Predicting Missing Forest Inventory Data using the Stochastic EM-Algorithm and Spatially Varying Coefficients

Johannes Dreesman

Johannes.Dreesman@nlga.niedersachsen.de

Niedersächsisches Landesgesundheitsamt

Roesebeckstr. 4-6, 30449 Hannover

Köln, December 3rd, 2004

The Data

The forest inventory

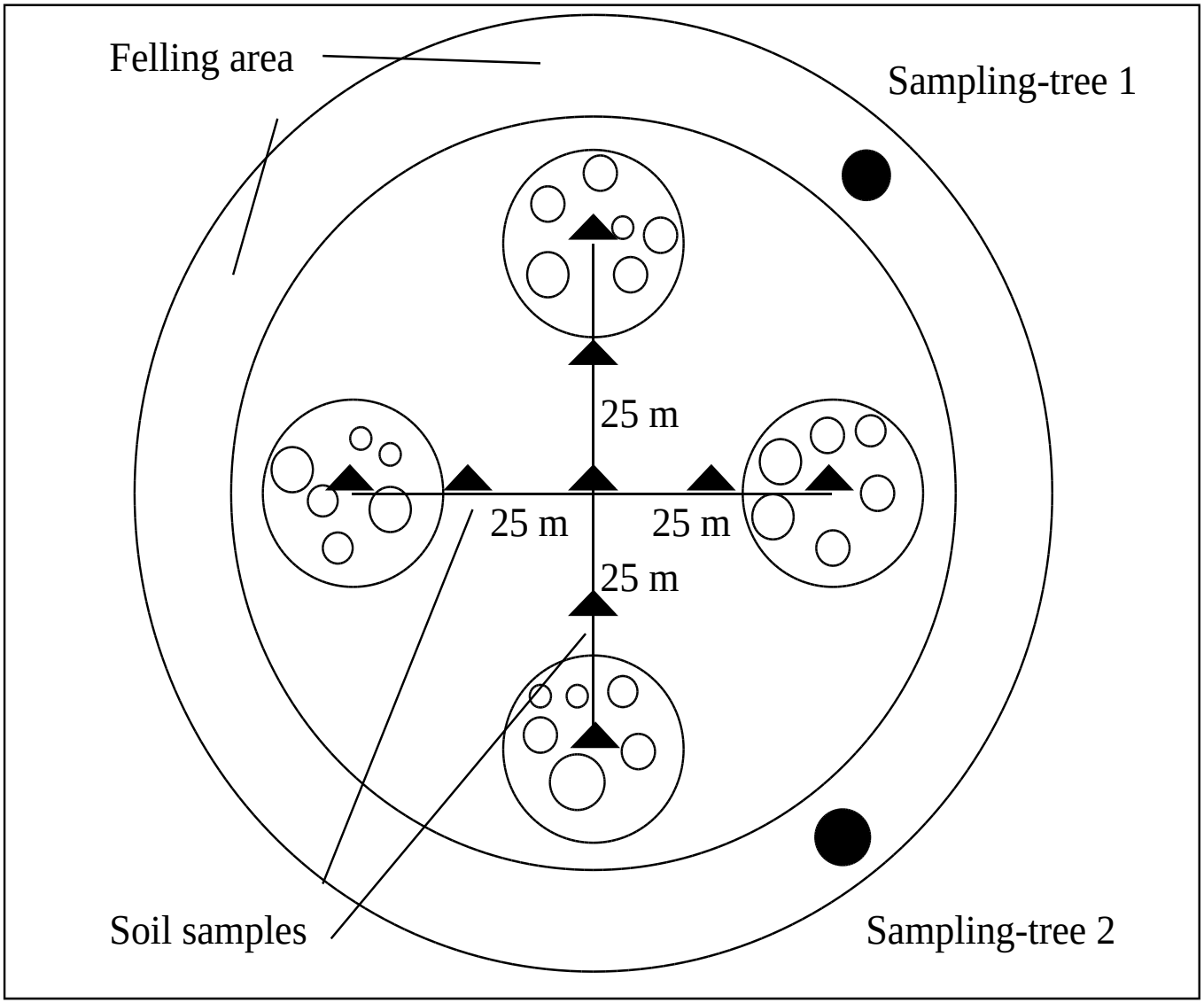
- Periodical inventory in the German state Baden-Württemberg.
- Goal is the detection and interpretation of the distribution patterns of damage to forests.
- Integral part is the analysis of needles, which gives information about the nutrition and pollution state.

The author is grateful to the Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg for providing the data.

