Kriging vs Geographically Weighted Regression for analysing farm experiments

Shih Ching Fu

GRDC

SAGI
Statistics for the Australian Grains Industry
National | North | West | South

Curtin University

Outline

- Precision agriculture
- Large-scale on-farm strip experiments
- Modelling spatial heterogeneity
- Comparison Results
- Recommendations
- Q & A

Precision Agriculture

- Response of crops to inputs is likely to vary spatially within a field.
- Targeted application of chemicals is more efficient and sustainable.

Aerial map of wheat farm in Kojonup. Source: Google Maps.

Precision Agriculture

- Response of crops to inputs is likely to vary spatially within a field.
- Targeted application of chemicals is more efficient and sustainable.

1: Where should inputs be applied for greatest effect?

Treatment effect estimation

2: What is the expected result?

Yield prediction

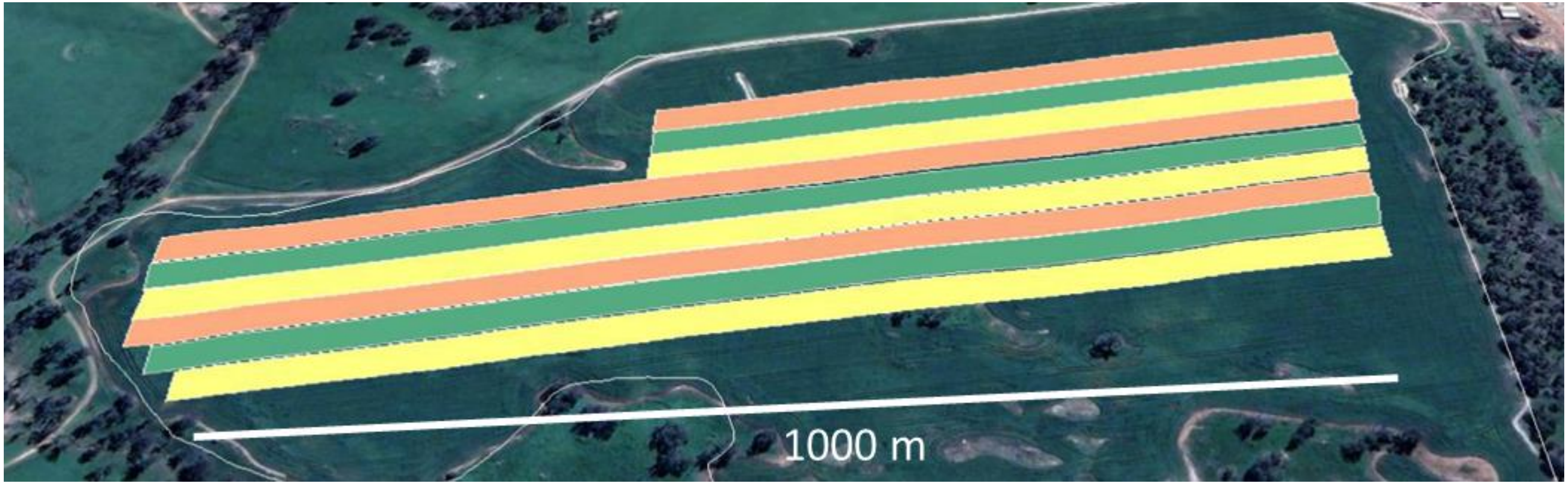
Aerial map of wheat farm in Kojonup. Source: Google Maps.

On-farm strip experiments

- Understand the effect of treatments β_1, e.g., fertiliser.

$$\text{yield} = \beta_0 + \beta_1 \times \text{treatmentRate} + \varepsilon$$

- At a scale that is meaningful and practical.



Kojonup Winter Wheat Trial. Source: SAGI-West, GRDC.

Spatial Heterogeneity

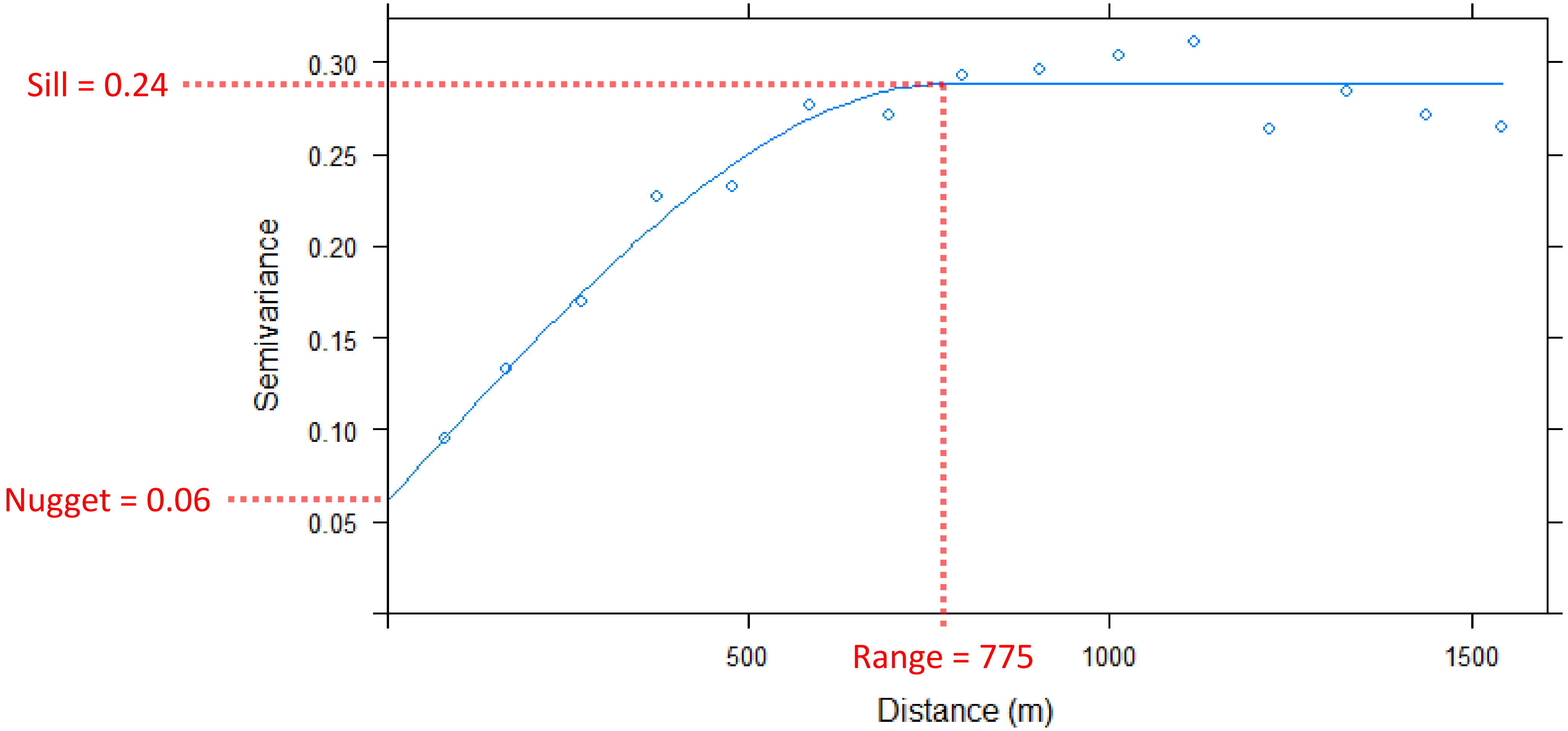
- Non-stationarity

  The response of crops, even without treatment, varies with location.

- Auto-correlation

  The response of crops in one location is related to the response at nearby locations.

Variogram


Sill = 0.24
Nugget = 0.06
Range = 775
0.30
0.25
0.20
0.15
0.10
0.05
Semivariance
500
1000
1500
Distance (m)

