Spatial Pattern Analysis of Settlement Locations Using Logistic Regression

[1]Pariwate Varnakovida, [2]Arwa Mohamed, [2]Suwiwat Witchakool

[1]Department of Geography
Center for Global Change and Earth Observations
Michigan State University

[2]Department of Statistics
Michigan State University

Abstract

The interactions between people and the environment are complex and dynamic, making direct, one-to-one relationships between two specific phenomena very rare, if not nonexistent. The complexity of these relationships means that linking people with change over a wide area has become one of the most difficult tasks in understanding land use and land cover change.

Though there are many different types of land in the tropics that people are currently working to develop, the priority for selecting areas for development and intensive human use varies between settings because spatially heterogeneous variables make some locations more desirable or susceptible to change than others; accessibility, certain biophysical, and social characteristics encourage or discourage intensive human use. The biophysical factors which contribute to an area undergoing human induced LULCC are legion, but among the most important are:

- Elevation, due to the close relationship between elevation and ecological community type.
- Topography, which directly influences the ease of use and access to an area, as well as vegetation cover.
- Presence of water resources, because human settlement and development of an area significantly depends on easy access to water.

In this paper we plan to illustrate how these characteristics come into play in the area surrounding the district of Nang Rong, Thailand, the site of ongoing LULCC, over the past 50 years. However, it is first worthwhile to overview the basic spatial factors and determinants of these changes. We applied multiple spatial logistic regressions for this research for a preliminary study. The models could easily be applied to any rural setting in Thailand, and many other locations

Study site

Thailand is located in Southeast Asia and bordered by Burma in the West and North, Laos and Cambodia to the North and East, and Malaysia to the South. The case study area is the Nang Rong district, which is an agriculturally marginalized 1,300 km^2 environment in the province of Burirum, on the Korat Plateau. The physical environment in Nang Rong is characterized by relatively infertile soils, poor drainage, and inconsistent precipitation level. Human settlement in Nang Rong takes the form of nuclear households clustered in villages, the number of which has increased from 83 to 352 between 1950 and 2000 due to socio-economic and environment-landscape (Crawford, 2000) change in the district. The rapid change over the past 50 years has substantially influenced human-environment relationships, such as deforestation and migration.

Research problem:

How do biophysical characteristics (elevation and stream), road network, and previous settlement pattern relate to village settlement pattern in Nang Rong, Thailand?

Available Data:

The case of Nang Rong offers an opportunity to utilize a rich collection of interlinked data sets, other information, human-environment theory and concepts, and socio-economic ideas put together through many projects, some of which were undertaken at Mahidol University in Bangkok, Thailand, and others undertaken at the University of North Carolina at Chapel Hill (Permission for the use of these data sets and information has been obtained). Base data for this project include geographic and social coverages derived from social surveys, Thai government maps, aerial photography, and satellite imagery. Roads and streams were digitized from the aerial photo mosaic.

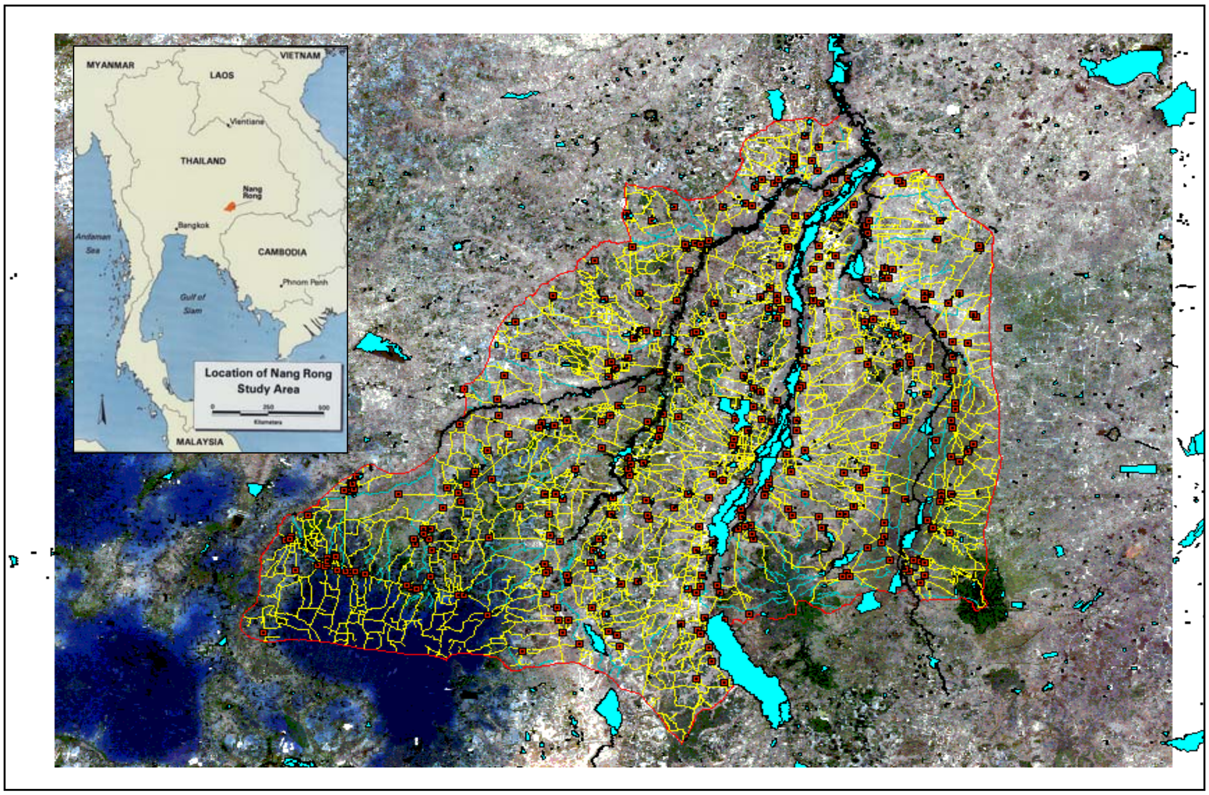


Figure 1: The study site, village locations and road network, Nang Rong, Thailand and

The following GIS layers were use:

- Location of the village centers (red color)
- Road network
- Streams, rivers, and water bodies
- Digital Elevation Model (DEM)

Hypotheses:

- ❖ The center of the village is related to the following:
    - Road density
    - Slope
    - Distance from the nearest or second nearest existing village center (dist1, dist2).

- Distance from river
- Distance from water body
- Low elevation

Each of the listed hypotheses will be tested step by step in a multiple logistic regression model.

The following Kernel shows village intensities over the years from 0-1900, 1901-1951, and 1951 to 2000. As it shows that the center of the villages were concentrated along with the existing streams and the rivers. Before 1901, the center of villages seemed to be along the main rivers, Lum Nang Rong and Lum Praey Mas. From 1901 to 1950, villages tended to increase along the Lum Nang Rong. Next, from 1951 to 2000, village increased all over Nang Rong area but concentrated along the rivers and streams.

