International Journal of Geology, Agriculture and Environmental Sciences
Volume – 6 Issue –6 December 2018
Website: www.woarjournals.org/IJGAES *ISSN: 2348-0254*

Evaluation of Deterministic Spatial Interpolators with myGeoffice©: The Utah Grasshoppers Case

Joao Negreiros[1], Jenny Phillips-Lao[2]

[1]School of Business & Law, University of Saint Joseph,
Estrada Marginal da Ilha Verde, 14-17, Macao, China.

[2]School of Business & Law, University of Saint Joseph,
Estrada Marginal da Ilha Verde, 14-17, Macao, China.

Abstract: ***The knowledge of spatial distribution of grasshoppers can be very relevant for agricultural planning purposes. On the other hand, the comparison of spatial interpolators for efficiency and reliability reasons is also a key factor to understand interpolation maps outcomes (versus reality). At last, but not least, the use of open Web geographical tools to disseminate true spatial inferential methods to address spatial issues is still quite limited (if none) in high schools and universities, particularly in Geography subjects. If the latter can be addressed with myGeoffice©, the first issue will use the Utah, USA, dataset (58 samples) to layout the spatial distribution of grasshoppers and understand the counties that are more pro to this kind of agriculture infestation. Inverse Distance Weighted (IDW), Moving Average (MA), Multi-quadratic, Inverse Multi-quadratic and Nearest Neighbor (NN) will produce interpolated surfaces of grasshopper's properties. Efficiency of spatial interpolators was assessed in this writing based on the prediction error's statistics derived from the difference between the estimation and the real samples on a cross-validation procedure. Remarkably, results show that NN was the most accurate one when compared with the remaining deterministic approaches at sample's locations.***

Keywords: myGeoffice©, Spatial deterministic interpolators, Grasshoppers, Utah.

1. Introduction

The advent of the Internet in 1992 was one of the major greatest achievements of humanity, leading our planet to be reduced to a simple global village. As expected, human daily behaviors are changing accordingly [14]. Quite often, parents ask themselves about what are their kids doing (for hours and hours) on their smartphones. As expected, the only plausible answer is "a lot of things". Curiously, if these digital natives are not connected, a guilt feeling is assumed in a fairly and shy way (the syndrome of missing out fear) since any friend may accuse him/her of "I called you and you did not answer me". As well, if these youngsters do not hold Facebook, WhatsApp, Line, Twitter or WeChat, they felt marginalized by their social community and friends. Sharing the latest news, gossips, pictures, YouTube links or selfies are "a must" on our present digital social life [15].

Some parents setup the "no technology day per week" regarding their own children while a few schools adopted the no Wi-Fi policy for students. The French psychoanalyst Serge Tisseron setup the 3, 6, 9, 12 rule for screens management regarding the present Z generation: No computer screens before age 3, no Internet before 6, no electronic gaming until 9 and no unsupervised Internet before 12.

Can information technologies improve student's grades? Half true. The number of slept hours and real study decreased, the sharing of sexting, grooming and cyberbullying increase whilst the forever-digital footprint of our personal history became a reality. As well, online games and shopping activities followed the same latter pattern. The time concept changed, that is, one hour studying Math is equivalent to 9 hours in Instagram [7]. On the other hand, wonderful Websites with multimedia allows us to learn everything in a second and clear all our misconceptions and confusions in any science field. Geography, in general, and Geographical Information Systems (GIS), in particular, is not an exception to this new learning trend.

myGeoffice© is an innovative WWW platform for GIS users which includes the Moran scatterplot, Kriging with measurement error and several nugget-effect solutions, sampling declustering based on the nearest neighborhood analysis, geographical weighted regression (GWR), Dijkstra shortest path, raster image processing, index of Knot and Mantel, Kernel Gaussian density, sequential simulation and deterministic interpolators, for instance [5]. Yet, myGeoffice© is not a comprehensive statistical package in the traditional of solving everyone's problems. [6].

In this research article, the first aim is to promote attention and curiosity in young students, teachers and other fans of Geography for myGeoffice©, a free WebGIS 2.0 tool, to resolve geographical problems. Globally, it is expected that readers may comprehend and take the below full comprehensive deterministic spatial interpolation issue into their classroom for discussion and construct new knowledge in a spatial way of thinking.

The second goal is to make a comparison of the main spatial deterministic interpolators available at present, that is, Multi and Inverse Multi-quadratic Radial Base Functions (RBFs), Inverse Distance Weighted (IDW), Moving Average (MA), Nearest Neighbor (NN), Voronoi and Thiessen with myGeoffice©. For that, a grasshopper real dataset of 2014 from Utah, USA, will be used to assess their accuracy.

The third ambition regards about the understanding of grasshoppers outbreaks, a major agriculture issue in certain regions of the planet. According to [9], several difficulties can be mentioned towards this problem: A) Control effects frequently use a single treatment to manage all species with varying success; B) The fluctuation in population involves both space and time factors; C) Information gained from laboratory studies may not be applicable under field conditions. For example, natural reproductive rates are much lower than the rates obtained in studies of caged grasshoppers; D) Grasshoppers are managed on various spatial scales; E) The grasshopper species is composed of 600 species with particular habitat requirements. Thus, better grasshopper identification and forage biomass estimation techniques are needed.

Besides the present section that introduces the background of this theoretical study, this research paper consists of six other main sections. The second one briefly highlights the main characteristics of six deterministic interpolators while mygeoffice.org is portrayed next. Section four, five and six presents the three main phases on a typical case study of spatial analysis: exploratory, autocorrelation and interpolation. The conclusion ends this research article before the references section.

2. Spatial Deterministic Interpolators: Review

Visiting every location in a study area to measure the height, magnitude or concentration of a phenomenon is usually difficult or expensive. Instead, dispersed sample input point locations can be selected and an expected value can be assigned to all other locations. This is the key reason of the existence of spatial interpolators.

