Synopsis of an Internet Spatial (Decision Support) Prototype

João Negreiros¹, Marco Painho¹, Tiago Oliveira¹, Aguilar, M.A.², Aguilar, F.J.²

¹Instituto Superior de Estatística e Gestão de Informação – Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal
²Departamento de Ingeniería Rural - Escuela Politécnica Superior, Universidad de Almería, La Cañada de San Urbano s/n 04120 Almería, Spain

Abstract - SAKWeb© pretends to be the first E-Learning research development of spatial autocorrelation and Kriging interpolation for geostatisticians. It is expected to be available in 2009 as a worldwide Web platform (named myGeoOffice.org) for all research institutes and universities that needs to carried out truly spatial stochastic and deterministic interpolation analysis. This will include simulation, morphologic geostatistics and cost analysis. At present, major spatial autocorrelation and Kriging interpolation are already offered. Founded on 1997 grasshopper dataset of Colorado, SAKWeb© major features are reviewed here with the perspective of a promising E-Learning alternative.

Keywords: GIS, SDSS, E-Learning, SAKWeb©, Spatial autocorrelation, Kriging.

1 Preamble

Progress in digital data acquisition and storage technology has resulted in the growth of huge databases. This has occurred in all areas of human endeavor, from the mundane (such as supermarket transaction data, credit card usage records, telephone call details, and government statistics) to the more exotic (such as images of astronomical bodies, molecular databases and medical records) domains. If spatial databases are at stake, the interest has grown in the possibility of extracting from them information that might be of value to the spatial user. The discipline concerned with this task has become known as spatial analysis.

Geographical Information System (GIS) can be thought of as an Executive Information System (EIS) that focuses on data access and reporting functionality for problems that involve spatial dimensions in a spatial manner [Guiseppi, 2003]. In this case, pertinent data would also include spatial information, such as regions location and their connectivity. Certainly, GIS directly supports the intelligence phase of spatial decision making. Thus, several frameworks have been developed to describe this spatial decision making process: GIS, spatial extensions of commercial statistical products and independent research software. Unexpectedly, the Internet choice was almost never undertaken.

Under the electronic business view, Internet technology is providing the enterprise infrastructure because it can be used to make information flow seamlessly from one part of the organization to another. As well, Internet standards can be used to link dissimilar systems, such as ordering and logistics tracking that previously could not communicate with each other. Yesterday, the cost of comparison shopping was very high, because people had to physically travel from store to store. At present, Internet has changed this relationship. Once everyone is connected electronically, information about products and services can flow on its own directly and instantly to consumers [Laudon, Laudon, 2002] although this situation might create a channel conflict with the firm’s traditional channels. Thus, Internet shrinks information asymmetry. In the same way, using Internet and Web multimedia capabilities, companies can quickly and inexpensively provide detailed product information specific to each customer to very large numbers of people simultaneously [Evans, Wurster, 2000], the richness-reach concept.

Similarly, Internet help B2C companies to create and capture profit in new ways by providing the foundation for new products and services such as people can make on-line bids for items like rock concert tickets or antiques, the pure-play business model. This includes m-commerce, electronic payment systems and customer support via e-mail. Another current trend is Web personalization, the capability to present to each customer a modified Web page based on that person’s purchase history.

B2B e-commerce generates efficiencies by enabling companies to electronically locate suppliers, solicit bids, place orders, and track shipments in transit. That is, trading partners can directly communicate with each other, bypassing intermediaries and inefficient multilayered procedures. This means simpler business processes, fewer employees, and much flatter organizations than in the past in a redesigned and more closely integrated framework.

From the science organization, the synergy effect of the online educational material (bibliographies, electronic newsgroups, W3 sites and E-learning software) in which the overall information may be greater than the sum of its parts is a conviction. Under the GIS view, the greatest W3 impact is to close the data access gap among users (from the preserve of the few to the right of the many) with free and direct retrieval of spatial analysis tools.

Globally, GIS holds two major components: data and model. The first element contains all spatial and non-spatial attributes. The model base holds spatial analysis processes.
and other specific tools, report maps, particular guidance for selecting decision alternatives, problem relationships and advice in interpreting possible outcomes. What-if, goal-seeking and other types of sensitivity analyses can also be used to extend or modify the original analyses and evaluations [Guiseppi, 2003]. Certainly, it is this feedback loop founded on spatial maps that increases users confidence recommendations and enable the decision maker to better explain, justify and communicate its decisions during implementation.

If GIS holds this distinct contribution to handle spatial data, what is, then, the difference between GIS and a Spatial Decision Support System (SDSS)? No direct answer can be given. The simplest perspective on SDSS definition is that a GIS is implicitly a Decision Support System (DSS), as a GIS can be used to support decision making [Keenan, 2002]. Mennecke [1997] sees SDSS as an easy-to-use subset of GIS, which incorporates facilities for manipulating and analyzing spatial data. Confirmed by Armstrong and Densham [1990], GIS lacks the modeling component needed to be accepted as a DSS. However, GIS is regarded as a form of DSS generator [Sprague, 1980] to which models can be added to make a specific DSS [Keenan, 1996].

