Coregionalized Single and Multi-resolution Spatially-Varying Growth Curve Modelling

Sudipto Banerjee and Gregg A. Johnson

sudiptob@biostat.umn.edu

University of Minnesota

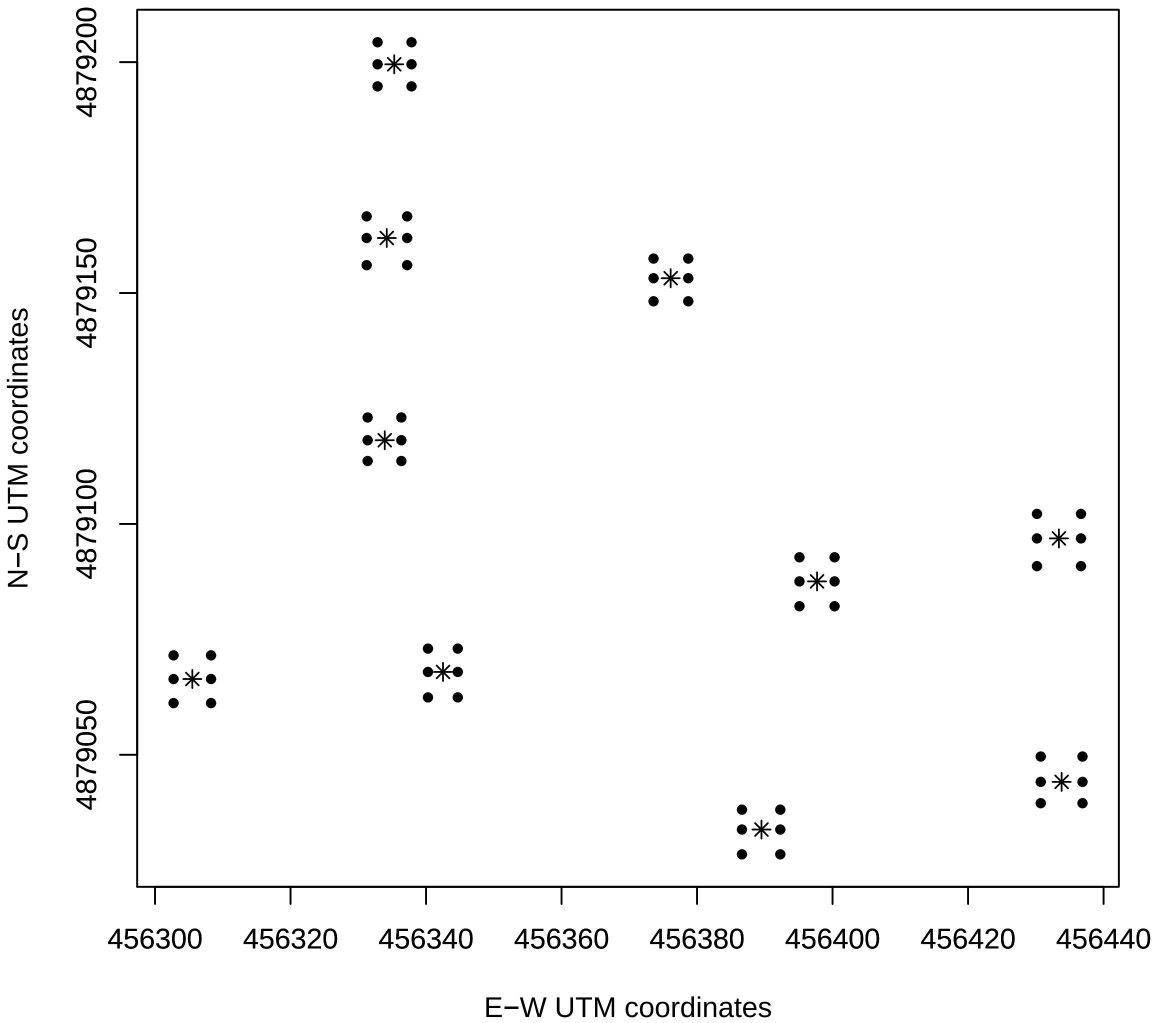
Foxtail - the villain

Damage to crops from weed infiltration estimated at $1.2mn last year in the US. Foxtail accounts for about 40%

Weed growth studies

- Better understand conditions conducive to weed growth.
- Study the growth patters of weeds.
- Capture *spatial variation* in growth patterns
- Help identify “no-cultivation” zones.
- Designed experiments for growing weeds:
  - Based on soil and terrain analysis a set of plots are chosen
  - Each site has “replicates” (subplots).
  - Fields were tilled (eliminate existing plants); natural seedbed prepared.
  - Germination and emergence takes about 7–14 days. After this, weed density in each subplot is measured and monitored through time.