By definition, interpolation predicts values for cells in a raster format from a limited number of sample dataset and it can be used to predict unknown values for any geographic point data such as elevation, soil properties, rainfall, population density, chemical concentrations and noise levels [1]. For Mother Nature, it is quite common that, on average, values at points close together in space are more likely to be similar than points further away. This became known as the First Law of Geography or Tobler's Law. It is the capability to use this local information that really makes the difference among spatial deterministic interpolators.

There are two types of techniques for generating raster surfaces: A) Deterministic, which use a mathematical function to predict unknown values and result in hard classification of the features to be estimated; B) Geo-statistics that produce confidence limits to the prediction accuracy but are more difficult to execute since more parameters need to be set. In this article, only the first approach will be covered although myGeoffice© also computes Ordinary/Indicator Kriging and Gaussian and Indicator geo-simulation.

2.1 Inverse Distanced Weighted

Inverse Distance Weighted (IDW) only accounts for distance relationships only (see Figure 1 left). This method assigns weights in an averaging function, based on the inverse of the distance (raised to some power) to every data points located within a given search radius (centered on the estimate point). However, a common concern with all these linear interpolators is the conditional bias issue: under-estimation of high values and over-estimation of low values (smoothing out of the dips and the humps) as Figure 1 (right) shows.

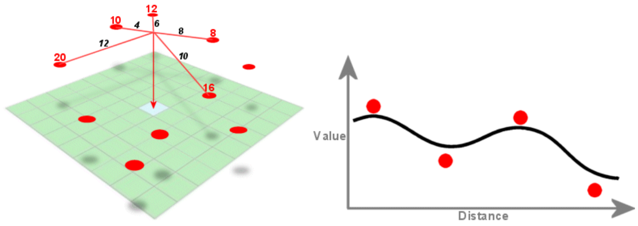


Figure 1: As stated by [10], the interpolated map also depends on data clustering (IDW works better if samples are evenly distributed) and outlier presence because these estimations are strongly influenced by the closest samples (source: geography.hunter.cuny.edu).

The higher the function power, the more weight will be given to closer samples in conformity with Tobler's Law (all things are related, but nearby things are more related than distant ones), where the best guess is the measured value at closest observations. Yet and again, as more samples are included into the weighted linear combination, the resulting estimates become less variable, which leads to fewer extremes. Regarding programming, the resulting map first sets a grid over the area as the estimated value at each node is calculated. However, in some of those grid nodes, there are already samples implying an infinity value since distance equals zero. Therefore, two special cases for this calculation should be taken into consideration: (1) For most software, the observed values are copied over, forcing this technique to be an exact interpolator; (2) With the moon hole-effect, a zero estimated value is assigned due to samples lacking within its neighborhood. Another key issue of IDW involves the search of the optimal IDW (power function or best number of samples to include, in particular). Quite often, this decision is based on the cross-validation procedure, that is, by using the original samples and their predictions, the best power parameter is the one that holds the lowest root-mean-square prediction error.

2.2 Multi-Quadratic and Inverse Multi-Quadratic

By using Figure 2 as an example to interpolate x0, Multi-quadratic uses the following mathematical expression to achieve this interpolation aim: Z(x0)=b1×sqrt(d1×d1+c×c) +b2×sqrt(d2×d2+c×c)+b3×sqrt(d3×d3+c×c)+b4×sqrt(d4×d4+c ×c), where sqrt() means square root, c denotes the shape or smooth factor and d1, d2, d3 and d4 are the four considered distances between the four available samples and the estimation point. Therefore, b1, b2, b3 and b4 weights are estimated according to the four equations system:

- Z(x1)=b1×sqrt(d1,1×d1,1+c×c)+b2×sqrt(d1,2×d1,2+c×c)+b 3×sqrt(d1,3×d1,3+c×c)+b4×sqrt(d1,4×d1,4+c×c)
- Z(x2)=b1×sqrt(d2,1×d2,1+c×c)+b2×sqrt(d2,2×d2,2+c×c)+b 3×sqrt(d2,3×d2,3+c×c)+b4×sqrt(d2,4×d2,4+c×c)
- Z(x3)=b1×sqrt(d3,1×d3,1+c×c)+b2×sqrt(d3,2×d3,2+c×c)+b 3×sqrt(d3,3×d3,3+c×c)+b4×sqrt(d3,4×d3,4+c×c)
- Z(x4)=b1×sqrt(d4,1×d4,1+c×c)+b2×sqrt(d4,2×d4,2+c×c)+b 3×sqrt(d4,3×d4,3+c×c)+b4×sqrt(d4,4×d4,4+c×c)

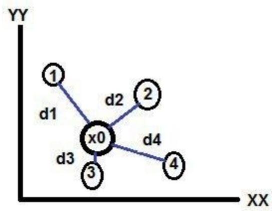


Figure 2: Logically, Z(x1), Z(x2), Z(x3) and Z(x4) represents the four observations values for those sites whereas $d_{y,z}$ equals the distance between site y and site z.

The major difference between Multi and Inverse Multi-quadratic RBFs is the internal computation formula since all remaining concepts are equivalent [11]. Under myGeoffice©, the Inverse Multi-quadratic applies the square root inverse of the sum of the squares (1/sqrt(c×c+d×d)) of the following two parameters: the c shape factor and the d distance between the available sample and the estimation site.

2.3 Moving Average

This estimation gridding is accomplished with the traditional average of all samples and whose distance between the observation and the estimation point is under the considered radius. As expected, the search radius may become wider or shorter according to each researcher resulting in different outcomes. Analogous to other interpolators, if no samples are available on the search radius for any particular site prediction, the assigned estimation will be the global sample average

2.4 Nearest Neighbor

The Nearest Neighbor gridding approach assigns the value of the nearest point to each grid node. Compared with other major interpolation techniques, this methodology suffers less from the smooth effect and it can be effective for filling data if grid maps present large holes due to the lack of available observations [13].

2.5 Voronoi & Delaunay

Both boundary drawings use external landscape features to delineate land units, assuming variations only at borders, whereas variation is null within the limits. Mostly, spatial autocorrelation is ignored between categories, where the quick assessment with sparse data is, hence, a fundamental intrinsic feature of these interpolators. Computationally, Voronoi and Thiessen vector polygons (see Figure 3) applies the best information about an unvisited point by using the nearest data prediction point, a weighted linear combination approach where all weights are given to the closest sample. Thus, polygon output depends on the sample layout, a point-in-polygon computation problem resolved by the semi-line algorithm.