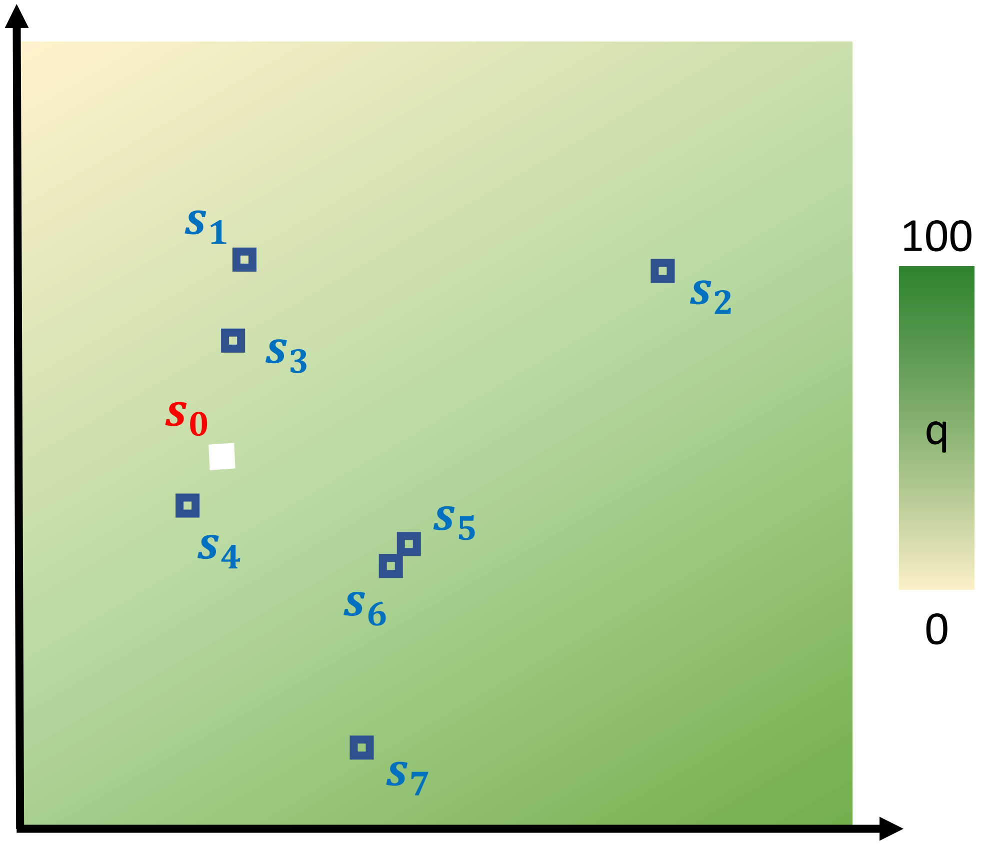
Regression Kriging

- Estimate the value at a target location by the weighted average of the known observations

$$\hat{z}(s_0) = \sum_{k=1}^{p} \hat{\beta}_k \cdot q_k(s_0) + \sum_{i=1}^{n} \lambda_i \cdot e(s_i)$$

$\hat{z}(s_0)$ is the interpolated value at target location s_0
$e(s_1), \cdots, e(s_n)$ are residuals at locations s_i
$q_1(s_0), \cdots, q_p(s_n)$ are explanatory variables at s_i
$\hat{\beta}_1, \cdots, \hat{\beta}_p$ are regression coefficients
$\lambda_1, \cdots, \lambda_n$ are kriging weights

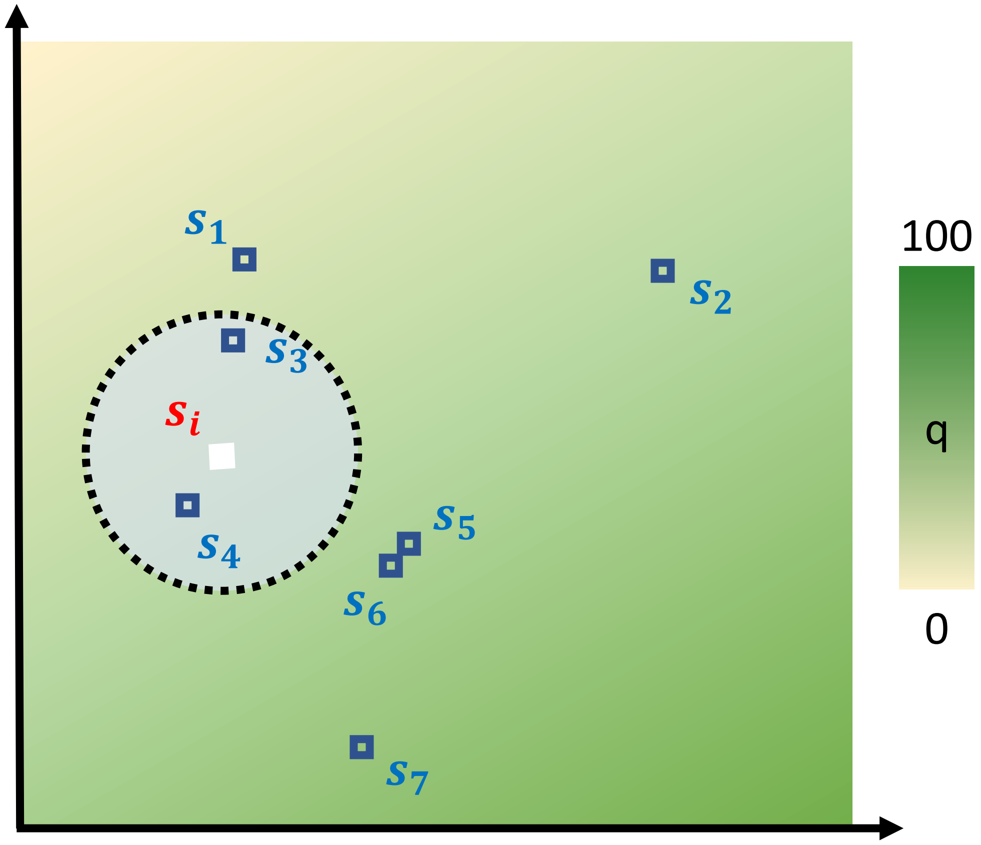
✘ $\hat{\beta}_k$ coefficients are <u>not</u> spatially varying



Geographically Weighted Regression (GWR)

- Fit a regression model using data from within a window of the target location.

$$\hat{z}(s_i) = \hat{\beta}_0(s_i) + \sum_{k=1}^{p} \hat{\beta}_k(s_i) \cdot q_k(s_i) + \varepsilon_i$$



$\hat{z}(s_i)$ is the fitted value at location s_i
$e_i \sim N(0, \sigma^2)$ are residuals at locations s_i
$q_1(s_0), \cdots, q_p(s_n)$ are explanatory variables at s_i
$\hat{\beta}_1(s_i), \cdots, \hat{\beta}_p(s_i)$ are regression coefficients

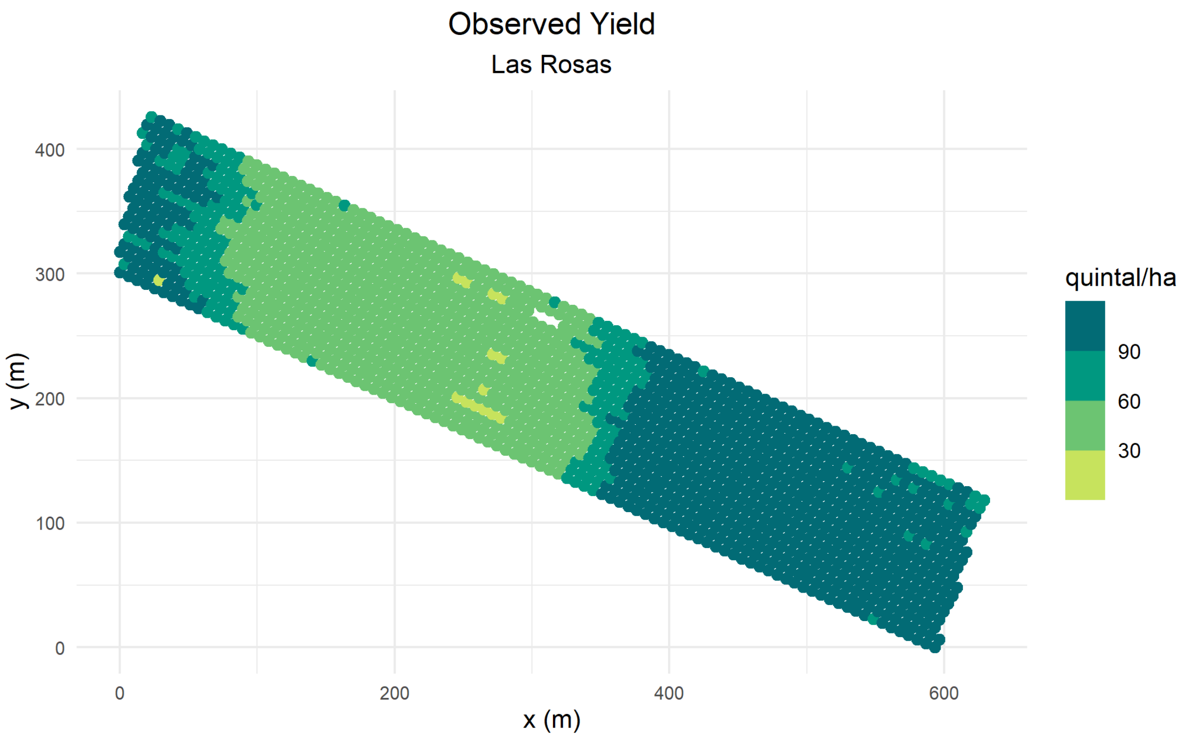
✓ $\hat{\beta}_k$ coefficients are spatially varying