Clark and Hosking [1986] see spatial analysis as spatial modeling of a decision support such as GADS (Geodata Analysis and Display System) for solid waste spatial planning. In conjunction with the network and spatial analysis of GIS modules, the Decision Support System Location Planner© analyzes market saturation, retail facilities accessibility, population mobility and demand-supply prediction based on demographic and socio-economic attributes, warehouse locations, distance or travel time between sites and expenditure flows between demand and supply chains [Arentze et al., 1998]. Once again, according to Openshaw [1998], an emphasis of SDSS is a convenient distraction to hide a lack of the relevant GIS technology. That is, if GIS results becomes a major input player required for some type of information decision maker then the GIS might be said to be acting directly as a SDSS. If GIS focus on specific problems with the lack interest of GIS features outside that domain, GIS cannot be regarded as a SDSS. Under both visions, SAKWeb© belongs to the second one.

In previous papers (see references), the four main reasons to develop SAKWeb© project was analyzed: (1) Geostatistical innovations, (2) E-Learning issues, (3) Poor knowledge among stochastic GIS issues, (4) Challenge to build an alternative framework to the traditional geo-modules, geo-statistical and geo-independent software such as Geostatistical Analyst of ESRI®, GSLib® or SAS-GIS®. Facing technologies (ASP®, PHP®, WebChart®, Flash®, Java Applets® and JavaScript®) and user evaluation, both issues were undertaken, too. Thus, this essay presents some global functionality and interfaces using an ancient agriculture problem of Colorado, USA: Grasshoppers infestation. To highlight some SAKWeb© features become a main concern of this paper. The last two sections summarize future prospects of this new Web platform as an E-Learning framework and some final grasshopper thoughts.

2 SAKWeb© Inner View

There are a hundred of species of grasshopper and, occasionally, they develop into a plague. For instance, major periods of drought are correlated with sunspot cycles. According to the University of Wyoming [2002], a decrease in the number of sunspots is followed by a fall in precipitation which in turn is followed by grasshopper population increases. Pfadt [1977] suggests that population densities double annually until the fifth year when densities triple or quadruple. Because of its negative historical impact on the rangelands, grasshopper habits quite often conflict with human interests. For instance, grasshoppers eat approximately one-half of their body weight in green forage per day.

The region of interest is the suitable plains of Colorado (cf. 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.

![Figure 1 - Topographic (left) and city map (right) of Colorado.](image)

The capability to retrieve data from MS-Excel® into SAKWeb© can be achieved by the Insert/Name/Define option although plain files can be used, as well. All processing records must respect the following pattern: Sample ID, Coordinate_x, Coordinate_y and Sample_value. With the First Order Trend Surface with Join Count Residuals option, this general linear trend model is shown in a 3D perspective embracing nine statistics: b0, b1 and b2 parameters of the linear polynomial regression, total,
regression and residual variances, adjusted $R^2$, raw $R^2$ and $F$ test (cf. figure 2). The Descriptive Measures and Nearest Neighbourhood Analysis presents the sample histogram and the cumulative one. A series of univariate descriptive measures are also displayed such as standard error of the average, variance, coefficient of variation, skewness, estimated global mean, standard deviation of distance, region density and expected mean nearest distance for a random arrangement.

According to the box-whisker plot, a highly skewness value was founded, a situation followed by many natural phenomena. According to Skinner [1999], the lowest density of grasshoppers when compared with the previous 5 years are closely related with the precipitation of 1997 since moist conditions increase mortality rates. Likewise, the eastern plains of Colorado reveal higher densities because of the drier climate which creates more favorable conditions for the grasshopper population.

SAKWeb© Moran I Analysis option depicts the Moran I correlogram for several distances and determines the distance lags for the two highest statistics of Moran I (cf. figure 3). Actually, this correlogram indicates that there is a smaller spatial pattern of density then previous years [Negreiros, 2004].

At the heart of geostatistics, the variogram indicates how different the values are as distance increases in a specific direction. Three major parameters describe this tool: (1) the range or lag distance at which all successive values are independent of each other, (2) the sill, the variogram value corresponding to the range (a length measure of the heterogeneity correlation), and (3) the nugget-effect, the gradual change near the origin indicating strong spatial continuity, while a sudden decrease in the variogram points out a scale variability situation. To fulfill this goal, the Variogram Setup requires six parameters (lag distance, tolerance lag, number of lags, horizontal bandwidth, principal axis angle and tolerance angle) to estimate the experimental variograms.

User interactivity was not forgotten, either. With the mouse control, for instance, the Local Interactive Statistics holds the capability to select a continuous subset of points to compute local statistics (i.e. average, standard deviation, coefficient of variation and kurtosis), which helps checking local heteroskedasticity within the dataset.
Forgetting the geostatistical view, the state of Colorado can be described as one with small spatial patterns and big concerns for local land owners. Several negative and positive local spots in a scatter arrangement can be found in the Moran location scatterplot, leading to a experimental variogram structure with a short range. Furthermore, the central-east region of Wray, Limon and Cheyenne Wells presents some grasshopper infestation fluctuations which hold a negative impact on the accuracy of the final estimates. According to Isaaks and Srivastava [1989], these transitional regions are more disposed to misclassification for a particular cutoff limit, with major economic consequences, than regions classified as exhibiting positive or negative patterns.

