

# Analysis of soil at India

*Facundo Muñoz*

*March 1, 2016*

## Contents

<b>1 Data description</b>	<b>2</b>
<b>2 Cost-based distances</b>	<b>6</b>
<b>3 Analysis of Calcium</b>	<b>7</b>
<b>4 Analysis of Copper</b>	<b>15</b>
<b>5 Analysis of Ferrum</b>	<b>23</b>
<b>6 Analysis of Potassium</b>	<b>31</b>
<b>7 Analysis of Magnesium</b>	<b>39</b>
<b>8 Analysis of Zinc</b>	<b>47</b>
<b>9 Analysis of food remains</b>	<b>55</b>
<b>10 Analysis of living room</b>	<b>63</b>
<b>11 Analysis of enclosed spaces</b>	<b>71</b>
<b>12 Analysis of burningareas</b>	<b>79</b>
<b>13 Analysis of ashes</b>	<b>87</b>
<b>14 Conclusions</b>	<b>95</b>

geoRcb package v.1.7.6

Here we compare the outcome of a classical kriging against a cost-based kriging which takes into account the presence of a barrier.

# 1 Data description

```
'data.frame': 70 obs. of 16 variables:
$ sample      : Factor w/ 70 levels "JIN10","JIN100",...: 28 45 66 1 16 22 25 26 27 29 ...
$ Area        : Factor w/ 2 levels "Inside","Veranda": 1 1 1 1 1 1 1 1 1 1 ...
$ Side         : Factor w/ 2 levels "Fireplace","Storage": 2 2 2 2 1 2 1 1 2 2 ...
$ x            : num 13 12.5 13.5 13.5 14 13.5 14 14 14 ...
$ y            : num -11.5 -12 -12 -13 -11 -11.5 -10 -10.5 -13.5 -12.5 ...
$ Ca           : num 2.95 3.4 4.3 5.7 3.97 4.5 3.3 3.2 2.81 5.07 ...
$ Cu           : int 15 13 15 13 14 17 20 18 14 14 ...
$ Fe           : num 1.21 1.18 1.3 1.36 1.04 1.27 1.23 1.43 1.41 1.27 ...
$ K            : num 0.28 0.32 0.31 0.21 0.27 0.27 0.18 0.33 0.21 0.27 ...
$ Mg           : num 0.54 0.51 0.57 0.55 0.5 0.53 0.49 0.56 0.45 0.6 ...
$ Zn           : int 46 31 36 28 32 33 36 41 36 35 ...
$ foodremains : num 2.24 2.27 2.47 2.05 2.21 ...
$ livingroom   : num 1.73 1.74 1.88 1.48 1.65 ...
$ enclosedspaces: num 1.03 1.12 1.34 1.57 1.26 ...
$ burningareas : num 0.911 0.969 1.016 0.651 0.895 ...
$ ashes        : num 1.43 1.57 1.77 1.65 1.59 ...
```

	sample	Area	Side	x	y
JIN10	: 1	Inside :47	Fireplace:36	Min. : 7.50	Min. :-13.50
JIN100	: 1	Veranda:23	Storage :34	1st Qu.: 9.50	1st Qu.:-12.00
JIN101	: 1			Median :11.50	Median :-11.00
JIN102	: 1			Mean :11.53	Mean :-10.99
JIN103	: 1			3rd Qu.:13.50	3rd Qu.:-10.00
JIN106	: 1			Max. :16.00	Max. :-8.00
(Other)	:64				
	Ca	Cu	Fe	K	
Min.	:0.660	Min. : 7.00	Min. :0.810	Min. :0.1400	
1st Qu.	:2.490	1st Qu.:11.00	1st Qu.:1.050	1st Qu.:0.2300	
Median	:2.945	Median :12.00	Median :1.190	Median :0.2600	
Mean	:3.106	Mean :12.39	Mean :1.175	Mean :0.2679	
3rd Qu.	:3.697	3rd Qu.:14.00	3rd Qu.:1.300	3rd Qu.:0.2900	
Max.	:5.700	Max. :20.00	Max. :1.500	Max. :0.6300	
	Mg	Zn	foodremains	livingroom	
Min.	:0.220	Min. :17.00	Min. :1.339	Min. :1.134	
1st Qu.	:0.390	1st Qu.:26.00	1st Qu.:1.826	1st Qu.:1.433	
Median	:0.460	Median :32.00	Median :2.006	Median :1.549	
Mean	:0.453	Mean :31.77	Mean :2.000	Mean :1.552	
3rd Qu.	:0.520	3rd Qu.:35.75	3rd Qu.:2.198	3rd Qu.:1.659	
Max.	:0.660	Max. :62.00	Max. :2.675	Max. :1.895	
		NA's :3	NA's :7		
	enclosedspaces	burningareas	ashes		
Min.	:0.2324	Min. :0.5497	Min. :0.9181		
1st Qu.	:0.7993	1st Qu.:0.7675	1st Qu.:1.3193		
Median	:0.9824	Median :0.8420	Median :1.4119		
Mean	:1.0086	Mean :0.8475	Mean :1.4126		
3rd Qu.	:1.2152	3rd Qu.:0.9204	3rd Qu.:1.5366		
Max.	:1.6038	Max. :1.1363	Max. :1.8757		
		NA's :5	NA's :4		

Figures 1 and 2 display the raw data, and an exploratory smoothed surface.

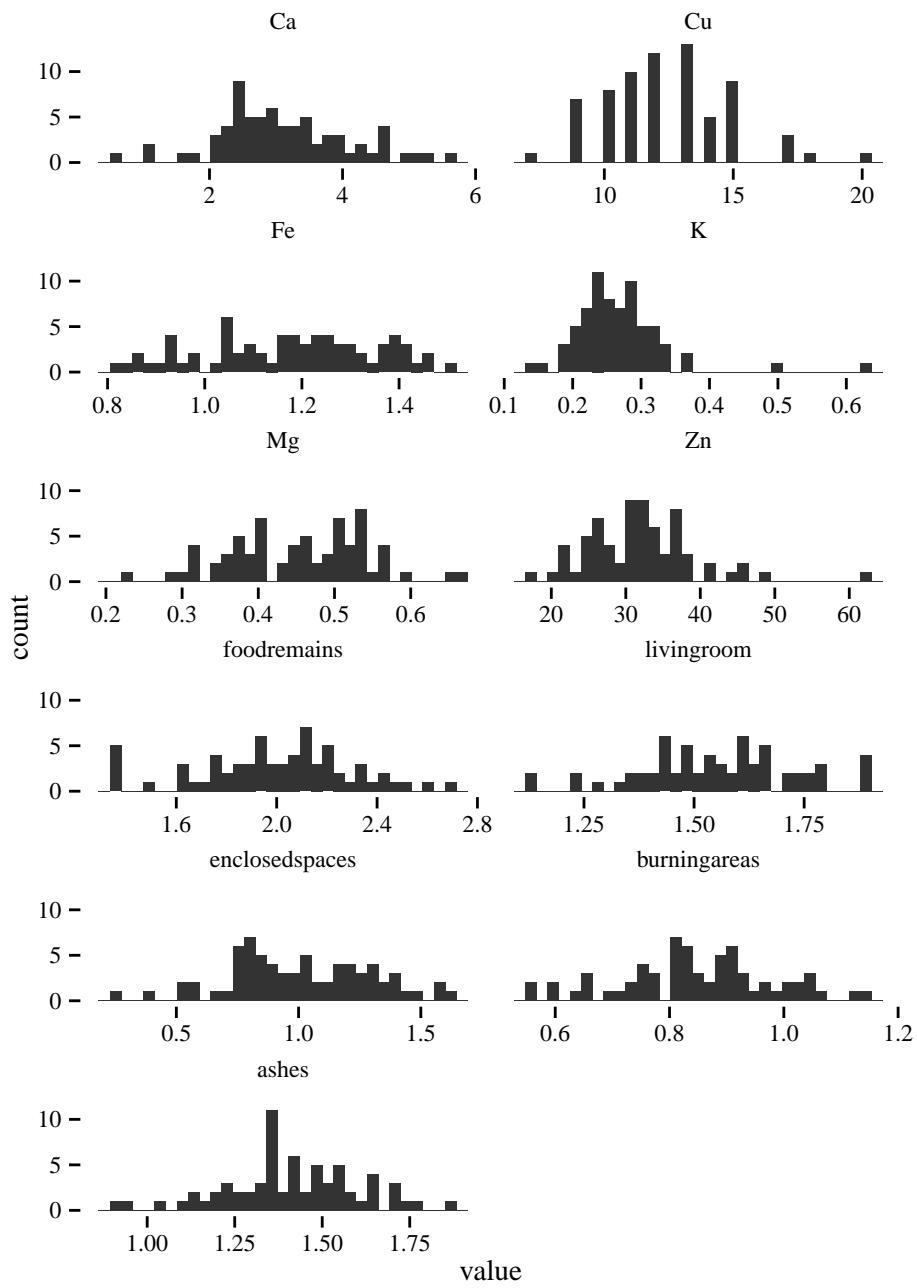


Figure 1: Histograms of measured variables.

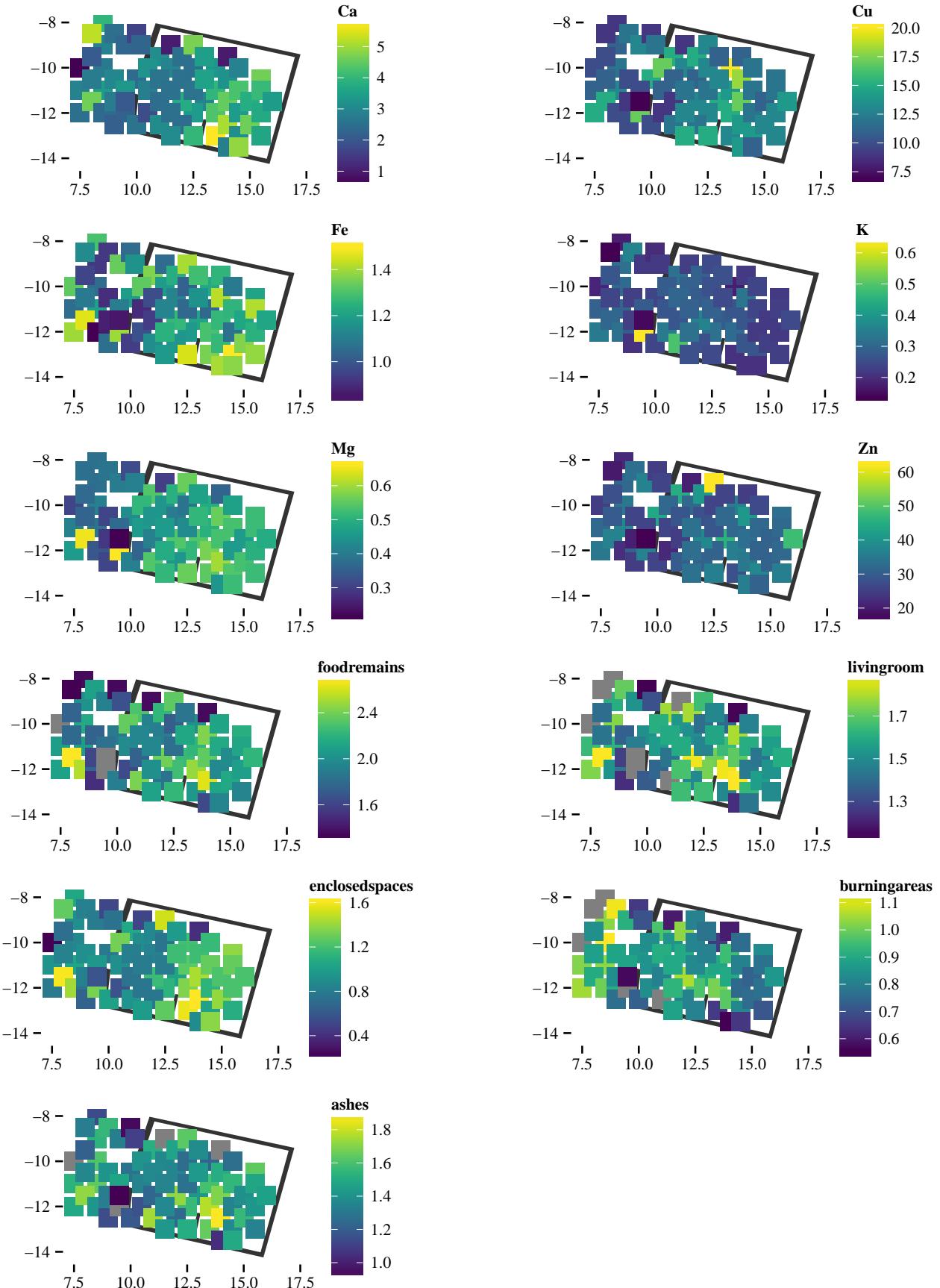


Figure 2: Measurement locations and observed values

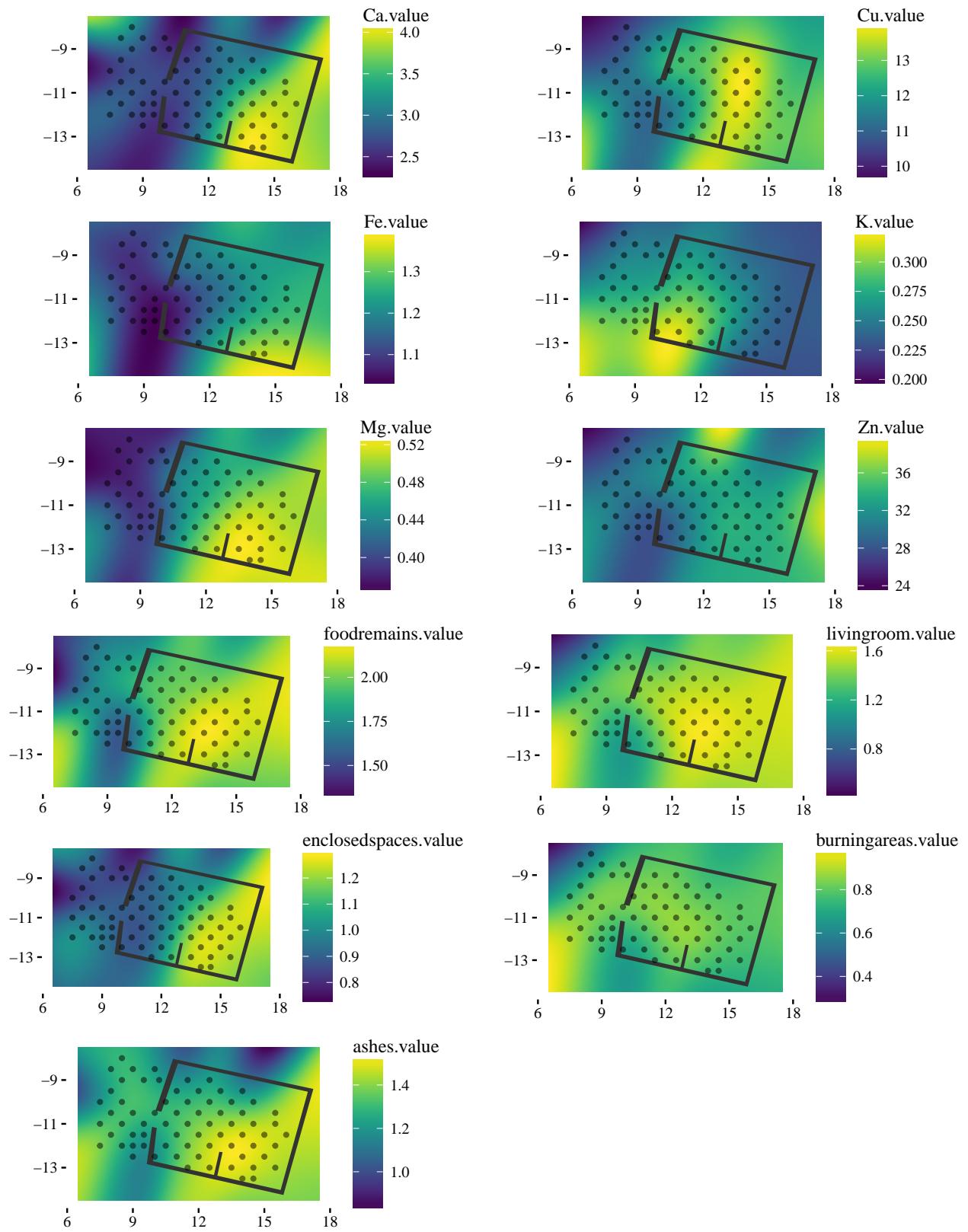


Figure 3: Exploratory kernel smoothing of the measurements

## 2 Cost-based distances

Here we set up the cost-based surface, and compute some cost-based maps, for verifications purposes.

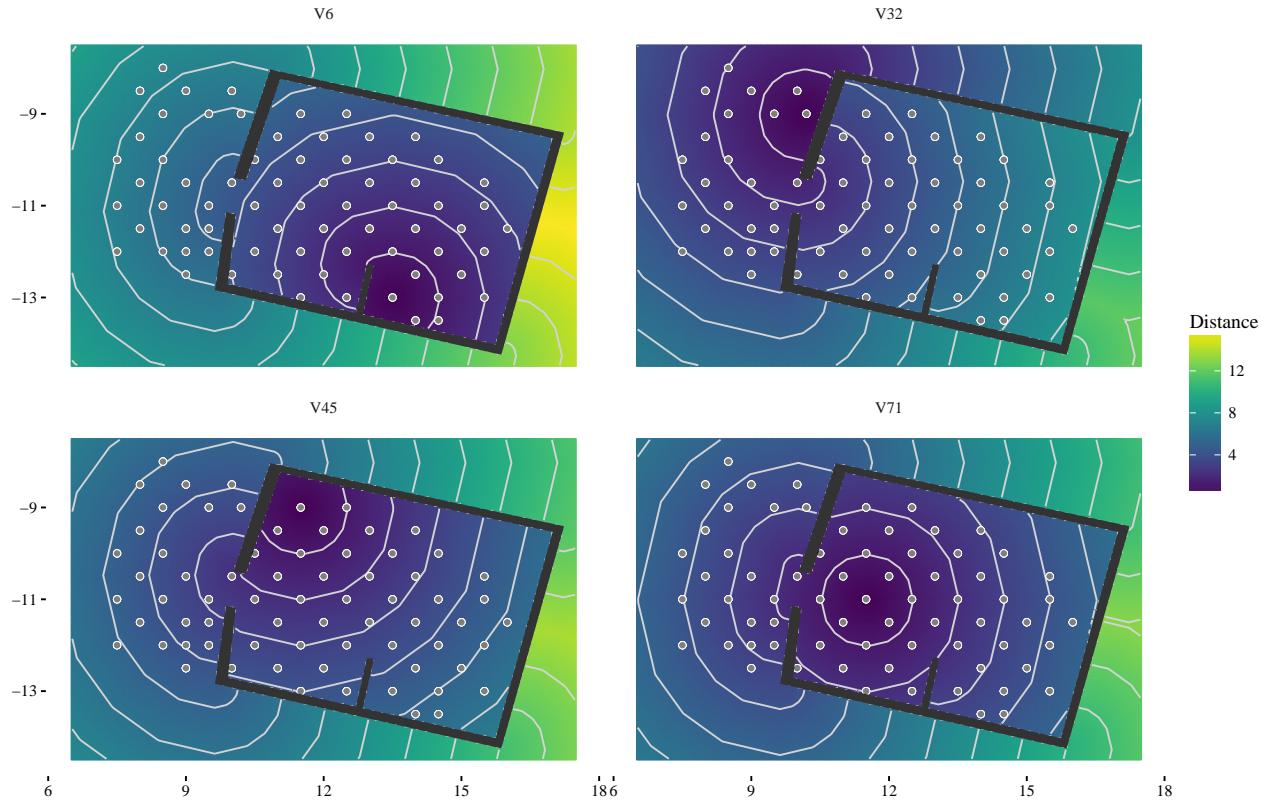


Figure 4: Some cost-based maps to selected observations.

### 3 Analysis of Calcium

#### 3.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

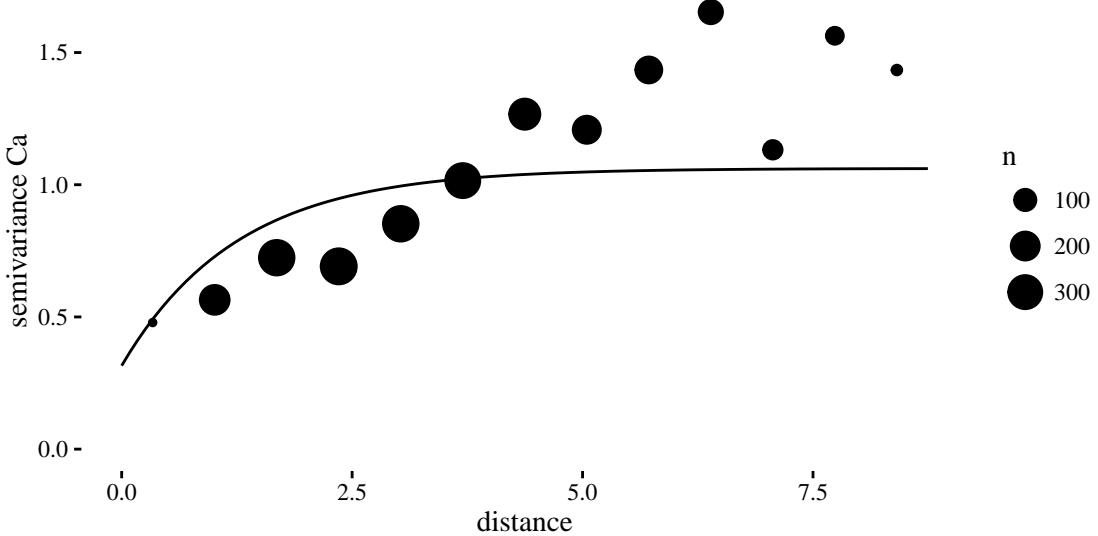


Figure 5: Empirical variogram and fitted model.

#### 3.2 Cost-based kriging

#### 3.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	3.12	3.17
Nugget	0.32	0.60
Partial sill	0.75	0.85
phi	1.25	6.53
Pract. range	3.75	19.56
Log-likelihood	-89.25	-89.82

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

#### 3.4 Leave-one-out Cross Validation (LOOCV)

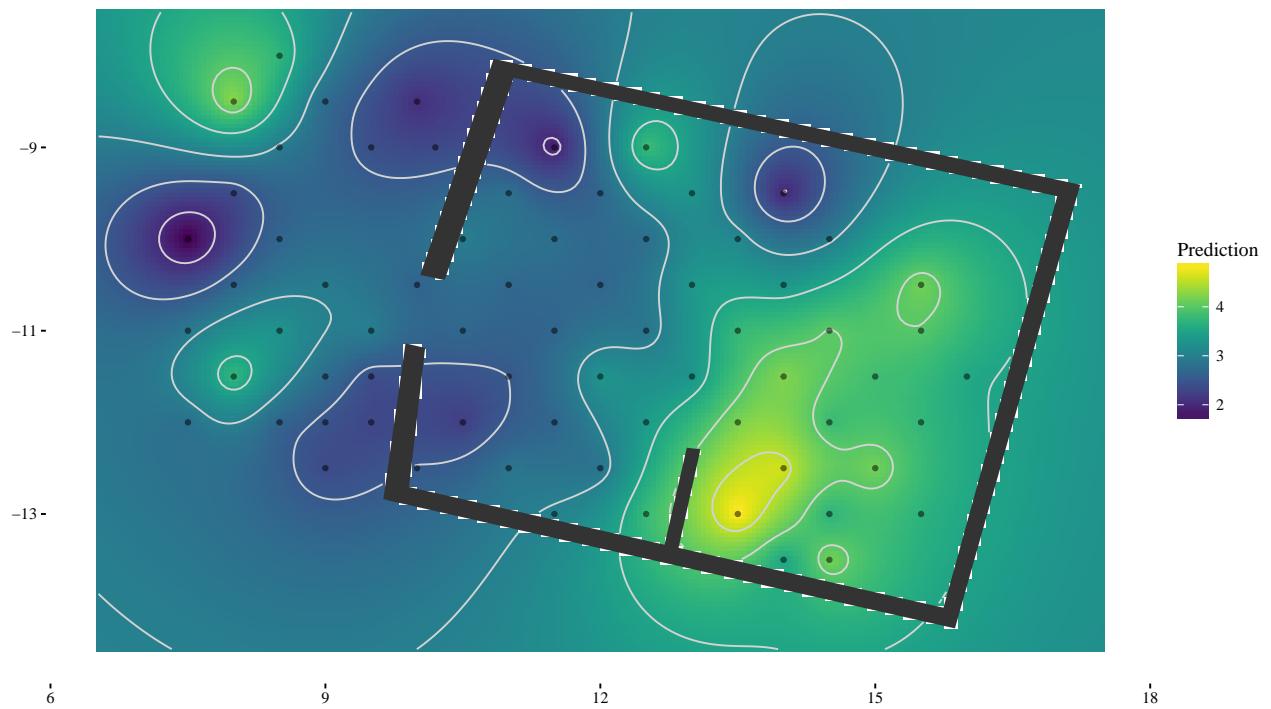


Figure 6: Euclidean kriging prediction

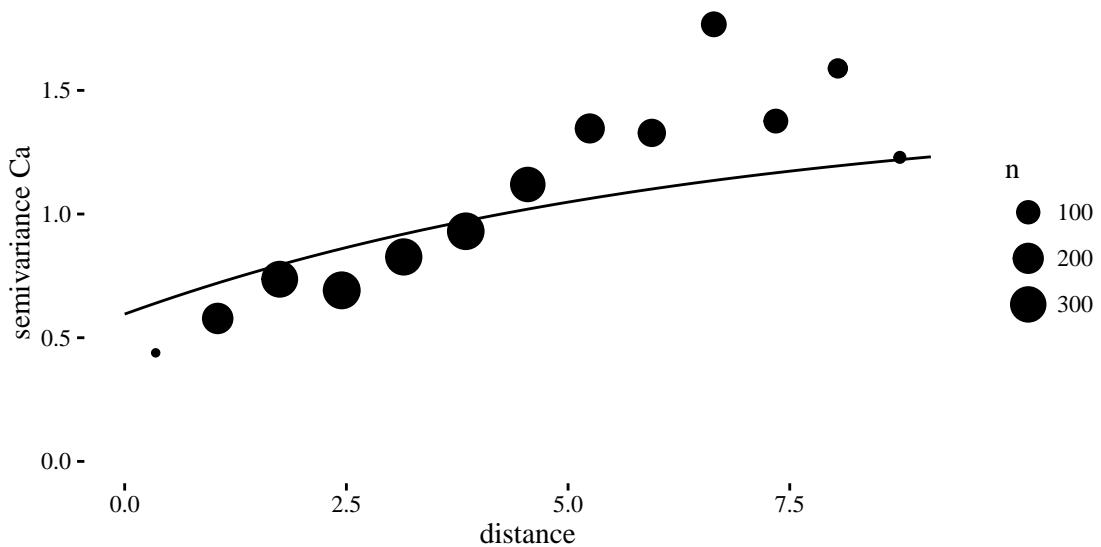


Figure 7: Empirical cost-based variogram and fitted model.

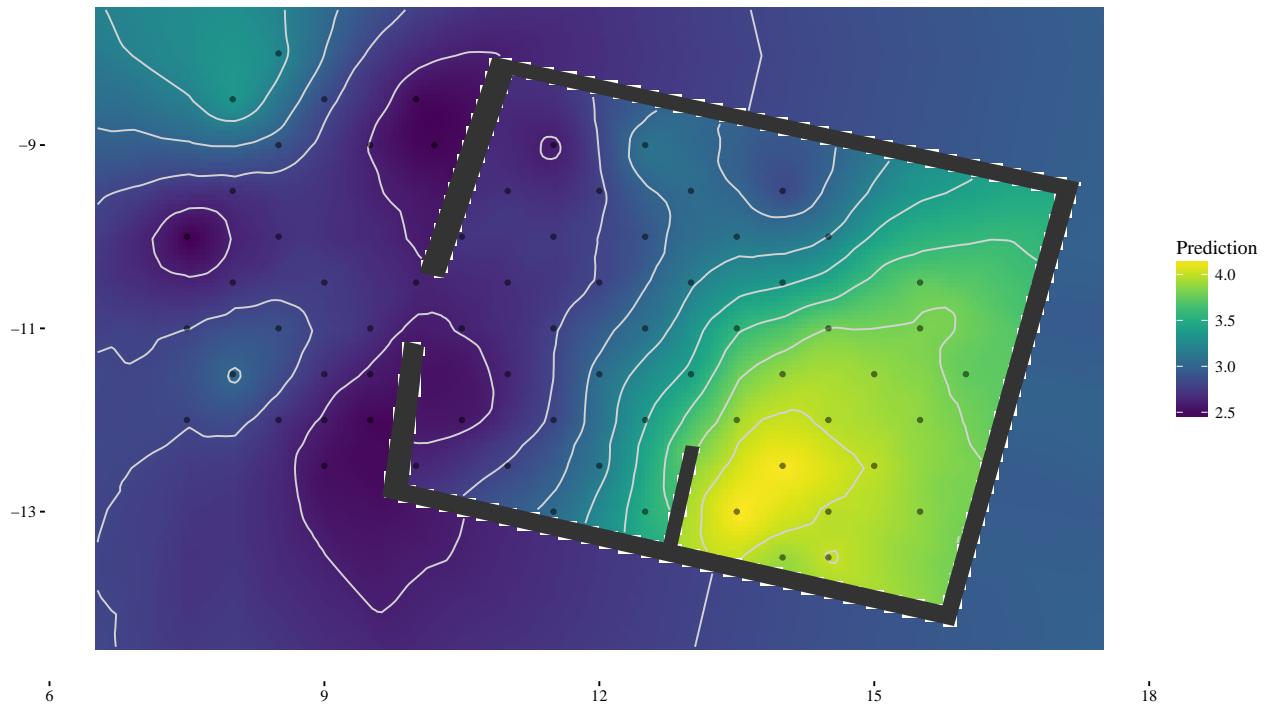


Figure 8: Cost-based kriging prediction

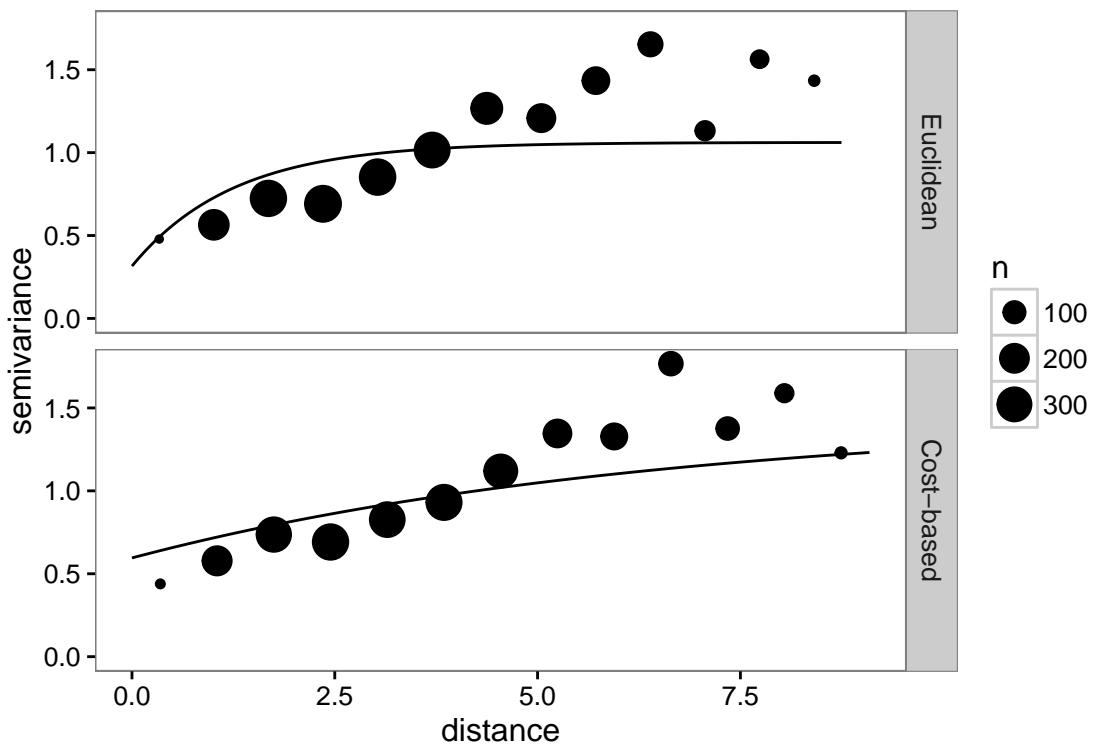


Figure 9: Empirical variogram and fitted models by method for Calcium.

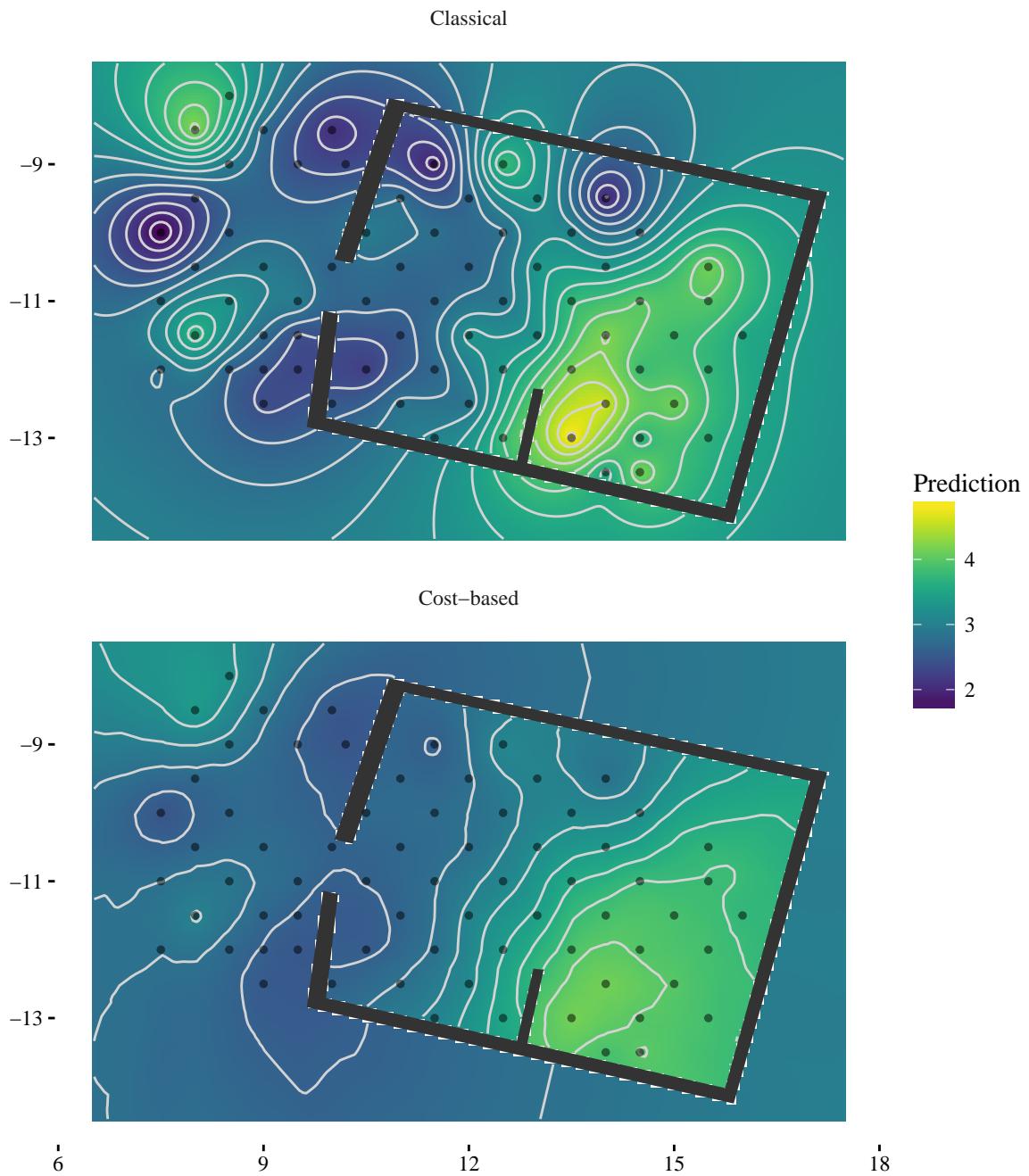


Figure 10: Comparison of Kriging estimates.

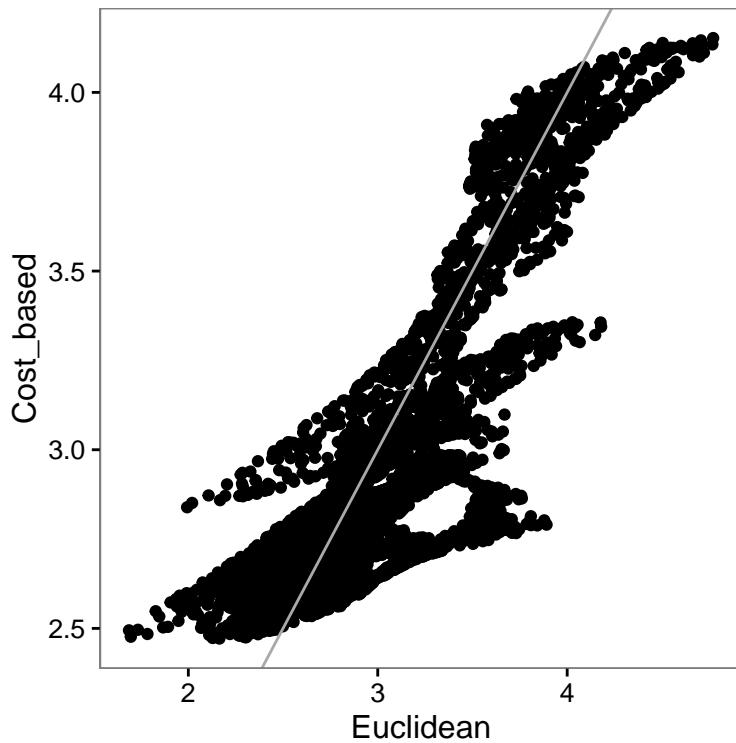


Figure 11: Pointwise comparison of predictions by method.

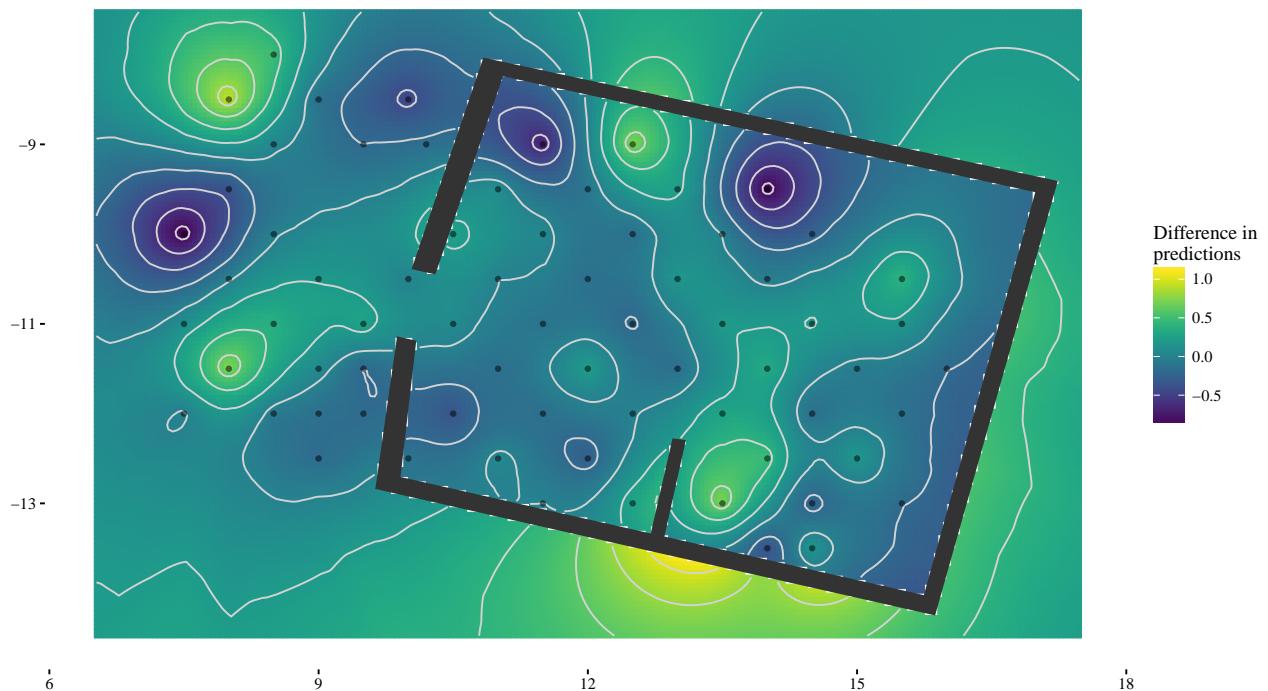


Figure 12: Difference between the Euclidean and the cost-based predictions.

