Grasshopper Outbreaks in Colorado: A Spatial Decision Support with SAKWeb©

Negreiros, J.[1], Damásio, M.[1], Painho, M.[2], Oliveira, T.[2], Aguilar, F.J.[3], Aguilar, M.A.[3]

c8057@isegi.unl.pt, mjdamasio@ulusofona.pt, painho@isegi.unl.pt, toliveira@isegi.unl.pt, faguilar@ual.es, maguilar@ual.es

[1]Universidade Lusófona, Campo Grande 376, Lisboa, Portugal

[2]Instituto Superior de Estatística e Gestão de Informação – Universidade Nova de Lisboa, Campus de Campolide, Lisboa, Portugal

[3]Departamento de Ingeniería Rural - Escuela Politécnica Superior, Universidad de Almería, La Cañada de San Urbano s/n 04120 Almería, Spain

Abstract: To analyze SAKWeb© (Negreiros, Painho, 2006) capability to perform true spatial tasks in a W3 decision support environment and to formulate a positive contribution to the grasshopper outbreak issue in Colorado, USA, are the two major objectives of this paper. This includes the same consideration of Skinner's prediction (1999) study for the Colorado 1994 grasshopper where spatial decision analysis was performed with SPlus® 4.0 while ArcInfo® 3.5 did the mapping. It is hoped do confirm some results, enhance new ones and confirm the first goal through the second one. To highlight close relationships between the Moran I correlogram and the variogram parameters are equally stressed, as well.

Keywords: SAKWeb©, Spatial decision analysis, Colorado grasshopper plague, Interpolation.

1. Preamble

A grasshopper is an important agricultural pest which can be found in a variety of crops and ecosystems. There are a hundred of species of grasshopper and, occasionally, they develop into a plague. For Capinera and Sechrist (1999), a decrease in the number of sunspots is followed by a fall in precipitation which in turn is followed by grasshopper population increases. According to the same source, major epidemics occurred in 1936-38, 1957-58 and 1980-82. This is simply a confirmation that major droughts in the western United States tend to occur at 22 year intervals, followed by minor ones at eleven year intervals. Pfadt (1977) suggests that population densities double annually until the fifth year when densities triple or quadruple. Ecologically, a grasshopper outbreak is costly and

complicated because unnecessary treatment wastes money but failure to treat, when treatment is needed, may damage the local economy. Birds are important grasshopper predators and serve to exemplify 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 cause reduction in bird numbers without direct evidence of bird mortality. Emigration from treated sites is, thus, implied.

Can infestations be predicted? This is a view legacy that fails to accept grasshopper structure heterogeneity (Skinner, 1999). For instance, grasshopper migration may occur because of a lack of suitable food. One approach is to look for temporal or spatial patterns in outbreak occurrences instead of the traditional hierarchical levels of biological organization. Another alternative is to look for correlation with auxiliary variables such as drought or sunspot activity in a spatial autoregressive model. The research presented here only concerns the former method emphasizing multiple spatial scales, because grasshopper concentration is related to both time and space.

Based on Skinner (1999) spatial data, the region of interest is the suitable plains of Colorado (regions A, B and C of figure 1). Very few samples were taken in the middle longitudes because that area corresponds to the Rocky Mountains of Colorado. Generally, there was over-sampling throughout the eastern plains (a region limited by Fort Collins, Boulder, Cripple Creek, Florence, San Luis, Trinidad, Springfield, Cheyenne Wells, Wray and Sterling) and the western area of Craig, Dinosaur, Grand Junction, Delta, Glenwood Springs and Steamboat Springs. SAKWeb© was chosen to spatially explore this 1994 grasshopper dataset in order to appraise its contribution to geostatistical software such as Regard®, GSLib®, GMS® or VarioWin® (Negreiros, 2004). To see spatial analysis as spatial modeling of a decision support for grasshopper infestation spatial planning is equally important.