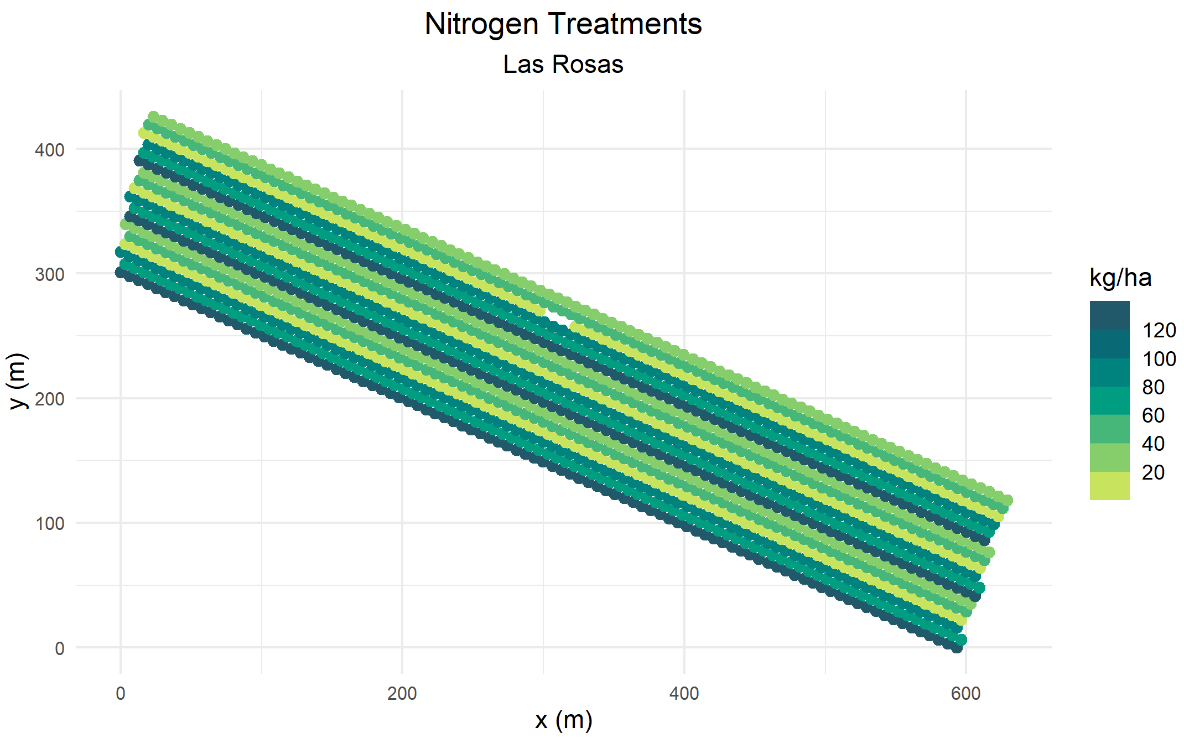
GWR+Kriging

GWR followed by (Simple) Kriging:

1. Fit GWR model using auxiliary variables
   - Yield prediction, $z(.)$
   - Treatment effect size, $\hat{\beta}_1$
2. Apply SK to the yield residuals from GWR
3. Add newly interpolated yield residuals back into the original GWR yield prediction
4. Obtain new yield residuals

