

Resumen

Este trabajo final de master trata acerca de la construcción de una librería de estimación geoestadística que permita la consideración explícita de la localización espacial de los datos de entrada y los resultados obtenidos mediante su georeferenciación a partir de sistemas de coordenadas geográficos o planos estándar. Para lograr dicho objetivo las entidades espaciales susceptibles de integrar en un análisis estimativo se conceptualizan como *objetos* computacionales que tienen asociados diferentes tipos de *propiedades* medibles, observables y/o calificables. A dichos objetos se les puede asociar ciertas localizaciones referidas a un origen de coordenadas geográfico. Así, desde el punto de vista computacional se utiliza la programación orientada a objetos con el fin de modelar entidades computacionales que representen dichas entidades espaciales reales.

Inicialmente, se trata el desarrollo de los objetos computacionales que manipulan la información de entrada y de salida a los algoritmos implementados. En general, dado que la información utilizada en la geoestadística se basa en datos (mediciones) localizados puntualmente o se encuentra densamente muestreada, los *objetos* computacionales espaciales que se modelan son entidades puntuales y campos formato raster. La creación de los *objetos* computacionales se implementan en *clases* individuales que hacen posible la definición de su propiedades, su localización y su sistema de georeferenciación en los *constructores* asociados. Dado que el *objeto* debe representar características espaciales definidas, para una misma entidad espacial (por ejemplo localizaciones puntuales de mediciones) se generan diferentes tipos de objetos de acuerdo con las dimensiones del análisis: 2D o 3D. Para la manipulación y operación con los objetos modelados computacionalmente se generan los respectivos *métodos de clase*, que pueden ser *privados* o *públicos*. Los *métodos públicos* pueden ser utilizados efectivamente para nuevo desarrollo del software, por lo cual a lo largo del trabajo se hace referencia exclusivamente a ellos.

El aspecto siguiente que se trata es la implementación de los algoritmos de análisis de información: el análisis estadístico descriptivo de las muestras de datos y el análisis estructural con el fin de proponer un modelo que explique el comportamiento espacial del fenómeno estudiado. Ambos procedimientos deben realizarse como paso previo a la ejecución de cualquier algoritmo geoestadístico de estimación. La implementación de las herramientas computacionales para el análisis de información se realiza en *clases* especiales, cuyos *parámetros* fundamentales de construcción son *objetos* de geográficos. Los *métodos de clase* implementan los algoritmos de cálculo propiamente dichos. En el caso del análisis estadístico se pueden realizar descripciones univariadas o bivariadas, mientras

que las *clases* desarrolladas para el análisis estructural permiten realizar estimaciones de variogramas experimentales y ajustes automáticos de modelos de variograma.

El aspecto más importante que trata este trabajo es la implementación informática de los diferentes algoritmos de estimación espacial. Con respecto a ello se realiza una evaluación crítica de las diferentes metodologías existentes para la estimación espacial de variables que se encuentran actualmente en la literatura. Las metodologías clásicas para la estimación espacial como los polígonos de Thiessen o el método del inverso cuadrático de la distancia, no consideran que la variable a estimar es aleatoria, por lo cual no puede cuantificarse la confiabilidad asociada con las estimaciones. Por el contrario, los algoritmos geoestadísticos consideran explícitamente la aleatoriedad de los fenómenos representados y, en virtud de lo anterior, permiten cuantificar la varianza relacionada con la estimación espacial. Igualmente, se presenta una revisión de diferentes librerías de estimaciones espaciales existentes y sus características se analizan en profundidad para incorporar sus fortalezas en la librería de algoritmos desarrollada en el presente trabajo.

