

II Simpósio Internacional de Climatologia

*Detecção e atribuição de causas para as mudanças climáticas
na América do Sul*

*Modelos, metodologias e tecnologias estatísticas
com potenciais aplicações em estudos de
mudanças climáticas*

Paulo Justiniano Ribeiro Jr

LEG: Laboratório de Estatística e Geoinformação, UFPR

em colaboração com: *MUITA GENTE!!!*

São Paulo

2-3 de Novembro de 2007

Outline

- Motivação: avaliação de estoques de carbono em solos

Outline

- Motivação: avaliação de estoques de carbono em solos
- Questões e modelos

Outline

- Motivação: avaliação de estoques de carbono em solos
- Questões e modelos
- Desafios metodológicos

Outline

- Motivação: avaliação de estoques de carbono em solos
- Questões e modelos
- Desafios metodológicos
- Disponibilizando Tecnologias

Outline

- Motivação: avaliação de estoques de carbono em solos
- Questões e modelos
- Desafios metodológicos
- Disponibilizando Tecnologias
- **Discussão**

Outline

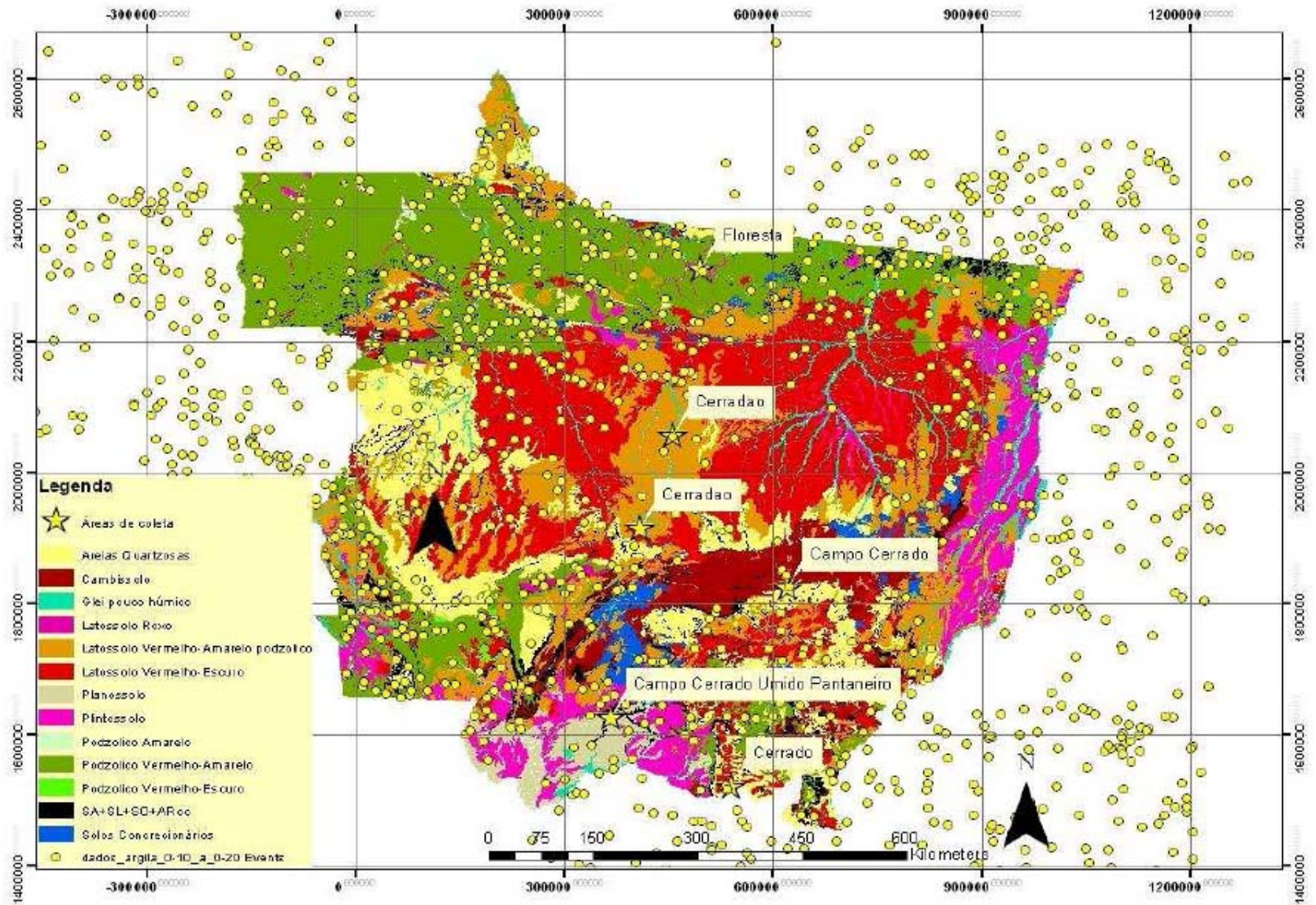
- Motivação: avaliação de estoques de carbono em solos
- Questões e modelos
- Desafios metodológicos
- Disponibilizando Tecnologias
- Discussão
- ...e tudo com um pouco de aleatoriedade e incerteza, é **claro!**

Avaliação de estoques de carbono

- Contatos/parceria com CENA e ESALQ (USP/Piracicaba)
- Impactos da mudança do uso da terra nos estoques de carbono
- BR : maior responsável mundial por emissão devido a mudança no uso da terra
- dois momentos: desmatamento e queimada, operações de preparo do solo e suas consequências (erosão e seus efeitos no aumento das emissões)