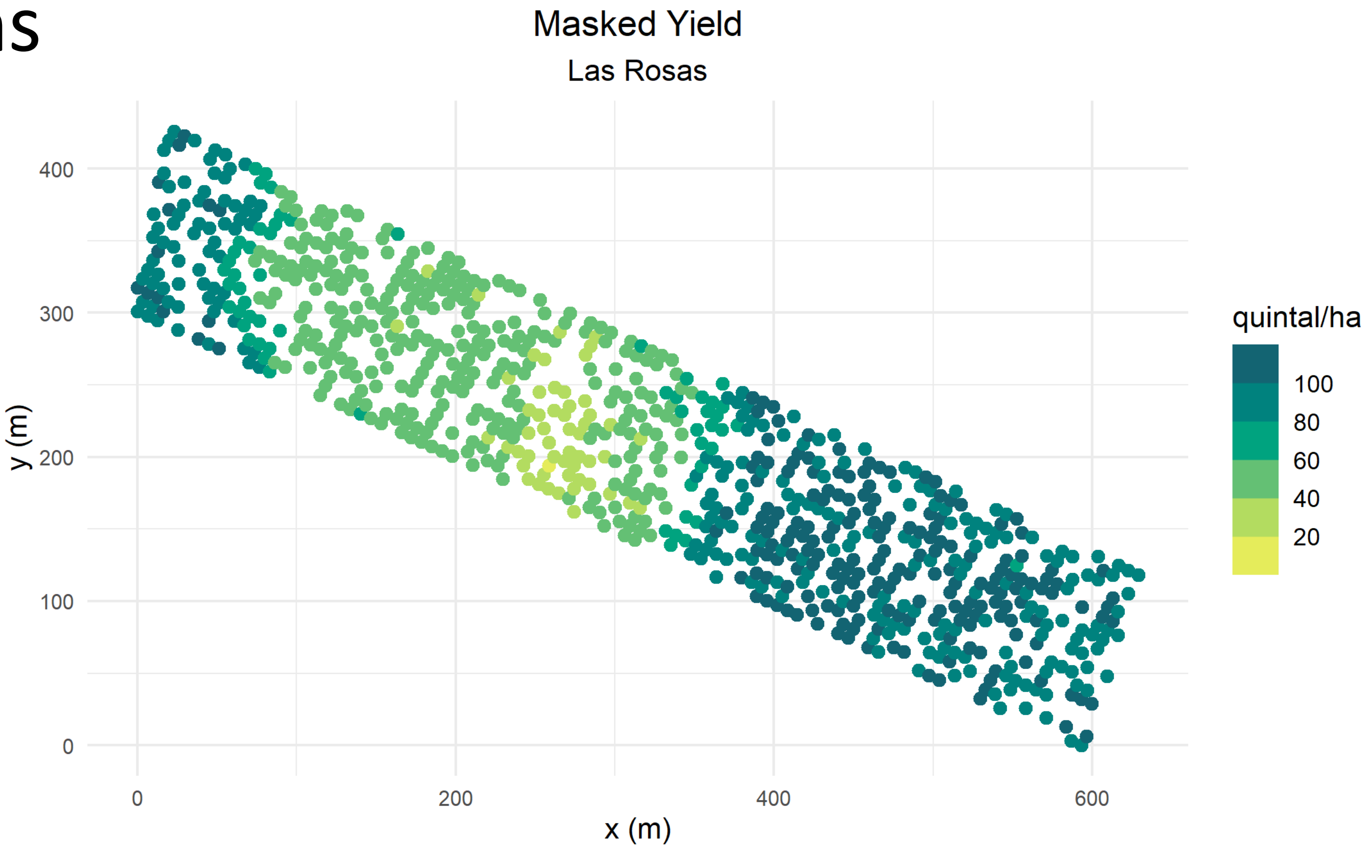
Las Rosas Cornfield Data

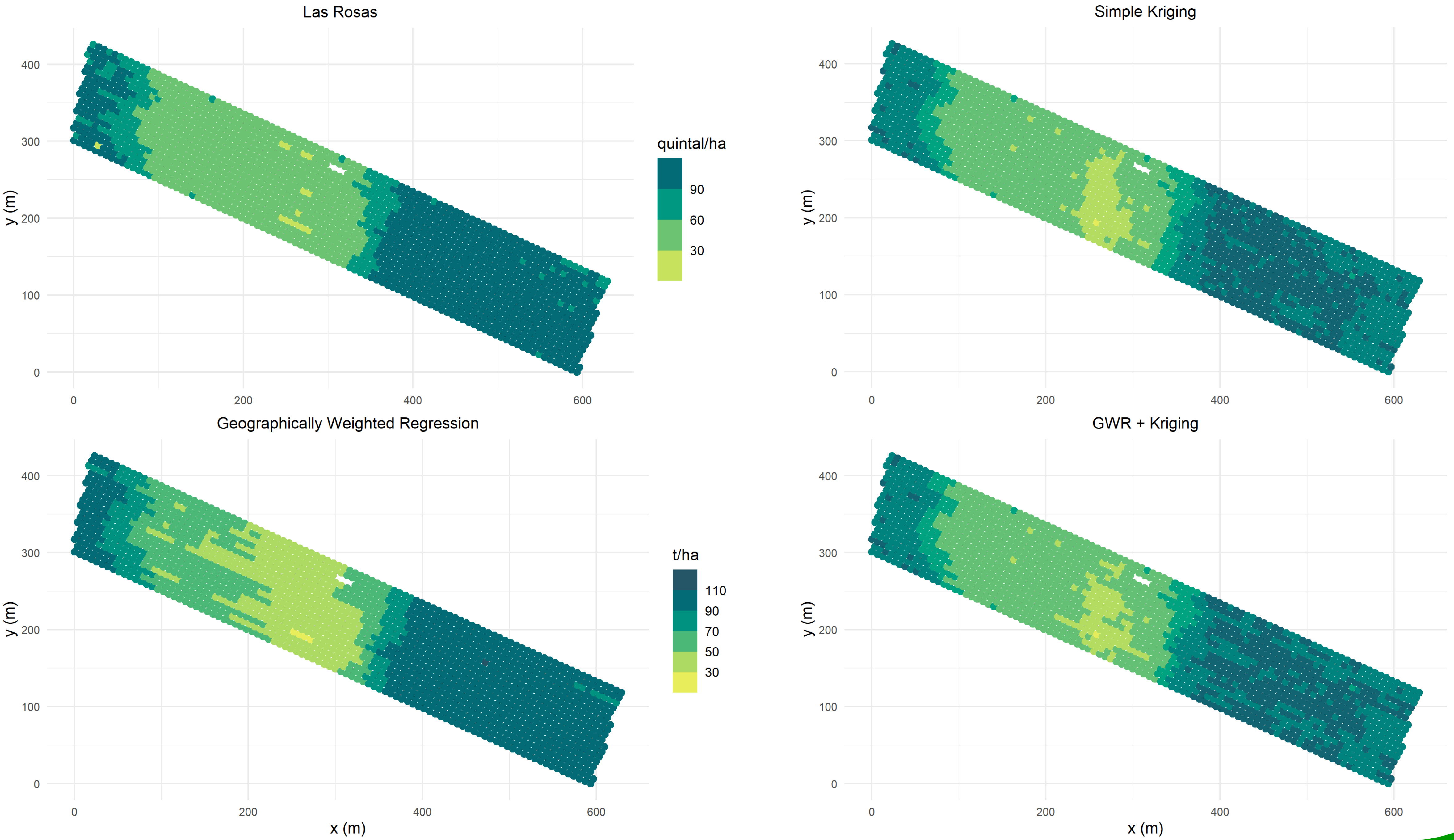




Yield Prediction Experiment

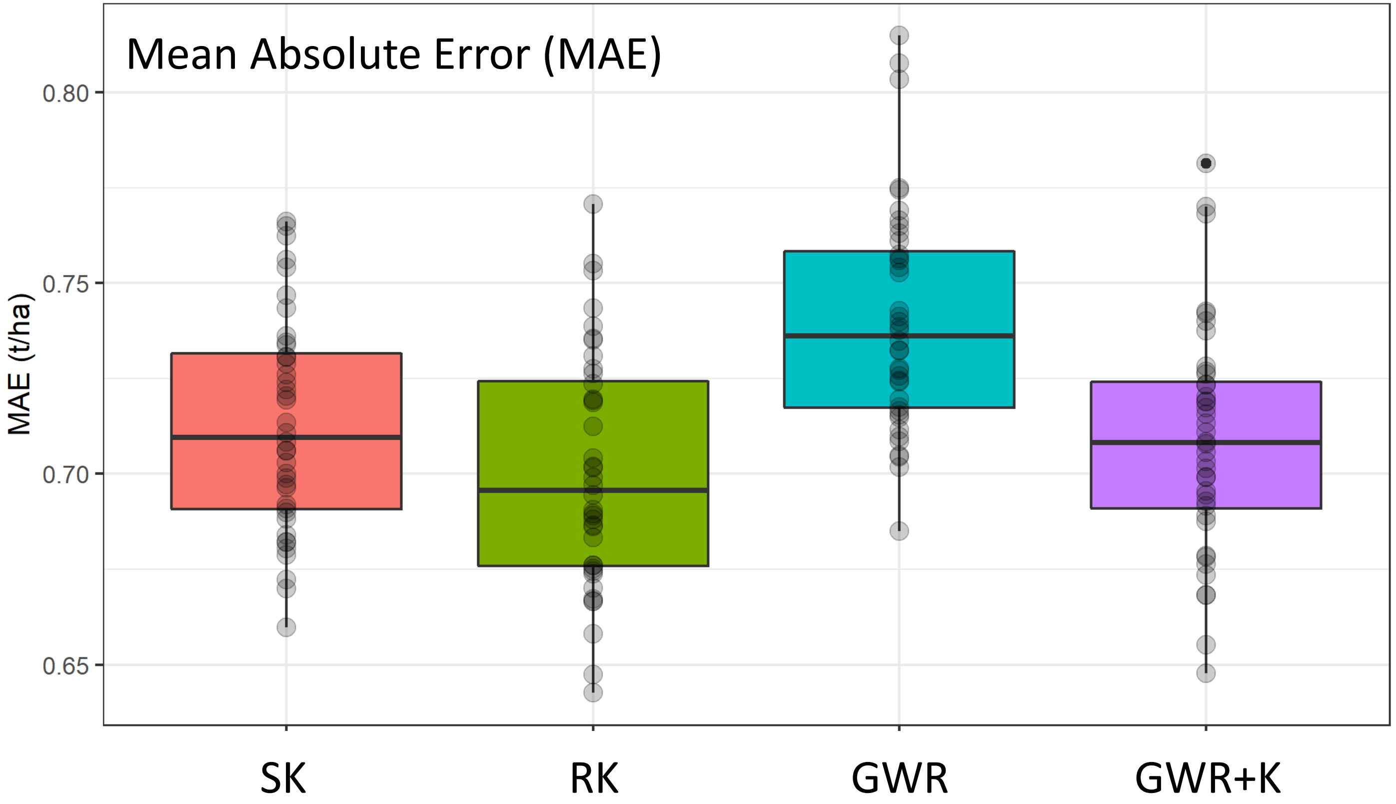
1. Randomly mask out locations
2. Predict missing values
   - Simple kriging
   - Regression kriging
   - GWR
   - GWR+K
3. Compare with known yields
   - MAE
   - RMSE
4. Repeat 40 times



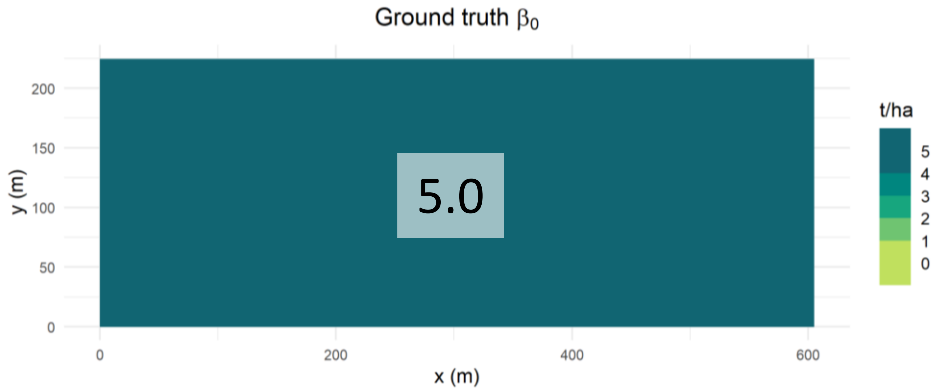

Las Rosas
quintal/ha
90
60
30
Simple Kriging
t/ha
100
80
60
40
20
Geographically Weighted Regression
t/ha
110
90
70
50
30
GWR + Kriging
t/ha
100
80
60
40
20
y (m)
x (m)
0
100
200
300
400
600

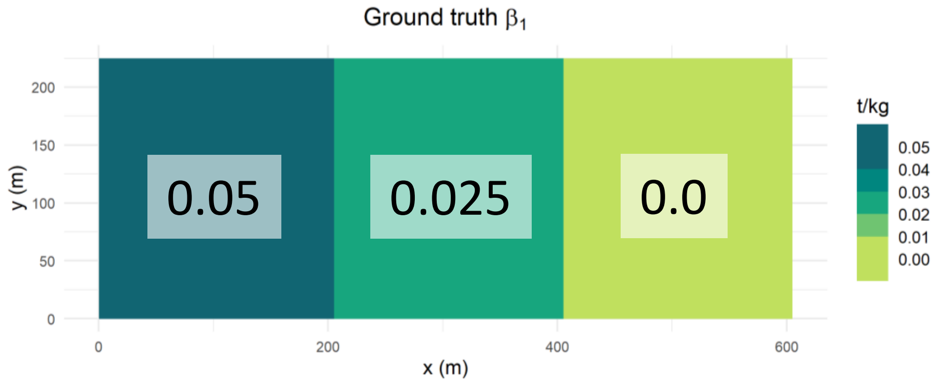
Yield Prediction Error

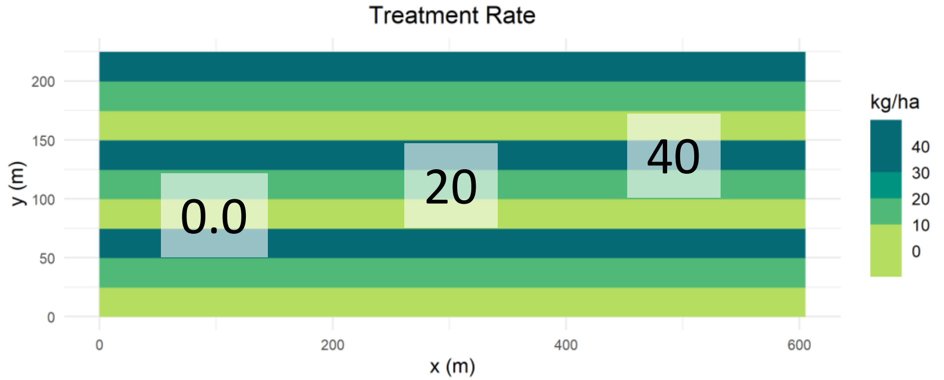
- SK, RK, and GWR+K have similar yield prediction accuracies
- GWR is distinctly worse
- GWR+K may have marginally better precision