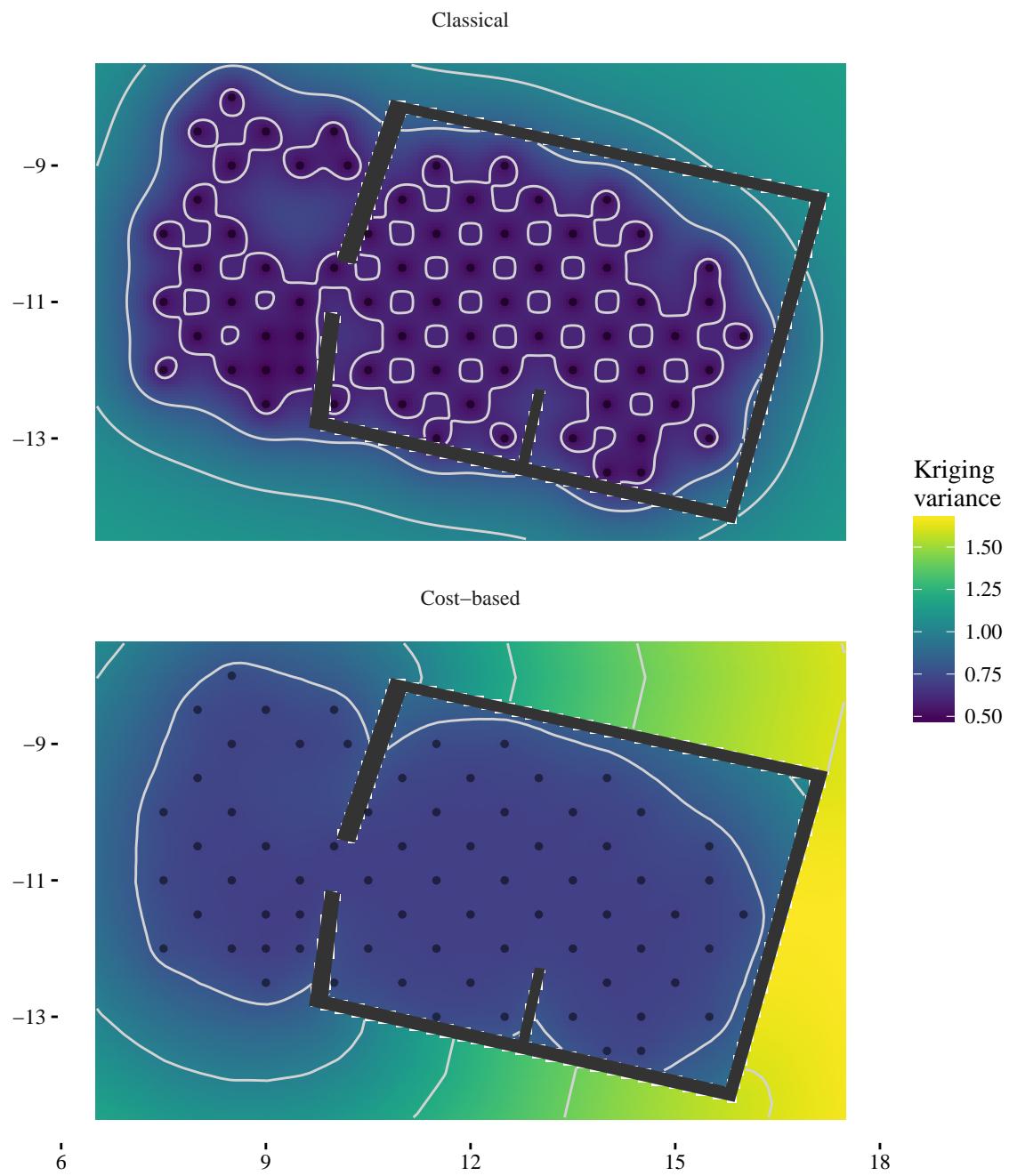


Figure 13: Comparison of prediction error by method.

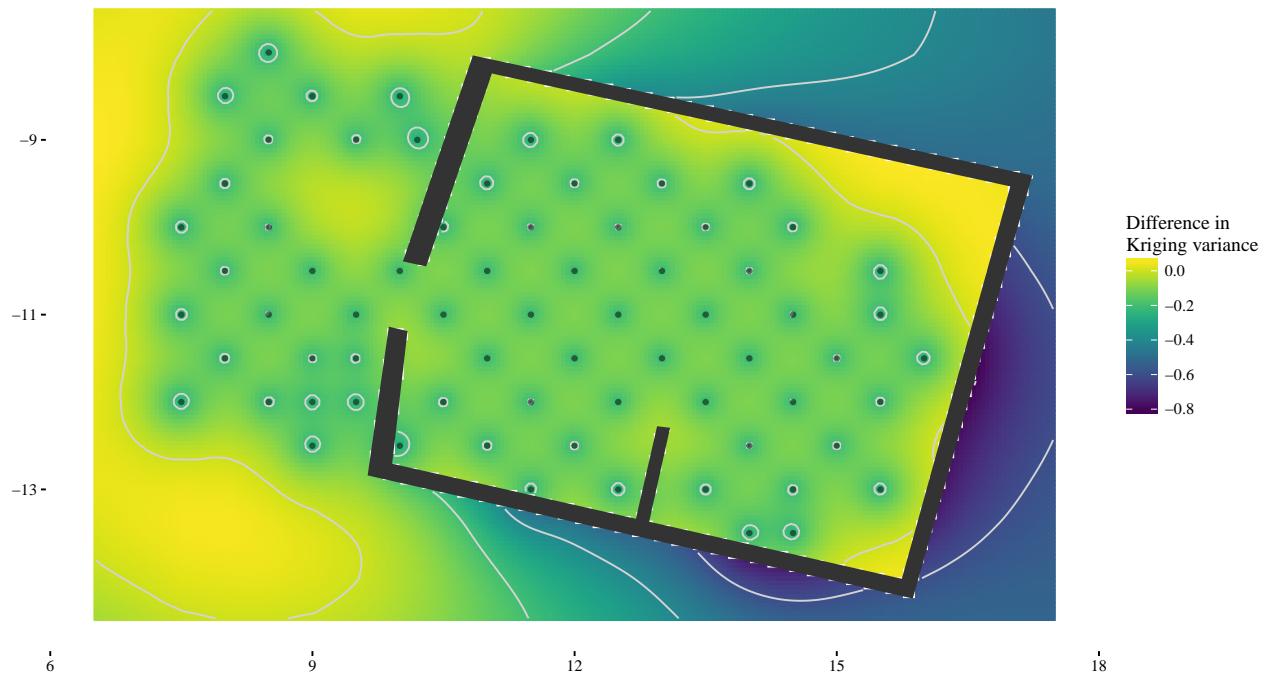


Figure 14: Difference between the Euclidean and the cost-based prediction errors

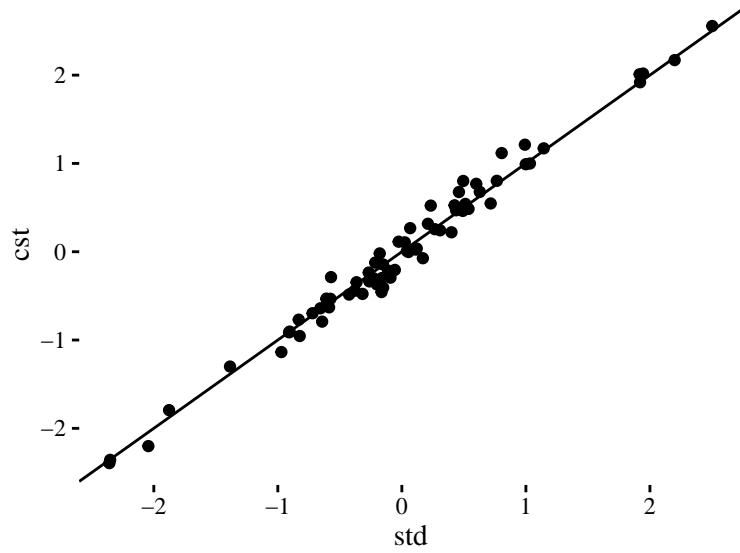


Figure 15: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	0.93
cst	0.96

## 4 Analysis of Copper

### 4.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

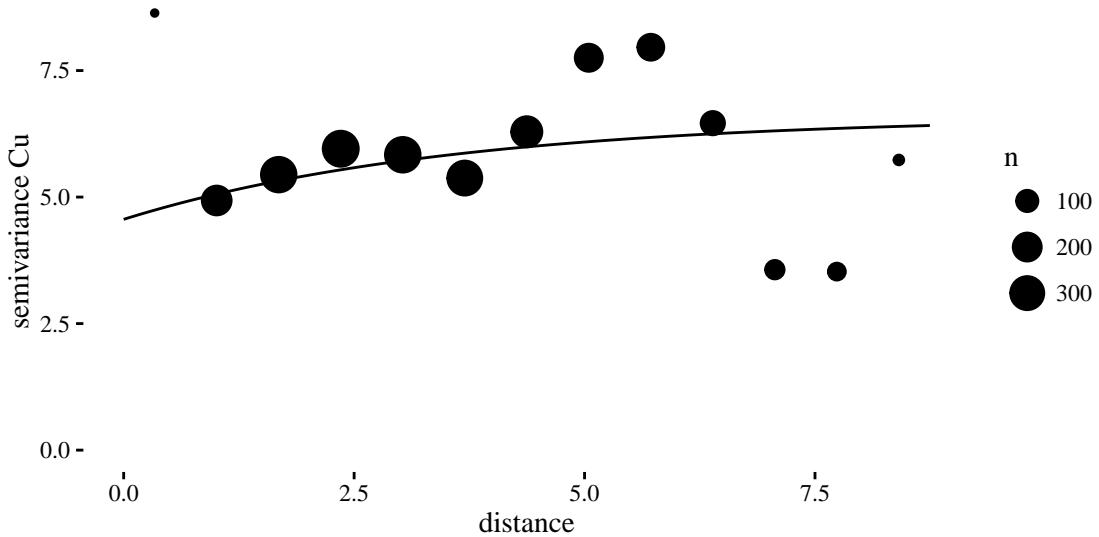


Figure 16: Empirical variogram and fitted model for Copper.

### 4.2 Cost-based kriging

### 4.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	12.13	12.14
Nugget	4.56	4.58
Partial sill	2.03	1.98
kappa	0.51	0.51
phi	3.58	3.81
Pract. range	10.73	11.43

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 4.4 Leave-one-out Cross Validation (LOOCV)

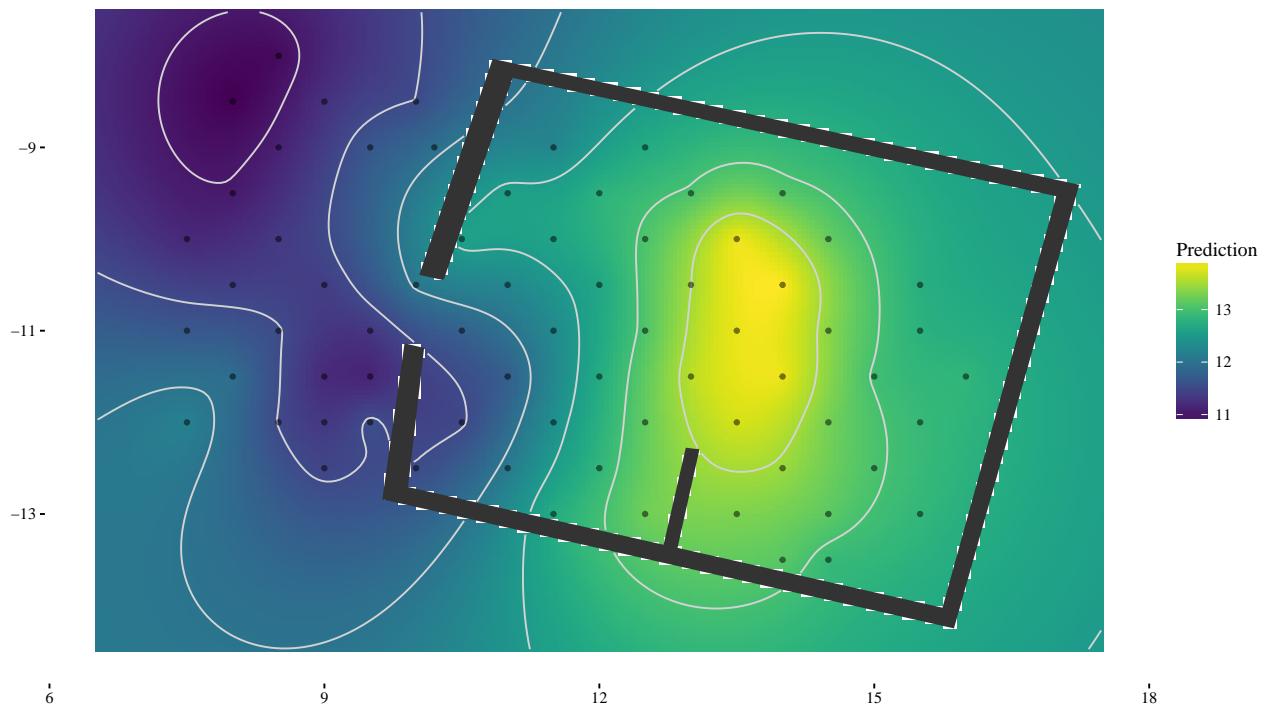


Figure 17: Euclidean kriging prediction for Copper.

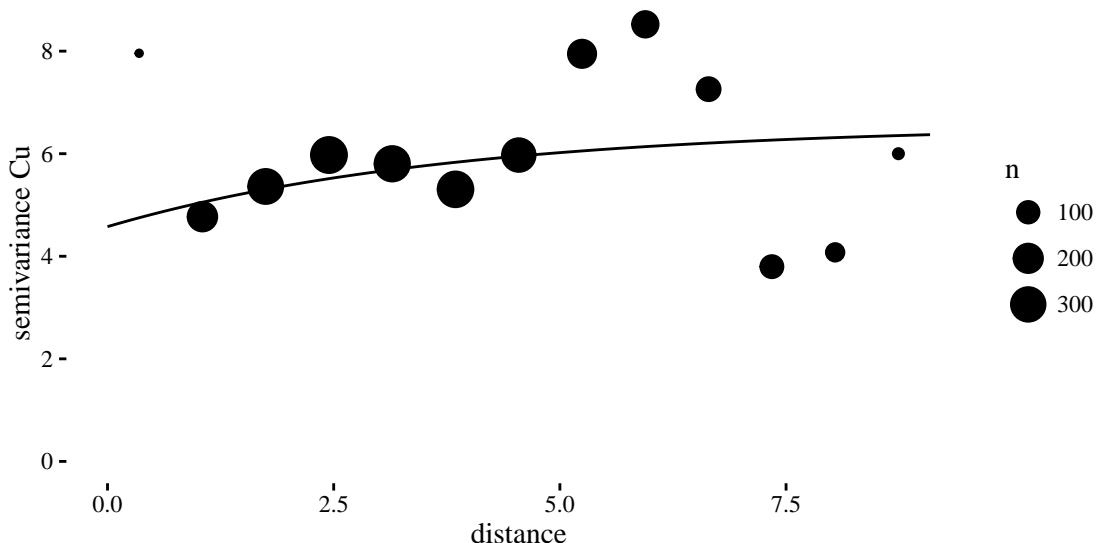


Figure 18: Empirical cost-based variogram and fitted model.

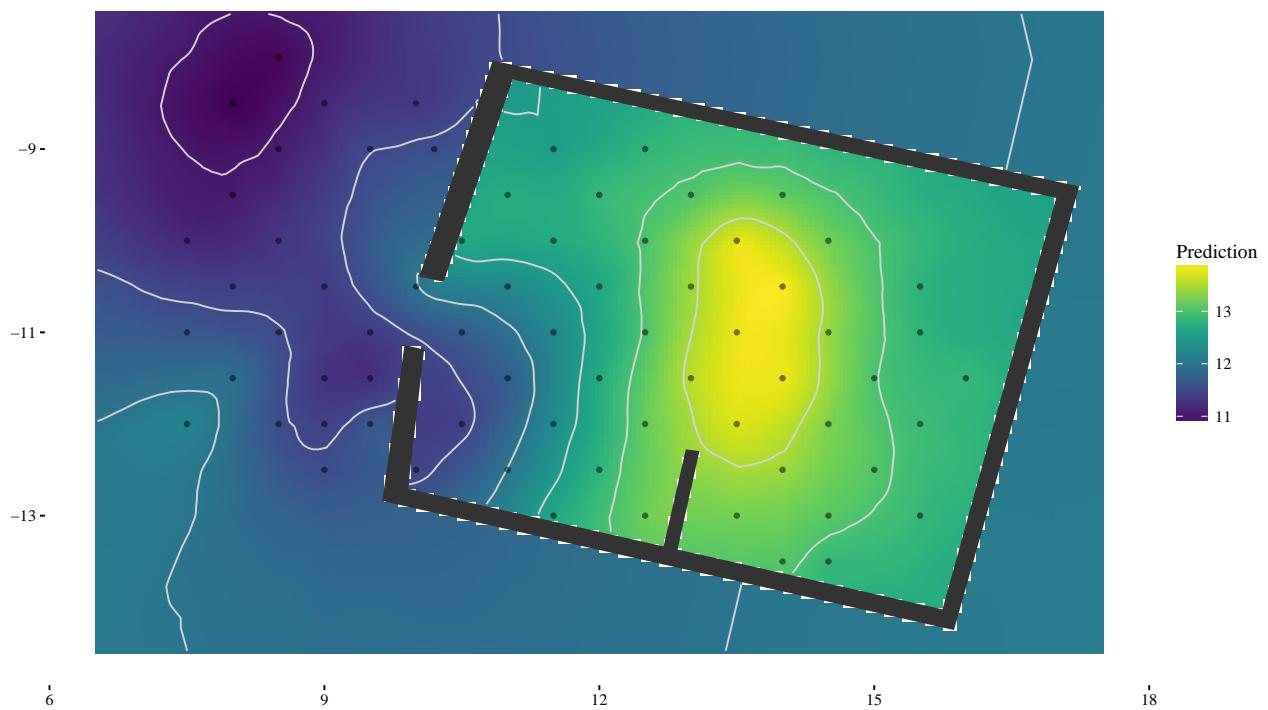


Figure 19: Cost-based kriging prediction

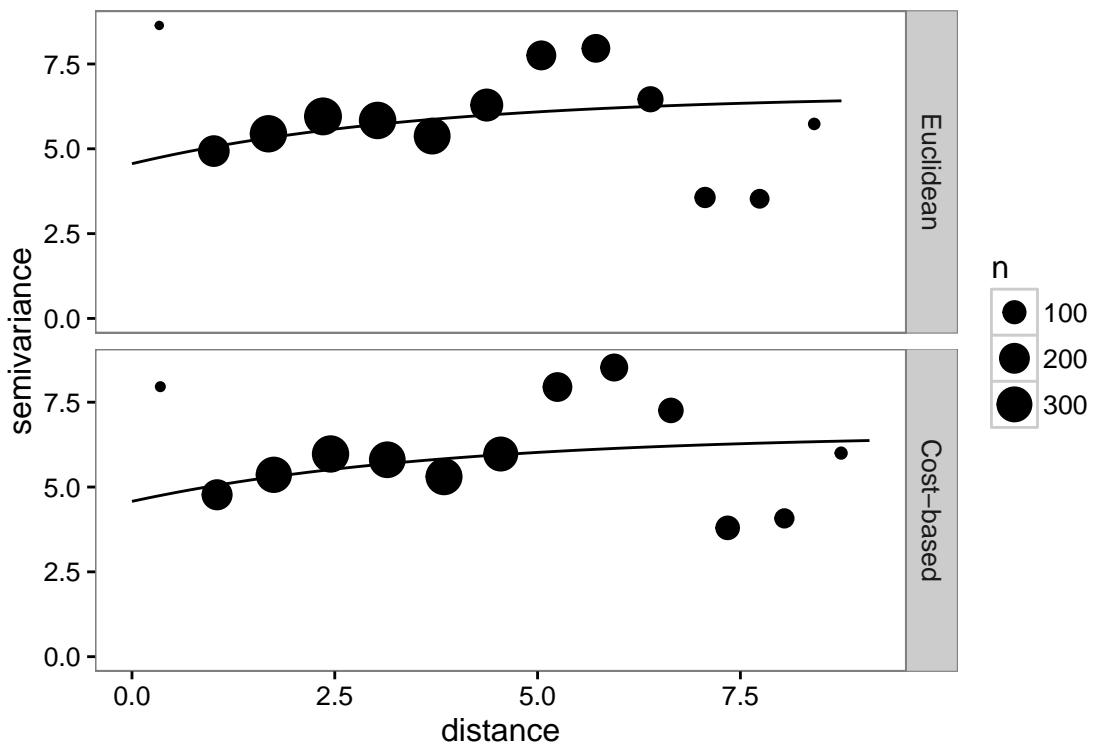


Figure 20: Empirical variogram and fitted models by method for Copper.

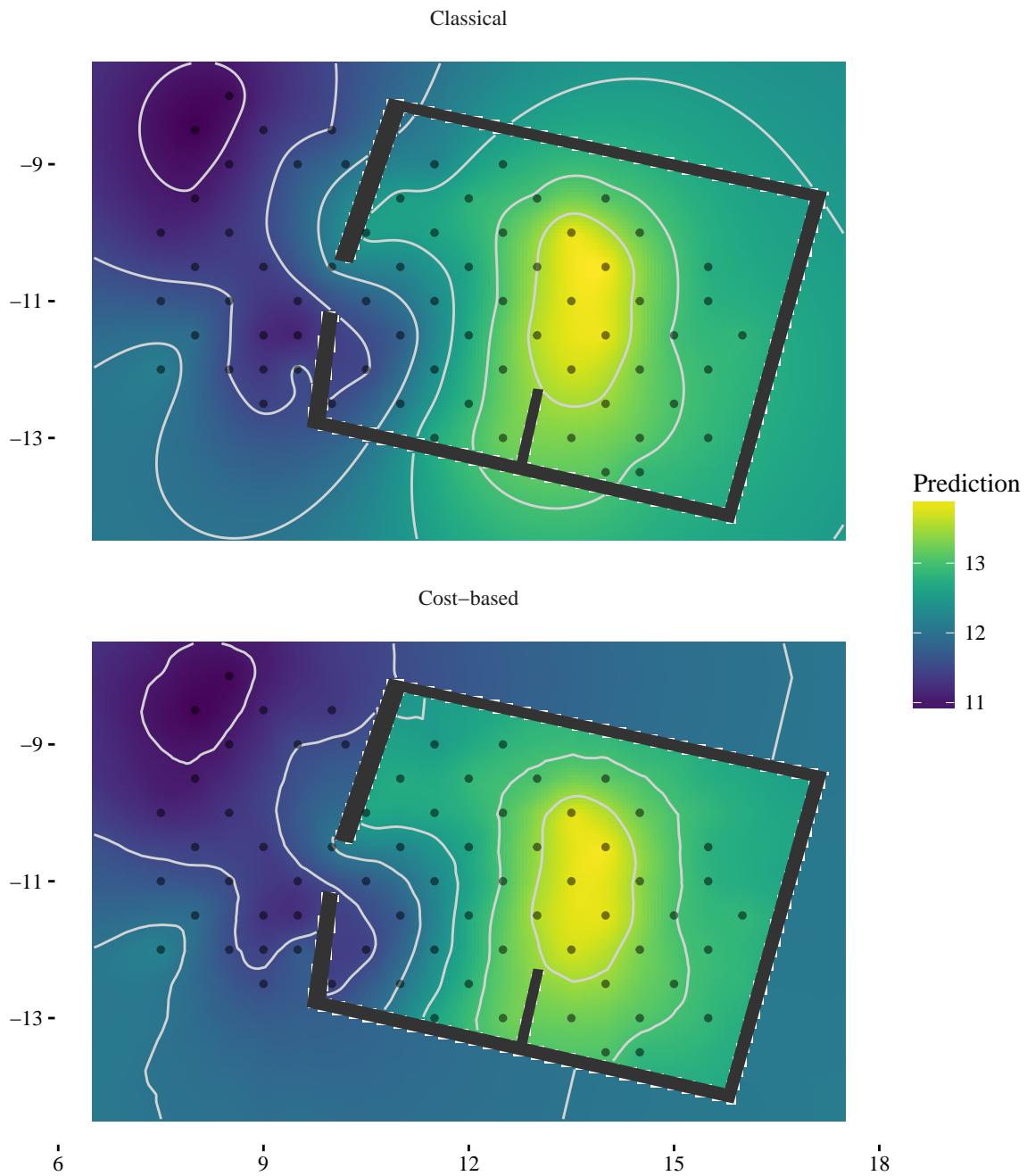


Figure 21: Comparison of Kriging estimates.

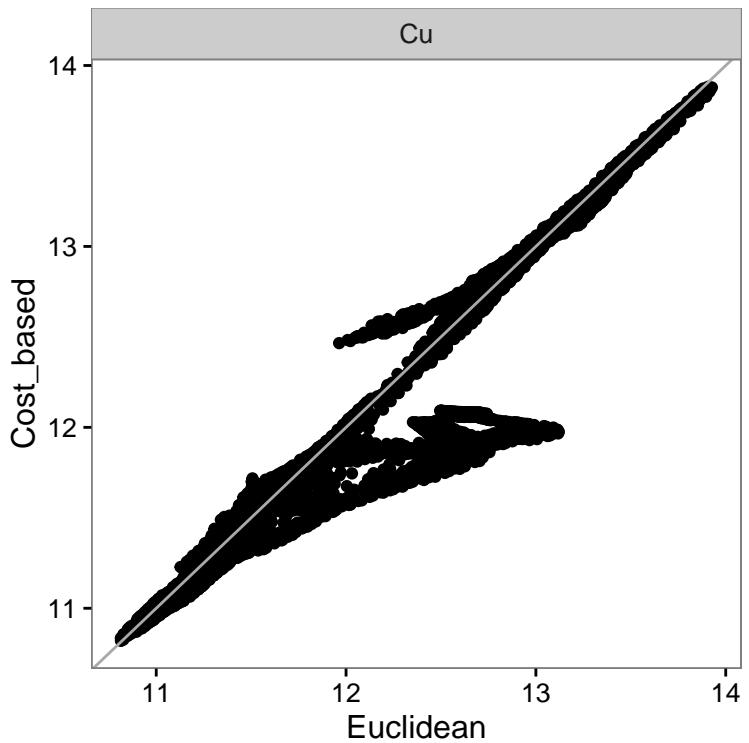


Figure 22: Pointwise comparison of predictions by method.

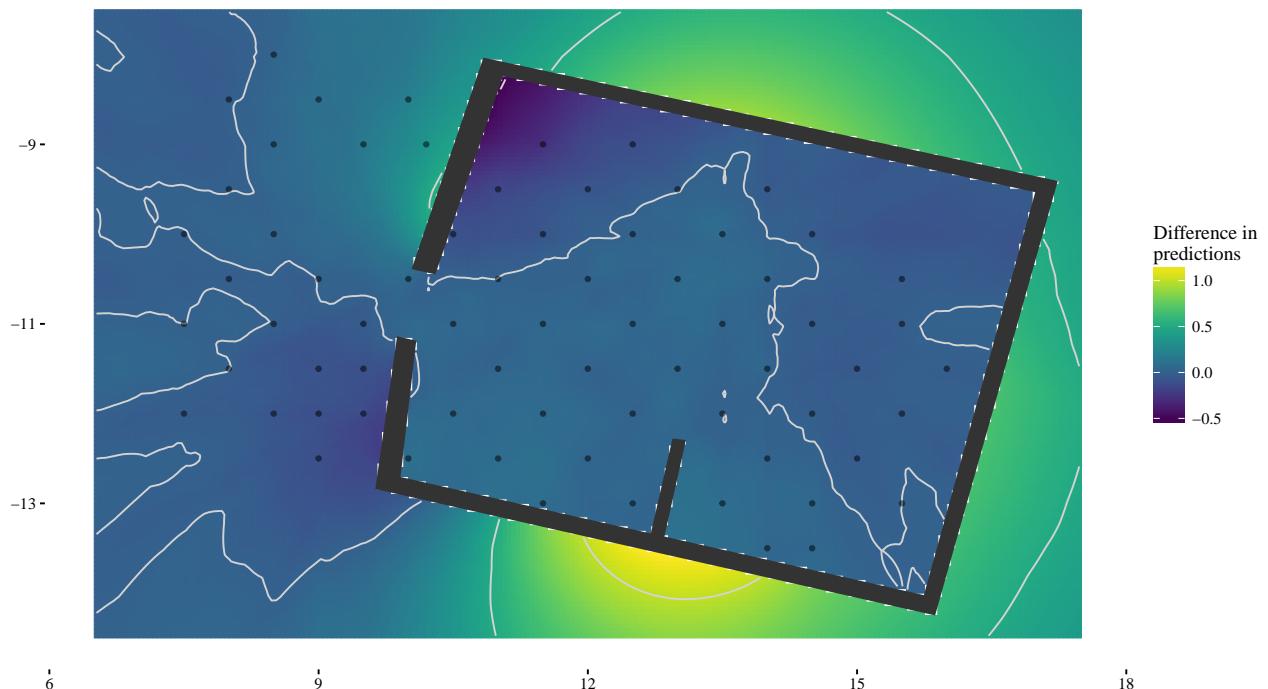


Figure 23: Difference between the Euclidean and the cost-based predictions.

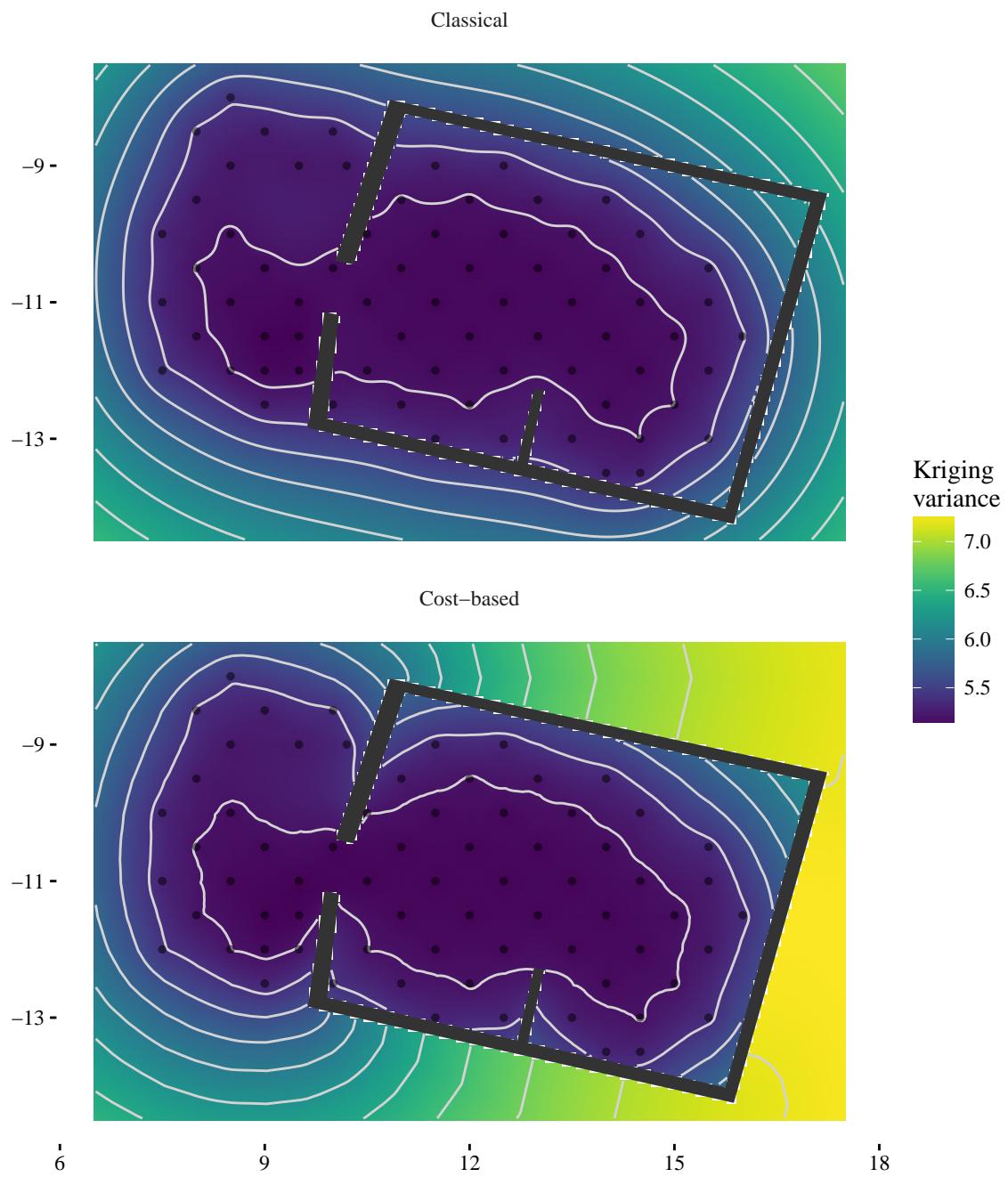


Figure 24: Comparison of prediction error by method.

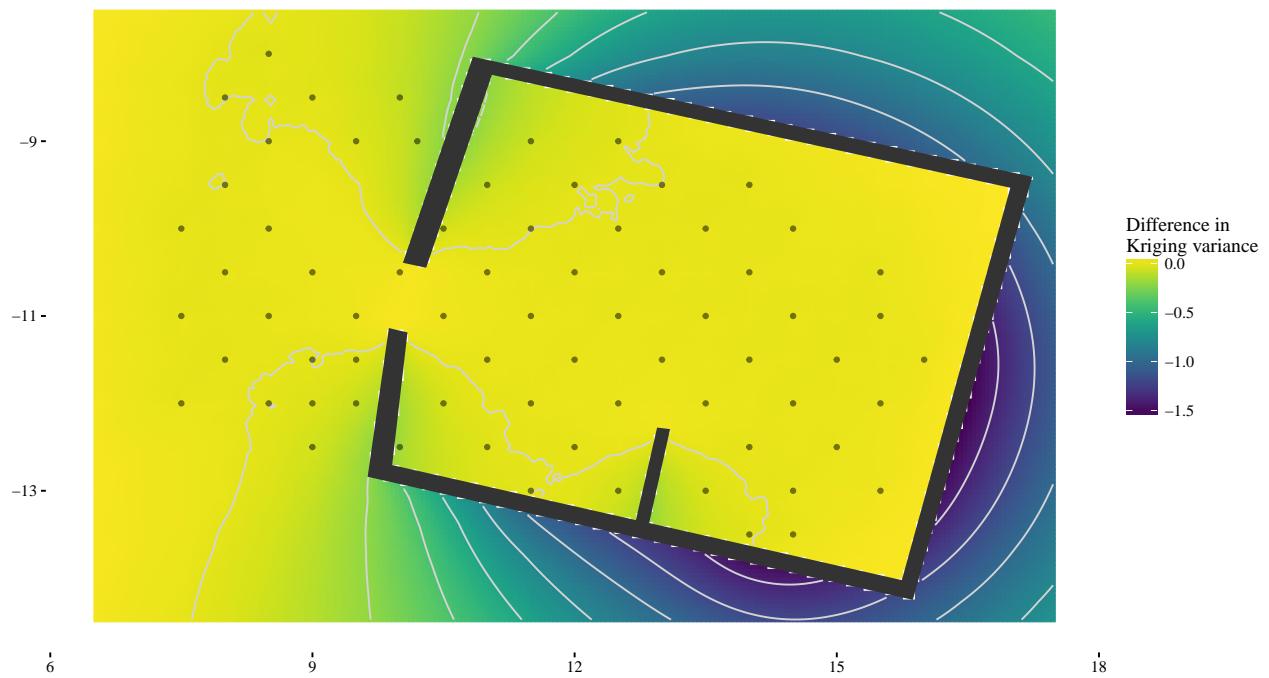


Figure 25: Difference between the Euclidean and the cost-based prediction errors

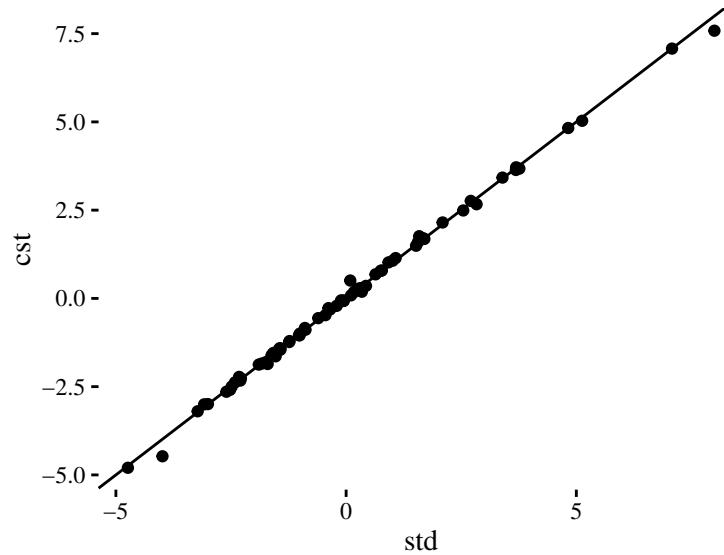


Figure 26: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	2.42
cst	2.41

## 5 Analysis of Ferrum

### 5.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

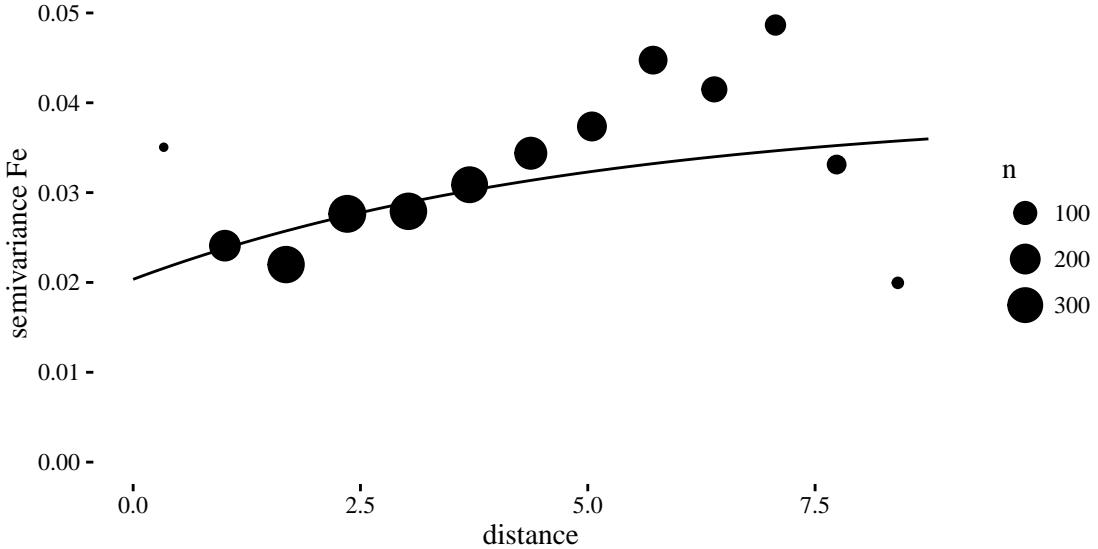


Figure 27: Empirical variogram and fitted model.