Figure 5 – The SAKWeb© Moran location scatterplot, a local exploratory spatial data analysis tool where each point represents samples locations.

However, its low global mean makes this year the most tranquil one for farmers and local state funding. In fact, the positive outliers in quadrant IV (cf. figure 5) are lower than in the previous years with a range of [2.9, 7.6] and a global average of 4.19 grasshoppers per m$^2$. Compared with the 1994 dataset, for instance, the range equals [5.4, 32.6] while the positive outlier average equals 11.12 grasshoppers per m$^2$.

The Global and Local Region Confidence Interval option estimates the minimum and maximum plume based on the rescaled OK variance for a particular confidence level. This option also determines the largest and the smallest regions, for the same confidence level, on the basis of a particular cutoff limit, although the smoothing effect of Kriging estimates can be inadequate for high or low threshold choices. On the basis of the Normal error distribution, sill variogram rescale and OK variance, SAKWeb© simulates the minimum and maximum plume for three confidence intervals (80%, 90% and 95%) against three OK models (OK with two structures and OK with without C0).

Figure 6 – A partial view of the Global Region Confidence Interval option of SAKWeb©.

Based on the U.S. Department of Agriculture, it is suggested that treatment is justified when grasshopper numbers reach approximately 9 per m$^2$. By overlapping figure 1 against figure 5, the central and east region of the Burlington, Limon 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.

At last, the 3D-2D OK Profile option is another attractive interactive option because it allows the user with a 3D surface generated by Ordinary Kriging to outline the 2D profile between 2 points (cf. figure 7).
3 The E-Learning Perspective

E-Learning does not pretend to replace traditional teaching methods but it constitutes a good complimentary one. According to Painho et al. [2002], the major features for this alternative are: (1) It targets an audience that live teaching cannot reach, particularly for long-life learning whose students may have a full time job or lived far away from urban centers; (2) Global cost reduction (both students and universities); (3) Teachers are forced to structure their scientific knowledge regarding their subjects; (4) Students can learn anytime anywhere at its own rhythm; (5) Implementing the technological structure that supports E-Learning platform is a scalable solution (it is always possible to increase the number of servers that support the course); (6) People without minimum technologies competences are excluded; (7) Teachers need to have the right competence and skills to reflect a strong motivation regarding students because there is a shorter interaction between both.

Unquestionably, E-Learning enables the possibility of reviewing traditional teaching methods and experience new learning approaches. Knowledge diffusion is facilitated and, thus, the use of Web applications such as SAKWeb can fulfil this need for any Bachelor or Master GIS degree, for instance. Expected to be available in 2009, myGeoOffice.org personifies this belief. Simple, universal and indicator Kriging, Cokriging, sequential and indicator simulation, morphological geostatistics and costs analysis is already on the way. Its future use by geostaticiens might constitute another major goal of this Web application. The preliminary analysis of the survey presented in Negreiros [2004], for instance, announces a strong confidence for this Internet solution. At last, it is an encouraging aspect the fact that students and colleagues from ISEGI-UNL, Portugal, and Universidad de Almeria, Spain, are both interested in contributing in its evaluation and participating in its technical problems resolution.

4 Grasshoppers Final Thoughts

Parallel to previous years, the central and east region of the Burlington and Cheyenne Wells districts registers high levels of grasshopper infestation. In addition, the west side of the Rocky Mountains in Colorado and the northern and west districts of Sterling, Yuma, Fort Morgan, Wiggins and Fort Collins do not suffer from any grasshopper. It is the cold wind of the permanent Aspen ice that is responsible for this pattern? Is there any other local pocket to be taken, though it is not shown by this sampling design? What are the disparities in the physical and biotic factors in the eastern and western region that may influence the grasshopper’s outbreaks in the east? Does it rain more or is it warmer in the eastern region than in the west? Should more samples be taken in the eastern districts to confirm the permanent low densities of grasshoppers? In terms of soil attributes such as salinity, clay content or organic matter, how does oviposition differ? Since there are an enormous variety of grasshoppers, do these results reflect all of them? Comparing levels of grasshopper infestations with those in other countries, is there any spatial characteristic to be revealed? Can the spatial user extrapolate the grasshopper infestation levels in adjacent states for this and the following years? Is there any preferred route for grasshopper plagues? Does the direction of the prevailing wind correspond to those routes? Is there any long-term cycle of grasshopper infestations that cannot be captured by the present dataset?

According to the University of Wyoming [2002], major epidemics occurred in 1936-38, 1957-58 and 1980-82, at approximately 22 year intervals. Smaller epidemics occurred midway between major outbreaks. 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 11-year intervals. Certainly, the added value of Geography is to discuss spatial autocorrelation and spatial interpolation processes to find space rules for other science experts in a continuum process context of detection.
5 References


