

Master thesis - Final presentation

Comparison of statistical methods and designs for a high throughput phenotyping experiment



Alexandre Bohyn

Faculty of Science
Department of Statistics

September 2019

Research question

How good is the SpATS model at estimating spatial variability in plant yield in a phenotyping platform?

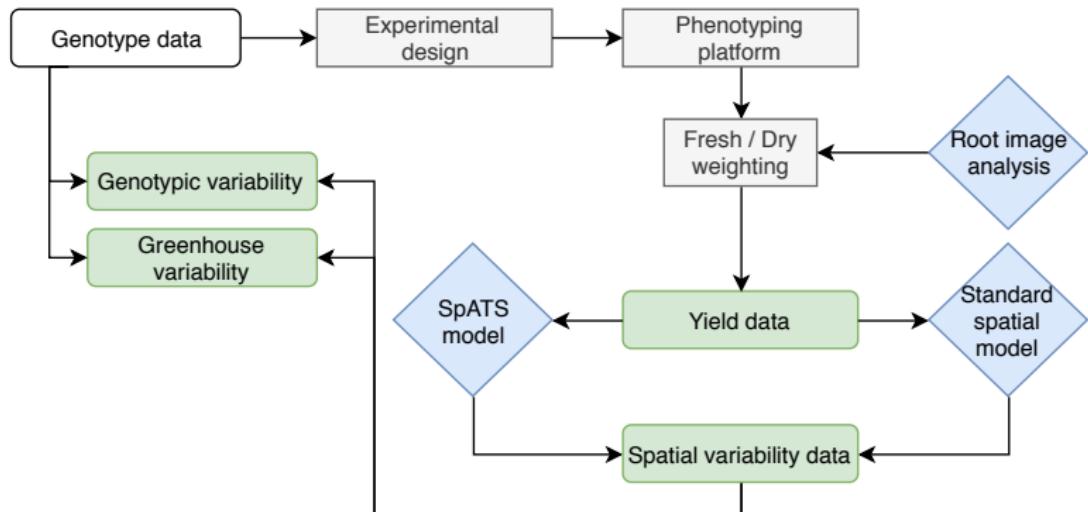
- Platform-specific variability
- How does the model compares to another one
- Genotype part of the yield variability

Objectives

4 main objectives, both regarding **statistics** and **agronomy**:

- Establish an adequate experimental design
- Compare the efficacy of two spatial models in the context of this platform
- Estimate the platform environmental variability
- Estimate the genotypic variability

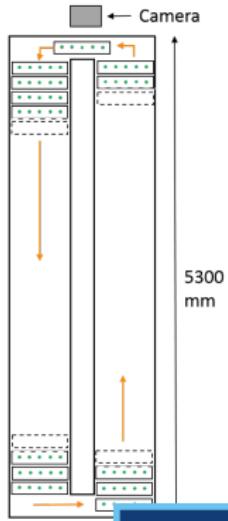
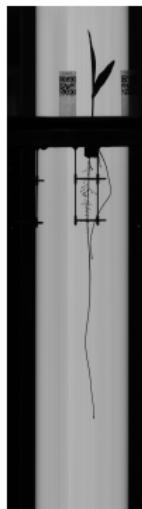
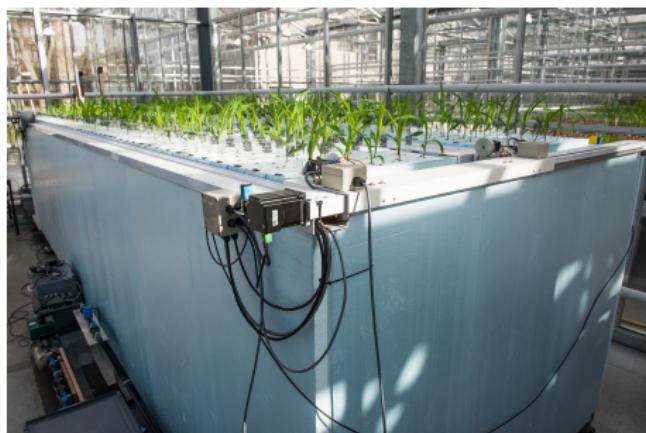
Methodology