Contexto

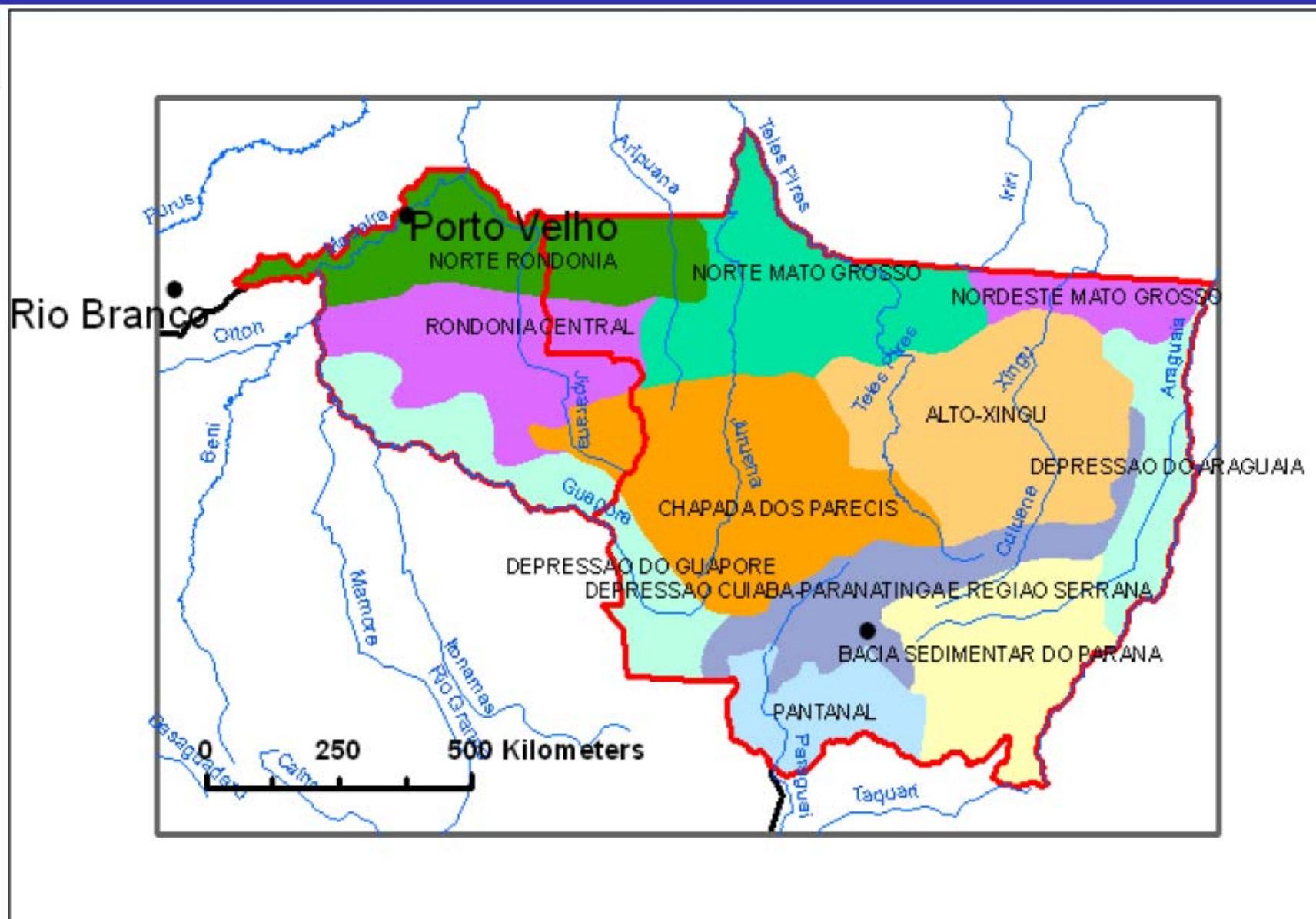
- Amazônia legal: foco em RO e MT
- solo como emissor de GEE (em áreas agrícolas predominam: CO₂, CH₄, NO₂)
- Pressuposto: equilíbrio dinâmico de emissões sob vegetação natural
- i.e., sem variação no estoque C orgânico no tempo,
- o mesmo não ocorre em áreas agrícolas
- C no solo como indicador de qualidade – avaliação buscando sistemas que melhorem estoques



Elementos do projeto (CENA/ESALQ)

Necessidades:

- avaliação de estoques "iniciais"
- medidas diretas de emissão: dificuldades operacionais (câmeras estáticas, operações manuais, limitações de acesso, etc)
- alternativa: remanescentes de sistema nativo e medidas em áreas agrícolas
- (pareamento)
- $C(Mg/ha) = \%C \times D \times Prof \times Fc$
- afetado por vários fatores: clima, topografia, tipos de solo, umidade, composição, etc

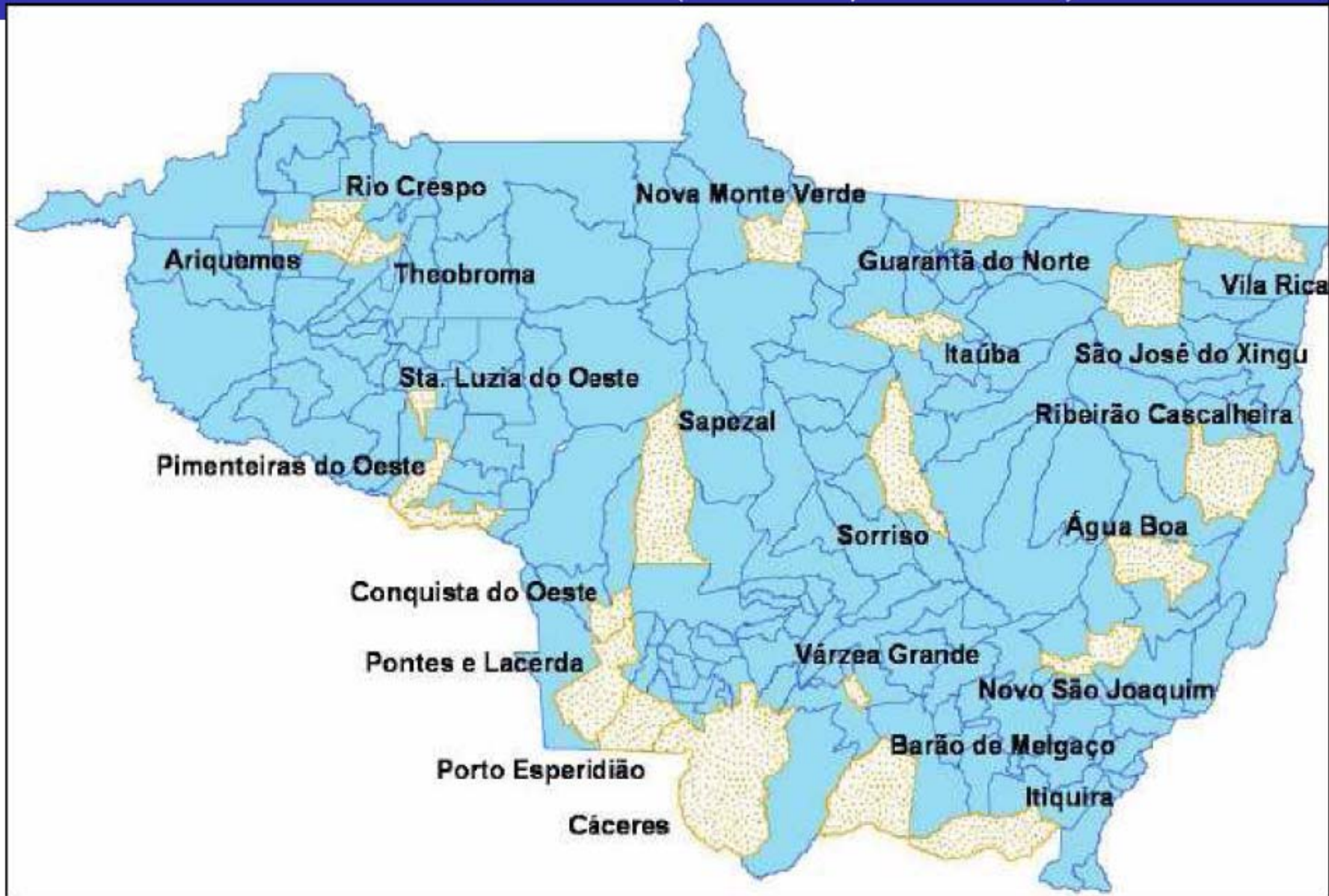


Elementos do projeto (CENA/ESALQ)