Así, para cada metodología de estimación se implementa una *clase*, es decir, se crea un *objeto* computacional que puede realizar procedimientos de estimación. Los *objetos* de estimación están caracterizados básicamente por la información usada, representada mediante un objeto computacional geográfico, y por el modelo de variación espacial construido para el atributo que se desea estimar. Como resultado los *métodos* arrojan objetos espaciales raster que representan un campo distribuido de los valores estimados en un área de estudio. Se implementan algunas metodologías deterministas: los polígonos de Thiessen y la estimación con ponderaciones basadas en el inverso de la distancia a un exponente; junto con estimadores geoestadísticos: krigado ordinario, krigado simple, krigado con una deriva externa, cokrigado simple, cokrigado ordinario, cokrigado colocalizado y cokrigado colocalizado bajo un modelo de Markov.

Finalmente, para posibilitar la interacción de un usuario de nivel intermedio con los algoritmos implementados, se muestra una interfaz gráfica de usuario construida para gestionar y administrar los procesos computacionales relacionados con la ejecución de los diferentes análisis integrados a la librería, que permite al usuario realizar el ingreso de datos y la extracción, visualización y análisis de los resultados. El software (librerías e interfaz) desarrollado se denomina ESTELLUS y está escrito completamente en lenguaje de programación orientado a objetos **Java** para posibilitar la implementación de la conceptualización propuesta. Además de lo anterior, **Java** hace posible el uso de una gran cantidad de librerías de acceso libre existentes, algunas de las cuales fueron adaptadas y utilizadas para propósitos de visualización de resultados y análisis de datos. ESTELLUS pueden ser usado efectivamente en estudios de diferente tipo: modelación hidrogeológica, análisis hidrológicos, estudios ambientales, análisis meteorológicos, estudios de la ciencia del suelo, análisis de riesgo, prospección minera, etc. Mediante la aplicación ESTELLUS para la resolución de un problema práctico sencillo en hidrogeología se demuestra su versatilidad para el análisis y visualización de información geográfica y sus amplias posibilidades de expansión futura en lo concerniente con la estimación y simulación estocástica espacial.

Resum

Este treball final de màster tracta sobre la construcció d'una llibreria d'estimació geostatística que permeti la consideració explícita de la localització espacial de les dades d'entrada i els resultats obtinguts per mitjà del seu georeferenciament a partir de sistemes de coordenades geogràfics o plans estàndard. Per aconseguir dita objectiva lassa entitats espacials susceptibles d'integrar en una anàlisi estimatiu es conceptualitzen com a *objectes* computacionals que tenen associats diferents tipus de *propietats* mesurables, observables y/o qualificables. Als dits objectes se'ls pot associar certes localitzacions referides a un origen de coordenades geogràfic. Així, des del punt de vista computacional s'utilitza la programació orientada a objectes a fi de modelar entitats computacionals que representen les dites entitats espacials reals.

Inicialment, es tracta el desenrotllament dels *objectes* computacionals que manipulen la informació d'entrada i d'eixida als algorismes implementats. En general, atés que la informació utilitzada en la geostatística es basa en dades (mediciones) localitzats puntualment o es troba densament muestreada, els *objectes* computacionals espacials que es modelen són entitats puntuals i camps format raster. La creació dels *objectes* computacionals s'implementen en classes individuals que fan possible la definició de les seues propietats, la seua localització i el seu sistema de georeferenciament en els *constructors* associats. Atés que l'*objecte* ha de representar característiques espacials definides, per a una mateixa entitat espacial (per exemple localitzacions puntuals de mediciones) es generen diferents tipus d'objectes d'acord amb les dimensions de l'anàlisi: 2D o 3D. Per a la manipulació i operació amb els objectes modelatges computacionalment es generen els respectius *mètodes de classe*, que poden ser *privats* o *públics*. Els *mètodes públics* poden ser utilitzats efectivament per a nou desenrotllament del programari, per la qual cosa al llarg del treball es fa referència exclusivament a ells.