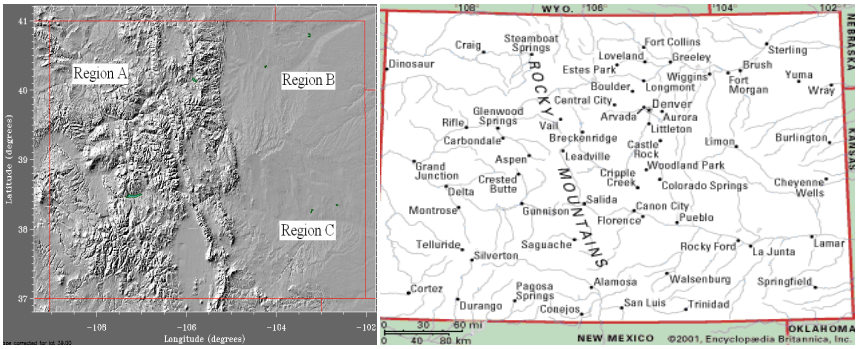


Figure 1- Topographic (left) and city map (right) of Colorado, USA.

This article is divided into further eight sections. Exploratory spatial data analysis (ESDA) is conducted in section two to determine whether grasshopper densities

exhibit spatial patterning: regional or large-scale versus field or small-scale. Spatial autocorrelation, variography and Moran location scatterplot are presented in sections 3, 4 and 5 to determine the patch size of spatial dependence among grasshopper counts. Quoting Chou (Chou, 1993), the description and analysis of spatial patterns is essential to understanding the spatial distribution of an ecological phenomenon and the processes that produce that distribution. If these spatial patterns are found then it is possible to describe grasshopper population growth and to give experts clues about the factors involved (and the decisions to be made) in the development of outbreaks. The relationship between the Moran I correlogram and the estimation of variogram parameters is also emphasized in the same section. Ordinary Kriging (OK) interpolates the map surface in section 6. The next section establishes the results of validation, cross-validation analysis and regional confidence interval procedures. By knowing the estimated and true value at sample points, the overestimation or underestimation biases of the OK technique can be determined and, thus, controlled. A research summary with concluding remarks is presented as well.

2. SAKWeb© ESDA

The 1994 summary aspatial statistics of SAKWeb© are depicted in table 1. Familiarity with the dataset is an asset in any statistical study (Isaaks, Srivastava, 1989). As expected, the preferential sampling issue is a reality on account of the frequency historical outbreaks or the value of agricultural productivity in A, B and C regions. Based on the nearest neighbourhood distance, the estimated global mean (EGM) and the estimated global variance (EGV) reveal different values when compared with the conventional average and variance. This happens because these last indexes are computed based on the nearest neighbour analysis in order to take into account the spatial samples distribution (SAKWeb© declustering method) and reflect, thus, a superior local and global distribution estimation.

The overall histograms present positive skewness, confirmed by the box-and-whisker plots (cf. figure 2, left). Since SAKWeb© does not cover any Box-Cox operations, no logarithm transformation was performed on the original dataset, a situation not pursued in Skinner's work. The frequency curve is peaked (leptokurtic) with a highly skewness value, a situation followed by many natural phenomenon.

Table 1. Aspatial descriptive and nearest neighborhood analysis of SAKWeb©

Number of samples	99	Coefficient of variation	1.45
Average (individuals per m^2)	4.9	Absolute average of deviation	4.6
Standard error of the average	0.71	Quartil 1	1.1
Variance	50.7	Quartil 2	2.1

Standard deviation	7.1	Quartil 3	5.8
Maximum	32.6	Nearest neighborhood index (0 to 2.15)	0.82
Median	2.1	Skewness	2.7
Minimum	0	Pearson	0.53
Mode	1.1	Kurtosis	7.7
Estimated Global Mean	5.06	Estimated Global Variance	57.01
EGM/Average ratio	1.03	EGV/Variance ratio	1.12
Coordx of mean center	683.7	Coordy of mean center	292.1
EGM/Average ratio	1.03	EGV/Variance ratio	1.12
Coordx of weighted geo-centroid	738	Coordy of weighted geo-centroid	243
Standard deviation distance	772.3	Region area (Km^2)	322525
Region density	0.0003	Standard deviation ellipse length (Y axis)	219.8
StdDev ellipse length (X axis)	402.8	StdDev ellipse angle (degrees)	66.6
Mean nearest distance	23	StdErr of the mean nearest neighbor distance	1.4
Expected mean nearest distance for a random arrangement	28.5	Expected nearest distance for a perfect uniform point pattern situation	61.3