### 5.2 Cost-based kriging

### 5.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	1.23	1.23
Nugget	0.02	0.02
Partial sill	0.02	0.02
phi	5.04	4.23
Pract. range	15.09	12.67
Log-likelihood	28.20	28.02

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 5.4 Leave-one-out Cross Validation (LOOCV)

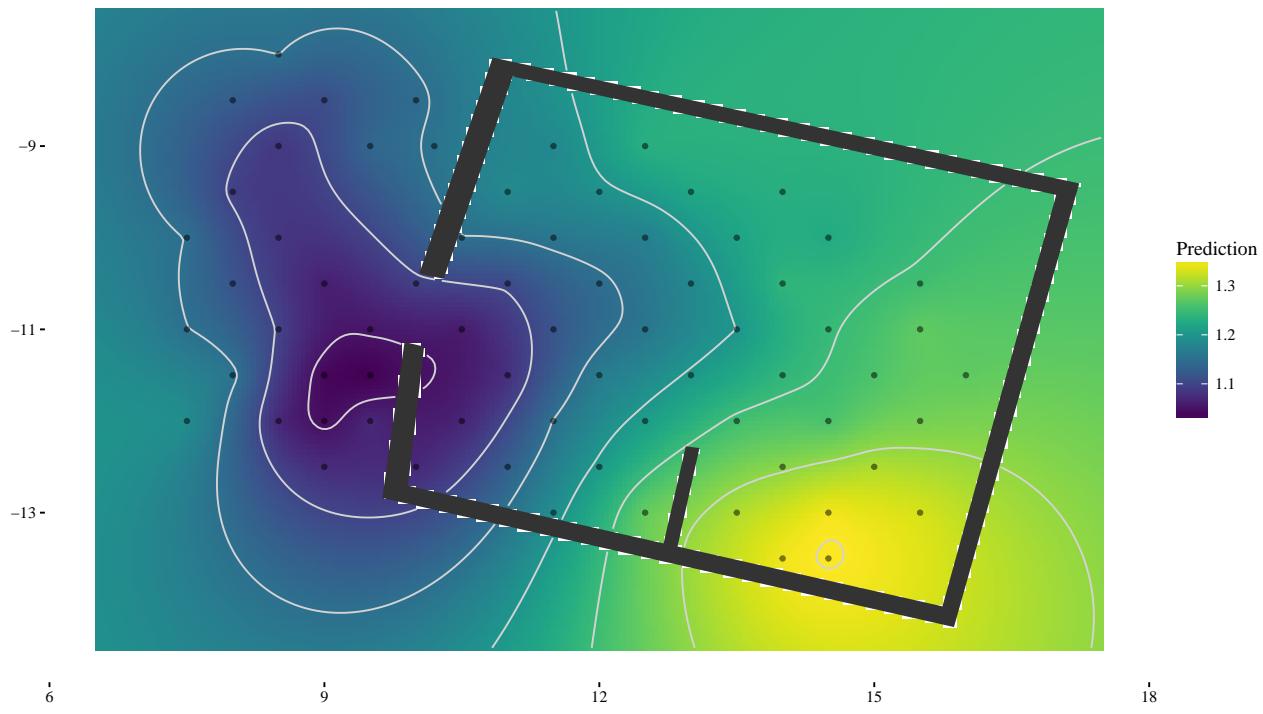


Figure 28: Euclidean kriging prediction

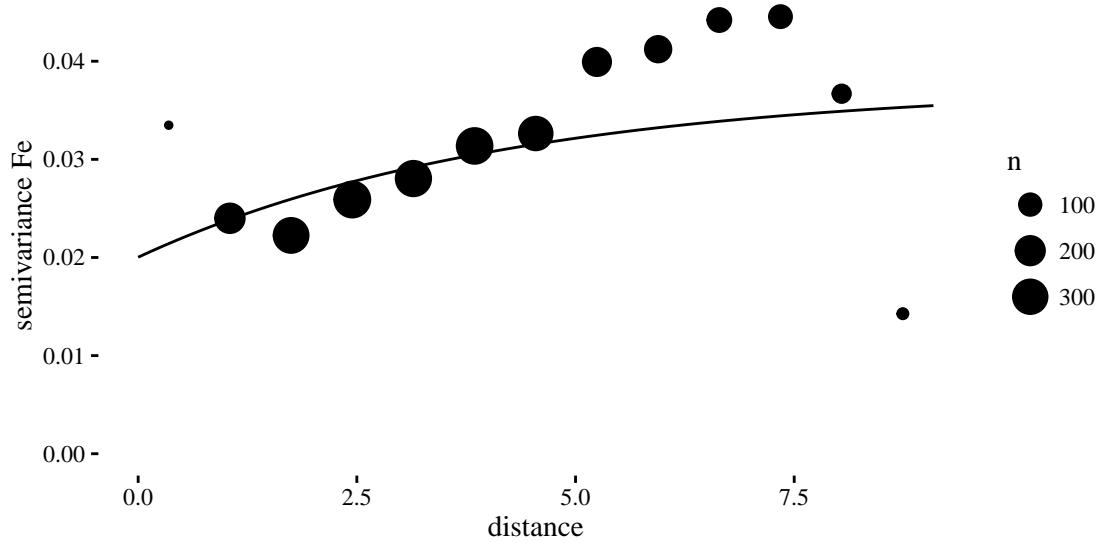


Figure 29: Empirical cost-based variogram and fitted model.

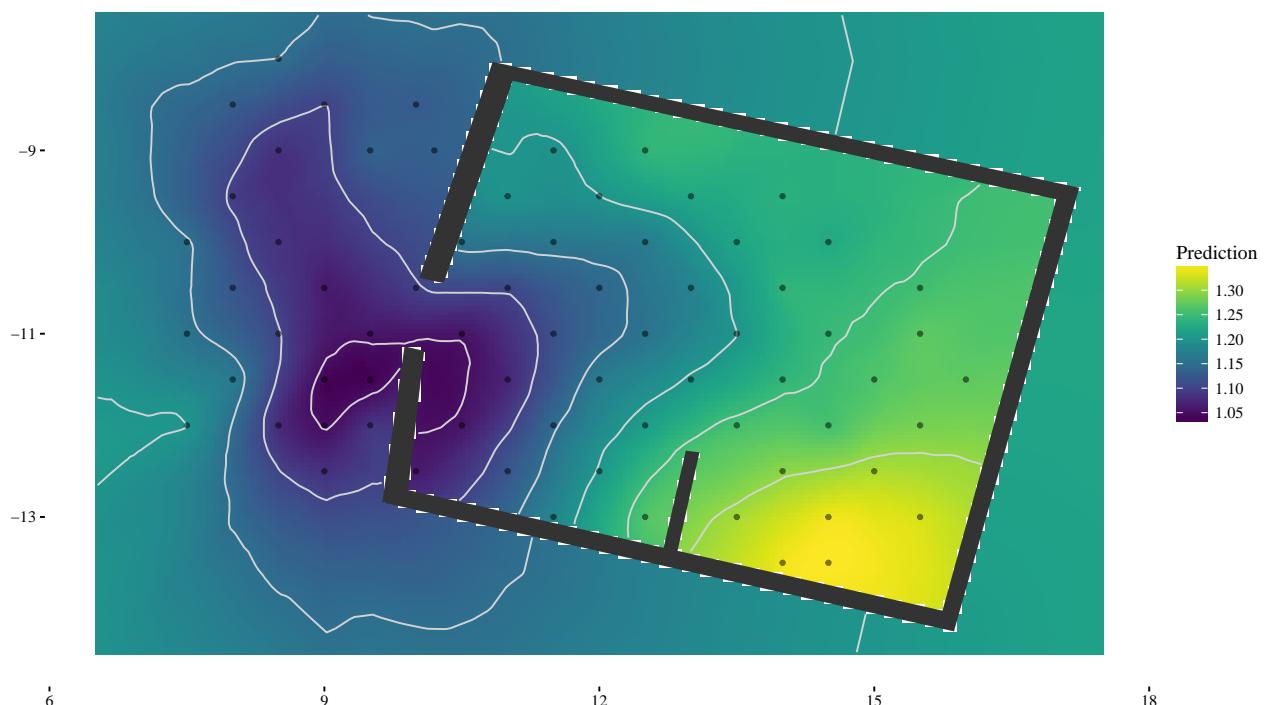


Figure 30: Cost-based kriging prediction

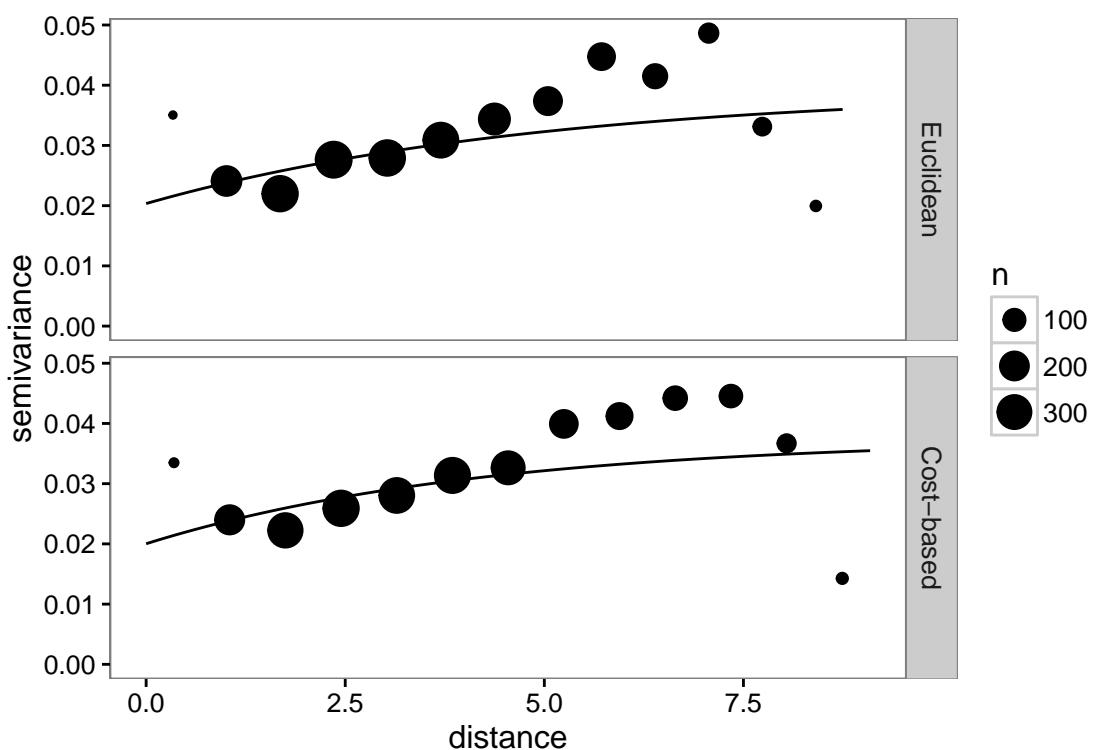


Figure 31: Empirical variogram and fitted models by method for Ferrum.

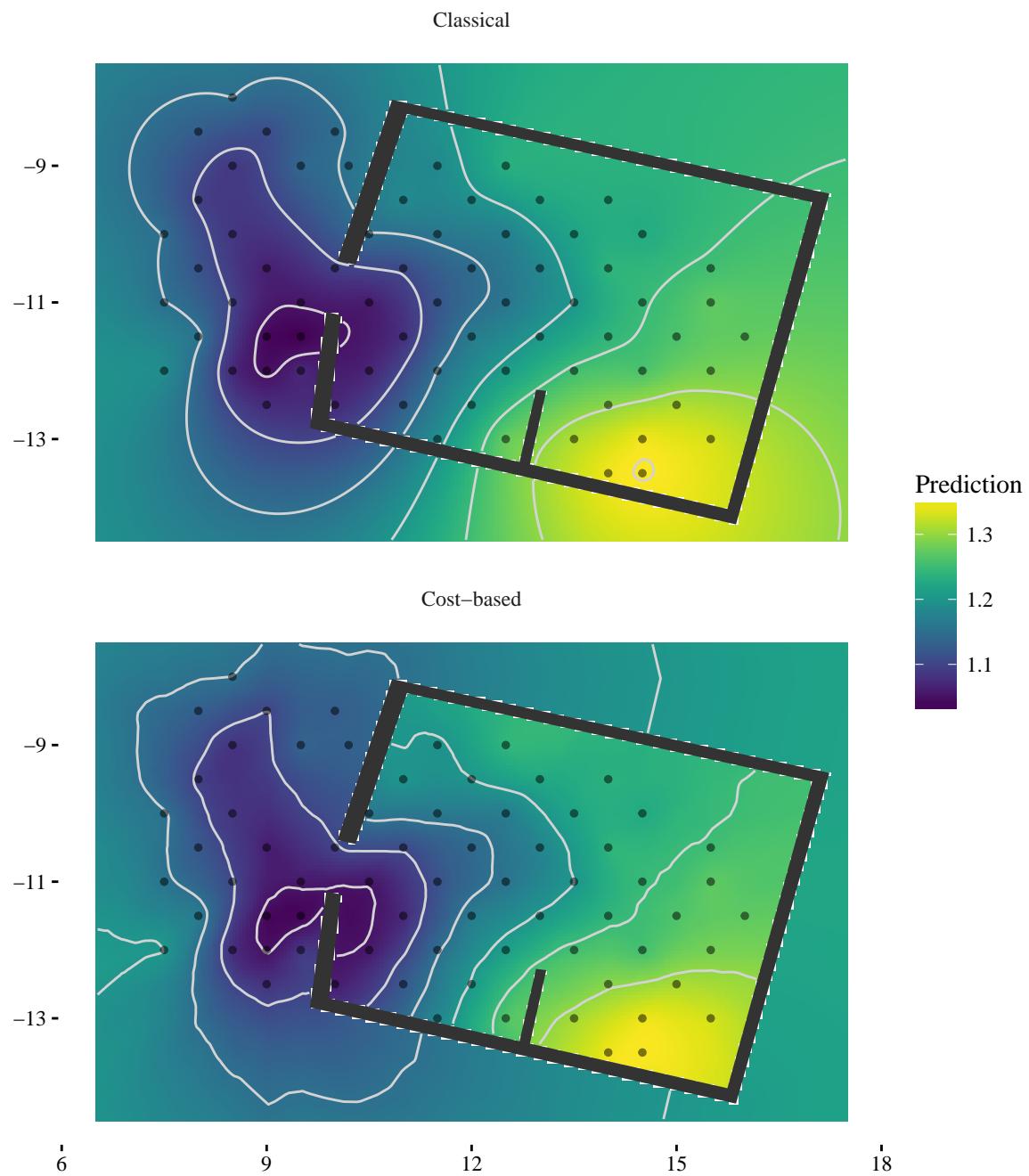


Figure 32: Comparison of Kriging estimates.

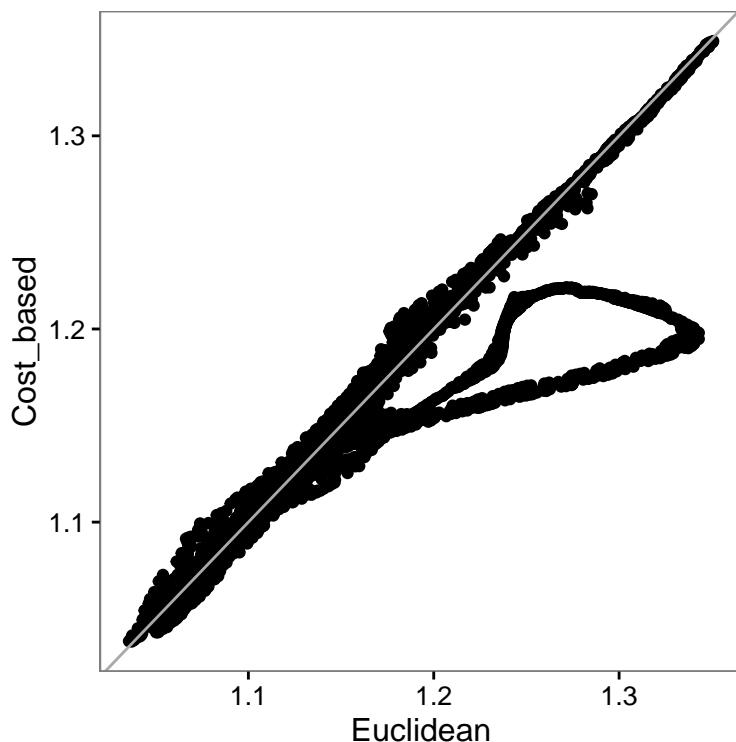


Figure 33: Pointwise comparison of predictions by method.

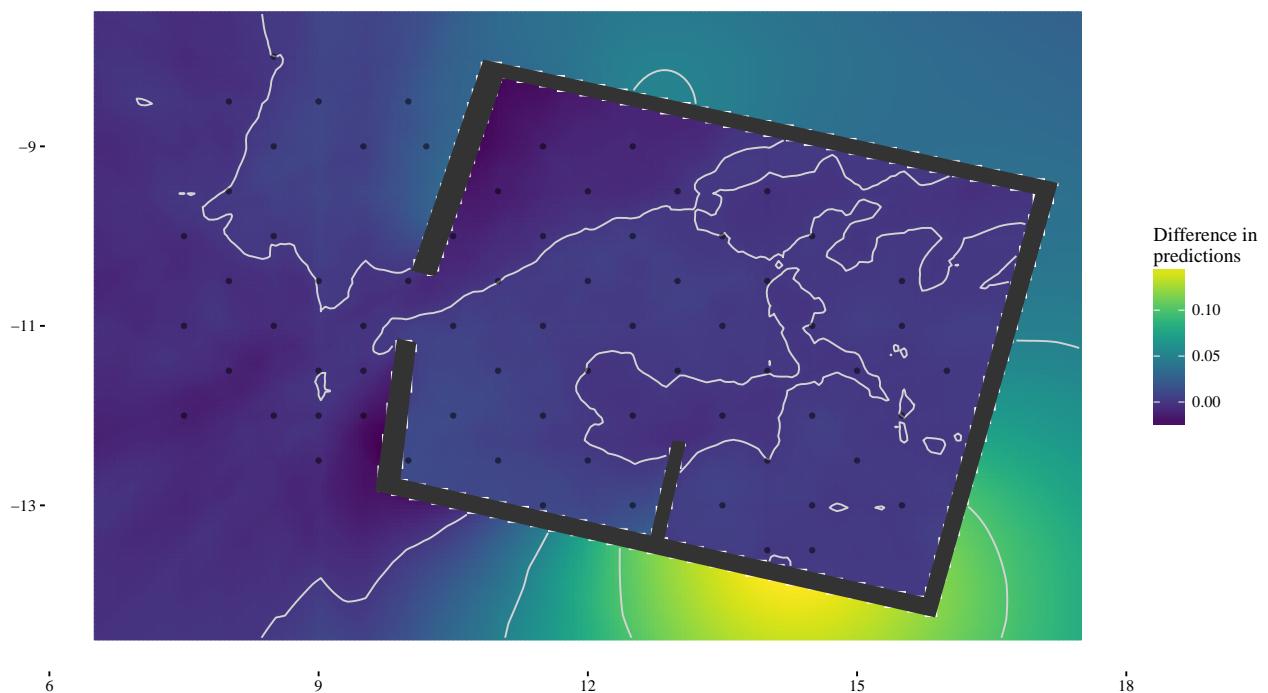


Figure 34: Difference between the Euclidean and the cost-based predictions.

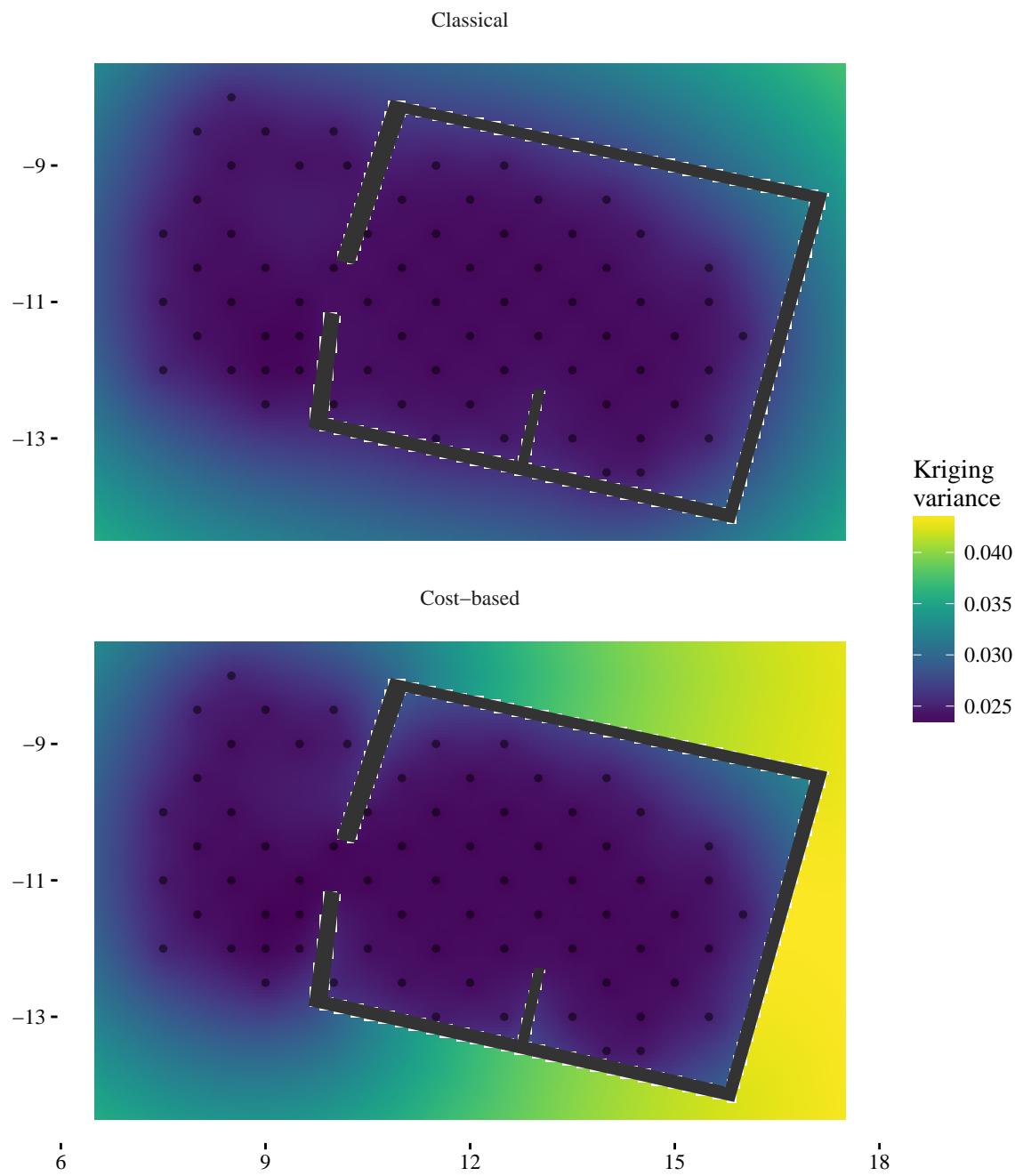


Figure 35: Comparison of prediction error by method.

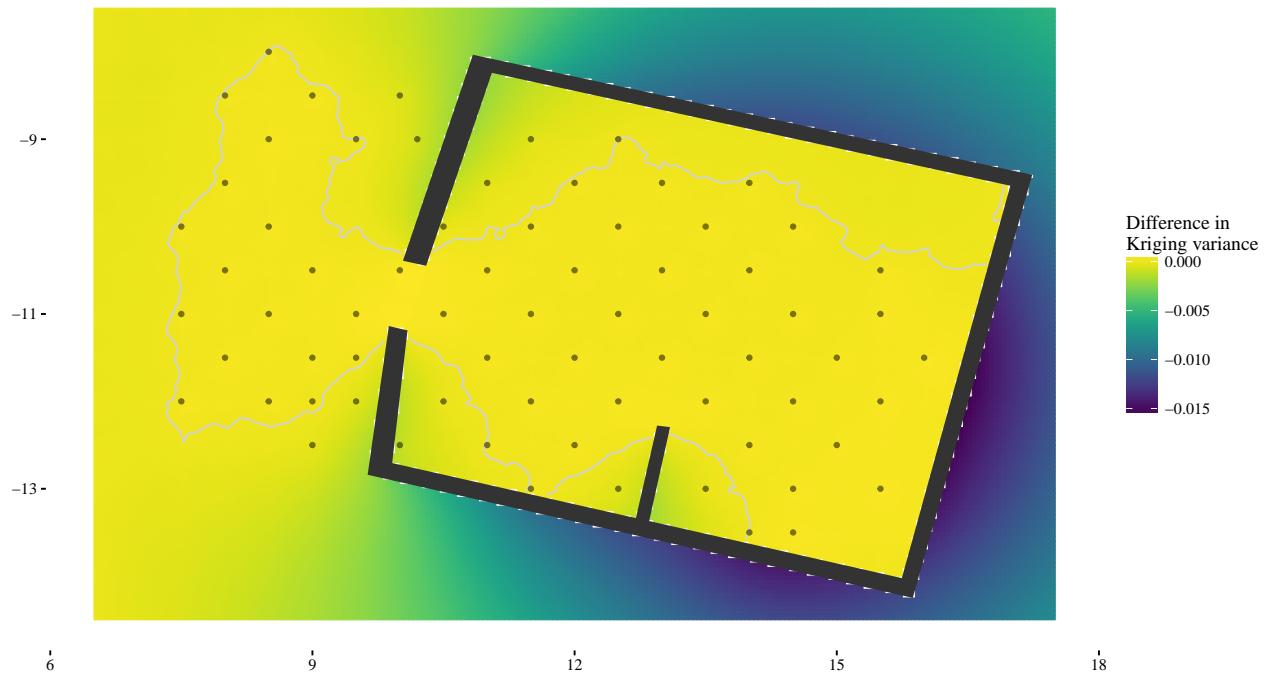


Figure 36: Difference between the Euclidean and the cost-based prediction errors

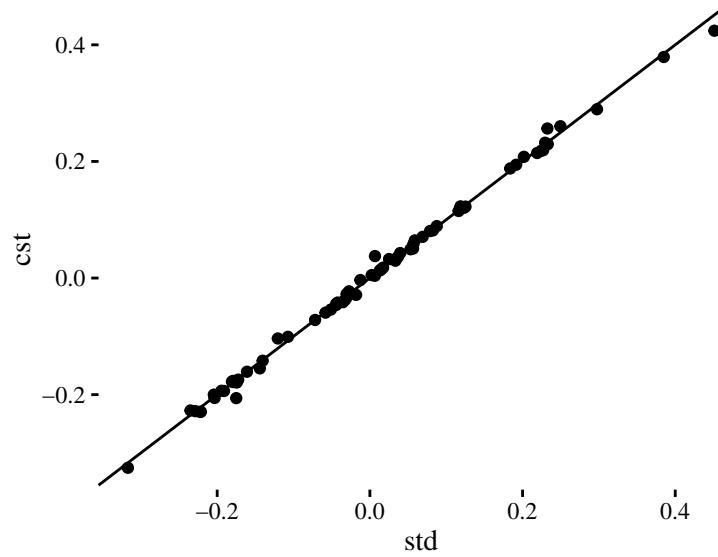


Figure 37: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	0.16
cst	0.16

## 6 Analysis of Potassium

### 6.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

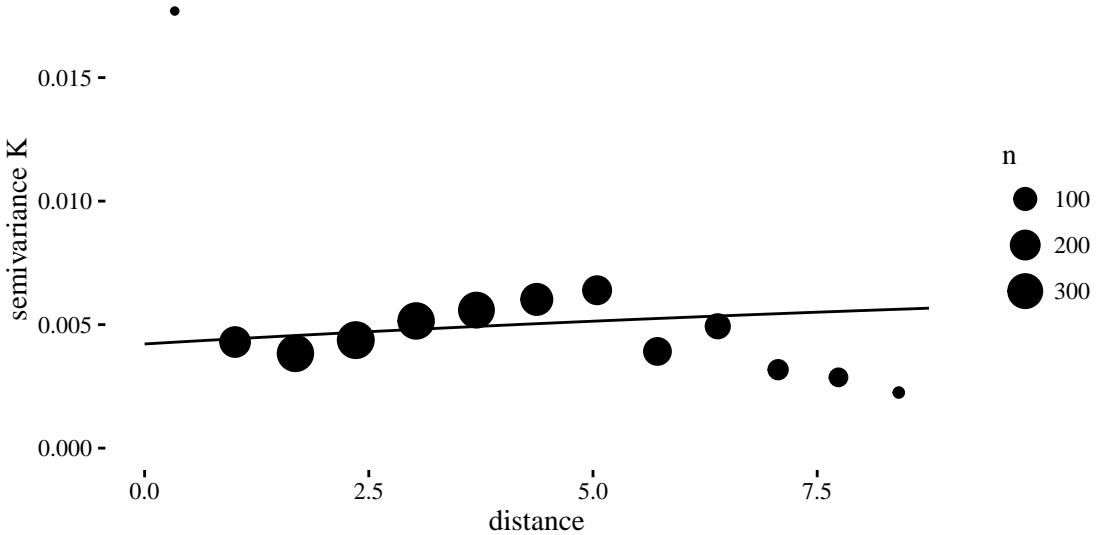


Figure 38: Empirical variogram and fitted model.

### 6.2 Cost-based kriging

### 6.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	0.25	0.25
Nugget	0.00	0.00
Partial sill	0.00	0.00
phi	16.23	15.13
Pract. range	48.61	45.32
Log-likelihood	87.28	86.98

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 6.4 Leave-one-out Cross Validation (LOOCV)

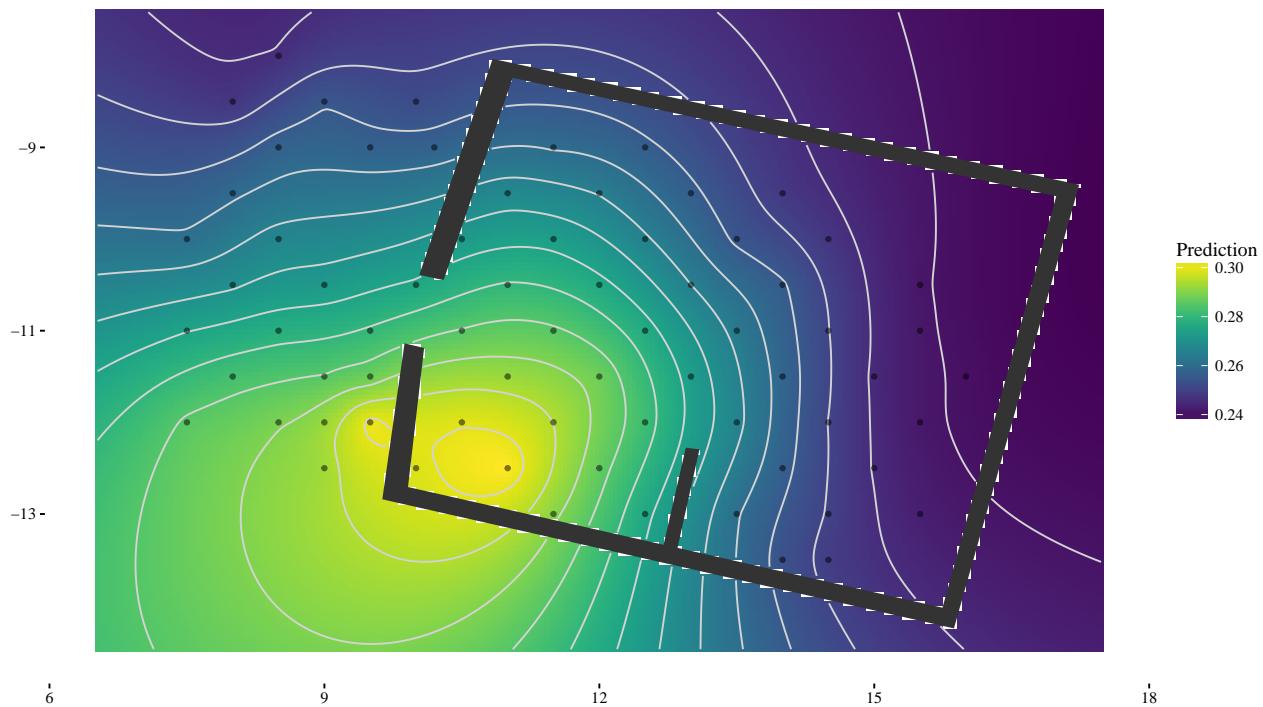


Figure 39: Euclidean kriging prediction

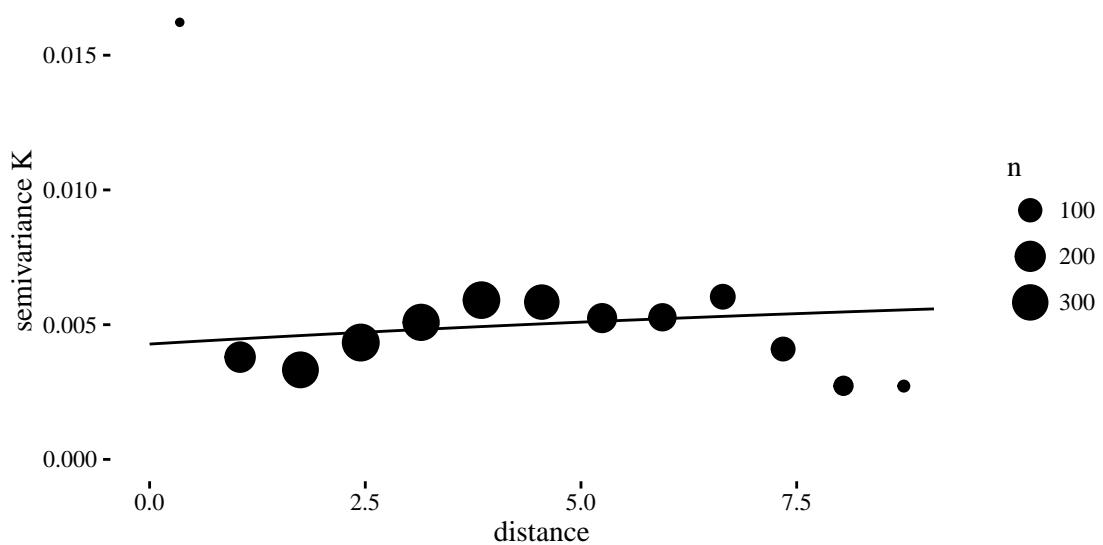


Figure 40: Empirical cost-based variogram and fitted model.

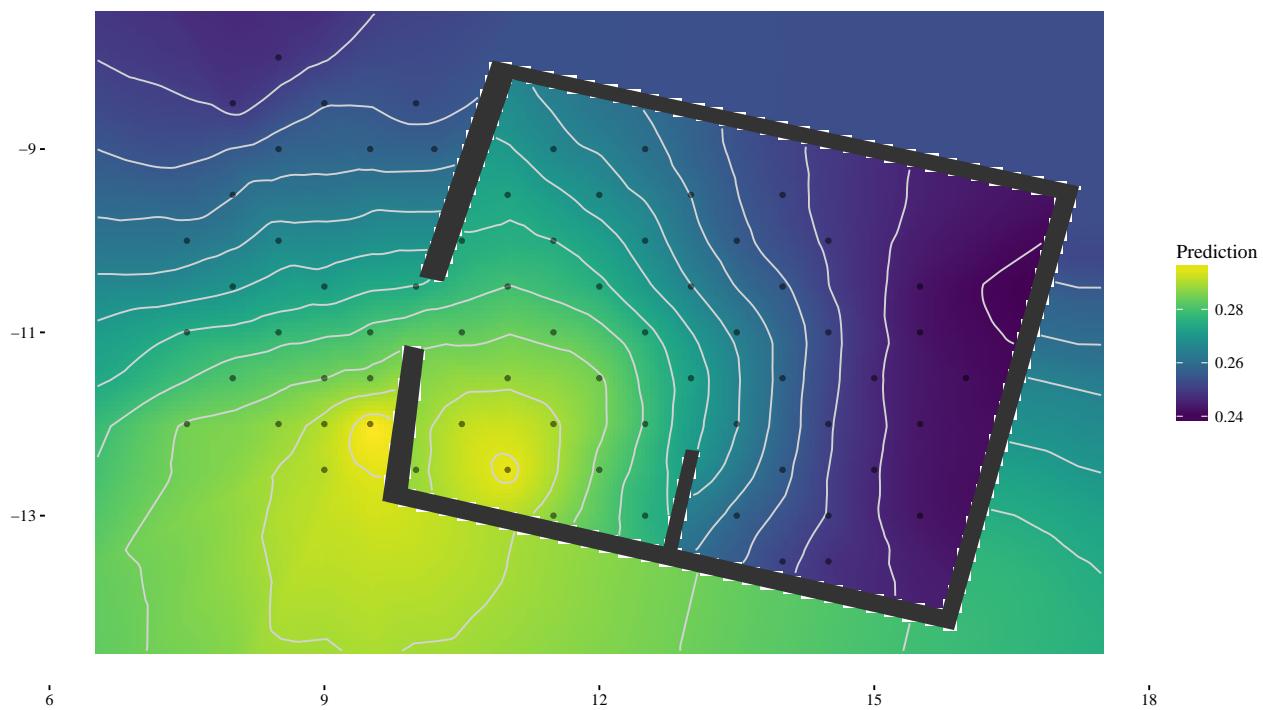


Figure 41: Cost-based kriging prediction

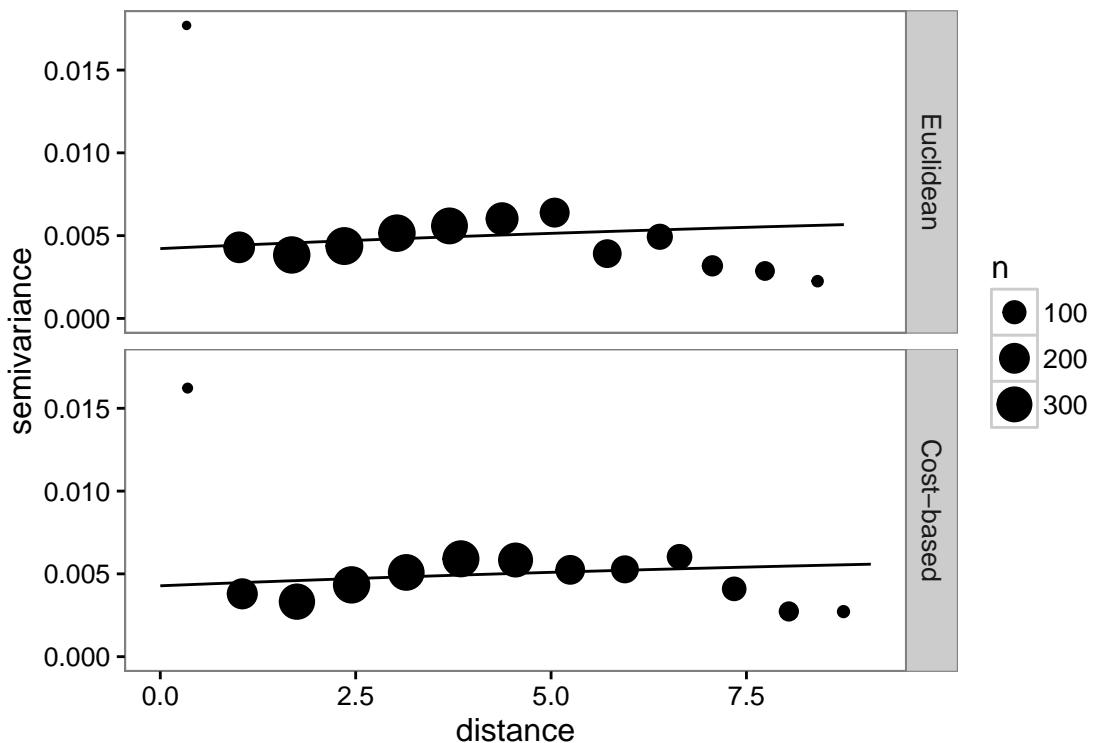
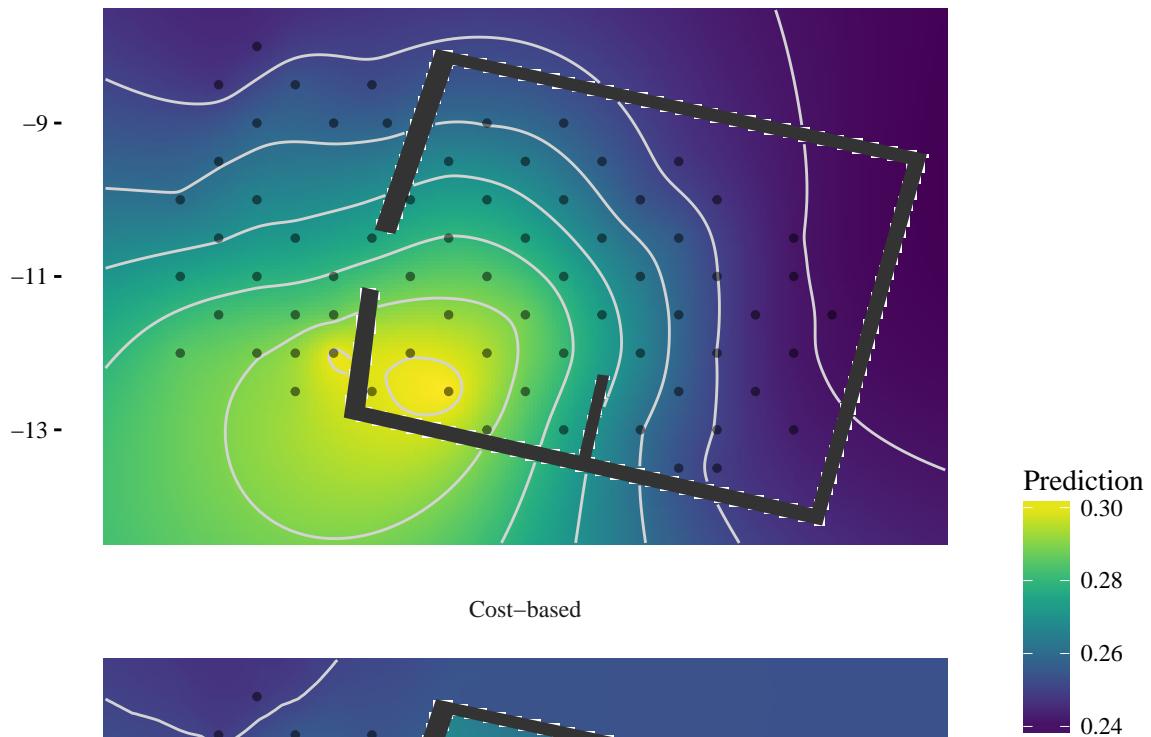


Figure 42: Empirical variogram and fitted models by method for Potassium.