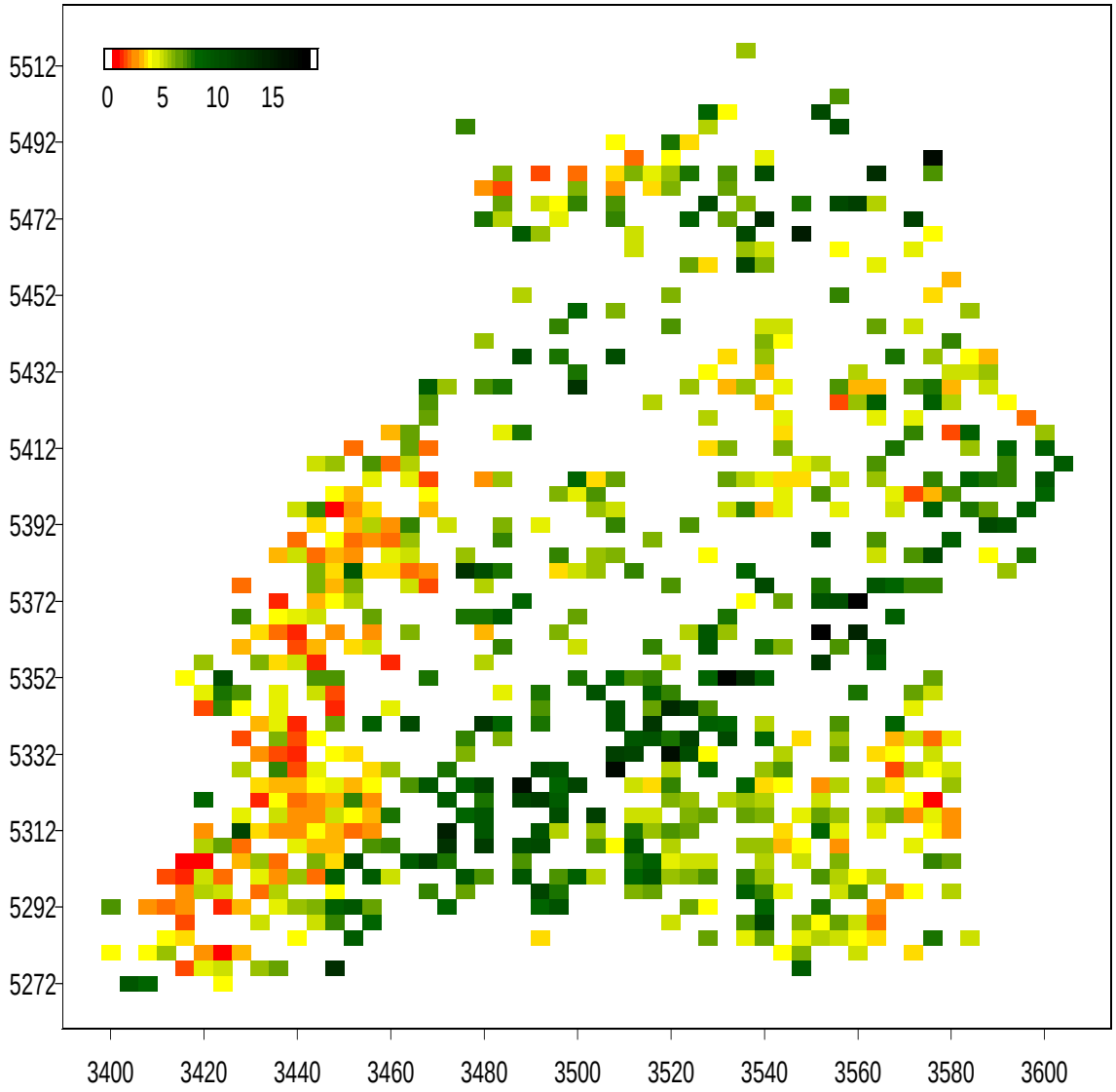
Design of inventory

- The data were obtained by sampling from the spruce- and pine-trees of Baden-Württemberg.
- The sampling points were chosen on a 4×4 km grid on the Gauß-Krüger-coordinate-system.
- If the target-point didn't fulfill the required criteria, an appropriate point was chosen in an objective manner within 300 meters.
- Needle-samples were drawn from two trees:
  - The trees belong to the same species (two spruces or two pines).
  - One tree should belong to the relativ vital trees (+ tree) and the other one to the less vital trees (− tree).

Satellite sample



Spatial distribution of observed values



The problem

At many lattice-points no values were available, because the sampling requirements were not fulfilled.

The task

At these points it is desired to get a prediction for the nutrient supply, preferably as an interval.

Crucial part is the consideration of the spatial dependence-structure.

