

# A Spatial Model of Received Signal Strength Indicator Values for Automated Collision Notification Technology

Eric M. Delmelle<sup>1</sup>, Peter A. Rogerson<sup>2,3,4,6</sup>, Mohan R. Akella<sup>5</sup>, Rajan Batta<sup>5,6</sup>  
Alan Blatt<sup>7</sup> and Glen Wilson<sup>7</sup>

December 9, 2005

<sup>1</sup>Department of Geography, University of Idaho, Moscow, ID 83843, U.S.A.

<sup>2</sup>Corresponding author. Tel: (716) 645-2722 xt. 53. Email: [rogerson@buffalo.edu](mailto:rogerson@buffalo.edu)

<sup>3</sup>Department of Geography, University at Buffalo (SUNY), Buffalo, NY 14261, U.S.A.

<sup>4</sup>Department of Biostatistics, University at Buffalo (SUNY), Buffalo, NY 14214, U.S.A.

<sup>5</sup>Department of Industrial Engineering, University at Buffalo (SUNY), Buffalo, NY 14260, U.S.A.

<sup>6</sup>Also associated with the National Center for Geographic Information and Analysis.

<sup>7</sup>Center for Transportation Injury Research, General Dynamics, Buffalo, NY 14225, U.S.A.

## Abstract

This paper reports on a spatial model that can be used to support evaluation of the effectiveness of Automated Collision Notification (ACN) technology. Such systems rely on wireless phones to transmit their crash messages. However, cellular coverage exhibits great variation in hilly and rural areas, and these are regions where ACN devices provide their greatest benefits. In this context, the accurate prediction of cell phone signal strength is of paramount importance and hence, an in-depth analysis of the variables influencing the cellular signal is necessary. In this paper, we first model the spatial variation of cell phone coverage using an interpolation method known as kriging. We focus on how particular factors such as slope, altitude, distance to the cell tower and topographic obstruction are used to improve the prediction of cell phone signal strength. The effect of these factors is evaluated in a multiple regression framework, followed by an analysis of the residuals, providing valuable information for cellular network providers. However, since the residuals show some spatial structure, we therefore analyze the merits of a more suitable method, namely spatial regression. We illustrate the approach using data on cell phone coverage in a hilly section of Erie county, New York. The model should (a) enhance our knowledge of the effect of particular factors on signal strength intensity, and (b) result in improved estimates of RSSI at unknown locations. The latter is important in establishing the effectiveness of Automated Collision Notification (ACN) technology.

**Keywords:** Received Signal Strength Indicator (RSSI), Automated Collision Notification (ACN), Geographical Information System (GIS), kriging, linear regression, spatial regression.

# 1 Introduction

## 1.1 Context of the study

The number of vehicle crashes continues to climb due to increases in traffic volume and extended travel times. Correspondingly, the fatality rate remains at unacceptably high levels. According to the National Highway Traffic Safety Administration (NHTSA 2004), there were 6,328,000 million motor vehicle crashes in the US in 2003. In rural areas fatality rates are disproportionately high. There is a substantial body of literature regarding the impact of Emergency Medical Service (EMS) response time and time to definitive care on trauma victim outcomes. Terms such as 'platinum ten minutes', 'golden hour' (Jacobs *et al.* 1984; Lerner and Moscati 2001) and 'silver day' (Blow *et al.* 1999) have been coined to describe the importance of time in treating trauma injuries. Approximately 50% of fatalities occur within minutes of the injuries, 30% within hours, and 20% within days to weeks (Petri *et al.* 1995; Sampalis *et al.* 1993 and Pepe *et al.* 1987). To evaluate technologies that might reduce crash-related fatalities and response time, the NHTSA sponsored the Automated Collision Notification (ACN) Field Operational Test Program from 1995 to 2000. ACN explored the ability of in-vehicle equipment to reliably sense and characterize crashes, and automatically transmit crash location and crash severity data to the proper public safety agencies. Since the NHTSA-sponsored ACN program, several more advanced commercial systems have been introduced, such as the recently announced next generation of OnStar. As a minimum, the emergency message transmitted by the vehicle provides the geographic location of the crash and an indication of the type and severity of the crash. Once the data is received by the Public Safety Answering Point it is used to dispatch and route emergency services (fire, police and EMS) to the scene. The wireless call must transmit at least a relatively short burst of data containing the vehicle location. More robust systems provide additional data that can be used to characterize the crash severity and estimate potential occupant injuries.

Robust cellular communications are essential for ACN effectiveness. Received Signal Strength Indicator (RSSI) is a measure of the bond between the customer's cellular phone and the cellular service provider's tower or cell site. The strength of this bond determines

the likelihood of completing a phone call within that specific cell. Many factors, including distance to the nearest cell tower, foliage cover, presence of buildings, topographical features, and loading of the cellular system contribute to the integrity of a cellular connection. In most urban and suburban areas, signal strength is more than adequate for cellular communications. In rural, hilly or forested areas, the signal strength may exhibit significant variations at a scale of a few hundred feet. Sometimes the signal strength may drop low enough to be inadequate for establishing a cellular connection, or may cause a delay in establishing a call.

## 1.2 Contributions

The ACN device can reduce the time between a crash and the notification to the EMS by generating a call automatically, provided the signal strength is strong enough at the location of the crash. Mayday systems offer the greatest promise in rural, hilly and remote areas, since crashes can otherwise go unreported for long periods of time (Walton and Meyer 2000). It is the goal of this paper to assess the role of topographic factors on cell phone coverage, incorporating information that is cell-tower dependent such as the distance separating a point to its nearest base station, but also the terrain obstructions along the path connecting those two points. In turn, this topographic information is combined with existing variables into a multiple regression framework, and we evaluate (1) how well the new model predicts the signal strength and (2) the impact of each variable on signal strength. A predictive spatial model for RSSI values will be useful in (a) improving the assessment of the benefit of ACN systems and (b) assisting in the placement of new cell towers aimed at improving emergency coverage (see Akella *et al.* 2005).

## 1.3 Preliminary studies

We undertook some preliminary studies and these are summarized in Akella *et al.* (2003). We empirically sampled RSSI on major and minor roads in Erie County NY (Figure 1a), from November 1999 through April 2000. A set of 6864 data points was collected for the analog band of the 800MHz cellular frequency (Figure 1b). The test apparatus main components are the In-Vehicle Module (IVM), which is a small "black box" containing a

GPS Board, modem, non-volatile memory and handset control circuitry. The apparatus was installed in a test vehicle and utilized magnetic mount GPS and cell antennas. While the vehicle is in motion, the IVM records RSSI data by automatically re-initializing the transceiver every .8 km based on GPS readings. RSSI is reported by the transceiver in 5dB increments, ranging from -53dB to -123dB. Values of -89dB and under denote low RSSI values affecting call completion. The data is stored as a separate text file, containing for each sample point the day, longitude, latitude, RSSI value and speed of the vehicle (Table 1).

INSERT TABLE 1 ABOUT HERE

## 1.4 Structure of the paper

In the next section, we use kriging to estimate the value of RSSI for locations where it is unknown. In section 3, we use regression analysis to assess the effects of slope and elevation on RSSI values. Capitalizing on the findings from the regression residuals, we are able to compute cell tower distances and obstructions occurring along the way. In section 4, we define the regression model, and a case study area is selected to test our model. The impact of each variable on cell phone signal strength is analyzed in a regression framework. Unfortunately, multiple regression ignores the spatial structure of the variables, and we therefore review a more suitable regression method for that purpose. We compare the benefits and drawbacks of spatial regression to multiple regression based on the magnitude of the residuals and their spatial pattern. Some thoughts on reducing the amount of spatial pattern in the residuals is suggested. The final section of the paper is devoted to a discussion of the implications for cellular network providers.