Classical



Cost-based

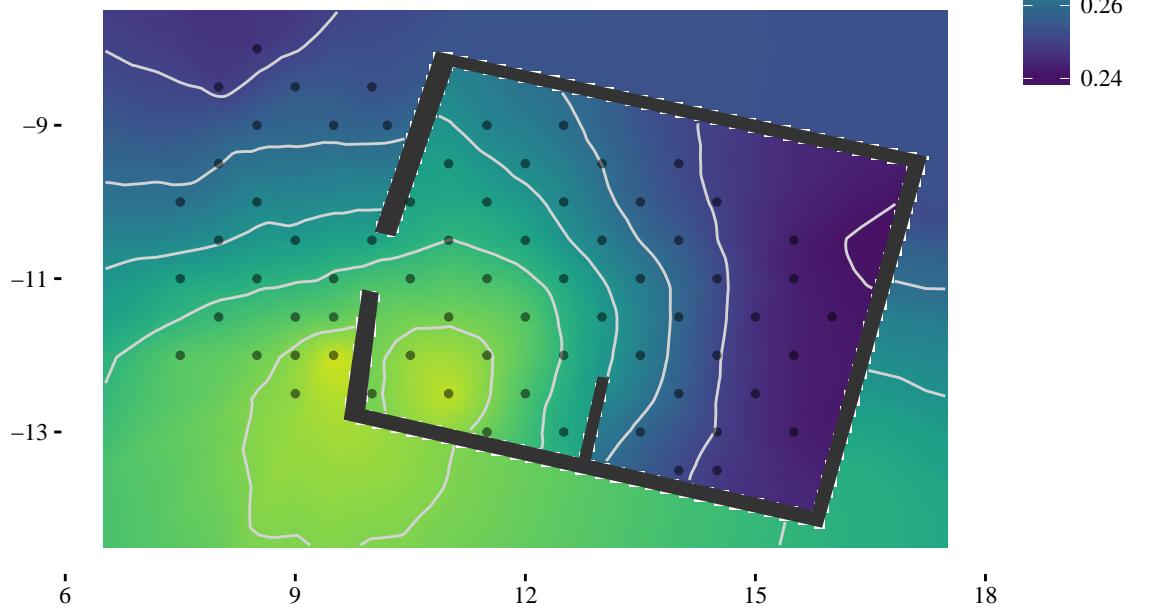


Figure 43: Comparison of Kriging estimates.

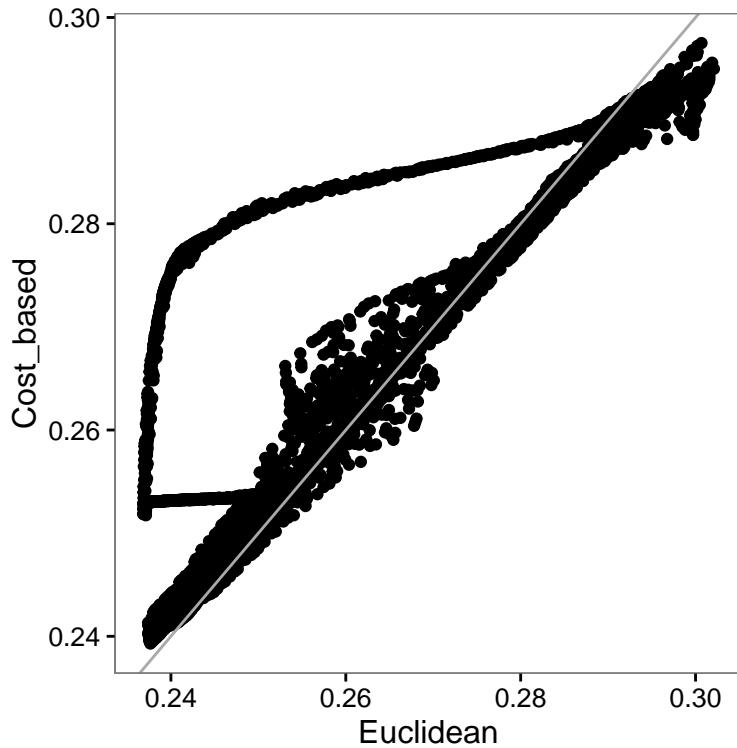


Figure 44: Pointwise comparison of predictions by method.

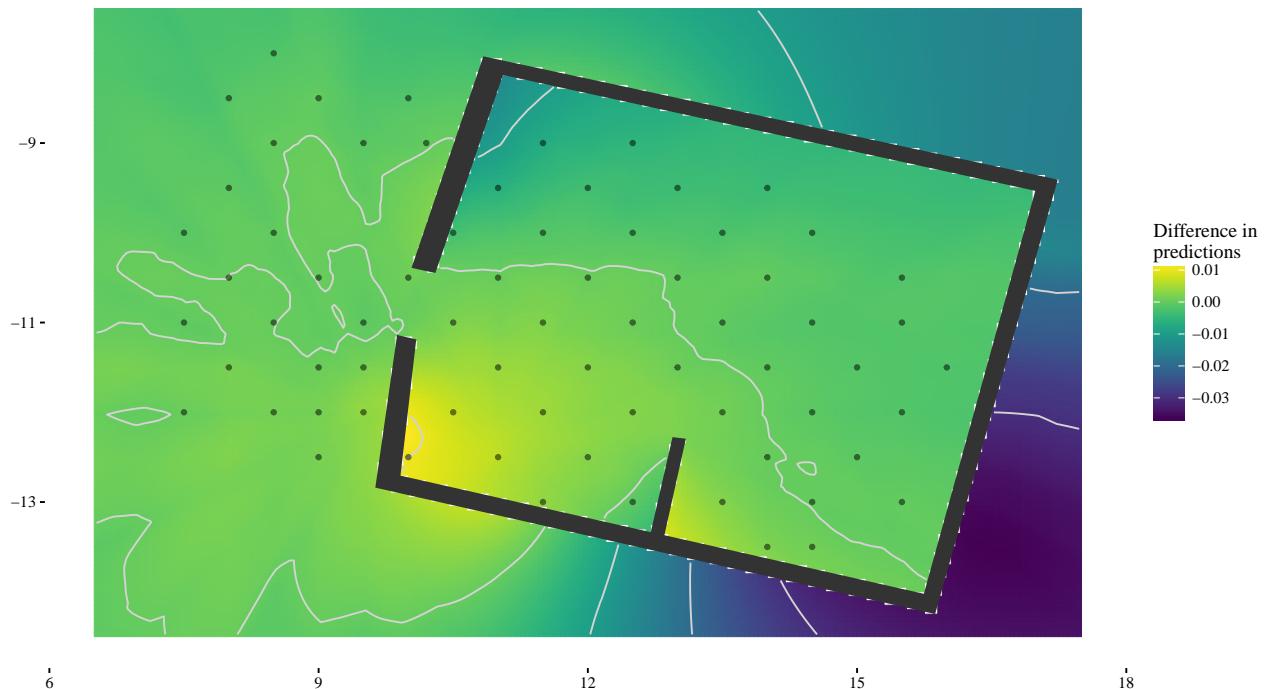


Figure 45: Difference between the Euclidean and the cost-based predictions.

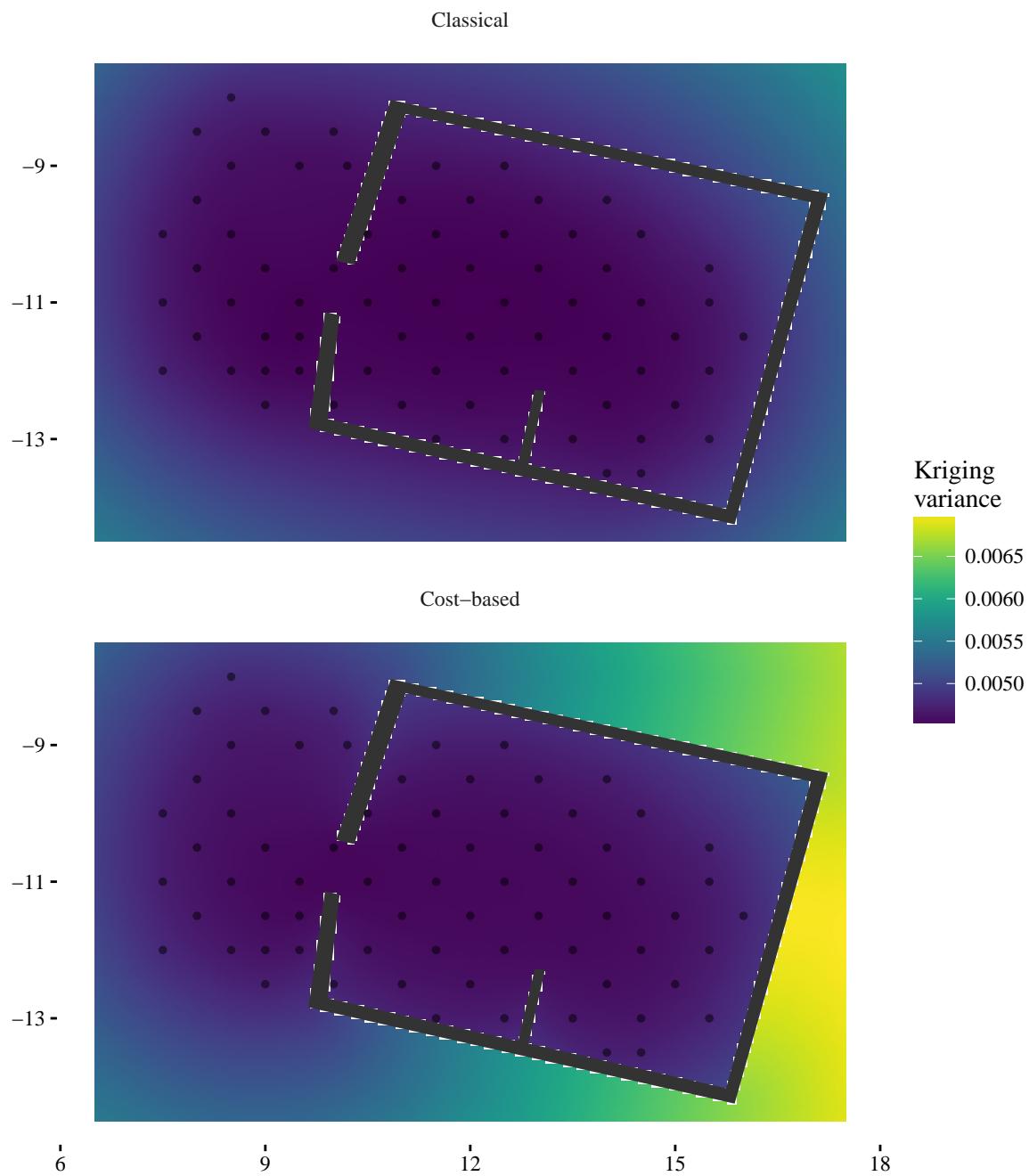


Figure 46: Comparison of prediction error by method.

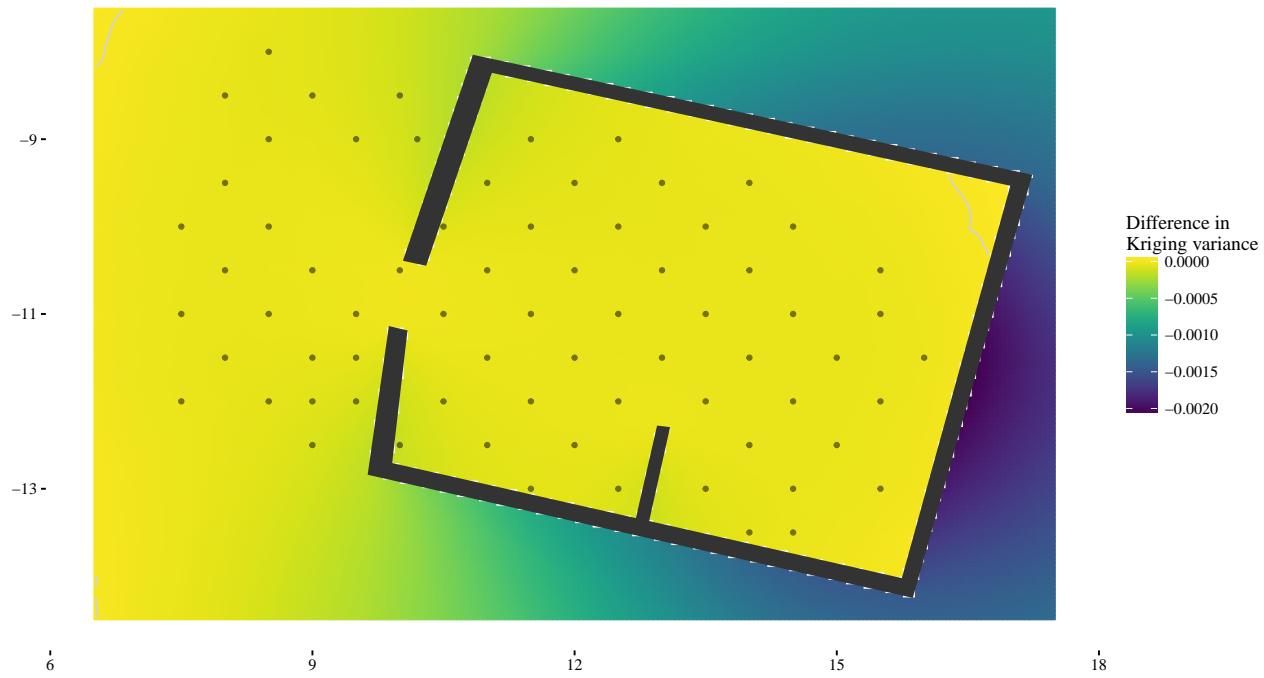


Figure 47: Difference between the Euclidean and the cost-based prediction errors

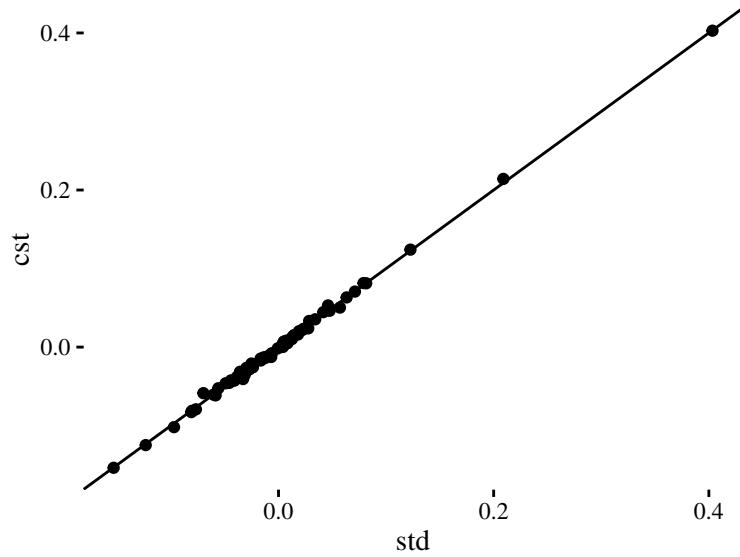


Figure 48: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	0.07
cst	0.07

## 7 Analysis of Magnesium

### 7.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

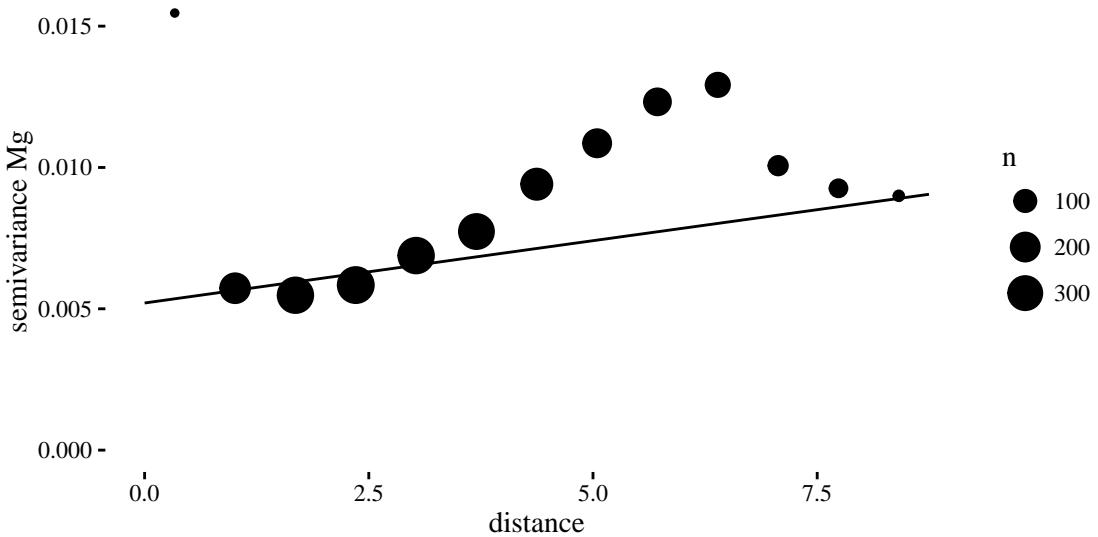


Figure 49: Empirical variogram and fitted model.

### 7.2 Cost-based kriging

### 7.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	0.45	0.45
Nugget	0.01	0.01
Partial sill	0.30	0.28
phi	668.30	706.67
Pract. range	2002.05	2116.99
Log-likelihood	78.19	78.02

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 7.4 Leave-one-out Cross Validation (LOOCV)

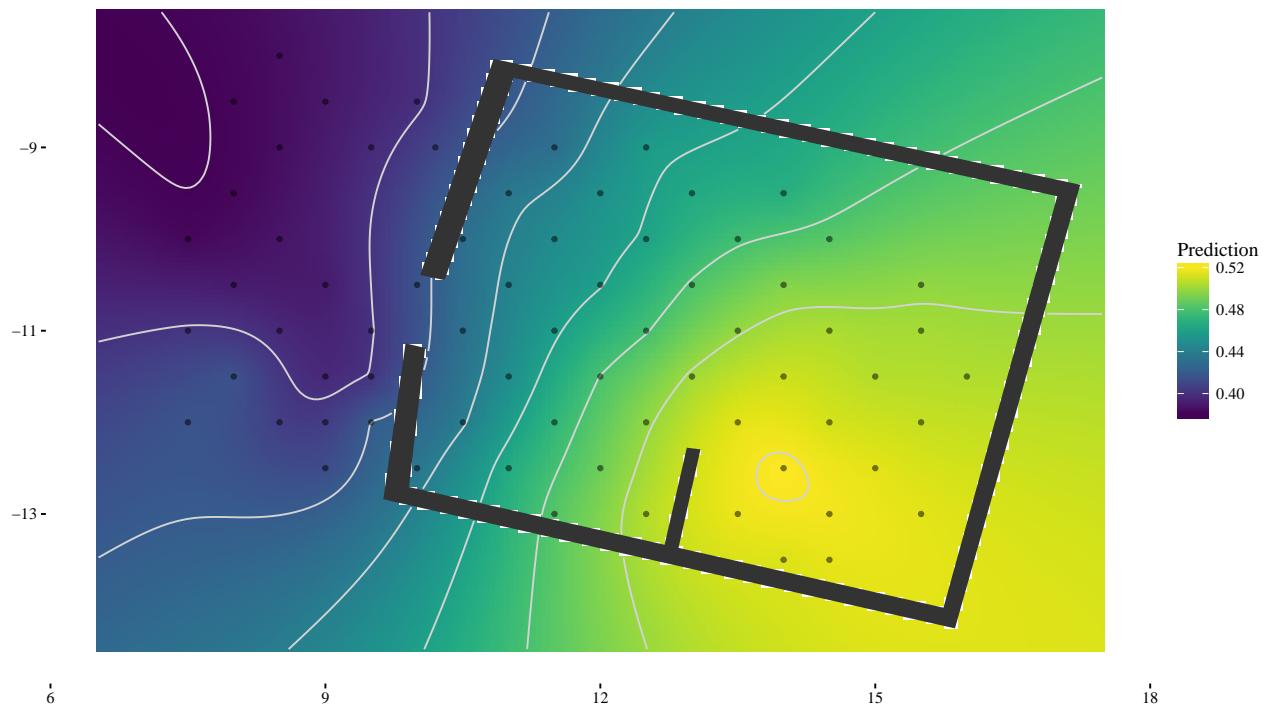


Figure 50: Euclidean kriging prediction

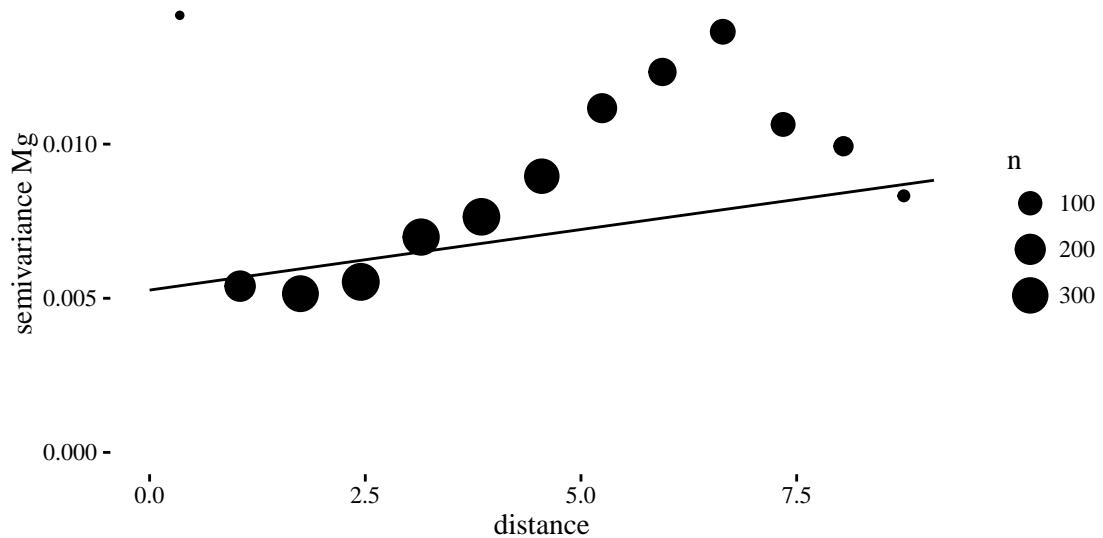


Figure 51: Empirical cost-based variogram and fitted model.

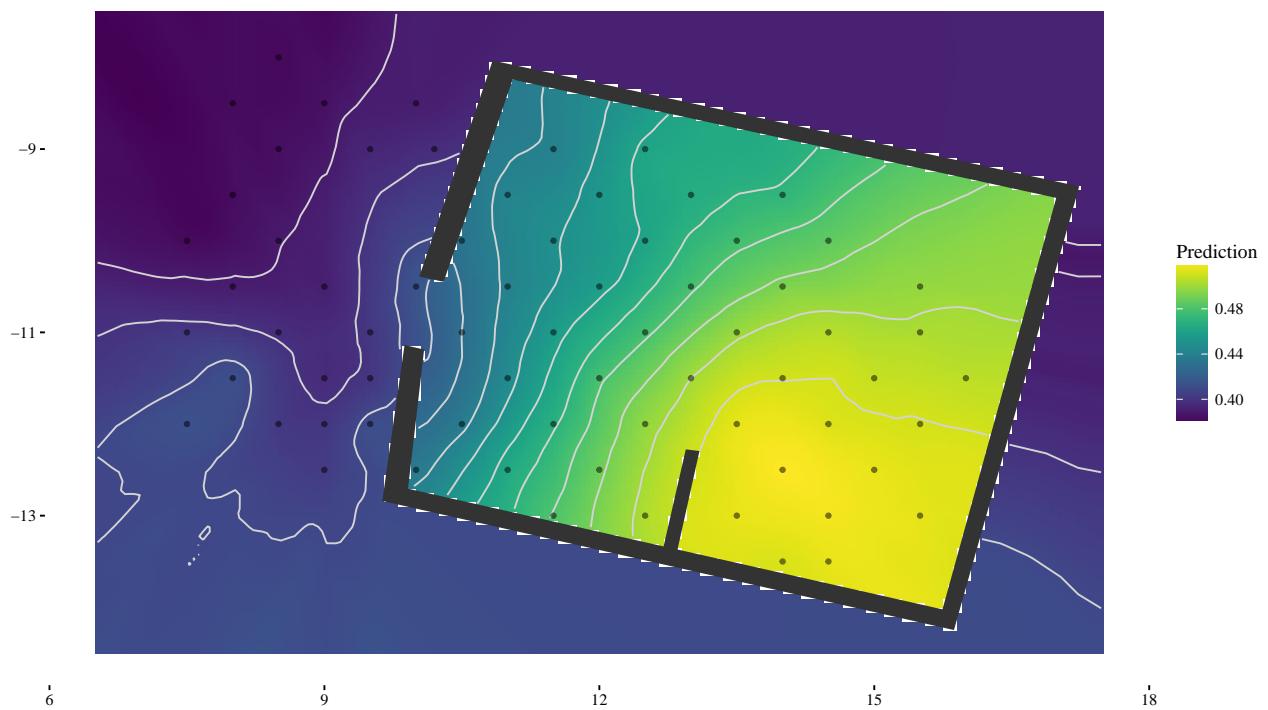


Figure 52: Cost-based kriging prediction

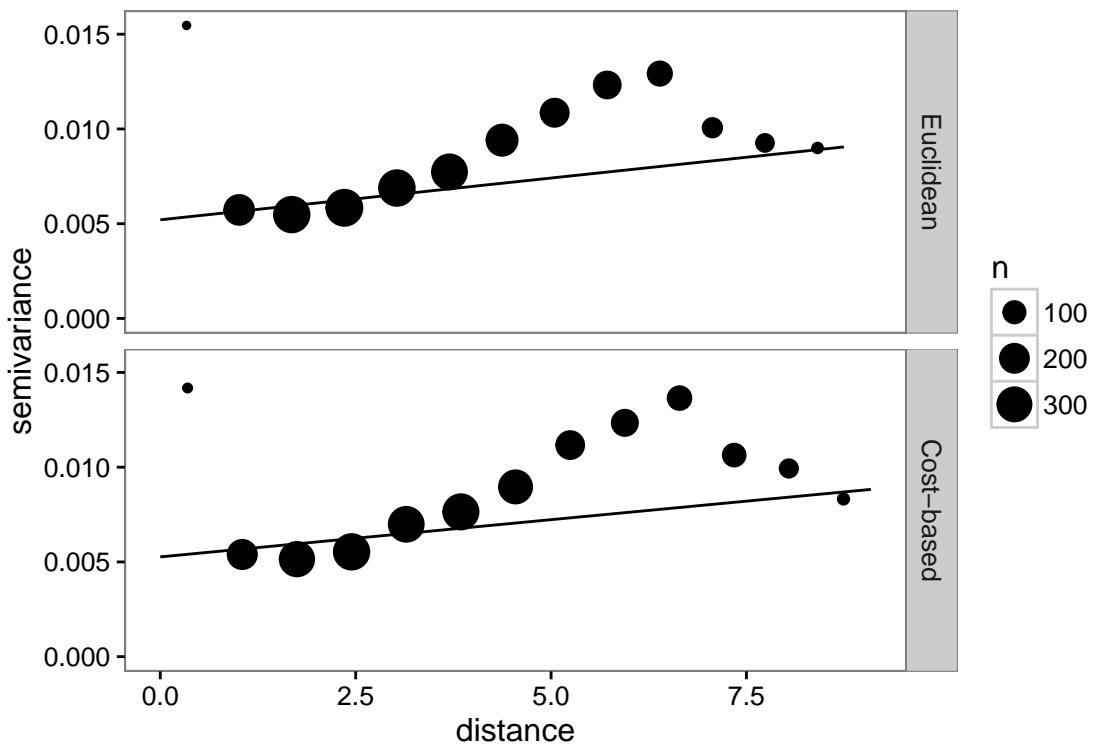
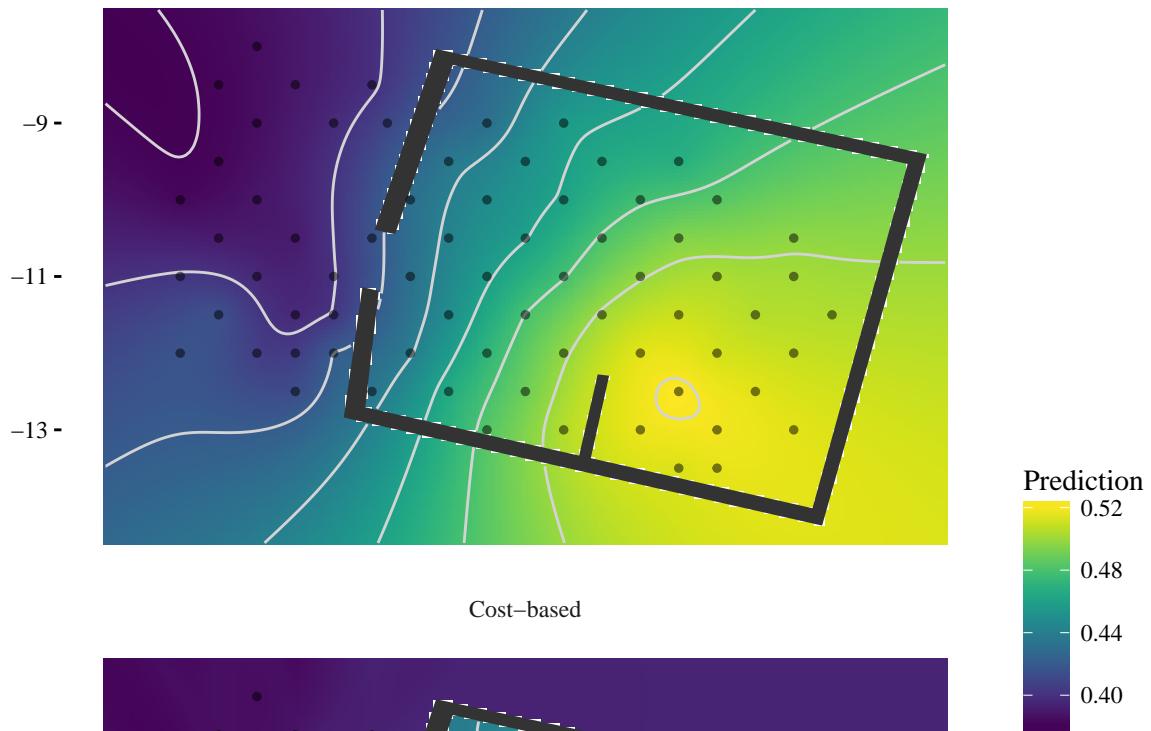


Figure 53: Empirical variogram and fitted models by method for Magnesium.

Classical



Cost-based

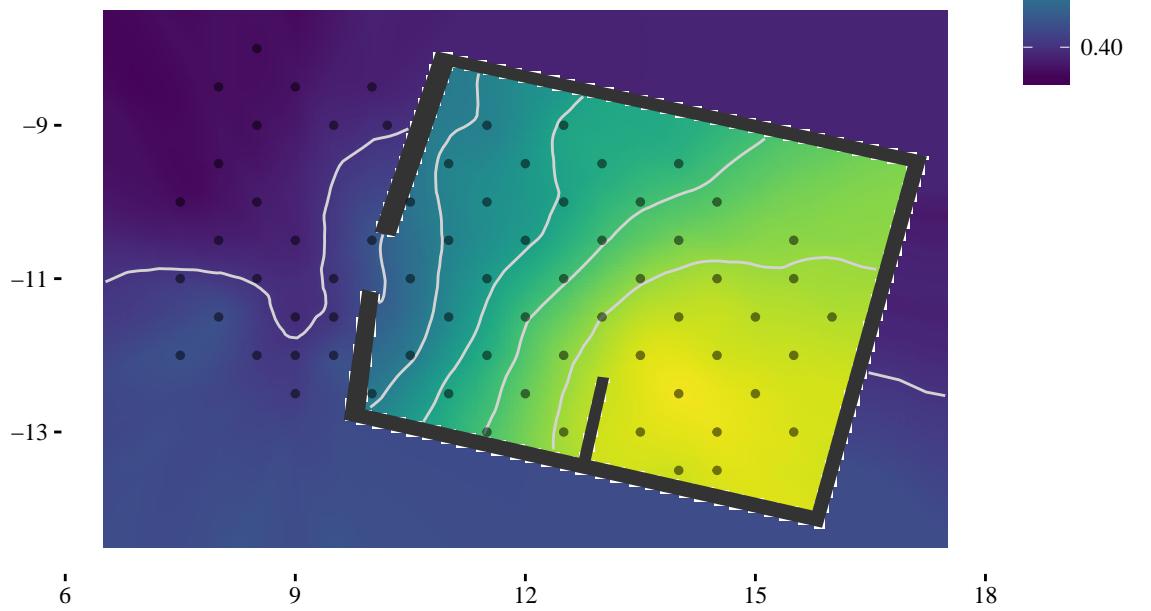


Figure 54: Comparison of Kriging estimates.

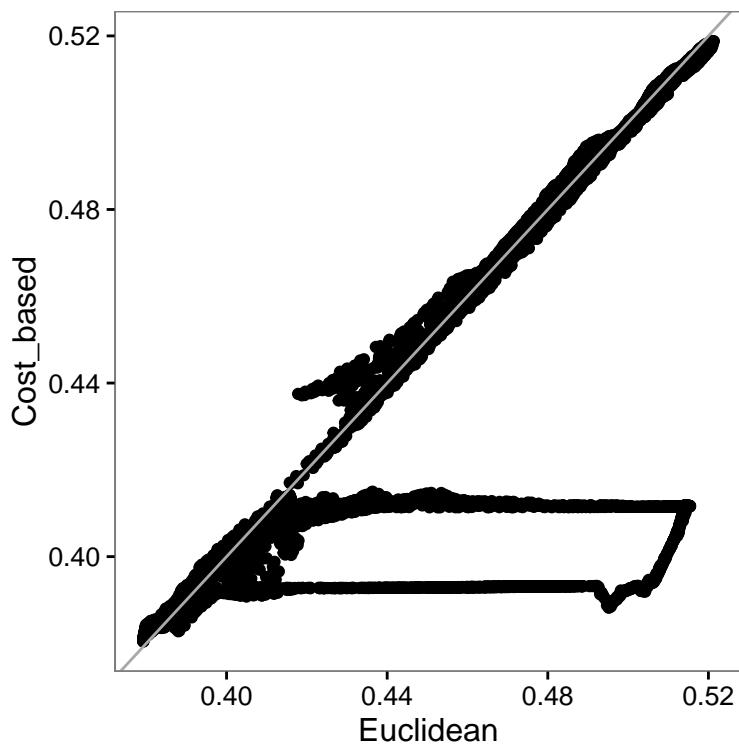


Figure 55: Pointwise comparison of predictions by method.

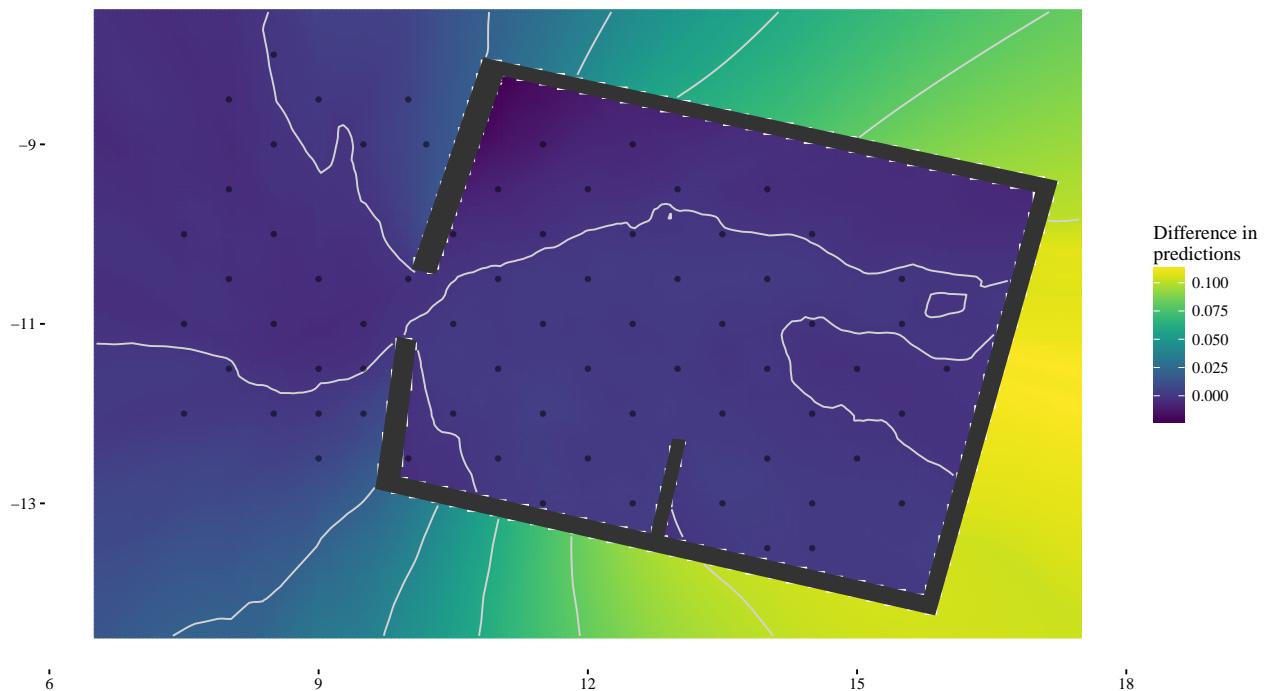


Figure 56: Difference between the Euclidean and the cost-based predictions.

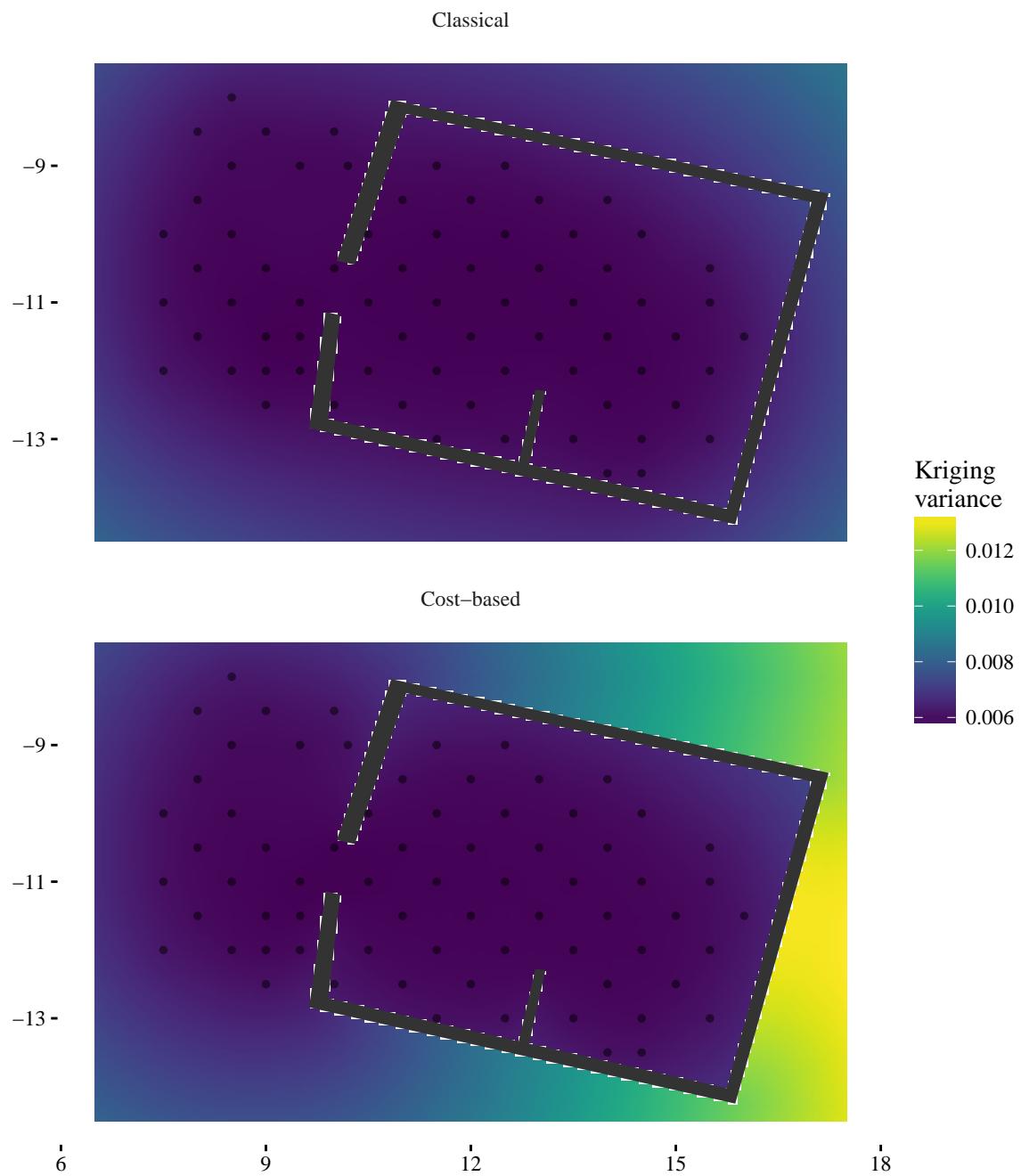


Figure 57: Comparison of prediction error by method.