L'aspecte següent que es tracta és la implementació dels algorismes d'anàlisi d'informació: l'anàlisi estadístic descriptiu de les mostres de dades i l'anàlisi estructural a fi de proposar un model que explique el comportament espacial del fenomen estudiat. Ambdós procediments han de realitzar-se com a pas previ a l'execució de qualsevol algorisme geostatístico d'estimació. La implementació de les ferramentes computacionals per a l'anàlisi d'informació es realitza en *classes* especials, els *paràmetres* fonamentals de construcció de la qual són *objectes* de geogràfics. Els *mètodes de classe* implementen els algorismes de càlcul pròpiament dites. En el cas de l'anàlisi estadística es poden realitzar descripcions univariadas o bivariadas, mentres que les *classes* desenrotllades per a

l'anàlisi estructural permeten realitzar estimacions de variogramas experimentals i ajustos automàtics de models de variograma.

L'aspecte més important que tracta este treball és la implementació informàtica dels diferents algoritmes d'estimació espacial. Respecte a això es realitza una avaluació crítica de les diferents metodologies existents per a l'estimació espacial de variables que es troben actualment en la literatura. Les metodologies clàssiques com els polígons de Thiessen o el mètode de l'invers quadràtic de la distància, no consideren que la variable a estimar siga aleatòria, per la qual cosa no pot quantificar-se la confiabilitat associada amb les estimacions. Al contrari, els algoritmes geoestadístics consideren explícitament l'aleatorietat dels fenòmens representats i, en virtut de l'anterior, permeten quantificar la varianza relacionada amb l'estimació espacial. Igualment, es presenta una revisió de diferents llibreries d'estimacions espacials existents i les seues característiques s'analitzen en profunditat per a incorporar les seues fortaleces en la llibreria d'algoritmes desenrotllada en el present treball.

Així, per a cada metodologia d'estimació s'implementa una *classe*, és a dir, es crea un *objecte* computacional que pot realitzar procediments d'estimació. Els *objectes* d'estimació estan caracteritzats bàsicament per la informació usada, representada per mitjà d'un objecte computacional geogràfic, i pel model de variació espacial construït per a látribut que es desitja estimar. Com resultat els *mètodes* tiren objectes espacials raster que representen un camp distribuït de ho valores estimats en una àrea d'estudi. S'implementen algunes metodologies deterministes: els polígons de Thiessen i l'estimació amb ponderacions basades en l'invers de la distància a un exponent; junt amb estimadors geoestadístics: krigeado ordinari, krigeado simple, krigeado amb una deriva externa, cokrigeado simple, cokrigeado ordinari, cokrigeado colocalizado i cokrigeado colocalizado davall un model de Markov.

Finalment, per a possibilitar l'interacció d'un usuari de nivell intermedi amb els algoritmes implementats, es mostra una interfície gràfica d'usuari construïda per a gestionar i administrar els processos computacionals relacionats amb l'execució de les diferents anàlisis integrats a la llibreria, que permet a l'usuari realitzar l'ingrés de dades i l'extracció, visualització i anàlisi dels resultats. El programari (librerías i interfaz) desenrotllat es denomina ESTELLUS i està escrit completament en llenguatge de programació orientat a objectes Java per a possibilitar la implementació de la conceptualització proposada. A més de l'anterior, Java fa possible l'ús d'una gran quantitat de llibreries d'accés lliure existents, algunes de les quals van ser adaptades i utilitzades per a propòsits de visualització de resultats i anàlisi de dades. ESTELLUS pot ser usat efectivament en estudis de diferent tipus: modelació hidrogeològica, anàlisis hidrològiques, estudis ambientals, anàlisis meteorològiques, estudis de la ciència del sòl, anàlisi de risc, prospecció minera, etc. Per mitjà de l'aplicació de ESTELLUS per a la resolució d'un problema pràctic senzill en hidrogeologia es demostra la seua versatilitat per a l'anàlisi i visualització d'informació geogràfica i les seues àmplies possibilitats d'expansió futura en el concernent amb l'estimació i simulació estocàstica espacial.