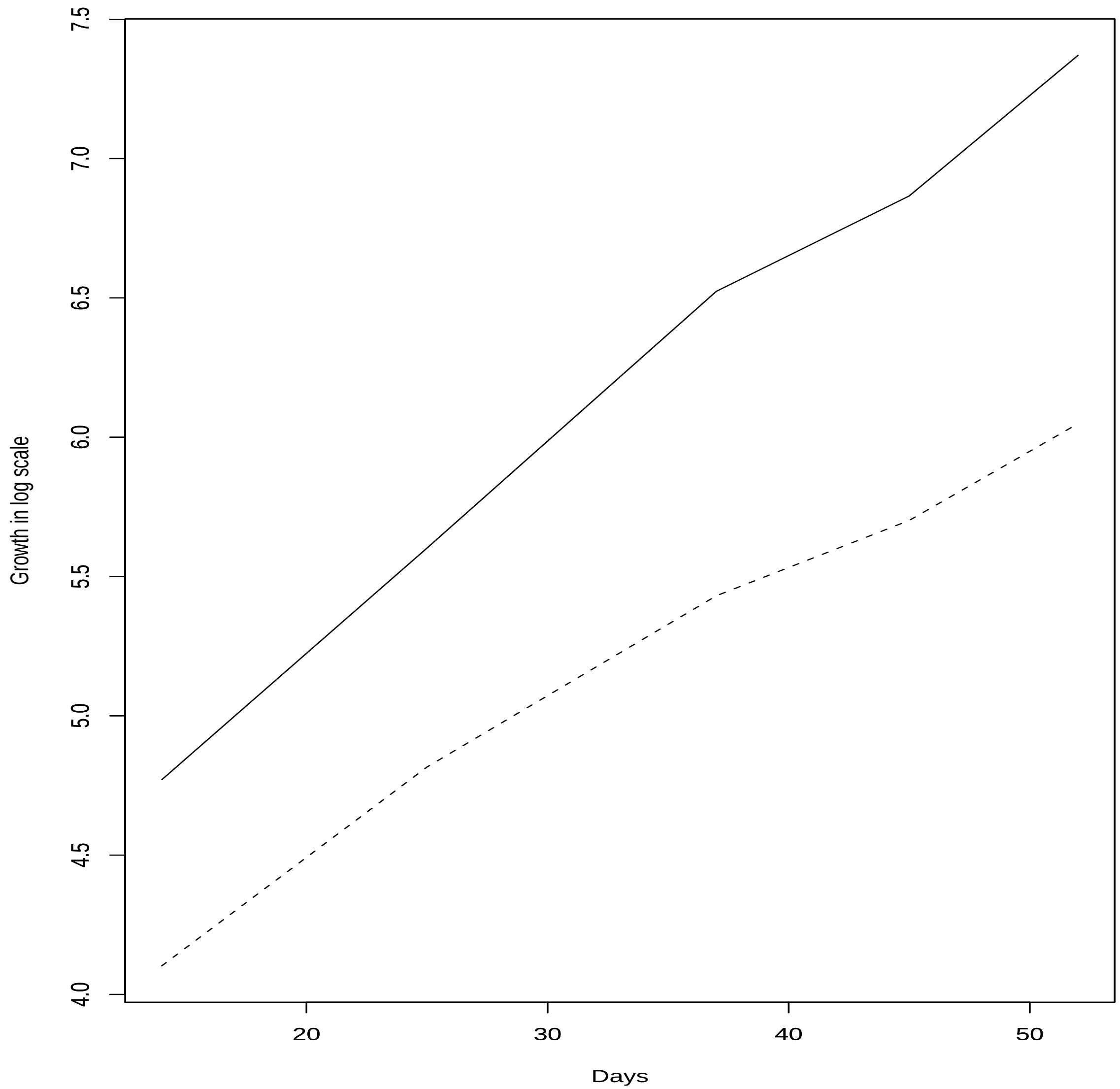
Weed Growth locations: Spatial Domain



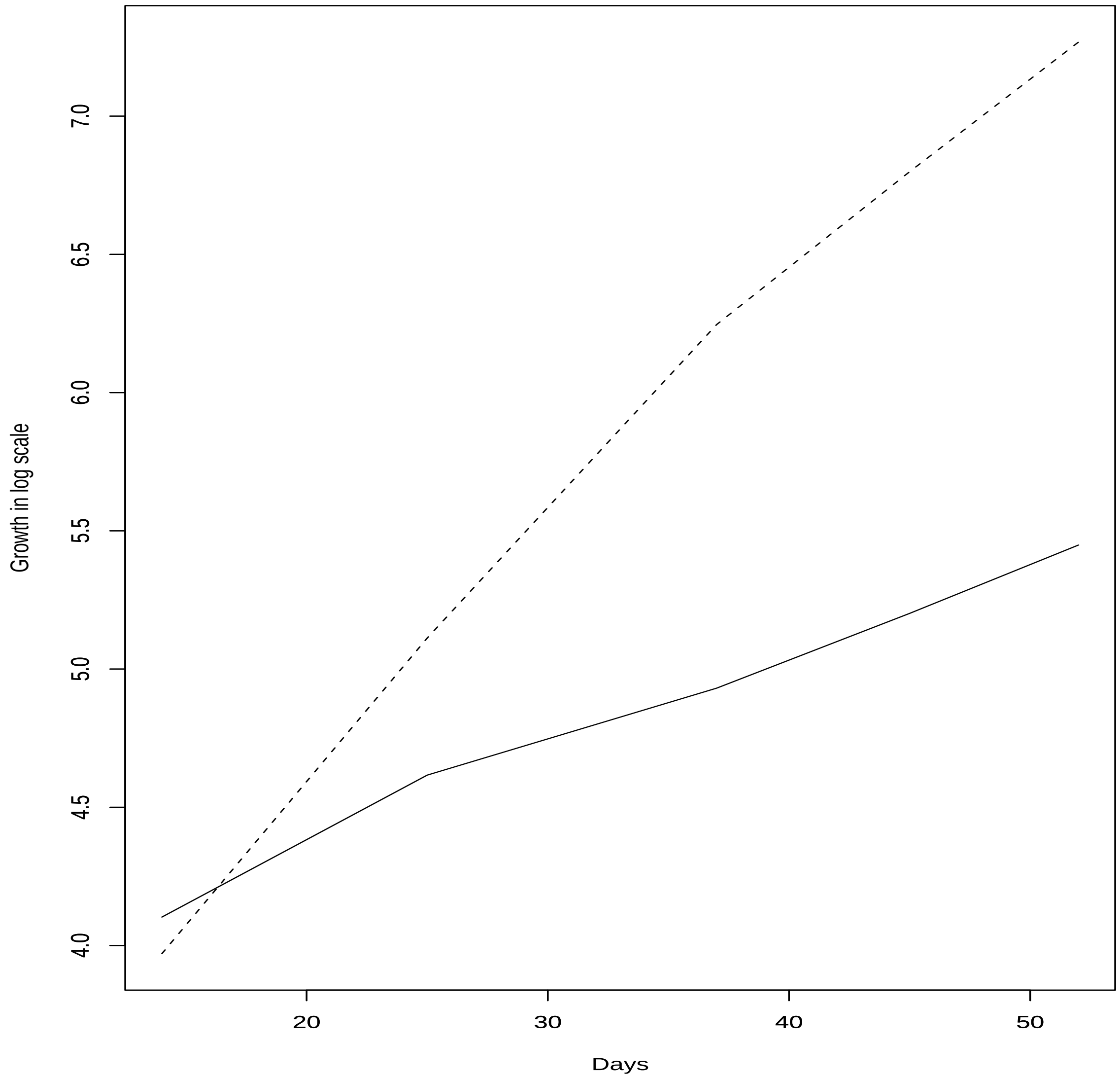
Main plots and Subplots

- Our spatially-replicated experimental setup for weed-growth involves $N_s = 10$ main plots, each associated with a rectangular array comprising $N_r(s) = 6$ subplots (replicates) on which weed density is monitored on a daily basis (DAP: days after planting).
- Traditional analysis:
  - Longitudinal modelling using growth curves.
  - Scant regard for spatial nature of weed growth.
- Objectives: Model growth curves as spatially correlated. Account for similarity in growth patterns for closer locations
- Predict baseline growth and growth rate over domain