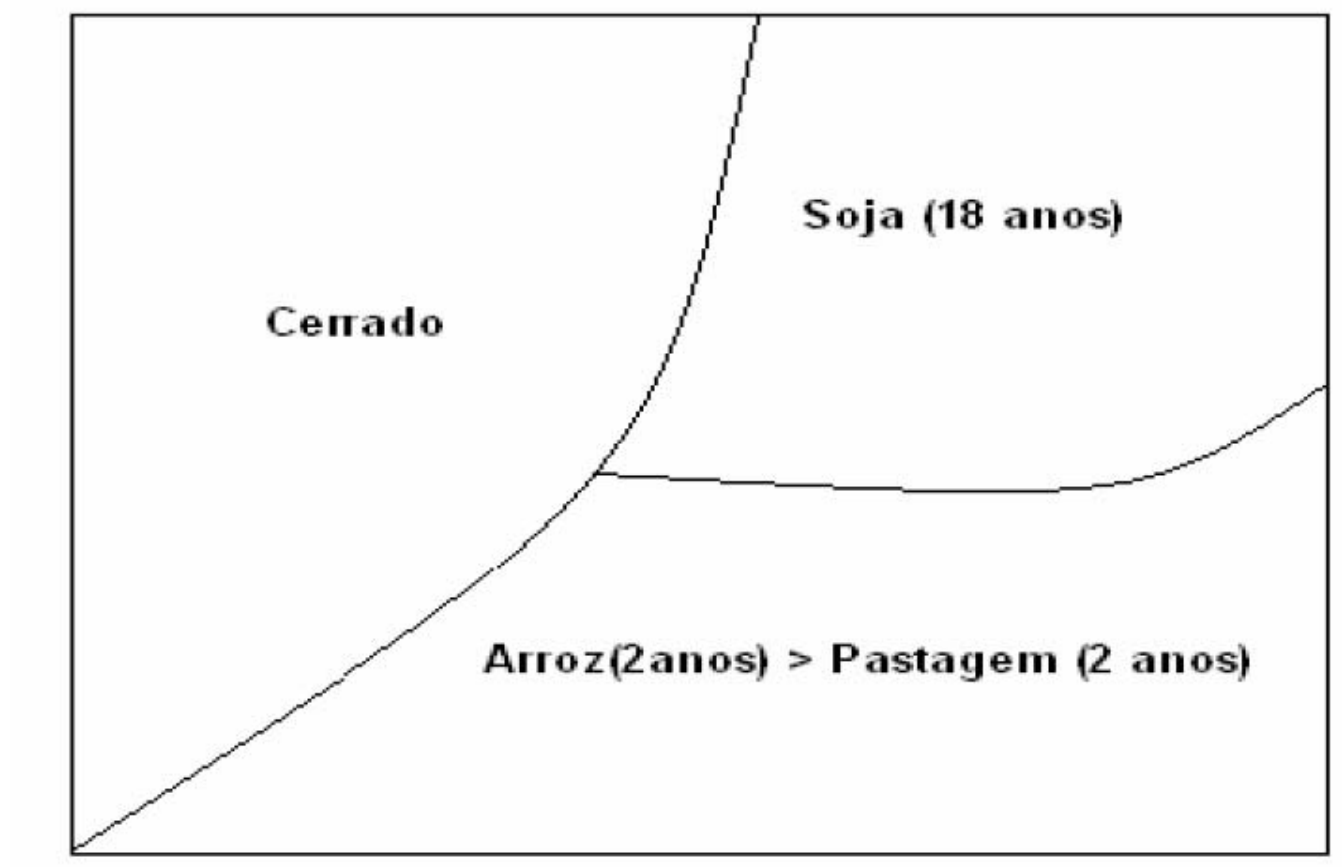
Estudo:

- RO e MT
- **Nível 1: (fixo)** 11 zonas biogeoclimáticas obtidas pela sobreposição de planos de informação (solos, veg. nativa, geologia, clima e relevo)
- **Nível 2: (aleatório)** sorteados 2 municípios em cada zona
- **Nível 3: (aleatório?)** seleção de cronosequências em cada município (desbalanceado)
- **Nível 4:** pares na cronosequência: nativo vs uso(s) agrícola
- FIGS

Elementos do projeto (CENA/ESALQ)



Elementos do projeto (CENA/ESALQ)



Elementos do projeto (CENA/ESALQ)

Características:

- dados sobre: estoques de C, tempo de mudança no uso
medidas de C em áreas agrícolas e nativa
- quantidades principais: C_i, t_i, U_i, d_{ij}
- geometria variando no tempo
- extrapolação para macrozona
- 33 cronosequências

Modelagem

Algumas questões a serem tratadas:

- **modelos hierárquicos**

Modelagem

Algumas questões a serem tratadas:

- modelos hierárquicos
- quantificação de efeitos sob diferentes cenários de uso

Modelagem

Algumas questões a serem tratadas:

- modelos hierárquicos
- quantificação de efeitos sob diferentes cenários de uso
- **dinâmica temporal**

Modelagem

Algumas questões a serem tratadas:

- modelos hierárquicos
- quantificação de efeitos sob diferentes cenários de uso
- dinâmica temporal
- estimativas e avaliação de *incertezas*

Modelagem

Algumas questões a serem tratadas:

- modelos hierárquicos
- quantificação de efeitos sob diferentes cenários de uso
- dinâmica temporal
- estimativas e avaliação de *incertezas*
- avaliação do modelo hierárquico (*e cuidados com falácia ecológica*)

Extensões, desafios e analogias!

- modelagem espacial (ADD-1)

A benchmark model

- $[Y(x_i) | S(x_i)] \sim D_E(\mu_i, (\tau^2))$
- $h(\mu_i) = \sum_{j=1}^p d_{ij}\beta_j + S(x_i) + Z_i$
- covariates d_{ij} , regression parameters β , independent random effects Z_i ;
- $S(x)$ is a (stationary) gaussian process with variance σ^2 and **correlation function** $\rho(\cdot)$.

Inference and Prediction

■ Likelihood function

$$L(\theta) = \int_{\mathbb{R}^n} \prod_i^n f(y_i; h^{-1}(s_i)) f(s | \theta) ds_1, \dots, ds_n$$

Inference and Prediction

- Likelihood function

$$L(\theta) = \int_{\mathbb{R}^n} \prod_i^n f(y_i; h^{-1}(s_i)) f(s | \theta) ds_1, \dots, ds_n$$

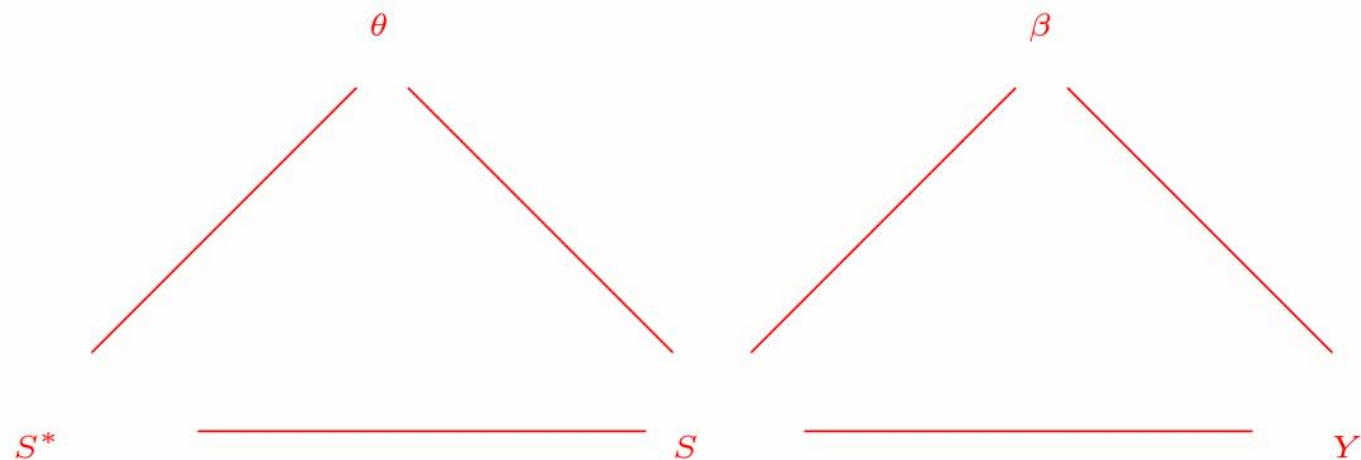
- Involves a high-dimensional (numerical) integration

Inference and Prediction

- Likelihood function

$$L(\theta) = \int_{\mathbb{R}^n} \prod_i^n f(y_i; h^{-1}(s_i)) f(s | \theta) ds_1, \dots, ds_n$$

- Involves a high-dimensional (numerical) integration
- MCMC algorithms can exploit the conditional independence structure of the model



Some words on spatio-temporal models

- may depend on time resolution and space resolution, e.g.;
 - multivariate time series
 - multivariate random fields;
 - random fields as priors;
 - geostatistical modelling.