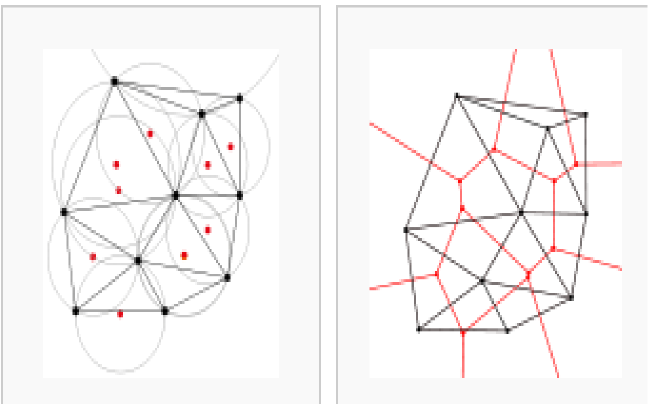


Figure 3: Relationship between the Delaunay (left) and Voronoi (right) triangulation. For reference but not covered by myGeoffice©, Splines is another deterministic interpolator that uses a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points.

3. mygeoffice.org

myGeoffice© is not a full Web GIS software package. Its functionality is depicted on Figure 4 where ten critical procedures implement event analysis, sequential geostatistical simulation, Kriging and deterministic interpolators among other options. Two newly developed spatial autocorrelation measures were also developed: the Moran location scatterplot and the variance scatterplot [5]. As well, it is possible to save generated images for post-image processing with Internet Explore® by using JHLabs®, an Web editor built on Swing technology, and Lightbox® (Java Applet)

Point analysis concerns the study of spatial arrangements of points in space under a particular time-frame whose applications includes a wide range of areas such as astronomy, ecology, biology and epidemiology. A case control study that compares point patterns of living organisms to determine if there were significant differences in their arrangements is a good example of this set of techniques. Cluster proximities matrix, Dijkstra shortest path, elevation grid mapping, Kruskal minimum spanning tree (MST), point buffering and Prim MST are six point analysis extra functions covered by myGeoffice© [5].

Figure 4: Mind map of myGeoffice.org

4. Grasshopper's Outbreaks in Utah, USA: Exploratory Analysis

Grasshoppers are an important agricultural pest that may be found in a variety of crops and ecosystems, where females may deposit from eight to twenty-five egg masses, each containing 20 to 120 eggs. Major outbreaks occur without any apparent explanation, with extensive damage during outbreak years. As expected, the cost of grasshopper control by means of a reduction in forage is also an important direct expense that ranchers must incur [8]. Quoting both authors, despite more than 120 years study of rangeland grasshopper ecology and population dynamics, we are still unable to predict when or where the next grasshopper outbreak will occur.

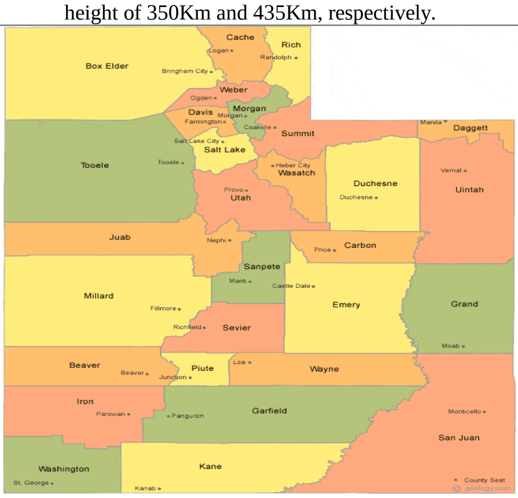
Birds are significant grasshopper predators and serve to exemplify, as well, some of the problems associated with insecticide use. Materials such as toxaphene and diazinon cause high levels of bird mortality following their application to rangeland. Others such as propoxur and azinphosmethyl cause reduction in bird numbers without direct evidence of bird mortality [8]. In other words, insecticides used for grasshopper control degrade the environment.

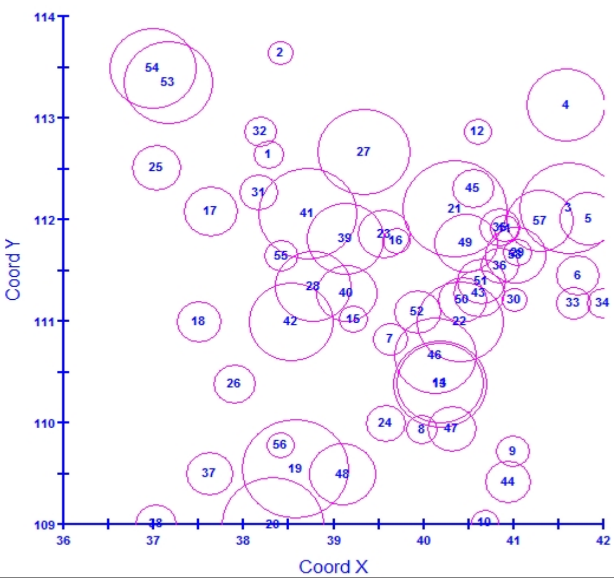
The cost of grasshopper control by means of a reduction in forage is also an important direct expense that ranchers must suffer. However, indirect costs must also be included because of the reduction in weight gain by cattle and relocation costs. Moreover, grasshoppers consume from 6% to 12% of the available forage in the western United States, although in some localities they consume essentially all available forage. At last, grasshoppers eat approximately one-half of their body weight in green forage per day [2]. Stating these researchers, with a grasshopper population of seven or eight per square meter in a four hectares field, grasshoppers consume as much forage as a cow. Still, can infestations be predicted or, at least, be mapped its current location?