When dealing with the indexes presented in table 1, the nearest average distance between samples is 23 Km with a standard error of the mean of 1.4 Km. For the same area, if all points had been arranged randomly then the expected mean nearest neighbour distance would be 28.5 Km. Likewise, the eastern plains of Colorado reveal higher densities because of the drier climate which creates more favorable conditions for the grasshopper population. The bivariate scatterplot (cf. figure 2, right) reveals a weak relationship between density and latitude-longitude for this year with high densities located mostly between 102º-104º West and 37º-39º North.

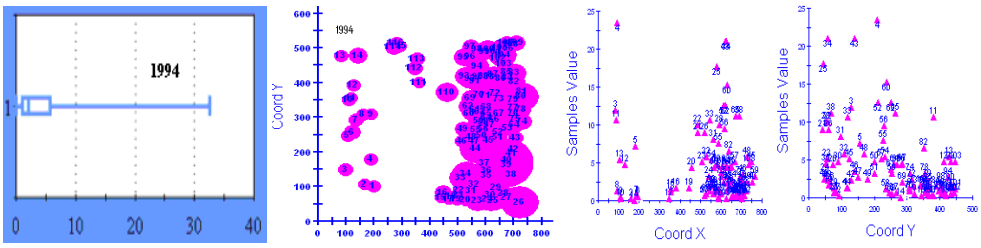


Figure 2. The box-and-whisker (left), 2D data posting (center) and bivariate (right) maps of SAKWeb©

3. SAKWeb© Spatial Autocorrelation

Variography was performed first with the goal to assess the Moran I correlogram leading, hence, to a first attempt of describing spatial continuity. Most real world patterns are somewhere between a random pattern and a clustered pattern. Grasshopper densities in Colorado follow this tendency. The Moran I correlogram does not reveal a significant positive spatial autocorrelation for which the highest autocorrelation measure is roughly 70-80 Km (cf. figure 3). Thus, a wider variogram range should be estimated for two reasons: A) The Moran I correlogram reveals a wider patch area for its highest value. B) The Moran I is low, giving the idea that stronger variation among neighbors is expected and interpolation results might be worse. Concerning grasshopper prevention and confirming Skinner work, this year is particularly important for local land owners because of the scatter of epidemics within the Colorado state due to the lack of spatial autocorrelation homogeneity.

A possible cause of this positive spatial autocorrelation on a larger scale can be a result of the smaller number of observations for a wider spatial area (300,000 Km^2, on average). There may be poor resolution and changes in grasshopper densities on scales of less than 30 Km but it is not possible to detect them because the average distance between samples is about 23 Km. If there are small hot spots then it is not possible to detect them. In fact, the length of the standard deviation ellipse, a sampling bias measure, indicates that this effect is more pronounced longitudinally than in latitudinal (3 times more so). Hence, different interpolation maps are expected according to the different nugget-effects, particularly in sampling zones. In regions without any observations, the difference among OK approaches should be minimal.