Some words on spatio-temporal models

- may depend on time resolution and space resolution, e.g.;
 - multivariate time series
 - multivariate random fields;
 - random fields as priors;
 - geostatistical modelling.
- *model-based approach and geostatistical approach*

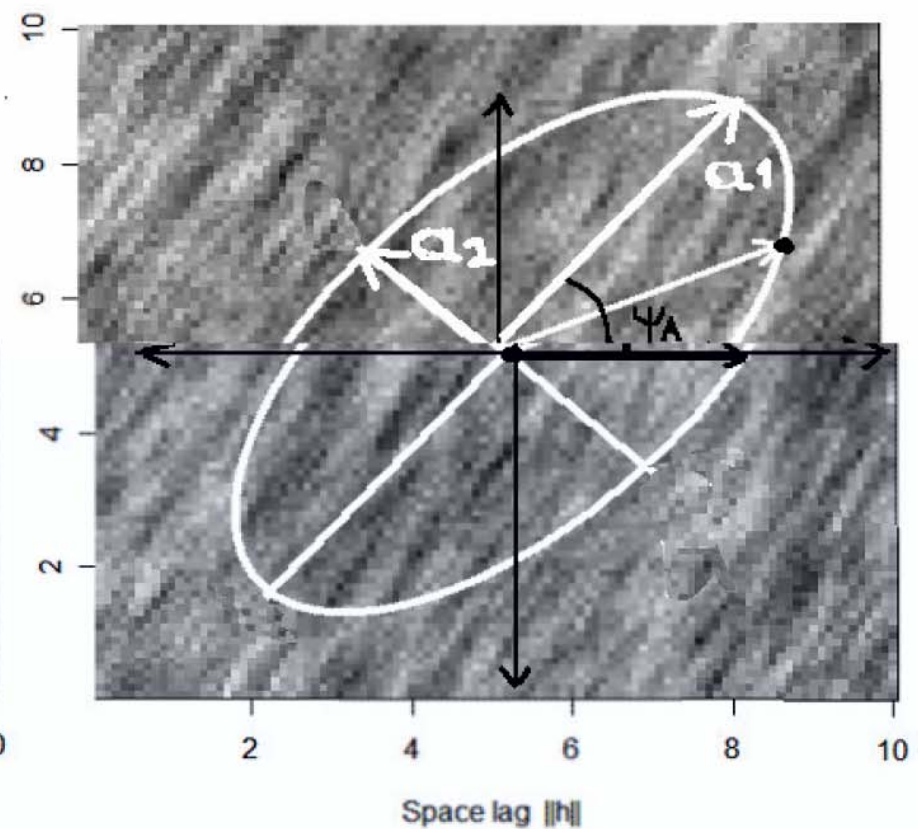
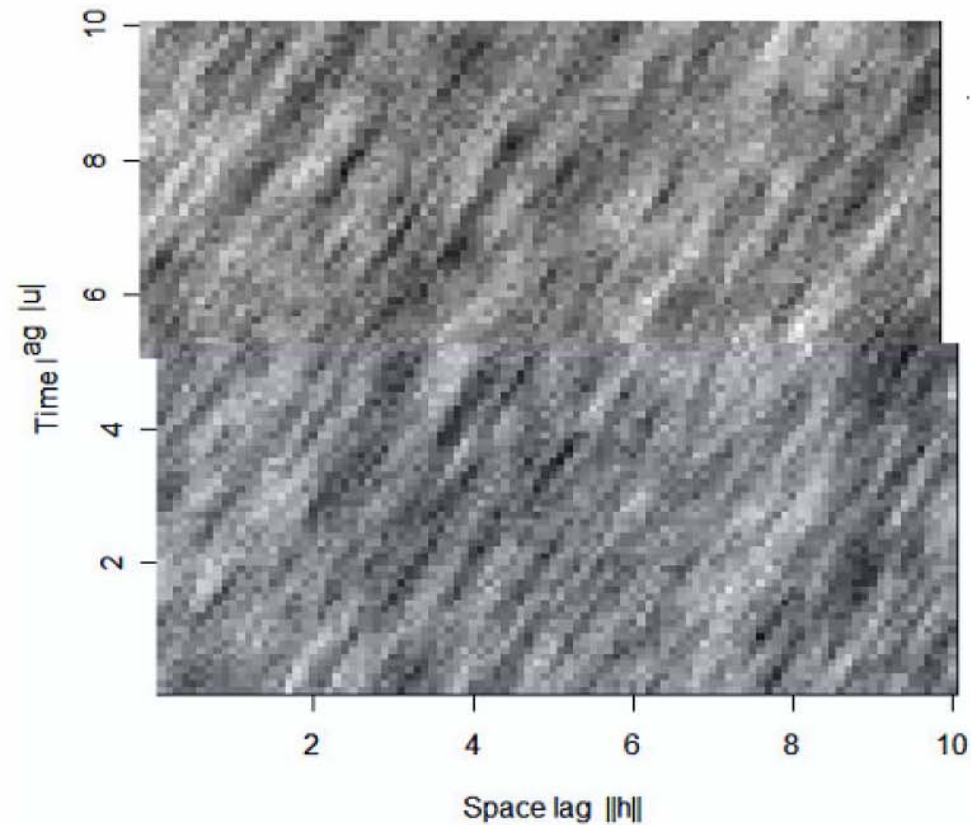
Separability

- Separable models;
 - $\text{Cov}(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = \text{Cov}(Z(\mathbf{s}_1, \mathbf{s}_2)) + \text{Cov}(Z(t_1, t_2))$
 - $\text{Cov}(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = \text{Cov}(Z(\mathbf{s}_1, \mathbf{s}_2)) \text{Cov}(Z(t_1, t_2))$,
 - with \mathbf{s}_1, t_1 and $\mathbf{s}_2, t_2 \in \mathbf{R}^2 \times \mathbf{R}$.

Separability

- Separable models;
 - $\text{Cov}(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = \text{Cov}(Z(\mathbf{s}_1, \mathbf{s}_2)) + \text{Cov}(Z(t_1, t_2))$
 - $\text{Cov}(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = \text{Cov}(Z(\mathbf{s}_1, \mathbf{s}_2)) \text{Cov}(Z(t_1, t_2))$,
 - with \mathbf{s}_1, t_1 and $\mathbf{s}_2, t_2 \in \mathbf{R}^2 \times \mathbf{R}$.
- non-separable geostatistical models
 - Cressie & Huang (1999)
 - Gneiting (2002)
 - others ...

