

Relatório científico da Oficina PROCESSOS ESTOCÁSTICOS APLICADOS À ESTATÍSTICA ESPACIAL

Local. Núcleo de Modelagem Estocástica e Complexidade, São Paulo (NUMEC)

Data 25, 26 e 27 de outubro de 2006.

Organização Núcleo de Modelagem Estocástica e Complexidade

Instituto de Matemática e Estatística

Instituto do Milênio para o Avanço Global da Matemática

Comissão organizadora: Alexandra M. Schmidt (IM-UFRJ) e Pablo Ferrari (IME-USP)

Importância da reunião para o cenário científico ou tecnológico Estatísticas espaciais aparecem em diversos modelos fundamentais no cenário científico e tecnológico. Problemas como o estudo do desmatamento da Amazonia, a disseminação de doenças infectocontagiosas, localização de áreas de insegurança pública, planejamento de trânsito nas grandes capitais, estudo de comportamentos atmosféricos requerem ferramentas teóricas e práticas que vão além das estatísticas clássicas ou séries temporais que estudam modelos unidimensionais.

A base teórica são os chamados campos aleatórios, que descrevem o comportamento de variáveis aleatórias indexadas pelos pontos de uma região espacial que pode ser discreta ou contínua. A análise pode ser estática ou dinâmica. Em ambos os casos são desenvolvidas ferramentas específicas para o estudo de propriedades dos campos como o comportamento ergódico, a existência de inhomogeneidades, o crescimento de regiões, etc. O quadro geral é a mecânica estatística de campos aleatórios no caso estático e a evolução de sistemas com muitas componentes no caso dinâmico. Esse é o tema central do projeto temático sob a direção de um dos organizadores do evento.

Dada a importância da estatística espacial tanto do ponto de vista teórico quanto pelas aplicações, o NUMEC conjuntamente com o Instituto de Matemática da USP organizaram uma oficina de 3 dias com os especialistas do Brasil e da região sobre o tema de modelamento espacial e suas estatísticas.

Relatório O Workshop foi elaborado de modo a ter a participação de importantes pesquisadores nacionais e internacionais da área de Processos Estocásticos, Estatística

Espacial e de áreas aplicadas ao estudo de dados espacialmente referenciados. Além dos 10 pesquisadores convidados, houve 40 inscritos entre professores de diferentes instituições brasileiras (USP, UFPR, UNESP, UNICAMP), alunos de Doutorado e Mestrado de diferentes Programas de Pós-Graduação do Brasil (IME-USP, IM-UFRJ, IME-UNICAMP, ESALQ-USP), além de alguns pesquisadores internacionais.

As apresentações cobriram as principais áreas da Estatística espacial, a saber: geostatística, dados de área e processos pontuais.

O Workshop, em geral, manteve um clima bastante informal, com apresentações de 50 minutos de cada pesquisador convidado. Os pesquisadores convidados apresentaram trabalhos em desenvolvimento, dando oportunidade para que várias discussões surgissem em cada apresentação.

O Workshop criou um ambiente muito propício para que pesquisadores de diferentes áreas mostrassem seus trabalhos e encontrassem pontos de interesse em comum. Foi muito interessante a participação de pesquisadores do INPE, que colocaram perguntas práticas importantes. Deste contato, ficou claro que ainda é preciso desenvolver o contato entre profissionais da área de Probabilidade e Estatística com as outras áreas da ciência. O intuito é criar e estudar modelos e procedimentos de inferência que resolvam problemas práticos da forma mais realista possível.

Na sessão de encerramento do workshop decidiu-se realizar um segundo workshop no ano 2007, para que o grupo possa mostrar o desenvolvimento dos problemas abordados e continuar incentivando novas parcerias de trabalho.

Pesquisadores Convidados: Alexandra M. Schmidt (IM-UFRJ)

Antonio Miguel V. Monteiro (DPI-INPE)

Camilo Rennó (DPI-INPE)

Dani Gamerman (IM-UFRJ)

Hedibert F. Lopes (GSB-University of Chicago)

Marcelo A. Costa (UFMG)

Nancy Garcia (IME-Unicamp)

Pablo Ferrari (IME-USP)

Paulo Justiniano Ribeiro Jr (DE-UFPR)

Ricardo Fraiman (Universidad de San Andrés)

Programação: Chegada dos participantes : 25/10 à tarde

Abertura e mesa redonda “Estatística Espacial” as 18h no dia 25/10 no NUMEC, com a presença dos conferencistas convidados.

Cada convidado teve 50 minutos de apresentação mais 15 minutos de discussão para incentivar discussões de temas em aberto na área apresentada.

Resumo das palestras ALEXANDRA M. SCHMIDT (IM-UFRJ)

TÍTULO: Estruturas de covariância flexíveis para processos espaciais e espaço-temporais

RESUMO: O desafio na modelagem de processos espaciais e espaço-temporais está na descrição da estrutura de covariância do fenômeno sob estudo. Aqui, nos concentraremos em descrever modelos que resultem em estruturas de covariância flexíveis, tanto para processos espaciais, univariados e multivariados, como para processos espaço-temporais.