Hierarchical Markov random field model

(Besag, Green, Higdon and Mengerson: Bayesian Computation and Stochastic Systems. Statistical Science, 1995)

Definitions

1. Lattice:

$$L = \{(i,j)\} = L_{obs} \cup L_{miss},$$

2. Observed mean nutrient concentration (of + and − tree):

$$\mathbf{Y}_{L_{obs}} = \{Y_{(i,j)}\}_{(i,j)\in L_{obs}},$$

3. Unobservable true nutrient supply:

$$\mathbf{X}_{L} = \{X_{(i,j)}\}_{(i,j)\in L}.$$

Model hierarchy

1st level: $Y_{(i,j)}\big|x_{(i,j)} \overset{ind}{\sim} N(x_{(i,j)}, \sigma^2) \qquad \forall (i,j) \in L_{obs},$

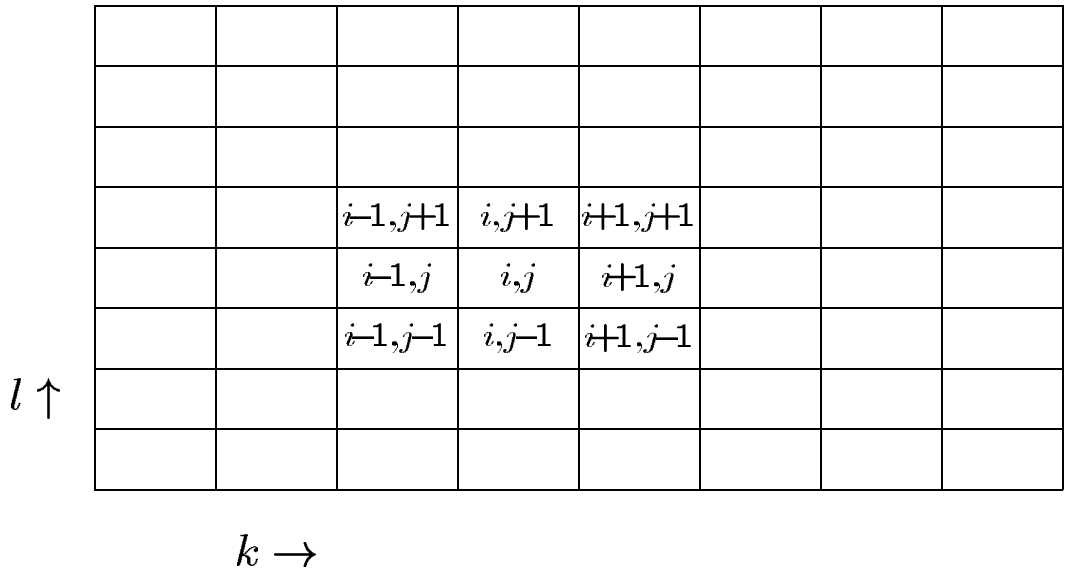
2nd level: $X_{(i,j)}\big|\mathbf{x}_{-(i,j)} \sim N(\eta_{(i,j)}, \tau^2) \qquad \forall (i,j) \in L,$

where $\eta_{(i,j)} = \beta_0 + \beta_1 \cdot (x_{i-1,j} + x_{i+1,j}) + \beta_2 \cdot (x_{i,j-1} + x_{i,j+1}) +$

$$\beta_3 \cdot (x_{i-1,j-1} + x_{i+1,j+1}) + \beta_4 \cdot (x_{i-1,j+1} + x_{i+1,j-1})$$

$$= \mathbf{Z}_{(i,j)}(\mathbf{x}) \cdot \boldsymbol{\beta}.$$

X = CAR (Conditional Autoregressive Model) $\in$ Markov random fields.



Spatially varying coefficients

In order to treat the nonstationarity it is additionally allowed that the coefficients of the model vary smoothly across the lattice.

$$\Rightarrow \text{Replace } \beta \text{ with } \{\beta(i,j)\}_{(i,j)\in L},$$
$$\tau^2 \text{ with } \{\tau^2(i,j)\}_{(i,j)\in L},$$
$$\text{and } \sigma^2 \text{ with } \{\sigma^2(i,j)\}_{(i,j)\in L_{obs}}.$$

(Dreesman and Tutz: Nonstationary conditional models for spatial data based on varying coefficients. The Statistician, 2001)

The task (revisited)

Determine the posterior distribution of $\mathbf{X}$ and obtain credible intervals from this distribution.

Estimating the hyperparameters

Estimation problem

$\{\beta(i,j)\}_{(i,j)\in L}$, $\{\tau^2(i,j)\}_{(i,j)\in L}$ and $\{\sigma^2(i,j)\}_{(i,j)\in L_{obs}}$.