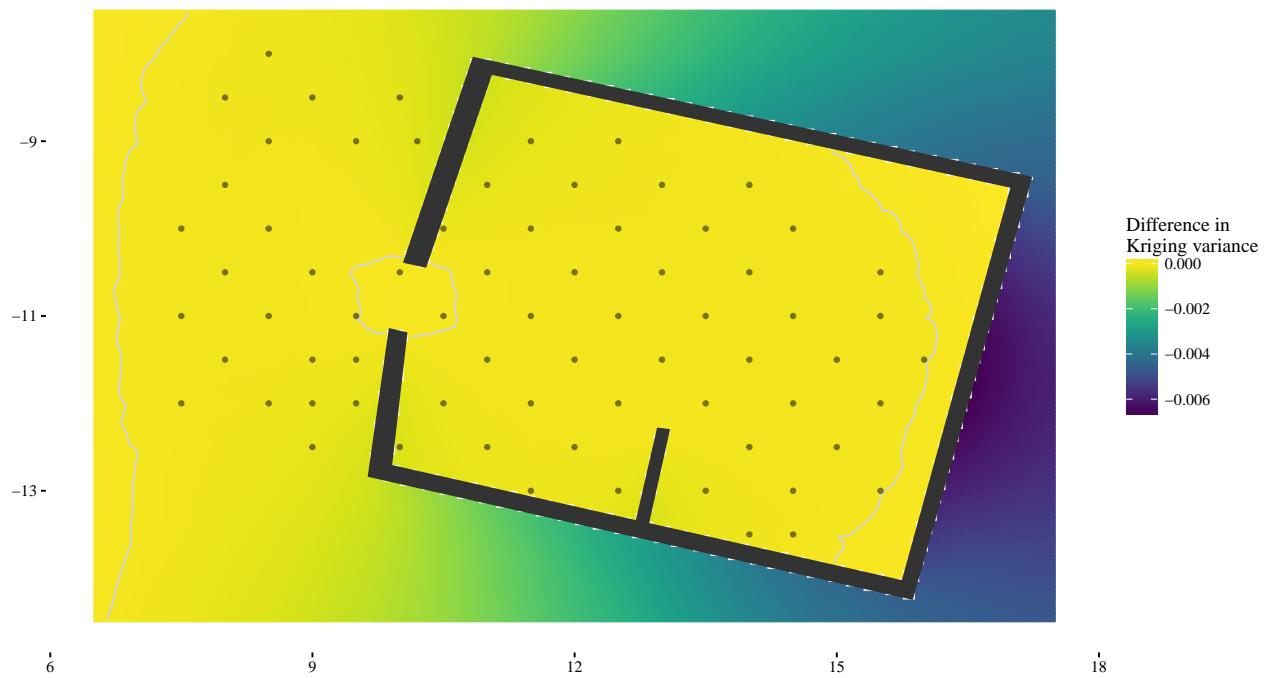


Figure 58: Difference between the Euclidean and the cost-based prediction errors

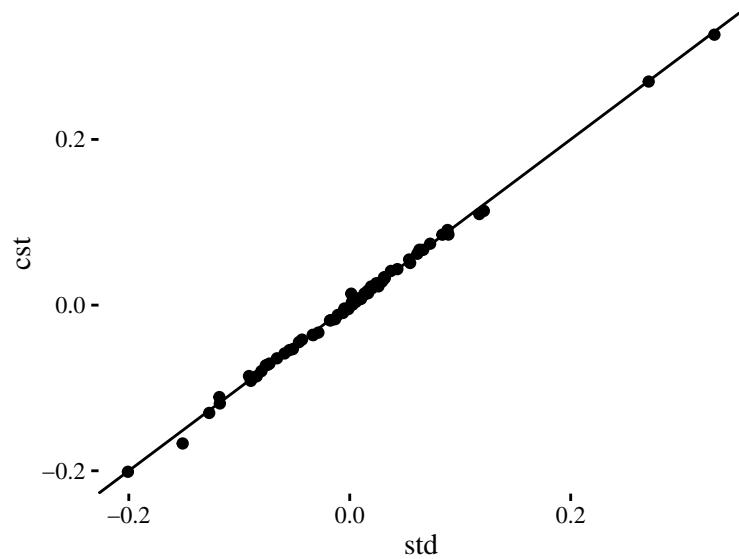


Figure 59: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	0.08
cst	0.08

## 8 Analysis of Zinc

### 8.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

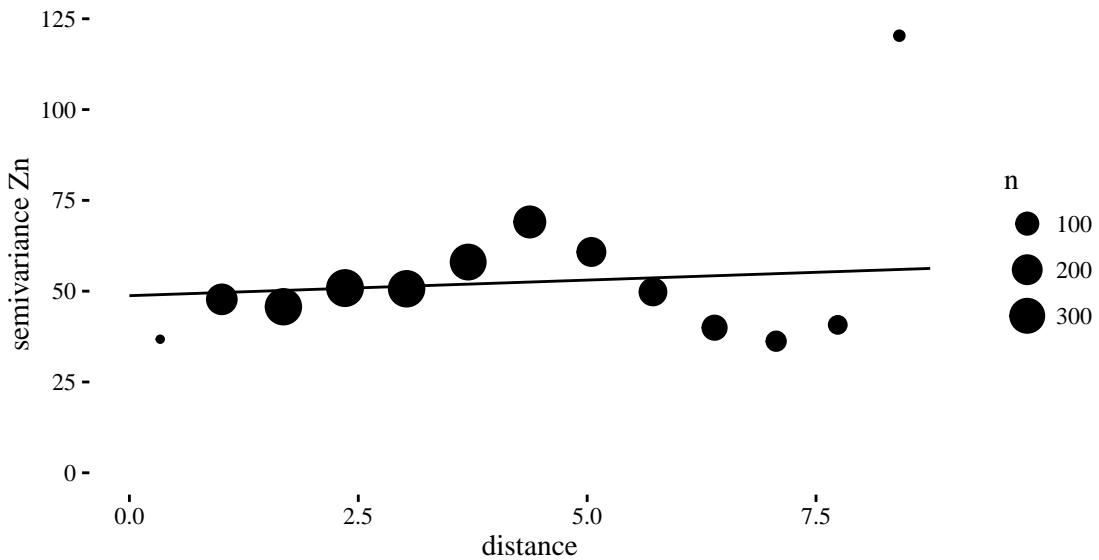


Figure 60: Empirical variogram and fitted model.

### 8.2 Cost-based kriging

### 8.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	31.71	31.63
Nugget	48.76	48.23
Partial sill	1791.06	2072.77
phi	2083.70	2244.35
Pract. range	6242.21	6723.48
Log-likelihood	-233.61	-233.43

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 8.4 Leave-one-out Cross Validation (LOOCV)

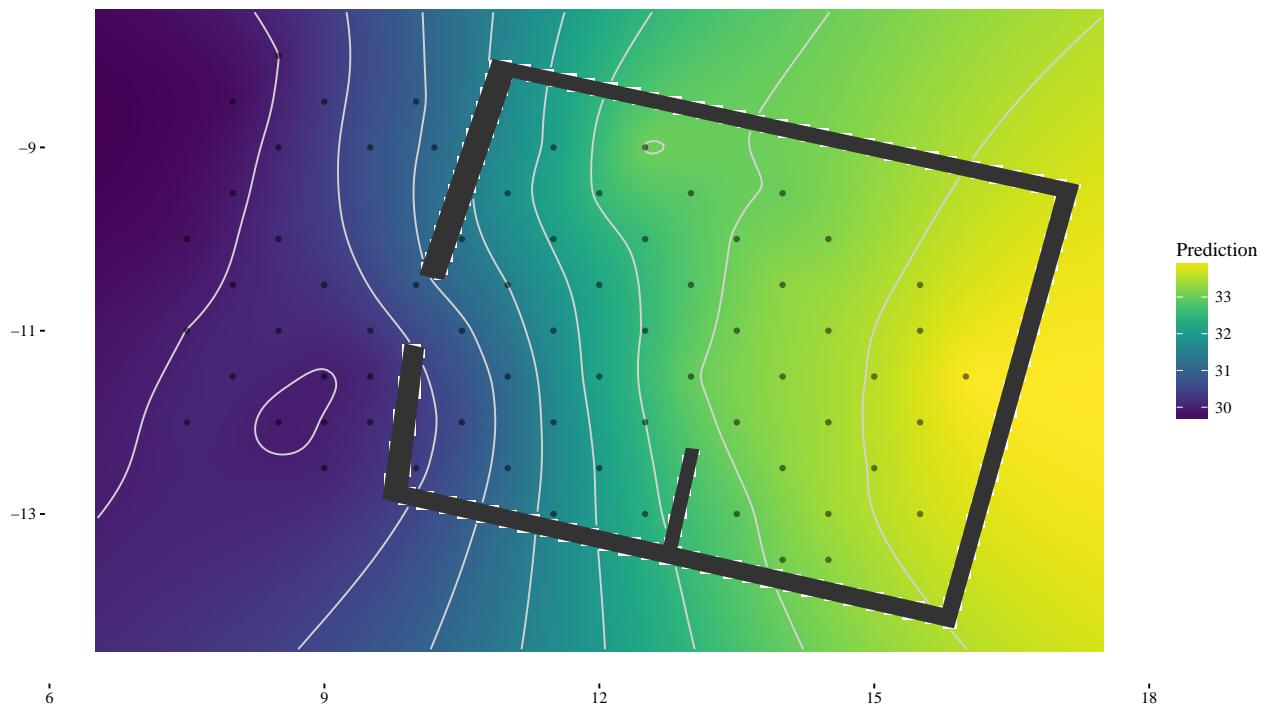


Figure 61: Euclidean kriging prediction

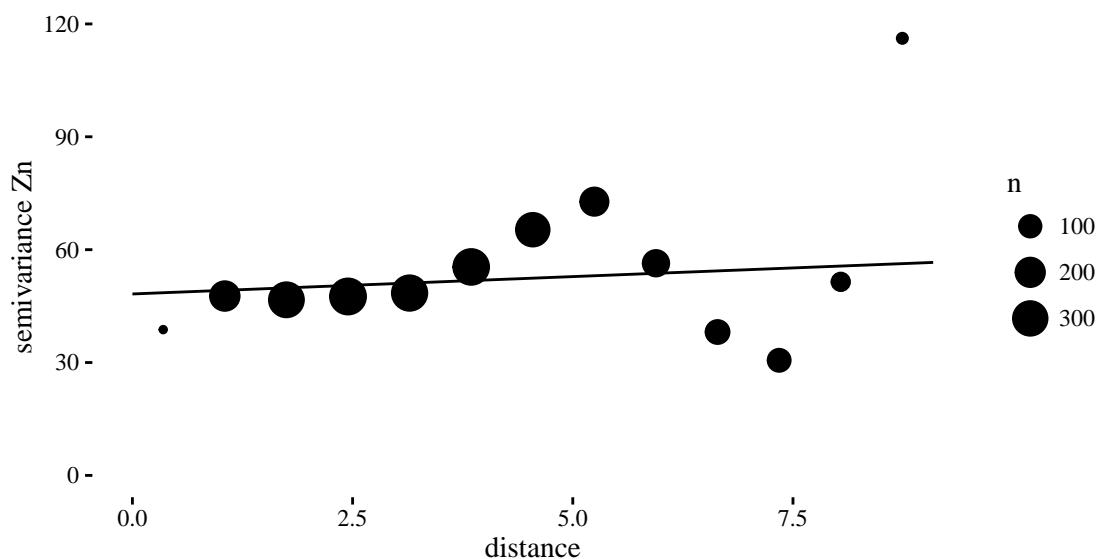


Figure 62: Empirical cost-based variogram and fitted model.

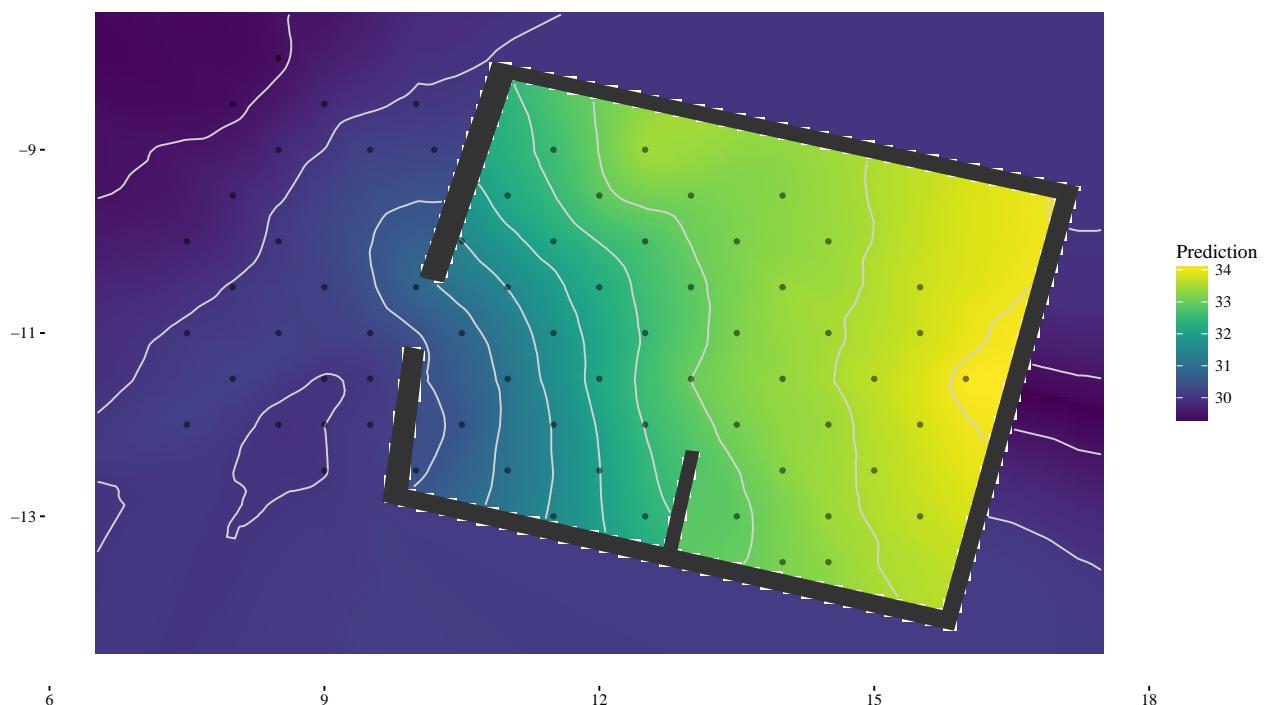


Figure 63: Cost-based kriging prediction

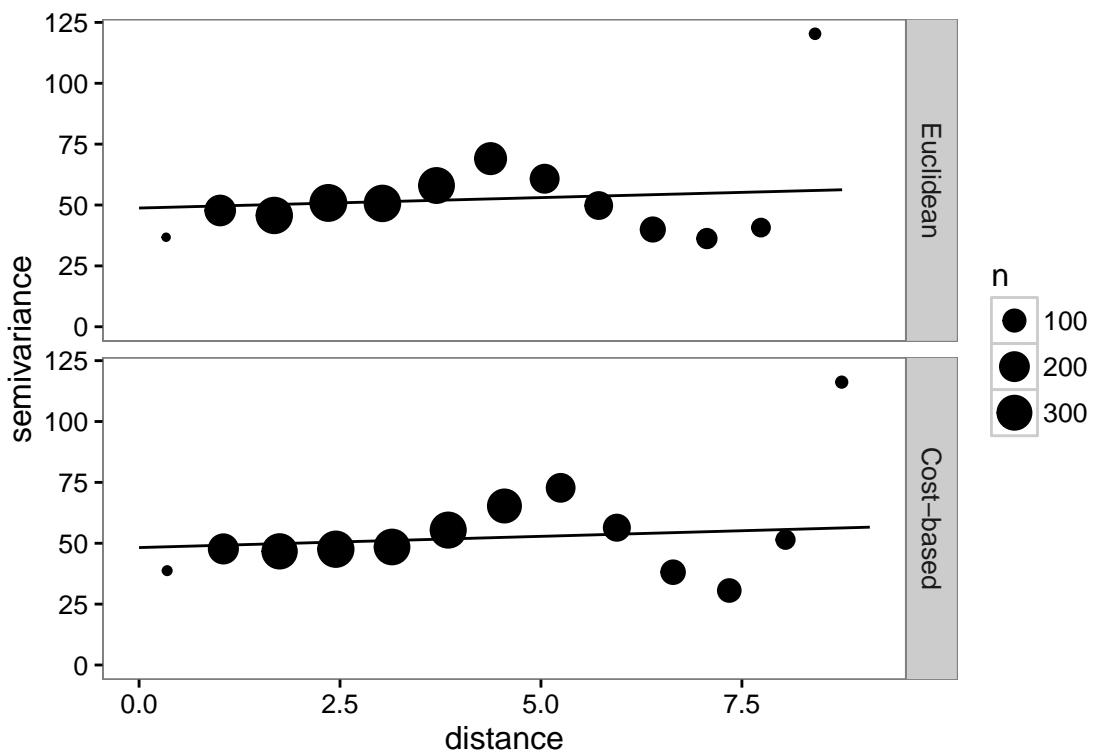


Figure 64: Empirical variogram and fitted models by method for Zinc.

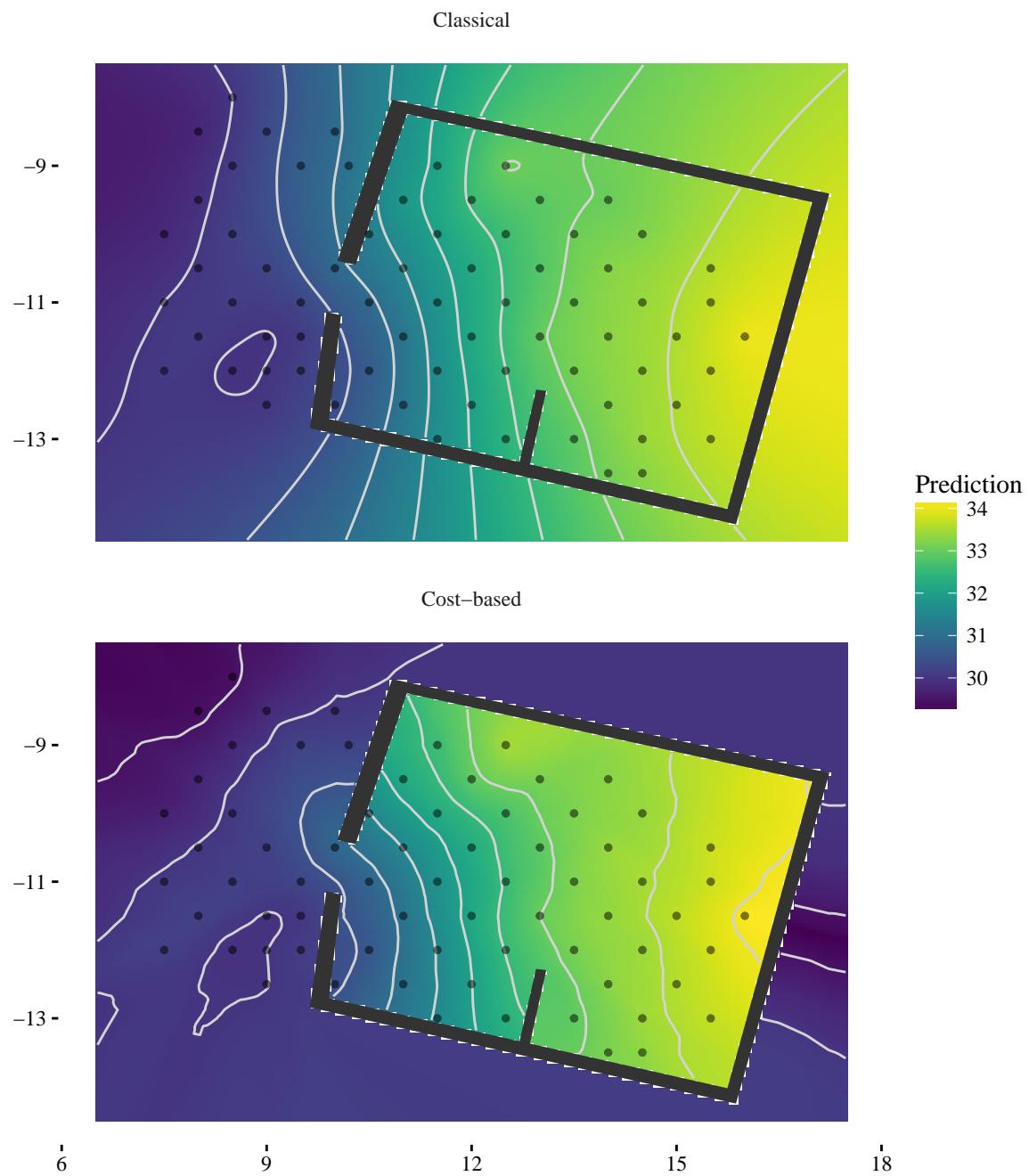


Figure 65: Comparison of Kriging estimates.

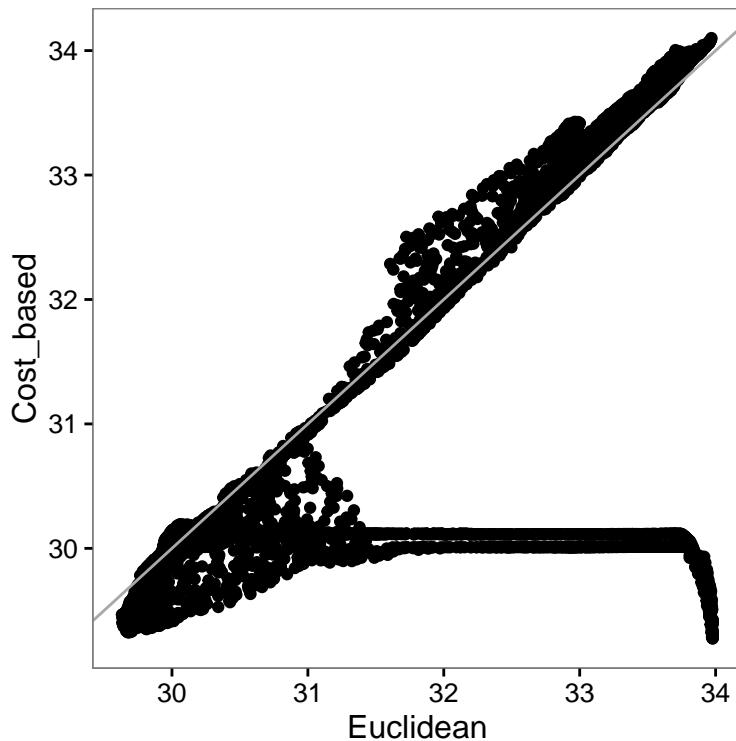


Figure 66: Pointwise comparison of predictions by method.

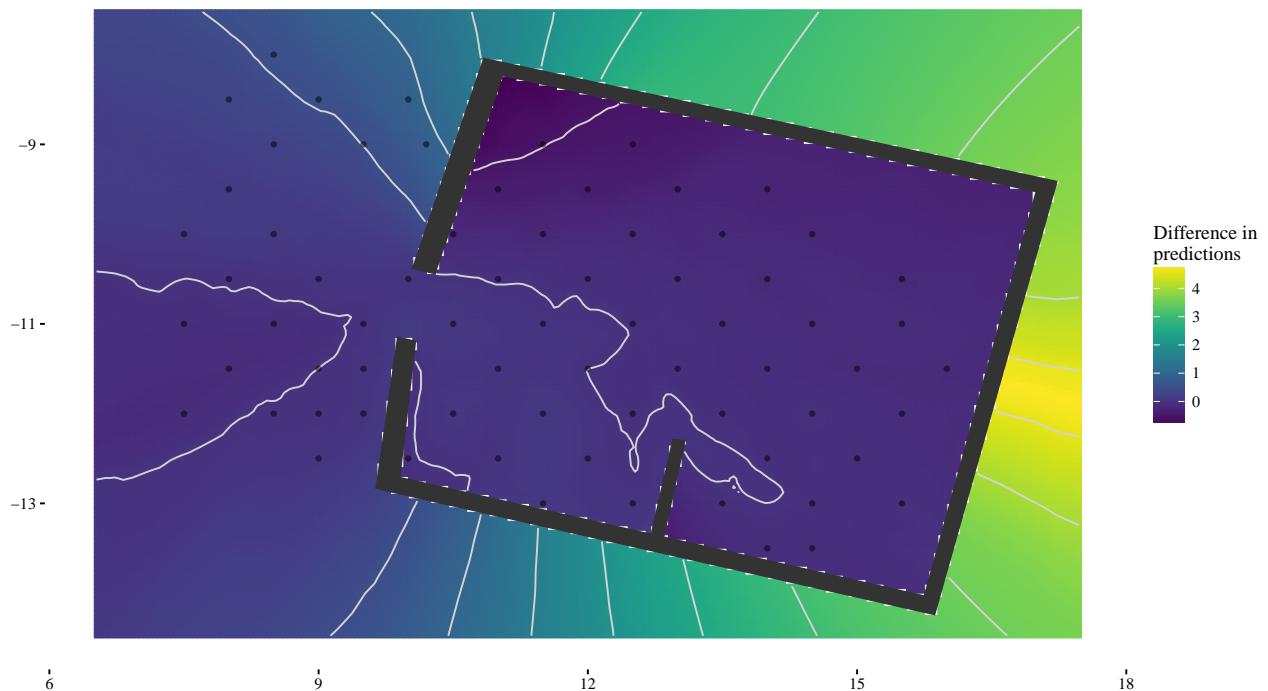


Figure 67: Difference between the Euclidean and the cost-based predictions.

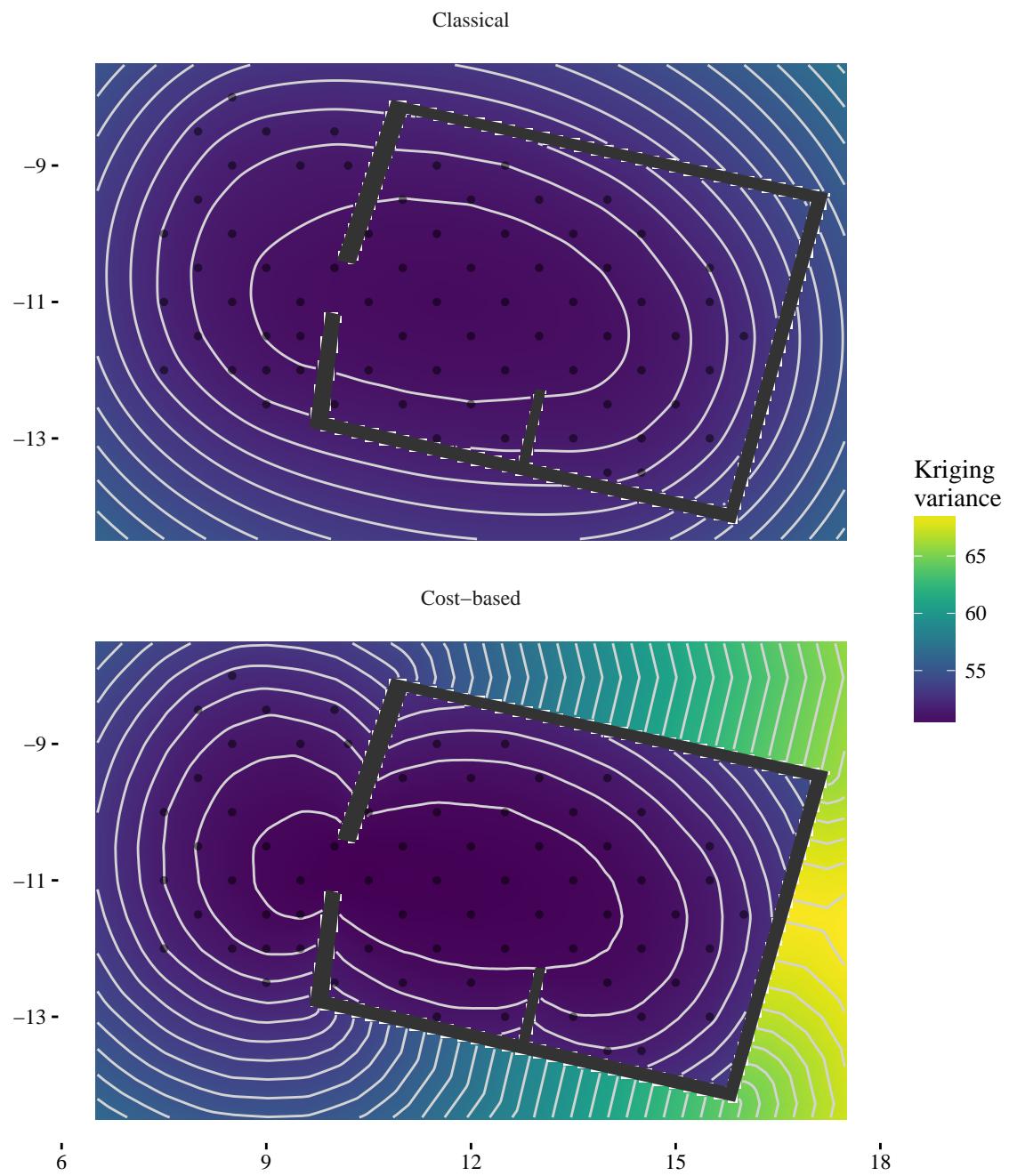


Figure 68: Comparison of prediction error by method.

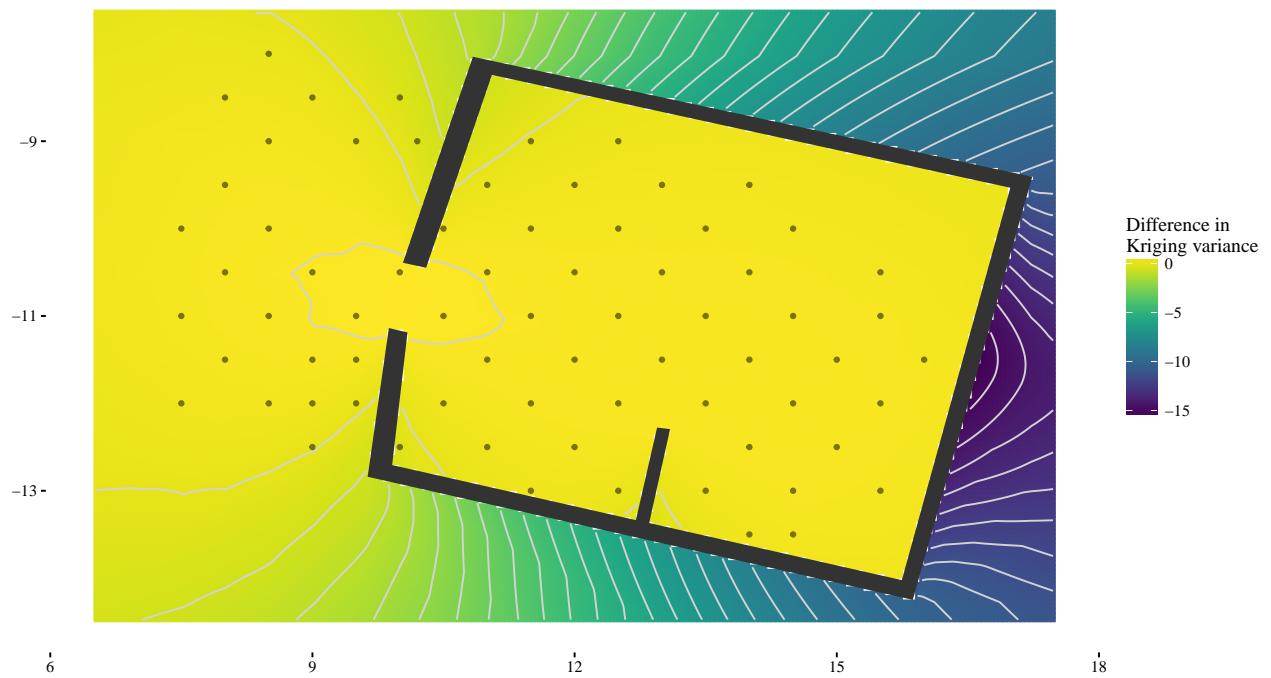


Figure 69: Difference between the Euclidean and the cost-based prediction errors

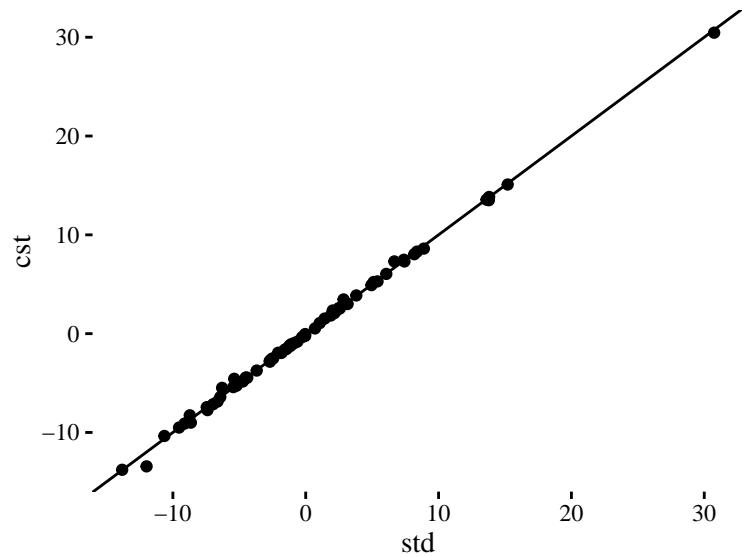


Figure 70: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	7.23
cst	7.23

## 9 Analysis of food remains

```
as.geodata: 3 points removed due to NA in the data
```

```
[1] 14 59 60
```

### 9.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

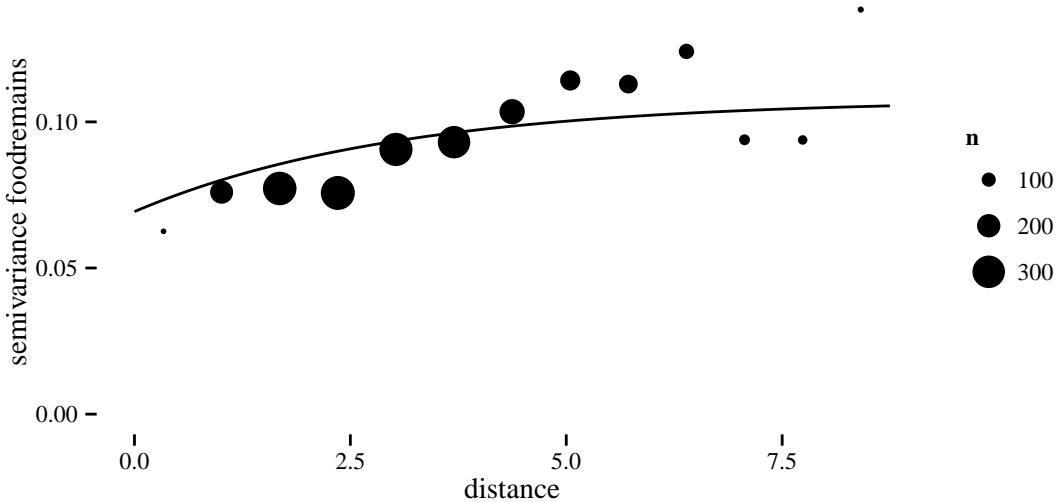


Figure 71: Empirical variogram and fitted model.

### 9.2 Cost-based kriging

### 9.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	1.95	1.94
Nugget	0.07	0.07
Partial sill	0.04	0.04
phi	3.04	2.94
Pract. range	9.12	8.81
Log-likelihood	-12.84	-12.77

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

In the main area, the prediction errors are practically the same with both approaches. In most of the area, the Cost-based approach is slightly more accurate (median 1%) achieving up to 1.4% of improvement in accuracy in the rightmost corners.

The drop of about -5% prediction variance for the Euclidean prediction in the other corners and the veranda area are a consequence of incorrectly assuming that observations at the other side of the wall are *close*.

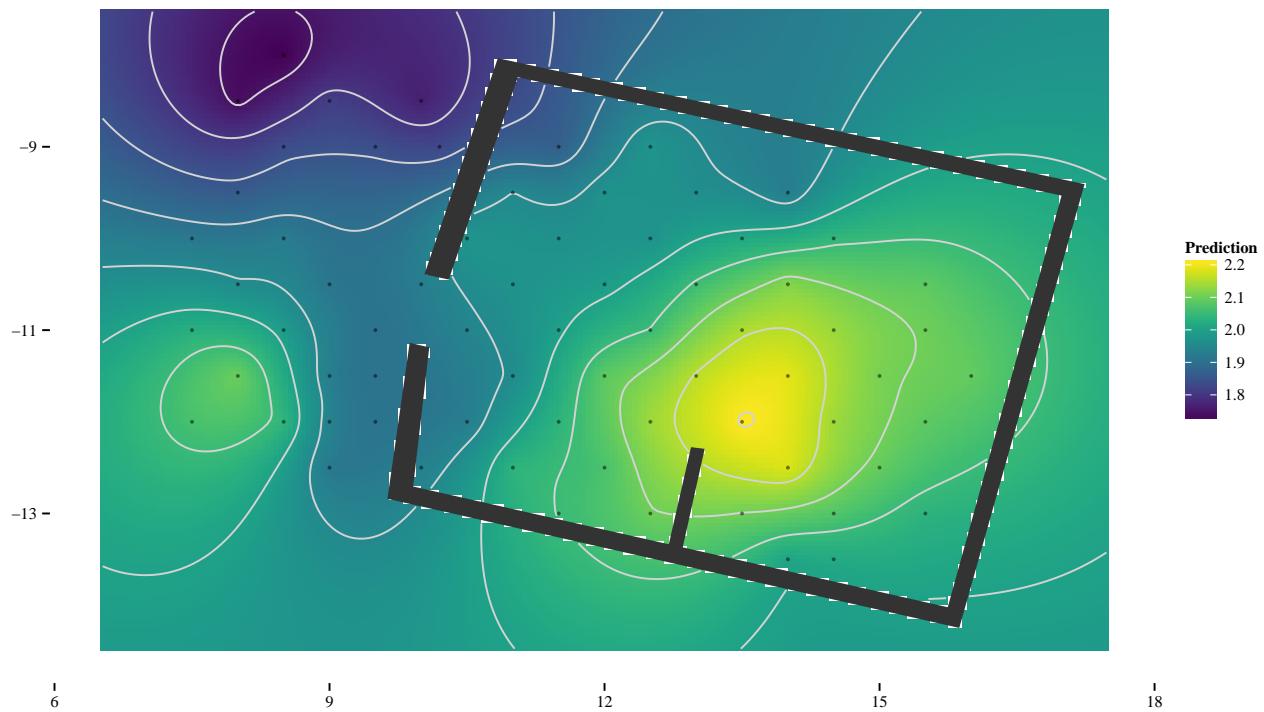


Figure 72: Euclidean kriging prediction

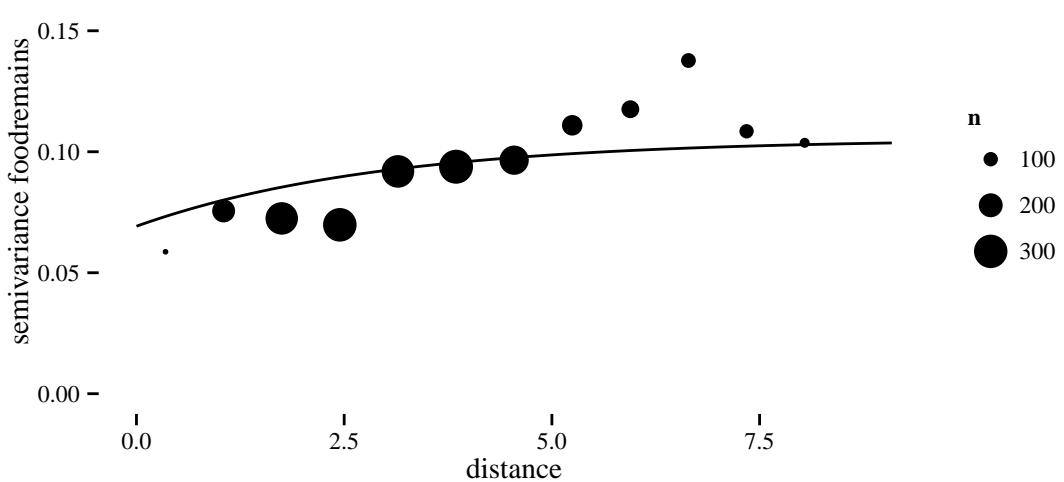


Figure 73: Empirical cost-based variogram and fitted model.