ANTONIO MIGUEL V. MONTEIRO (INPE)

TÍTULO: Tecnologias da informação espacial e estatística: experimentos com territórios digitais

RESUMO: "O Brasil não conhece o Brasil". E isto é muito grave. As implicações são diretas para a gestão, organização e ordenamento do território brasileiro, com reflexos diretos na vida dos municípios e de cada cidadão. Como dizia o saudoso Mestre Milton Santos, "O território é o dado essencial da condição da vida cotidiana". Então como conhecer o Brasil sem conhecer os Territórios Brasileiros? Esta palestra vai procurar mostrar que esta não é uma tarefa exclusiva dos geógrafos e cientistas sociais, e tenta tratar a questão de como podem Cientistas com origem nas Ciências Naturais auxiliar na construção e acúmulo de conhecimento sobre a realidade brasileira a partir do conceito de Territórios Digitais e de suas representacionais matemático-computacionais. Vamos ilustrar o tipo de problemas que enfrentamos, as ferramentas que podem nos ajudar, o pouco já feito e o muito que precisamos ainda fazer. Esperamos mostrar oportunidades de trabalhos de pesquisa e desenvolvimento em diferentes áreas da Ciência da Computação e das Ciências Matemáticas e Estatísticas e da necessidade de sua estreita colaboração. Apresentaremos exemplos de um ambiente aberto promovido pelo INPE e parceiros que é nossa plataforma computacional para o desenvolvimento dos experimentos com o conceito aqui apresentado, a TerraLib, e um exemplo de interação entre grupos das ciências médicas e sociais, matemáticos e estatísticos e grupos de Engenharia e Computação para responder questões relevantes nos territórios da saúde pública, da segurança pública, dos indicadores de exclusão e segregação social em cidades e na modelagem da dinâmica de uso e cobertura da terra na Amazônia Brasileira.

CAMILO D. RENNÓ (DPI-INPE)

TÍTULO: Modelos Hidrológicos Distribuídos

Os modelos hidrológicos são ferramentas úteis para o entendimento do comportamento hidrológico de bacias hidrográficas. Estes modelos deveriam representar grande parte dos processos hidrológicos a fim de produzir resultados realistas. Entretanto, os modelos são apenas uma aproximação da realidade e requerem uma boa inicialização das variáveis de estado e uma grande quantidade de dados para representar apropriadamente os processos en-

volvidos. Esta apresentação visa discutir as diferentes abordagens utilizadas na construção de modelos hidrológicos distribuídos. Inicialmente, será apresentado o ciclo hidrológico e seus componentes. Aspectos relativos a discretização espacial e temporal serão também discutidos. Especificamente, serão apresentados os resultados do desenvolvimento do sistema SASHI (Sistema de Análise e Simulação Hidrológica). O enfoque principal é mostrar as diferenças das variáveis que alimentam modelos deste tipo.

DANI GAMERMAN (IM-UFRJ)

TÍTULO: Modelos Gaussianos Dinâmicos em Processos Pontuais

Espaço-Temporais: Estudo da Infecção Gastrointestinal em Hampshire (Reino Unido)

RESUMO: O projeto AEGISS (Ascertainment and Enhancement of Gastrointestinal Infection Surveillance and Statistics) registrou 7126 casos de infecção gastrointestinal não-específica no condado de Hampshire (Reino Unido) nos anos de 2001 a 2003. Um caso foi definido como qualquer chamada telefônica a um serviço de orientação médica relatando sintomas da doença. Cada caso foi identificado pela localização residencial da pessoa (coordenadas geográficas) e pela data da chamada. Diggle et al. (2005) modelam os casos diários dos dois primeiros anos com processos espaço-temporais descritos através de equações diferenciais estocásticas (Brix e Diggle, 2001). A intensidade do processo pontual é decomposta de forma multiplicativa em componentes determinísticos representando as variações puramente espacial e puramente temporal e em um componente estocástico modelado como um processo de Cox log-gaussiano estacionário. A inferência via verossimilhança mostra ser uma tarefa difícil neste contexto, levando os autores a recorrer a estimadores de momentos. Neste trabalho apresentamos a análise dos casos dos três anos através de um modelo espaço-temporal (Reis, 2006) especificado por uma seqüência de superfícies de log-intensidade espaciais ligadas através do tempo e modeladas por processos gaussianos dinâmicos (Gamerman et al., 2006). A inferência é feita sob a abordagem Bayesiana, que incorpora naturalmente a característica estocástica da função de intensidade nos processos pontuais de Cox. Os parâmetros do modelo são estimados a partir de amostras de suas distribuições a posteriori, obtidas através de métodos Monte Carlo via cadeias de Markov, como o algoritmo de Metropolis-Hastings.

