# MYGEOFFICE©: OVERVIEW OF A WEB GEOGRAPHICAL INFORMATION SYSTEM

**JOAO NEGREIROS**

University of Saint Joseph, Macao, China
E-mail: joao.garrot@usj.edu.mo

**Abstract**- Mygeoffice© is a Geographical Information System for the Web. Analogous to any SaaS (Software as a Service), this paper highlights the first Web application to offer spatial autocorrelation measures, stochastic simulation, variography, Indicator Kriging, for instance. By exploitingASP (Active Server Pages), PHP(Personal Home Page), Java Script, Java Applets and IIS(Internet Information Service) capabilities, mygeoffice.org was designed in an attractive and straightforward way for GIS users, in general, and environmental users, in particular.

**Index Terms**- Geographical Information Systems, Web GIS, Geo software, myGeoffice©.

## I. INTRODUCTION

This paper introduces myGeoffice.org, an innovative WWW platform for Geographical Information System (GIS) users. It is commonly accepted that GIS is an integrated technology that has resulted from the fusion of various systems, moving from a cartography view to one of virtual reality modeling. At first, spatial analysis shifted from a classical exploratory and inferential statistical framework to a spatial analysis methodology embracing only non-spatial data. Overlay, networking, measuring, raster and vector buffering, spatial search, raster spreading or contouring operations were the following step in a common sense spatial approach (analytical cartography analysis). In its present second phase, advanced spatial analysis is in demand of interpolation, extrapolation, stochastic simulation and spatial-time analysis where the use of both topology and non-spatial information is mandatory. Without it, spatial analysis loses its identity. After all, spatial data is special [1].

Statistical packages are a possible software implementation to relate GIS with the existing statistical packages such as SAS-GIS®, Glim®, Systat® or Minitab®. The difficulty involved in this solution is that statistical packages do not provide the GIS functionality needed because they were not developed to handle spatial data and the structure of spatial data switching back and forth [2]. The requirement to learn more than one package can also be problematic.

A second implementation strategy is to add statistical functionality into commercial GIS by modular integration such as Geo statistical Analyst of ESRI®. A major advantage is good software documentation and ready availability of spatial analysis functionality. Curiously, users prefer to buy complete systems. According to [3], ninety-two people were surveyed in some prestigious Portuguese institutions and 80% of them felt more comfortable working with a known commercial product than unknown independent software. The common user trusted the software company trademark attached to it blindly because (probably) of the brand marketing that lies behind the software.

For simple descriptive analysis, the potential of spatial analysis within GIS is of little interest. Yet, regarding nearest neighbor methods, connectivity measures directly derived from the topology matrix can be extremely important. After all, road network distances may be more relevant than straight line distances. The W weighted matrix specification for autoregressive functions is another example of the greatest linkage potential. Since topological relationships are already stored within GIS, this goal can be achieved with little effort. On the other hand, major GIS module deficiencies include their financial costs and the availability of new tools will only depend on software developers and not on the research community.

With independent geo software, the titanic issue is the effort required to rewrite a complete package from scratch, reflecting a well-known dilemma: customization versus time/money savings of existing software modules. GRASS® and SAGA® have been considered two of the finest pieces of GIS independent software nowadays. Another highlight should be given to SPRING®, a non-cost software developed by the National Institute for Space Research, Brazil, which includes a number of remarkable innovations such as an object-oriented data model for spatial data and an object-oriented data query and manipulation language.

A careful review of all pros and cons of these three software implementations (GIS, independent software and spatial extensions of statistical products) does not give much of a clear trend for the near future and, in the end, the future is partially controlled by the past. Quite often, software solutions start on the end-users needs. Because of the object embedding capability of Windows®, Linux® and Macintosh®, the potential of computer languages of today overlaps quite often, leading to the possibility, for instance, of assessing regression estimators or visualizing the same plots/maps with different computer codes. Although certain tools fit best for a specific goal, programmers common use their background and field of expertise in their creation process, reflected on several types of implementations. And as the new academic methods for spatial analysis emerge, more programming is needed.

In terms of speed and performance, executable languages are certainly the programmers' best choice especially when integrated with the standards of .NET and Java. From the user's viewpoint, this fuzzy trend becomes clear though. The aim for non-expert and expert users is different: the former are looking for simple descriptive statistical tools and data property summarization while the latter requires spatial auto-regression and hypothesis testing. According to [4], GIS for utility applications requires network processing operations while geological ones demand 3D views. Thus, the same GUI should offer various possibilities for a wide range of users.

Software should encourage scientific computing visualization where statistical views are interpreted, validated and explored in a dynamic and linked graphic mapping. Still, these graph operations on the Internet are still quite difficult to achieve. Another Web complexity is the access to topology information (myGeoffice© suffers from these constraints, for instance).