## 2 Interpolating cell phone coverage using kriging

Interpolation of data from sampled locations is necessary to estimate RSSI values at unknown locations. The Inverse Distance Weighted (IDW) technique used in Akella *et al.* (2003) is based on the assumption that the RSSI value at an unsampled point can be approximated as a deterministic weighted average of existing data points, where the weights are inversely proportional to the distance away from the sample points (Mitas

and Mitasova 1999). However, as Burrough (1986) points out, there are better ways to estimate the weights than as a deterministic function of distance. It is desirable to avoid the presence of local extremes at data point locations, a pitfall of the IDW technique. Kriging is better than traditional interpolation techniques since the weights are chosen to optimize the interpolation function, and to minimize the prediction error. Central to kriging is the semivariogram  $\hat{\gamma}(h)$ , summarizing the variance of values separated by a particular distance lag ( $h$ ), where  $d(h)$  is the number of pairs of points for a given lag value, and  $y(\mathbf{s}_i)$  is the measured RSSI value at location  $\mathbf{s}_i$ . The semivariogram is estimated as

$$\hat{\gamma}(h) = \frac{1}{2d(h)} \sum_{|\mathbf{s}_i - \mathbf{s}_j| = h} \left( y(\mathbf{s}_i) - y(\mathbf{s}_j) \right)^2 \quad (1)$$

The semivariogram is characterized by a nugget effect  $a$  and a sill  $\sigma^2$  where  $\hat{\gamma}(h)$  levels out. The nugget effect is the spatial dependence at micro scales (Cressie 1991). Once the lag distance exceeds a certain value  $r$ , called the range, there is no spatial dependence between sample sites anymore; usually, an exponential model  $\gamma(h)$  is fitted to the experimental semivariogram.

$$\gamma(h) = a + (\sigma^2 - a)(1 - e^{-\frac{3h}{r}}) \quad (2)$$

The interpolated, kriged value at a location  $\mathbf{s}$  in space is given as a weighted mean of surrounding values, where each value is weighted according to the semivariogram model:

$$\hat{y}(\mathbf{s}) = \sum_{i=1}^W w_i(\mathbf{s}) y(\mathbf{s}_i) \quad (3)$$

where  $W$  is the set of neighboring points that are used to estimate the interpolated value at location  $\mathbf{s}$ , and  $w_i(\mathbf{s})$  is the weight associated with each surrounding point. Usually, kriging is performed on a set of grid nodes  $\mathbf{s}_g$  ( $g = 1, 2, \dots, G$ ). ArcMap (ESRI, Redlands, CA) was used to estimate an exponential semivariogram from sample RSSI data. Data values are spatially dependent up to a distance of approximately  $r=1200\text{m}$ , with a nugget effect of  $\sigma^2=42.61\text{dB}$ . Figure 1c shows the interpolated RSSI map for Erie county, suggesting that existing cellular coverage is adequate to support mayday system technology in most urban and suburban locations. Some areas in the southern part of Erie county show substantially

lower signal strength values however, most likely due to topographic variation. In addition, the rural northeast of the county is also characterized by low RSSI values.

INSERT FIGURE 1 ABOUT HERE

### 3 Problem Statement and Initial Regression Model for RSSI

In this paper, we hypothesize that the spatial variation of cell phone coverage is influenced by the presence of many factors. We attempt therefore to evaluate the impacts of geographic variables on RSSI. We hypothesize that topographic variables such as slope, altitude, and the obstructions occurring along the way between the receiver and the base station (buildings, hills) may impact RSSI significantly. Obstructions in general, and buildings in particular are difficult to account for. Additional factors such as distance to the cell tower, the height of the cell tower and the receiver's antenna, the base station transmitting power and vegetation should be accounted for as well. Four essential questions that merit further investigation are:

1. Which covariates influence the primary variable?
2. By how much do these variables affect the primary variable?
3. How can this information improve upon the prediction of cell phone signal strength?
4. Which methods are appropriate to account for this information?

As previously discussed, one issue pertains to by how much we can improve upon the prediction of RSSI, and to how much we can reduce the residuals. If, in addition to slope and altitude, we account for variables that depend on the location of the cell tower (such as distance to the cell tower, power radius, and the line-of-sight), we may be successful in reducing the spatial patterns in the residuals, a desirable goal for regression analysis. There are several alternatives for incorporating this information in a statistical model. A multiple regression framework can yield valuable results at first, but more significant results can be obtained by capitalizing on spatial information provided by nearby samples. For computation purposes, we apply these techniques to a case study in southern Erie county.

We first construct our prediction model by carefully choosing explanatory variables and analyzing how multiple regression performs on predicting the dependent variable. We also analyze spatial patterns in the residuals using Moran’s  $I$  statistic. One objective is to reduce the residual pattern using a more suitable regression technique accounting for the information provided by nearby samples. The merits of each of the two techniques are compared.

### 3.1 Impact of slope and altitude on RSSI

We hypothesized that RSSI becomes less predictable and, on average, lower, in areas where there is a strong altitude variation. We used a set of Digital Elevation Models (DEMs) for Erie county in our analysis, provided by the Cornell University Geospatial Information Repository (CUGIR). A DEM is a layer of small cells (10meters×10meters), each having an altitude value (see Figure 3a). Slopes were derived from the DEMs, and expressed in angular degrees (see Figure 3b). The slope for a raster cell is calculated from the eight nearest neighbors in a 3×3 cell window using the average maximum technique. The algorithm used to compute the slope  $S$  for a cell  $e$  can be formulated as follows (Burrough and McDonnell 1998)

$$\tan S_e = \sqrt{\left\{ \frac{\partial z}{\partial x} \right\}^2 + \left\{ \frac{\partial z}{\partial y} \right\}^2} \quad (4)$$

where  $z$  is the altitude while  $x$  and  $y$  represent the coordinate axes. To calculate  $\left\{ \frac{\partial z}{\partial x} \right\}$  and  $\left\{ \frac{\partial z}{\partial y} \right\}$ , consider the following example wherein the slope of the central cell  $e$  has to be computed. We have a 3×3 moving window over 8 neighboring cells  $a, b, c, d, f, g, h$  and  $i$ .

INSERT FIGURE 2 ABOUT HERE

To analyze whether there is any relation between these two factors and RSSI, we performed a multiple regression analysis (Equation 5), where  $x_s$  and  $x_a$  denote the slope and altitude variables respectively and  $\mathbf{s}_i$  is a row vector denoting the location of an existing RSSI data point.  $\hat{y}(\mathbf{s}_i)$  is the dependent variable RSSI. The initial specification is:

$$\hat{y}(\mathbf{s}_i) = b_0 + b_s x_s(\mathbf{s}_i) + b_a x_a(\mathbf{s}_i) \quad (5)$$

As a result of the regression analysis, the following coefficients were obtained:

$$\hat{y}(\mathbf{s}_i) = -68.084dB - .65 x_s(\mathbf{s}_i) - .021 x_a(\mathbf{s}_i). \quad (6)$$

This means that RSSI would drop by approximately 6.5dB when the slope angle increases by 10 degrees. Similarly, a 100 meter increase in altitude would decrease the signal by approximately 2.1dB. Note that these parameters may not be constant throughout Erie county, showing strong discrepancies between flat and hilly areas.

In a regression framework, the residuals, denoted  $u$ , are calculated as the difference between the observed and the predicted values.

$$u(\mathbf{s}_i) = y(\mathbf{s}_i) - \hat{y}(\mathbf{s}_i) \quad (7)$$

When the value of  $u(\mathbf{s}_i)$  is positive -i.e.  $y(\mathbf{s}_i) \geq \hat{y}(\mathbf{s}_i)$ -, we have underpredicted the primary variable. If the residuals from this regression are interpolated (Figure 3c), one can observe that high positive residuals values may correspond with existing cell towers, because RSSI values at these locations have been underpredicted. We did not include cell tower location information in the initial specification because the location of all cell towers was initially unknown. However, incomplete information regarding cell tower locations was obtained from the Federal Communication Commission (FCC 2001) and interestingly most of these locations are locations where residuals are strongly positive. The rationale is that RSSI values in the proximity of cell towers show strong values, but the multiple regression ignores this. For validation purposes, we drove through five of these strong positive residual areas and in all five of the cases, there was a cell tower near the center of the zone.

INSERT FIGURE 3 ABOUT HERE

### 3.2 Capitalizing on Cell Tower Location

Capitalizing on the findings from the previous regression, it is possible in a GIS to compute the distance between any RSSI data point and its nearest cell tower. Additionally, we can determine whether any point in the study area can be viewed from a cell tower or not,

given a visibility distance and the height of the cell tower. This visibility analysis accounts for obstructions along the path separating any point in the study area and cell towers (see Clarke 1995). Resulting from a terrain obstruction analysis is a binary surface with values of 1 when a location is visible from a cell tower, and 0 otherwise (Figure 4a). This is a simplification since in real circumstances the signal is attenuated and reflected depending on terrain and building characteristics (Figure 4b). We have performed this analysis for the entire county. In general, a greater area is visible in the northern part of the county and along the shoreline, since the area is relatively flat (Figure 3a). As we move southeast, variation in the visibility appears as a result of the topography.