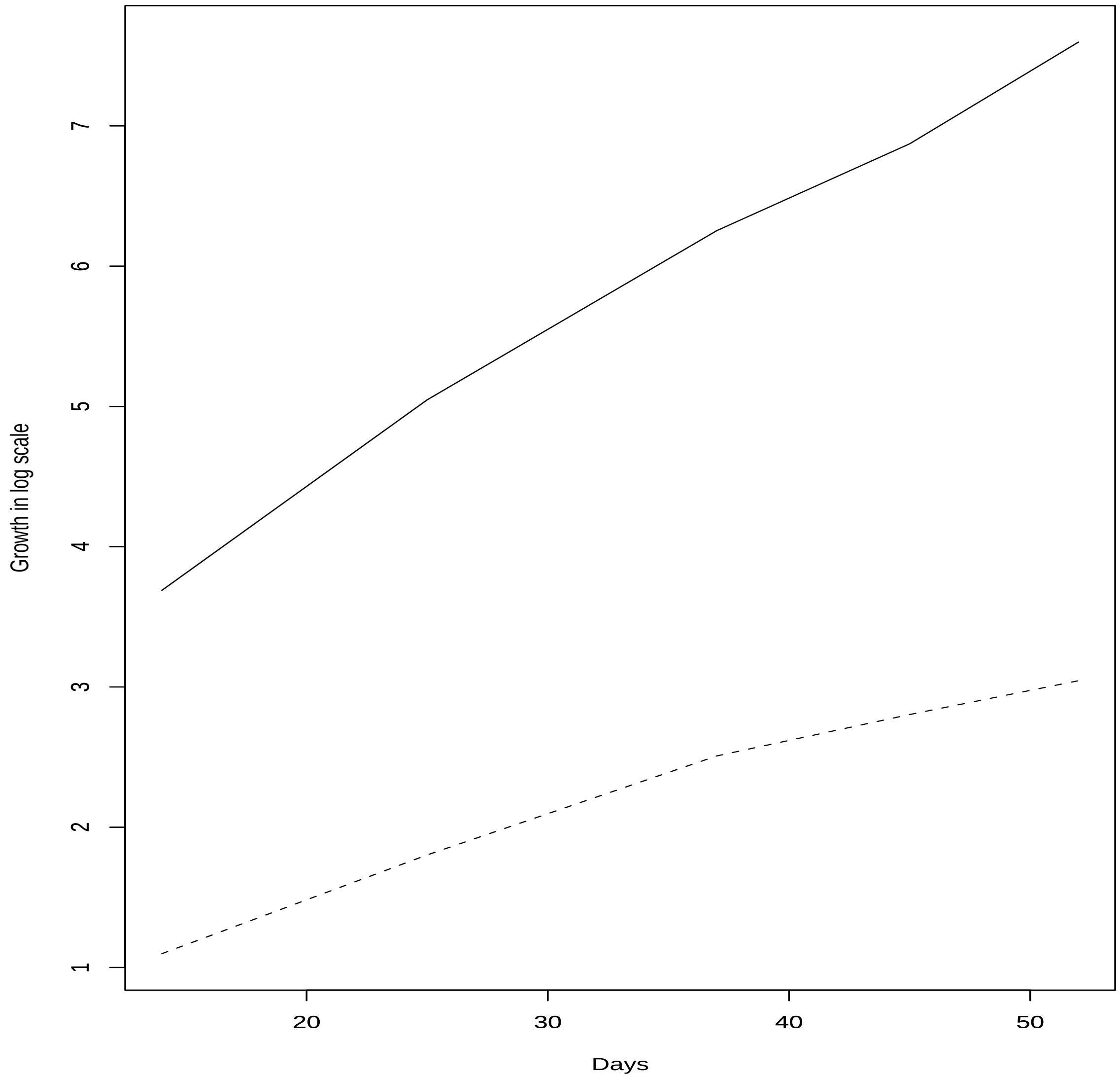
Micro-level variation



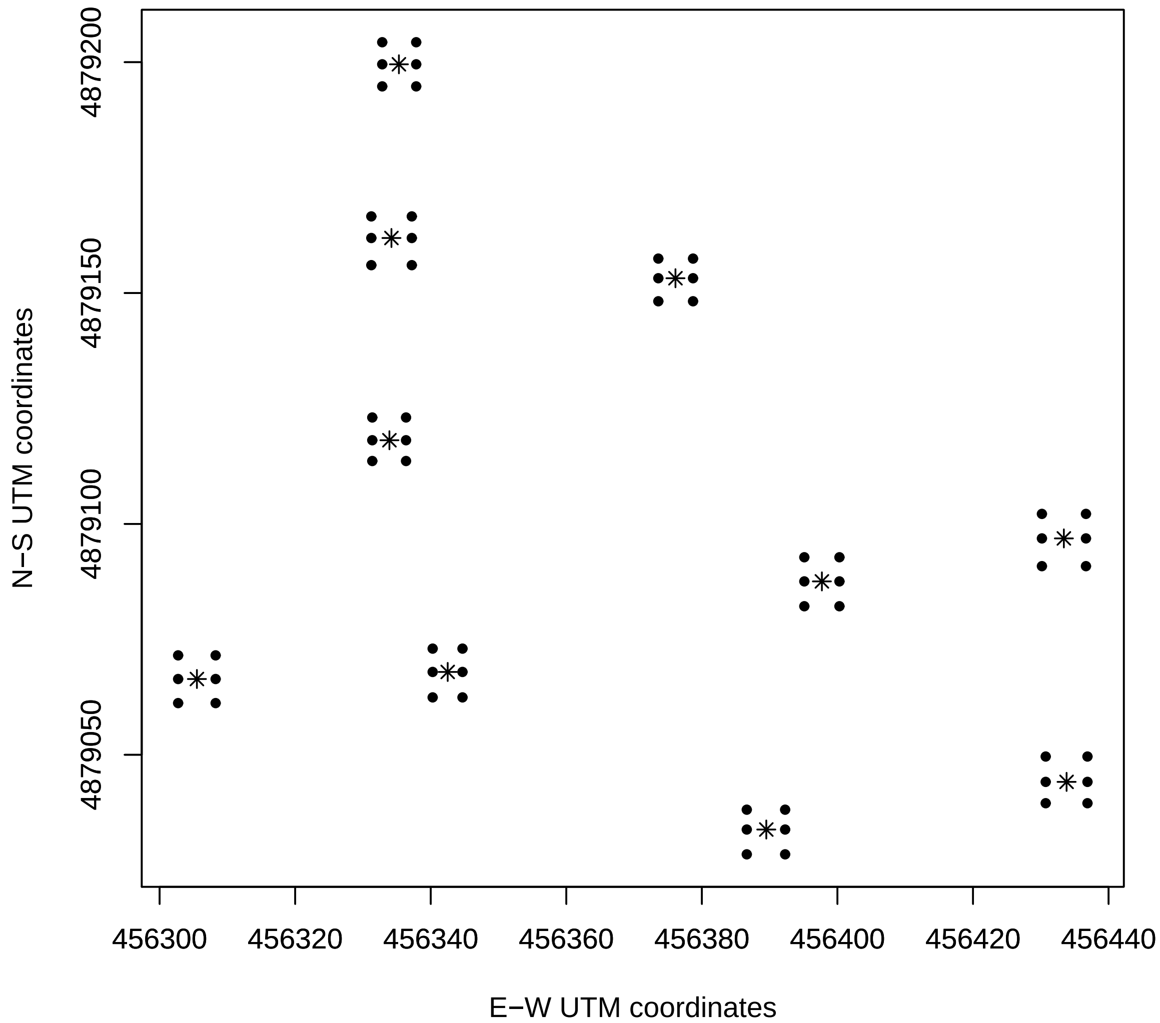
Subplots with same ROW-COL effect



Completely general Macro effect



Spatial Questions?



Non-spatial growth curve models

- Let Y_{it} be the response (weed density in the log scale recorded for location i in time t). Our basic growth curve model is

$$Y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + f_i(t) + \epsilon_{it},\ i = 1, \ldots, N,\ \epsilon_{it} \sim N(0, \tau^2)$$

Non-spatial growth curve models

- Let Y_{it} be the response (weed density in the log scale recorded for location i in time t). Our basic growth curve model is

$$Y_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta} + f_i(t) + \epsilon_{it},\ i = 1, \ldots, N,\ \epsilon_{it} \sim N(0, \tau^2)$$

- Examples:

Model 1a: $f_i(t) = \alpha_0 + \alpha_1 t;\ \boldsymbol{\alpha} \sim N(\mathbf{0}, Diag(\sigma_0^2, \sigma_1^2))$

Model 1b: $f_i(t) = \alpha_0 + \alpha_1 t;\ \boldsymbol{\alpha} \sim N(\mathbf{0}, \Lambda)$

Model 1c: $f_i(t) = \alpha_{0i} + \alpha_{1i} t;\ \boldsymbol{\alpha}_i \sim N(\mathbf{0}, \Lambda)$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$ and $\boldsymbol{\alpha}_i = (\alpha_{0i}, \alpha_{1i})$.

Single resolution models

- Similar growth patterns in proximate locations. Growth curves likely to be *spatially associated*.
- Ignore nested structure.

$$Y_t(s) = \mathbf{x}_t(s)^T \boldsymbol{\beta} + f(s,t) + \epsilon_t(s),\ \epsilon_t(s) \sim N(0,\tau^2)$$

where $f(s,t) = \alpha_0(s) + \alpha_1(s)t$.

- $\boldsymbol{\alpha}(s) = (\alpha_0(s), \alpha_1(s)) \sim GP(\boldsymbol{\mu}(s), \Gamma_\alpha(\cdot))$.
- When process means are constant across sites:

$$\begin{aligned} f\left(s,t\right) &= \mu_0 + \mu_1 t + \tilde{\alpha}_0\left(s\right) + \tilde{\alpha}_1\left(s\right)t \\ &= g\left(t\right) + \tilde{f}\left(s,t\right). \end{aligned}$$