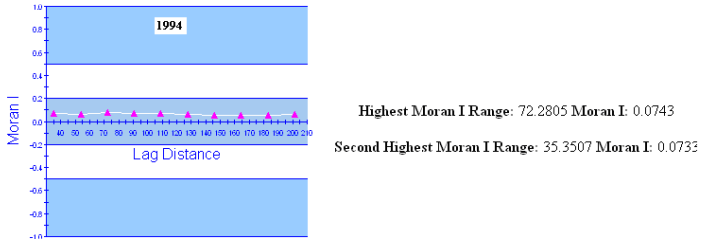


Figure 3. Moran's I correlogram of SAKWeb©

These inferences lead this analysis to the moving rectangular window for a local summary descriptive assessment including data heteroskedasticity. The need for this SAKWeb© capability is that the variogram assumes that the interpolation surface has a constant mean, a situation almost never true. Certainly, there are major differences among results, as table 2 shows. For instance, C region holds half of the grasshopper density mean of B region. The most natural way to consider accounting for the effect of the local mean is to define separate regions and treat the data within each region as a separate population (Isaaks, Srivastava, 1989). This means the local, general or pairwise relative variograms. Yet, SAKWeb® still does not cover this functionality. Moreover, the lack of samples is usually a major problem for this strategy. Definitely, the quasi-stationary assumption is a sampling density compromise.

Table 2. Aspatial descriptive results according to each region

Region	Samples Number	Mean	Standard Deviation	Skewness	Kurtosis
A	18	4.01	6.06	2.26	5.58
B	31	6.16	5.56	1.5	1.82
C	54	3.17	3.46	1.89	3.38

4. SAKWeb© Variography

If spatial autocorrelation is not significant for any distance then variography and Kriging should not be performed because the sample average is the best estimation for any location. Since this dataset do not exhibit this trend then it makes sense to carry on with the interpolation process. Otherwise, this procedure would have to finish here. In this case, the directional spherical variograms reveals an unstructured and periodic continuity: the major axis is 220 Km (and also reflected with the omnidirectional one) while the minor equals 80 Km, the distance where the Moran I correlogram is closer to zero (cf. figure 4). A long distance range was, thus, setup as an obscure average of the several directional behaviors. The sill

equals 75-80 Km, the nugget-effect is 10 Km while the anisotropy angle direction becomes 45º.

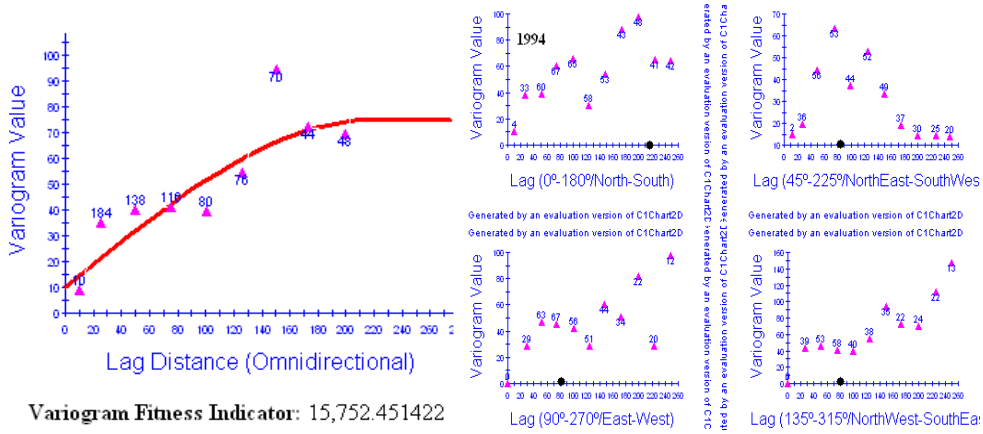


Figure 4. The omnidirectional and directional (0º, 45º, 90º and 135º) spherical variogram of SAKWeb©

Note that there is a close match between the variogram range (minor axis) and the lag distance of the Moran I correlogram. Another issue is related with the sampling method. If the variogram range changed to small distances then the OK estimation and OK variance for the Rocky Mountains in Colorado would give non-sense results because of the lack of observations within that range. By considering a range of 100Km, for example, it would be quite impossible to find samples for the variogram lags at the Rocky Mountains region. As figure 5 shows, nonsense results are given.