Describing non-separable models using anisotropy



Gneiting (2002)

- *Fourier free approach*
- *ingredients*: monotone function $\phi(u)$, $u \geq 0$ and positive function $\psi(u)$, $u \geq 0$
- a valid spatio-temporal covariance function in $R^d \times R$ is given by

$$C(\mathbf{h}, u) = \frac{\sigma^2}{(\psi(u)^2)^{d/2}} \phi\left(\frac{\|\mathbf{h}\|^2}{\psi(|u|^2)}\right)$$

- with some conditions for $\phi(u)$ and $\psi(u)$

Some choices for $\phi(u)$ and $\psi(u)$

Completely monotone functions $\phi(u), u \geq 0$

$$\phi(u) = \exp(-cu^\gamma), c > 0, 0 < \gamma \leq 1$$

$$\phi(u) = (1 + cu^\gamma)^\nu, c > 0, 0 < \gamma \leq 1, \nu > 0$$

$$\phi(u) = (2^\nu \Gamma(\nu))^{-1} (cu^{1/2})^\nu \mathbf{K}_\nu(cu^{1/2}), c > 0, \nu > 0$$

$$\phi(u) = 2^\nu (\exp(cu^{1/2}) + \exp(-cu^{1/2}))^\nu, c > 0, \nu > 0$$

Positive functions $\psi(u), u \geq 0$

$$\psi(u) = (au^\alpha + 1)^\beta, a > 0, 0 < \alpha \leq 1, 0 \leq \beta \leq 1$$

$$\psi(u) = \ln(au^\alpha + b) / \ln(b), a > 0, b > 1, 0 < \alpha \leq 1$$

$$\psi(u) = (au^\alpha + b) / (b(au^\alpha + 1)), a > 0, 0 < b \leq 1$$

An example

- choosing:

- $\phi(u) = \exp(-cu^\gamma), c > 0, 0 < \gamma \leq 1$
- $\psi(u) = (au^\alpha + 1)^\beta, a > 0, 0 < \alpha \leq 1, 0 \leq \beta \leq 1$

- the covariance function is given by

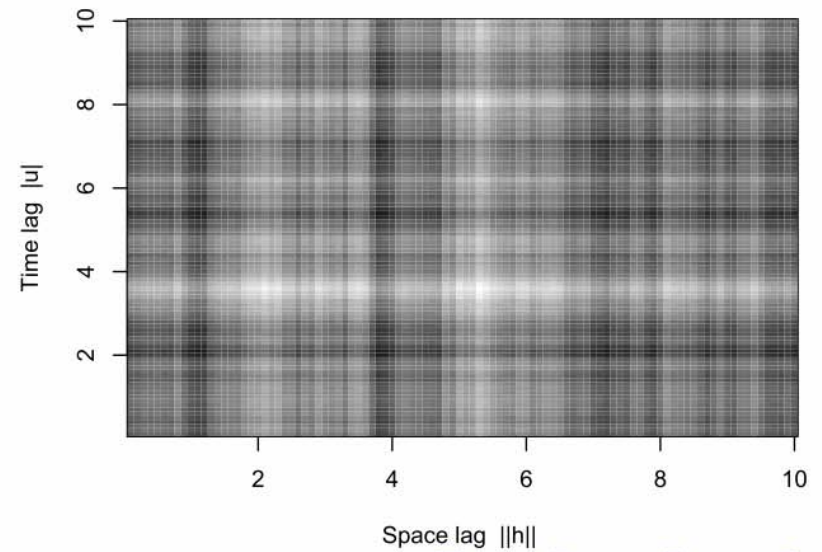
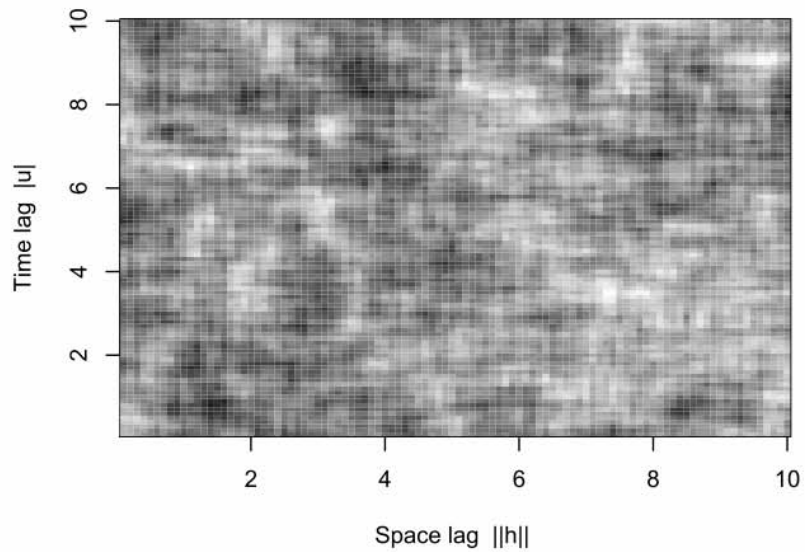
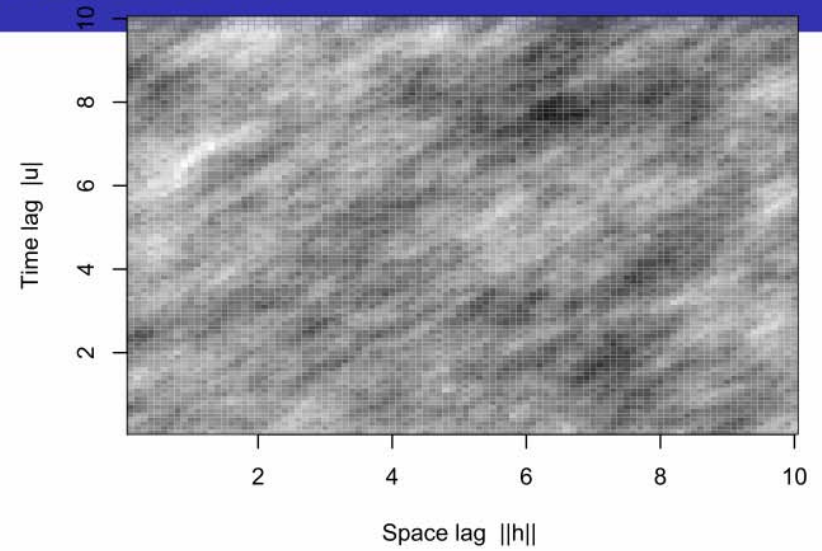
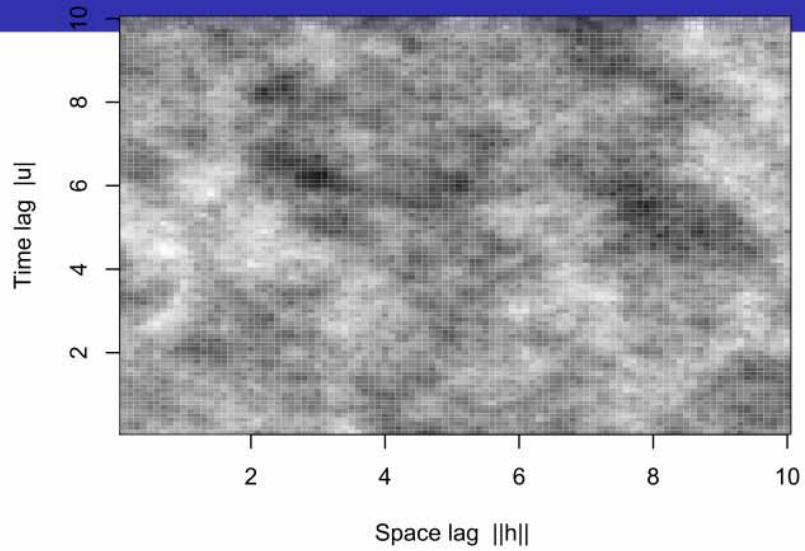
$$\text{Cov}(\mathbf{h}, t) = \frac{\sigma^2}{(a|t|^{2\alpha} + 1)^{\frac{\beta d}{2}}} \exp\left\{-\frac{c\|\mathbf{h}\|^{2\gamma}}{(a|t|^{2\alpha} + 1)^{\beta\gamma}}\right\}.$$

- taking $\beta = 0$, and multiplying by $(at^\alpha + 1)^\delta$:

$$\text{Cov}(\mathbf{h}, u) = \frac{\sigma^2}{(a|t|^{2\alpha} + 1)^{\delta + \frac{\beta d}{2}}} \exp\left\{-\frac{c\|\mathbf{h}\|^{2\gamma}}{(a|t|^{2\alpha} + 1)^{\beta\gamma}}\right\},$$

- so with $\beta = 0$ it reduces to a separable form

Some simulations



Extensões, desafios e analogias!