Cross-covariance function

- Cross-covariance function:

$$\Gamma_\alpha(s, s') = \begin{pmatrix} Cov(\alpha_0(s), \alpha_0(s')) & Cov(\alpha_0(s), \alpha_1(s')) \\ Cov(\alpha_1(s), \alpha_0(s')) & Cov(\alpha_1(s), \alpha_1(s')) \end{pmatrix}$$

- Dispersion matrix from cross-covariance function:

$$\Sigma_\alpha = [\Gamma_\alpha(s_i, s_j)]_{i,j=1}^N$$

- A "valid" $\Gamma_\alpha(\cdot) \Rightarrow \Sigma_\alpha$ is symmetric and positive definite.
- $\Gamma_\alpha(s, s')$ for $s \neq s'$ need not be positive definite or even symmetric. But:
  - $\Gamma_\alpha(s, s') = \Gamma_\alpha^T(s', s)$ for symmetry of Σ_α.
  - $\lim_{s \to s'} \Gamma_\alpha(s, s')$ is symmetric and positive definite.

Coregionalization

- Univariate: Bochner's Theorem characterizes covariance functions. Multivariate: Characterizing cross-covariance matrices is more difficult.
- Constructive approach: $\boldsymbol{\alpha}(s) = \boldsymbol{\mu}(s) + A\mathbf{v}(s)$.

$$\mathbf{v}(s) \sim GP(\mathbf{0}, \Gamma_{\mathbf{v}}(\cdot)) \Rightarrow \boldsymbol{\alpha}(s) \sim GP(\boldsymbol{\mu}(s), \Gamma_{\alpha}(\cdot))$$

  with $\Gamma_\alpha(s - s') = A\Gamma_{\mathbf{v}}(s - s')A^T$.
- Simple structure for $\mathbf{v}(s)$:

$$\Gamma_{\mathbf{v}}(s - s') = \begin{pmatrix} \rho_0(s - s') & 0 \\ 0 & \rho_1(s - s') \end{pmatrix}$$

- $\Gamma_\alpha(0) = AA^T$, so A identifies with $\Gamma_\alpha(0)^{1/2}$ (e.g. lower-triangular).

Examples

- Model 2a: $\alpha_0(s)$ and $\alpha_1(s)$ are independent processes with possibly different correlation structures.

$$\text{Model 2a: } \Gamma_\alpha(s-s') = \begin{pmatrix} \sigma_0^2\rho_0(s-s') & 0 \\ 0 & \sigma_1^2\rho_1(s-s') \end{pmatrix}.$$

- Model 2b: associates $\alpha_0(s)$ and $\alpha_1(s)$ through an "intrinsic" specification so that

$$\text{Model 2b: } \Gamma_\alpha(s-s') = \Lambda\rho(s-s'),$$

where $\Lambda = \Gamma_\alpha(0)$.