Estimation of $\{\sigma^2(i,j)\}$

1. At each $(i,j) \in L_{obs}$ from the two observations the unbiased variance estimate for $Y_{(i,j)}$ is calculated.

2. The estimates are smoothed in order to reduce the error.

Estimation of $\{\beta(i,j)\}$ and $\{\tau^2(i,j)\}$ using the stochastic EM-algorithm

1. Generate starting values $\mathbf{X}_L^{(0)}$, $\{\widehat{\beta}^{(0)}(i,j)\}$ and $\{\widehat{\tau^2}^{(0)}(i,j)\}$.

2. Update iteratively.

In iteration k:

S-step: Draw $\mathbf{x}_L^{(k)}$ from the posterior distribution

$$f\left(\mathbf{x}_L \middle| \mathbf{Y}_{L_{obs}}, \{\widehat{\beta}^{(k-1)}(i,j)\}, \{\widehat{\tau^2}^{(k-1)}(i,j)\}\right).$$

M-step: Calculate $\{\widehat{\beta}^{(k)}(i,j)\}$ and $\{\widehat{\tau^2}^{(k)}(i,j)\}$ from $\mathbf{x}_L$.

3. Calculate ergodic means:

$\widehat{\beta}_{SEM}(i,j)$ and $\widehat{\tau^2}_{SEM}(i,j)$ from $\widehat{\beta}^{(k)}(i,j)$ and $\widehat{\tau^2}^{(k)}(i,j)$, $k = 1, \ldots, K$.

Consider a burn-in-period.

Properties

1. The E-step of the classical EM-algorithm, which would require intractable integration, is replaced with simulation

2. MCMC-algorithm

3. In the case of ML-estimation in the exponential family, the SEM-estimate converges to the ML-estimate

(Diebolt and Ip: Stochastic EM: Method and Application. In: Markov Chain Monte Carlo in Practice, Chapman & Hall, 1996)

A closer look at the S-step

The new realisation $\mathbf{x}_{L_{obs}}^{(k)}$ is drawn componentwise from $\mathbf{x}_{L_{obs}}^{(k-1)}$.

At points from $\mathbf{L_{obs}}$: Bayes' theorem is needed to calculate the posterior distribution.

$$X_{(i,j)}\Big|\mathbf{x}_{-(i,j)} \sim N(\eta_{(i,j)}, \tau^2(i,j))$$

$$\text{with } \eta_{(i,j)} = \mathbf{Z}_{(i,j)} \cdot \boldsymbol{\beta}(i,j)$$

$$\text{and } Y_{(i,j)}\Big|x_{(i,j)} \sim N(x_{(i,j)}, \sigma^2(i,j))$$

$$\Rightarrow X_{(i,j)}\Big|\mathbf{x}_{-(i,j)},\ y_{(i,j)} \sim$$

$$N\left(\frac{1}{\tau^2(i,j)+\sigma^2(i,j)} \cdot (\tau^2(i,j) \cdot y_{(i,j)} + \sigma^2(i,j) \cdot \eta_{(i,j)}),\ \frac{\tau^2(i,j) \cdot \sigma^2(i,j)}{\tau^2(i,j)+\sigma^2(i,j)}\right).$$

At points from $\mathbf{L_{miss}}$: The prior distribution of $X_{(i,j)}$ at a point from L_{miss} in only based on the X-values in the neighbourhood:

$$X_{(i,j)} \Big| \mathbf{x}_{-(i,j)} \sim \; N(\eta_{(i,j)}, \tau^2(i,j))$$

$$\text{with } \eta_{(i,j)} = \mathbf{Z}_{(i,j)} \cdot \boldsymbol{\beta}(i,j).$$

Instead of the corresponding parameters the current estimates $\widehat{\boldsymbol{\beta}}^{(k-1)}(i,j)$ and $\widehat{\tau^2}^{(k-1)}(i,j)$ as well as $\widehat{\sigma^2}(i,j)$ are plugged in.

A closer look at the M-step

Estimation principle: Local Pseudolikelihood

(Weighted least squares)

Calculation of the posterior distribution using MCMC

Required posterior density of the Markov-random-field:

$$f\left(\mathbf{x}_L \middle| \mathbf{y}_{L_{obs}}\right).$$

Estimate:

$$f\left(\mathbf{x}_L \middle| \mathbf{y}_{L_{obs}}, \{\widehat{\boldsymbol{\beta}}_{SEM}\}, \{\widehat{\tau^2}_{SEM}\}, \{\widehat{\sigma^2}\}\right).$$

- Analytical calculation is hardly possible.

- MCMC-sampling is easy and doesn't require additional implementation subsequent to the stochastic EM-algorithm.

- Just carry out M further iterations of the S-step with $\{\widehat{\boldsymbol{\beta}}_{SEM}\}$ and $\{\widehat{\tau^2}_{SEM}\}$.