The summary statistics of myGeoffice© for the present grasshoppers 2014 dataset of Utah, USA, are presented in Table 1. Several deductions should be highlighted:

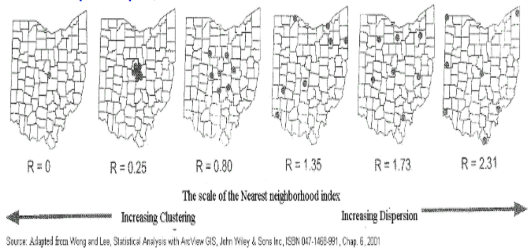
- The range of the count distribution of grasshoppers is quite high (12927-692=12235) with a Skew index of 0.669 (right hand tail) and a Kurtosis of -0877 (less values in the tails than the Normal distribution);
- The sampling location by the researcher preferences holds a 33.19 degrees direction from the North;
- The distances among samples varies quite significantly (from 0.0141 to 6.145 units, where 1 unit of myGeoffice© output represents 63 Km on the ground) and confirmed by the nearest neighborhood R index of 1.13 (the lowest this parameter tends to zero, the more concentrated the samples are layout);
- The difference between the conventional average and the estimated global mean (EGM) is quite low, meaning that the nearest neighborhood declustering approach, in this case, is not appropriate to get the bias weight of spatial representation of the input samples. Hence, the sample geographical locations do not influence the traditional descriptive measures such as variance, standard deviation or coefficient of variation. In fact, the ratio between EGM and the conventional average is close to one.

Table 1: According to Wikipedia (2018), the state of Utah is a western U.S. state defined by its vast expanses of desert and the Wasatch Range Mountains. It holds a population of 3.1 million (219,887Km2) with a length width and height of 350Km and 435Km, respectively.





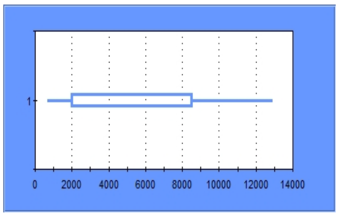
NEAREST NEIGHBORHOOD ANALYSIS
Mean distance = 2,25383001528033
Mean nearest distance = 0,360037803192084
Standard error of the mean nearest neighbor distance = 2,17047666145735E-02
Minimum distance among samples = 1,41421356237332E-02
Maximum distance among samples = 6,14556750837544
Expected mean nearest distance for a random arrangement = 0,316227766016838
Expected nearest distance for a perfect uniform point pattern situation = 0,679592442836146
Nearest neighborhood R index (0 to 2.31) = 1,13853950185043



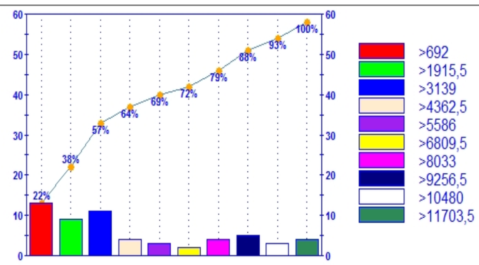
Source: Adapted from Wong and Lee, Statistical Analysis with ArcView GIS, John Wiley & Sons Inc, ISBN 047-1468-991, Chap. 6, 2001

COORDINATE'S RANGE
XX Minimum = 36,99
XX Maximum = 41,99
YY Minimum = 109
YY Maximum = 113,64

Univariate Descriptive, Nearest Neighbor Analysis, Histogram and 2D Posting



DESCRIPTIVE STATISTICS
Sample Size: 58
Conventional Average or Mean: 5189,48275862069
95% Confidence Interval of the Mean: [4233,18522155748 ; 6145,7802956839]
Conventional Standard Error of the Average: 487,906906664901
Conventional Variance: 13807082,6751361
Conventional Standard Deviation: 3715,7882979438
Maximum: 12927
Median: 4122,5
Minimum: 692
Mode: 1280
Coeficient of Variation: 0,71602286215735
Absolute Average of Deviation: 3165,32342449465
First Quartil: 2000
Second Quartil: 4122,5
Third Quartil: 8476
Interquartil Range: 6476
Skewness Index: 0,669320041446022
Pearson Index: 1,05212742092548
Kurtosis Index: -0,877960019645513

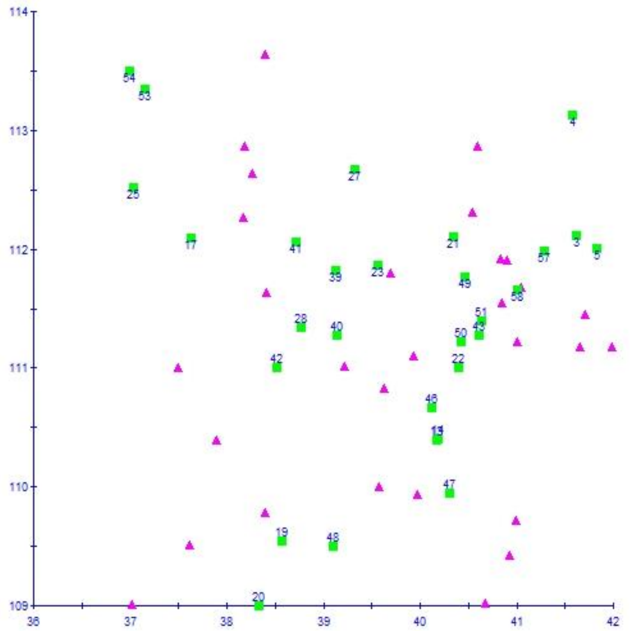


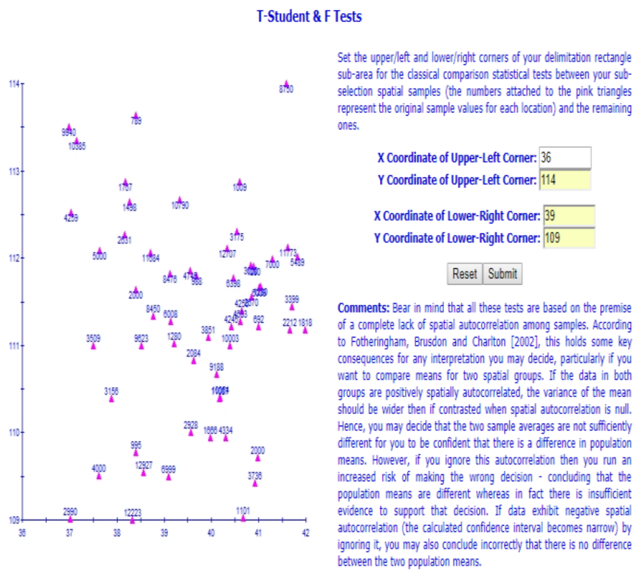
OVERALL MEAN & VARIANCE STATISTIC (NEAREST NEIGHBORHOOD DECLUSTERING)
Estimated Global Mean (EGM)= 5205,64483290816
Estimated Global Variance (EGV) = 13295567,8991212
EGM/Conventional_Average = 1,00311439020789
EGV/Conventional_Variance = 0,962952725926959

