

II Workshop em Clima, Estatística e Saúde

Mudanças Climáticas Globais – cenários de produção e propagação de doenças

Tecnologias computacionais e ambiente para modelagem estatística no monitoramento de ovos do mosquito da dengue em Recife, PE

Paulo Justiniano Ribeiro Jr

LEG: Laboratório de Estatística e Geoinformação, UFPR

em colaboração com: *Wagner Bonat, Henrique Dallazuanna, Miguel Monteiro e ... muita gente*

Natal, RN

26-27 de Maio de 2008

Eventos em sociedades estatísticas:

2º Workshop on Stochastic Processes Applied to Spatial Statistics:

- Multi-scenario analysis and stochasticity in environmental prediction
- NUMEC - Núcleo de Modelagem Estocástica e Complexidade
- USP, December 5-8, 2007
- *This multidisciplinary workshop aims to put together specialists from different fields of science and discuss some current topics of research in the environmental science.*

53ª RBRAS

- Região Brasileira da Sociedade Internacional de Biometria
- 14–16 Maio 2008, Lavras, MG
- <http://www.rbras.org.br>

Eventos em sociedades estatísticas:

18^o SINAPE

- Associação Brasileira de Estatística
- 28/07 a 01/08 de 2008
- Seção sobre estatística e meio ambiente
- <http://www.redeabe.org.br>

55^a RBRAS, IBC-Floripa-2010

- Evento da Sociedade Internacional de Biometria
- Dezembro de 2010, Florianópolis, SC
- <http://www.rbras.org.br>

Tópicos

- Experimento Dengue/Recife/SAUDAVEL
- Disponibilizando Metodologias e Tecnologias
- Disponibilizando Resultados
- Discussão

Tópicos

- sistema de aquisição de dados

Tópicos

- sistema de aquisição de dados
- análises e relatórios frequentes/sistemáticos

Tópicos

- sistema de aquisição de dados
- análises e relatórios frequentes/sistemáticos
- análises estáticas e dinâmicas

Tópicos

- sistema de aquisição de dados
- análises e relatórios frequentes/sistemáticos
- análises estáticas e dinâmicas
- conformidade com princípios de inovação e acesso

Tópicos

- sistema de aquisição de dados
- análises e relatórios frequentes/sistemáticos
- análises estáticas e dinâmicas
- conformidade com princípios de inovação e acesso
- explorar possibilidades de modelagem (estatística)

Tópicos

- sistema de aquisição de dados
- análises e relatórios frequentes/sistemáticos
- análises estáticas e dinâmicas
- conformidade com princípios de inovação e acesso
- explorar possibilidades de modelagem (estatística)
- *habeas data, habeas software . . . habeas código*

Tópicos

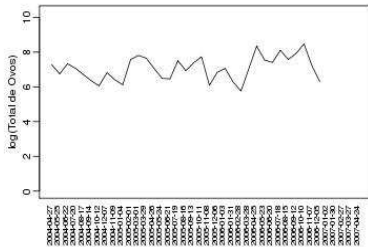
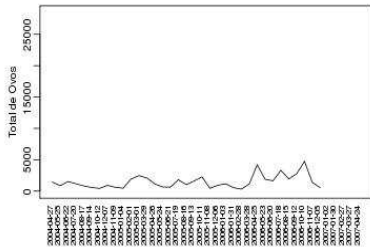
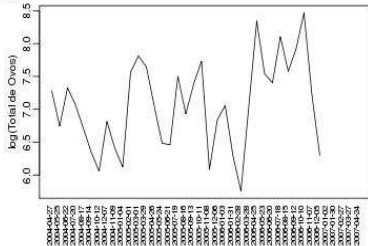
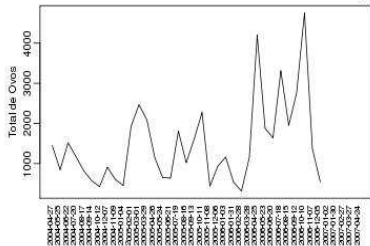
- sistema de aquisição de dados
- análises e relatórios frequentes/sistemáticos
- análises estáticas e dinâmicas
- conformidade com princípios de inovação e acesso
- explorar possibilidades de modelagem (estatística)
- *habeas data, habeas software . . . habeas código*
- *reproducible Research*