HEDIBERT F. LOPES (GSB-CHICAGO)

TÍTULO: Spatial Dynamic Factor Models

RESUMO: A new class of space-time models derived from standard dynamic factor models is proposed. The temporal dependence is modeled by latent factors while the spatial dependence is modeled by the factor loadings. Factor analytic arguments are used to help identifying temporal components that summarize most of the spatial variation of a given region. The temporal evolution of the factors is described in a number of forms to account for different aspects of time variation such as trend and seasonality. The new structure implies nonseparable space-time variation to observables, despite its conditionally independent nature, while reducing the overall dimensionality, and hence complexity, of the problem. Conditionally conjugate and reference-type prior distributions are entertained

for the parameters of the spatial components. Different covariance structures are also entertained. Conditionally on the number of common factors, inference is performed by standard Gibbs and Metropolis-Hastings steps. The number of factors is treated as another unknown parameter and fully Bayesian inference is performed by a reversible jump Markov Chain Monte Carlo algorithm. The new class of models is tested against two synthetic and one real data sets. The real data was obtained from Clean Air Status and Trends Network (CASTNET) and refers to atmospheric concentration of sulfur dioxide weekly observed at 24 monitoring stations from 1998 to 2004. The factor model decomposition is shown to capture important aspects of spatial and temporal behavior of the data.

MARCELO A. COSTA (UFMG)

TÍTULO: Detecção de conglomerados espaciais utilizando o método de crescimento de árvore com restrição

RESUMO: A detecção de conglomerados espaciais tem como objetivo a delimitação de uma região geográfica na qual a hipótese de ocorrência aleatória de um determinado evento pontual é rejeitada. Tal informação é de extrema relevância em estudos epidemiológicos. Este artigo apresenta um método de detecção de conglomerados espaciais no qual a estrutura de vizinhança espacial é agregada ao processo de crescimento e busca de conglomerados, possibilitando a detecção de conglomerados de geometria arbitrária. Os métodos tradicionais de varredura espacial restringem à geometria de busca a conglomerados de geometria circular, resultando em uma detecção parcial ou superestimação do conglomerado. Restrições durante o processo de crescimento são sugeridas para evitar conglomerados de tamanho excessivo e geometria muito irregular. Uma avaliação do poder de detecção do método para conglomerados com geometria arbitrária é realizada utilizando dados simulados. Resultados de detecção de conglomerados espaciais em dados de crimes são apresentados para a região de Belo Horizonte.

NANCY L. GARCIA (IME-UNICAMP)

TÍTULO: Penalized maximum likelihood approach for an inverse problem for a Poisson point process.

RESUMO: In this talk we will consider a particular example of an inverse problem for a Poisson point process. In this case, it is possible to obtain exact formulas for the mean measure and the intensity function of the process. Then, a penalized maximum likelihood estimation technique is used. Simulation results indicate the procedure to be consistent.

PABLO FERRARI (IME-USP)

TÍTULO: Grafos fatores de processos de Poisson

RESUMO: Um grafo fator de um processo pontual é um grafo que tem como vértices os pontos do processo e que é construído em forma invariante por isometrias. Construções de grafos fatores para processos unidimensionais devidas a Ferrari, Landim Thorisson, Peres e Holroyd, Last and Haveling e Timar serão revistas. Há importantes consequências dessas construções para a teoria de medidas de Palm para processos estacionários.

PAULO JUSTINIANO RIBEIRO Jr. (DEs-UFPR)

TÍTULO: Avaliando a evolução da Morte Súbita dos Citrus

RESUMO: A Morte Súbita dos Citrus (MSC) é uma doença qua atinge certas combinações e enxerto/porta-enxerto em árvores de citrus já afetando parte da principal região citrícola brasileira e com potencial de espalhamento nos próximos anos. A doença foi detectada pela primeira vez em 1999 e descrita em 2001. O agente causador ainda não está determinado embora evidências apontem para uma variação do vírus da Tristeza dos Citrus. Dada a situação atual de alto risco e potencial prejuízo a doença tem sido foco de atenção por porte dos órgãos ligados a produção de citrus. Diversos levantamentos de dados e estudos tem sido realizados a fim de compreender melhor os padrões de dispersão espacial da doença tanto a pequena escala em talhões, quanto na larga escala, na região produtora. Nesta apresentação descreve-se o problema e as questões de relevância científica e estatística associadas. Resultados obtidos até o momento serão apresentados e serão discutidas estratégias para futuras análises.

RICARDO FRAIMAN (Universidad de San Andrés, Argentina)

TÍTULO: On projection-based depth notions with applications to high-dimensional and functional data analysis.

RESUMO: We introduce two new concepts of data depth for infinite dimensional spaces. We provide some applications for high-dimensional and functional data. Particular attention is given to The classification problem for functional data.