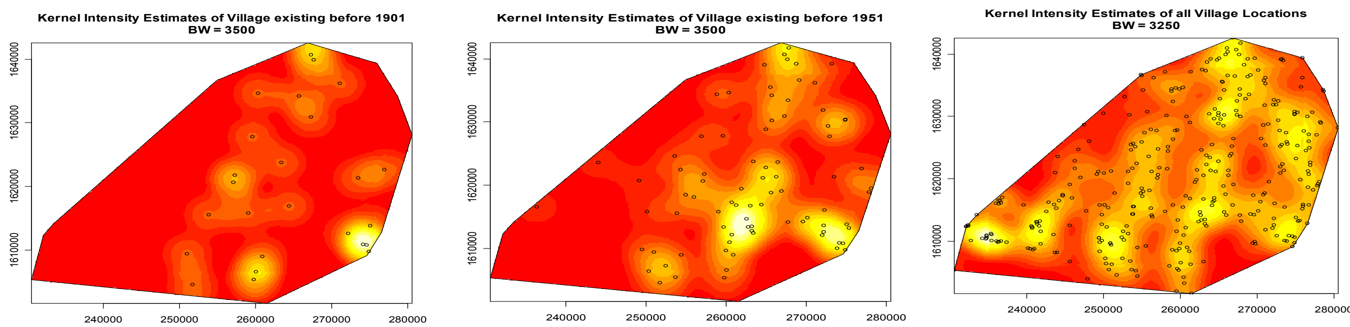


Figure 2: Kernal intensity of villages in different time period

Techniques and methods:

A. Spatial Processing using ArcGIS and ArcInfo (Figures are in appendices session)

- **Road development**

  The road density was calculated using Hawth's Analysis Tools, an extension for ESRI's ArcGIS. LINEDENSITY algorithm was used to calculate the density of linear features in the neighborhood of each output grid cell. Density is calculated in units of length per unit of area. The default unit of area is one square map unit. Thus, the units of line density are mapunits per square mapunit, or 1/mapunits. For map units of meters, density would thus be meters (of line length) per square meter (Silverman, 1986).

- **Streams and water bodies**

  Euclidean distance was used to create a distance surface. For each cell, the distance is calculated to each source cell by calculating the hypotenuse with the x-max and y-max as the other two legs of the triangle. This calculation derives the true Euclidean, not cell, distance. The shortest distance to a source is determined and if it is less than the specified maximum distance, the value is assigned to the cell location on the output raster.

- **Existing settlement pattern**

  Distance between points tools from Hawth's Analysis Tools were used to create various tables representing nearest and second nearest distances, and summary statistics of distances between random-to-village and village-to-village.

- **Elevation**

  Elevation was used to assess the local physical structure of the landscape, and to generate slope convexity/concavity, and terrace structure.

B. Multiple Spatial Logistic regressions: Processing in R

Our typical approach to measure spread of settlement locations is to use radial buffers around the village centroids. Rindfuss et al. (2002) illustrated that 2.0 km radial buffer were used to approximate the village size. The argument is that ambiguously titled forest cover converted into crop land by young adults and that they would be more likely to stay in their home village, and 2.0 km is approximately the distance that someone could travel. 351 village points were used and buffered. 352 were randomly generated within the study area to incorporate non-existing village data into our model. The first random was generated after the existing village is buffered, while the second random were conditioned not to be closer than 200 meters from each other. Assigning the same buffer to every village in every time point will miss the variation. However, this project is only an initial experiment that tries to explore the pattern of village settlement.

Multiple spatial logistic regressions were used to solve the problem given the limitation in the dependent variable (discrete not continuous). Logistic regression uses transformation technique to transform dependent variable, which is binary variable, to be continuous random variable. This analysis attempts to identify whether village occurred or not. In addition, this analysis attempts to predict a dependent variable on the basis of independent variables and to determine the percent of variance in the dependent variable explained by the independents. A binary variable (coded 0,1) were used for the dependent variable. All the area within the buffer was assigned to 1 and the other 351 (that will be randomly located) outside the buffer were assigned to zero. We compared models with the interaction and without the interaction among independent variables. The first nearest village and the second nearest village distances were examined for each with/without interactions. The experiment was replicated in the second randomization to validate the results. The goal of the regressions is to identify the factors that contribute significantly (at 0.05 level) to the explanation of the variability in the village settlement distribution. In this way, it is possible to distinguish which factors have relevance for the spatial pattern to increase the settlement areas in Nang Rong. Residuals from the models were investigated and reported using residual for coefficient surface model.

The spatial multiple logistic regression model were as follows:

$$\ln\left(\frac{p}{1-p}\right) = a + \sum_{i=1}^{n} b_i * \vec{x}_i$$

where, $\vec{x}_i$ ***'s:*** a vector of 703 observations of all variables which may be effects to village settlement

ln : the natural logarithm, log(exp), where exp = 2.71828.

p: the probability that the event Y (village) occurs, p(Y = 1).

p/(1-p): the "odds ratio."

ln[p/(1-p)]: the log odds ratio, or "logit."

Odds (of having a village) = $\exp\left\{a + \sum_{i=1}^{n} b_i * \vec{x}_i\right\}$

Then use the stepwise (backward) to remove all non-significant variables to get the following model:

$$\ln\left(\frac{p}{1-p}\right) = a + \sum_{i=1}^{n} b_i * Factor_i$$ for those *i's* who are significant.

Results

A. First randomization

1) The first nearest distance Based Model without interaction among independent variables:

Variables	Full Model		Reduce Model	
	Coefficient	p-value	Coefficient	p-value
Intercept	-1.47000	0.31476	-0.54000	0.24320
Distance1	-0.00084	0.00130 **	-0.00084	0.00089 ***
Elevation	0.00412	0.61415		
Slope	-0.21770	0.50636		
Road density	1362	< 2e-16 ***	1365	< 2e-16 ***
Direction river	0.00154	0.20214		
Distance river	-0.00058	0.00109 **	-0.00057	0.00125 **
Distance waterbody	-0.00146	2.54e-09 ***	-0.00141	1.14e-09 ***
Null deviance	974.56	df 702	974.56	df 702
Residual deviance	408.97	df 695	411.15	df 698
AIC	424.97		421.15	

Running the logistic regression fitting in all the variables proposed: 1) road density. 2) distance from the nearest existing village, 3)distance from river and stream, 4) distance from water body , 5)elevation, and 6)slope. We found that the following variables were significant:

- Distance1 (the first nearest neighbor village to village or non-village, with negative relation)
- Road density (with positive relation)
- Distance from river (with negative relation)
- Distance from water body (with negative relation)

Village occurrence = dist1 + road_density + dist_river + dist_waterbody

This is the R code result showing the estimates and the level of significance. As listed below distance1, distance from waterbody and distance from river have a negative relationship with the existence of the village while road density has a positive relationship.

**glm(formula = z ~ dist1 + road_density + dist_river + dist_waterbody
, family = binomial, data = data_logistic)**

Deviance Residuals:

```
 Min        1Q     Median       3Q       Max
-3.73955  -0.42220  -0.00496   0.32572   3.67513
```

Testing Autocorrelation for model A (without interaction);

Maron's I and Geary's C:

Computing I and C with equal weights proximity matrix (K = 2) on the residuals from the Model shows that there is no autocorrelation. Both are not significant at the .05 level and so we don't reject the null hypothesis that there is no autocorrelation.