Experimento Dengue/Recife/Saudavel

- 80 a 100 armadilhas/ovitrampas em cada um de 6 bairros
- 4 grupos de ovitrampas: em cada bairro 25% das armadilhas por semana
- contagem do número de ovos por ovitrampa
- informações sobre local de coleta
- ambiente: temperatura e umidade

Experimento Dengue/Recife/Saudavel

BT104



Vendo os dados mais de perto

- ver código/dados
- mapas suavizados

Questões iniciais

- relacionar contagens de ovos com condições locais e ambientais

Questões iniciais

- relacionar contagens de ovos com condições locais e ambientais
- efeitos marginais *versus* ambientais

Questões iniciais

- relacionar contagens de ovos com condições locais e ambientais
- efeitos marginais *versus* ambientais
 - efeito das condições locais (dado ambiente) $Y \sim Amb + Loc$

Questões iniciais

- relacionar contagens de ovos com condições locais e ambientais
- efeitos marginais *versus* ambientais
 - efeito das condições locais (dado ambiente) $Y \sim Amb + Loc$
 - efeito das condições ambientais (dado local) $Y \sim Loc + Amb$

Questões iniciais

- relacionar contagens de ovos com condições locais e ambientais
- efeitos marginais *versus* ambientais
 - efeito das condições locais (dado ambiente) $Y \sim Amb + Loc$
 - efeito das condições ambientais (dado local) $Y \sim Loc + Amb$
- efeitos espaciais e/ou temporais

Questões iniciais

- relacionar contagens de ovos com condições locais e ambientais
- efeitos marginais *versus* ambientais
 - efeito das condições locais (dado ambiente) $Y \sim Amb + Loc$
 - efeito das condições ambientais (dado local) $Y \sim Loc + Amb$
- efeitos espaciais e/ou temporais
- relações dentro e entre bairros

Questões iniciais

- relacionar contagens de ovos com condições locais e ambientais
- efeitos marginais *versus* ambientais
 - efeito das condições locais (dado ambiente) $Y \sim Amb + Loc$
 - efeito das condições ambientais (dado local) $Y \sim Loc + Amb$
- efeitos espaciais e/ou temporais
- relações dentro e entre bairros
- relações para o município

Questões iniciais

- relacionar contagens de ovos com condições locais e ambientais
- efeitos marginais *versus* ambientais
 - efeito das condições locais (dado ambiente) $Y \sim Amb + Loc$
 - efeito das condições ambientais (dado local) $Y \sim Loc + Amb$
- efeitos espaciais e/ou temporais
- relações dentro e entre bairros
- relações para o município
- **predição, monitoramento, sistema de alerta (dinâmica)**

Algumas análises

Total de Observacoes: 19068

Total de Armadilhas: 564

Total de Bairros: 6

Total de Ovos: 14,829,557

Dois Irmãos:

- período: 04/01/2005 a 16/05/2006
- tipo de imóvel, presença de quintal, água ligada à rede geral, água canalizada no cômodo, frequência de abastecimento, recipientes pequenos com/sem tampa, recipientes grandes com/sem tampa
- umidade do ar, precipitação, temperatura máxima e mínima
- defasagens de até 12 semanas, agrupadas por semana

Modelo 1: No Ovos \sim Ambiente

- defasagens distribuídas:
 $\sim \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_2 X_{t-2} + \dots$
- semi/não-paramétricos, funções de transferência, passeios aleatórios, CAR, etc
- covariáveis da análise fatorial
- binomial negativa

Tabela: Ajuste do Modelo 1

Causas de Variação	Estimativas	Erro Padrão	P-valor
Intercepto	1.818	1.174	0.121
TEMP.1	-0.150	0.049	0.002
TEMP.3	0.257	0.038	$3.95e^{-11}$
UMI.2	0.0382	0.008	$2.38e^{-06}$

