Modelos Geoestatísticos

Paulo Justiniano Ribeiro Jr.

12^{o} SEAGRO e 52^{a} RBRAS

Paulo J. Ribeiro Jr.
LEG: Laboratório de Estatística e Geoinformação
Universidade Federal do Paraná
C.P. 19.081
Curitiba, Paraná, Brasil
81.531-990
paulojus@ufpr.br

Prefácio

Métodos para análise de dados que incorporem a informação da posição no espaço têm estado presentes na literatura estatística desde suas origens, ainda que de forma incipiente. Entretanto, nas últimas décadas tem havido uma verdadeira explosão no desenvolvimento de metodologias e tecnologias em várias áreas para o tratamento da dados espacializados. Ao conjunto destas metodologias dá-se o nome de *análise espacial*. A área de estatística é uma entre as quais se percebe um pronunciado desenvolvimento de métodos e recursos para tratamento de tais dados, o que é explicado pela facilidade de coleta e referenciamento dos dados, a capacidade de armazenagem, o desenvolvimento de recursos tais como os sistemas de informações geográficas (SIG's), a possibilidade de integração de aplicativos e códigos computacionais, e, talvez principalmente, a crescente capacidade de processamento que permite métodos anteriormente impraticáveis serem agora desenvolvidos e aplicados com relativa simplicidade e rapidez. As implementações computacionais se multiplicam e a cultura cada vez mais forte de software livre permite rápido desenvolvimento, adaptação, prototipação e teste de idéias e evolução rápida e incremental dos recursos disponíveis. Por outro lado as mais diversas áreas substantivas do conhecimento tem gerado questões científicas cada vez mais desafiadoras e detalhadas em relação à necessidade de explicação espacial de fenômenos. De ciências humanas às médicas, meteorológicas às ambientais, ecológicas às agronômicas, geológicas às florestais, entre outras; todas têm tido na análise espacial de fenômenos um dos tópicos de interesse. Dentro do amplo universo de métodos de análise espacial estão as metodologias estatísticas adequadas para tais estudos, coletivamente tratadas por *estatística espacial*.

O termo *Geoestatística* se refere a uma sub-área de estatística espacial na qual os dados consistem de um número finito de medições relacionadas a um fenômeno subjacente espacialmente contínuo. Por exemplo, considere um conjunto de medidas de temperatura máxima coletadas em um dia em estações meteorológicas de um certo país. A superfície subjacente é dada pelo o conjunto de temperaturas em toda a área, considerada então uma realização de processo estocástico no espaço. A amostra consiste de apenas observações em alguns pontos desta superfície, nas localizações das estações. Outros termos são utilizados na literatura para identificar a mesma categoria de problemas, tais como *dados com referência pontual* ou *dados de variação espacial contínua*.

As particulares e múltiplas terminologias empregadas na área são reflexo do fato que o tema teve origens em diferentes áreas do conhecimento. Em particular o termo geoestatística foi cunhado por Georges Matheron e colaboradores em Fontainebleau, França, para descrever seu trabalho voltado para problemas de predição surgidos em mineração. A formulação teórica mostrou-se bastante geral e foi propagada e adaptada para diversas outras áreas de aplicação. Entretanto, tais desenvolvimentos ocorreram desconectados da corrente principal de estatística espacial dentro da comunidade estatística que incluem, por exemplo trabalhos como de Kolmogorov, Whittle, Bartlett e Watson em processos espaciais e Matérn, com ênfase em aplicações em silvicultura. Com o tempo as conexões e contatos entre as áreas foram se estabelecendo, como por exemplo traçadas por Brian Ripley em seu texto *Spatial Statistics* relacionando o método conhecido como *krigagem simples* com predição de mínimos quadrados em modelos lineares, seguido por diversos outros autores como Journel, Kitanidis, Omre, entre outros, que esclareceram conexões entre métodos geoestatísticos e conceitos usuais de estatística. Noel Cressie em seu vasto livro de 1993 classifica então geoestatística como uma das três grandes áreas de estatística espacial, sendo as outras variação espacial discreta e processos pontuais. Desde então a interação entre áreas foi crescente com uma profícua cross-fertilização de idéias e metodologias.

Mesmo com esta integração, algumas práticas tendem ainda a refletir as origens do tema levando a adoção de diferentes abordagens. Em particular métodos de inferência baseados declaração explícita de modelos e em inferência baseada na verossimilhança (incluindo aqui métodos Bayesianos), largamente aceitos na comunidade estatística, não são uma unanimidade, principalmente em áreas aplicadas com tradição de longa data no uso de métodos geoestatísticos. Neste curso seguimos esta abordagem por métodos de inferência baseados na verossimilhança seguindo a declaração explicita de modelos paramétricos. São mostradas aplicações em diferentes tipos de dados e os procedimentos centrais aqui discutidos estão implementados nos pacotes geoR e geoRglm do ambiente de computação estatística R (*http://www.r-project.org*).

O conteúdo deste material segue estritamente o conteúdo do recém publicado livro *Model based Geostatistics* (Diggle & Ribeiro Jr, 2007, Springer). Em particular estas notas são uma versão adaptada do material apresentado em conjunto com Peter Diggle na XXIII IBC (2006), em disciplina de pós-graduação oferecidas na ESALQ/USP e UFPR e no curso de verão oferecido no IME/USP em 2007.

Tendo em vista o ativo desenvolvimento da área o material discutido aqui pode ser considerado de nível introdutório. Os modelos e métodos aqui discutidos guardam as idéais e conceitos centrais que permitem a compreensão e desenvolvimento de extensões da metodologia para atender às necessidade de modelagem de fenômenos cada vez mais complexos. Dentre as áreas de desenvolvimento ativo mas ainda por responder de forma mais satisfatória e geral às necessidades levantadas por problemas reais eu destaco modelagem espaço temporal, modelagem multivariada, modelos não estacionários e métodos para tratamento de grandes bases de dados. Por necessidade de desenvolvimento entendo não apenas a proposição de modelos, estudos de suas propriedades e limitações, validações e comparações, mas também a disponibilização de implementações computacionais que não seja específicas do problema ou conjunto de dados em questão permitindo a adoção de tais metodologias numa variedade de problemas concretos, bem como a integração de tais implementações com outros recursos de análise espacial.

Agradeço às instituições que proporcionaram as oportunidades de apresentar e desenvolver o material, bem como à comissão organizadora do 12^o SEAGRO e 52^a reunião anual da RBRAS pelo convite e oportunidade de apresentar este curso em um evento de nossa sociedade, o que muito me honra dados minhas ligações de longa data com a RBRAS. Agradeço de forma geral a colegas, colaboradores em trabalhos e projetos e orientados com os quais tenho tido a oportunidade de discutir e aprender. Destaco aqui que a parte do material e implementação computacional referente a modelos lineares espaciais geoestatísticos deve-se à parceria com Ole Christensen. O material aqui apresentado foi desenvolvido em trabalho em parceria com meu orientador, co-autor e amigo Peter J. Diggle. Eventuais incorreções e imprecisões são de minha inteira responsabilidade.

Paulo Justiniano Ribeiro Jr
Curitiba, PR, Junho de 2007