Monte-Carlo simulation of Moran's I (k=2):
data: rm12.nodist2.noint.logis$residuals
weights: data2.sw
number of simulations + 1: 100
statistic = -0.0017, observed rank = 15, p-value = 0.85
alternative hypothesis: greater

Monte-Carlo simulation of Geary's C (k=2):
data: rm12.nodist2.noint.logis$residuals
weights: listw2U(data1.sw)
number of simulations + 1: 100
statistic = 0.6181, observed rank = 6, p-value = 0.06
alternative hypothesis: less

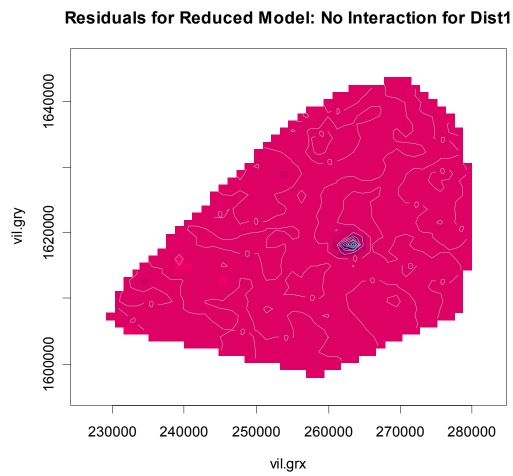


Figure 3

2) The first nearest distance Based Model with interaction among independent variables:

Variables	Dist1 Model		Dist2 Model	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Intercept	1.47E+00	0.01893 *	-1.72E+00	4.23e-08 ***
Distance i	-2.52E-03	4.31e-10 ***		
Road density	1.35E+03	< 2e-16 ***	1.52E+03	< 2e-16 ***
Distance river	1.14E-04	0.69848	-6.52E-04	0.000152 ***
Distance waterbody	-3.35E-03	2.12e-11 ***	-1.64E-03	2.27e-13 ***
Dist i*distance waterbody	1.64E-06	5.22e-10 ***		
Distance river*distance waterbody	-9.12E-07	0.00247 **		
Null deviance	974.56	df 702	974.56	df 702
Residual deviance	371.85	df 696	423.04	df 699
AIC	385.85		431.04	

Village occurrence = dist1 + road_density + dist_river + dist_waterbody + dist1:dist_waterbody + dist_river:dist_waterbody

We found that the following variables were significant:
- Distance1(the first nearest neighbor village to village or non-village, with negative relation)
- Road density(with positive relation)
- Distance from water body(with negative relation)
- Dist1 & dist_waterbody(with positive relation)
- Dist_river & dist_waterbody(with negative relation)

In this model distance1, distance from waterbody and the interaction of (dist_river with dist_waterbody) have a negative relationship with the existence of the village while road density and the interaction of (dist1 with dist_waterbody)have a positive relationship.

glm(formula = z ~ dist1 + road_density + dist_river + dist_waterbody + dist1:dist_waterbody + dist_river:dist_waterbody, family = binomial, data = data_logistic)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.660705	-0.376687	-0.003157	0.243032	2.863843

Testing Autocorrelation for model B (with interaction);
Maron's I and Geary's C:

Computing I and C with equal weights proximity matrix (K = 2) on the residuals from the Model shows that there is no autocorrelation. Both are not significant at the .05 level and so we don't reject the null hypothesis that there is no autocorrelation.

Monte-Carlo simulation of Moran's I (k=2):
data: rm1.int2.logis$residuals
weights: data2.sw
number of simulations + 1: 100
statistic = -7e-04, observed rank = 56, p-value = 0.44
alternative hypothesis: greater

Monte-Carlo simulation of Geary's C (k=2):
data: rm1.int2.logis$residuals
weights: listw2U(data2.sw)
number of simulations + 1: 100
statistic = 0.5059, observed rank = 8, p-value = 0.08
alternative hypothes

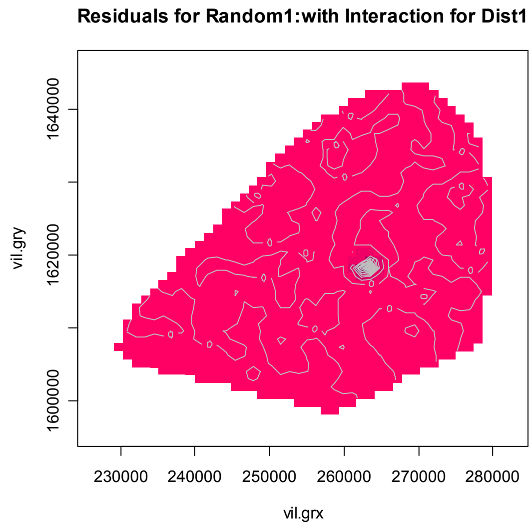


Figure 4

3) The second nearest distance based model (distance2) without interaction among independent variables:

Variables	Full Model		Reduce Model	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Intercept	-1.83E+00	0.212104	-1.72E+00	4.23e-08 ***
Distance2	-2.85E-04	0.205207		
Elevation	2.44E-03	0.761516		
Slope	-2.73E-01	0.386691		
Road density	1.46E+03	< 2e-16 ***	1.52E+03	< 2e-16 ***
Direction river	1.53E-03	0.199315		
Distance river	-6.07E-04	0.000712 ***	-6.52E-04	0.000152 ***
Distance waterbody	-1.59E-03	8.94e-11 ***	-1.64E-03	2.27e-13 ***
Null deviance	974.56	df 702	974.56	df 702
Residual deviance	418.16	df 695	423.04	df 699
AIC	434.16		431.04	

In this model, distance2 were considered instead. As shown in the results below, dist_river and dist_waterbody have a negative relationship with the existence of the village, while road density has a positive relationship.

- Road density (with positive relation)
- Distance from river (with negative relation)
- Distance from water body (with negative relation)

Village occurrence = road_density + dist_river + dist_waterbody

glm(formula = z ~ road_density + dist_river + dist_waterbody, family = binomial, data = data_logistic)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.002702	-0.429137	-0.009968	0.364905	3.024811

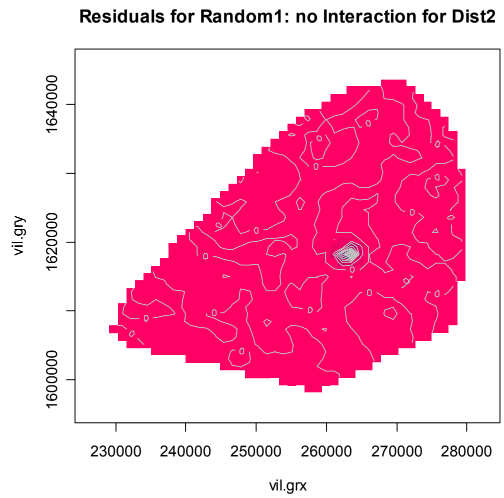


Figure 5

Testing Autocorrelation for model A(without interaction);
Maron's I and Geary's C:

Computing I and C with equal weights proximity matrix (K = 2) on the residuals from the Model shows that there is no autocorrelation. Both are not significant at the .05 level and so we don't reject the null hypothesis that there is no autocorrelation.

Monte-Carlo simulation of Moran's I(K=2)
data: rm2.int.logis$residuals
weights: data2.sw
number of simulations + 1: 100
statistic = -2e-04, observed rank = 84, p-value = 0.16
alternative hypothesis: greater

Monte-Carlo simulation of Geary's C (K=2)
data: rm2.int.logis$residuals
weights: listw2U(data2.sw)
number of simulations + 1: 100
statistic = 0.5018, observed rank = 13, p-value = 0.13
alternative hypothesis: less

4) The second nearest distance based model (distance2) with interaction among independent variables:

The model yields the same result as in the model without interaction.