INSERT FIGURE 4 ABOUT HERE

## 4 A Spatial Model of RSSI

Cell phone signal strength varies with distance to the cell tower (Macario 1997). In ideal conditions, in free space where there are no obstructions between the cell tower and the location where the signal is measured, the signal loss in dB at a distance  $d$  from a cell tower is given by Krishnamachari (1999), where  $x_d$  denotes the distance to the nearest cell tower.

$$Loss(d) = b_0 + b_d \ln(x_d) \tag{8}$$

We have empirically determined the value of the coefficients  $b_0$  and  $b_d$  in Equation 8. We have concentrated our effort around a cell tower in a remote area in the northern part of Erie county, minimizing the risks of interference with nearby cell towers from the same provider. We have measured RSSI values at gradually increasing distances to the cell tower. RSSI values drop from -50dB in the immediate vicinity of the cell tower to about -70dB at 1200m and about -85dB at 4000m. At a conservative distance of 5000m, RSSI becomes weak enough that establishing a phone call is questionable. Since we know that the relationship between RSSI values and distance to the nearest cell tower follows a negative exponential

curve, we can rewrite Equation 8 as:

$$\widehat{RSSI}(\mathbf{s}_i) = b_0 e^{b_d x_d(\mathbf{s}_i)} \quad (9)$$

$$\ln(-\widehat{RSSI}(\mathbf{s}_i)) = \ln(-b_0) + b_d x_d(\mathbf{s}_i) \quad (10)$$

$$\widehat{y}(\mathbf{s}_i) = \ln(-b_0) + b_d x_d(\mathbf{s}_i) \quad (11)$$

The following variables are the explanatory variables:

$x_s$  = slope in degrees at an RSSI data point

$x_a$  = altitude value at an RSSI data point

$x_r = 1$ , if RSSI is in range of the cell tower radius, and 0 otherwise

$x_v = 1$ , if RSSI is visible from the nearest cell tower, and 0 otherwise

$x_d$  = distance between an RSSI point and its nearest cell tower

The regression model is set as follows:

$$\widehat{y}(\mathbf{s}_i) = \ln(-b_0) + b_s x_s(\mathbf{s}_i) + b_a x_a(\mathbf{s}_i) + b_r x_r(\mathbf{s}_i) + b_{rd}(x_r x_d(\mathbf{s}_i)) + b_v x_v(\mathbf{s}_i) \quad (12)$$

where  $(x_r x_d(\mathbf{s}_i))$  is an interaction term because we only consider the distance effect if an RSSI point is within the range of a cell tower. At this point, we have not included vegetation information since data on trees is difficult to obtain and to interpret. For the same reason we have not incorporated information on the presence of buildings.

#### 4.1 Choice of a study area

Since the computational time to calculate the results for the entire county (6864 RSSI data points) was very large, we decided to limit the analysis to a representative area in the close neighborhood of a cell tower. This area is located in the southern part of Erie county (see square area in Figure 3) and characterized by significant altitude variation. In Figure 5, the triangle in the middle of each map is the cell tower. Map 5a represents the altitude, and the slope is depicted in Figure 5b. Map 5c results from the cell tower anisotropic line-of-sight analysis, with a radius of 5000 meters, and a cell tower height of 35 meters. We performed

this analysis within ArcInfo (ESRI, Redlands, CA), a Geographical Information System (GIS) using a set of AMLs, the ArcInfo programming language. Visibility is computed for a coverage radius of 5000m, a tower height of 35m and a receiver’s antenna of 2 meters above the ground. This height of 2 meters corresponds approximately to an antenna’s height placed on the roof of a car. Finally, map 5*d* is an interpolated map of the RSSI values, where stronger values are observed in the vicinity of the cell tower. There are 229 RSSI data points in the study area. All of the regression results were obtained using Matlab v. 6.5 and SPSS v.11. Results in the form of maps were created in ArcMap.

INSERT FIGURE 5 ABOUT HERE

## 4.2 The multiple regression approach

We are interested in the significance of each explanatory variable, summarized in Table 2. All variables are significant, except the slope coefficient  $b_s$ , at the 5%-level. Since the dependent variable is  $\ln(-RSSI)$ , negative coefficients should be interpreted as having a positive impact on the received signal strength. An increase in altitude has a positive impact on the RSSI value. This is in contradiction with the results of Equation 6, where an increase in the altitude value had a negative impact. But intuitively, in hilly areas, the higher the altitude, the more likely the call will go through as the obstructions to the nearest cell tower decrease. An existing RSSI point visible from its nearest cell tower has a positive impact on the RSSI. Similarly, an increasing distance to the cell tower decreases the RSSI value. RSSI values increase when a sample point is within the coverage radius or when it is visible from the cell tower. An increasing slope will have a negative impact on the signal strength. For comparison purposes, we performed a multiple regression with slope and altitude only for this study area. Both variables are more significant, but the  $R^2$  is .095. The  $R^2$ -values are unadjusted for degrees of freedom. Adding the range-distance term improves the  $R^2$  to .154. Consequently, the significance of the altitude term  $b_a$  decreases in favor of the distance-range term. The slope term  $b_s$  is slightly more significant. Adding the visibility term  $b_v$  and the range term  $b_r$  increases the magnitude of the  $R^2$  to .282. The importance of slope and altitude has diminished in favor of the new variables. The standard error of the estimate, defined in Equation 13, is another expression for the standard deviation of

the residuals:

$$s_e = \sqrt{\sum_{i=1}^n \frac{(y(\mathbf{s}_i) - \hat{y}(\mathbf{s}_i))^2}{n-2}} \quad (13)$$

Having only altitude and slope in our model yields a  $s_e$ -value of .20649, while adding the interaction term accounting for the distance effect if an RSSI point is within the range of a cell tower decreases the  $s_e$  value to .18895. Adding visibility and range explanatory variables drops the  $s_e$ -value to .18520, an improvement over the initial standard error of approximately 12%.

INSERT TABLE 2 ABOUT HERE

**Analysis of the residuals.** Figure 6 on the right shows the interpolated residuals for the multiple regression, after taking the back transform of  $\ln(-\widehat{RSSI})$ . As expected, a significant amount of spatial pattern remains in the residuals from the multiple regression. For comparison purposes, the map on the left displays the residuals from a multiple regression using slope and altitude only. Also of interest is whether there are still spatial patterns in the residuals. Moran's  $I$  (Moran 1948, 1950) is a measure of the degree of spatial autocorrelation among data points:

$$I = \frac{n \sum_i \sum_j w_{\mathbf{s}_{ij}} (u_{\mathbf{s}_i} - \bar{u})(u_{\mathbf{s}_j} - \bar{u})}{(\sum_i \sum_j w_{\mathbf{s}_{ij}}) \sum_j (u_{\mathbf{s}_j} - \bar{u})^2} \quad (14)$$

with  $n$  being the number of observations and the mean of the residuals denoted  $\bar{u}$ . The weight  $w_{\mathbf{s}_{ij}}$  is a measure of spatial proximity between points  $\mathbf{s}_i$  and  $\mathbf{s}_j$ :

$$w_{\mathbf{s}_j} = e^{-\beta d_{\mathbf{s}_{ij}}^2} \quad (15)$$

where  $d_{\mathbf{s}_{ij}}^2$  is the squared distance between location  $\mathbf{s}_i$  and point  $\mathbf{s}_j$ . The solid curve on Figure 8 to the right displays Moran's  $I$  values for the residuals of the multiple regression, plotted for a range of  $\beta$  values. Note that  $\beta$  is relatively small, since we are using squared meters units. The value of Moran's  $I$  is greater than zero, denoting some positive spatial autocorrelation. This correlation increases as we increase the value of  $\beta$ , i.e. when relatively more weight is given to nearby locations than further away locations. In general, we wish

Moran’s  $I$  to be close to zero, implying no pattern in the residuals; this is an assumption of regression, and is important for accurate estimates of the regression coefficients. Very noteworthy is the fact that the Moran’s  $I$  curve is increasing up to a maximum  $\beta$  value of .00005, and then decreases. An interesting objective is whether we can reduce significantly this maximum Moran’s  $I$ . A solution consists of using an alternative regression method that accounts for the spatial structure of the residuals, namely spatial regression. Using this additional information reduces the spatial autocorrelation in the residuals.