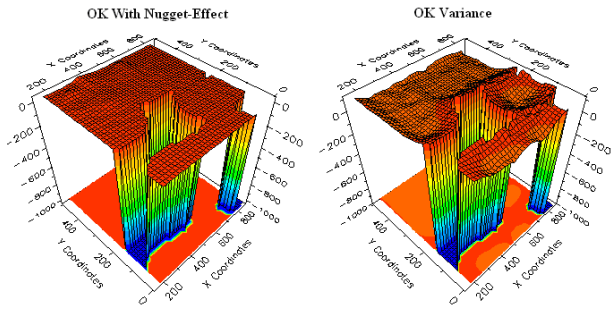


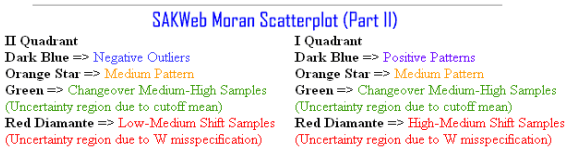
Figure 5. The OK estimation and OK variance of SAKWeb© for a 100 Km range

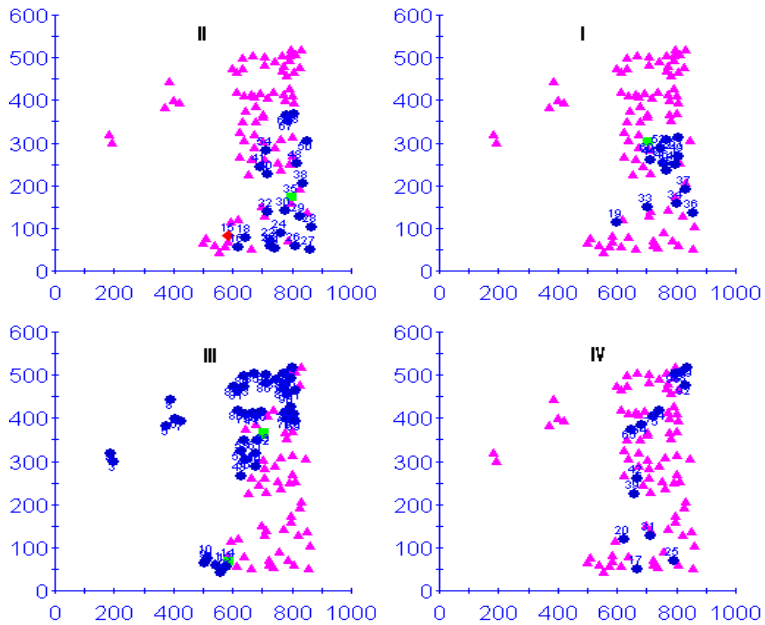
However, the most motivating issue with this approach is that the Moran I correlogram can be a useful tool for the determination of the variogram factors. It seems that the neighborhood view of spatial autocorrelation measures can work out as a best parameter guess for the distance view of spatial autocorrelation measures.

This happens because the variogram fitness can turn out to be a complicated problem on account of the uncertainty and personal thinking that underlies this procedure. Other inferences between the Moran I correlogram and the variogram parameters can, thus, be stated (Negreiros, 2004): A) With isotropic variograms, there is a direct relationship between the lag distance of the Moran I correlogram and the variogram range (x-axis). B) With geometric anisotropy, there is a close match between the minor range of the variogram and the distance of the Moran I correlogram at which reaches zero value. C) There is an inverse relationship between the positive value of the Moran I correlogram and the variogram sill (y-axis). D) Although it can be hard to detect spatial autocorrelation for shorter lag distances due to the lack of observations, a strong Moran I correlogram for small distances suggests a lower nugget-effect. E) Unlike the Moran I correlogram, the comparison among variograms sill is valid if and only if the sample sites have similar values. Quite often, lower densities lead to lower dissimilarity values. F) No variography or Kriging should be carried out for a zero and negative Moran I correlogram for all lag distances since it does not respect Tobler's Law. G) Deeper slopes in the Moran I correlogram indicate a shift from a Spherical to an Exponential model while flat profiles with a high spatial autocorrelation reveal a shift from a Spherical to a Gaussian one.