B. Second randomization with constrained 200 meters from each other.

1) The first nearest distance based model without interaction among independent variables:

Variables	Full Model		Reduce Model	
	Coefficient	p-value	Coefficient	p-value
Intercept	-3.17E+00	0.0260 *	-2.79E+00	0.0407 *
Distance1	-1.25E-03	4.40e-07 ***	-1.23E-03	5.47e-07 ***
Elevation	1.99E-02	0.0103 *	1.88E-02	0.0135 *
Slope	-6.39E-01	0.0396 *	-6.34E-01	0.0409 *
Road density	1.19E+03	< 2e-16 ***	1.19E+03	< 2e-16 ***
Direction river	1.21E-03	0.3033		
Distance river	-6.77E-04	6.69e-05 ***	-6.81E-04	6.08e-05 ***
Distance waterbody	-1.57E-03	1.04e-10 ***	-1.53E-03	1.40e-10 ***
Null deviance	974.56	df 702	974.56	df 702
Residual deviance	419.94	df 695	421.01	df 696
AIC	435.94		435.01	

Running the logistic regression, fitting in the variables proposed, we found that the following variables were significant:

Village occurrence = dist1 + elevation + slope + road_density + dist_river + dist_waterbody

- Distance1 (first nearest neighbor to village or non-village, with negative relation)
- Elevation (with positive relation)
- Slope (with negative relation)
- Road density (with positive relation)
- Distance from river (with negative relation)
- Distance from water body (with negative relation)

glm(formula = z ~ dist1 + elevation + slope + road_density +
dist_river + dist_waterbody, family = binomial, data = data_logistic)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.957901	-0.436025	-0.005791	0.321242	3.818477

The R code results showing the estimates and the level of significance. Distance1, distance from river, slope and dist from waterbody have a negative relationship with the existence of the village while road density and elevation have a positive relationship.

Testing Autocorrelation for model A(without interaction);
Moran's I and Geary's C:

Computing I and C with equal weights proximity matrix (K = 2) on the residuals from the Model shows that there is no autocorrelation. Both are not significant at the .05 level and so we don't reject the null hypothesis that there is no autocorrelation.

Monte-Carlo simulation of Moran's I (K=2):

data: rm12.noint.logis$residuals
weights: data2.sw
number of simulations + 1: 100

statistic = -0.0033, observed rank = 27, p-value = 0.73
alternative hypothesis: greater
Monte-Carlo simulation of Geary's C(K=3):
data: rm12.noint.logis$residuals
weights: listw2U(data2.sw)
number of simulations + 1: 100
statistic = 0.7941, observed rank = 30, p-value = 0.3
alternative hypothesis: less

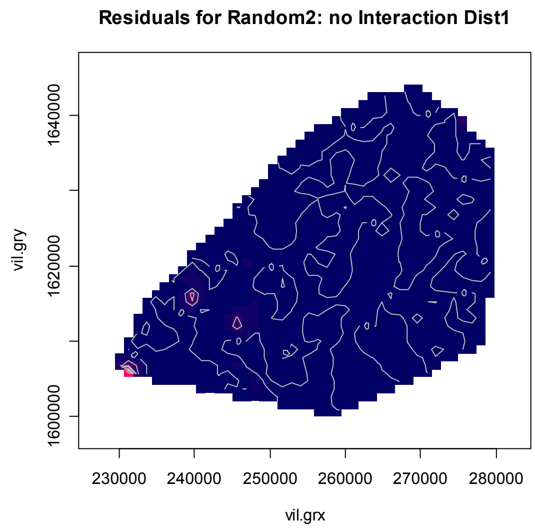


Figure 6

2) The first nearest distance based model with interaction among independent variables:

Variables	Dist1 Model	
	Coefficient	p-value
Intercept	4.34E+00	0.165566
Distance i	-8.82E-04	0.133518
Elevation	-2.56E-02	0.123703
Slope		
Road density	-6.44E+02	0.606986
Distance river	-4.50E-05	0.871749
Distance waterbody	-3.69E-03	1.98e-13 ***
Dist i*road density	-7.39E-01	2.86e-05 ***
Dist i*distance river		
Dist i*distance waterbody	1.70E-06	1.26e-10 ***
Elevation*road density	1.60E+01	0.019646 *
Slope*distance waterbody		
Distance river*distance waterbody	-9.63E-07	0.000967 ***
Null deviance	974.56	df 702
Residual deviance	362.75	df 693
AIC	382.75	

The logistic regression analysis showed that these variables were significant:

Village occurrence = dist1 + elevation + road_density + dist_river + dist_waterbody + dist1*road_density + dist1*dist_waterbody + elevation*road_density + dist_river*dist_waterbody, family = binomial, data = data_logistic)

- Distance from water body(with negative relation)
- Dist1 & road_density(with negative relation)
- Dist1 & water_body(with positive relation)
- Elevation & road_density(with positive relation)
- Dist_river & dist_water body(with negative relation)

The R code results show the estimates and the level of significance. As listed below, dist_waterbody, the interaction of: (dist1 & road density) and (dist_river & dist_waterbody) all have a negative relationship with the existence of the village while the interaction of (dist1 & dist_waterbody) and (elevation & road_density) have a positive relationship. Note that there are few terms (such as, dist1, elevation, etc) that are not significant but cannot be removed from the model because they have a significant interaction with other terms. The model shows below with the interaction terms.

glm(formula = z ~ dist1 + elevation + road_density + dist_river + dist_waterbody + dist1:road_density + dist1:dist_waterbody + elevation:road_density + dist_river:dist_waterbody, family = binomial, data = data_logistic)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7846103	-0.4091512	-0.0003913	0.1694263	2.9802044

Testing Autocorrelation for model B (with interaction);
Maron's I and Geary's C:

Computing I and C with equal weights proximity matrix (K = 2) on the residuals from the Model shows that there is no autocorrelation. Both are not significant at the .05 levels and so we don't reject the null hypothesis that there is no autocorrelation.

Monte-Carlo simulation of Moran's I (K=2):
data: rm1.int4.logis$residuals
weights: data2.sw
number of simulations + 1: 100
statistic = -0.0374, observed rank = 5, p-value = 0.95
alternative hypothesis: greater

Monte-Carlo simulation of Geary's C(K=2):
data: rm1.int4.logis$residuals
weights: listw2U(data2.sw)
number of simulations + 1: 100
statistic = 1.2262, observed rank = 91, p-value = 0.91
alternative hypothesis: less

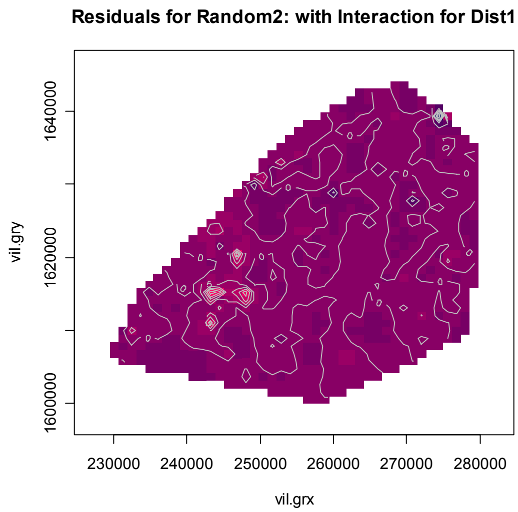


Figure 7

3) The second nearest distance based model without interaction among independent variables:

Variables	Full Model		Reduce Model	
	Coefficient	**p-value**	**Coefficient**	**p-value**
Intercept	-3.38E+00	0.0108 *	-3.08E+00	0.0152 *
Distance2	-5.05E-04	0.0201 *	-4.90E-04	0.0229 *
Elevation	1.71E-02	0.0161 *	1.62E-02	0.0201 *
Slope	-7.19E-01	0.0174 *	-7.14E-01	0.0182 *
Road density	1.28E+03	< 2e-16 ***	1.28E+03	< 2e-16 ***
Direction river	9.60E-04	0.3981		
Distance river	-7.56E-04	1.17e-05 ***	-7.59E-04	1.07e-05 ***
Distance waterbody	-1.68E-03	8.17e-13 ***	-1.65E-03	9.94e-13 ***
Null deviance	974.56	df 702	974.56	df 702
Residual deviance	443.01	df 695	443.73	df 699
AIC	459.01		457.73	

Village occurrence = dist2 + elevation + slope + road_density + dist_river + dist_waterbody

The logistic regression analysis showed that these variables were significant:

- Distance2 (with negative relation)
- slope(with negative relation)
- elevation
- dist_water_body(with positive relation)
- road_density(with positive relation)
- Dist_river(with negative relation)

The R code results show the estimates and the level of significance. As listed below, distance2, slope, dist_waterbody and dist_river have a negative relationship with the existence of the village while elevation and road_density have a positive relationship.

glm(formula = z ~ dist2 + elevation + slope + road_density + dist_river + dist_waterbody, family = binomial, data = data_logistic)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.48085	-0.43725	-0.01064	0.38670	3.24094

Number of Fisher Scoring iterations: 6

Testing Autocorrelation for model A (without interaction);
Moran's I and Geary's C:

Computing I and C with equal weights proximity matrix (K = 2) on the residuals from the Model shows that there is no autocorrelation. Both are not significant at the .05 levels and so we don't reject the null hypothesis that there is no autocorrelation.

Monte-Carlo simulation of Moran's I
data: rm2.noint.logis$residuals
weights: data2.sw
number of simulations + 1: 100
statistic = 0.005, observed rank = 87, p-value = 0.13
alternative hypothesis: greater

Monte-Carlo simulation of Geary's C
data: rm2.noint.logis$residuals
weights: listw2U(data2.sw)
number of simulations + 1: 100
statistic = 1.324, observed rank = 95, p-value = 0.95
alternative hypothesis: less

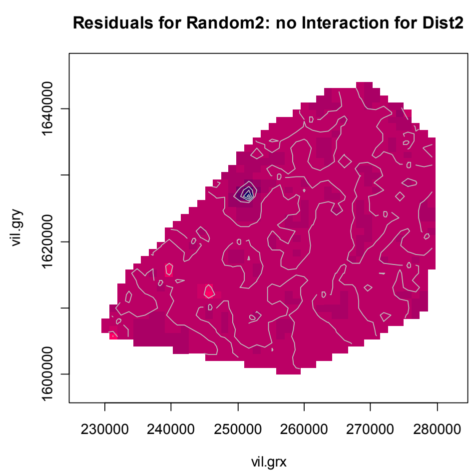


Figure 8

4) The second nearest distance based model with interaction among independent variables:

Variables	Dist2 Model	
	Coefficient	p-value
Intercept	3.31E+00	0.29988
Distance i	-1.28E-04	0.79801
Elevation	-2.23E-02	0.18407
Slope	-1.47E+00	0.00456 **
Road density	-2.47E+02	0.84245
Distance river	-1.01E-03	0.03994 *
Distance waterbody	-3.16E-03	1.08e-07 ***
Dist i*road density	-7.18E-01	5.00e-06 ***
Dist i*distance river	4.89E-07	0.02673 *
Dist i*distance waterbody	8.97E-07	6.14e-05 ***
Elevation*road density	1.52E+01	0.02599 *
Slope*distance waterbody	7.85E-04	0.03353 *
Distance river*distance waterbody	-9.65E-07	0.00143 **
Null deviance	974.56	df 702
Residual deviance	392.97	df 690
AIC	418.97	

Village occurrence = dist2 + elevation + slope + road_density + dist_river + dist_waterbody + dist2*road_density + dist2*dist_river + dist2*dist_waterbody + elevation*road_density + slope*dist_waterbody + dist_river*dist_waterbody

The logistic regression analysis showed that these variables were significant:

- slope(with negative relation)
- dist_river(with negative relation)
- dist_water_body(with negative relation)
- dist2 & road_density(with negative relation)
- dist2 & dist_river(with positive relation)
- dist2 & dist_waterbody(with positive relation)
- elevation & road_density(with positive relation)
- slope & dist_waterbody(with positive relation)
- dist_river & dist_waterbody(with negative relation)

The R code results show the estimates and the level of significance. As listed below, slope, dist_waterbody, dist_river (dist2 & road_density) and (dist_river & dist_waterbody)all have a negative relationship with the existence of the village while the interaction of (dist2 & dist_river), (dist2 & dist_waterbody) and (elevation & road_density) have a positive relationship.

glm(formula = z ~ dist2 + elevation + slope + road_density + dist_river + dist_waterbody + dist2:road_density + dist2:dist_river + dist2:dist_waterbody + elevation:road_density + slope:dist_waterbody + dist_river:dist_waterbody, family = binomial, data = data_logistic)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.173067	-0.403581	-0.007028	0.248529	2.929279

Testing Autocorrelation for model B(with interaction);
Moran's I and Geary's C:

Computing I and C with equal weights proximity matrix (K = 2) on the residuals from the Model shows that there is no autocorrelation. Both are not significant at the .05 level and so we don't reject the null hypothesis that there is no autocorrelation.

Monte-Carlo simulation of Moran's I
data: rm2.int2.logis$residuals
weights: data2.sw
number of simulations + 1: 100
statistic = 0.0278, observed rank = 93, p-value = 0.07
alternative hypothesis: greater
Monte-Carlo simulation of Geary's C
data: rm2.int2.logis$residuals
weights: listw2U(data2.sw)
number of simulations + 1: 100
statistic = 1.0164, observed rank = 56, p-value = 0.56
alternative hypothesis: less

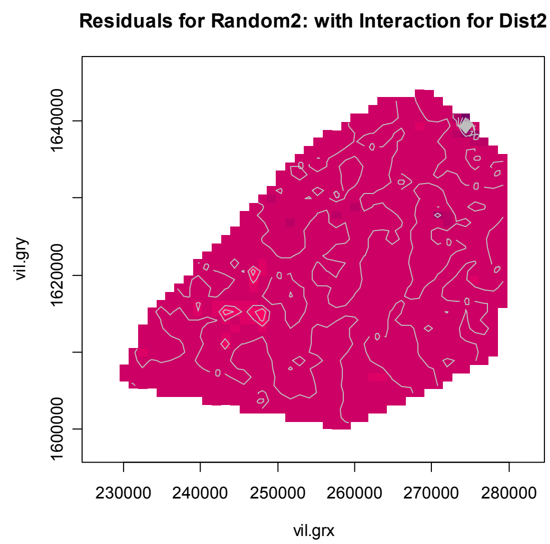


Figure 9

Conclusions and Discussions:

In general for the first randomization, distance1 models have smaller AIC than distance2 models, but the difference is not significance. AIC is commonly used to compare models, where the lower the AIC, the better. To conclude, road density, distance to water body, distance to river, and distance to nearest existing village were significantly correlated to the central of the village within the Nang Rong district. The second nearest variable was not significant when substituted with the first nearest village. Moran's-I and Geary's-C were not significant for K=2, indicating no autocorrelation for first randomization. Model with interactions improved AIC, however, the interpretation may be ambiguous. The residual map highlighted the middle part of the study areas where have to further investigate.