Furthermore, the computer GIS environment should not embrace a completely automatic procedure. This means users must be aware of decisions at each stage of being confronted with an explicit interface. As a result, the stages have to respect a linear sequence where practitioners begin at the beginning and end at the end, progressing from left to right across menu choices. Therefore, what the users appreciate and interact with is the graphical user interface (GUI), because users do not care about the technical structure, as long as the results are trustworthy, prompt and compatible with their operating system and hardware. Users want intuitive and easy-to-use software in order to give immediate results without having to read pages of documentation.

The standard Web browser, regardless of the background computer code adopted, fulfills this strategy quite well. It is cost free and already provided with any operating system. Does the user need to see a Word document? It can be viewed in a browser. Does the user need to work on an Excel spreadsheet? It can be opened in the browser. Does the user need to find a local or a network file? The browser can search for him/her. Ultimately, the user will work with all available software in the same way, regardless of the location of the data and its purpose: in the beginning, the trend was to find the data with the browser (Web 1.0). At present, just run the programs/services from the Internet (Web 2.0).

This represents a significant advance on interfaces because users will no longer have to worry about the software location and the technical knowledge necessary to connect to their data. Even today, the keyboards available on the market hold special keys for the browser such as home, search, back, mail and refresh. The standard Web browser is the interface. The capability to undercover technical implementation to the final user via WWW is essential. After struggling for years for digital information, spatial analysis needs to concentrate on what the information means and sharing it through the World Wide Web. The shift from mainframes to Apache® intranets and wireless transmissions via PDAs, laptops or smart phones is already the way. In effect, the wireless Web is developing furiously nowadays enabling millions of people to access the Internet while on the go. Therefore, improving spatial exploration by applying the available tools of spatial analysis within the Internet becomes critical.

## II. MYGEOFFICE©

As an evolution of SAKWeb©, myGeoffice© is an Internet software that provides access to a wider GIS audience by undertaking the Web solution platform which includes the Moran scatterplot, Kriging with measurement error and several nugget-effect solutions, 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. myGeoffice© is not a comprehensive statistical package in the traditional of solving everyone's problems. Written for an Internet Information Server® (IIS) environment, it was developed with the philosophy of being a learning tool for individuals with limited geo statistical knowledge. myGeoffice© deals with interpolation and simulation in conjunction with spatial association measures in a Web continuum process (second menu) and a loose of local spatial functions (first menu), as well. From this viewpoint, an element of its originality and innovation can, thus, be appreciated.


**Fig.1.Homepage of myGeoffice.org**

For space visualization purposes, the near future concerns about the creation of virtual reality worlds and 3D Web pages which offer an additional possibility compared to 2D representations. The use of virtual reality in major fields such as medicine, military flight simulators and industrial applications has been increased significantly.

By definition, X3D is the integration of a 3D Application Program Interface (API) in a Web environment using XML concepts. Its main goal is to allow the creation of Web virtual worlds with interactivity where VRML 97 was its predecessor. It was developed by the WebW3 consortium (web3d.org/x3d) for integration of 3D graphics and multimedia contents. Broadly, lower bandwidth, navigation and exploration are some features of this API. Specifically, X3D supports the inclusion of graphics (polygonal geometry, lighting, materials, textures, hierarchical transformations), animations (timers, interpolators to drive continuous simulations), audio and video, user interactions (mouse picking and dragging, keyboard input), navigation (cameras, collisions, visibility detection), scripting for programming (Scene Access Interface) and the ability to compose a single X3D scene with other assets located in the World Wide Web.

Certainly, geo referenced information may benefit from X3D concepts which may increase users' engagement and motivation to acquire spatial information. Following this trend, a plausible future of myGeoffice© is to generate a 3D space travelling based on the present generated 2D maps is highly desirable, if further advancements of this Software as a Service (SaaS) will be taken.

A central component of GIS is technology management. "There is a world market for maybe five computers" was a famous statement by the IBM® chairman in 1943. He moved on a few years later and IBM® was saved. Ken Olson, from DEC® Digital Corporation, the second largest computer manufacturer in the seventies, did not survive ("There is no reason why anyone would want a computer in their home"). When he woke up, Open VMS® was already dead and DEC® had been bought by Compaq® Corporation. Bill Gates recovered from the "640K is enough for everyone" statement over time.