5. SAKWeb© Moran Location Scatterplot

The lack of stationarity within this 1994 grasshopper dataset uncovered by the Moran location scatterplot of SAKWeb© confirms the long range and nugget-effect of the variogram fitness procedure. According to figure 6, quadrant III highlights two separate regions with the quadrant I locations (somewhere in the middle of the Eastern Colorado state). Hence, a second order polynomial should be applied for a detrend operation.





III Quadrant
Dark Blue => Negative Patterns
Orange Star => Medium Pattern
Green => Changeover Medium-Negative Samples
(Uncertainty region due to cutoff mean)
Red Diamante => Low-Medium Shift Samples
(Uncertainty region due to W misspecification)

IV Quadrant
Dark Blue => Positive Outliers
Orange Star => Medium Pattern
Green => Changeover Positive-Medium Samples
(Uncertainty region due to cutoff mean)
Red Diamante => High-Medium Shift Samples
(Uncertainty region due to W misspecification)

Figure 6. The Moran location scatterplot of SAKWeb©

Positive and negative outliers are spread out in the map without any cluster patterning, which confirms the imperfect behavior of the experimental variogram. Since the global average of grasshopper densities is quite high and the global number of observations considered as a positive pattern is less than half of quadrant III, this suggests that sites in quadrant I are high or sites in quadrant III are quite low. Nevertheless, the existence of a strong infestation gap between both quadrants and sub-regions is a correct inference: 15.76 per m^2 against 2.27 per m^2.

Site 15 (Walsenbarg ward) of quadrant II represent an uncertainty site with a low intensity (0.3/m^2) but with neighbors that register a medium grasshopper density. If this situation was not considered then this sample will be shown as true negative outlier (low grasshopper concentrations surrounded by high values). Furthermore, the false outlier 35 (La Janta ward) represents an average infestation (4.3/m^2) with close high densities (a true outlier classification if a conventional Moran scatterplot was created).

6. SAKWeb© Ordinary Kriging

Interpolation quality is a result of the number and distribution of the known points used and how well the mathematical function correctly models the phenomena. Thus, the choice of the appropriate model is essential if reasonable results are to be obtained. Once the structural analysis (variogram) is made, the estimation can be achieved more accurately using the rate at which the variance between points changes over space.

Then again, OK variance is an excessively spatial operation although this uncertainty measure is considered essential for most users to assign confidence intervals. Nevertheless, some type of transformation must be applied to take into account the discrepancy between the sample variance and the variogram sill. This happens because, most times, the variogram sill overestimates the global sample variance. With SAKWeb©, each sample weight relies on the nearest neighbor distance between all samples and the estimated one. The estimated global mean, EGM, and the estimated global variance, EGV, are computed on the basis of this rule. Naturally, if the EGV is greater or less than the sill then variogram rescaling must be setup (note that Kriging estimates do not change with this rescaling operation). With the 1994 grasshopper, this trend is followed by the higher sill of the variogram (75) versus the EGV (57). This means a rescale operation of 0.76.

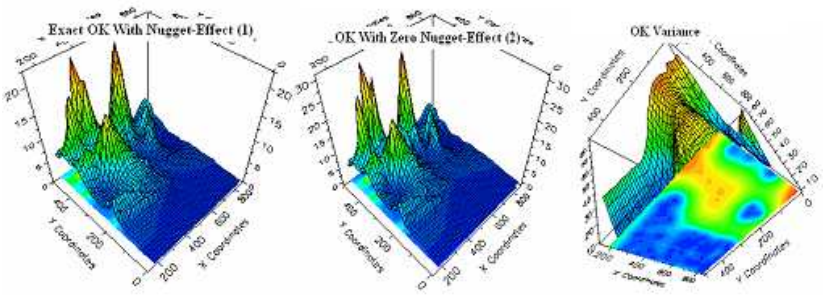


Figure 7. SAKWeb© OK (left and center) estimation and OK variance (right) with a rescaled variogram sill factor

As expected, higher OK variances are expected where there is a lack of observations (red color of figure 9, right). Technically, the existence of a nugget-effect (Co) reflects a higher OK variance, on average, against OK without Co. Another pattern is that OK with zero nugget-effect achieves a wider range for the final interpolation. At last, as the nugget-effect gets higher, the interpolation values become smoother and closer to the overall mean. Geographically, two triangle areas are of major concern: Sterling-Wray- Morgan and Cheyeanne-Burlington-Limon.