Fatores - temperatura máxima

Carga Fatorial	Factor1	Factor2	Factor3
Semana 1	0.9808	-0.1799	0.0117
Semana 2	0.9152	-0.1335	0.1008
Semana 3	0.7789	-0.0726	0.2388
Semana 4	0.6652	-0.0263	0.3462
Semana 5	0.4625	0.0197	0.5299
Semana 6	0.3216	0.1194	0.5867
Semana 7	0.2220	0.2333	0.5752
Semana 8	0.1359	0.4048	0.4842
Semana 9	0.0795	0.6043	0.3287
Semana 10	0.0559	0.7245	0.2030
Semana 11	0.0098	0.8607	0.0706
Semana 12	-0.0950	0.8775	0.0837

Modelo 2: No Ovos ~ Ambiente + Local

Causas de Variação	Estimativas	Erro Padrão	P-valor
Intercepto	1.55	1.099	0.156
TEMP.1	-0.160	0.046	0.00052
TEMP.3	0.268	0.036	$2.01e^{-13}$
UMI.2	0.0395	0.007	$1.84e^{-07}$
RES.PEQ.SEM	0.574	0.103	$3.78e^{-08}$
RES.PEQ.COM	0.178	0.06	0.007
FREQ ABAST	0.420	0.088	$2.36e^{-06}$
AGUA REDE	-0.292	0.098	0.0031

Modelo 3: No Ovos \sim Ambiente + Local + Espaço + Tempo

Causas de Variação	Estimativas	Erro Padrão	P-valor
Intercepto	1.209	0.980	0.217
TEMP.1	-0.161	0.041	$8.7e^{-05}$
TEMP.3	0.284	0.032	$< 2e^{-16}$
UMI.2	0.0378	0.006	$2.39e^{-08}$
RES.PEQ.SEM	0.426	0.104	$2.39e^{-08}$
Funções Suaves	EDF	F	P-valor
Coordenadas	24.346	10.61	$< 2e^{-16}$
Tempo	6.71	3.39	0.00039

- $Y \sim$ Binomial Negativa
- Modelo aditivo generalizado (GAM)

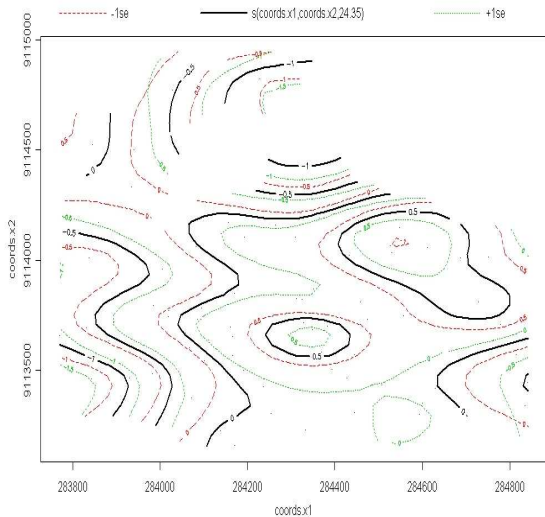
$$g(E(Y)) = \beta_0 + \sum_a \beta_a x_a + \sum_l \beta_l z_l + s(c_x, c_y) + s(t)$$

Qual modelo?