In second simulation with 200 m. distance constrained between random points, road density, distance to water body, distance to river, and distance to nearest existing village, slope, and elevation were significantly correlated to the central of the village within the Nang Rong district. The second nearest v was also significant when substituted with the first nearest village with those same other independents in the model. Moran's-I and Geary's-C were not significant for K=2, indicating no autocorrelation for first randomization. Model with interactions improved AIC similar to the first randomization. The residual map based on distance1 shows significantly different between with/without interactions in the Southwestern region. The model without interaction captures a better variability than with interaction.

In conclusion, human settlement in Nang Rong takes the form of nuclear households clustered in villages. Road density, distance to water body, distance to river, and distance to nearest existing village were significantly correlated to the central of the village within the Nang Rong district. This can be inferred that people tended to settle closer distance to river and water body, high road density, and further away from previous existing village center. However, with a constraint 200 meters apart of randomization points, the results show that the second nearest village, slope and elevation may be additional functions to the settlement pattern. The political structures of the region, and in particular, localized governance, may have prevented and may be continuing to inhibit development along some corridors. The orientation and geographic extent of villages' functional territories may depend on more than existing urban centers, transportation, and accessibility. From a biophysical perspective, natural barriers such as a river, riparian corridor (Varnakovida and Messina 2004), or a water impoundment may influence the size, orientation, and distance of lands farmed by households living within a particular village. Also, the nature of the topography surrounding the village settlement has implications for LULCC dynamics; rain fed rice is primarily grown in the lowlands, and field crops such as cassava and sugar cane in uplands. The intermediate terraces are regarded, as transitional areas where rice might be extended during periods of suitable rainfall and/or crop prices, or field crops might similarly be grown in the high and middle terraces where environmental conditions permit and/or economic opportunity encourages their cultivation. A high-resolution digital elevation model would have prevented the annually flooded areas from being selected for urbanization, but such data was unavailable for the study area.

According to statistic results, in general, the models with the interaction terms have a lower AIC. To pick one between the two models (with/without interaction) based on the AIC, there is really no big difference between the two models. However, the significant of variables with interaction were not really meaningful. The models without interaction have better residual plots. The residual plots for the model with the interaction has lots of pockets while the other plots for the model without interaction did not show a lot of pockets. Therefore, we agreed that the simple model is preferred over the one with the interaction terms. Models based on distance1 have lower AICs than those based on distance2, but they are not better in a significant matter.

The model developed for this research is only preliminary study. Further investigation could be researched especially the closest distance from the existing village variable relating to Central Place Theory. However, the models could easily be applied to any rural setting in Thailand, and many other locations as well, especially with place-specific adaptations. Other questions may be a worthwhile topic of research as well; while it is difficult to foresee a scenario whereby the district of Nang Rong experiences mass out-migration as it has in the past, it is certainly possible given significant economic collapse and may be a worthwhile future research question. Even so, it is more likely that Nang Rong will continue to aggressively expand both spatially and by population density creating a larger impact footprint throughout the region.

References

Crawford, T.W., 2000. Human-Environment Interactions and Regional Change in Northeast Thailand: Relationships between socio-Economic, Environment, and Geographic Patterns. Doctoral Dissertation, Department of Geography, University of North Carolina-Chapel Hill.

Crawford. 2002. Spatial Modeling of Village Fuctional Territories to Support Population-Environment Linkages. In Linking People, Place, and Policy A GIScience Approach, 91-111: Kluwer Academic Publishers.

Evans, T., and E. F. Moran. 2002. Spatial integration of social and biophysical factors related to landcover change. Population and Development Review 28:165-186.

Entwisle, B., S. J. Walsh, R. R. Rindfuss, and A. Chamratrithirong. 1998. Land-use/Land-cover and population dynamics, Nang Rong, Thailand. In People and Pixels: Linking Remote Sensing and Social Science, eds. D. Liverman, E. F. Moran, R. R. Rindfuss and P. C. Stern, 121-144. Washington, D.C., USA: National Academy Press.

Hall, C. A. S., Y. Qi, G. Pontius, and J. Cornell. 1995. Modelling spatial and temporal patterns of tropical land use change. Jornal of Biogeography 22 (4/5):753-757.

Irwin, E. G., and J. Geoghegan. 2001. Theory, data, methods: Developing spatially explicit economic models of land use change. Agriculture, Ecosystems and Environment 85:7-23.

Messina, J. P., and S. J. Walsh. 2001. 2.5D morphogenesis: Modeling landuse and landcover dynamics in the Ecuadorian Amazon. Plant Ecology 158 (1):75-88.

Nelson, G. 2002. Introduction to the special issue on spatial analysis for agricultural economists. Agricultural Economics 27:197-200.

Pan, W. K. Y., S. J. Walsh, R. E. Bilsborrow, B. G. Frizzle, C. M. Erlien, and F. Baquero. 2004. Farm-level models of spatial patterns of land use and land cover dynamics in the Ecuadorian Amazon. Agriculture, Ecosystems and Environment 101:117-134.

Parnwell, M. J. G. 1988. Rural Poverty, Development and the Environment: The case Study of Northeast Thailand. . Journal of Biogeography 15:199-208.

Rindfuss, R. R., et al. 2002. Continuous and Discrete: Where They Have Met in Thailand. . In Linking People, Place, and Policy A GIScience Approach, 7-37: Kluwer Academic Publishers.

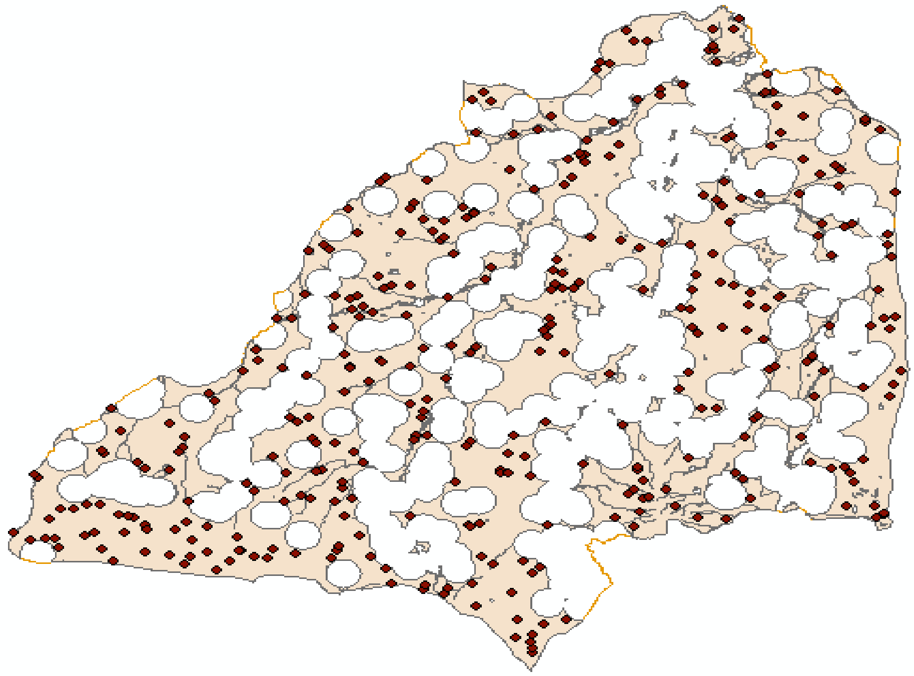
Silverman, B.W. Density Estimation for Statistics and Data Analysis. New York: Chapman and Hall, 1986.