- Model 2c: coregionalized model

$$\text{Model 2c: } \Gamma_\alpha(s-s') = A\Gamma_{\mathbf{v}}(s-s')A^T;\ A = \Gamma_\alpha(0)^{1/2}.$$

Multi-resolution models

- Nested micro-level spatial associations within a global macro field. Spatial referencing, is an ordered pair (r, s), treated as r residing within s
- We now write the model as:

$$Y_t(r,s) = \mathbf{x}_t^T(r,s)\boldsymbol{\beta} + f[(r,s),t] + \epsilon_t(r,s);\ \epsilon_t(r,s) \sim N(0,\tau^2).$$

- A fully general nested spatial hypothesis:
  - correlation between the intercept and slope for each location (r, s),
  - coefficients arising from the subplots within each main plot (or array) s,
  - the association for coefficients arising in different main plots.

contd.

- First, for each main plot s we construct a local bivariate, zero-centered, stationary process on the subplots, $\mathbf{u}_s(r) = (u_{0s}(r), u_{1s}(r))$:

$$\Gamma_{\mathbf{u}_s}(r - r') = \begin{pmatrix} \rho_{0s}^{mic}(r - r') & 0 \\ 0 & \rho_{1s}^{mic}(r - r') \end{pmatrix}$$

- Define the micro-level covariance structure on $\boldsymbol{\alpha}(r, s)$

$$\boldsymbol{\alpha}(r, s) = \boldsymbol{\mu}(r, s) + B_s(r)\mathbf{u}_s(r).$$

  This leads to: $\Gamma_{\alpha(s)}(r, r') = B_s(r)\Gamma_{\mathbf{u}_s}(r - r')B_s(r)^T$.

- Similar to a coregionalized single-resolution model over the subplots in s, where $B_s(r)$ is a space-varying linear transformation at the micro level.

contd.

- Main plot level equation:

$$\boldsymbol{\alpha}(s) = \boldsymbol{\mu}(s) + \mathcal{B}_s \mathbf{u}_s$$

- Further coregionalize: $\mathcal{B}_s \mathbf{u}_s = A(s)\mathbf{v}(s)$ with

$$\Gamma_{\mathbf{v}}[(r,s),(r',s')] = \begin{pmatrix} \rho_0^{mac}(s-s') & 0 \\ 0 & \rho_1^{mac}(s-s') \end{pmatrix}$$

- Identifiability resolved through:

$$\mathcal{B}_s \Gamma_{\mathbf{u}}(s,s') \mathcal{B}_{s'}^T = A(s)\Gamma_{\mathbf{v}}(s-s')A(s')^T$$

and $\Gamma_{\boldsymbol{\alpha}}(s,s) = A(s)A(s)^T = \mathcal{B}_s \Sigma_{\mathbf{u}_s} \mathcal{B}_s^T$.

Examples

- Assume $B_s(r) = B$: the covariance between the intercept and slope is constant across space.
- Model 3c: Fully coregionalized model

$$\text{Model 3c: } \Gamma_\alpha(s - s') = A(s)\Gamma_{\mathbf{v}}(s - s')A(s)^T,$$

where $A(s) = (I_{N_r} \otimes B)\Sigma_{\mathbf{u}_s}^{1/2}$.

- Model 3b: Same correlation functions for micro processes

$$\text{Model 3b: } \Gamma_\alpha(s - s') = A\Gamma_{\mathbf{v}}(s - s')A^T;\ A = [\Sigma^{mic}]^{1/2} \otimes B.$$

- Model 3a: Fully separable model:

$$\text{Model 3a: } \Gamma_\alpha(s - s') = \rho(s - s')\Sigma^{mic} \otimes BB^T.$$

Model fitting

- For each of the models, the data equation can be cast into the following first-stage mixed model framework:

$$\mathbf{Y} = X\boldsymbol{\beta} + \tilde{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon};\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \tau^2 I)$$

- MCMC: a Gibbs sampler with Metropolis steps - better to marginalize over $\boldsymbol{\alpha}$.
- Identifiability of the spatial decay parameters is generally weak: set priors for macro and micro parameters relative to the size of their domains.
- Prediction and Interpolation:

$$\int P(\boldsymbol{\alpha}(s_0)|\boldsymbol{\alpha}, \Omega,\ Data)P(\boldsymbol{\alpha}|\Omega\ Data)P(\Omega|\ Data)d\Omega d\boldsymbol{\alpha}$$

Model Comparisons

- Deviance Information Criteria (DIC):

$$D(\Omega) = -2 \log L(Data|\ \Omega)$$
$$\overline{D(\Omega)} = E_{\Omega|\mathbf{Y}}[D(\Omega)]$$
$$p_D = \overline{D(\Omega)} - D(\bar{\Omega});\ \bar{\Omega} = E_{\Omega\,|\,\mathbf{Y}}[\Omega]$$
$$\text{DIC} = \overline{D(\Omega)} + p_D.$$