INSERT FIGURE 6 ABOUT HERE

### 4.3 Spatial regression

A major shortcoming of the multiple regression framework is that the RSSI value at location  $\mathbf{s}_i$  is derived only from explanatory variables at  $\mathbf{s}_i$ , regardless of the information available at surrounding sample points  $\mathbf{s}_j$ . An assumption of the ordinary least squares regression analysis is that the residuals are independent of each other; the value of one error is not affected by the value of another error. In spatial regression however, we attempt to derive better estimates for the coefficients by using observations at neighboring sites. Spatial regression is specified similarly to the usual multiple regression model, with the exception that the residuals are modeled as functions of the surrounding residuals (Bailey and Gatrell 1995). To estimate the spatial regression model (Rogerson 2000), regressions of  $\hat{y}$  on  $\hat{x}$  are tried for a variety of  $\rho$  and  $\beta$  values beginning at zero, using a grid search method (Figure 7). The best values of  $\beta$  and  $\rho$  are those minimizing the standard error of the estimate  $s_e$ .

$$\hat{y}(\mathbf{s}_i) = y(\mathbf{s}_i) - \rho \sum_{\mathbf{s}_j=1}^m w_{\mathbf{s}_j} y(\mathbf{s}_j) \quad (16)$$

$$\hat{x}(\mathbf{s}_i) = x(\mathbf{s}_i) - \rho \sum_{\mathbf{s}_j=1}^m w_{\mathbf{s}_j} x(\mathbf{s}_j) \quad (17)$$

**Calibration of the parameters.** To account for information provided by nearby samples, it is important to weight these samples according to their distance to initial observation  $i$ , as defined in Equation 15. However, observations far away may not be significant. Lowest  $s_e$ -values for spatial regression were obtained for a neighborhood radius of 5000 meters.

Consequently we reformulated Equation 15 as follows:

$$w_{\mathbf{s}_j} = \begin{cases} e^{-\beta d_{\mathbf{s}_{ij}}^2} & \text{if } d_{\mathbf{s}_{ij}} \leq 5000m \\ 0 & \text{otherwise} \end{cases}$$

INSERT FIGURE 7 ABOUT HERE

**Analysis of the regression coefficients.** We have summarized the results of the spatial regression in Table 3. All variables are significant, except the slope coefficient  $b_s$ , at the 5%-level. The interpretation of the coefficients is similar to the multiple regression case. However, the significance level ( $t$ -value) of the six explanatory variables is slightly greater than in Table 2, as is the  $R^2$ -value.

INSERT TABLE 3 ABOUT HERE

#### 4.4 Comparing the different techniques

Both multiple and spatial regression give global estimates of the slope coefficients (i.e. for the entire study area), which allows comparison of their significance levels. Spatial regression performs slightly better than multiple regression in terms of the standard error of the estimate  $s_e$ .

**Cross-validation.** Cross validation gives us some idea of how well the regression model predicts the values at unknown locations. In this case, it withholds one data sample, and then makes a prediction for the same data location. The predicted and observed values at that location are compared. Cross validation (CV) takes the form:

$$CV = \sum_{i=1}^n \left( \hat{y}(\mathbf{s}_i) - \hat{y}(\neq \mathbf{s}_i)(\beta) \right)^2 \quad (18)$$

where  $\hat{y}(\neq \mathbf{s}_i)(\beta)$  denotes the fitted value of  $\hat{y}(\mathbf{s}_i)$  with the observation for point  $\mathbf{s}_i$  omitted from the calibration process. As  $\beta$  becomes very large, the model is calibrated only at samples in the very close neighborhood of  $\mathbf{s}_i$ . We have computed the cross validation results, taking the back transform of  $\ln(-RSSI)$ . We found a CV-value of 49,202dB for the

spatial regression method and a value of 50,809dB for the multiple regression case. Spatial regression improves on ordinary least squares, although by relatively small amounts.

INSERT FIGURE 8 ABOUT HERE

**Spatial pattern in the residuals.** Figure 8 on the left shows the interpolated residuals for the the spatial regression. In comparison with Figure 6, there has been a decrease in the range of the residuals and a slight decrease of the positive residuals around the cell tower. Figure 8 on the right displays the Moran's  $I$  values for the two regression methods for a range of  $\beta$  values. Spatial regression performs better than multiple regression. Moran's  $I$  curves increase up to a maximum  $\beta$  value of approximately .0005 and then decrease. Spatial regression shows significantly lower Moran's  $I$  for that particular value of  $\beta$ , which is an improvement on multiple regression.

#### 4.5 Reducing the residuals

Spatial regression performs better than regular ordinary least squares in reducing the size of residuals, and also their spatial pattern. Nevertheless, a great quantity of residuals is not explained since we have a limited number of explanatory variables. Incorporating the presence of buildings, cell tower transmitting power, antenna direction and vegetation may reduce the residuals and improve the spatial model, yet such information is difficult to collect. In the next two paragraphs, we address some issues pertaining to the effect of vegetation on cell phone signal strength.

**The presence of vegetation.** The signal generated by a cell tower tends to be attenuated by the presence of vegetation. This signal attenuation is due partly to the scattering and absorption from branches and foliage of trees. The signal is not completely lost however, due to multipath reflection that contributes to reduce the total attenuation (Golhirsh and Vogel 1998). The parameters affecting signal attenuation are (a) the signal frequency, (b) the distance between the base station and the receiver that goes through the forest canopy, (c) the type of trees, (d) the dimensions and shapes of the trees, (e) the angular and density distribution of the leaves, (f) the foliage density and (g) the height of the cell tower antenna

(Bertoni 2000, Tamir 1977). The attenuation or path loss coefficient is expressed as the decrease of decibels by meter through the canopy. As suggested by Comparetto (1993), very high frequencies (e.g., 40 GHz) show a much greater attenuation coefficient than very low frequencies (e.g. 200 MHz) when the vegetation is in full foliage. The recommended frequency scaling formulation pertaining to single trees in full foliage condition between 870 MHz and 1.6 GHz is given by Ulaby et al. (1990):

$$A(f_2) = A(f_1) \times \sqrt{\frac{f_2}{f_1}} \quad (19)$$

where  $A$  denotes the attenuation coefficient. For frequencies of 870MHz, the average path loss coefficient for vegetation in full foliage condition is equal to 1.3dB/m, with a maximum of 3.2dB/m for a Norway maple to a minimum of 0.6dB/m for a pine oak. With the 5dB rounding effect of the measuring device, it becomes difficult to detect minor changes in RSSI due to vegetation effects. However, repeated measurements behind large forested areas may indicate lower RSSI values with greater foliage.

**The impact of vegetation in the neighborhood of a cell tower.** Two empirical tests were performed to evaluate how the presence of vegetation impacts RSSI. Figure 9 illustrates the test area which is relatively flat, disregarding effects of altitude and slope. Repeated RSSI measurements were collected in the close vicinity of a cell tower in the Town of Alden at an approximate distance of 700m - 1200m. A total of 45 locations were sampled, some behind a single tree, some behind a group of trees and some behind a dense forest. For each point, we report information on (a) whether the cell tower was visible, (b) the vegetation between the cell tower and the measured point, and (c) the distance to the cell tower. Testing points have been classified into two categories of vegetation. The first category (group 1) includes sample points that have a path length through the canopy smaller than approximately 5m and the second category (group 2) includes sample points having a path length greater than 5m. For category 1, the predicted RSSI value at a distance of 1000m is equal to -55.174 dB. For the second group, the predicted value drops to -62.042dB which is almost 7dB lower than the estimate for group 1. This attenuation effect can be reduced when the cell tower antenna is very high.

**Seasonal impact of vegetation.** Also of interest is to evaluate the effect of foliage (and therefore only the leaves and tree branches) on RSSI. Goldhirsh and Vogel (1998) advance the idea that vegetation attenuation is also influenced by the elevation angle between the receiving and emitting antenna. The lower the angle, the longer the path length through the tree canopy and the greater the attenuation. With the 5dB rounding effect of the measuring device, it becomes extremely difficult to detect change in RSSI due to foliage. Seasonal growth effects of vegetation are less important than vegetation itself, yet should not be disregarded. Our experiments show that the RSSI value can drop by 5dB in areas of dense vegetation in a 50 day period of time because of the emerging vegetation.

INSERT FIGURE 9 ABOUT HERE

## 5 Conclusion

In this paper, we have described how RSSI values were collected in Erie County. Secondly, we have estimated RSSI values at unknown locations using kriging. The resulting map shows weaker RSSI in the southern areas of the county, which corresponds to areas of substantial elevation and slope variability. We next evaluated the influence of slope and elevation on varying RSSI values. We mapped the residuals based on the results from the multiple regression. Very strong positive residual areas denote underestimation of RSSI value, corresponding to locations of cell towers. We have used cell tower locations and Digital Elevation Models to perform visibility analysis, thereby accounting for terrain obstructions between a cell tower and an RSSI location point. Knowing cell tower locations allowed us to compute the distance for each RSSI data point to its nearest cell tower. We have formulated a regression model incorporating these variables, as well as altitude and slope, and we have tested our model in a typical study area.