- The M-step including the update of $\{\widehat{\boldsymbol{\beta}}\}$ and $\{\widehat{\tau^2}\}$ is omitted.

Every iteration produces a new realisation $\mathbf{x}_L^{(m)}$, $m = 1, \ldots, M$, from the desired posterior distribution.

The ergodic mean $\bar{\mathbf{x}}_M = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}_L^{(m)}$ of this MCMC-Sample for instance is an estimate for the expectation $E\left(\mathbf{X}_L \middle| \mathbf{y}_{L_{obs}}, \{\widehat{\boldsymbol{\beta}}_{SEM}\}, \{\widehat{\tau^2}_{SEM}\}\right)$.

The credible intervals are obtained from the sample distribution.

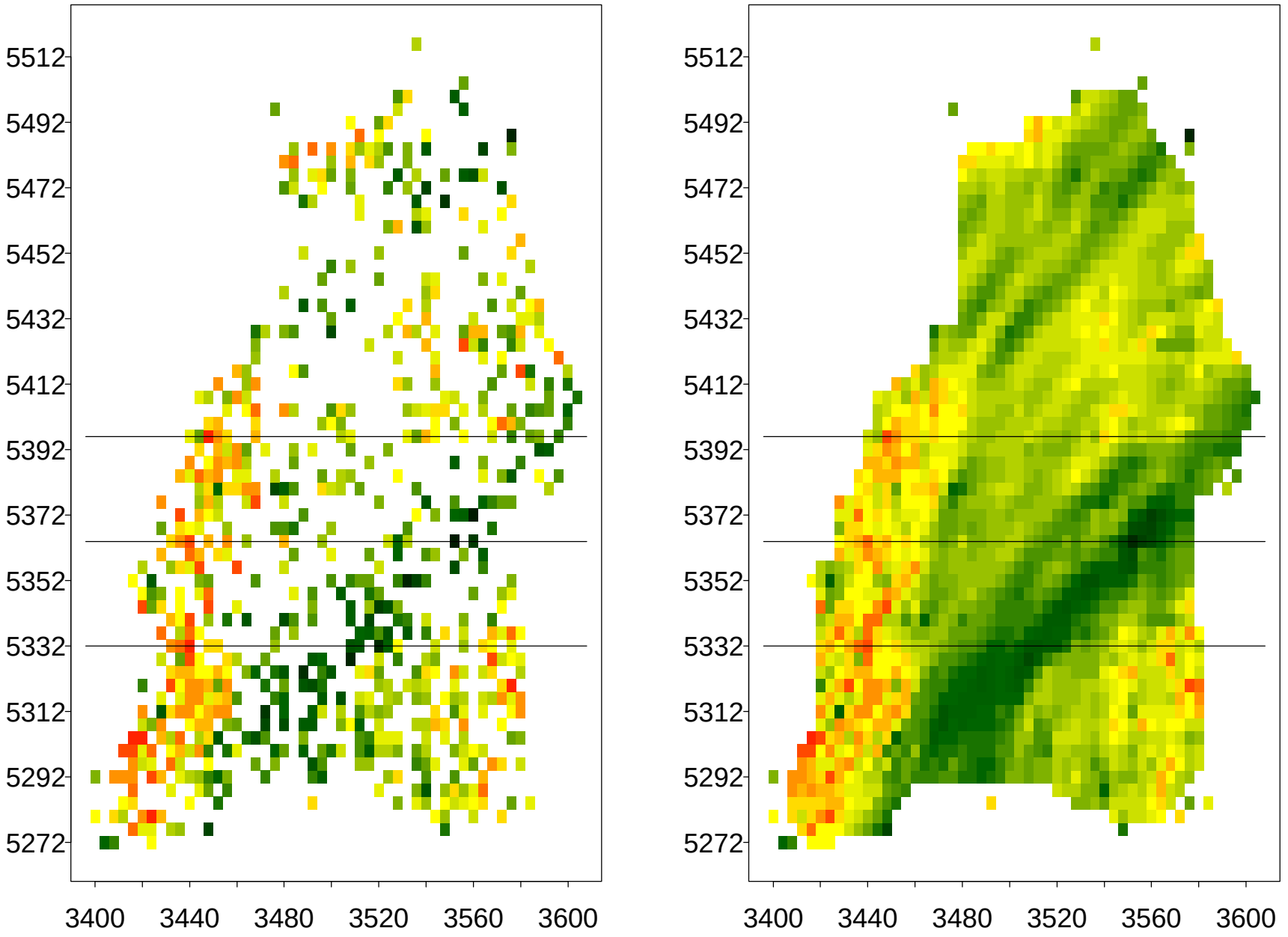
Results

- Properties of the algorithm.
  - Mixing properties
  - Speed of convergence of statistic
- Estimated posterior distribution:

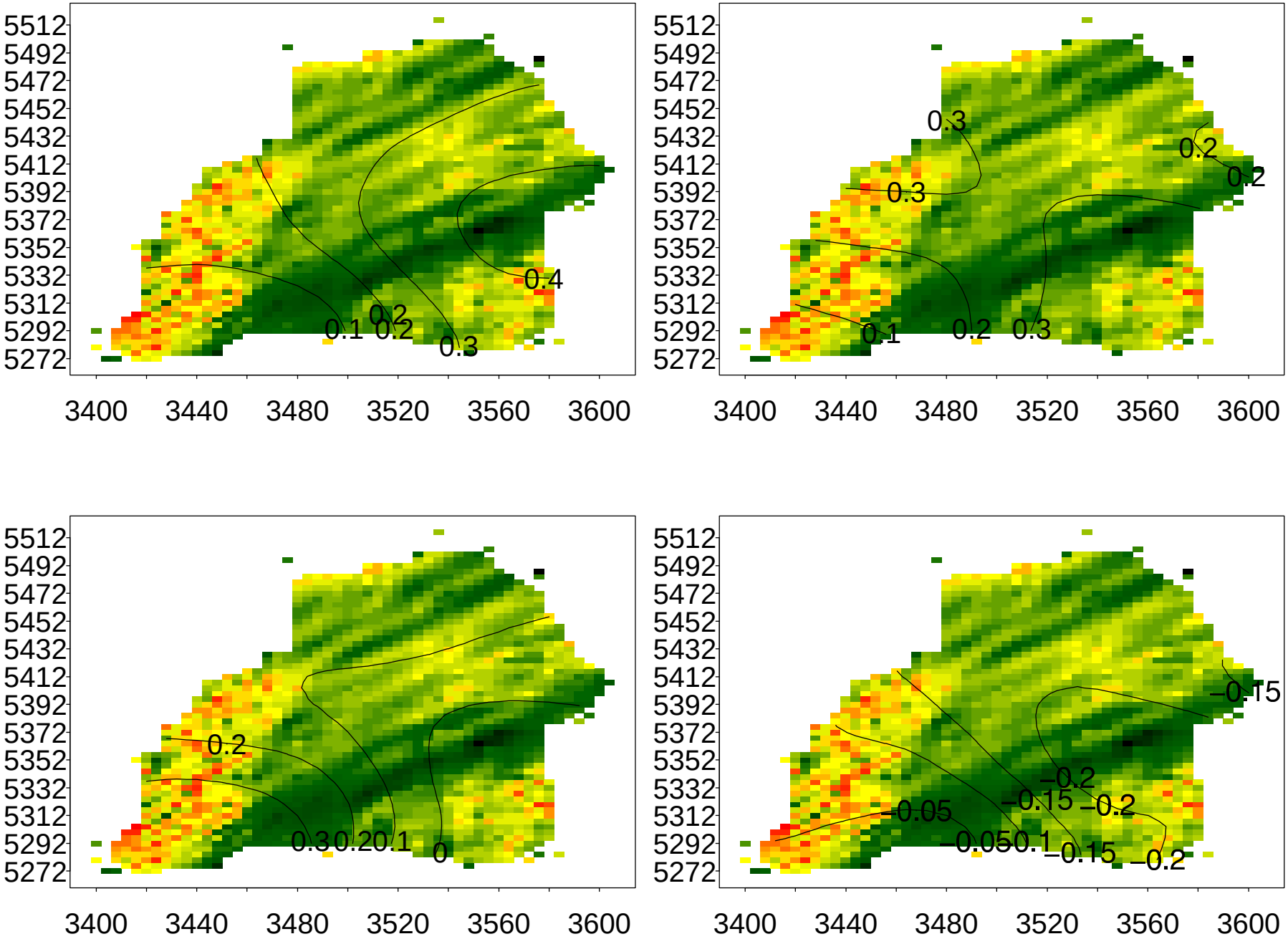
$$\widehat{f}\left(\mathbf{x}_L \middle| \mathbf{y}_{L_{obs}}, \{\widehat{\boldsymbol{\beta}}_{SEM}\}, \{\widehat{\tau^2}_{SEM}\}, \{\widehat{\sigma^2}\}\right)$$

  - Ergodic mean $\bar{\mathbf{x}}_M = \frac{1}{M} \sum_{m=1}^{M} \mathbf{x}^{(m)}$
  - Credible intervals

