (1990). *Time Series Forecasting Using Neural Networks Vs. Box-Jenkins Methodology*. UF-CIS Technical Report TR-90-3, Computer and Information Sciences Department, University of Florida, Gainesville, FL.

Tobler, W. R. (1969). “Geographical Filters and their Inverses.” *Geographical Analysis*, 1, 234-53.

Tobler, W. R. (1979). “Cellular Geography.” In Gale, S. and Olsson, G. (eds.): *Philosophy of Geography*. Dordrecht: Reidel.

Toffoli, T. and Margolus, N. (1987). *Cellular Automata Machines: A New Environment for Modeling*. MIT Press Series in Scientific Computations, Cambridge: MIT Press.

Trigg, D. W. and Leach, D. H. (1967). “Exponential Smoothing with an Adaptive Response Rate.” *Operational Research Quarterly* 18: 53-59.

Weiss, S. M. and Kulikowski, C. A. (1991). *Computer Systems That Learn*. Morgan Kaufmann, San Mateo.

White, H. (1988). “Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns.” *Proceedings of the IEEE International Conference on Neural Networks*, San Diego

Widrow, G. and Hoff, M. E. (1960). “Adaptive Switching Circuits.” *Institute of Radio Engineers, Western Electronics Show and Convention*