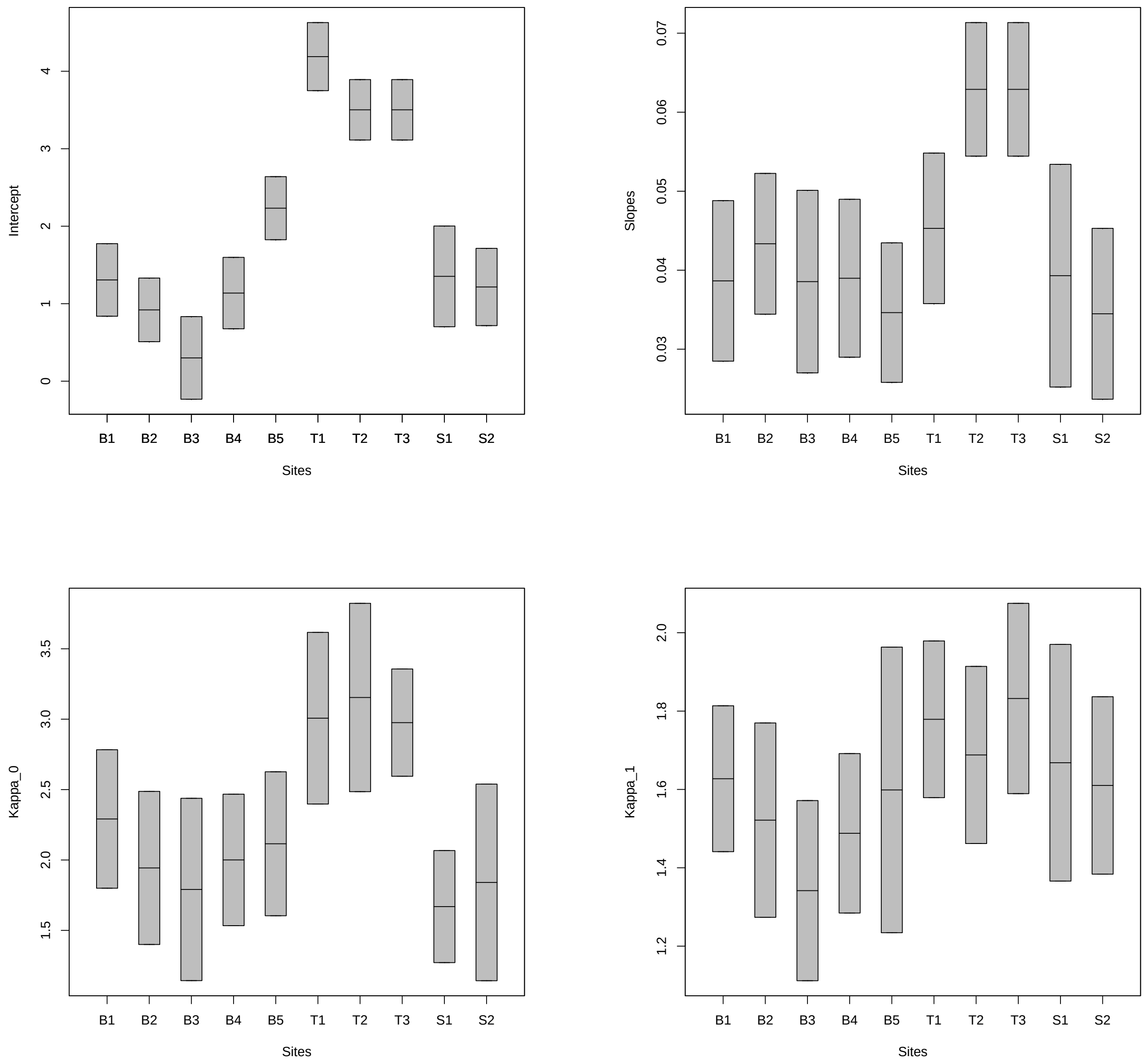
Model	pD	DIC	Model	pD	DIC	Model	pD	DIC
1a	7	979	2a	22.4	776	3a	23.2	738
1b	21.7	811	2b	23.1	771	3b	24.1	732
1c	22.3	802	2c	23.6	762	**3c**	**26.1**	**726**

Parameter estimates

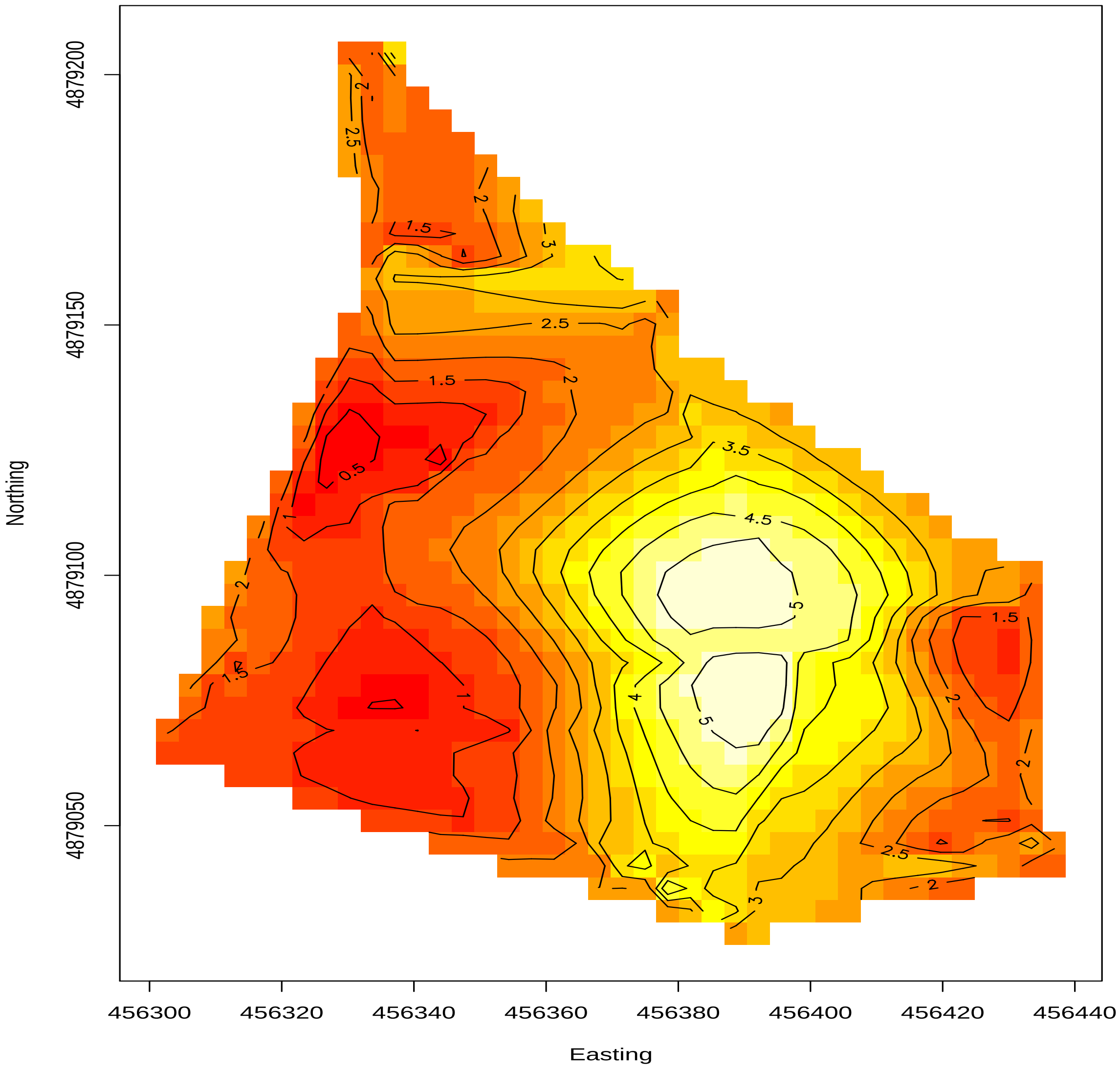
- Macro correlation function: Matern
- Micro correlation function: exponential