STATISTICS ON POINT DISTRIBUTION
Coordinate X of Mean Center = 39,6641379310345
Coordinate Y of Mean Center = 111,301896551724
Region Area (based on bounding rectangle) = 23,2
Region Density = 2,5
Standard Deviation Ellipse Length (XX axis) = 1,16019770671834
Standard Deviation Ellipse Length (YY axis) = 1,38687625055648
Standard Deviation Ellipse Degrees Angle (clockwise from YY axis or North) = 3,31954296103619

Under the T and F test option of myGeoffice©, a space comparison between the left (19 samples) and right (39 samples) counties was accomplished (see Table 2) since, in geographical analysis, it is often assumed that the levels of variability across the sub-regions are relatively uniform or stationary (there is only one spatial process across the entire region). Based on the present results, the parametric Student T-test equals 0.646 which is lower the critical one (2.457, df=30) for a 95% level of confidence. This means that both sets of groups comes from two identical Gaussian populations with the same mean. The F index (0.478) of ANOVA also confirms this perspective. Regarding the variance parameter, the same conclusion can be quantified, that is, the minor variance difference between both sub-regions (1.467 for a F one-tail test of df=19,39) is due to sampling error (acceptance of H0).

Table 2: The left above layout shows the spatial distribution of grasshoppers in Utah whose threshold equals the median value of 4122.5 (the pink triangles denotes lower counts and green squares symbolizes higher ones). It seems that in Utah grasshoppers outbreaks suffers between disperse and cluster pattern.





5. Grasshopper's Outbreaks in Utah, USA: Spatial Autocorrelation

The role of location (absolute coordinates and relative topology) holds two major implications on how statistical analysis should be carried out. Location leads to spatial dependence (correlation or variation that each neighbor holds in relation to a particular point) and spatial heterogeneity (clustering, concentration or proportion of neighborhood average in relation to a specific point) established by Tobler's First Law of Geography. Since regional differentiation respects the intrinsic uniqueness of each location, spatial autocorrelation can be viewed, hence, as a map pattern descriptor.

Autocorrelation also damages the ability to perform standard statistical hypothesis tests because the confidence interval estimated by the classical Pearson product moment is narrow, thus inducing to biased conclusions. The estimator standard errors are not be minimized and regression coefficients from least squares are unbiased as their variances are underestimated [4]. This occurs because new observations, under the lack of independence, do not each bring one full degree of freedom since the observer holds some prior knowledge at new locations reflecting information loss (a redundancy issue). As spatial autocorrelation approaches 1, the effective degrees of freedom approach 0 (the values become similar to the mean). As spatial autocorrelation approaches 0, the effective degrees of freedom tend to the total sampling number. As spatial autocorrelation approaches -1, the effective degrees of freedom increase beyond the number of observations.

One of the major steps in geo-statistics in order to quantify spatial autocorrelation correspond the inference of the variogram function. The most famous mathematical model are the bounded ones (Spherical, Exponential and Gaussian) because it increases with distance until they reach a maximum, named sill, at an approximate distance known as the range (see Figure 5). The sill equals the maximum variance and represents variability in the absence of spatial dependence. The range is the distance at which the spatial correlation vanishes, that is, observations separated by a distance larger than the range are spatially independent observations.

In theory, the variogram value at the origin (zero lag) should be zero. If it is significantly different from zero for lags very close to zero, this value is referred to as the nugget-effect. It represents variability at a point that cannot be explained by any spatial structure. The nugget results from high variability at short distances that can be caused by lack of samples or sampling inaccuracy. As well, anisotropy means the state of having different properties or behaviors along different directions (axes).

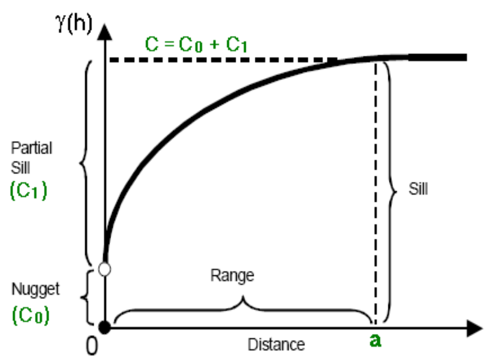
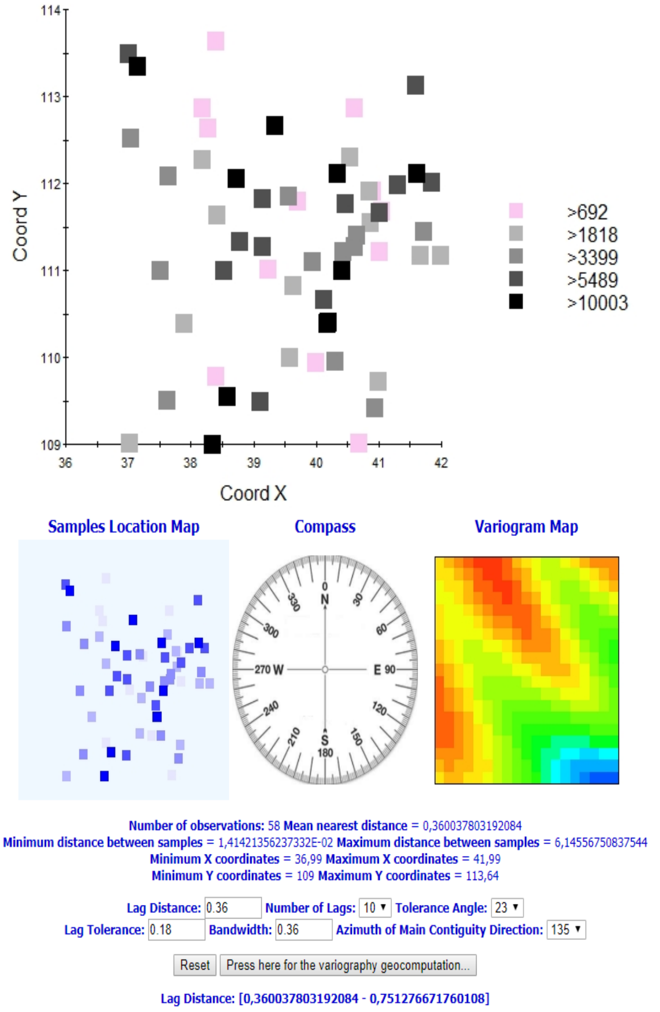