- modelagem espacial (ADD-1)
- **delineamentos**

Extensões, desafios e analogias!

- modelagem espacial (ADD-1)
- delineamentos
- combinação de modelos espaciais em diferentes escalas

Extensões, desafios e analogias!

- modelagem espacial (ADD-1)
- delineamentos
- combinação de modelos espaciais em diferentes escalas
- possibilidade de uso mais geral de cronosequências (HMM ?)

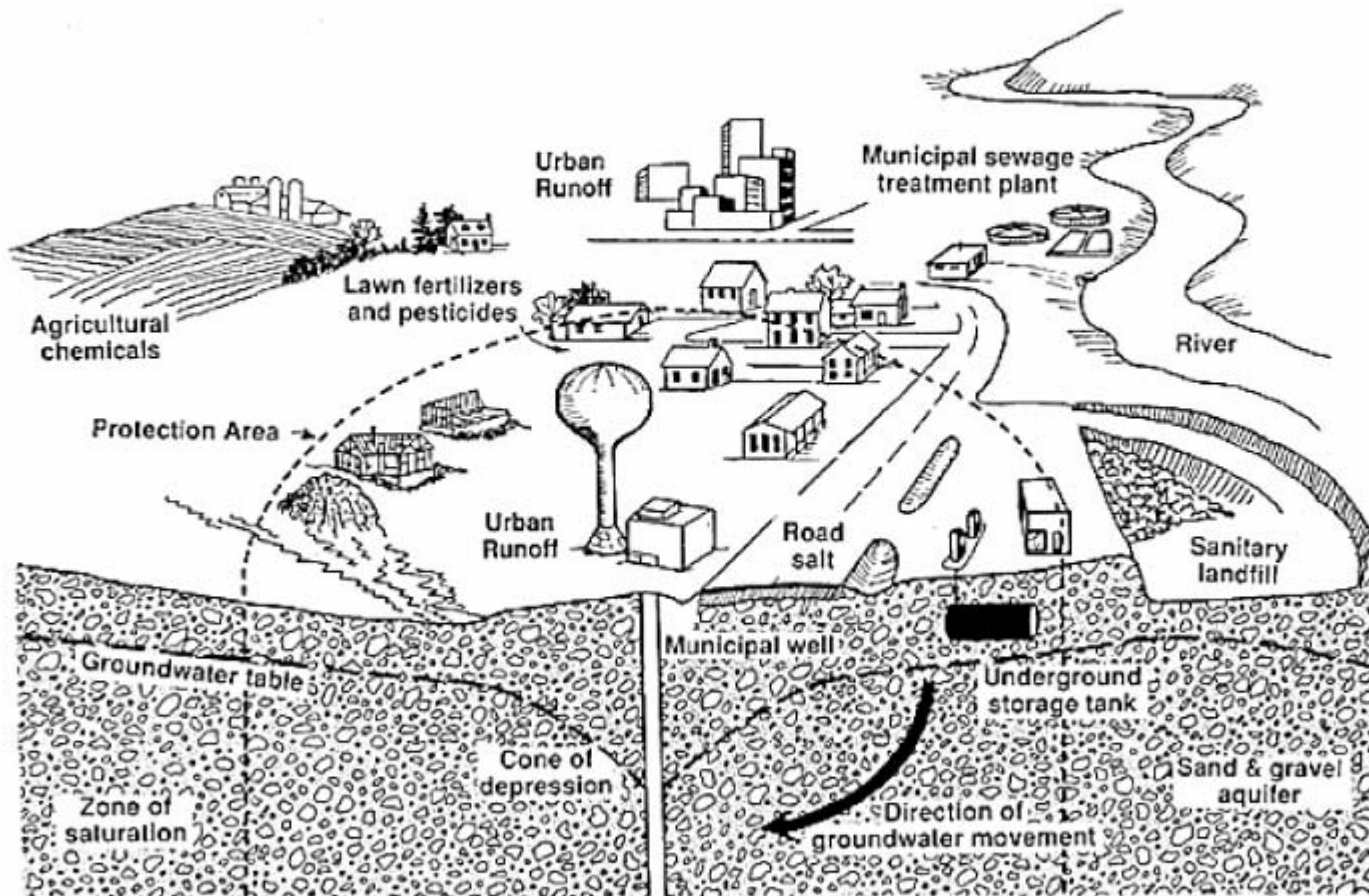
Extensões, desafios e analogias!

- modelagem espacial (ADD-1)
- delineamentos
- combinação de modelos espaciais em diferentes escalas
- possibilidade de uso mais geral de cronosequências (HMM ?)
- informações acessórias e modelagem conjunta para melhor descrição de toda área (ADD-2)

Extensões, desafios e analogias!

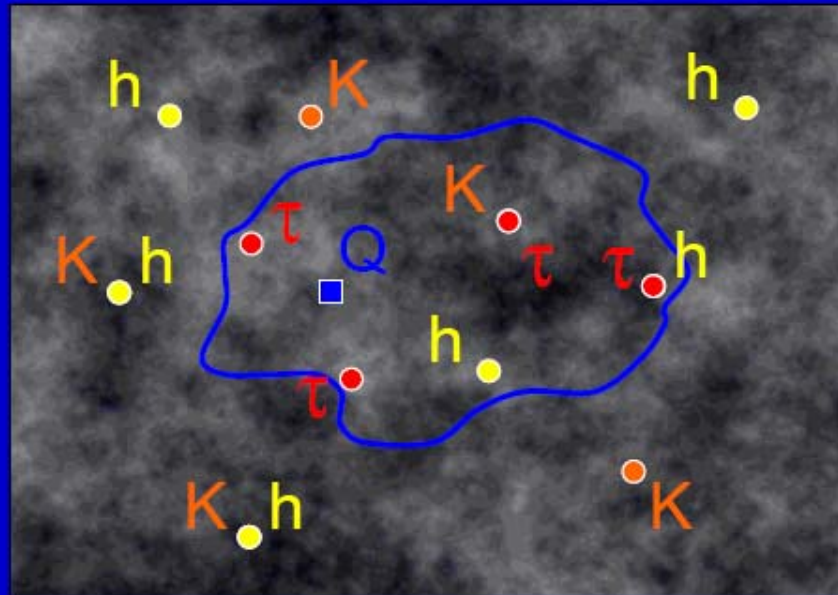
- modelagem espacial (ADD-1)
- delineamentos
- combinação de modelos espaciais em diferentes escalas
- possibilidade de uso mais geral de cronosequências (HMM ?)
- informações acessórias e modelagem conjunta para melhor descrição de toda área (ADD-2)
- combinação de modelos físicos e estocásticos e com avaliação de incertezas (ADD-3)