Steffen, W., A. Sanderson, P. Tyson, J. Jäger, P. Matson, B. Moore III, F. Oldfield, K. Richardson, H. J. Schellnhuber, B. L. Turner II, and R. J. Wasson. 2004a. Chapter 3 - The anthropocene era: How humans are changing the Earth System. In Global Change and the Earth System: A Planet Under Pressure, 81-141. New York, NY, USA: Springer-Verlag.

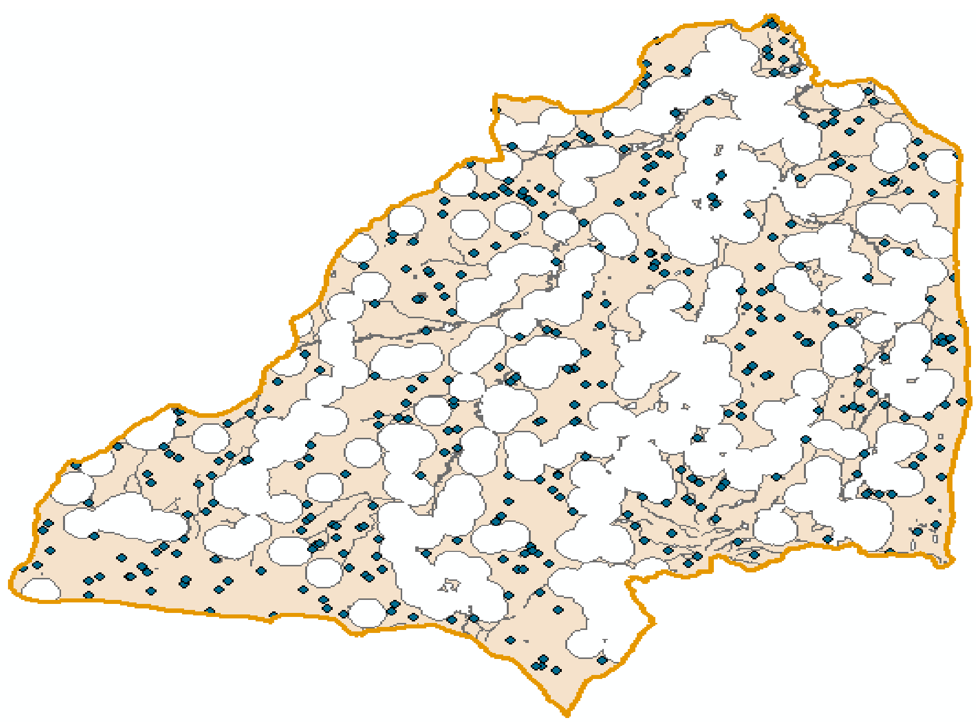
Varnakovida, P. and Messina, J., P., 2004. Urban Modeling and Landscape Prediction: The Case Study of Nang Rong, Thailand. Proceedings, Applied Geography Conference, (Montz, B.E. and Tobin, G.A., editors), 27:239-246.

Welsh, W.F., 2001. Agro-Ecological Sustainibility and Land Degradation Potential in Nang Rong, Thailand. Doctoral Dissertation, Department of Geography, University of North Carolina-Chapel Hill.