Figure 5: Example of a bounded variogram model.

Using an anisotropy angle of 33 degrees, the selected Exponential variogram for this grasshopper dataset holds the following parameters (see Table 3): Major range of 1.5 units (94Km), minor range of 1 unit (63Km), sill of 20 and a nugget-effect of 10. This also suggests that close to 2/3 of the grasshopper spatial autocorrelation happens on the first 31 Km and 21Km for 30 and 120 degrees, respectively (1/3 of the long and short range of the chosen Exponential variogram).

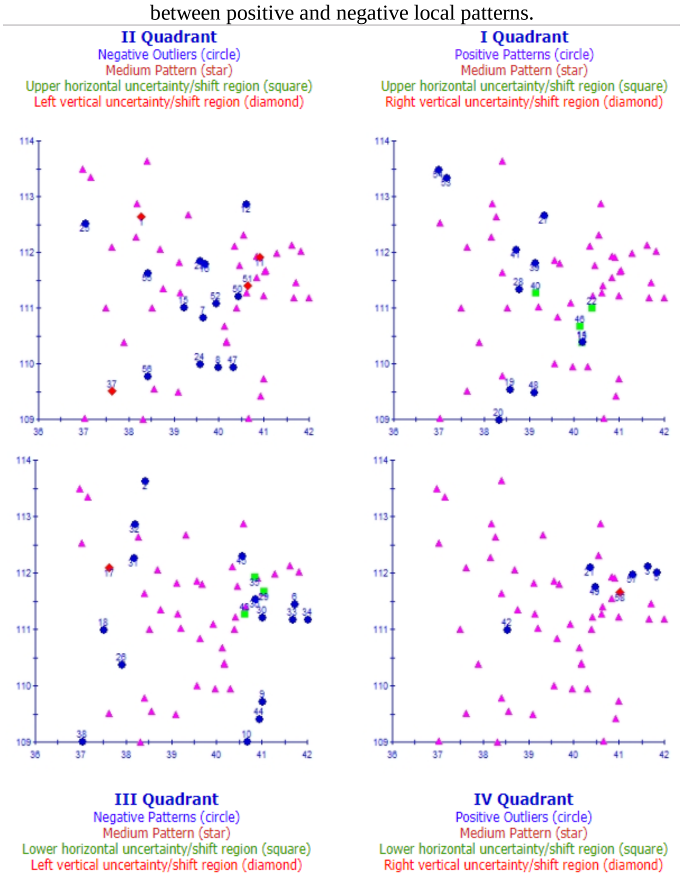
Table 3: Contrary to the temperature spatial phenomenon, this high ratio (0.5) between the nugget-effect and the sill reveals that the grasshopper's phenomenon is a non-spatial continuous one (relying on local spatial spots).



The Moran scatterplot allows the researcher to visualize and to identify the degree of spatial instability based on the bivariate regression coefficient of the spatial lagged variable (Wx), a weighted average of the neighboring values, against the original variable (x). The four quadrants, centered on the global mean, are composed of the x-axis, deviations from the original variable mean, and the y-axis, the average mean deviation of the neighborhood weight. As expected, this scatterplot is divided into four association types: upper right quadrant (high values above the mean surrounded by high values), lower left (high values surrounded by low values), upper left (low values surrounded by high values) and lower right (high values surrounded by low ones). According to Table 4 output, three major inferences can be drawn:

- Three pockets of counties (Box Elter – North West; Salt Lake, Davis and Morgan – Center; Garfield and Kane - South) with a high presence propensity of grasshoppers (I Quadrant). These are the regions that municipal authorities and agricultures should put more effort on grasshopper's control.
- Four pockets of counties (San Juan: South-East; Cache, Weber, Rich: North-Center; Carbon, Emery, Grand: Center-East; Millard, Beaver, Iron: South-West) can be considered as a safe place for crops production (III Quadrant).
- Positive and negative transitions and outliers sub-regions are spread along this state without any cluster patterning, which confirms the imperfect behavior of the experimental Exponential variogram.

Table 4: Quadrant II and IV indicates transitions and outliers districts between positive and negative local patterns.