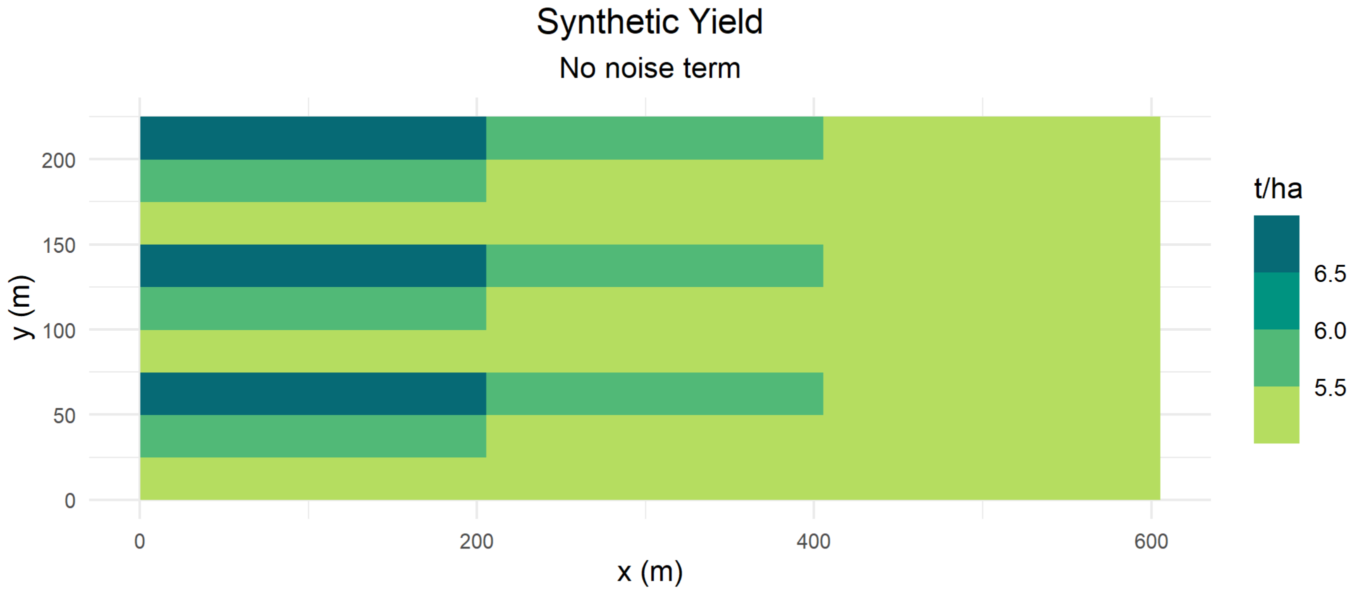
Treatment Effect Estimation: Synthetic Data



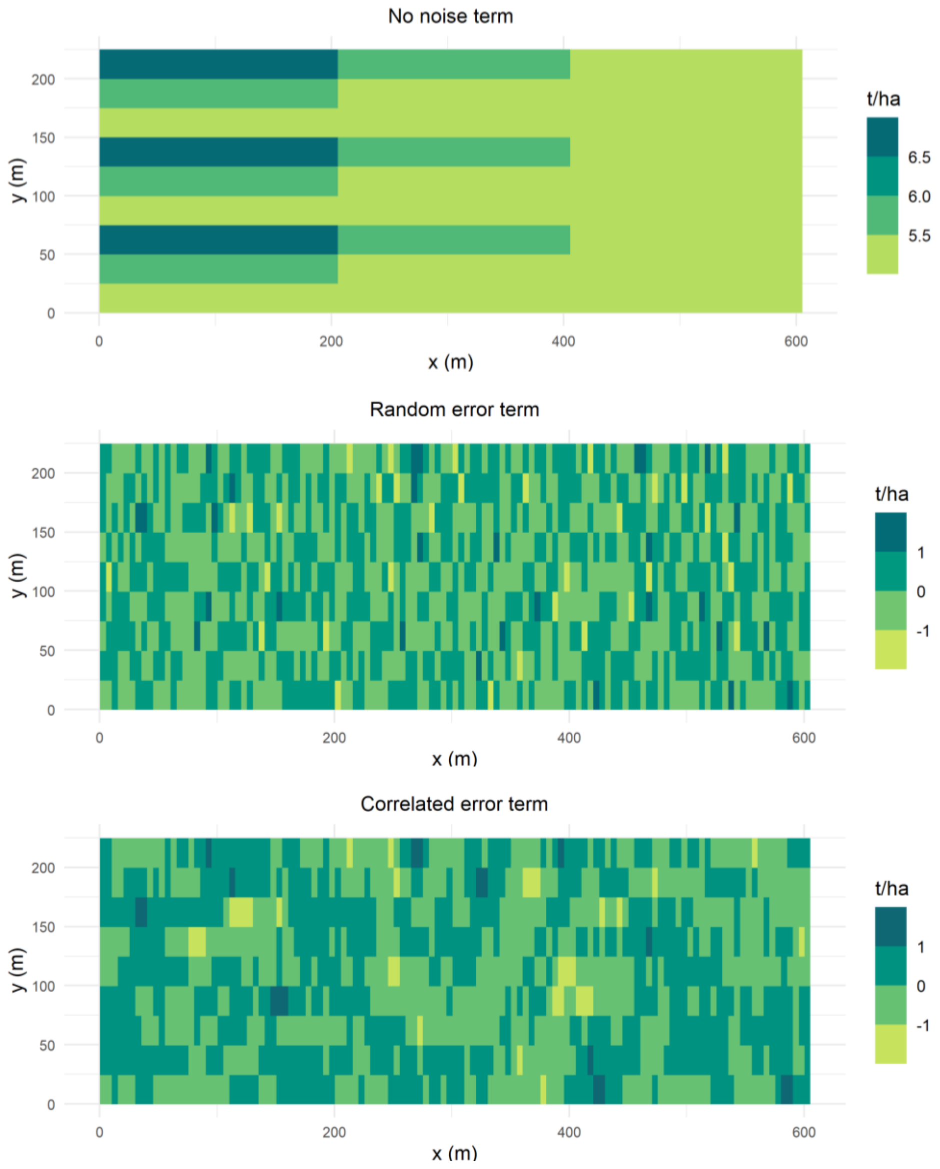




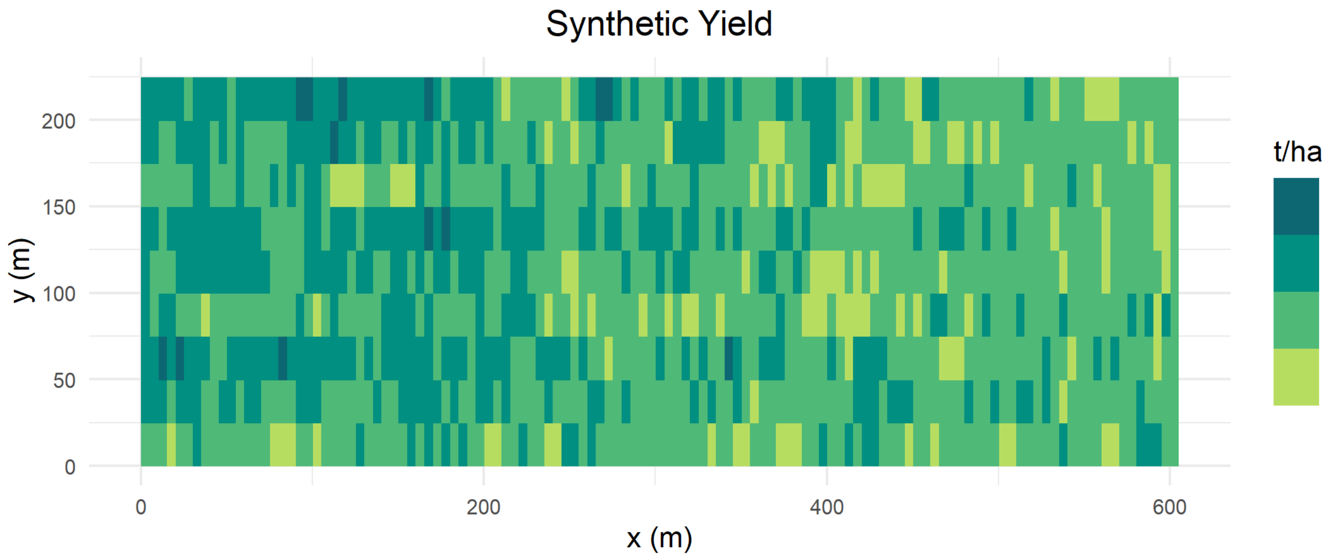
$$\text{yield} = \beta_0 + \beta_1 \times \text{treatmentRate}$$