Problem Definition



Objectives

- stochastic delineation of well capture zones



- ✓ quantify the uncertainty of the predictions
- ✓ assess the value of data

Spatial stochastic approach

- stochastic model to represent spatial variability
 - ✓ $Y(\mathbf{x}) = \log K(\mathbf{x})$ is a normally distributed stationary RSF

$$\langle Y(\mathbf{x}) \rangle = \mu_Y$$

$$\mathbf{V}_Y(\mathbf{r}) = \langle Y'(\mathbf{x}+\mathbf{r}) Y'(\mathbf{x}) \rangle = \sigma_Y^2 \rho_Y(\mathbf{r})$$

- ✓ prediction 

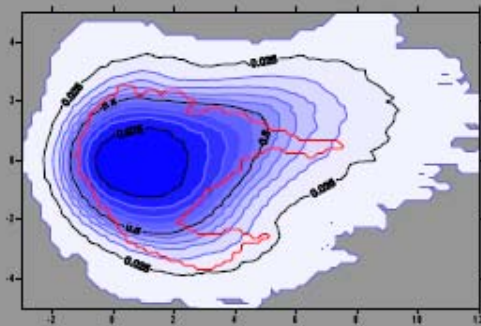
→ Uncertain parameters: $\theta \rightarrow \mu_Y, \sigma_Y^2$ and ϕ_Y

Aquifer properties

- hydraulic conductivity K
 - saturated thickness D
 - effective porosity n_e
- exhibit heterogeneity in space
- } transmissivity T



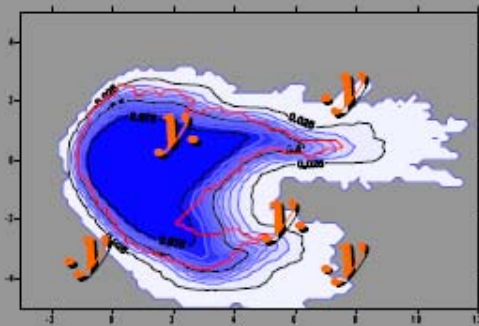
Conditioning on hydraulic conductivity measurements

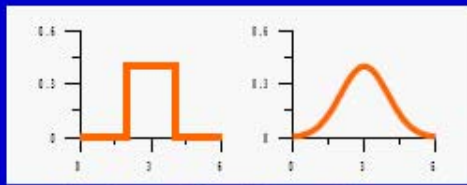


?

→

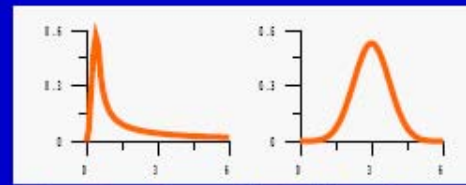
$y = \log K$





Prior distributions

$$y = \log K$$

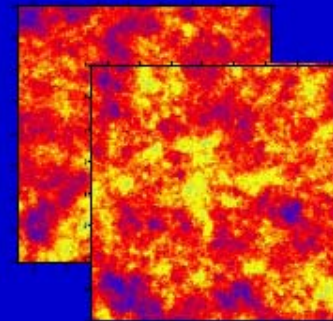


Posterior distributions

Monte Carlo

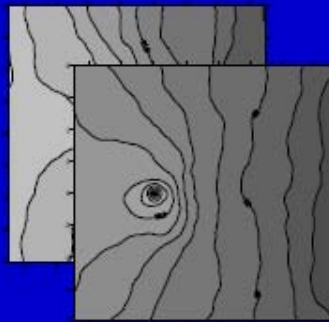
Sampled θ_i :
 $V(\sigma^2, \varphi)$ and μ

Conditional
simulation



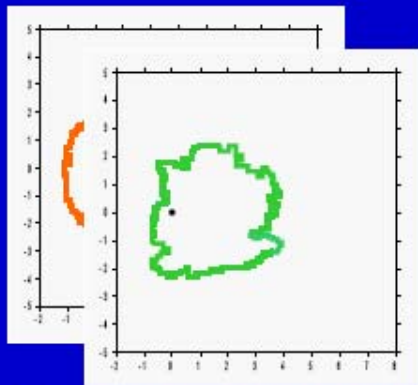
Conditional K -field

Flow simulation



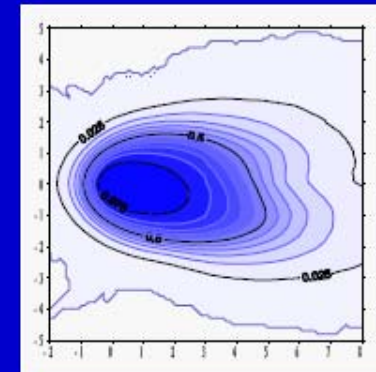
Head distribution

Particle
tracking



Capture zone $\{f(t)\}$

Statistical analysis



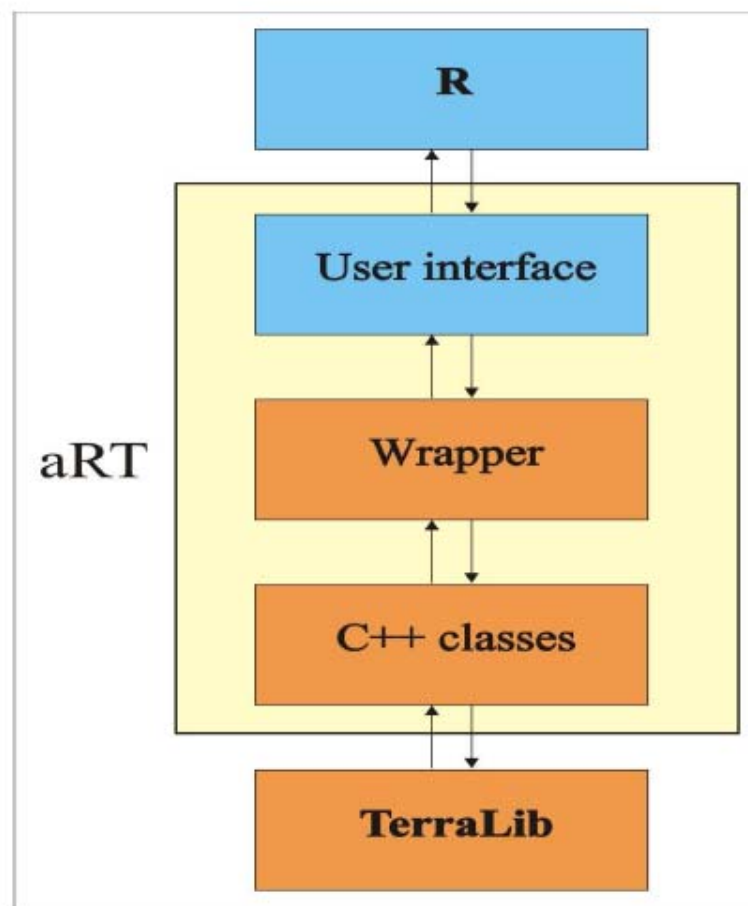
Stochastic capture
zone $\{f(t)\}$

Methodology

Extensões, desafios e analogias!