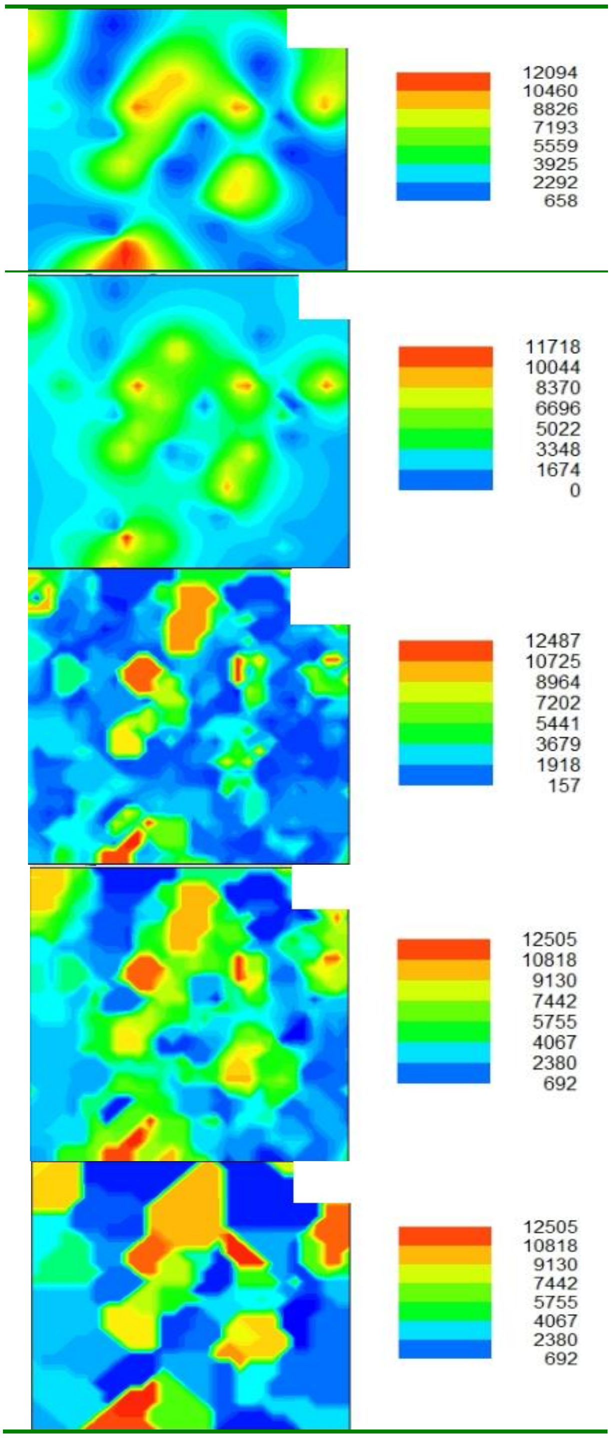
Like Sherlock Holmes, the questions that any researcher should address at this stage are, for instance, why Utah grasshoppers follow this spatial arrangement? Does the cold wind of Aspen, Colorado, shapes Utah situation? Is there any other local positive spatial pocket that should be unveiled because of the present sampling design? What are the physical and biotic disparities in the eastern, western, northern and southern regions that may influence grasshopper's outbreaks? Does rain influence grasshopper's activities? Does salinity, clay content or organic matter stimulates oviposit? Since there are an enormous variety of grasshoppers, do these results reflect all of them in the same way? By comparing levels of grasshopper infestations with other states, is there any spatial similarity characteristic to be disclosed? Can I extrapolate these grasshopper infestation levels to adjacent provinces? Is there any preferred route for grasshopper plagues? Does the direction of the prevailing wind correspond to those routes?

The prediction modelling of grasshopper dynamics should incorporate three major factors, a condition not covered by this research article: (a) The lack of food in a particular area creates pressure on its local neighbours; (b) Soil, slope aspect and vegetation have a major influence on an intermediate scale; (c) Climatic events have a greater impact on a larger one. For instance and according to [12], major epidemics occurred in 1936-38, 1957-58 and 1980-82, at approximately twenty-two year intervals. Smaller epidemics occurred midway between major outbreaks. This confirms that major droughts in the USA tend to occur at each 22 years, followed by minor ones of 11 years. Certainly, one of the add values of Geography, in general, and GIS, in particular, is to discuss autocorrelation and interpolation to find space and time rules for any spatial phenomenon.

6. Grasshopper's Outbreaks in Utah: Spatial Deterministic Interpolations

A wide diversity of spatial interpolations rises from the present dataset across Utah's state. The general trend reveals a concave drift from South to North (see the Second Order Polynomial Quadratic Trend Surface of Table 5). RBFs reveals a smoother map than all other interpolators do while Moving Average interpolates the highest values. As well, a sprinkle pattern of high and low values across Utah can be verified where the geographical center of Utah exhibits low-medium infestation values.

Table 5: Layout of myGeoffice© outputs (from top to bottom): Quadrant Global Multi-Quadric & Global Inverse Multi-Quadric RBFs; Inverse Distance Weighted & Moving Average; Nearest Neighborhood & Second Order Polynomial Quadratic Trend Surface; Voronoi & Delaunay Triangulation.



Cross-validation was used to evaluate the performance of the five deterministic interpolation methods (Table 6). As expected, all surfaces under and over predicted the grasshopper's real samples (maximum and minimum, respectively). Linear interpolation means smoothing, particular with RBFs. Unexpectedly, the simplest Nearest Neighbor deterministic approach outperformed better that the other ones, according to this criteria. Following this unforeseen trend, IDW performed worst, a situation not commonly found in other literature such as the one presented by [3] regarding the variability analysis of soil chemical properties in Cuamba, Mozambique.

Table 6: The Voronoi & Delaunay Triangulation outputs were not considered here in this analysis (myGeoffice© does not include cross-validation procedure for these approaches).

	Positive Sum of Errors	*Negative Sum of Errors*	*Total Sum of Errors*
Multi RBF	25757,57	-12881,09	38638,67
Inverse Multi RBF	36697,97	-11363,46	48061,44
Nearest neighbour	9497	-8748	18245
Moving Average	24846,2	-29231,76	54077,96
IDW	85775	-25734,42	111509,42

7. Conclusions

It is unrealistic to expect that only one factor totally accounts for trends in grasshopper abundance but weather appears to be a major driving force behind increases in grasshopper numbers [8]. For these researchers, grasshopper densities exhibits spatial patterning on two spatial scales: (a) The biotic factor on a local scale, including plant community composition, competition, density, natural enemies and population dynamics (predation and parasitism); (b) The abiotic factor on a global scale, covering habitat characteristics such as slope, aspect, soil properties, topography, precipitation and vegetation.