Autocorrelated Noise

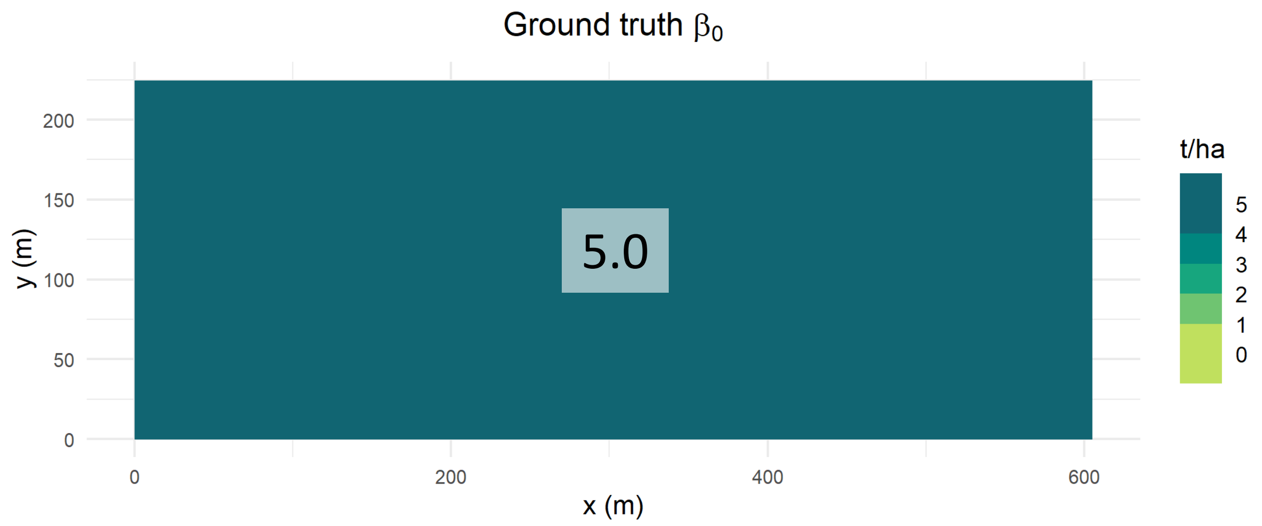


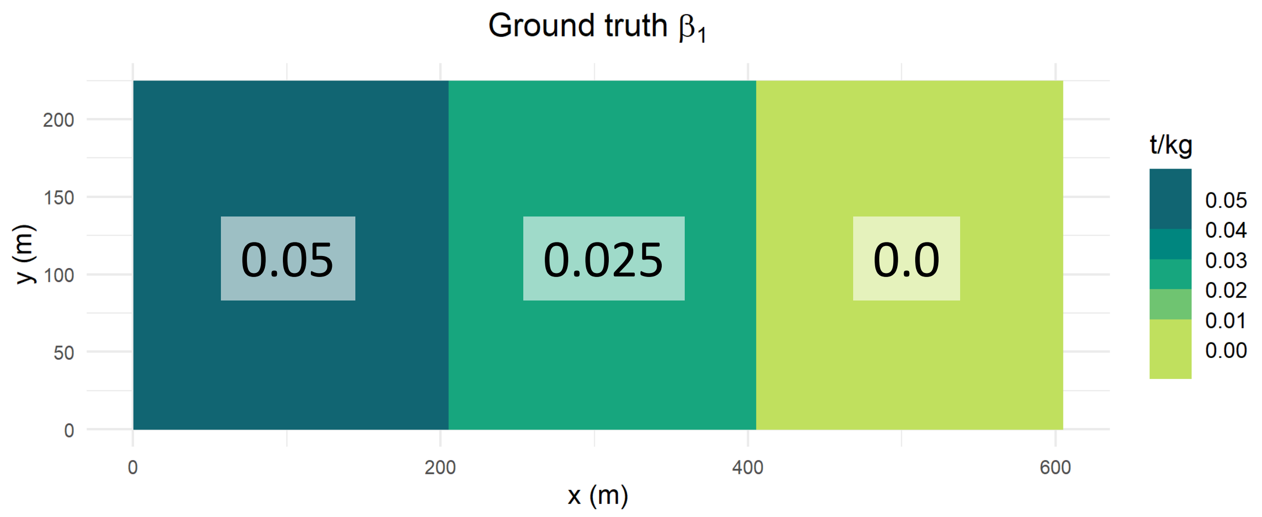
$$\text{yield} = \beta_0 + \beta_1 \times \text{treatmentRate} + \varepsilon$$



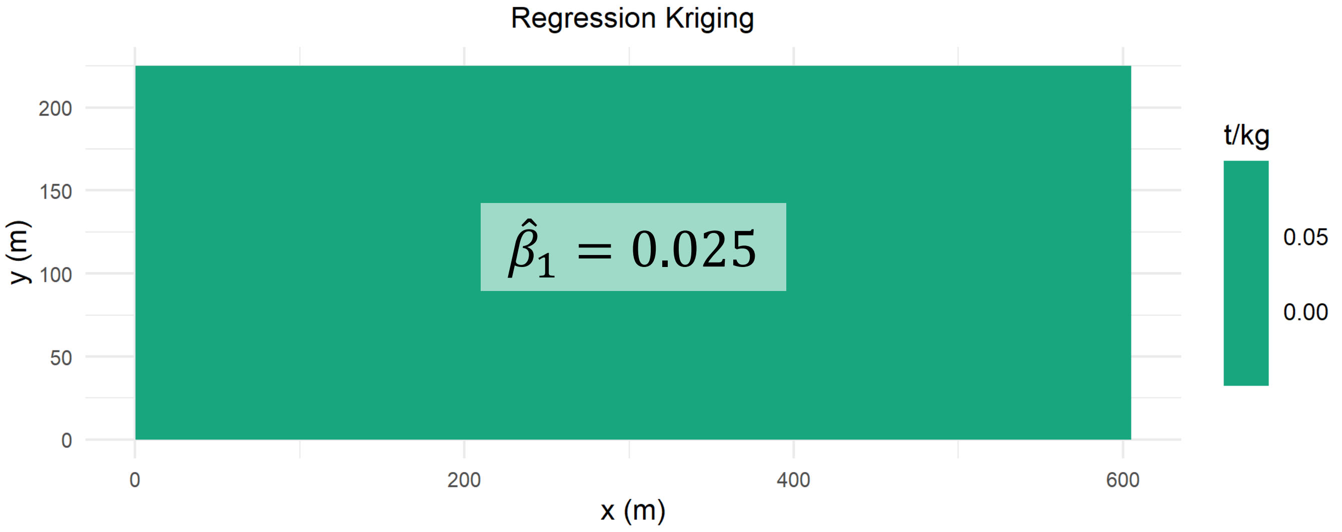
Treatment Effects Estimation Experiment

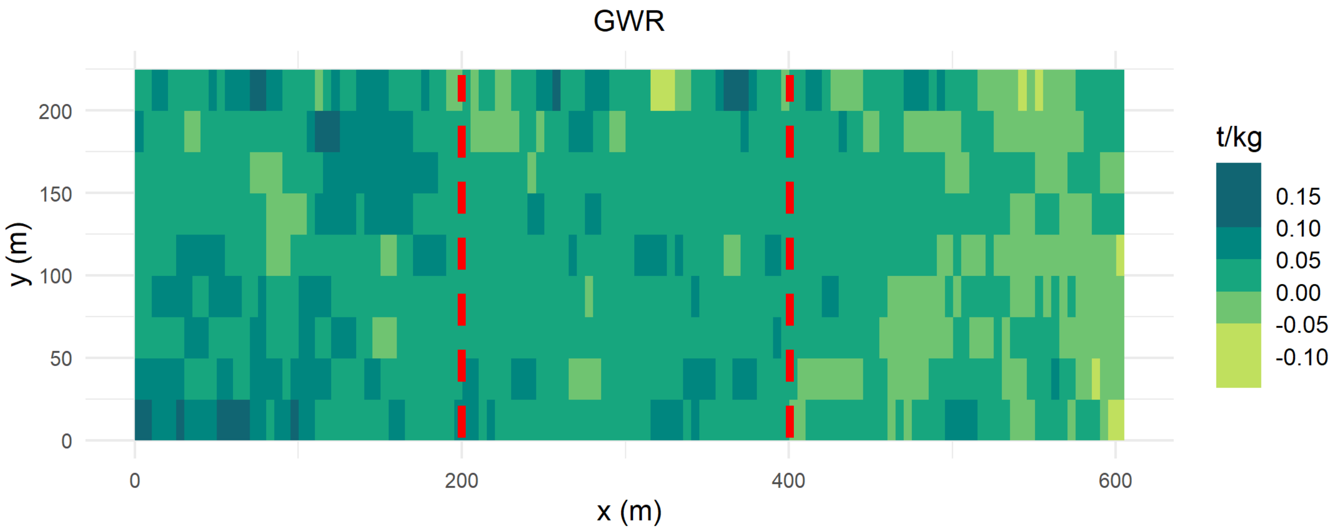
1. Use entire dataset
2. Fit regression model
   - Regression Kriging
   - GWR
3. Compare with known coefficients
   - $\beta_0 = 5$
   - $\beta_1 = 0.0, 0.025, 0.05$