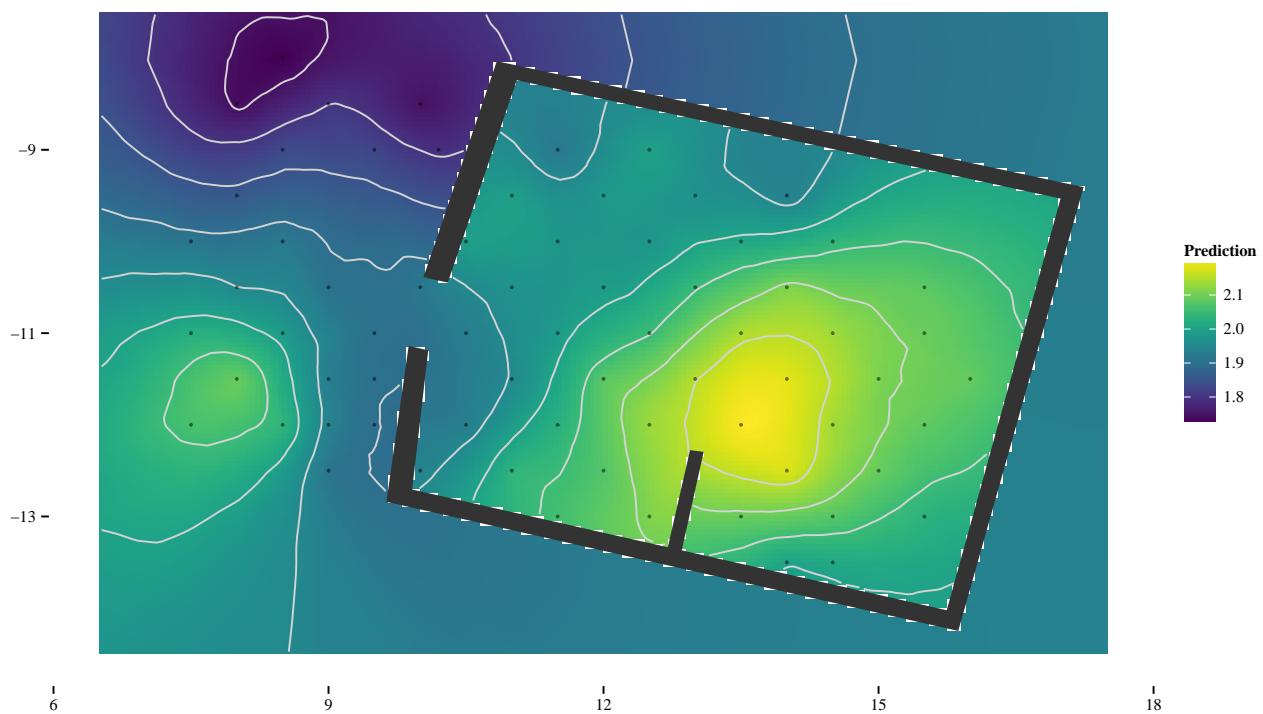


Figure 74: Cost-based kriging prediction

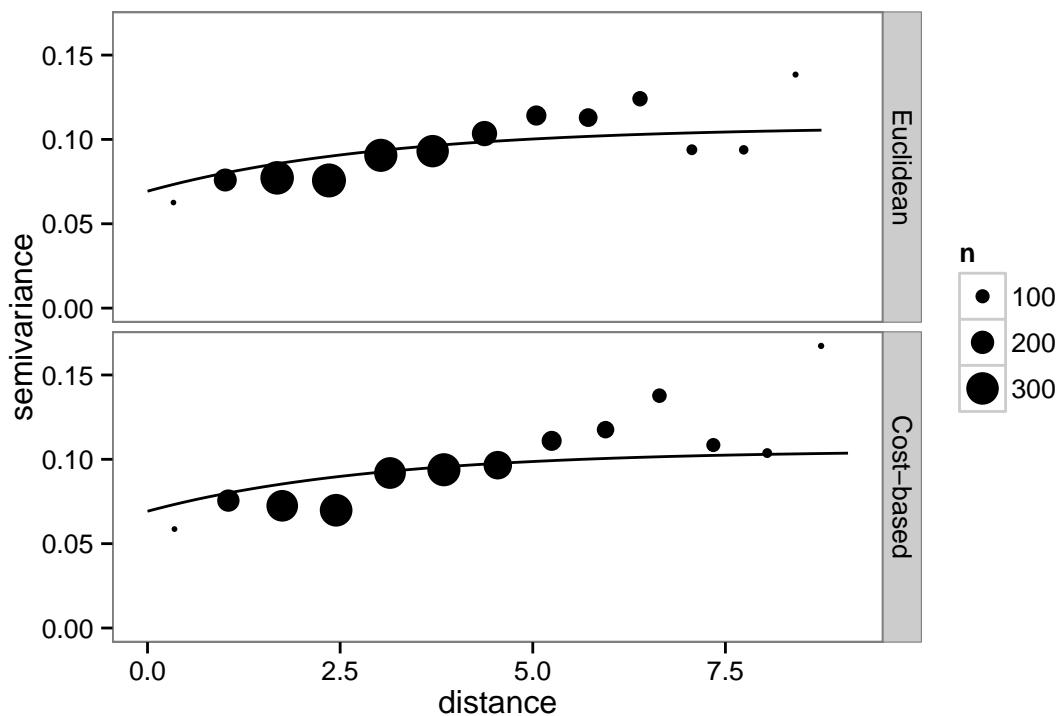


Figure 75: Empirical variogram and fitted models by method for food remains.

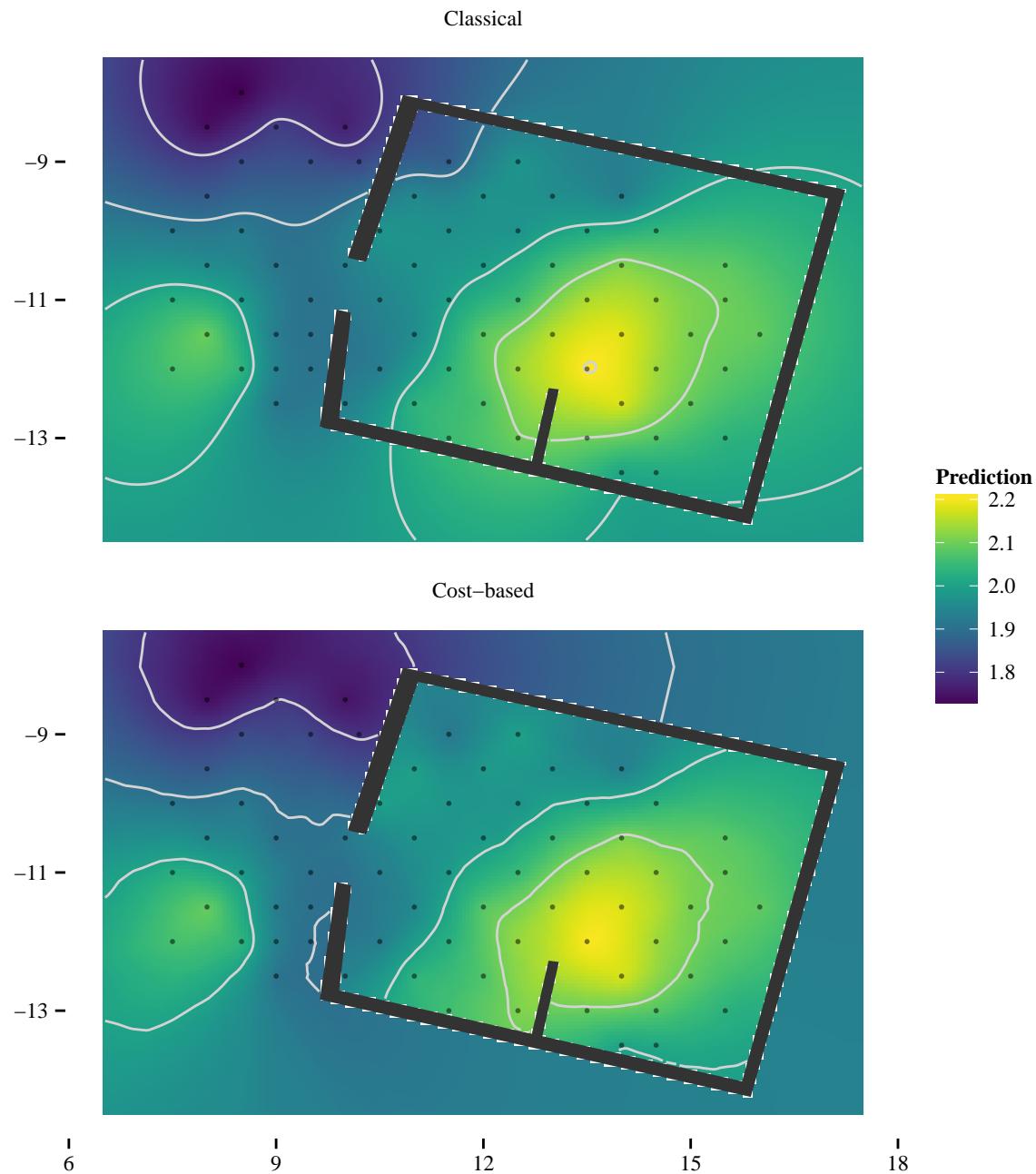


Figure 76: Comparison of Kriging estimates.

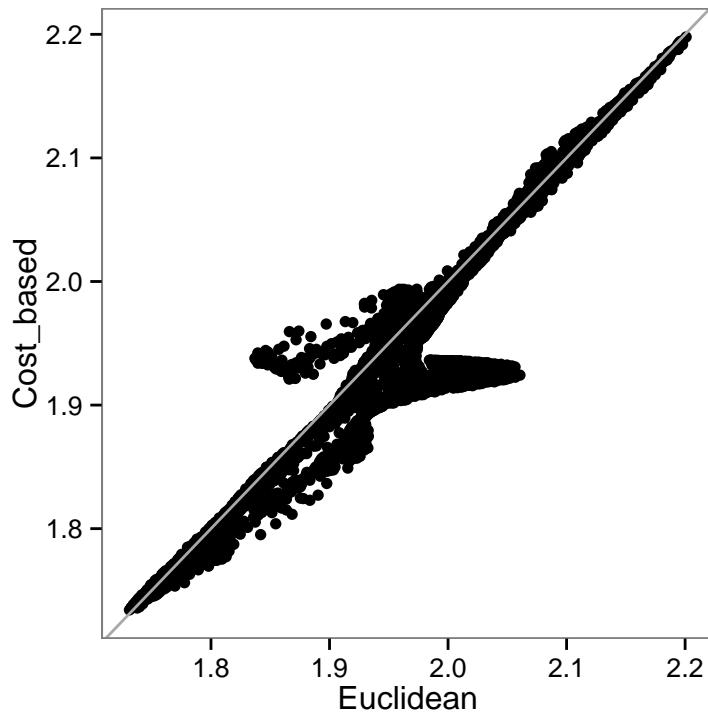


Figure 77: Pointwise comparison of predictions by method.

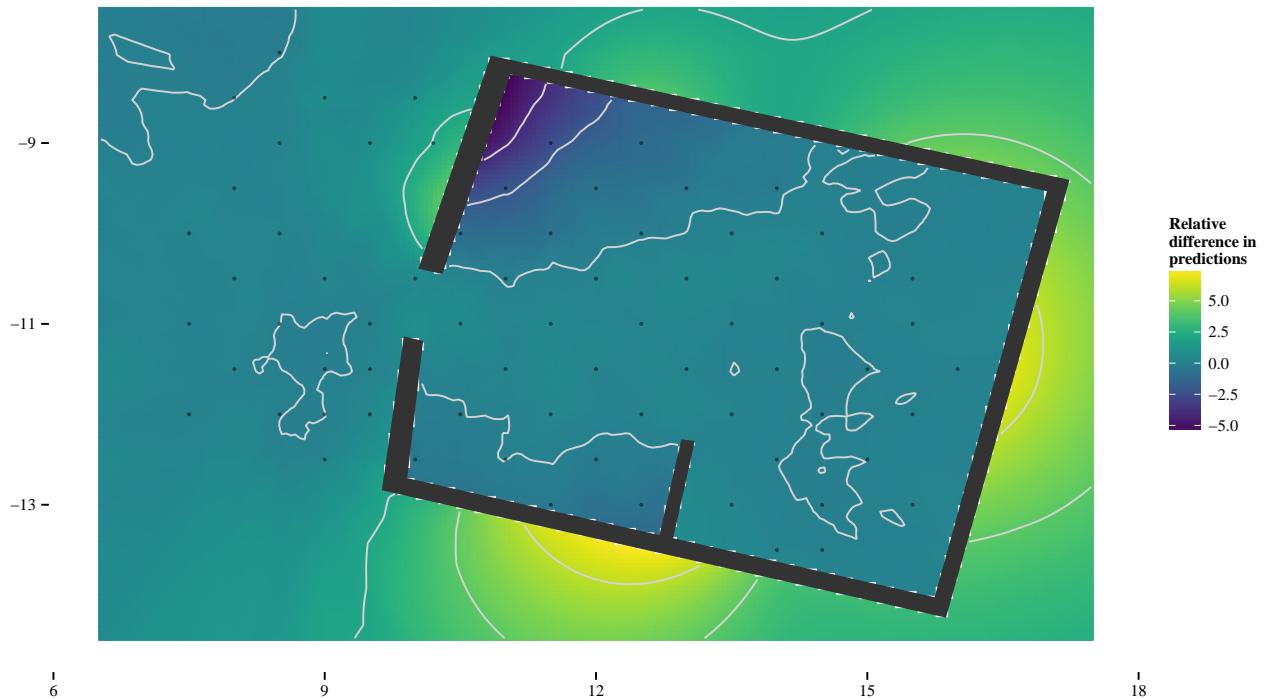


Figure 78: Relative Difference between the Euclidean and the cost-based predictions. Expressed as a percentage wrt Cost-based prediction.

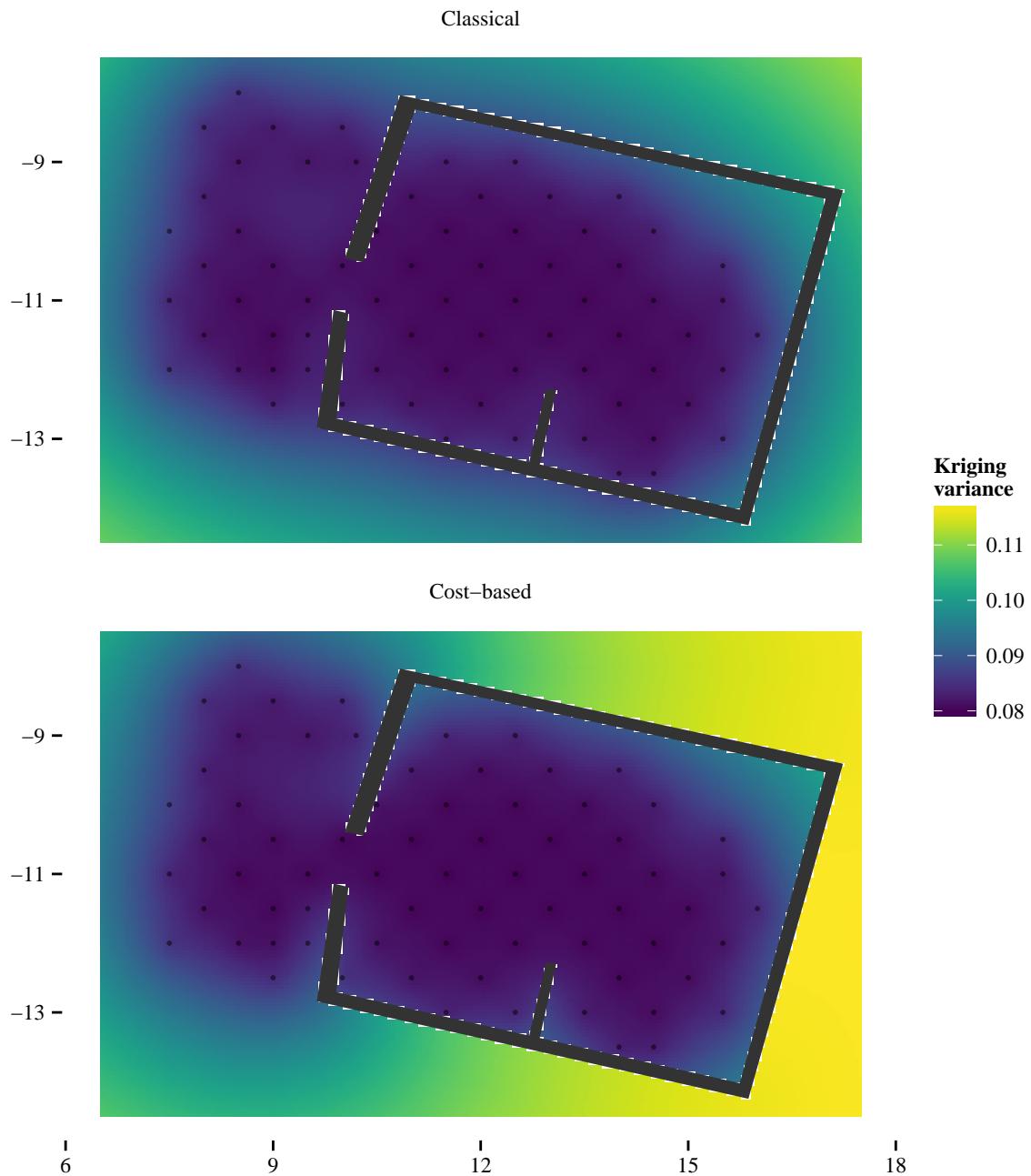


Figure 79: Comparison of prediction error by method.

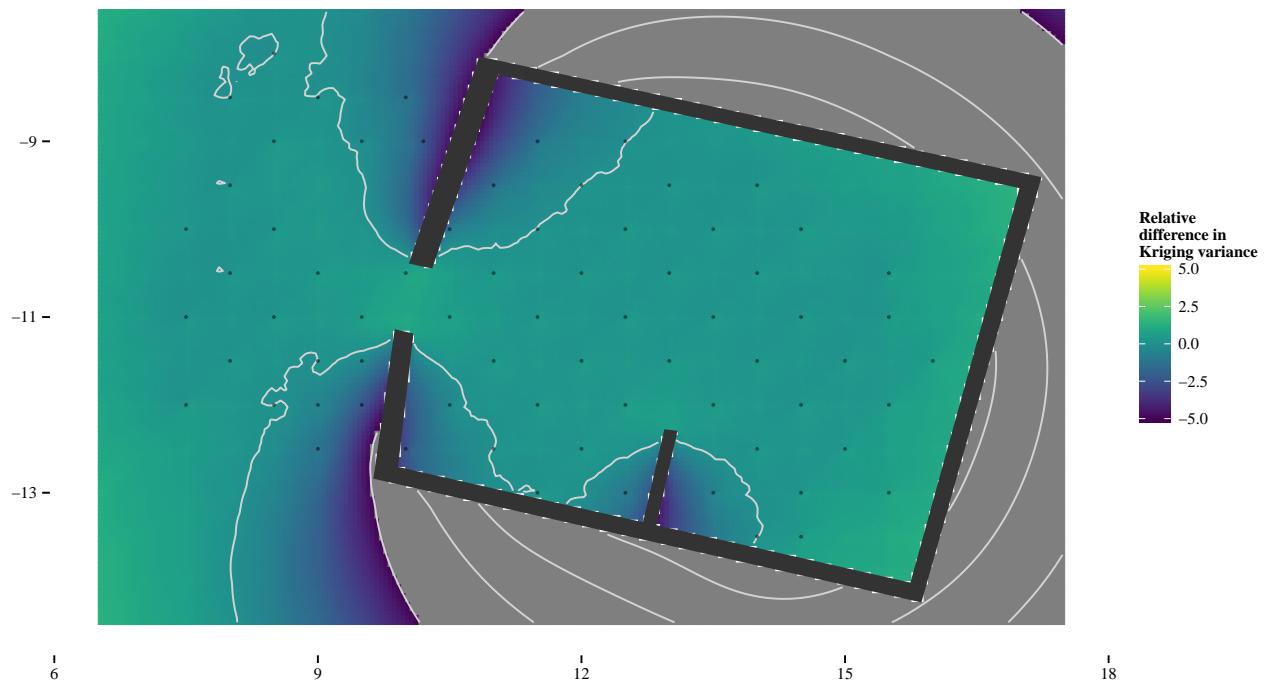


Figure 80: Relative difference between the Euclidean and the cost-based prediction errors. Expressed as a percentage wrt Cost-based variance.

Behind the walls, the Euclidean prediction error is unrealistically low (up to -23%) and is not represented.

#### 9.4 Leave-one-out Cross Validation (LOOCV)

method	rmse(error)
std	0.31
cst	0.31

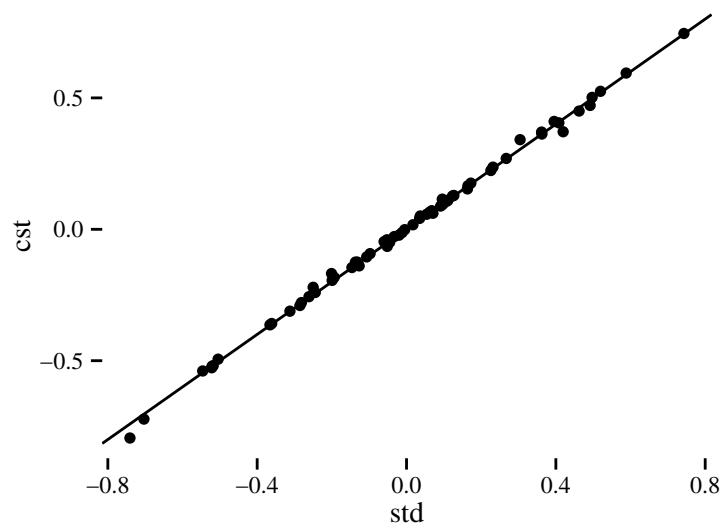


Figure 81: Pointwise leave-one-out prediction error by method.

## 10 Analysis of living room

as.geodata: 7 points removed due to NA in the data

### 10.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

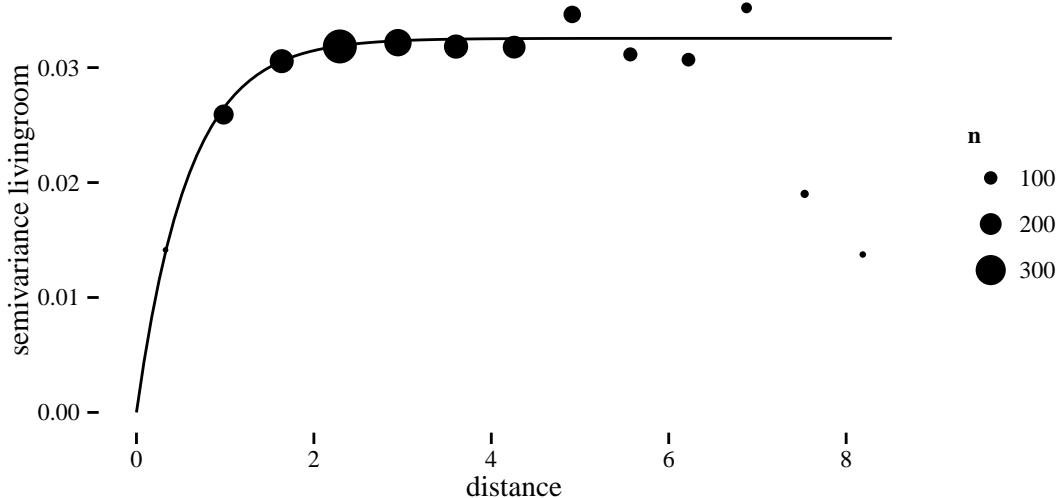


Figure 82: Empirical variogram and fitted model.

### 10.2 Cost-based kriging

### 10.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	1.54	1.53
Nugget	0.00	0.00
Partial sill	0.03	0.03
phi	0.58	0.66
Pract. range	1.75	1.99
Log-likelihood	23.38	23.68

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 10.4 Leave-one-out Cross Validation (LOOCV)

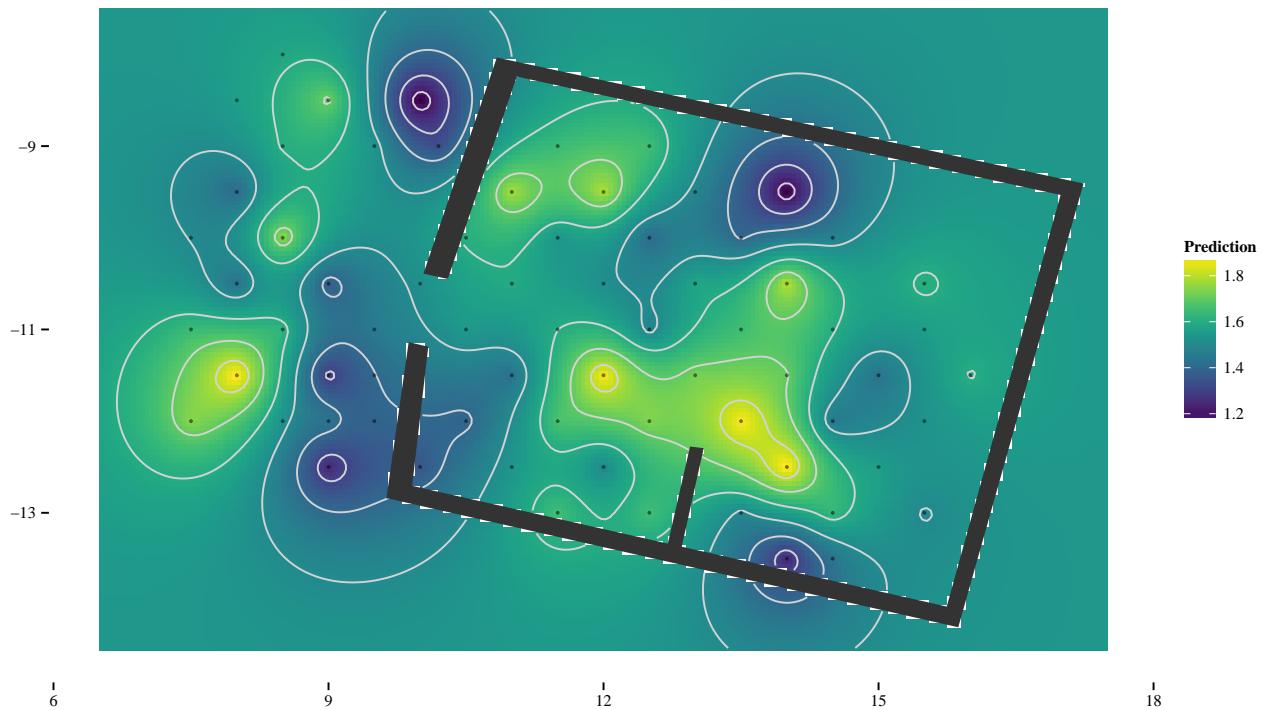


Figure 83: Euclidean kriging prediction

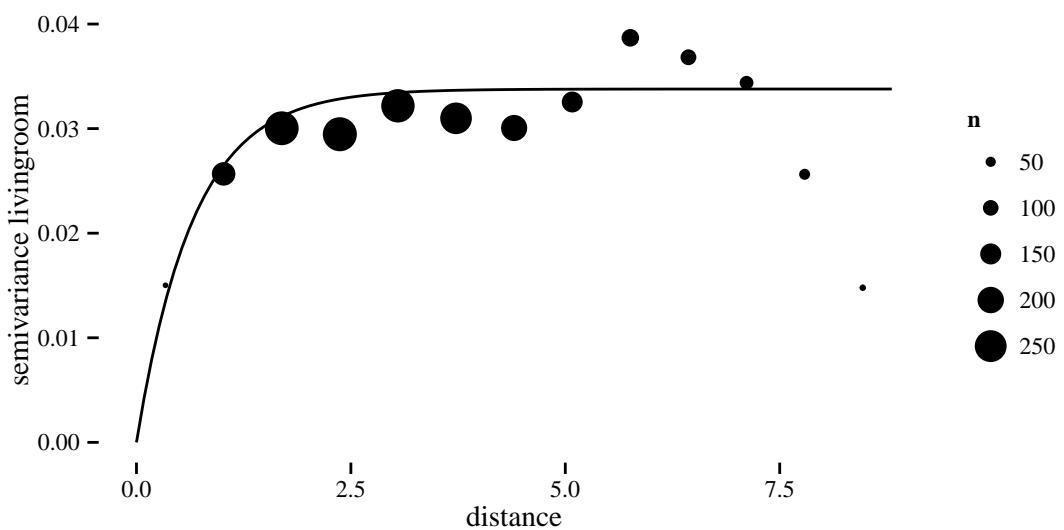


Figure 84: Empirical cost-based variogram and fitted model.

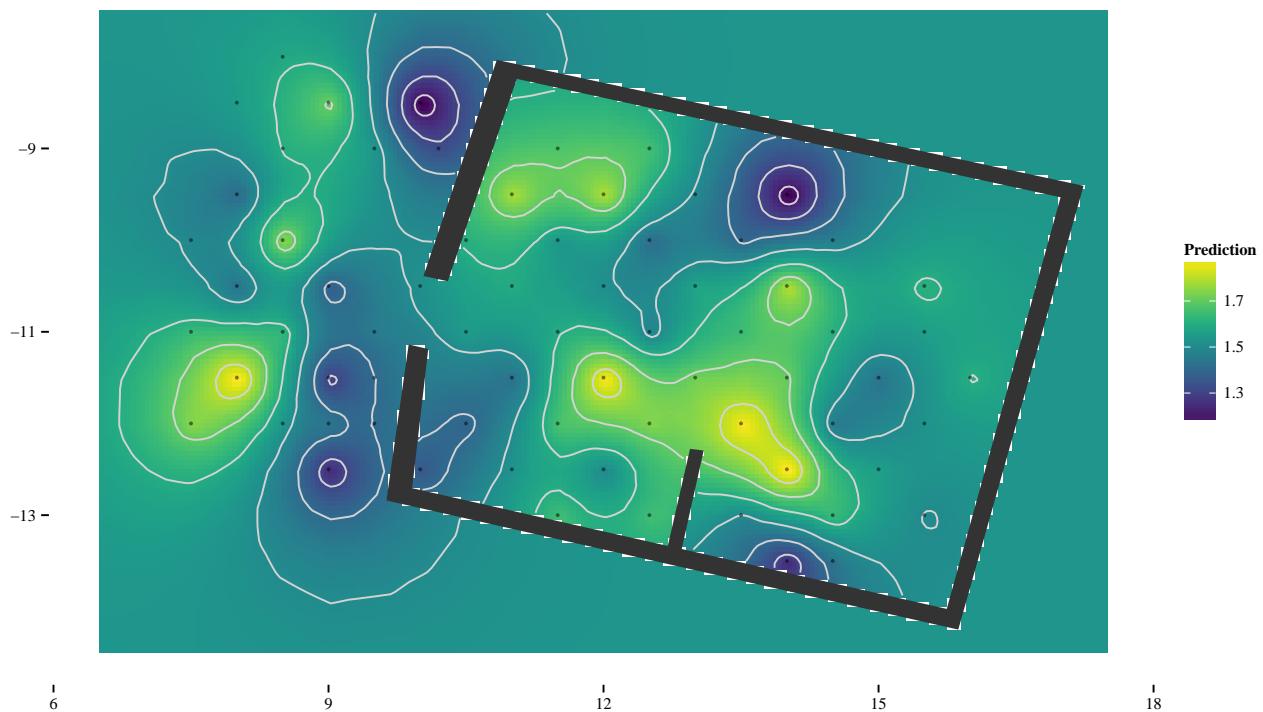


Figure 85: Cost-based kriging prediction

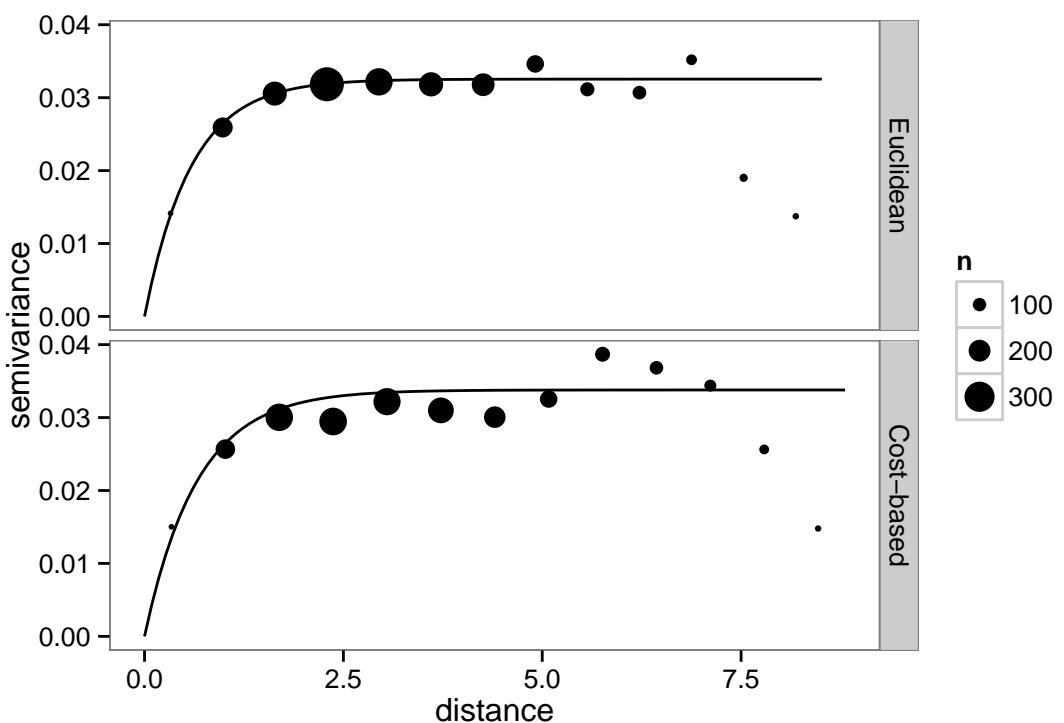


Figure 86: Empirical variogram and fitted models by method for living room.

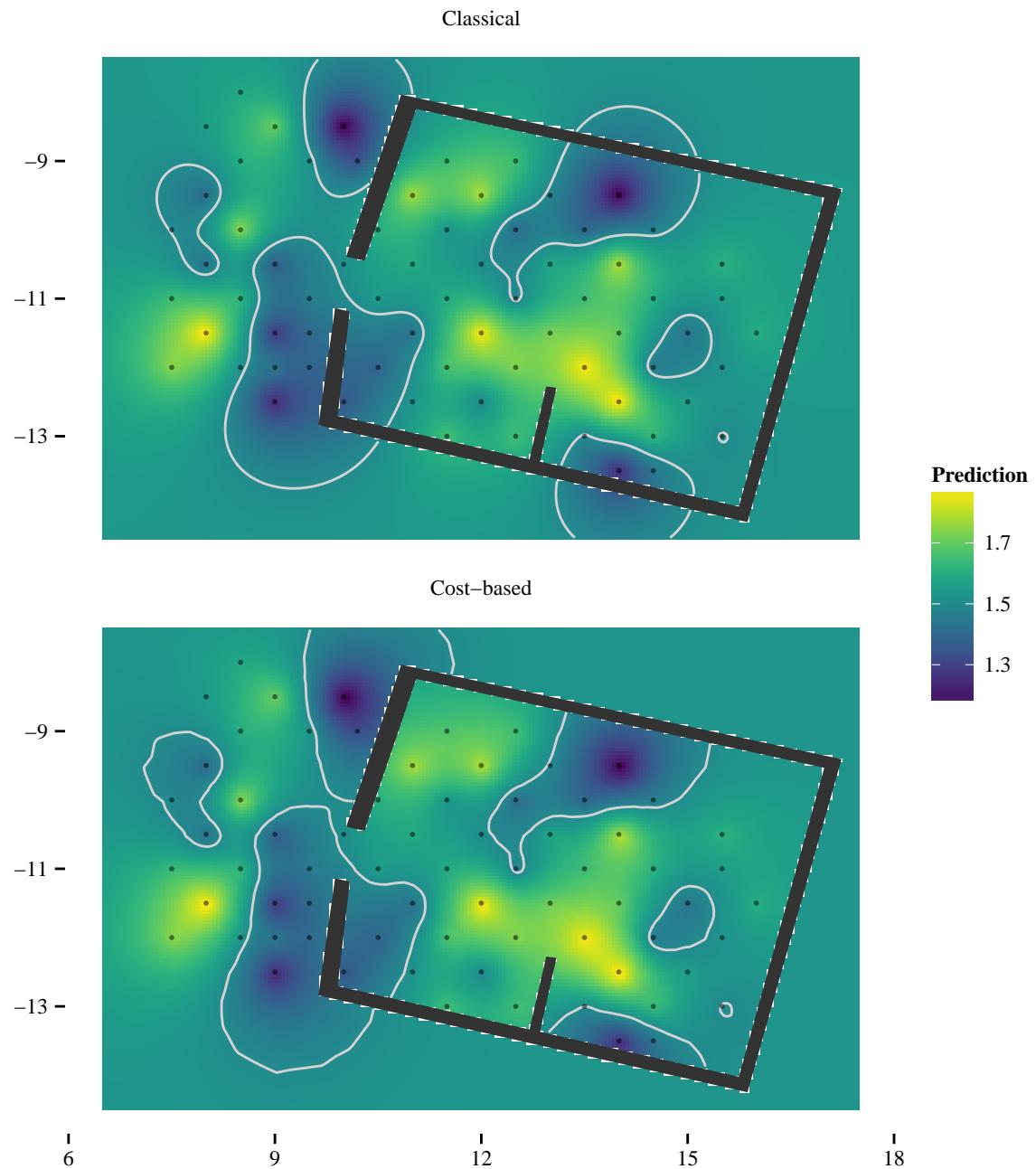


Figure 87: Comparison of Kriging estimates.

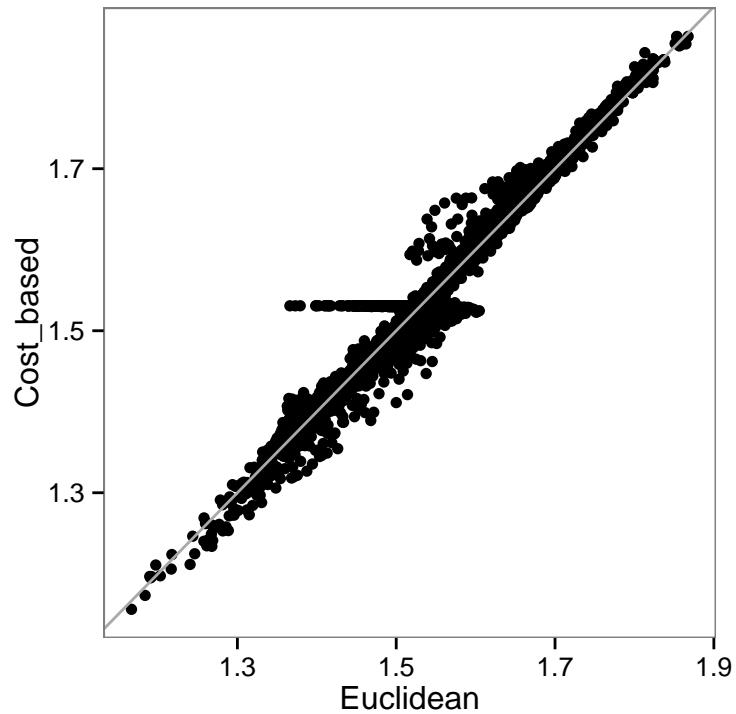


Figure 88: Pointwise comparison of predictions by method.