Naturally, a permanent observation and warning system throughout the effected region, to provide regular reports on these factors, may help individual landowners and counties, state and government authorities. In addition, the greater the positive spatial autocorrelation identified, the greater the intervention of these institutions should be. Although no cost comparisons are assessed here, the benefits of annual grasshopper surveys would most likely far outweigh the costs of the devastation the grasshoppers cause.

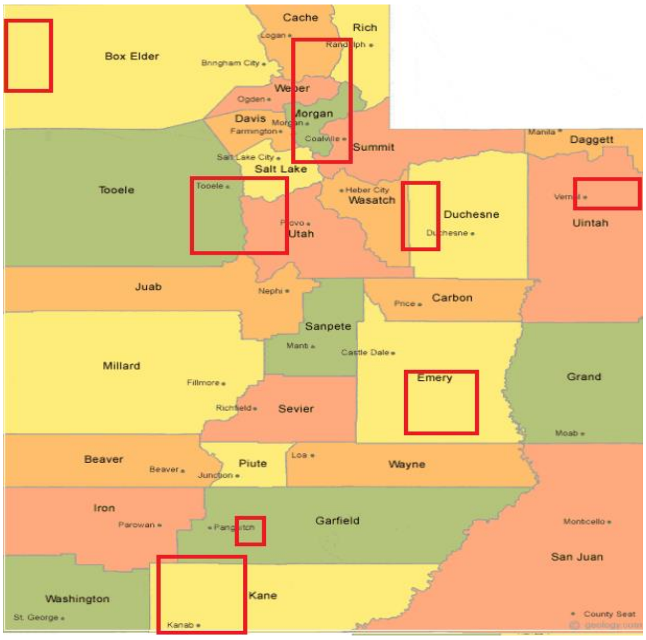


Figure 5: Sub-regions where major grasshopper's outbreaks occurred in 2014.

Regarding future predictions model, spatial-temporal scale must be incorporated in it, a limitation of this descriptive interpolation study with mygeoffice.org. For instance, grasshopper populations reach their annual peak in early July when they are in the adult stage and sharply decrease in the fall at the time of oviposit. Thus, any factor that influences the grasshopper during this period will have an obvious effect on the outbreak of the next year. Still, the goal of spatial autocorrelation is to explore, like a detective looking for evidences, while the clue's explanation ambition belongs to the ecology's experts.

References

[1] Appice, A., Ciampi, A., Malerba, D. & Guccion, P. (2013). Using trend clusters for spatiotemporal interpolation of missing data in a sensor network. Journal of Spatial Information Science, No 6.

[2] Berry, J., Kemp, W., Onsager, J. (2002). Hopper®, Version 4.0 Users Guide, www.usda.gov.

[3] Bofana, J. & Costa, A. (2017). Comparison of spatial interpolators for variability analysis of soil chemical properties in Cuamba (Mozambique). African journal of agricultural research. 12. 2153-2162. 10.5897/AJAR2016.12415.

[4] Ebdon, D. Statistics in geography. Blackwell, 1998.

[5] Negreiros, J. (2015). myGeoffice? Overview of a Geographical Information System. Proceedings of 14th ISERD International Conference, pp. 12-16, Kuala Lumpur, Malaysia, ISBN: 978-93-85832-13-0.

[6] Negreiros, J. (2017). Spatial Analysis Techniques Using MyGeoffice (Advances in Geospatial Technologies). IGI Global, ISBN-10: 9781522532705, 340 pages, USA.

[7] Norte. P., Negreiros, J., Correia, A. (2017). Cultivating Students¡¦ Reading Literacy Using Digital Lexile-Based Reading in a Chinese Primary School. CELDA - Cognition and Exploratory Learning in Digital Age, ISBN 978-989-8533-68-5, pp. 51-59, Vilamoura, Portugal.

[8] Skinner, K. & Child, R. (2000). Multivariate Analysis of the Factors Influencing Changes in Colorado Grasshopper Abundance. Rangelands, 22.

[9] Skinner, K. (2000). The Past, Present and Future of Rangland Grasshopper Management, Rangelands, 22.

[10] Souza, E. G., Bazzi, C. L., Khosla, R., Uribe-Opazo, M. A. & Reich, R. M. (2016). Interpolation type and data computation of crop yield maps is important for precision crop production. Journal of Plant Nutrition, 39(4), 531-538.

[11] Tabassum, N., Islam, S. M. R. & Huang, X. (2017). Novel Multirate Digital Filter for EEG on FPGA. In 2nd International Conference on Electrical & Electronic Engineering (ICEEE) (pp. 1-5). IEEE.

[12] University of Wyoming (2002). http://www.sdvc.uwyo.edu.

[13] Williams, M., Kuhn, W. & Painho, M. (2012). The influence of landscape variation on landform categorization. Journal of Spatial Information Science, No 5.

[14] Wong, K., Negreiros, J., Neves, A. (2016). Computer Literacy Teaching Using Peer-Tutoring: The Macao, China, Experience. Multidisciplinary Academic Conference on Education, Teaching and Learning, ISBN 978-80-88085-11-9, pp. 10-17, Czech Republic, Prague.

[15] Wong, K., Neves, A., Negreiros, J. (2017). Computer Literacy Teaching Using Peer Learning and Under the Confucian Heritage Cultural Settings of Macao, China. International Conference Educational Technologies, ISBN 978-989-8533-71-5, pp. 83-90, Sydney, Australia.

Author's Profile

Joao Negreiros Joao Negreiros is an Associate Professor of Saint Joseph University, Macao, China. He published several books and articles in English and Portuguese under the scope of Business, Technology, Education and Geographic Information Systems. His personal homepage is http://joaogarrot.rocks.

Jenny Oliveros Phillips-Lao is the Dean of the School of Business and Law at the University of Saint Joseph (USJ). She holds a PhD in Business Administration awarded with the highest honor of summa cum laude from USJ in 2015. As well, she is now a Post-Doctoral researcher at the Faculty of Human Science at the Catholic University of Portugal (UCP Lisbon) working on a research project bridging cognitive science in empathy and literature studies in catharsis focusing on the tragic theatre.