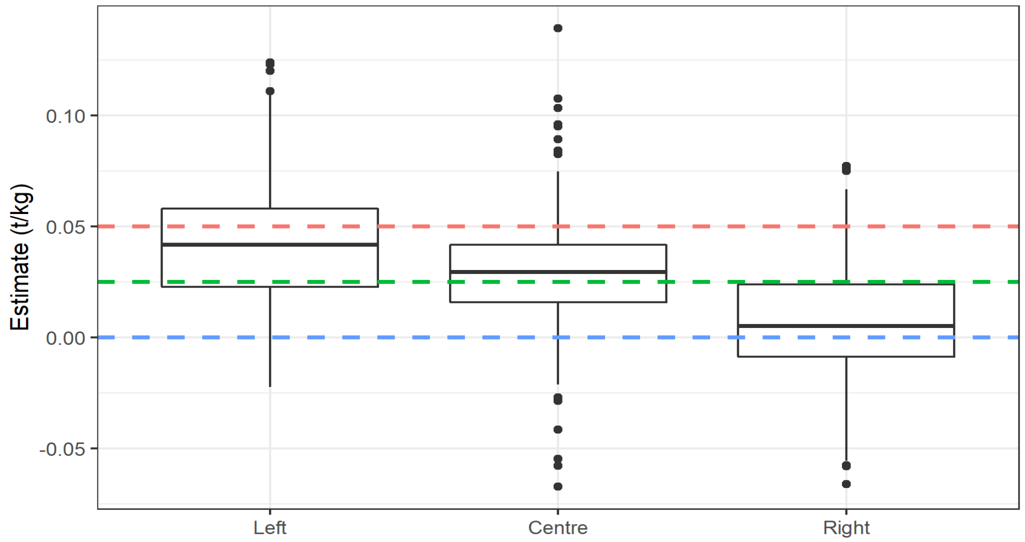


Estimated β_1

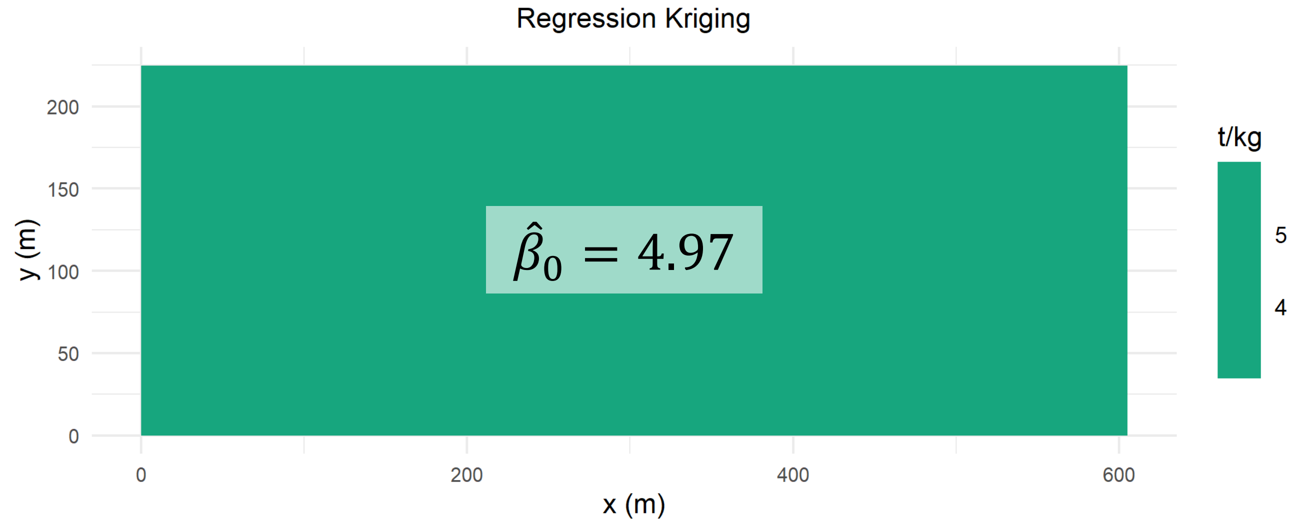


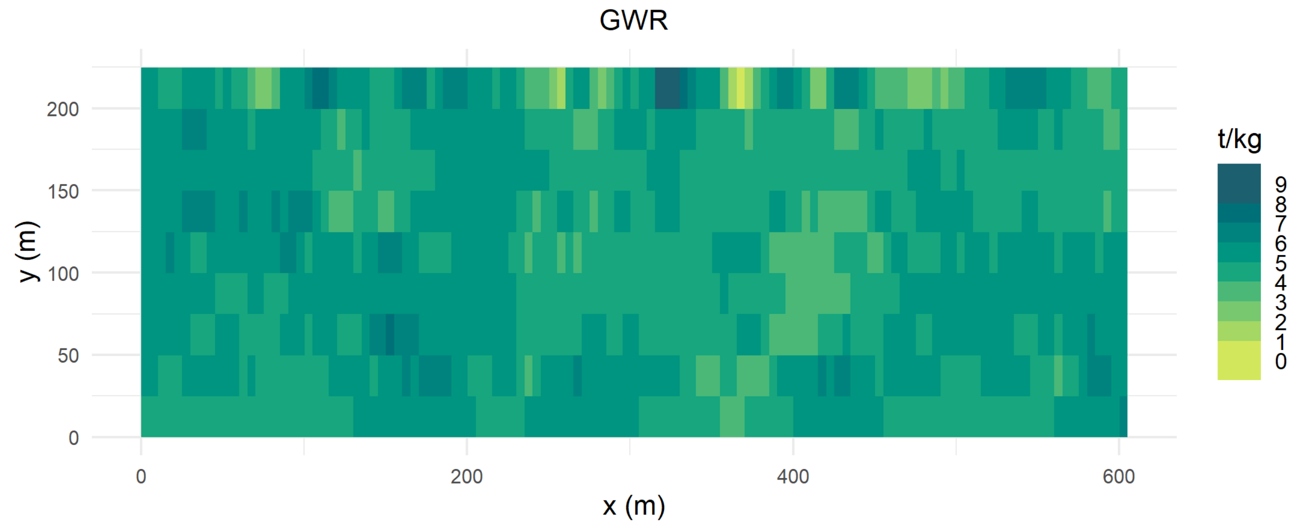


- GWR handles non-stationarity.
- GWR estimates are biased toward the average.
- Zones are unknown in practice.

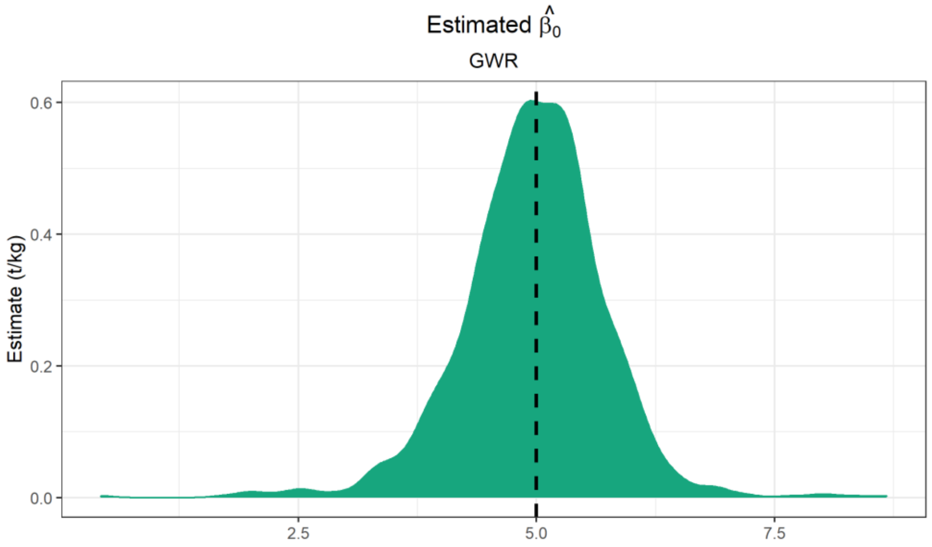


Estimated β_0





- RK returns a global estimate.
- GWR affected by autocorrelated noise.
- Edge artefacts in GWR result.



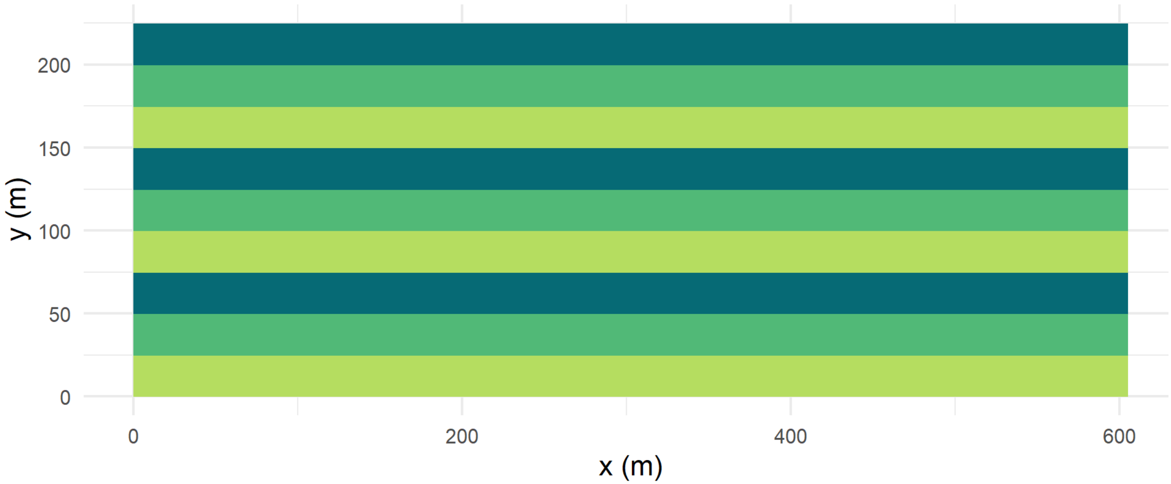
Recommendations

- Regression Kriging is sufficient for yield prediction.
- GWR+K provides both yield and treatment estimates in the face of autocorrelation and non-stationarity.