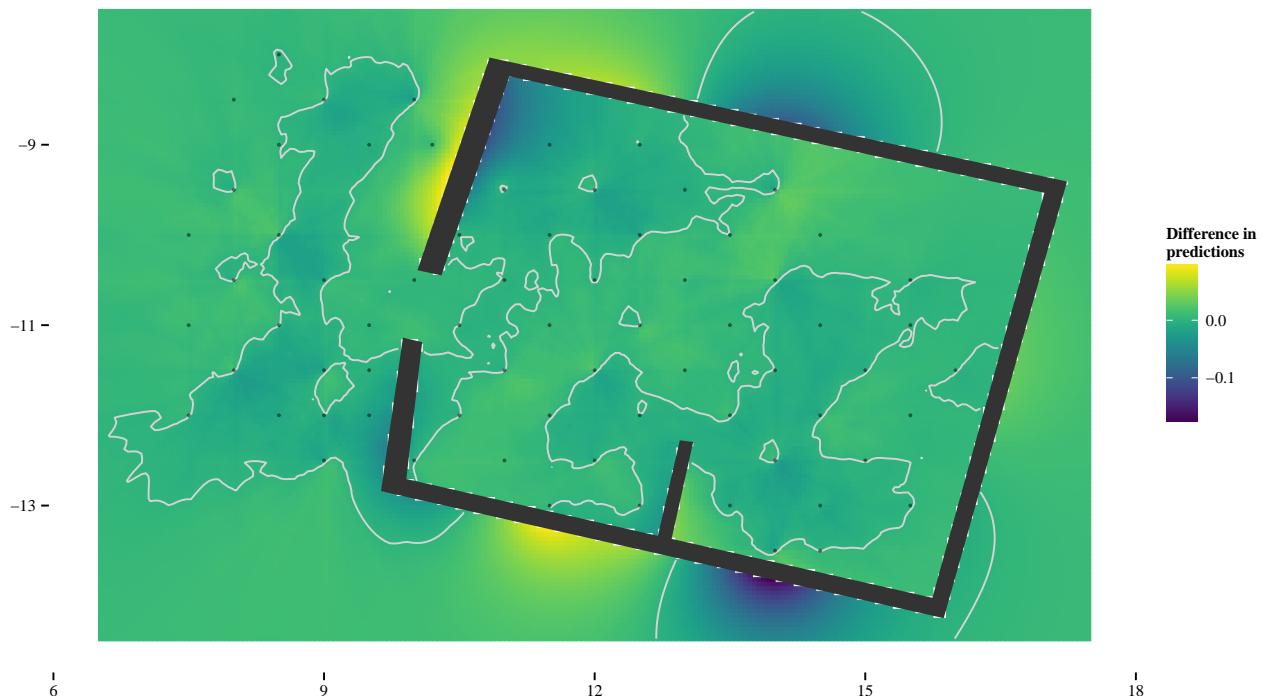


Figure 89: Difference between the Euclidean and the cost-based predictions.

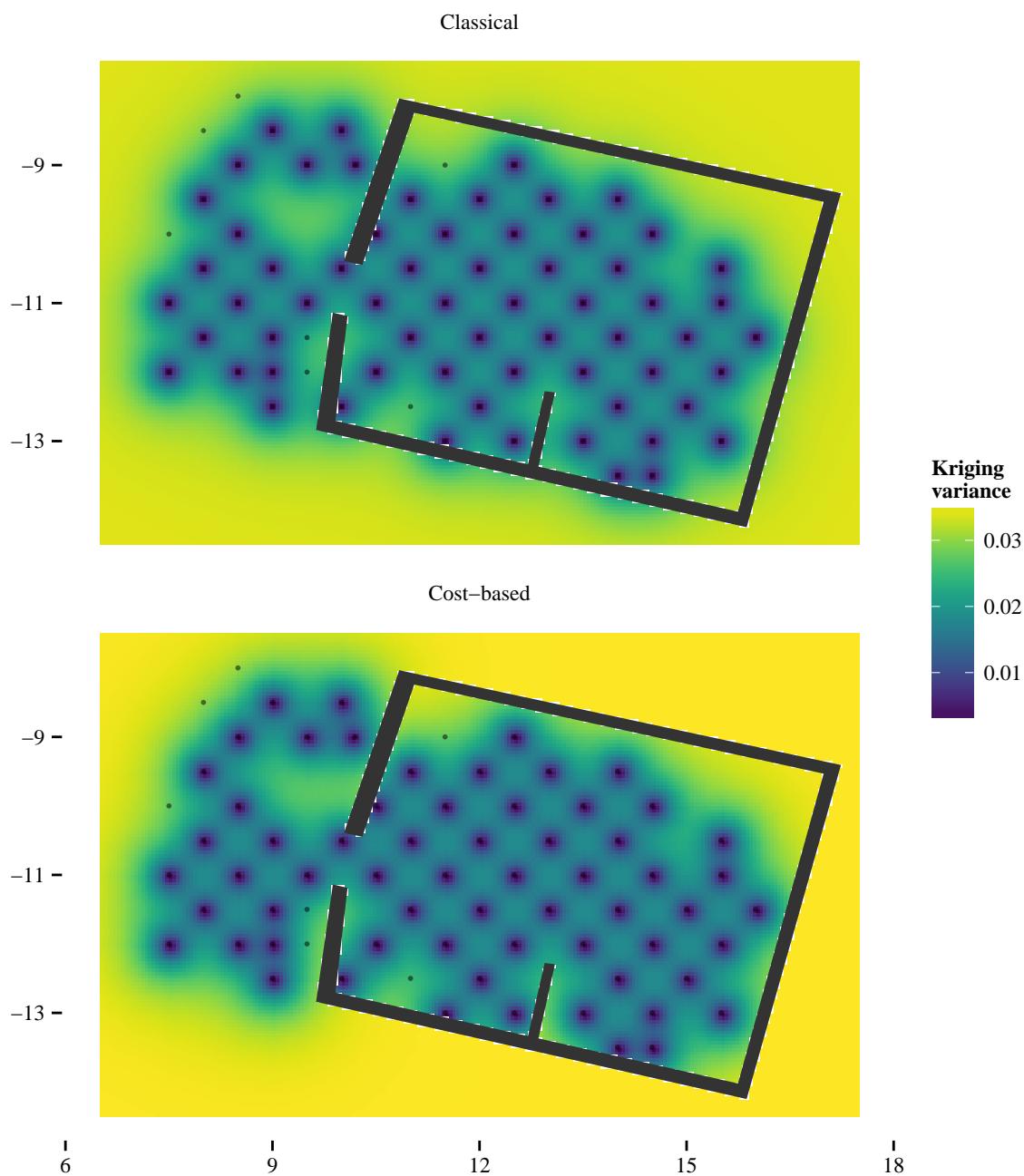


Figure 90: Comparison of prediction error by method.

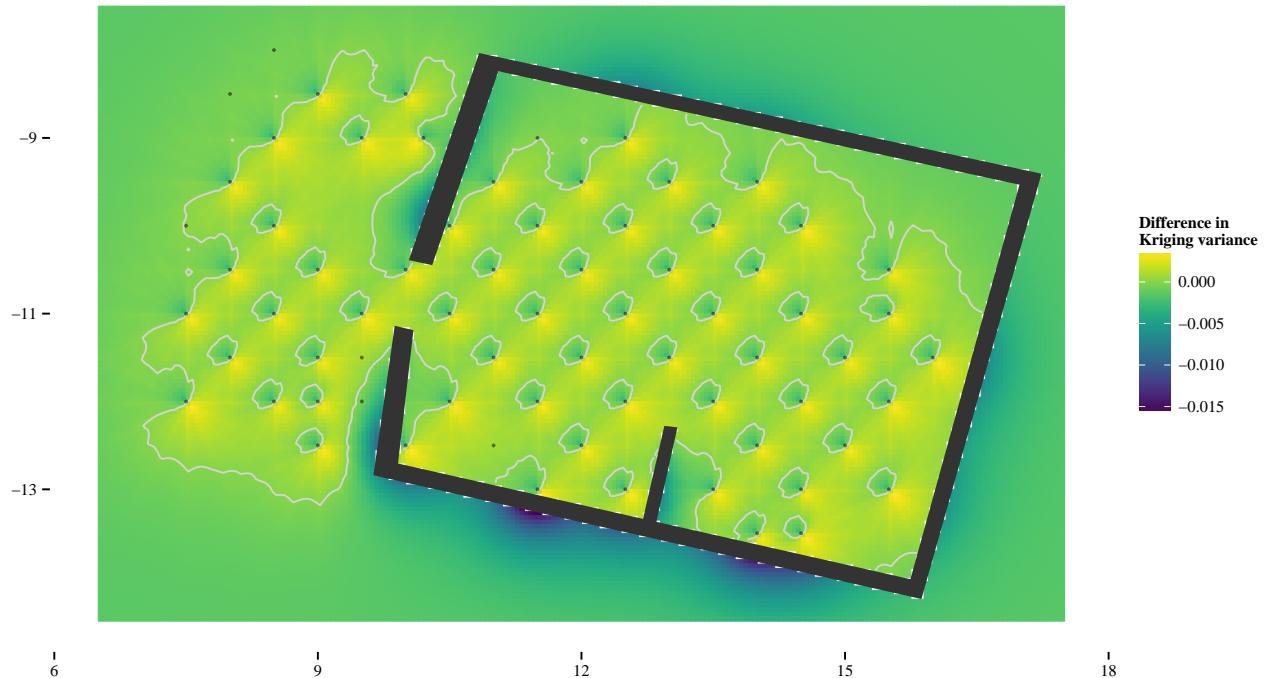


Figure 91: Difference between the Euclidean and the cost-based prediction errors

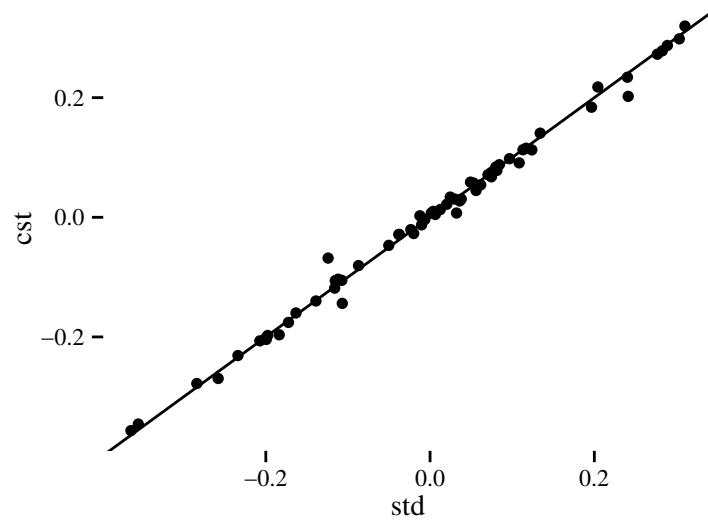


Figure 92: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	0.16
cst	0.16

## 11 Analysis of enclosed spaces

All these results look very much like Calcium results, at a different scale.

### 11.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

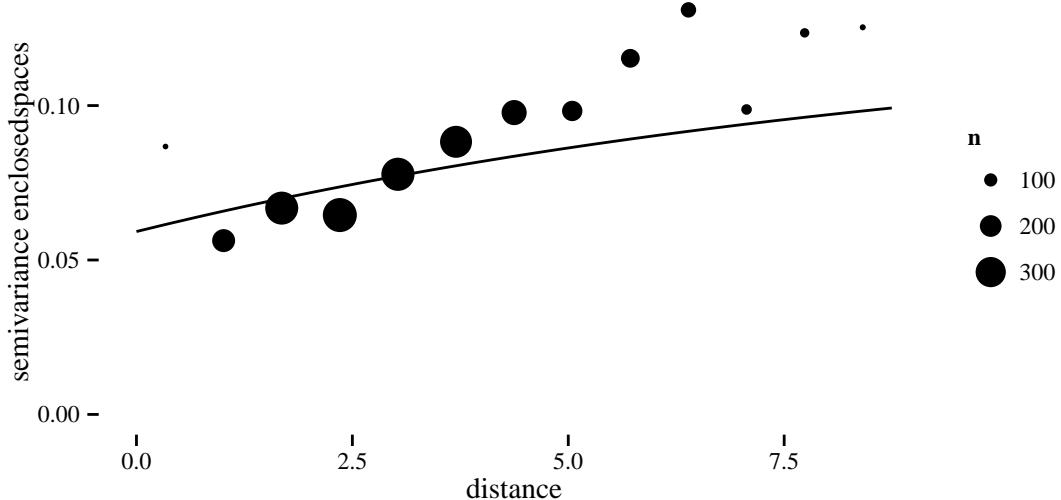


Figure 93: Empirical variogram and fitted model.

### 11.2 Cost-based kriging

### 11.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	1.03	1.04
Nugget	0.06	0.06
Partial sill	0.07	0.18
phi	9.84	33.34
Pract. range	29.46	99.87
Log-likelihood	-6.65	-6.69

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 11.4 Leave-one-out Cross Validation (LOOCV)

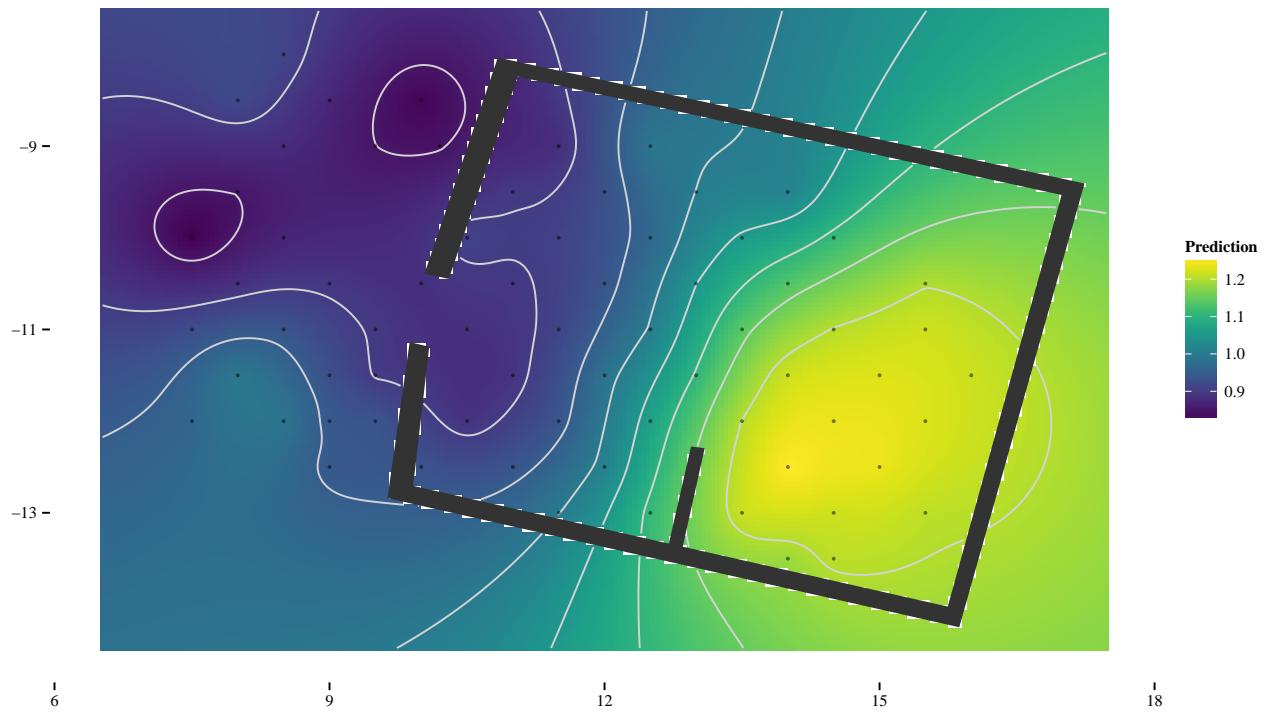


Figure 94: Euclidean kriging prediction

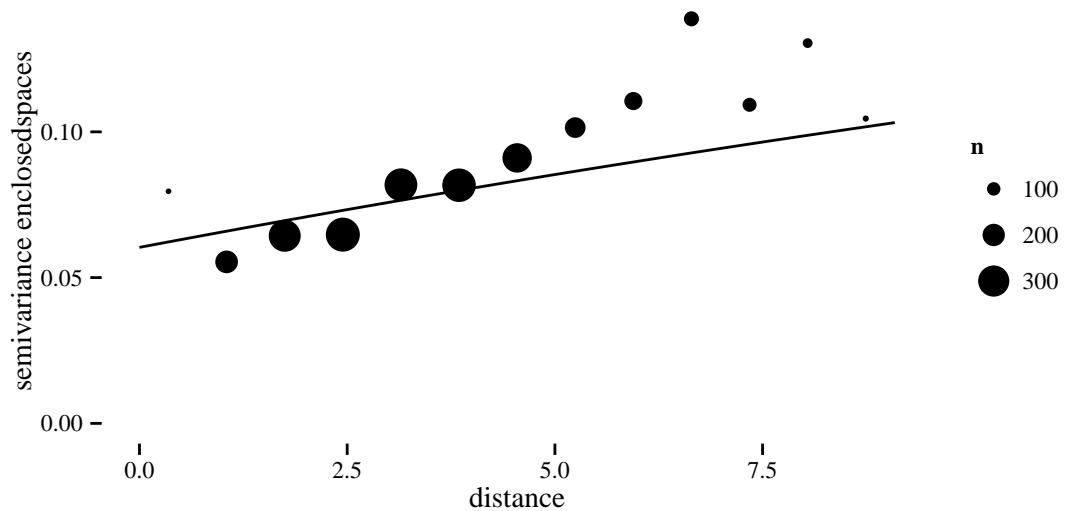


Figure 95: Empirical cost-based variogram and fitted model.

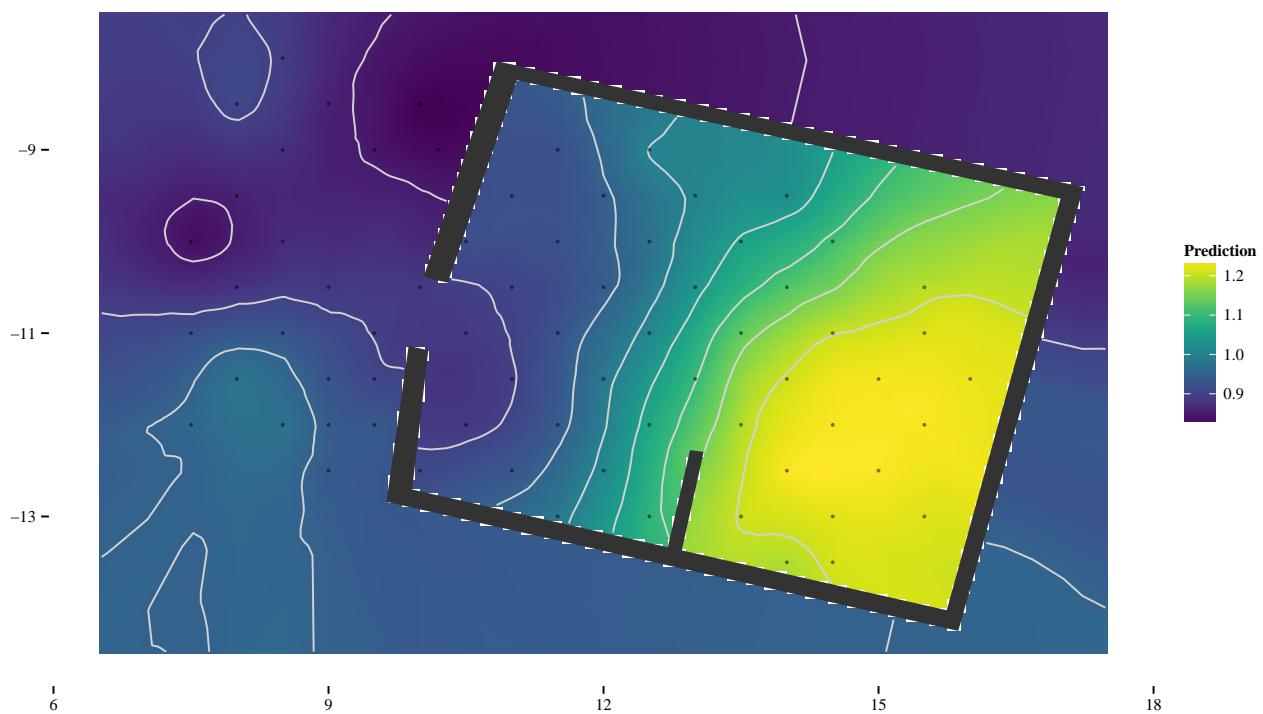


Figure 96: Cost-based kriging prediction

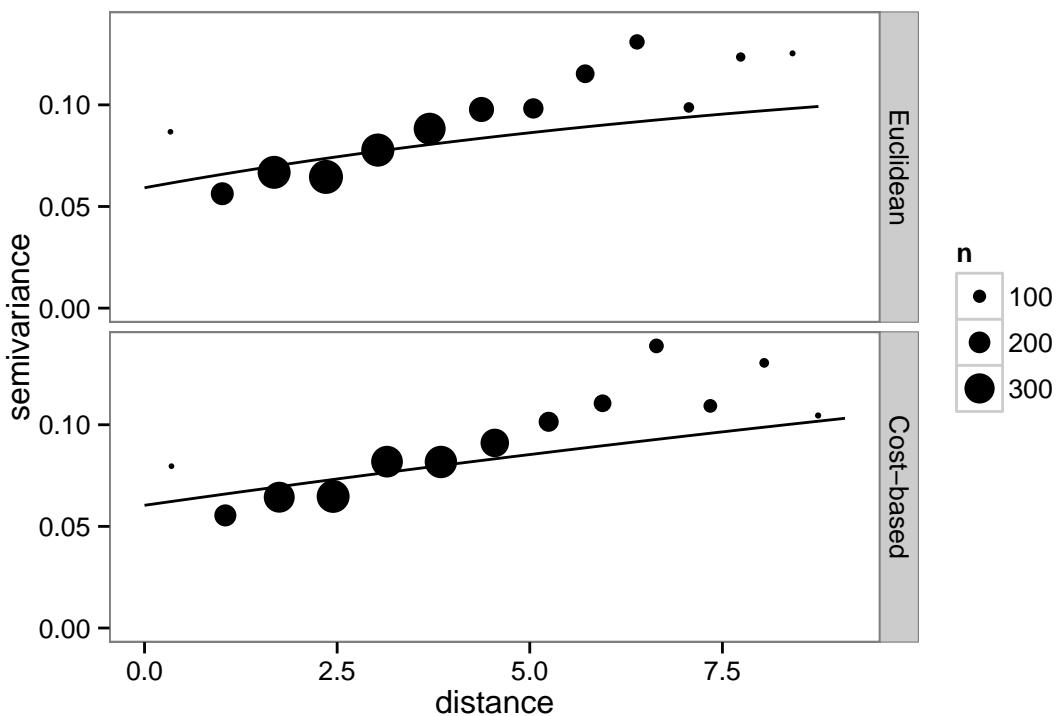


Figure 97: Empirical variogram and fitted models by method for enclosed spaces.

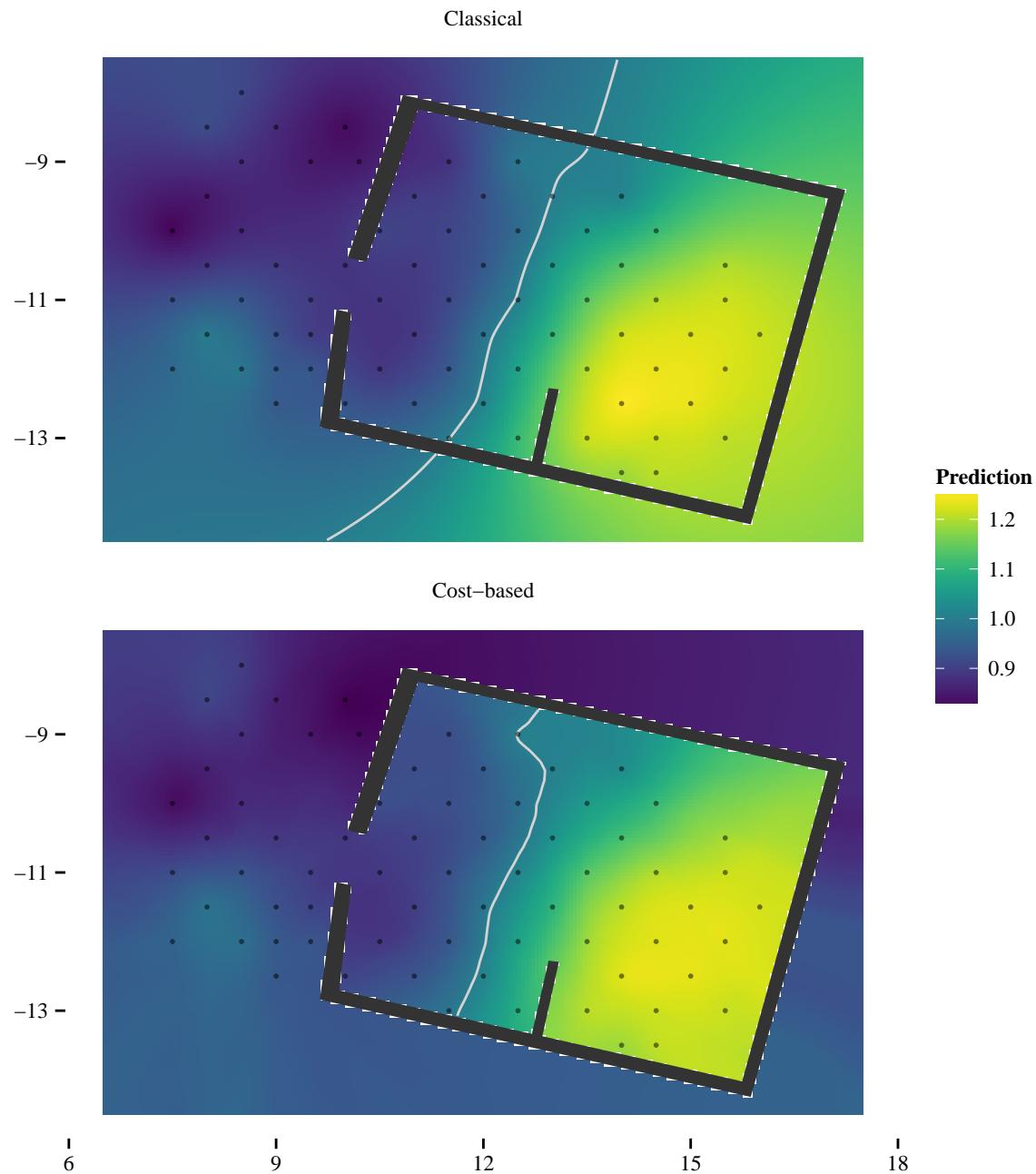


Figure 98: Comparison of Kriging estimates.

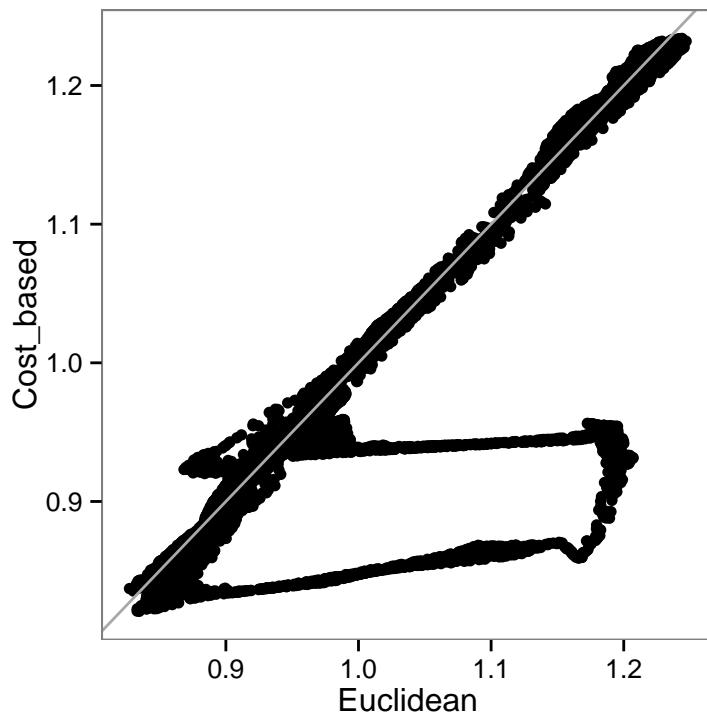


Figure 99: Pointwise comparison of predictions by method.

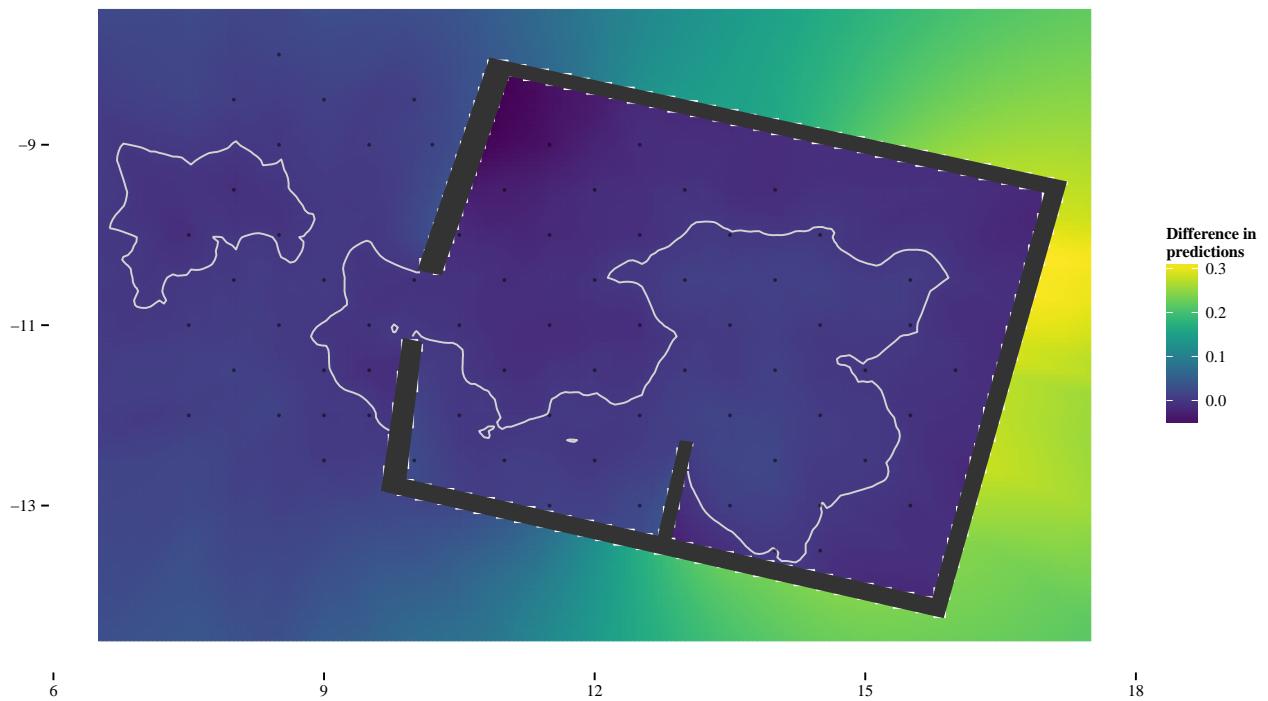


Figure 100: Difference between the Euclidean and the cost-based predictions.

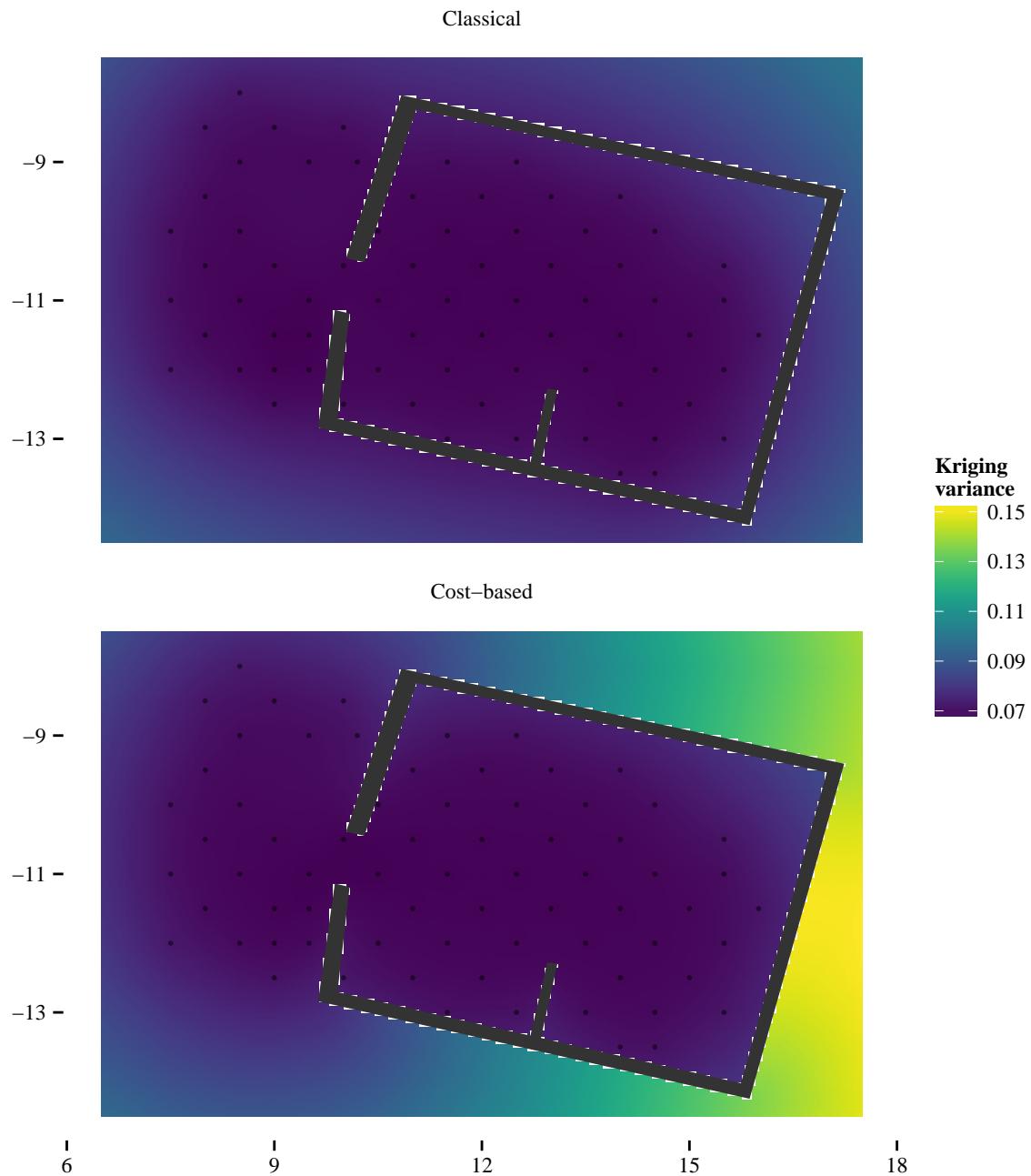


Figure 101: Comparison of prediction error by method.

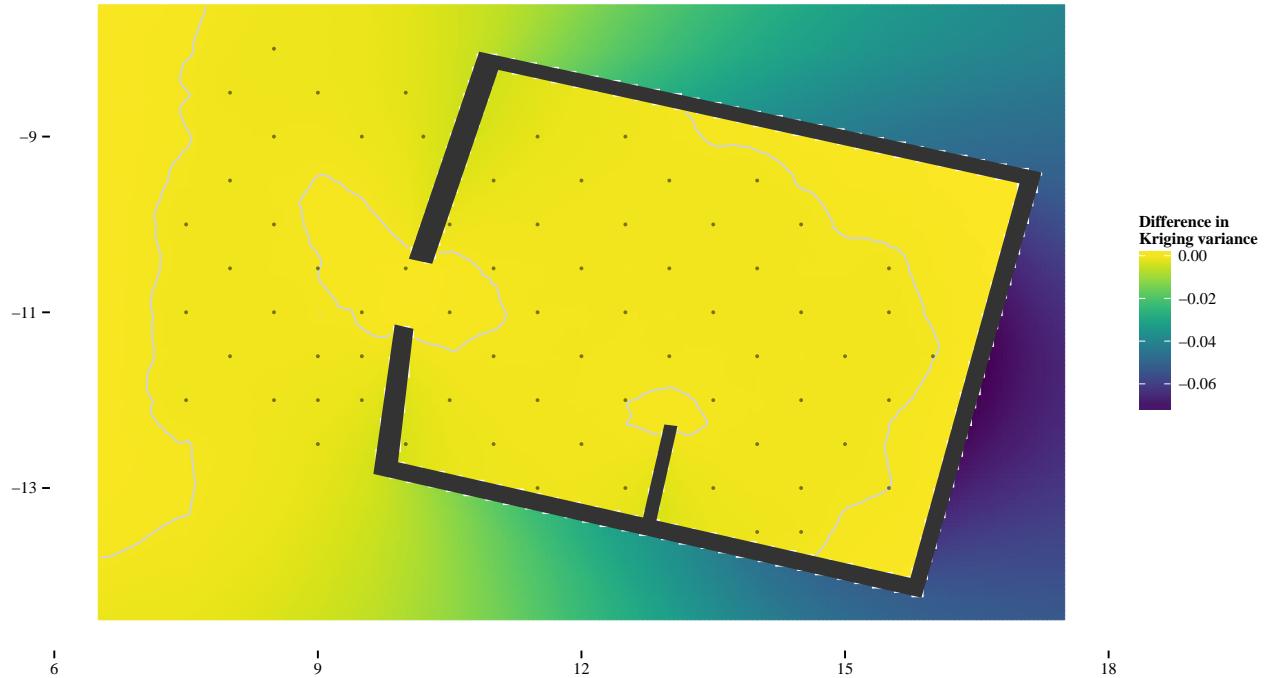


Figure 102: Difference between the Euclidean and the cost-based prediction errors

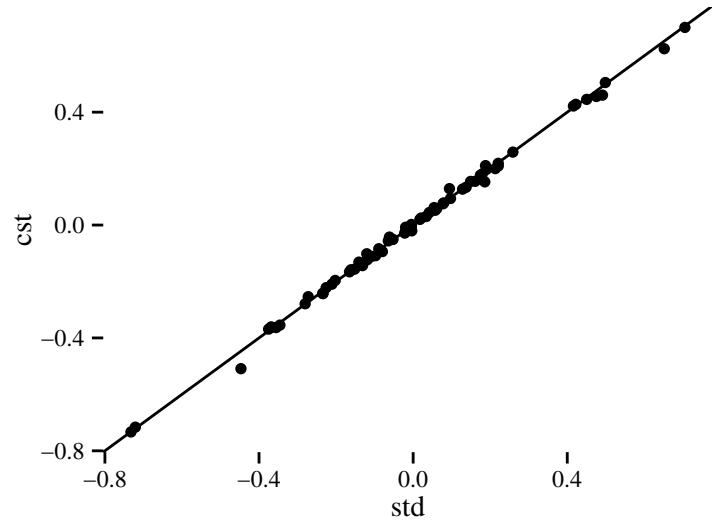


Figure 103: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	0.27
cst	0.27

## 12 Analysis of burningareas

as.geodata: 5 points removed due to NA in the data

### 12.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

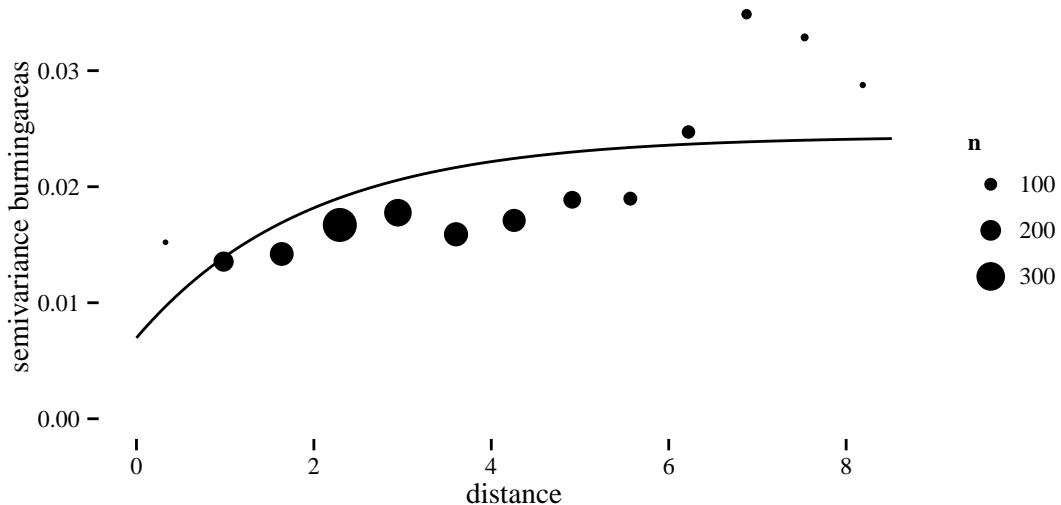


Figure 104: Empirical variogram and fitted model.