Abstract

This master thesis deals with the implementation of a geostatistical library for spatial estimation that allows to consider explicitly the spatial location of the data and the obtained results by using standard (cartesian or geographic) georeferencing coordinate systems. To achieve such goal, the spatial entities that are able to be used in an estimation analysis are conceptualized as computational *objects* or entities with some intrinsic *properties*: they possess different types of features that are able to be observed or qualified and they have associated some coordinates referred to a geographical reference system. Then, from a computational perspective, object oriented programming is used to model some computer objects that represent accurately such spatial entities.

Initially, the development of the computational objects to handle the input and output information to and from the estimation algorithms is presented. In general, because the information used in geostatistics is based in scattered-point data or is densely measured information, the modeled computational *objects* of spatial information are vectorial points and raster fields. The creation of the computational *objects* has been implemented in particular *classes*. Each *class* allows to define the spatial entity's properties, locations and georeferencing system inside of the associated *builders*. Because the *object* has to represent different spatial features for entities of the same type (for instance: scattered-pointed locations of measuring) different kinds of spatial objects are implemented according to the dimensions of the analysis: 2D or 3D. To handle and operate with the computationally modeled spatial objects, their associated *class methods* are developed. In objects oriented programming, *class methods* can be *public* if they are able to be used outside of the container class or *private* if they are able to be only inside the container class. Only *public methods* can be effectively used to create extensions of the actual version of the software.

Another subject treated in this thesis is about the implementation of algorithms for the description and analysis of the information: statistical description of the data sample and structural analysis to propose models of variability that explain the spatial behavior of the analyzed phenomena. Both of them have to be made as the previous step of the execution of any estimation procedure. Information analysis algorithms have been implemented in *classes* having geographical *objects* as fundamental *parameters*. The *class's methods* implement the analysis algorithms. In the particular case of the statistical analysis, univariate or bivariate descriptions are able to be performed. In the case of the structural analysis, algorithms for experimental variogram estimation and automatic fitting of variogram models are able to be performed.

The most important topic treated in the thesis is the computational implementation of some algorithms for the spatial estimation. Regarding to them, a critical a thorough literature review about spatial estimation methodologies is presented. The classical methodologies to perform spatial estimations, for instance, Thiessen's polygon or inverse distance interpolation, can not consider the randomness of the variable that has been estimated. In opposition, geostatistical algorithms can account the randomness of the represented phenomena and allow quantifying the reliability related with the spatial interpolation procedure. Also, a revision of the features of some computational libraries for spatial estimation is given. Their advantages and weakness are widely commented. In this work the main strengths of those libraries are adopted in the implementation of the estimation algorithms.

For each estimation methodology a *class* is implemented, it means that a computational *object* is created to be able to perform the estimation procedure. The estimation objects are characterized by two properties: the information used for estimation (spatial objects) and the model of spatial variability of the attribute. The results of the *class methods* are raster objects that represents a random field of the estimated variable over a given area. In this work, the most popular estimation algorithms have been implemented: Thiessen's polygons, inverse distance interpolation, simple kriging, ordinary kriging, simple cokriging, ordinary cokriging, collocated cokriging and collocated cokriging using a Markov model.

Finally, with the porpoise of achieve the interaction of the user with the implemented algorithms, the development of a graphical interface is presented. This interface manage the computational procedures related with the execution of the different types of analyses programmed in the library and also allows to input the data and extract, view and study the results. ESTELLUS is the name given to the developed software (algorithm library plus user's interface). It is written fully in **Java** object oriented programming language to be able to implement the conceptualization for the spatial objects. Also, **Java** makes possible to use a immense number of free access libraries, some of which were adapted to be used in ESTELLUS development, specially for data and results visualization and analysis of samples of information. ESTELLUS can be used effectively in different types of studies an researches: hydrogeological modeling, hydrological analysis, environmental studies, meteorological researches, soil science studies, risk analysis, minning prospection, etc. ESTELLUS is used to solve a practical problem of spatial estimation showing its versatility for analysis and visualization of geographical information and their possibilities of future growth. It is expected to integrate the simulations algorithms to the library in the future along with more utilities for handling the data and results.