7. SAKWeb© Validation Procedures

In 1977, the U.S. Department of Agriculture suggests that treatment is justified when grasshopper numbers reach approximately 9 per m² (Skinner, 1999). Next figure presents the global confidence interval assessed by SAKWeb©. Probabilistically, the central and east region of the Burlington and Cheyenne Wells districts are the ones that land owners, county, state and government authorities should be concerned about through a permanent observation and warning system. Although no cost comparisons are made here, the benefits of annual grasshopper surveys would most likely far outweigh the costs of the devastation the grasshoppers cause. Alternatively, the west side of the Rocky Mountains in Colorado and the west districts of Stering, Yuma, Fort Morgan, Wiggins and Fort Collins do not suffer from any grasshopper outbreak in this year.

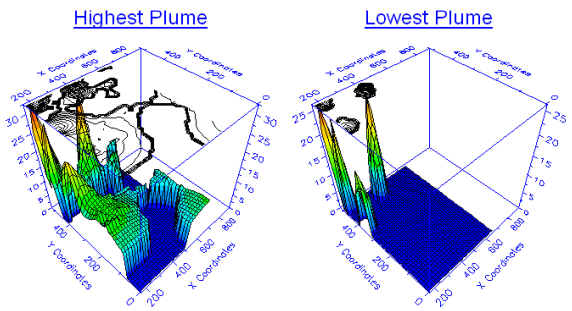


Figure 8. The global confidence interval for a level of 80% and a cutoff value of 9/m² with an OK model without Co

Although spatial data lives with uncertainty, science needs safe foundations. Uncertainty is a dimensionless parameter for which high values are bad and lower ones are optimal. Since different interpolation procedures may give dissimilar results and ground truth can never be known, it may be useful to know what the predicted chance of exceeding a given upper limit is, so decisions about expensive cleanup operations can be well founded, for instance. With agricultural applications, administrators might be interested to know-how much of the whole population would give a higher return than the value of a certain crop while, within environmental issues, supervisors might be looking at toxicity levels. Irrespective of the circumstances, the question is to determine how much of the population is likely to lie above or below a cutoff value.

If any site belong to quadrant II on an Estimated versus True grade plot, the risk α of wrongly declaring that a location is hazardous (health cost, unnecessary cleaning or false positive) becomes a reality in south-east of Colorado (in a scatter mode) and Denver, Wiggins and Fort Morgan districts. These conclusions are based on the U.S. Department of Agriculture threshold. If it falls within quadrant IV (risk β of

wrongly declaring that a location is safe), expensive consequences can be expected (false negative, potential ill health or remediation cost) at center and southern east of Colorado.

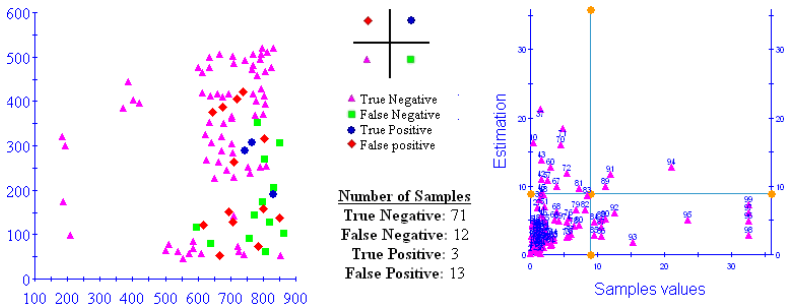


Figure 9. The grasshopper analysis of cross-validation of SAKWeb© (OK model with Co)