In the multiple regression, all of the variables with the exception of slope and the interaction term distance to the cell tower have a positive impact on RSSI. Both an increasing slope and/or an increasing distance to the nearest cell tower have a negative influence on the received signal strength. Adding the distance $\times$ range interaction term and the visibility factor had a significant effect on the  $R^2$ . Because multiple regression does not account for the

spatial structure of the explanatory variables, we have performed a spatial regression. The values of the coefficients are fairly close to those from ordinary least squares, but greater  $t$ - and  $R^2$ -values are obtained. We have compared the residuals from the two regression methods. In general, spatial regression performs better regular OLS: it has less spatial pattern in the residuals for small values of  $\beta$  and lower cross-validation values. Spatial regression revealed lower Moran's  $I$  at a  $\beta$ -value of .0005. The results from the regression analysis are beneficial because they allow us to predict the behavior of the signal strength where no measurements have been taken. This in turn allows to evaluate the impact of different geographic factors on the effectiveness of ACN devices. ACN will perform poorly in areas exhibiting sudden changes of topography. The promises of the ACN technology will suffer in areas that are not visible from cell towers, and in areas characterized by low elevation values. We noticed that rapid decrease in altitude reduces the RSSI value. This decrease is attenuated if the observer is still within the predefined range of the cell tower. Moreover, increasing slope values will delay call completion. The distance to the cell tower is the most significant variable, especially in the strict vicinity of the cell tower. As the distance to the nearest cell tower increases, the reliability of the device is increasingly subject to nearby obstructions. This information is also significant from a cell tower provider perspective since it could assist in optimally locating new base stations.

**Future work.** In this paper, we have suggested two different alternatives to incorporate topography-related factors into the prediction of cell phone signal strength. The maps of the residuals have shown that there are still important discrepancies between predicting the signal from covariates on the one hand and interpolating the signal from initial samples on the other. Further research is needed in the analysis of residuals: their range is still large: accounting for other variables such as buildings will reduce the range and produce a better regression model. In a second research stage, more data on vegetation should be collected over the entire county. The extensive use of aerial photography or remote sensing images provides additional information on the presence and type of vegetation at RSSI measurement points. This information could then be integrated in the regression model and would improve upon our predictions.

After a crash occurs, the vehicle is likely to be stationary. We have collected RSSI data at a fixed interval using a vehicle in motion. Depending on the speed of the vehicle, the emergency call can be assigned to a micro or macro cell depending upon its speed (Narasimhan and Cox 2001). Work is being undertaken along those lines, especially to evaluate the allocation of base station and channel assignment in the presence of hand-offs (Erdemir *et al.* 2005).

Both multiple and spatial regression estimate the coefficients of the explanatory variables globally, i.e. for the entire study area. It may also be useful to use geographically weighted regression (GWR) that provides locally linear regression estimates at every point  $i$ , using distance weighted samples (Brundson *et al.* 1996, 1999 and Fotheringham *et al.* 1998).

$$\hat{y}(\mathbf{s}_i) = a_{0\mathbf{s}_i} + \sum_{k=1}^p b_{k\mathbf{s}_i} x_{k\mathbf{s}_i} + \varepsilon_{\mathbf{s}_i} \quad (20)$$

For each location  $\mathbf{s}_i$ , we have a set of neighboring values  $\mathbf{s}_j$  that we use to estimate  $\hat{y}(\mathbf{s}_i)$  while omitting observation  $\mathbf{s}_i$ . The term  $b_{k\mathbf{s}_i}$  represents the value of the  $k^{th}$  parameter at location  $\mathbf{s}_i$ . The parameter  $p$  defines the number of explanatory variables. We need to estimate  $b_{k\mathbf{s}_i}$  on observations taken at sample points close to  $\mathbf{s}_i$ . GWR allows one to produce maps of these coefficients for the study area, allowing estimation of the influence of each explanatory variable locally.

## 6 Acknowledgements

We acknowledge the support of a grant from the Center for Transportation Injury Research at General Dynamics. Dr. Irene Casas (Geography Department, University at Buffalo) provided initial programming assistance.

## References

- Akella M., Bang C., Beutner R., Delmelle E., Wilson G., Blatt A., Batta R. and P. Rogerson** (2003). Evaluating the reliability of automated collision notification systems. *Accident Analysis and Prevention* 35: 349-360.
- Akella M., Batta R., Delmelle E., Rogerson P. and A. Blatt.** Using Geostatistical Methods to Decide an Additional Facility Location (*Submitted to Operations Research*)
- Bailey T. and A. Gatrell** (1995). *Interactive spatial data analysis*. Longman, 413p.
- Bertoni H.L.** (2000). Shadow Fading and the Effects of Terrain and Trees. *In: Bertoni H.L. Radio Propagation for Modern Wireless Systems*. Prentice Hall, Upper Saddle River, New Jersey, chapter 7, pp. 181-216.
- Blow O., Magliore L., Claridge J.A., Butler K. and J.S. Young** (1999). The golden hour and the silver day: Detection and correction of occult hypoperfusion within 24 hours improves outcome from major trauma. *Journal of Trauma, Injury, Infection, and Critical Care* 7: 964-969.
- Bose R.** (2001). A smart technique for determining base-station locations in an urban environment. *IEEE Transactions on Vehicular Technology* 50: 43-47.
- Brunsdon C., Fotheringham A.S. and M.E. Charlton** (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28: 281-298.
- Brunsdon C., Fotheringham A.S. and M.E. Charlton** (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science* 39: 497-524.
- Burrough P.A. and R.A. McDonnell** (1998). *Principles of Geographical Information Systems: Spatial Information Systems and Geostatistics*. Oxford University Press, New-York, pp. 147-166.
- Champion H.R., Augenstein J.S., Cushing B., Digges K.H., Hunt R., Larkin R., Malliaris A.C., Sacco W.J. and J.H. Siegel** (1998). Automatic Crash Notification: the Public Safety Component of the Intelligent Transportation System. *AirMed*, March/April, 36p.
- Clark D.E. and B.M. Cushing** (2002). Predicted effect of automatic crash notification on traffic mortality. *Accident Analysis and Prevention* 34: 507-513.
- Clarke K.** (1995) *Analytical and computer cartography*. Prentice-Hall. Second edition. 334p.

- Comparetto G.** (1993). The Impact of Dust and Foliage on Signal Attenuation in the Millimeter Wave Regime. *Journal of Space Communication* 11: 13-20.
- Cressie N.** (1991). *Statistics for spatial data*. Wiley, New York, USA, 900p.
- Cornell University Geospatial Data Information Repository (CUGIR)**  
Available on-line at: <http://cugir.mannlib.cornell.edu/>
- Erdemir E.T., Batta R., Spielman S., Rogerson P.A. and A. Blatt** (2005). Modeling Spatial Uncertainty in Facility Location Problems with Application to Aeromedical Emergency Services and Cellular Networks. Anticipated submission to *European Journal of Operations Research*.
- Evanco W.M.** (1999). The potential impact of rural mayday systems on vehicular crash fatalities. *Accident Analysis and Prevention* 3: 455-462.
- Federal Communications Commission (FCC)** (2001). Licensing Database Extracts - Cellular. Available on line at: <http://wireless.fcc.gov/geographic/data/db/cellular.zip>
- Fotheringham A.S., Charlton M.E. and C. Brundson** (1998). Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. *Environment and Planning A* 30: 1905-1927.
- Goldhirsh J. and W.J. Vogel** (1998). *Handbook of Propagation Effects for Vehicular and Mobile Satellite Systems*. The John Hopkins University (Applied Physics Laboratory) and the University of Texas at Austin (Electrical Engineering Research Laboratory).
- Jacobs L.M., Sinclair A., Beiser A. and R.B. D'Agostino** (1984). Prehospital Advanced Life Support: Benefits in Trauma. *Journal of Trauma* 24: 8-13.
- Krishnamachari B.** (1999). Global optimization in the design of mobile communication systems. *Master's thesis*, Electrical Engineering, Cornell University.
- Lerner E.B. and R.M. Moscati** (2001). The golden hour: Scientific fact or medical urban legend? *Academic Emergency Medicine* 8: 758-760
- Macario R.C.V.** (1997). *Cellular radio: Principles and design*. McGraw-Hill Series on Telecommunications, 2nd edition, 276p.
- Mitas L. and H. Mitasova** (1999). Spatial interpolation. In: Longley P.A., Goodchild M.F., Maguire D.J. and D.W. Rhind *Geographical Information Systems* (1), chapter 34. (Second edition), John Wiley and Sons.