- modelagem espacial (ADD-1)
- delineamentos
- combinação de modelos espaciais em diferentes escalas
- possibilidade de uso mais geral de cronosequências (HMM ?)
- informações acessórias e modelagem conjunta para melhor descrição de toda área (ADD-2)
- combinação de modelos físicos e estocásticos e com avaliação de incertezas (ADD-3)
- **proporções de emissões (ADD-4)**

aRT



Disponibilizando metodologias

- **O projeto aRT: API R-Terralib;**

Disponibilizando metodologias

- **O projeto aRT:** API R-Terralib;
- **SGBS:** Sistemas gerenciadores de bancos de dados (MySQL, postgres, etc);

Disponibilizando metodologias

- **O projeto aRT:** API R-Terralib;
- SGBS: Sistemas gerenciadores de bancos de dados (MySQL, postgres, etc);
- ambiente de SIG: Terralib;

Disponibilizando metodologias

- **O projeto aRT:** API R-Terralib;
- SGBS: Sistemas gerenciadores de bancos de dados (MySQL, postgres, etc);
- ambiente de SIG: Terralib;
- ambiente de prototipação de modelagem estatística e análises:R (+ linguagens: fortran, C, C++);

Disponibilizando metodologias

- **O projeto aRT:** API R-Terralib;
- SGBS: Sistemas gerenciadores de bancos de dados (MySQL, postgres, etc);
- ambiente de SIG: Terralib;
- ambiente de prototipação de modelagem estatística e análises:R (+ linguagens: fortran, C, C++);
- geração (automática) de relatórios e visualizações (Sweave/html);

Disponibilizando metodologias

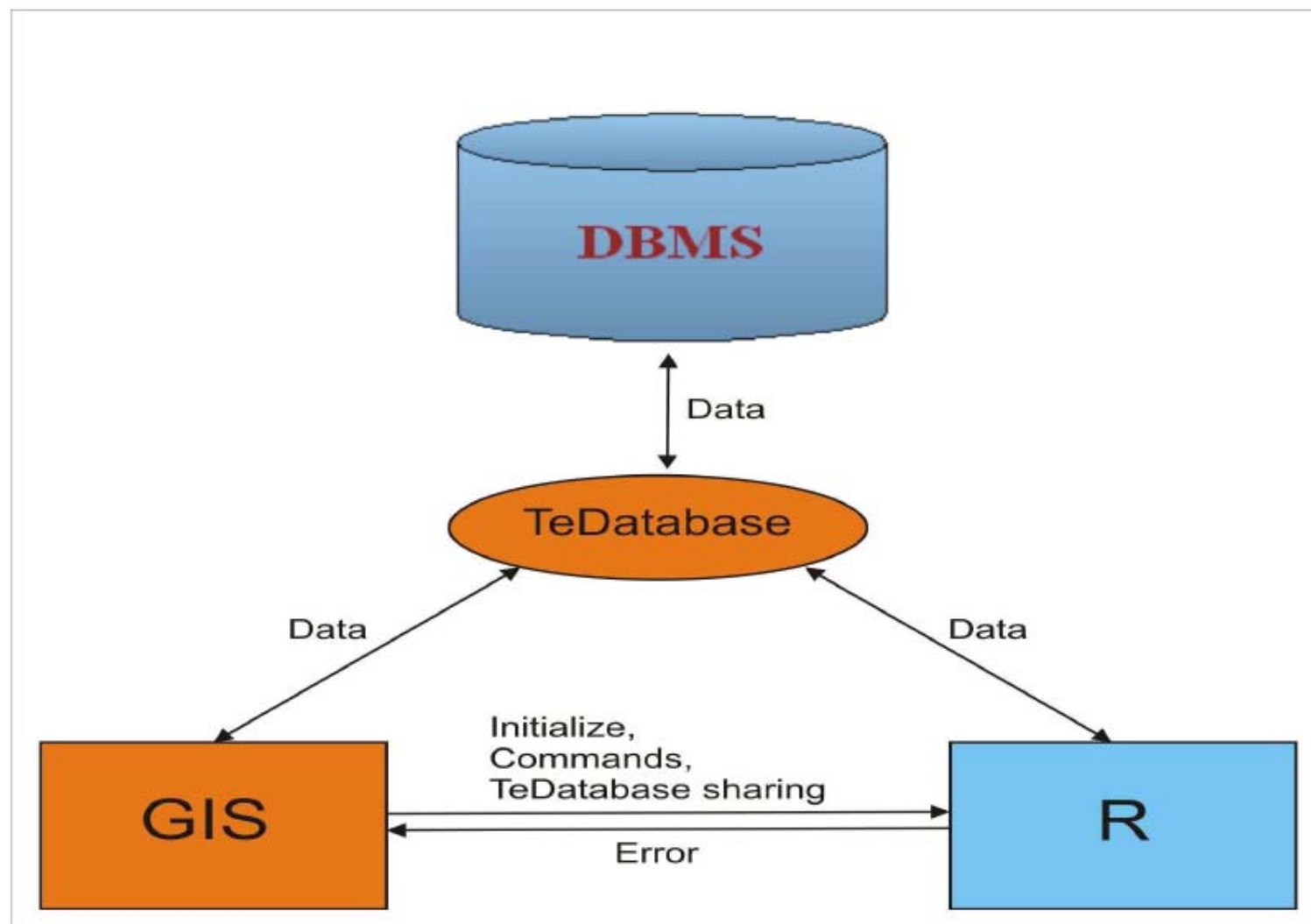
- **O projeto aRT:** API R-Terralib;
- SGBS: Sistemas gerenciadores de bancos de dados (MySQL, postgres, etc);
- ambiente de SIG: Terralib;
- ambiente de prototipação de modelagem estatística e análises:R (+ linguagens: fortran, C, C++);
- geração (automática) de relatórios e visualizações (Sweave/html);
- **integração de ambientes (aRT);**

Disponibilizando metodologias

- **O projeto aRT:** API R-Terralib;
- SGBS: Sistemas gerenciadores de bancos de dados (MySQL, postgres, etc);
- ambiente de SIG: Terralib;
- ambiente de prototipação de modelagem estatística e análises:R (+ linguagens: fortran, C, C++);
- geração (automática) de relatórios e visualizações (Sweave/html);
- integração de ambientes (aRT);
- **100% software livre**

DEMO DO aRT

Ilustrando o uso!

aRT

Algumas aplicações

- estruturas de dados grandes e complexas;

Algumas aplicações

- estruturas de dados grandes e complexas;
- entrada frequente de dados;

Algumas aplicações

- estruturas de dados grandes e complexas;
- entrada frequente de dados;
- atualização constante de análises e resultados;

Algumas aplicações

- estruturas de dados grandes e complexas;
- entrada frequente de dados;
- atualização constante de análises e resultados;
- geração (automática) de relatórios e visualizações;

Algumas aplicações

- estruturas de dados grandes e complexas;
- entrada frequente de dados;
- atualização constante de análises e resultados;
- geração (automática) de relatórios e visualizações;
- necessidade de tecnologias para suporte às análises e disponibilização de resultados.

Comentários

- Questões metodológicas: vasto campo para cross-fertilização;

Comentários

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

Comentários

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

Comentários

- 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 estatística;**

Comentários

- 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 estatística;
- *delivering statistics*: necessidade de disponibilizar tecnologias em formas acessíveis, reprodutíveis

Comentários

- 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 estatística;
- *delivering statistics*: necessidade de disponibilizar tecnologias em formas acessíveis, reprodutíveis
- ambientes de prototipação e testagem

Comentários

- 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 estatística;
- *delivering statistics*: necessidade de disponibilizar tecnologias em formas acessíveis, reprodutíveis
- ambientes de prototipação e testagem
- **incorporação vs (?) integração de ferramentas**

Comentários

- 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 estatística;
- *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*