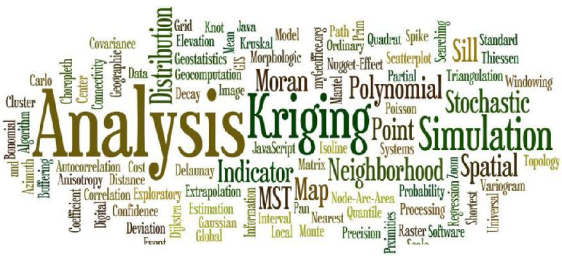


**Fig.2.The main word clouds closely related with myGeoffice© (created with wordle.net)**

It is hoped that myGeoffice© will not follow this trend. To make this project come to life, several different software packages were used such as Active Server Pages (ASPs 2.0), IIS, PHP®, Fireworks®, WebChart®, Java Applets®, JavaScript and HTML, Fortran and C language. An international domain name, mygeoffice.org, was another requirement to fulfill a full interaction of this Web 2.0 service around the globe. And if the Internet is part of our lives, in the end, the common practitioner will inherit all WWW intrinsic features such as the absence of site management obligations and installation procedures. After all, GIS expects the results to be displayed in appealing maps while the arduous work should be hidden in a W3 wireless environment. myGeoffice© follows this belief.

## III. MYGEOFFICE© MAIN FEATURES

MyGeoffice© is not a full GIS software package. Its functionality is depicted on the basis of two sub-menus and ten critical procedures that implement known event analysis, sequential geo statistical simulation, Kriging and deterministic interpolators among other options. Twonewly developed spatial autocorrelation measures are also developed: the Moran location scatterplot and the variance scatterplot.

The description of each procedure is outlined next and according to the next figure:

- Help, an overview of the aids given by myGeoffice©. As a rule of thumb, its philosophy towards aid consists of directly offering, in each option, a set of definitions, notes, links, examples and tips to help the regular practitioner fulfill his/her duty.
- By definition, raster image processing consists of the capability to perform some operations on raster pictures in order to get an enhanced illustration or to extract some useful information from it. This embraces analyzing and manipulating the matrix image, including data compression and image enhancement such as sharpening and restoration. Within myGeoffice©, it is possible to save generated imagesfor post-image processing purposes using JHLabs®, an Web editor built on Swing technology, and Lightbox®.

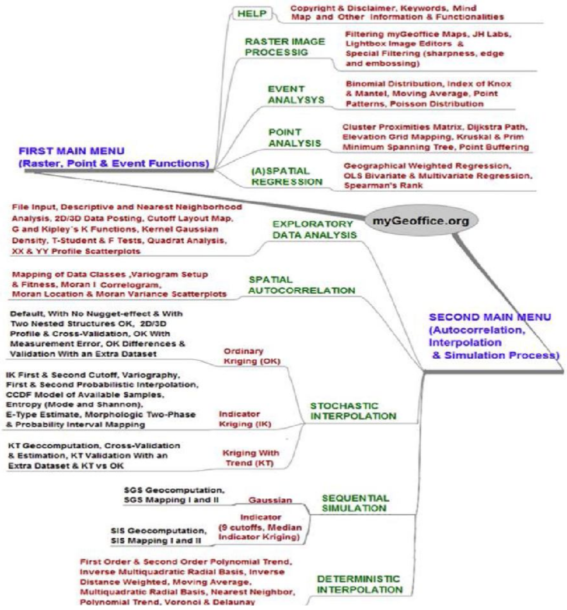


**Fig.3.The overall structure of myGeoffice©**

- Point analysis concerns the study of spatial arrangements of points in space (usually 2D) 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 functions covered by myGeoffice©.
- The aim of linear regression analysis is to uncover which variables contribute in a significant way for a linear relationship of the dependent variable. It is expected, among other factors, that regression errors are independent (within spatial processes, this aspect can be measured by the error term using the Localized Autocorrelation Diagnostic Statistic (LADS), Moran I and Joint Count Statistics, for instance). As a result, Ordinary Least Square (OLS) model breaks down quite often due to spatial autocorrelation existence.The global spatial lag model for stationary processes ($Y=\rho WY+X\beta+\varepsilon$, where W represents the neighborhood matrix while $\rho$ is the spatial autoregressive coefficient) and the spatial regimes (individual regression for each sub-region), developed within SpaceStat®, are two possibilities. An extra possibility is the local continuous variation framework computed by the Geographically Weighted Regression (GWR). The idea is to adjust a regression model (the βs of the regression model are not the same everywhere) by weighting the neighborhood observations. In this way, the estimation computation will reflect automatic adjustments, according to the distance and values of the available samples. myGeoffice© develops this matter while the next two options review the conventional bivariate and multivariate OLS regression, respectively. Spearman Rank Correlation for ordered data is also computed.