Parameters	Estimates: 50% (2.5%,97.5%)
Sand	-0.103 (-0.113,-0.084)
Phosphorus	-0.028 (-0.046,-0.010)
Potassium	-0.002 (-0.004,-0.001)
PDI	0.167 (0.139, 0.192)
Λ_{00}	0.380 (0.284, 0.521)
Λ_{11}	0.190 (0.141, 0.255)
$\Lambda_{01}/\sqrt{\Lambda_{00}\Lambda_{11}}$	-0.565 (-0.691,-0.390)
ϕ_0 (decay)	0.799 (0.216, 3.875)
ϕ_1 (decay)	0.211 (0.083, 1.179)
ν_0 (smoothness)	1.073 (0.665, 1.594)
ν_1 (smoothness)	0.687 (0.386, 1.197)
τ^2	0.002 (0.001, 0.004)

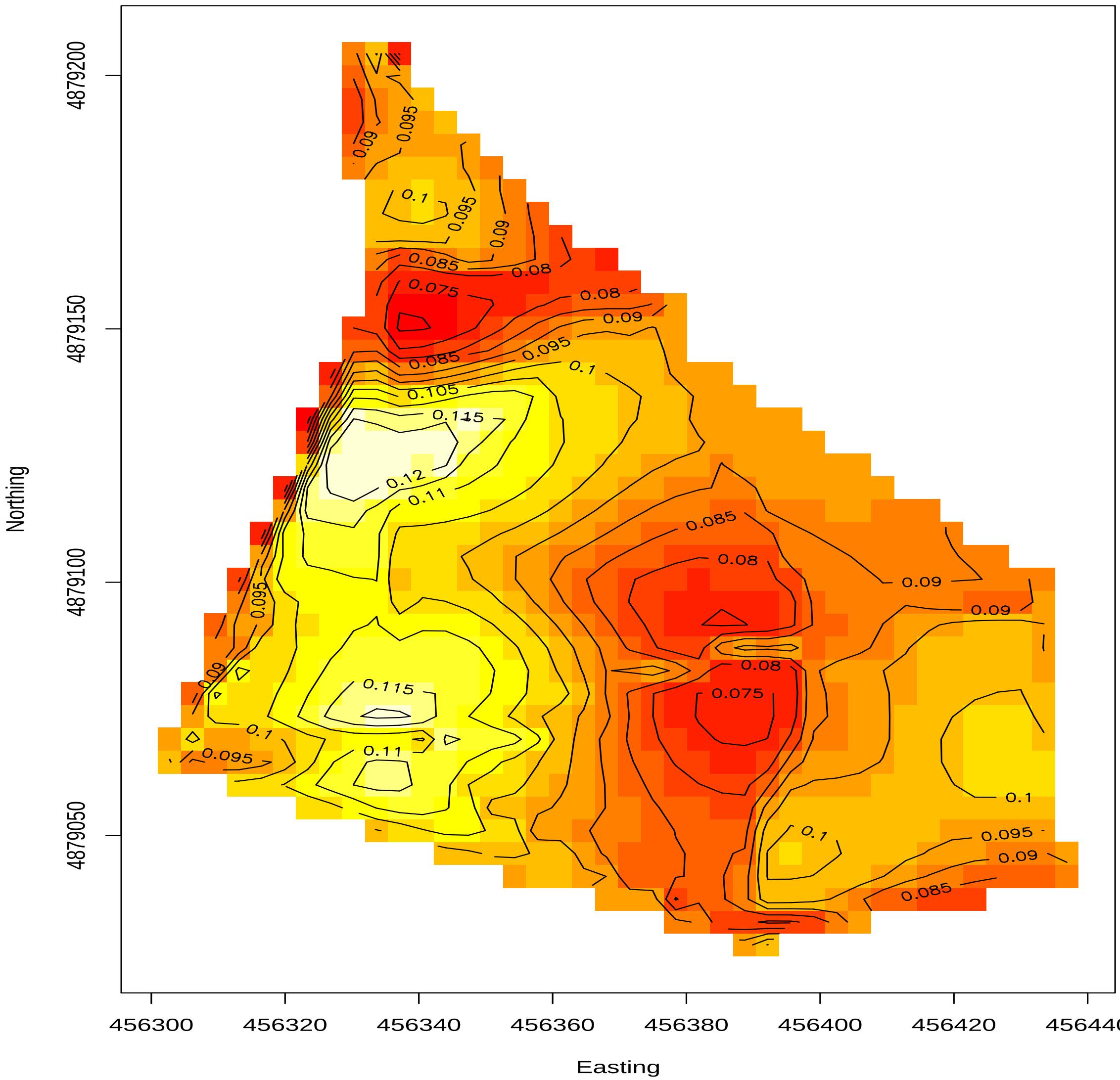
Sitewise Parameters: Posteriors



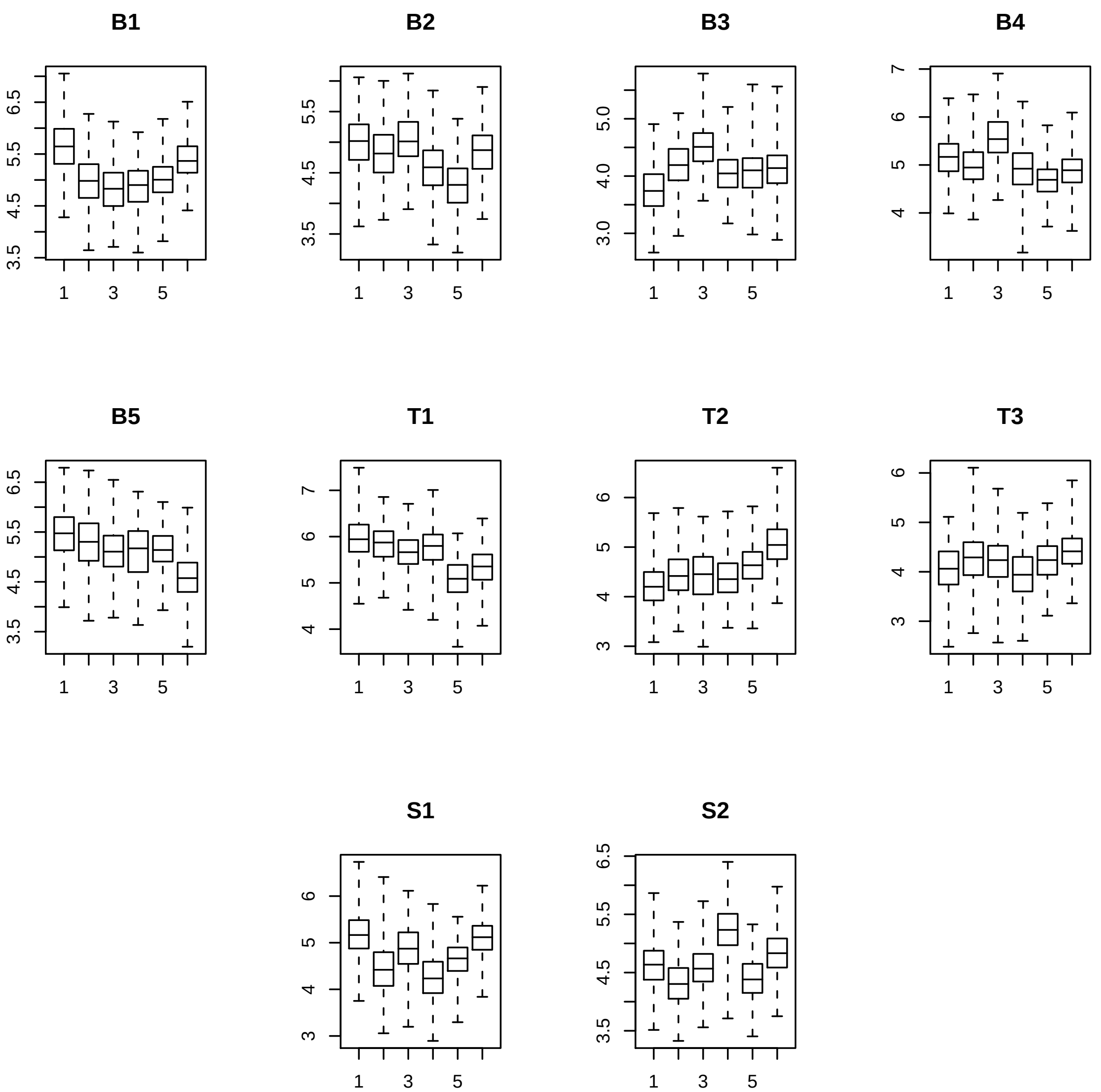