		Auto-correlation	
		Present	*Not Present*
Non-stationarity	*Present*	**GWR+K**	**GWR**
	Not Present	**RK**	**Either**

GWR Kernel

- Shape
  - Gaussian, Exponential, Bisquare, Tricube, Boxcar.
- Bandwidth
  - Consider strip dimensions



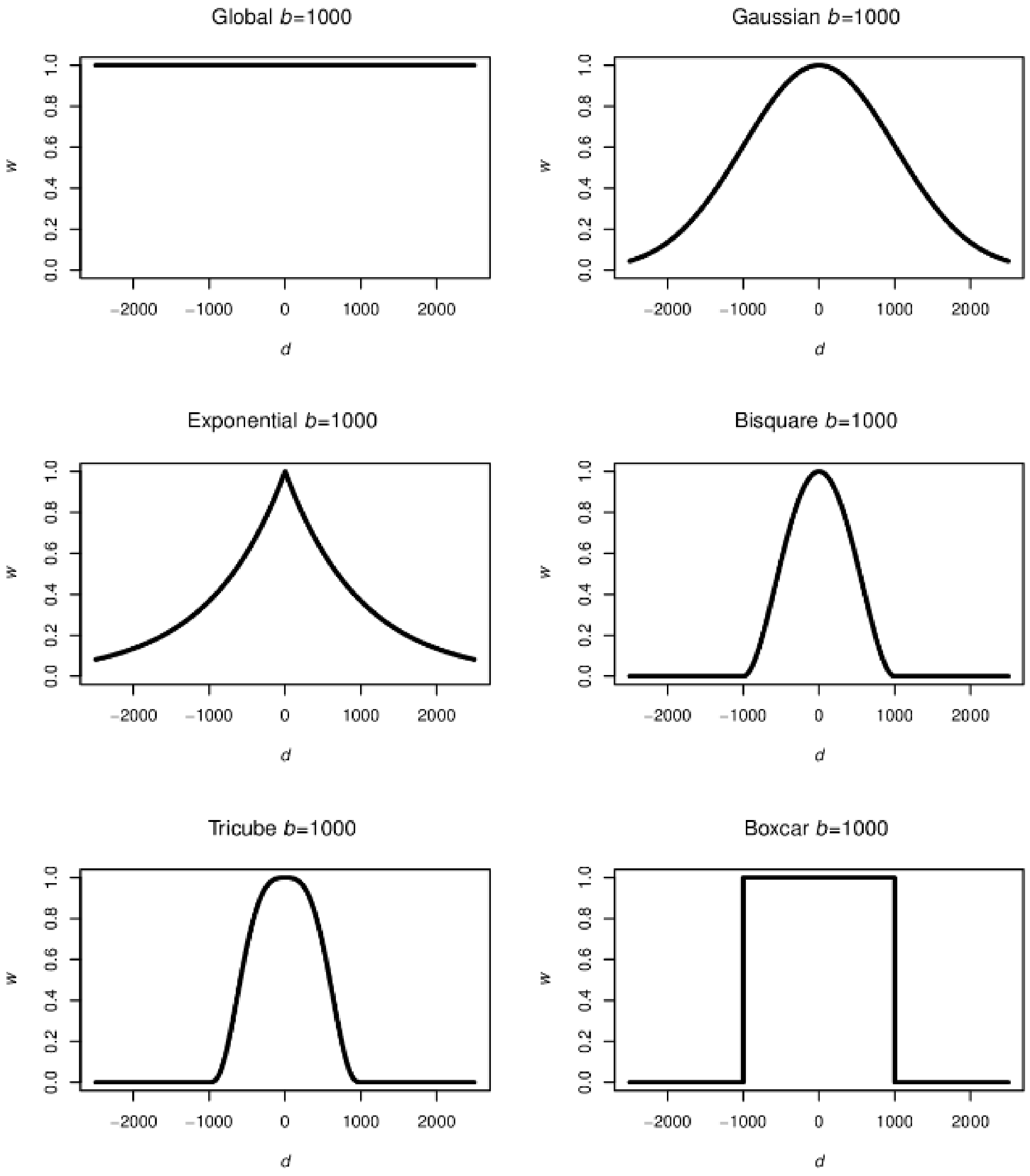
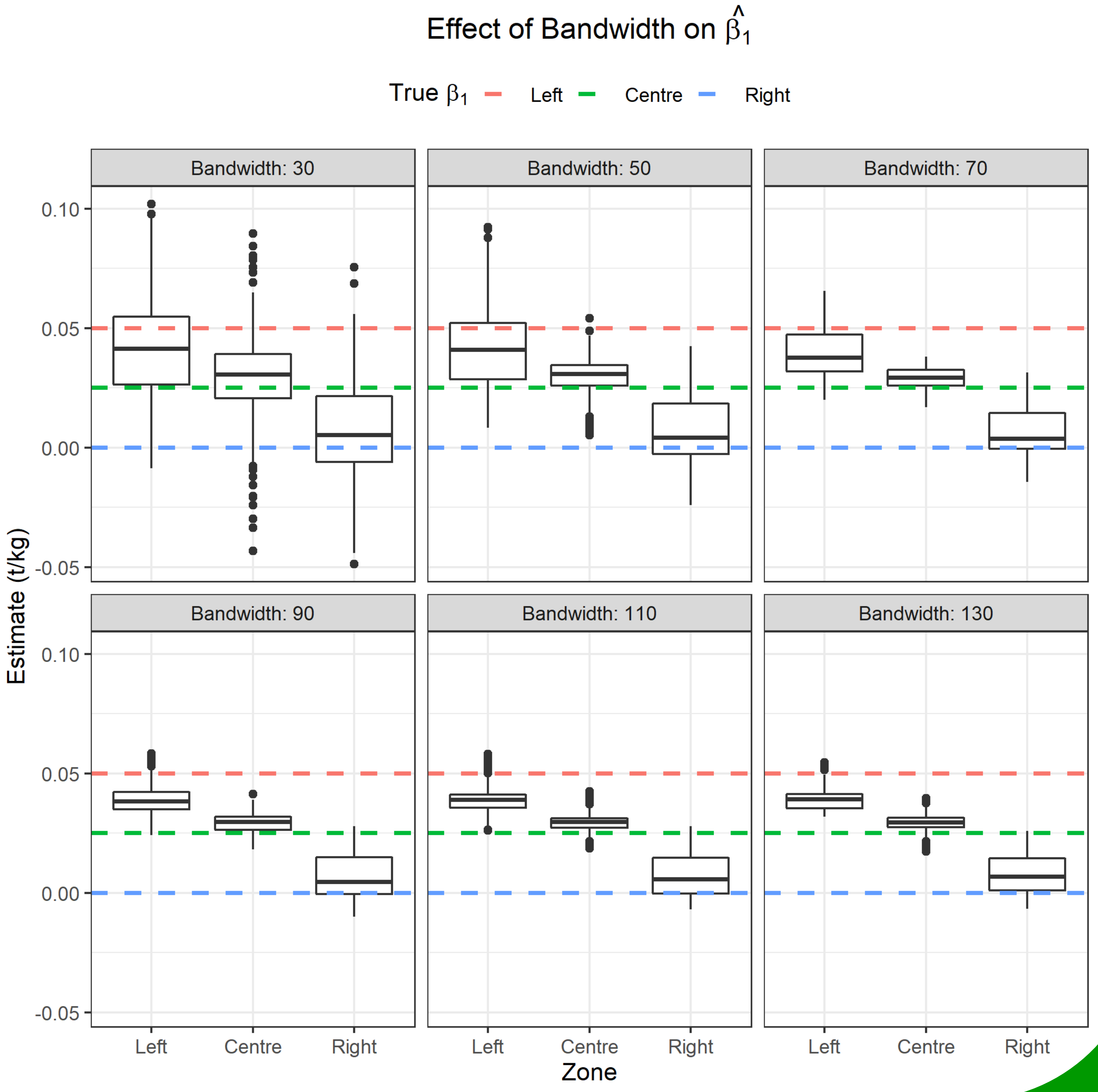


Figure: Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software, 63*(17), 1 - 50. doi: http://dx.doi.org/10.18637/jss.v063.i17

Effect of Bandwidth

As (boxcar) bandwidth size is increased:

1. Precision of treatment estimates improves.
2. Estimates are pulled towards global average.



Conclusion

- Precision agriculture promotes efficiency and sustainability by helping farmers target their interventions.
- On-farm strip experiments are conducted at large practical scales where we cannot ignore spatial heterogeneity.
- Spatial heterogeneity is mostly explained by non-stationarity and autocorrelation.
- A GWR+K approach produces good yield predictions and treatment effect estimates in the face of this heterogeneity.
- GWR kernel bandwidth must be chosen with care.

Thank you

GRDC

SAGI
Statistics for the Australian Grains Industry
National | North | West | South

Curtin University