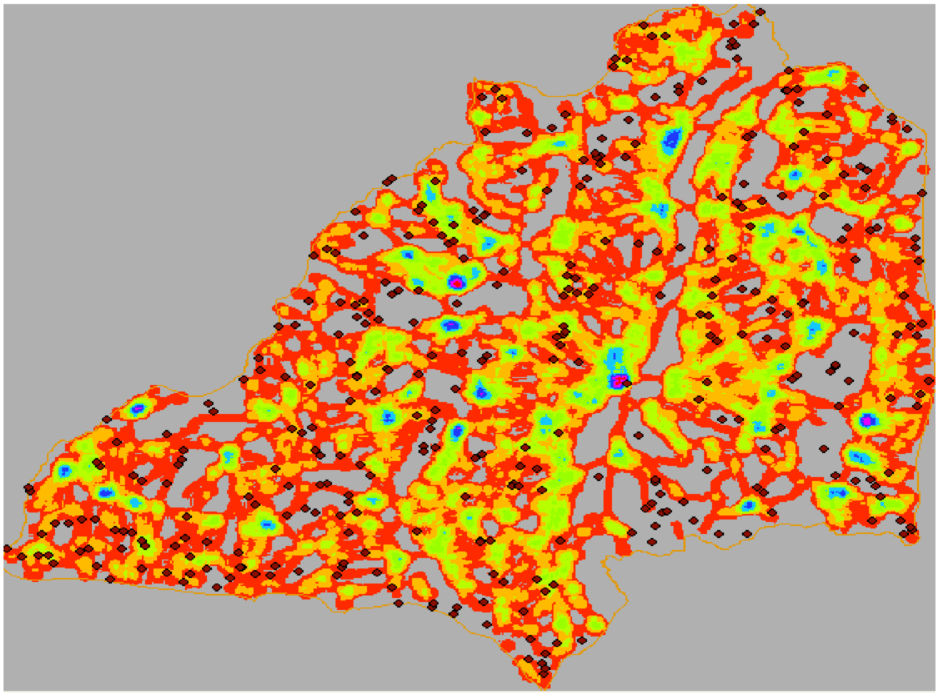
Appendices



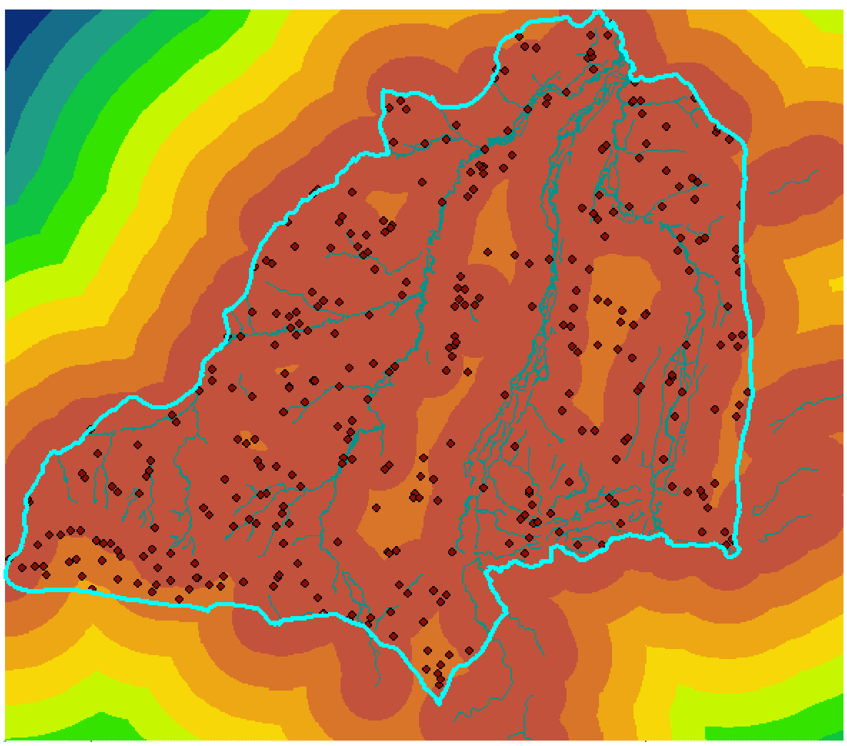
Appendix 1: First Randomization



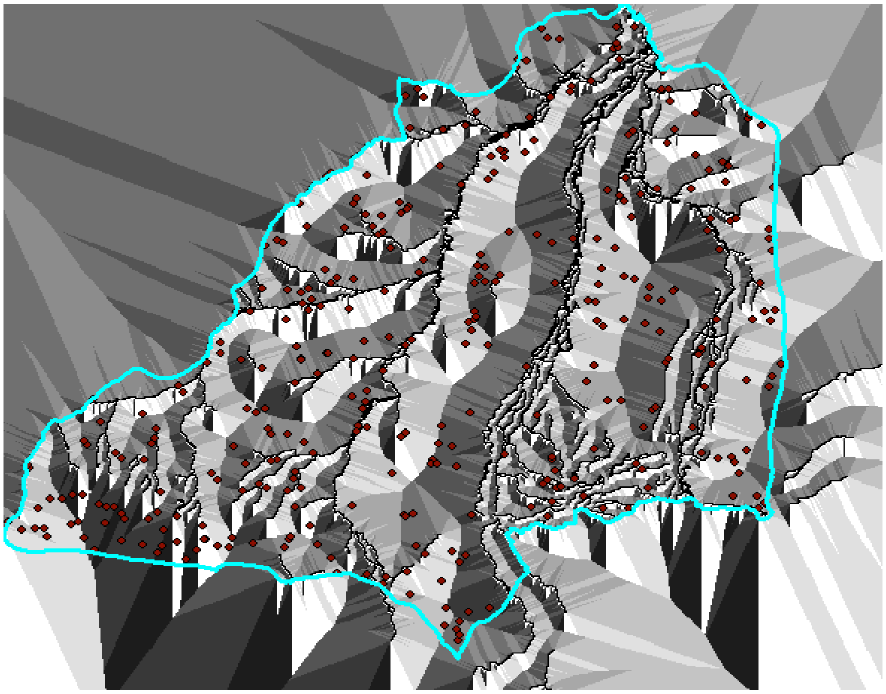
Appendix 2: Another random points with constraint 200 meters to each other



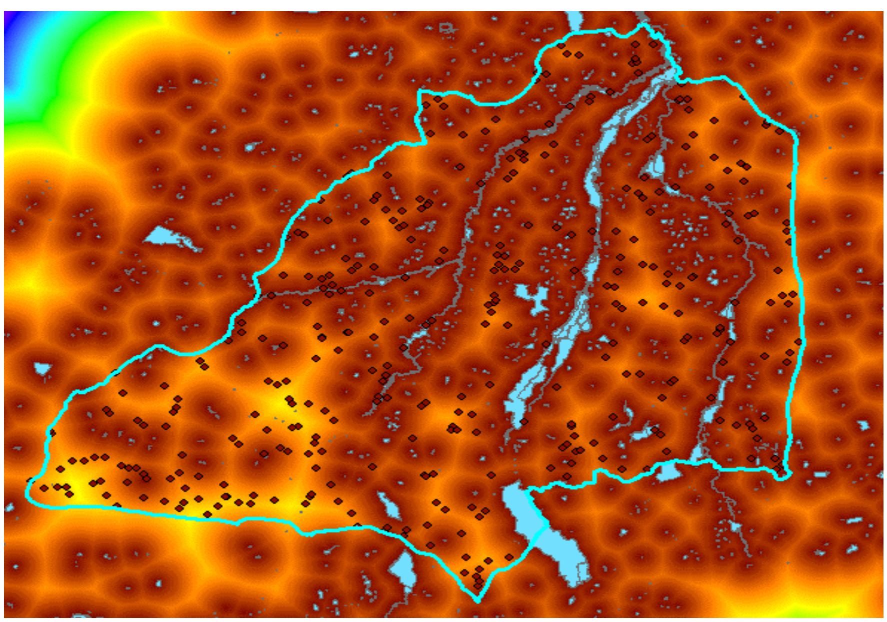
Appendix 3: Road density using line density in ArcGIS



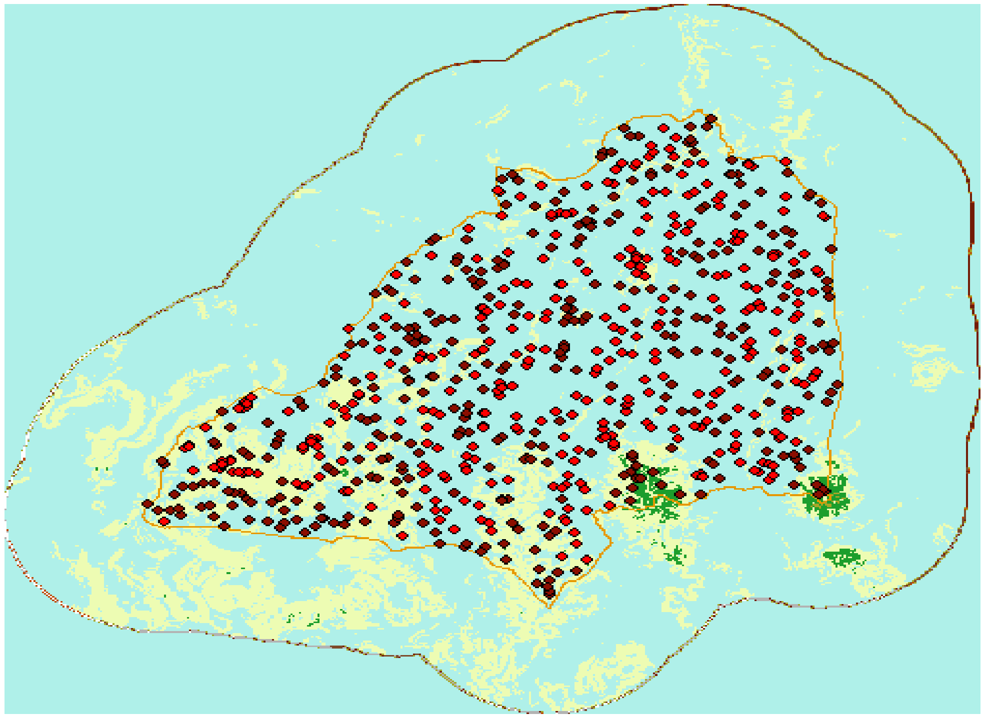
Appendix 4: Euclidean distance from river



Appendix 5: Euclidean distance direction from rivers



Appendix 6: Euclidean distance from water bodies



Appendix 7: Slope