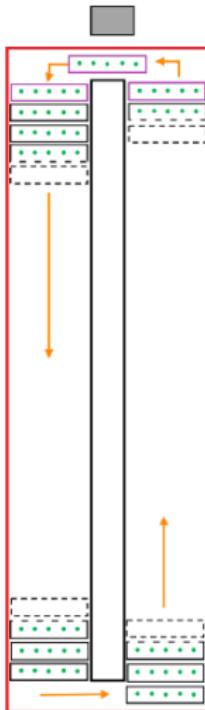
Phenotyping platform

Maize seeds are used - Germinated outside of the platform:

- 2 tanks of 495 plants
- Root scan every 2 hours
- 1 tank moving 24/7, 1 tank still



Design factors



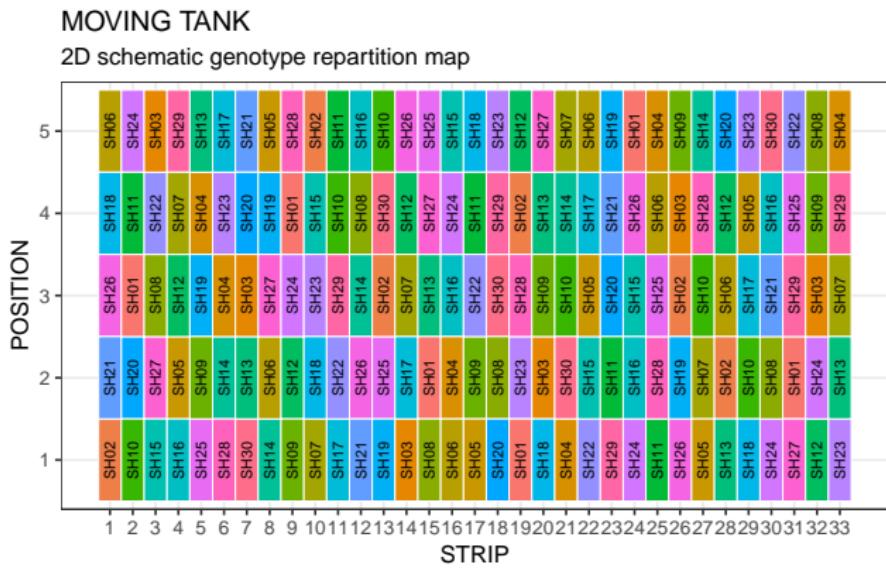
4 categorical factors:

- **Tank**: 2 levels (moving or still)
- **Strip**: 99 levels
- **Position** (on the strip): 5 levels
- Genotype: 30 levels

Hierarchical order in the categories → creates *blocking*:

- 2 wholes plots \approx 2 tanks
- 99 sub-plots \approx 99 strips

Design output



Spatial models

Spatial models can be re-written as **mixed models**:

$$\mathbf{y} = \underbrace{\mathbf{X}\boldsymbol{\beta}}_{\text{Fixed effects}} + \underbrace{\mathbf{Z}\mathbf{c}}_{\text{Random effects}} + \underbrace{\boldsymbol{\xi}}_{\text{Spatially dependant error}} + \underbrace{\boldsymbol{\varepsilon}}_{\text{Random error}}$$

→ Useful for variance estimation.

With 2 kinds of effect:

- **Fixed**: linked to environment (e.g. light distribution, slope)
- **Random**: cannot be linked to covariates (e.g. genotype effect, soil moisture)

Spatial variation

Spatial variation in \mathbf{y} can be decomposed in:

- Global trends (stationary): rain, temperature ...
- Local trends (non-stationary): patch of fertility, ...
- Extraneous variation: agricultural practices, ...