- Moran P.A.P.** (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B* 10: 245-251.
- Moran P.A.P.** (1950). Notes on continuous phenomena. *Biometrika* 37: 17-23.
- Narasimhan R. and D. C. Cox** (2001). Estimation of mobile speed and average received power in wireless systems using best basis methods. *IEEE Transactions on Communications* 49: 2172-2183.
- NHTSA** (2004). Traffic Safety Facts 2003. Overview, Department of Transportation.
- Pepe P.E., Wyatt C.H., Bickell W.H., Bailey M.L. and K.L. Mattox** (1987). The Relationship between Total Prehospital Time and Outcome in Hypotensive Victims of Penetrating Injuries. *Annals of Emergency Medicine* 16: 293-297.
- Petri R.W., Dyer A. and J. Lumpkin** (1995). The Effect of Prehospital Transport Time on the Mortality from Traumatic Injury. *Prehospital and Disaster Medicine* 10: 24-29.
- Rogerson P.** (2000) *Statistical methods for geography*. Sage Publication, 236p.
- Sampalis J.S., Lavoie A., Williams J.I., Mulder D.S., and Kalina** (1993). Impact of On-Site Care on Survival in Severely Injured Patients. *Journal of Trauma* 34: 252-261.
- Tamir T.** (1967). On Radio-Wave Propagation in Forested Environments. *IEEE Transactions on Antennas and Propagation* 15: 806-817.
- Ulaby F.T., Whitt M.W. and M.C. Dobson** (1990). Measuring the Propagation Properties of A Forest Canopy Using a Polarimetric Scatterometer. *IEEE Transactions on Antennas and Propagation* 38: 251-258.
- Walton S. and E. Meyer** (2000). *Proceedings of RATTTS conference on interpreting cellular coverage data*, Brandon, Missouri, 13-16 August 2000.

## LIST OF TABLES

- Table 1: Erie county RSSI data sample.
- Table 2: Coefficients of the explanatory variables for the multiple regression.
- Table 3 Coefficients of the explanatory variables for the spatial regression.

## LIST OF FIGURES

- Figure 1: Erie county, Western NY (*a*). RSSI sample points (*b*) and their interpolated values (*c*).
- Figure 2: Calculation of slope degree from altitude values.
- Figure 3: Altitude in Erie county (*a*) and its corresponding slope values (*b*). Map (*c*) represents the residuals from the multiple regression, including slope and altitude.
- Figure 4: Vegetation and reflection occurring in valleys attenuate the signal strength in Figure *a*. Figure *b* is a simplification of the model depicted in Figure *a* using a line-of-sight approach.
- Figure 5: The selected study area around the cell tower. Altitude in (*a*), slope values in (*b*), visibility in (*c*) and signal strength in (*d*).
- Figure 6: The interpolated residuals from the multiple regression with explanatory variables slope  $x_s$  and altitude  $x_a$  only on the left, and with added variables ( $x_d \times x_r$ ),  $x_r$  and  $x_v$  on the right.
- Figure 7: Standard error of the estimate. Optimal values found are  $\beta = .003298$  and  $\rho = .02566$ .
- Figure 8 Interpolated residuals for the spatial regression and variation of Moran's  $I$  value with varying  $\beta$ .
- Figure 9: Testing locations for the effect of vegetation, Town of Alden (Erie county).

# TABLES

ID	Date (mm/dd/yyyy)	Latitude ( $\phi$ )	Longitude ( $\lambda$ )	RSSI (dB)	Speed ( $\frac{ft}{sec}$ )
1	11/23/99	42.89	-78.75	-68	36.87
2	11/23/99	42.91	-78.73	-73	32.24
...	...	...	...	...	...
6863	4/24/00	42.93	-78.75	-58	48.19
6864	4/24/00	42.93	-78.76	-83	107.68

Table 1: Erie county RSSI data sample.

	$x_s, x_a$			$x_s, x_a$ $x_r x_d$			$x_s, x_a$ $x_r x_d, x_r, x_v$		
$R^2$	.095			.154			.282		
Model	$b$	$t$	$p$	$b$	$t$	$p$	$b$	$t$	$p$
$\ln(-b_0)$	4.723	41.029	.000	4.543	37.720	.000	4.615	40.858	.000
$b_a$	-.00117	-3.781	.000	-.00095	-3.119	.002	-.000608	-2.077	.039
$b_s$	.01130	2.182	.030	.01217	2.424	.016	.003138	.631	.529
$b_{rd}$	-	-	-	.0363	3.967	.000	.07334	6.343	.000
$b_r$	-	-	-	-	-	-	-.241	-3.814	.000
$b_v$	-	-	-	-	-	-	-.0846	-2.592	.010

Table 2: Coefficients of the explanatory variables for the multiple regression.

	<b>spatial regression</b>		
$R^2$	.285		
	coefficient	$t$	$p$
$\ln(-b_0)$	4.617	40.882	.000
$b_a$	-.000616	-2.100	.037
$b_s$	.003252	.654	.514
$b_r$	-.244	-3.877	.000
$b_{rd}$	.07424	6.420	.000
$b_v$	-.0844	-2.571	.011

Table 3: Coefficients of the explanatory variables for the spatial regression.

# FIGURES

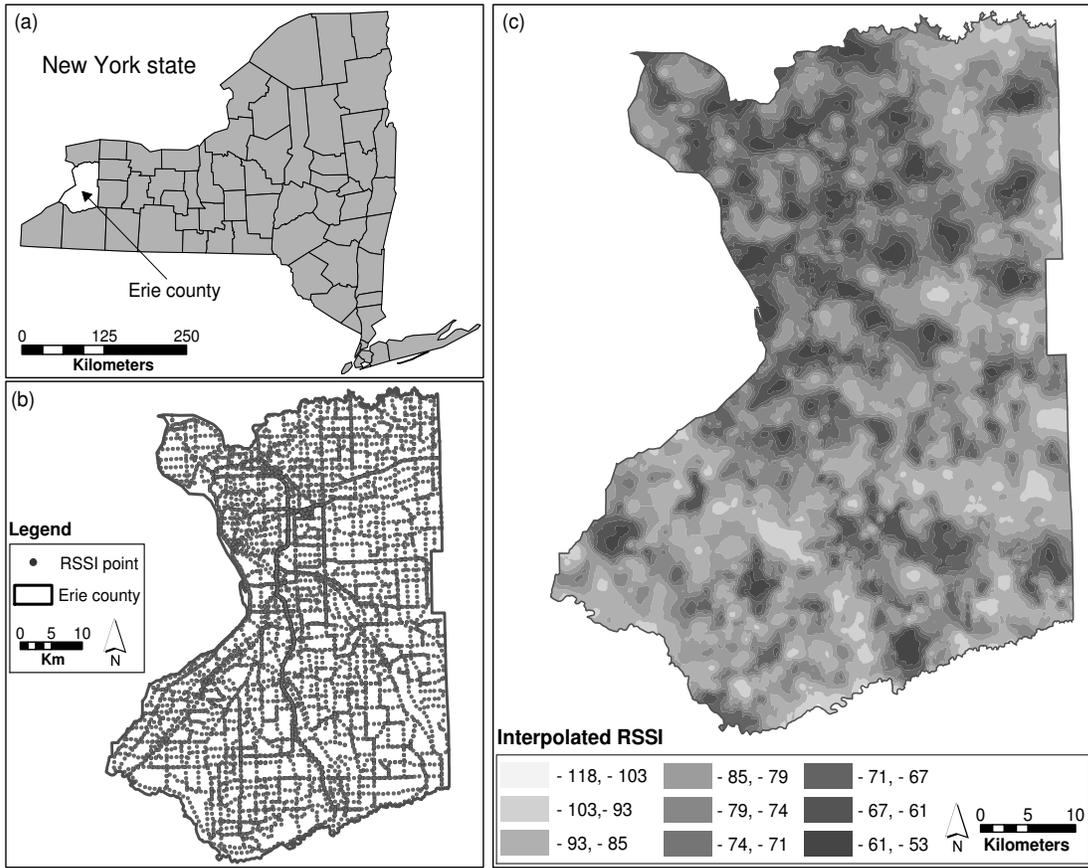
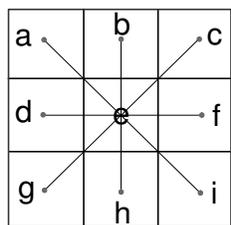


Figure 1: Erie county, Western NY (a). RSSI sample points (b) and their interpolated values (c).