**Fig. 4. Partial view of the first menu of myGeoffice.org**

- The goal of event analysis is to understand (and perhaps to predict) spatial occurrences that happened in a particular time-frame. Thus, five purposesare highlighted within myGeoffice©: Binomial for two outcome types of events (any phenomenon that can be observed such as rain or not rain), Knox and Mantel indicators, Moving average, Point patterns and Poisson distribution.
- Exploratory data analysis (EDA) of myGeoffice©tries to summarize datasets main characteristics, such as nearest neighborhood indexes, standard deviation, scatterplots or quadrat analysis, for instance.
- If the finding of spatial structures is fundamental for Geography then spatial autocorrelation subsist as a major technique for a better understanding of spatial relationships among spatial entities. myGeoffice©respects this conviction by covering variogram setup and fitness, Moran I correlogram, an innovative version of the conventional Moran scatterplot, and the recent Moran variance scatterplot [5].
- Spatial interpolation, in general, and stochastic Kriging, in particular, becomes then the next capability addressed by myGeoffice©. This means Ordinary Kriging (OK) and Kriging with Trend (KT). Both options hold five sub-options: geocomputation, estimation mapping, cross-validation procedure, validation with an extra dataset and KT versus OK comparison. At last, Indicator Kriging (IK) is also dissected with nine sub-options: First and second cutoff definition, first and second probabilistic interpolation maps, construction of the conditional cumulative distribution function, entropy of Shannon, E-type spatial estimation (including misclassification risks and economic classification), morphologic geostatistics and probabilistic interval mapping.


**Fig. 5. Partial view of the second menu of mygeoffice.org**

- Unlike Kriging, stochastic simulation does not aim for the minimization of local error variance but, instead, focuses on the reproduction of statistics such as the sample mean, variance, histogram or variogram model (in other words, it preserves the spatial variation of the studied attribute). In addition, it honors observation data values. Thus, stochastic simulation is increasingly preferred to Kriging for applications where spatial variation of the measured field must be preserved and the extreme behavior of the variable under study should not be overlooked [6].Conditional Cumulative Gaussian Simulation and Conditional Indicator Simulation are the two possibilities enclosed by myGeoffice©.
- The deterministic interpolation approaches of myGeoffice© regards the first and second order polynomial, multiquadratic and its inverse, inverse distance weight (IDW), moving average, nearest neighbor and triangulated irregular network (Voronoy and Delaunay).

## CONCLUSION

Nowadays, the common user begins to handle and understand spatial autocorrelation and Kriging, for instance, and the difficulty of getting specialized GIS software that covers these subjects. It is important to forget installation procedures and operating system requirements, to embrace a direct, friendly and known GUI with mobile access, to incorporate standard exchange for data input, to validate input parameters including suggestions and hints on-the-fly, to add new features available and to provide the capability of discussing these issues on an on-line basis with good hypermedia help [7].

With the advent of Web technology and modern wireless computing, it became necessary to develop a W3 service for GIS to understand the often complex spatial relationships that exist among the samples collected in space. myGeoffice© tries to respond to all these challenges. It is just hoped that, somehow, the end-userswill advocate this belief into reality.

## REFERENCES


[1] PennState. (2014). Lesson 1: Why Spatial Data are Special - Problems and Opportunities. Retrieved 2014, from Geography 586: www.e-education.psu.edu/geog586/l2.html.

[2] Negreiros, J., Painho, M.: SAKWeb© – Spatial Autocorrelation and Kriging Web Service, Geo-Environment & Landscape Evolution II, WIT Press, p79-89, ISBN 1-84564-168-X(Jun 06).

[3] Negreiros, J. (2004). SAKWeb - Spatial Autocorrelation and Kriging Web. Ph.D. (unpublished) Thesis.

[4] Worboys, M. (1996). GIS: a Computing Perspective. Taylor & Francis.

[5] Negreiros, J., Painho, M., Aguilar, F., Aguilar, M.: A Comprehensive Framework for Exploratory Spatial Data Analysis: The Moran Location and Variance Scatterplots. International Journal of Digital Earth, Taylor & Francis, Huadong, G. (Eds), ISNN 1753-8955, DOI 10.1080/17538940903253898, pp. 1-30 (Dec 09).

[6] Negreiros, J., Costa, A.C., Painho, M.: Evaluation of Stochastic Geographical Matters: Morphologic Geostatistics, Conditional Sequential Simulation and Geographical Weighted Regression. Trends in Applied Sciences Research, Science Alert - Asian Network for Scientific Information, 6 (3), pp. 237-255, ISSN 1819-3579, DOI 10.3923 (Jun 10).

[7] Negreiros, J., Painho, M.: SAKWeb© – Spatial Autocorrelation and Kriging Web Service (Part II), Internet Research 7.0: Internet Convergences, Brisbane, Australia (Sep 06).


★★★