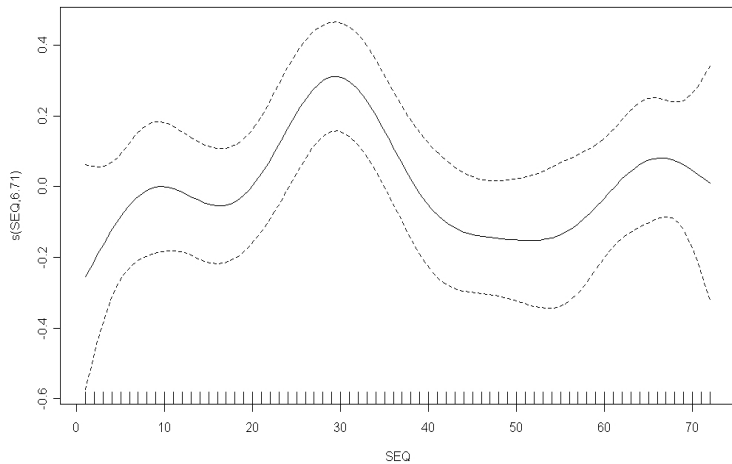
Tabela: AIC dos modelos

Modelo 1	Modelo 2	Modelo 3
19726.75	19622.66	19373.48

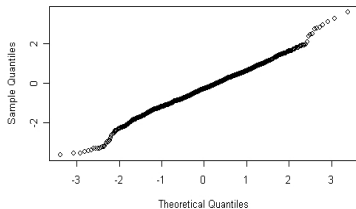
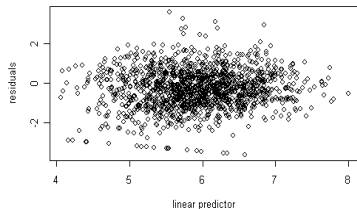
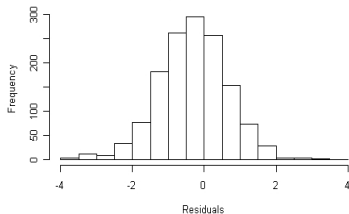
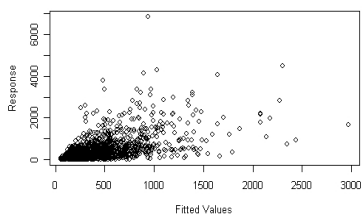
Termo espacial



Termo temporal

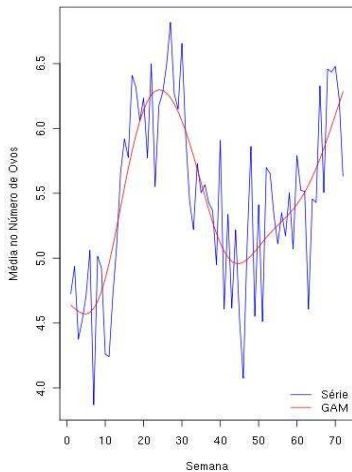


O modelo é bom? – resíduos

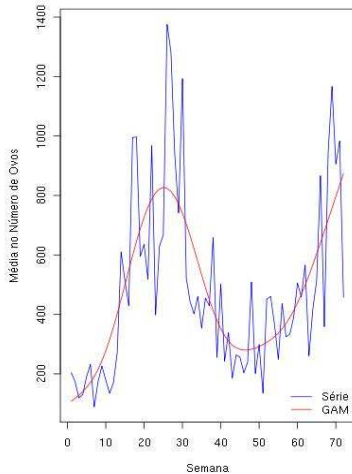
Normal Q-Q Plot**Resids vs. linear pred.****Histogram of residuals****Response vs. Fitted Values**

Predição do Modelo

Dois Irmãos
Logaritmo



Dois Irmãos



Discussão

- identificação de defasagens e interpretabilidade
- ortogonalidade dos efeitos ambientais e locais
- identificação de sequência *baixa umidade/temperatura* seguida de *alta umidade/temperatura*
- eliminação de redundâncias, seleção de medidas necessárias
- interpretação do efeito espacial *versus covariáveis* e quando usar cada
- detecção de covariáveis relevantes sem padrão espacial
- resultados interpretáveis e razoáveis!
- *modelo para ovos como estrutura (latente) de risco para casos de dengue ???*

Dados, dados, dados ??? !!!