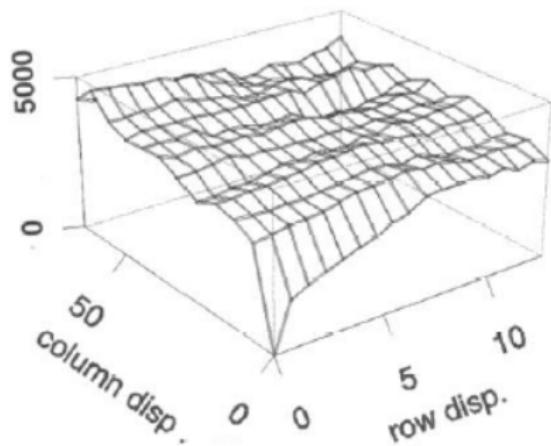
2 main approaches to model this spatial variation:

1. Spatial variance-covariance structures
2. Smoothing techniques

Standard spatial model (VCOV structure)

Random (\mathbf{c}) and fixed ($\boldsymbol{\beta}$) effects modelled using covariates
Spatial error $\mathbf{e} = \boldsymbol{\xi} + \boldsymbol{\varepsilon}$ as a 2D random process

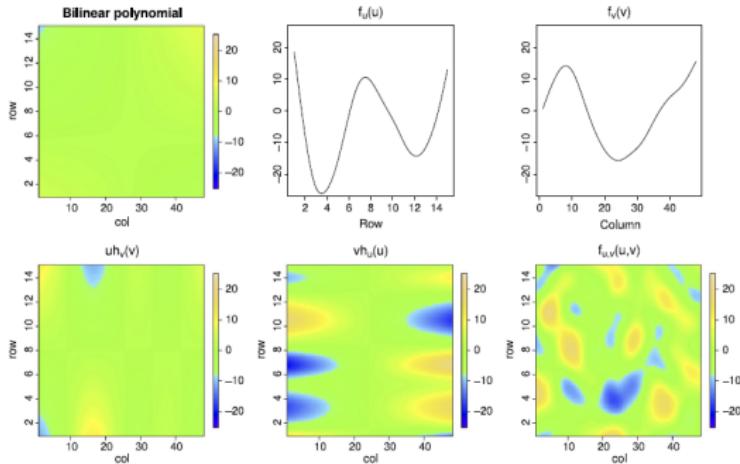
- Modelled using a semi-variogram
- Equivalent to a $AR(1) \times AR(1)$ model
 - only 2 parameters to estimate
- Elevated to a $LV \times LV$ model
 - robust to convergence issues



SpATS model (smoothing approach)

Use of a **smooth bivariate surface** to model global trends
AND local trends (no more spatial error term ξ)

→ Linear part, smooth part and interactions



Goal of the analysis

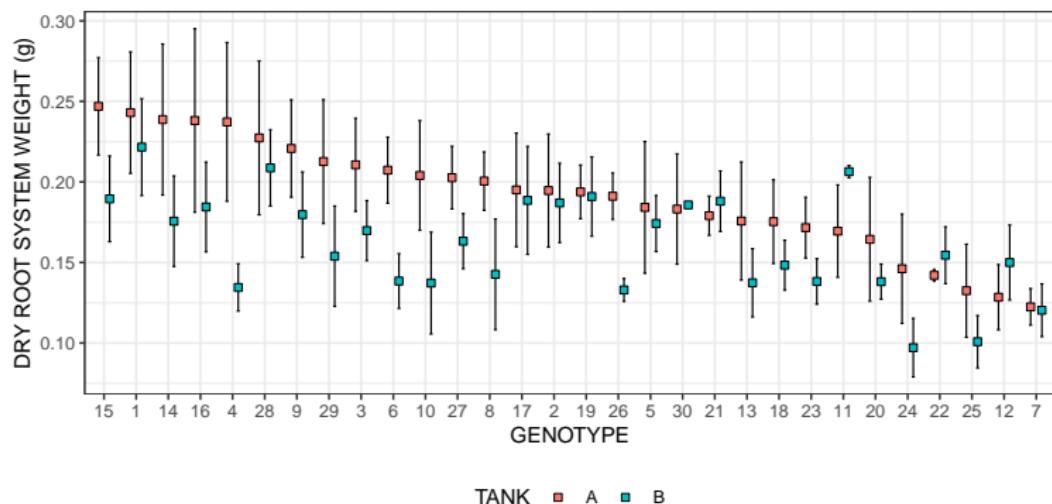
4 questions to answer:

1. Is there a significant difference between the moving and still tank ?
2. Do the models account for spatial trends ?
3. How different are the estimation of the two models ?
4. Is there a significant genotypic effect, picked up by the models ?

Using 4 variables:

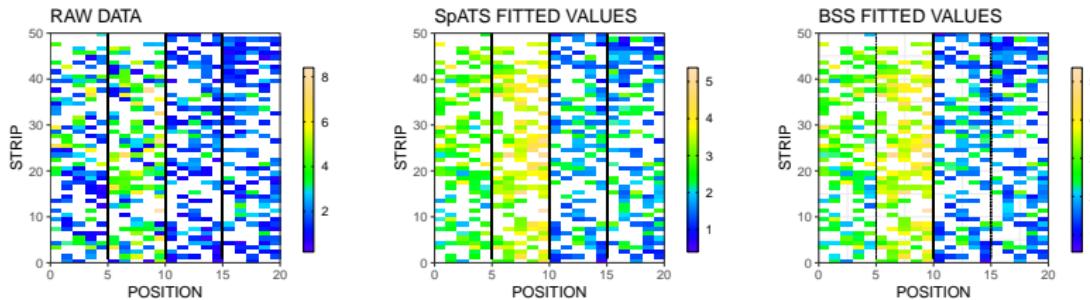
- **Fresh** weight for **root** and **leaf** system
- **Dry** weight for **root** and **leaf** system

1. Tank difference



T-tests revealed significant difference between tanks ($p < 0.05$)
→ For all variables
→ **NOT** for all genotypes

2. Model comparison



Model outputs show:

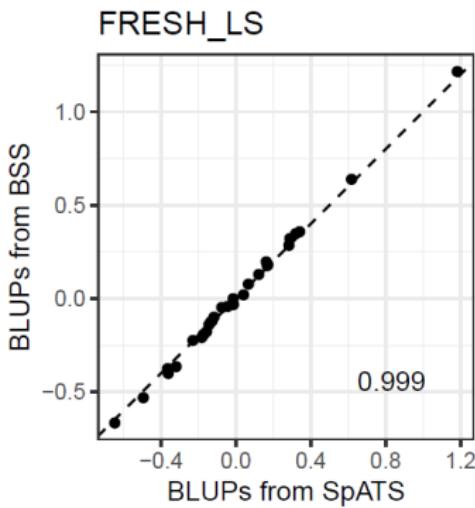
- Tank effect
- Highly similar values
- Clear spatial trends
 - mostly along the **STRIPS**
 - complex spatial patterns

⚠ More heterogeneity in
the moving tank
→ not picked up by the
model

3. Genotypic effect

- Models give similar values
- Large part of the variation attributed to the genotypes
→ Significant genotypic effect
- Ranking similar to the weights

⚠ Harder to interpret without context !



Conclusion

Main points:

- Significant tank effect
- Strong genotypes differences
- Similar output for both models
 - Better interpretability for SpATS
 - No iterative process
- Clear spatial trends
 - Better captured in the still tank

- ? Adapted to moving environment
- ? Results of a less-complex GLM
- ? Use of time dimension in the model