$$\frac{\partial z}{\partial x} = \frac{[(a + 2d + g) - (c + 2f + i)]}{8 \times \text{mesh spacing } x\text{-axis}}$$

$$\frac{\partial z}{\partial y} = \frac{[(a + 2b + c) - (g + 2h + i)]}{8 \times \text{mesh spacing } y\text{-axis}}$$

Figure 2: Calculation of slope degree from altitude values

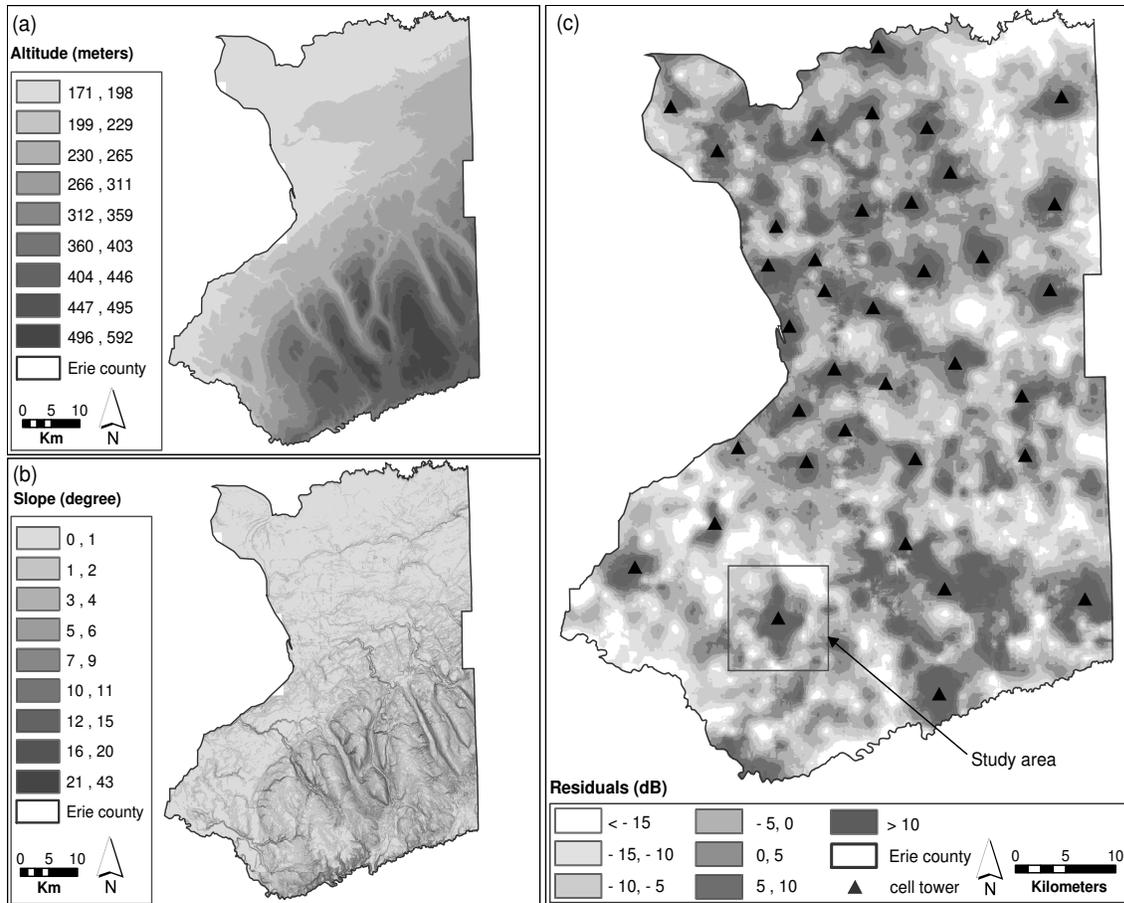


Figure 3: Altitude in Erie county (a) and its corresponding slope values (b). Map (c) represents the residuals from the multiple regression, including slope and altitude.

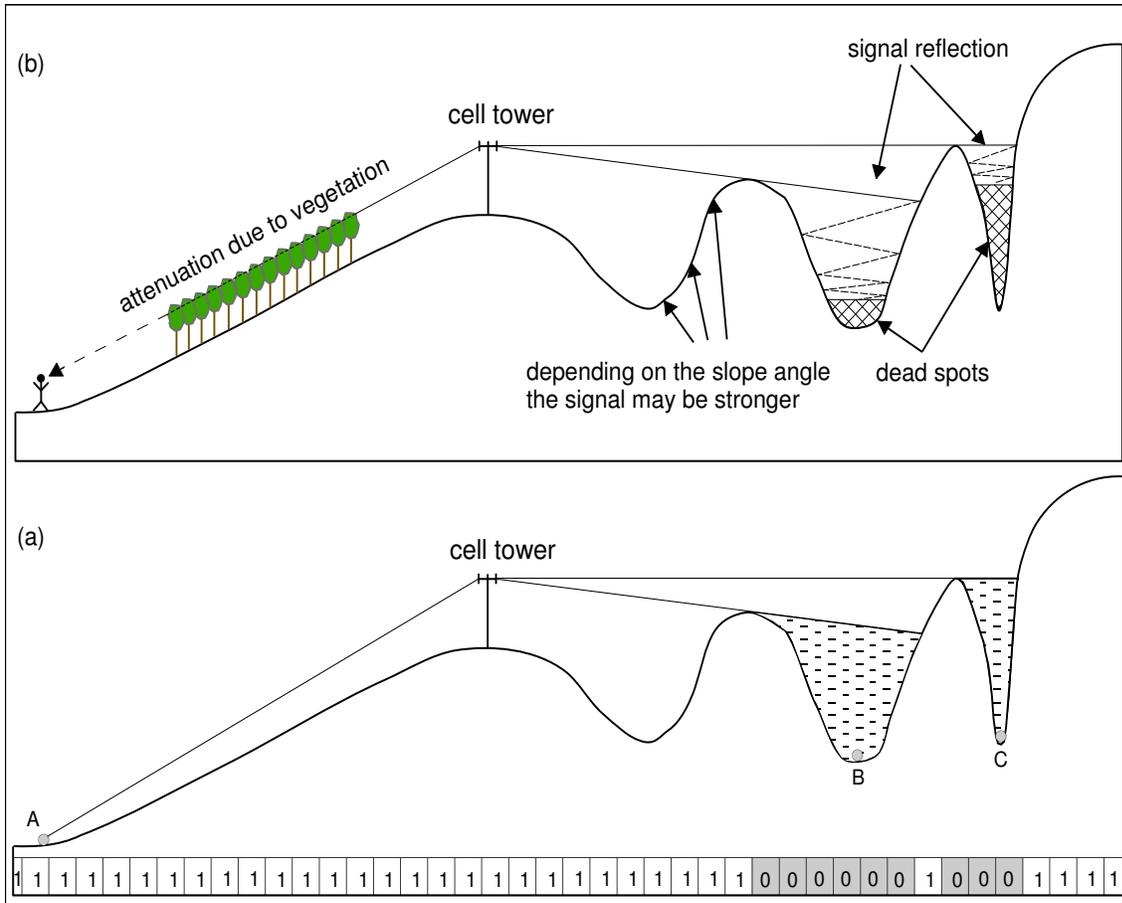


Figure 4: Vegetation and reflection occurring in valleys attenuate the signal strength in Figure *b*. Figure *a* is a simplification of the model depicted in Figure *b* using a line-of-sight approach.