8. Summary

It is unrealistic to expect that only one factor totally accounts for trends in grasshopper abundance, weather appears to be a major driving force behind increases in grasshopper numbers. Grasshopper density in Colorado exhibits spatial patterning on two spatial scales, implying the influence of at least two processes in the structuring of their density in space. The question is, then, what are the factors that create and influence both scales? In essence, there are two major categories (Skinner, 1999): A) The biotic factor, on local scale, including plant community composition, competition, density, natural enemies and population dynamics and predation and parasitism. B) The abiotic factor, on global scale, covering habitat characteristics such as slope, aspect, soil properties, topography, precipitation and vegetation. Thus, the efforts to model grasshopper dynamics on any scale should incorporate both factors. Yet, these factors are likely to differ in scale where processes at lower levels are included in higher ones in a hierarchical model: A) The lack of food in a particular area creates pressure on its local neighbors; 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.

With the present dataset, positive spatial autocorrelation was detected on scales up to 75-80 Km. However, the current grasshopper survey for local infestations is likely to be missed. Therefore, the sampling for monitoring purposes should be done at small spatial lags to capture the variability in abundance on both the field and regional scales. However, some geo-questions require further investigation: Can anyone extrapolate the grasshopper infestation levels in adjacent states for this year and next? Is there any preferred route for grasshopper plagues? Does the

direction of the prevailing wind correspond to those routes? What are the links between food preference location and these spatial patterns shown in figure 7?

Any predicted model should incorporate a temporal-spatial scale because grasshopper fluctuation in population involves both space and time factors. For instance: A) 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 oviposition. B) On the field scale, local land owners control grasshoppers by spraying fields and roadsides. On a regional scale, grasshopper management is coordinated by national institutions (Skinner, 1999).

As regards Geography role, its added value is to discuss spatial autocorrelation and spatial interpolation processes to find space rules for other science experts in the context of a process of detection (Negreiros, Painho, 2005). The goal of spatial autocorrelation is to explore, like detective looking for evidence and clues, and not to describe, an analyst evaluating the clues. The latter ambition belongs to the ecology experts.

References

Capinera, J., Schrist, T. (1999). Grasshoppers of Colorado. Retrieved April 10, 2008, from http://www.wygisc.uwyo.edu/grasshopper/ghcotoc.htm# ghcocontents

Chou, Y. (1993). Critical Issues in the Evaluation of Spatial Autocorrelation. *Lecture Notes in Computer Science,* 716, 421-433, Springer.

Getis, A. & Ord, J. (1992). The Analysis of Spatial Association by use of Distance Statistics. Geographical Analysis, 24, 189-206.

Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. 483 pp, Oxford University Presss.

Griffith, D. & Layne, J. (1999). *A Case Book for Spatial Statistical Data Analysis*. 506 pp., Oxford University Press.

Isaaks, E. & Srivastava, R. (1989). *An Introduction to Applied Geostatistics*. 561 pp, Oxford University Press.

Negreiros, J. & Painho, M. (2005). The Web Platform for Spatial Statistical Analysis. The Portuguese Conference of Information Systems 05 (#92).

Negreiros, J. & Painho, M. (2006). SAKWeb© – Spatial Autocorrelation and Kriging Web Service. Geo-Environment and Landscape Evolution II, 79-89, WIT Press.

Negreiros, J. (2004). SAKWeb-Spatial Autocorrelation and Kriging Web. Unpublished Thesis, ISEGI-UNL.

Pfadt, R. (1977). Some Aspects of the Ecology of Grasshopper Populations Inhabiting the Shortgrass Plains. Insect Ecology, 73-79, University of Minnesota Techical Bulletin.

Skinner, K. (1999). Spatial and Multivariate Analysis of Colorado Rangeland Grasshopper Abundances: Pattern and Process. Unpublished Thesis, Colorado State University.