Interpolated intercept plot



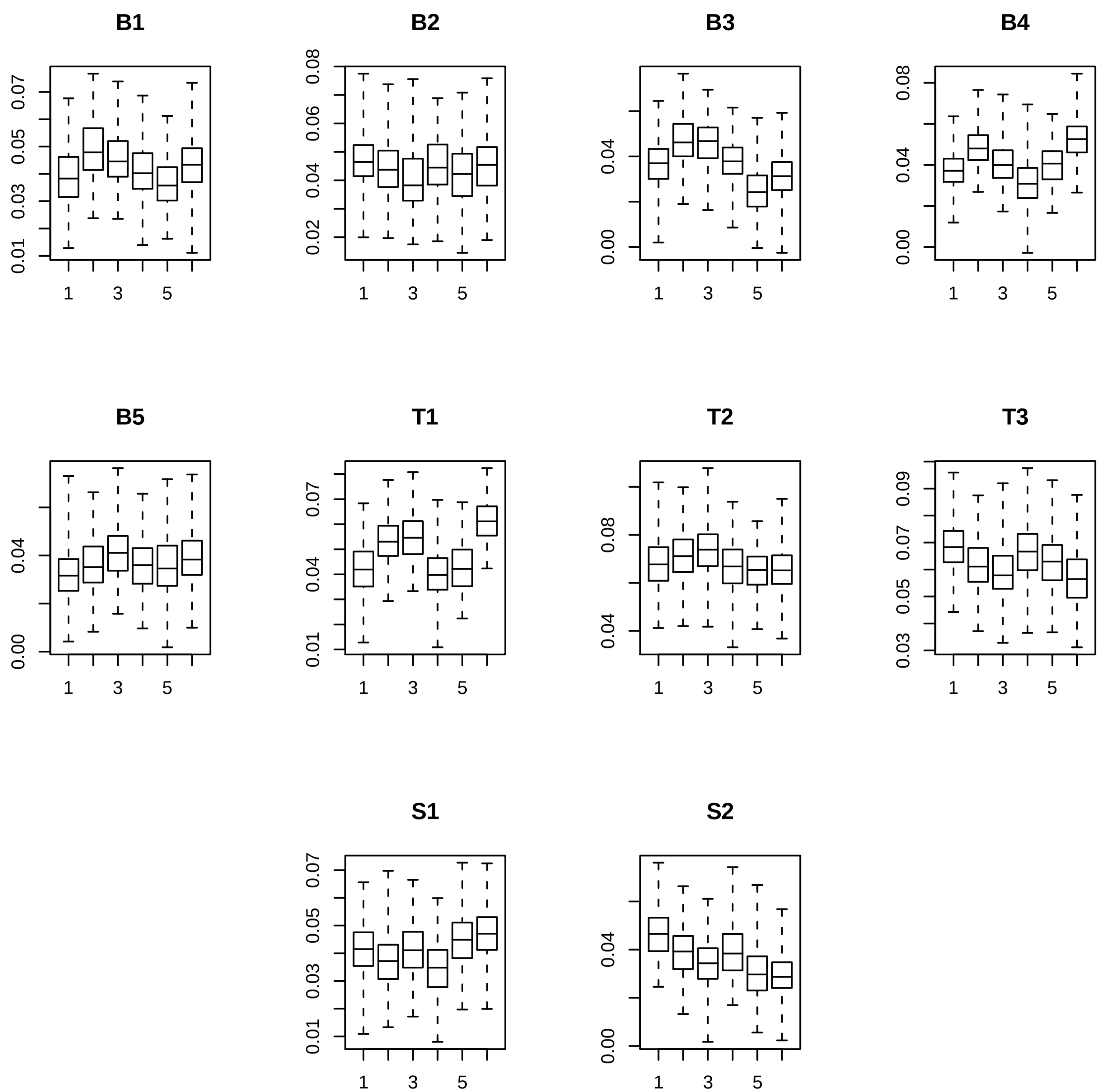
Interpolated slope plot



Sitewise variation: intercept process



Sitewise variation: slope process



Concluding remarks

- Modelled growth patterns using spatial processes for coefficients
- A space-varying linear transformation approach leads to extremely flexible classes of models
- Can be cast into a computationally feasible template, as long as number of plots is not too large
- Alternative approach: Model growth as a spatio-temporal process
- Investigate process gradients – spatial and temporal
- Thank You