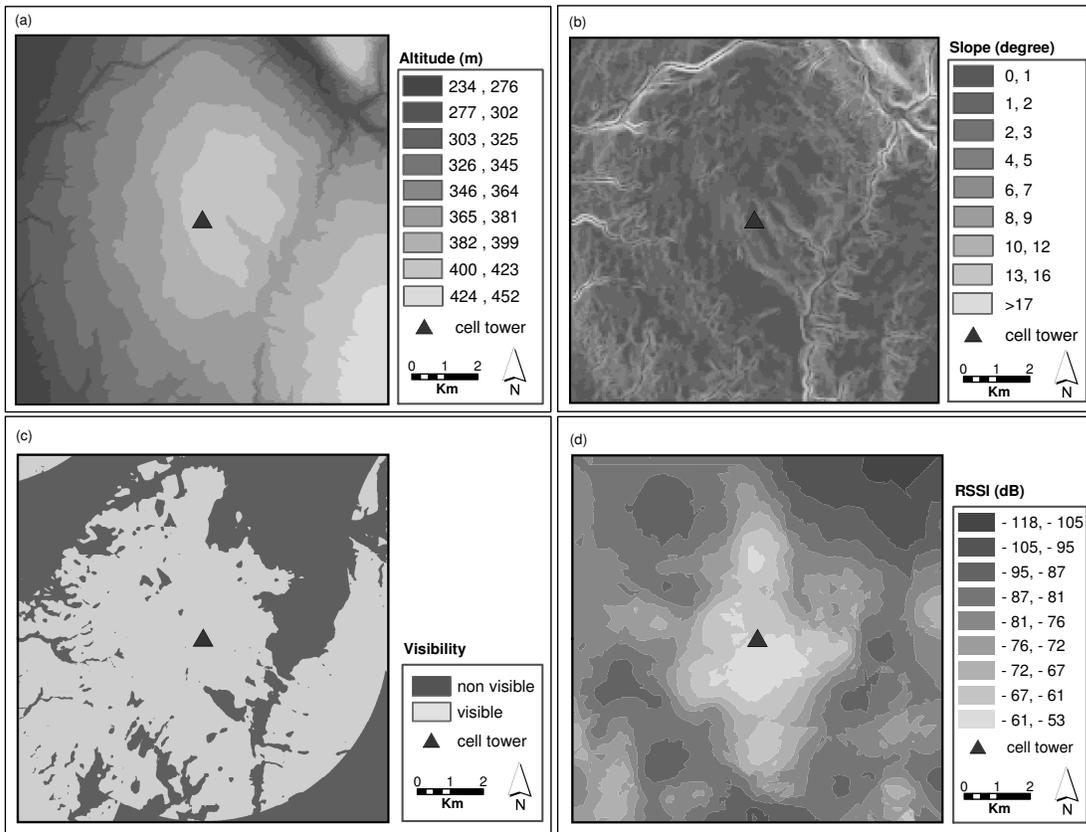


Figure 5: The selected study area around the cell tower. Altitude in (a), slope values in (b), visibility in (c) and signal strength in (d).

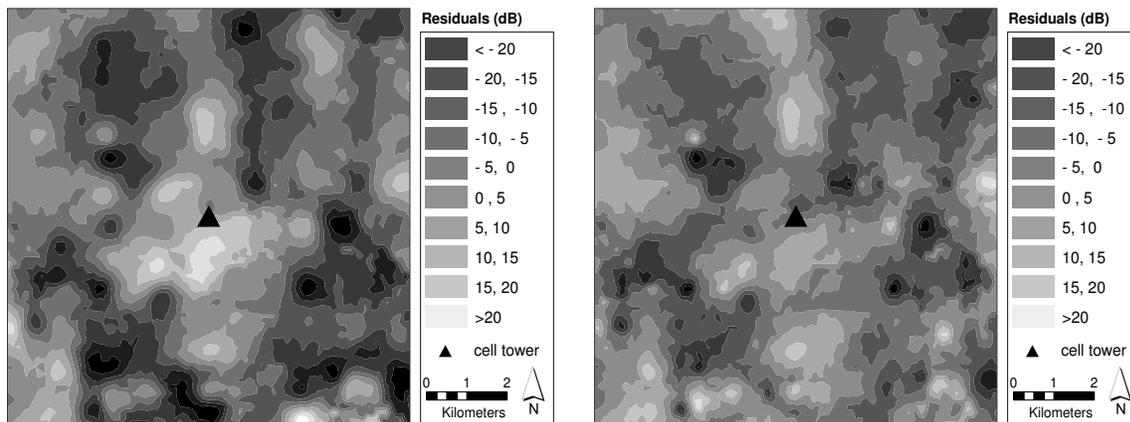


Figure 6: The interpolated residuals from the multiple regression with explanatory variables slope  $x_s$  and altitude  $x_a$  only to the left, and with added variables  $(x_d \times x_r)$ ,  $x_r$  and  $x_v$  to the right.

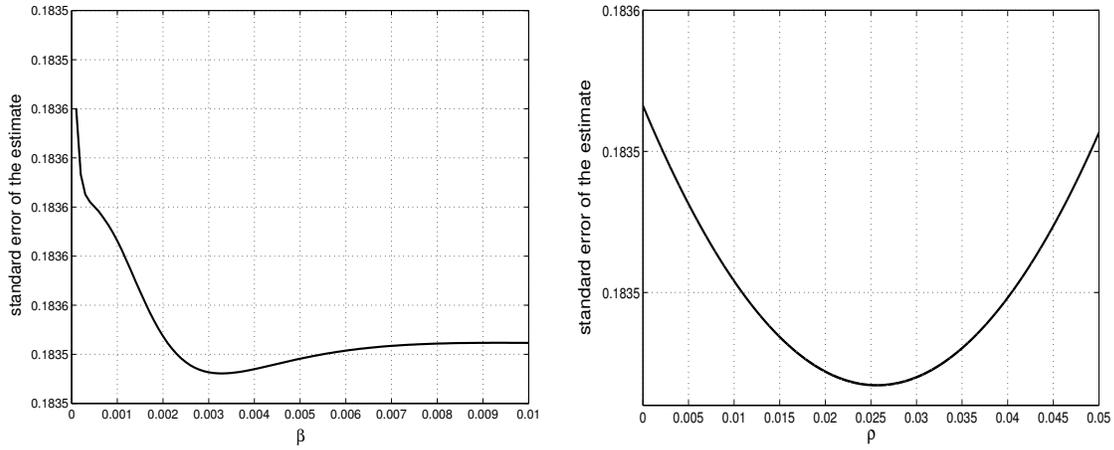


Figure 7: Standard error of the estimate. Optimal values found are  $\beta = .003298$  and  $\rho = .02566$ .

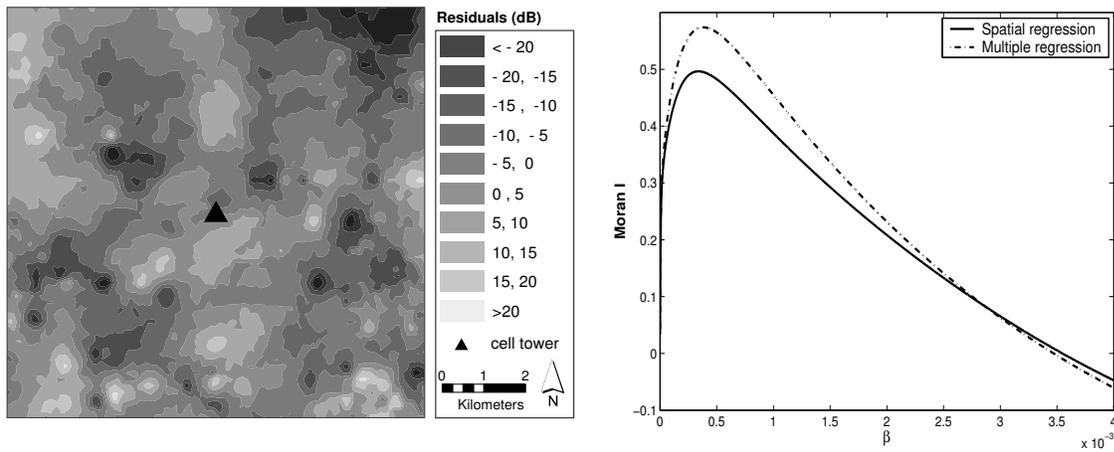
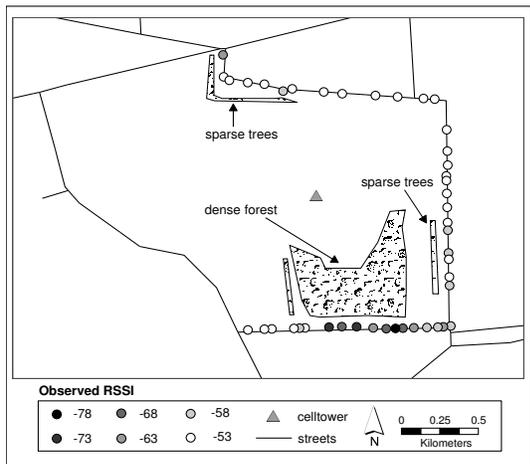


Figure 8: Interpolated residuals for the spatial regression and variation of Moran's  $I$  value with varying  $\beta$ .



$$\hat{y}(s_i)_{<5m} = -44.174 \text{ dB} - .11 x_d(s_i)$$

$$\hat{y}(s_i)_{>5m} = -38.342 \text{ dB} - .0237 x_d(s_i)$$

Figure 9: Testing locations for the effect of vegetation, Town of Alden (Erie county).