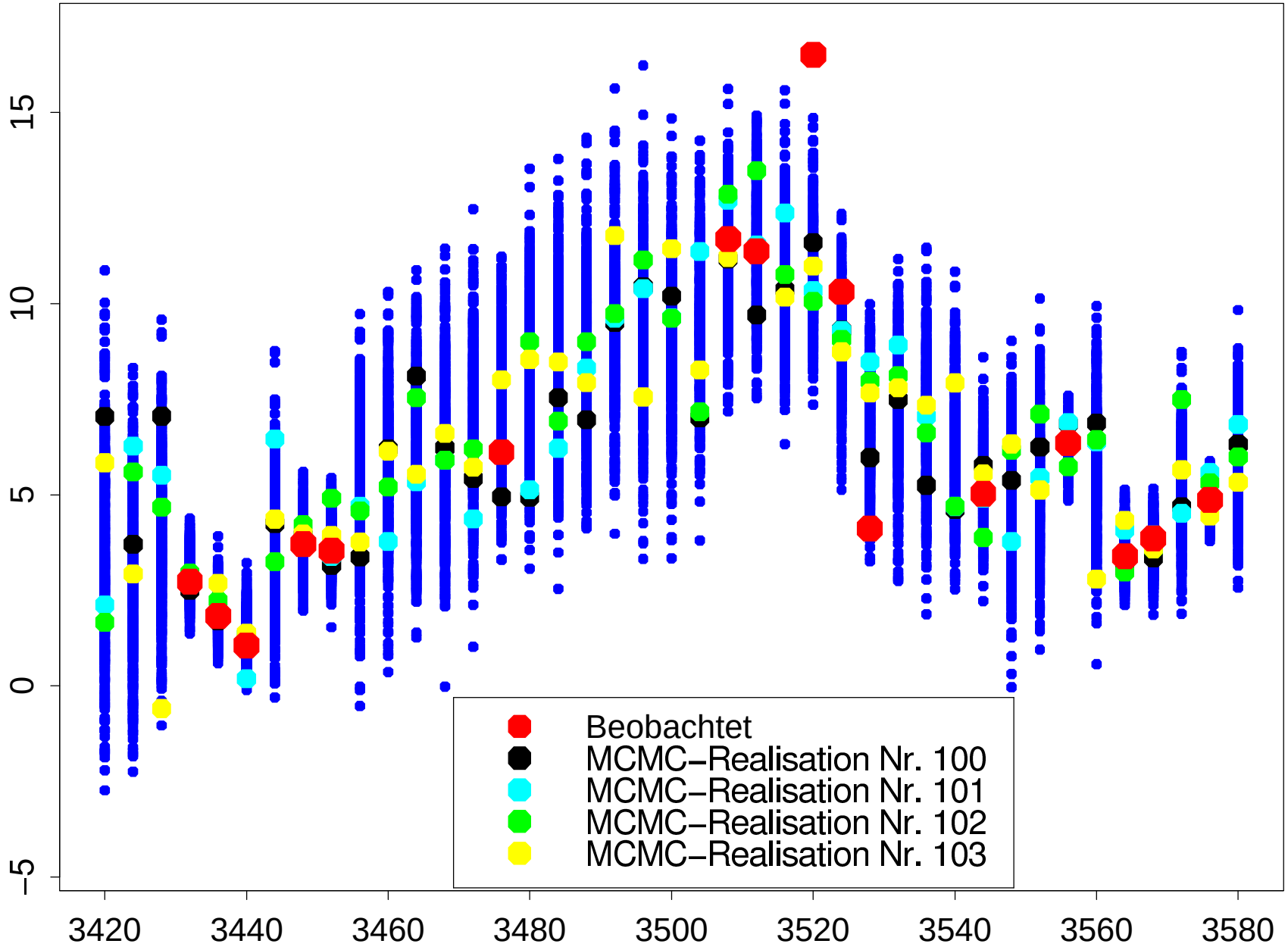
Observed values (left) and posterior mean (right)