Mas . . . *devagar com o andar*
. . . *que os **dados** são de barro!*

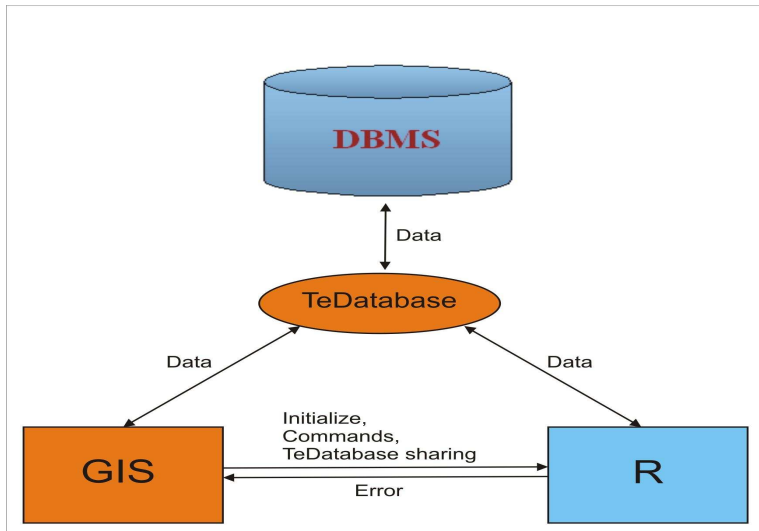
dato real é f . . .

Disponibilizando metodologias e tecnologias

- Implementação de Metodologias:
 - ambiente de prototipação de modelagem estatística e análises:
 - R e diversos pacotes
 - R (+ linguagens: fortran, C, C++, jags);
 - nossos próprios pacotes para adaptações e extensões (RDengue, ...)
- Estrutura computacional:
 - Linux (cross-plataforma);
 - SGBS: Sistemas gerenciadores de bancos de dados (MySQL, postgres, etc);
 - Tecnologias de SIG: Terralib e TerraView
 - **O projeto aRT**: API R-Terralib;
 - integração de ambientes (aRT);
 - geração (automática) de relatórios e visualizações (Sweave/html);
- 100% **software livre**

Disponibilizando resultados

- Visualizações em SIG
- Relatórios WEB
- tudo automatizado
- ... transparência/auditabilidade de código, possibilidade de re-análises, etc

aRT

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;
- interesses especiais em metodologias de estatística espacial;

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;
- interesses especiais em metodologias de estatística espacial;
- em particular formas de combinar modelos físicos e estocásticos;

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;
- interesses especiais em metodologias de estatística espacial;
- em particular formas de combinar modelos físicos e estocásticos;
- **necessidade de desenvolvimento e extensão de metodologias estatísticas;**

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;
- interesses especiais em metodologias de estatística espacial;
- em particular formas de combinar modelos físicos e estocásticos;
- necessidade de desenvolvimento e extensão de metodologias estatísticas;
- *delivering statistics*: necessidade de disponibilizar tecnologias em formas acessíveis, reprodutíveis

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;
- interesses especiais em metodologias de estatística espacial;
- em particular formas de combinar modelos físicos e estocásticos;
- necessidade de desenvolvimento e extensão de metodologias estatísticas;
- *delivering statistics*: necessidade de disponibilizar tecnologias em formas acessíveis, reprodutíveis
- ambientes de prototipação e testagem

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;
- interesses especiais em metodologias de estatística espacial;
- em particular formas de combinar modelos físicos e estocásticos;
- necessidade de desenvolvimento e extensão de metodologias estatísticas;
- *delivering statistics*: necessidade de disponibilizar tecnologias em formas acessíveis, reprodutíveis
- ambientes de prototipação e testagem
- **incorporação vs (?) integração de ferramentas**

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;
- interesses especiais em metodologias de estatística espacial;
- em particular formas de combinar modelos físicos e estocásticos;
- necessidade de desenvolvimento e extensão de metodologias estatísticas;
- *delivering statistics*: necessidade de disponibilizar tecnologias em formas acessíveis, reprodutíveis
- ambientes de prototipação e testagem
- incorporação vs (?) integração de ferramentas
- **ambientes computacionais, software livre, reproducible research**

Discussão

- Questões metodológicas: vasto campo para cross-fertilização;
- interesses especiais em metodologias de estatística espacial;
- em particular formas de combinar modelos físicos e estocásticos;
- necessidade de desenvolvimento e extensão de metodologias estatísticas;
- *delivering statistics*: necessidade de disponibilizar tecnologias em formas acessíveis, reprodutíveis
- ambientes de prototipação e testagem
- incorporação vs (?) integração de ferramentas
- ambientes computacionais, software livre, *reproducible research*
- **RACIONALIZAÇÃO DO USO DE RECURSOS HUMANOS**