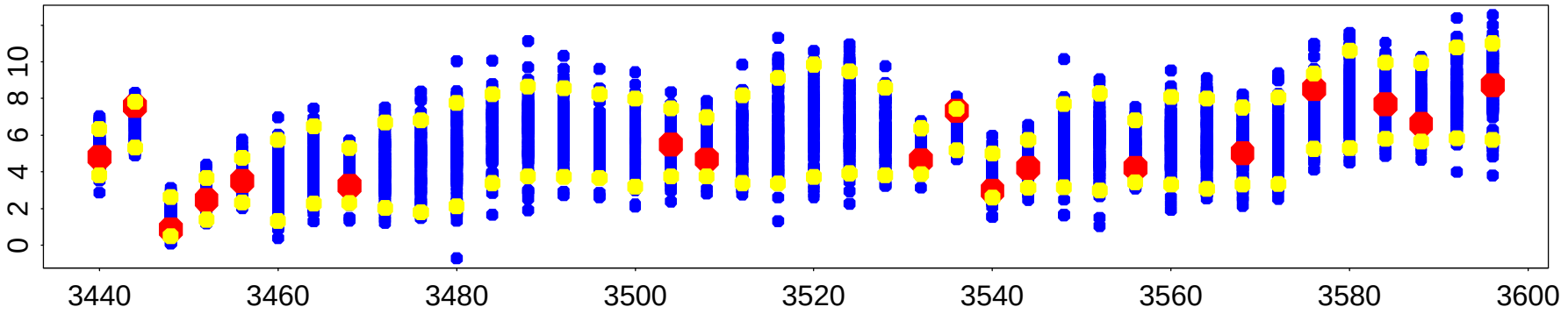
Spatially varying coefficients β_1 (top-left), β_2 (tr), β_3 (bl), and β_4 (br)

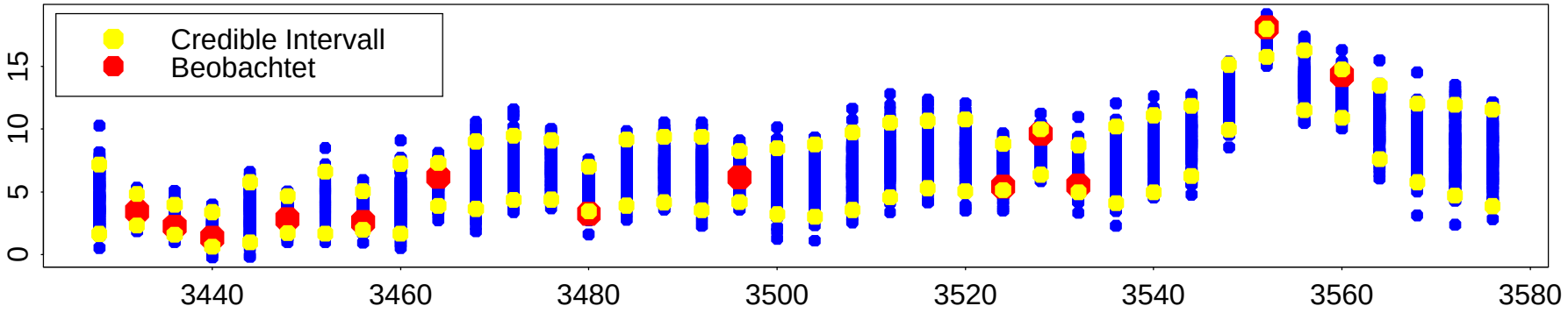


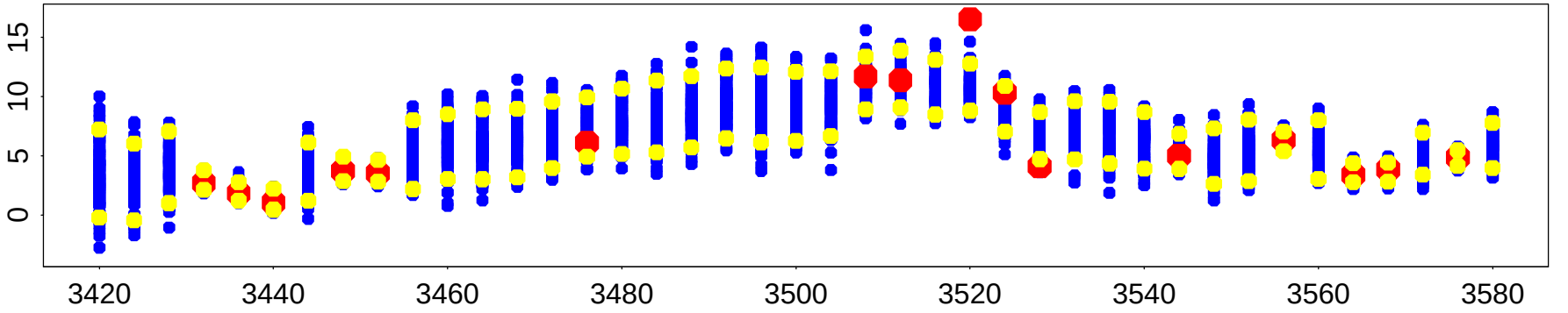
Simulated values g Ca/kg - cross-section at high-value 5332



Credible intervals at high-values 5396, 5364, and 5332 (top to bottom)







Conclusions

- Variation of coefficients plausible
- Good mixing properties
- Plausible credible intervals
- Improve representation of credible intervals
- Compare with other methods