### 12.2 Cost-based kriging

### 12.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	0.83	0.80
Nugget	0.01	0.01
Partial sill	0.02	0.02
phi	1.94	1.85
Pract. range	5.80	5.55
Log-likelihood	43.09	42.91

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 12.4 Leave-one-out Cross Validation (LOOCV)

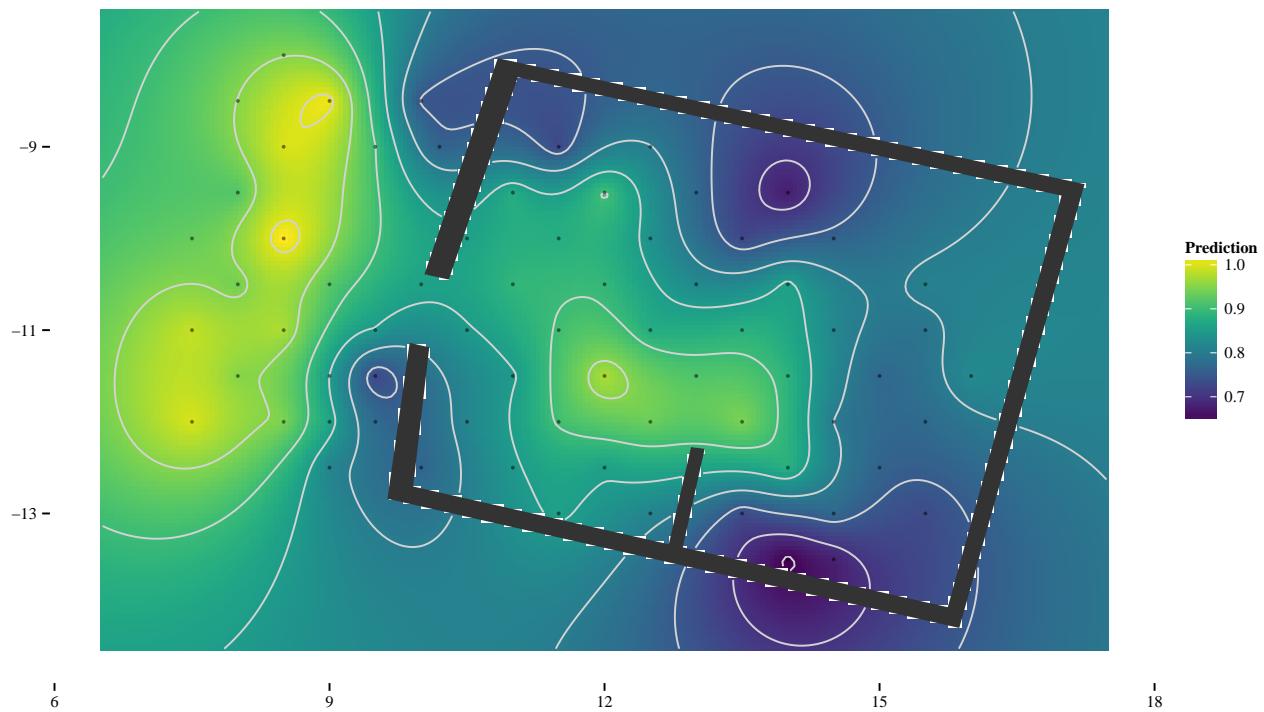


Figure 105: Euclidean kriging prediction

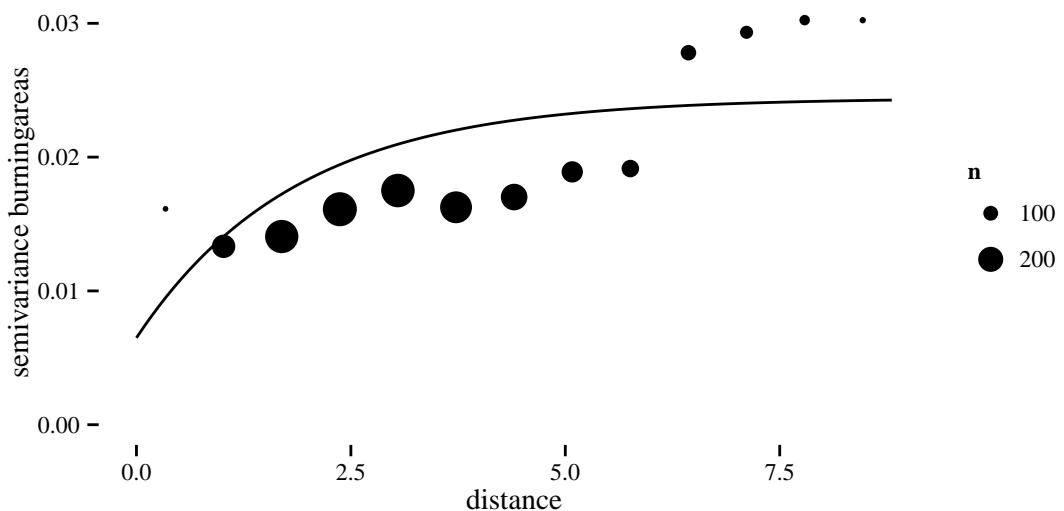


Figure 106: Empirical cost-based variogram and fitted model.

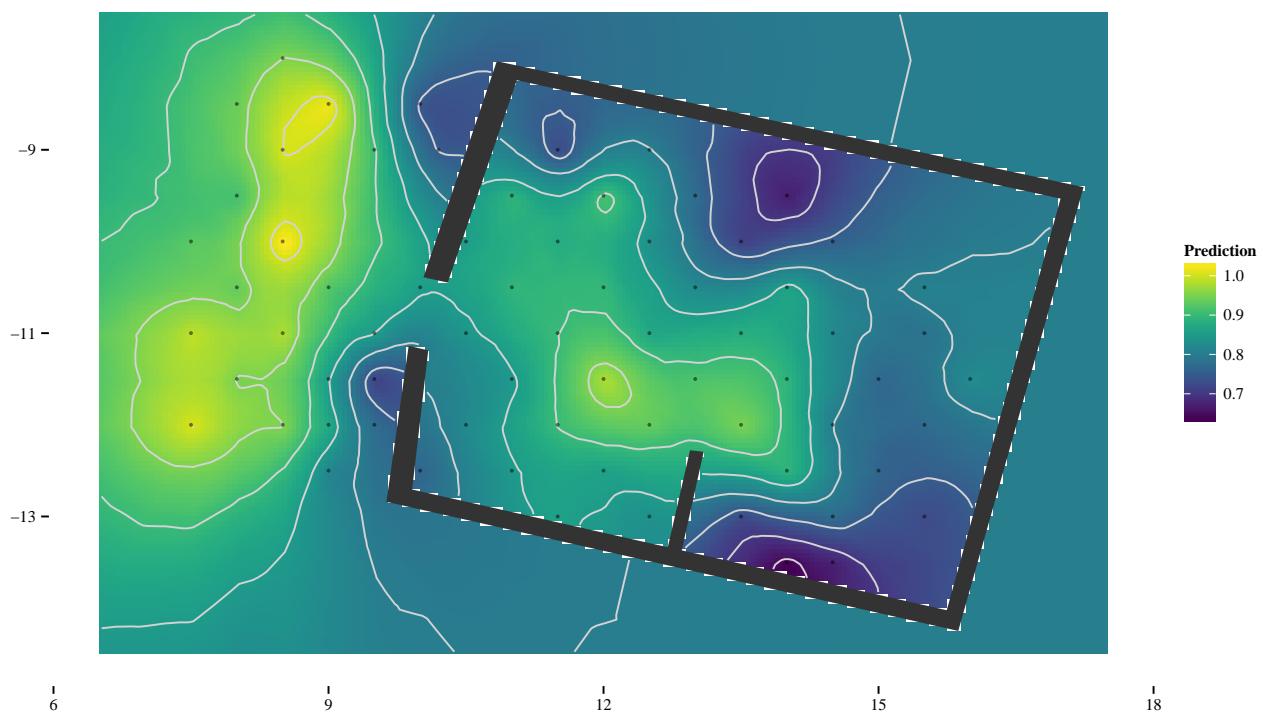


Figure 107: Cost-based kriging prediction

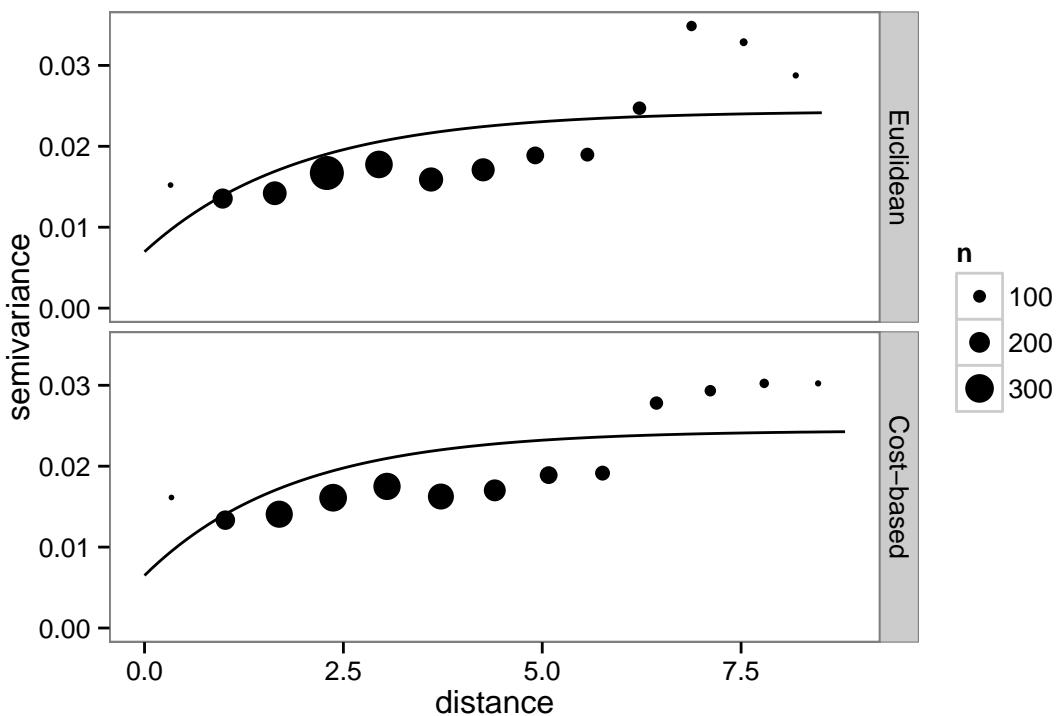


Figure 108: Empirical variogram and fitted models by method for burningareas.

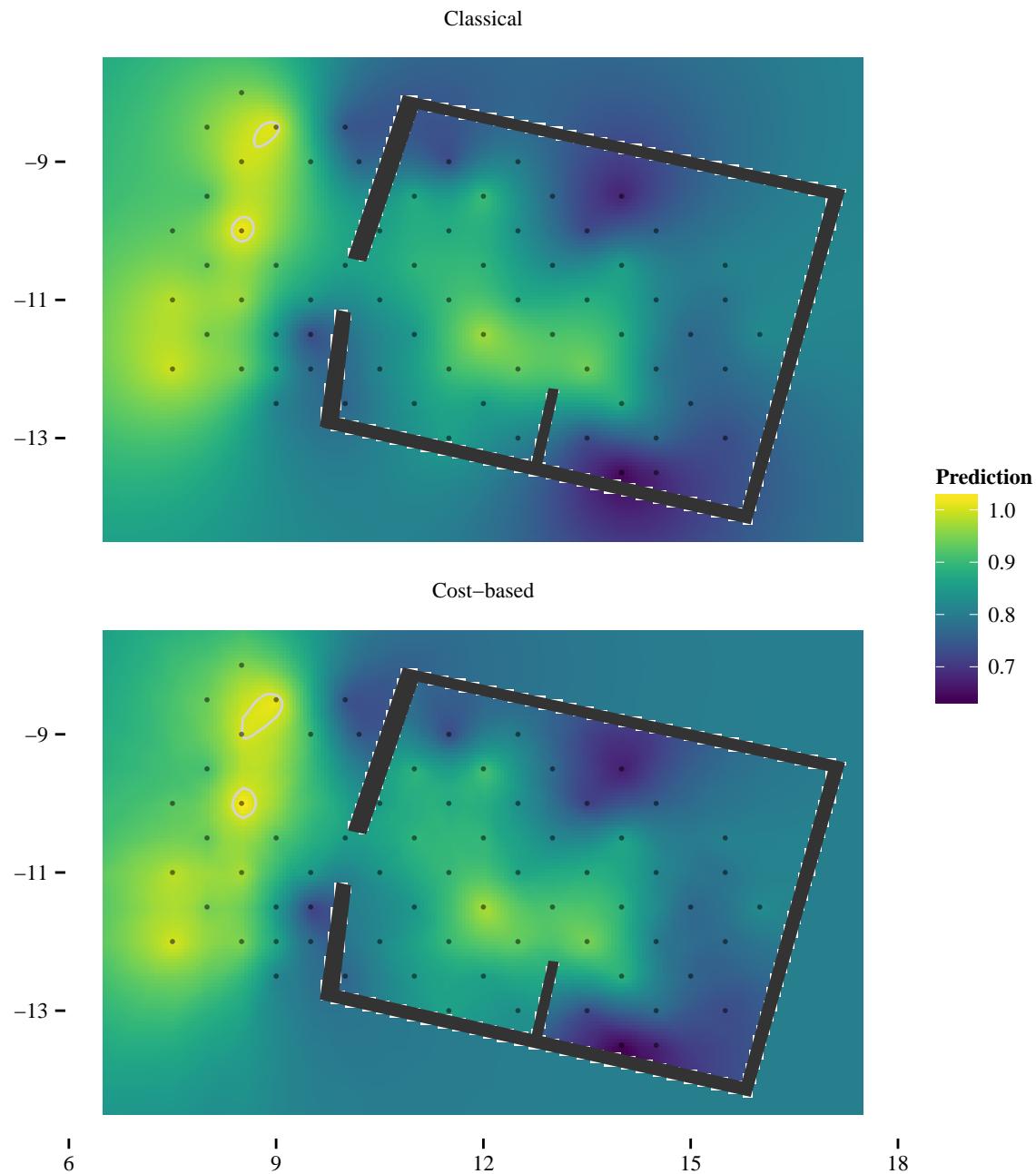


Figure 109: Comparison of Kriging estimates.

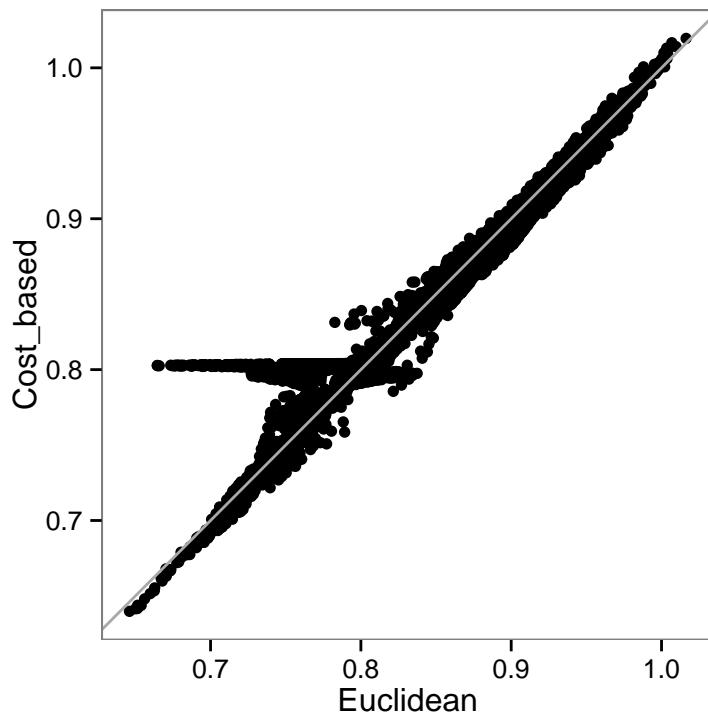


Figure 110: Pointwise comparison of predictions by method.

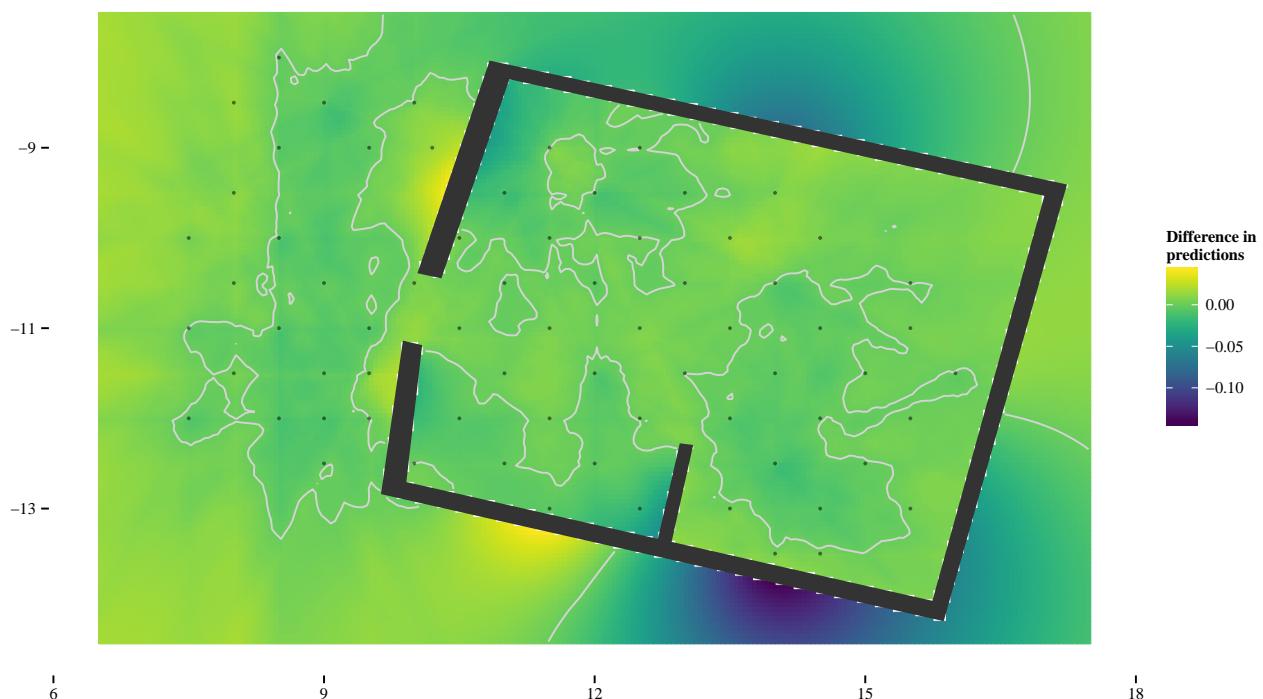


Figure 111: Difference between the Euclidean and the cost-based predictions.

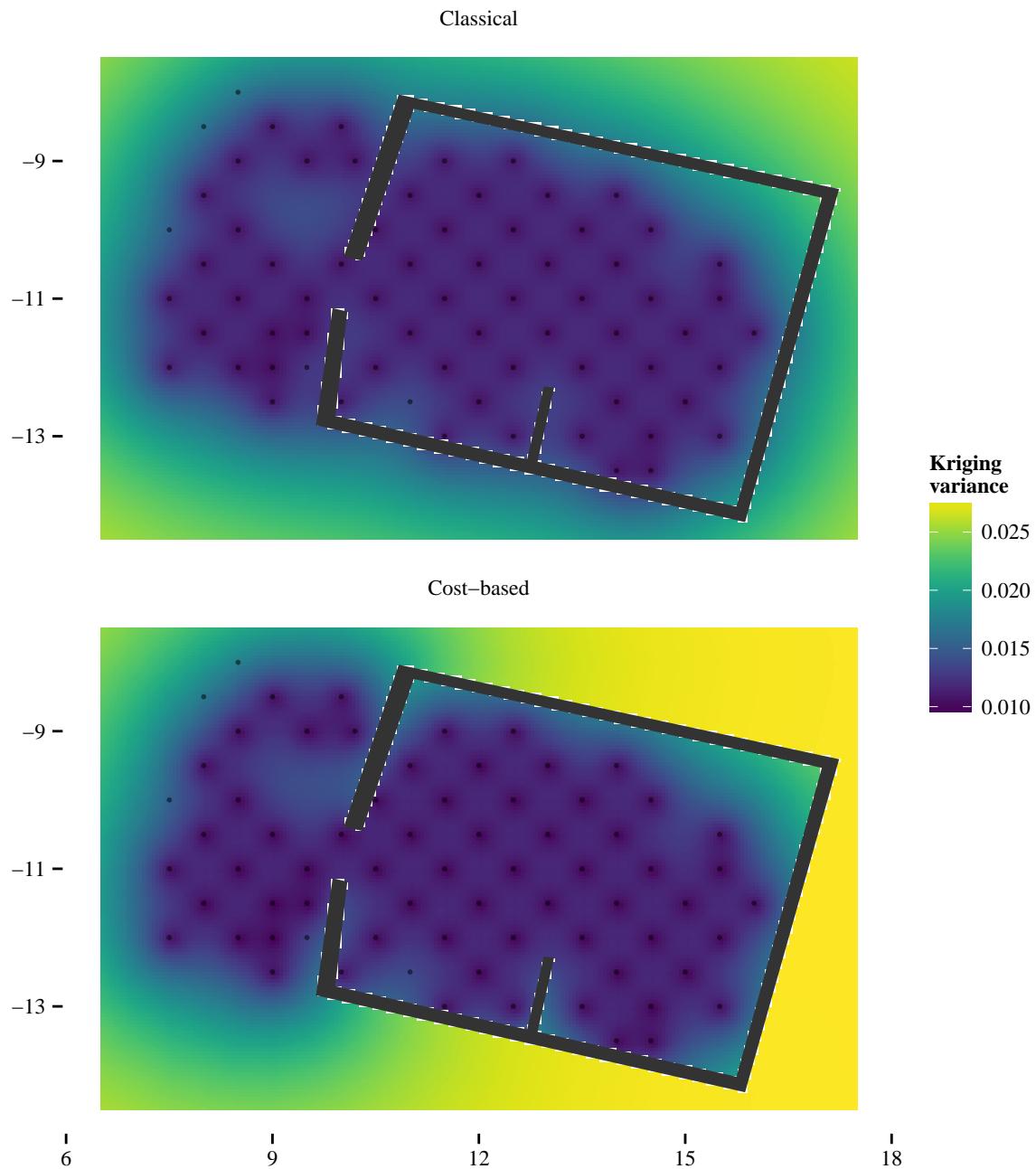


Figure 112: Comparison of prediction error by method.

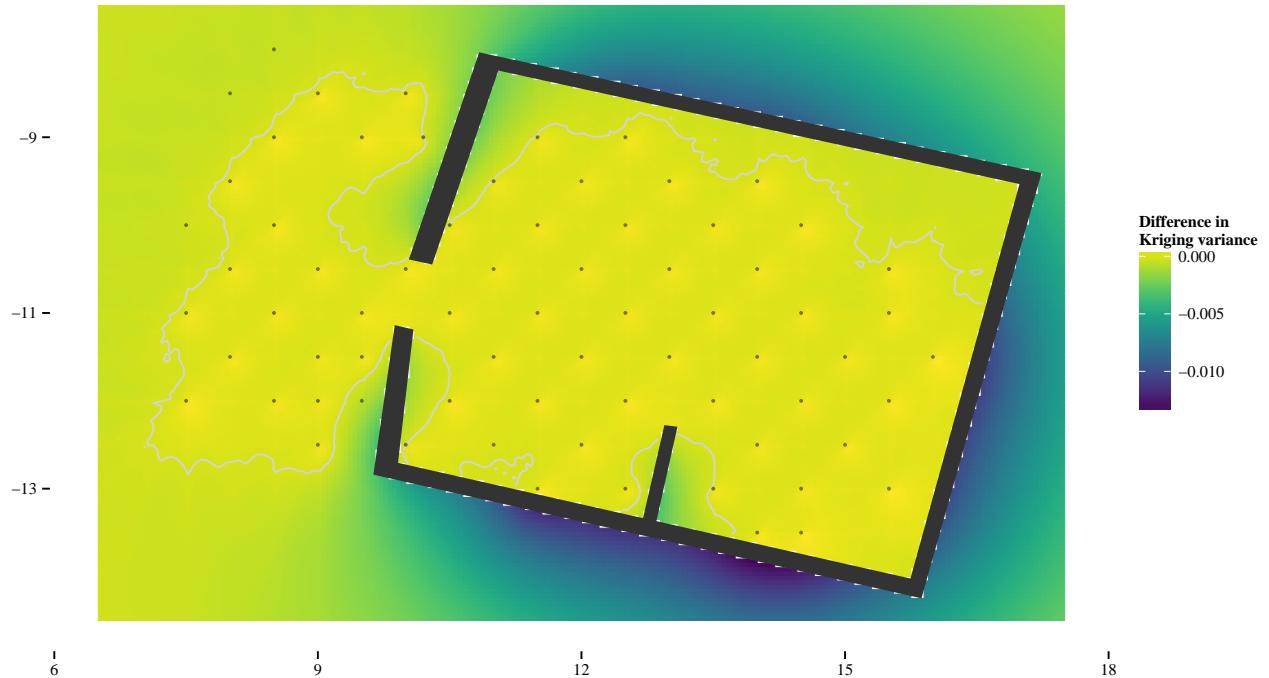


Figure 113: Difference between the Euclidean and the cost-based prediction errors

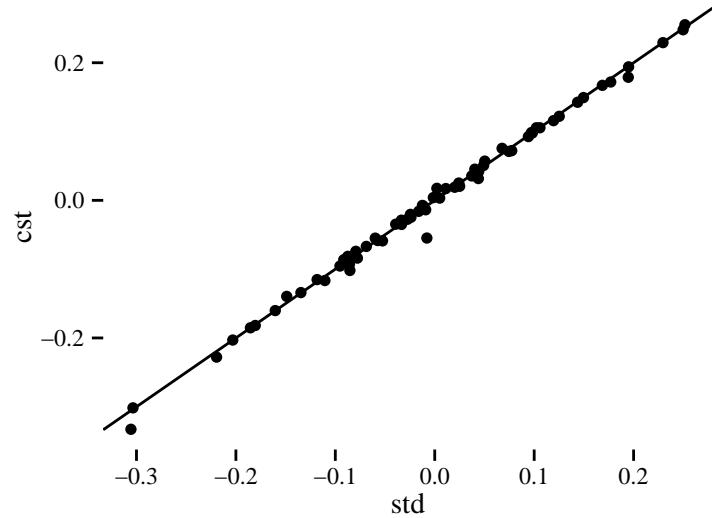


Figure 114: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	0.12
cst	0.12

## 13 Analysis of ashes

as.geodata: 4 points removed due to NA in the data

### 13.1 Euclidean kriging

The variogram model is Exponential. We choose to estimate the nugget effect, which may account for measurement error, for example.

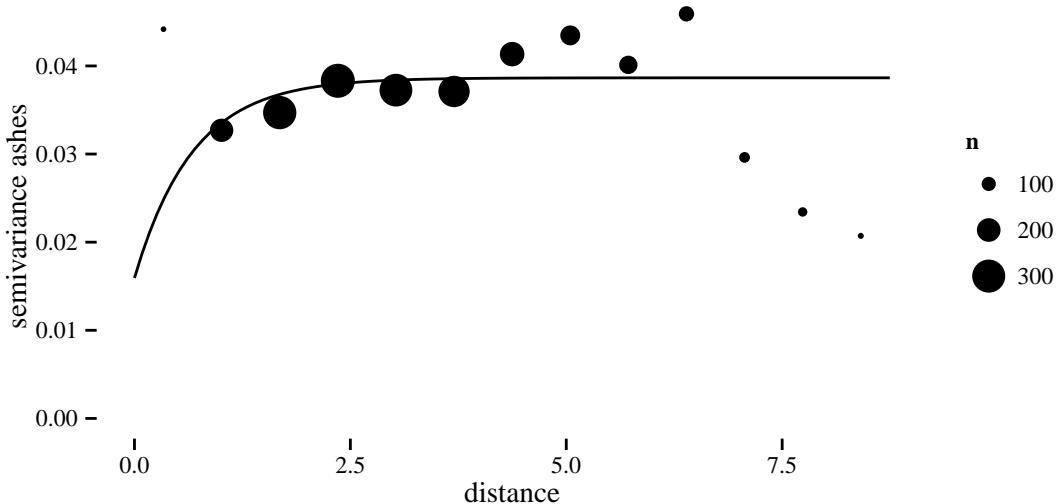


Figure 115: Empirical variogram and fitted model.

### 13.2 Cost-based kriging

### 13.3 Comparison of method outcomes

	Euclidean	Cost_based
Intercept	1.40	1.39
Nugget	0.02	0.02
Partial sill	0.02	0.02
phi	0.68	1.10
Pract. range	2.02	3.29
Log-likelihood	16.51	16.39

In the scatter plot, the horizontal patterns correspond to predictions on observed values. Otherwise, the differences are negligible.

Near the observations, the cost-based approach has a larger prediction error due to its increased estimation of the nugget (i.e. short-range variance). In the main area, the prediction errors are practically the same with both approaches. Behind the walls, the Euclidean prediction error is unrealistically low.

### 13.4 Leave-one-out Cross Validation (LOOCV)

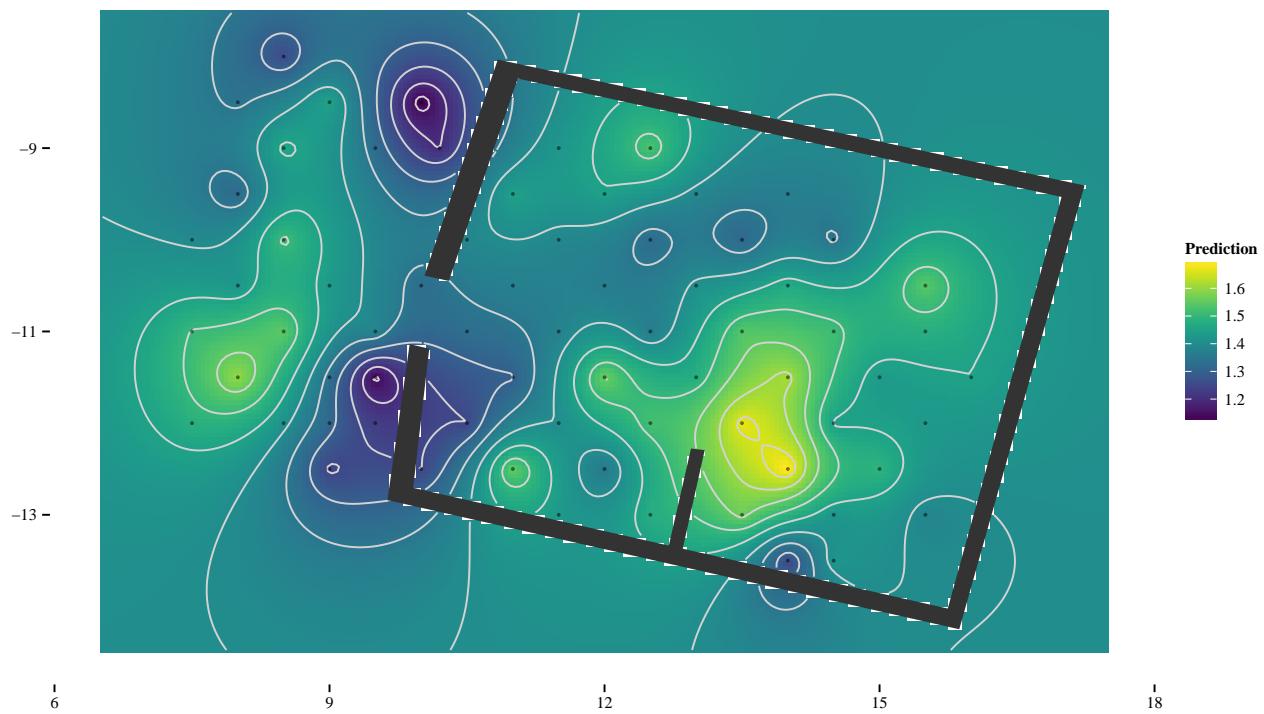


Figure 116: Euclidean kriging prediction

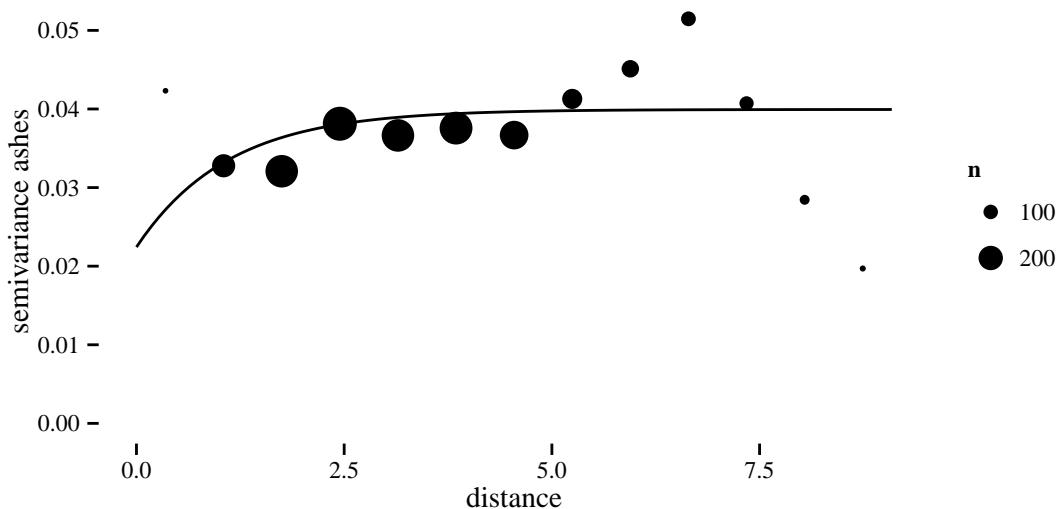


Figure 117: Empirical cost-based variogram and fitted model.

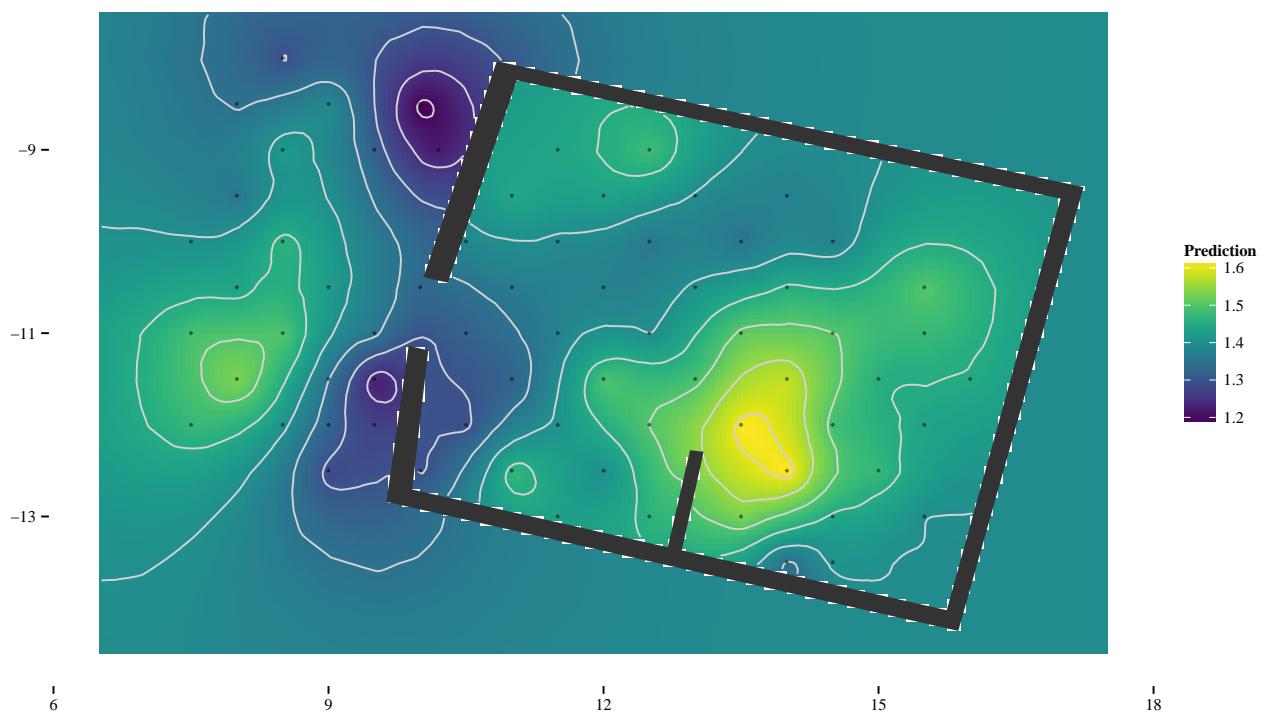


Figure 118: Cost-based kriging prediction

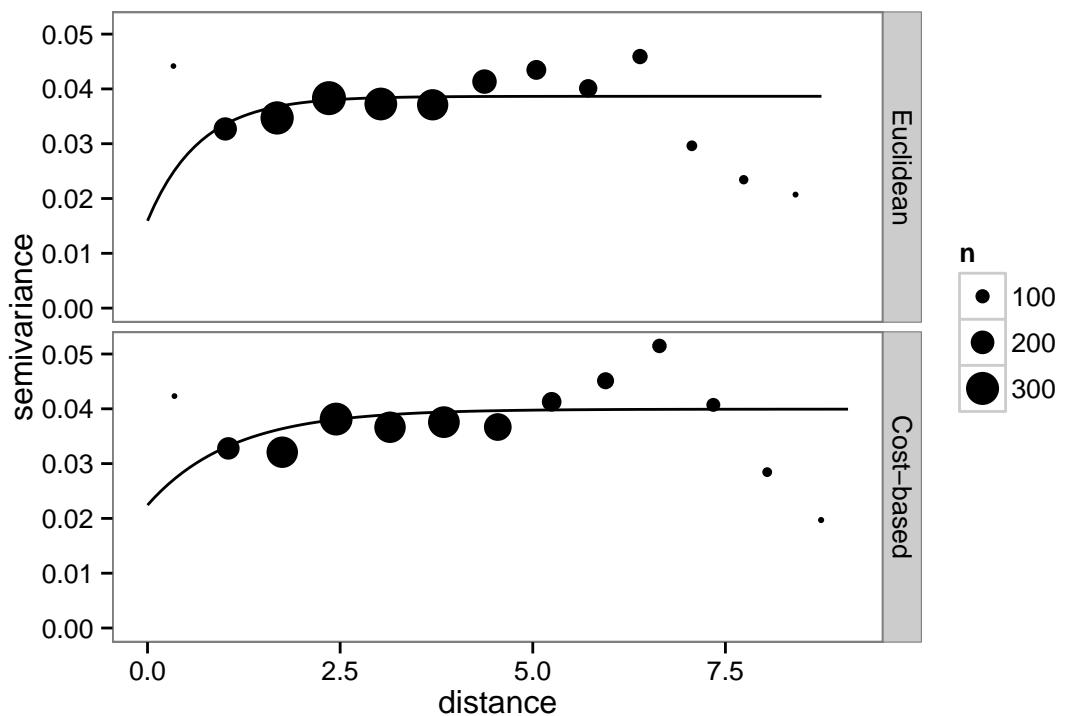


Figure 119: Empirical variogram and fitted models by method for ashes.

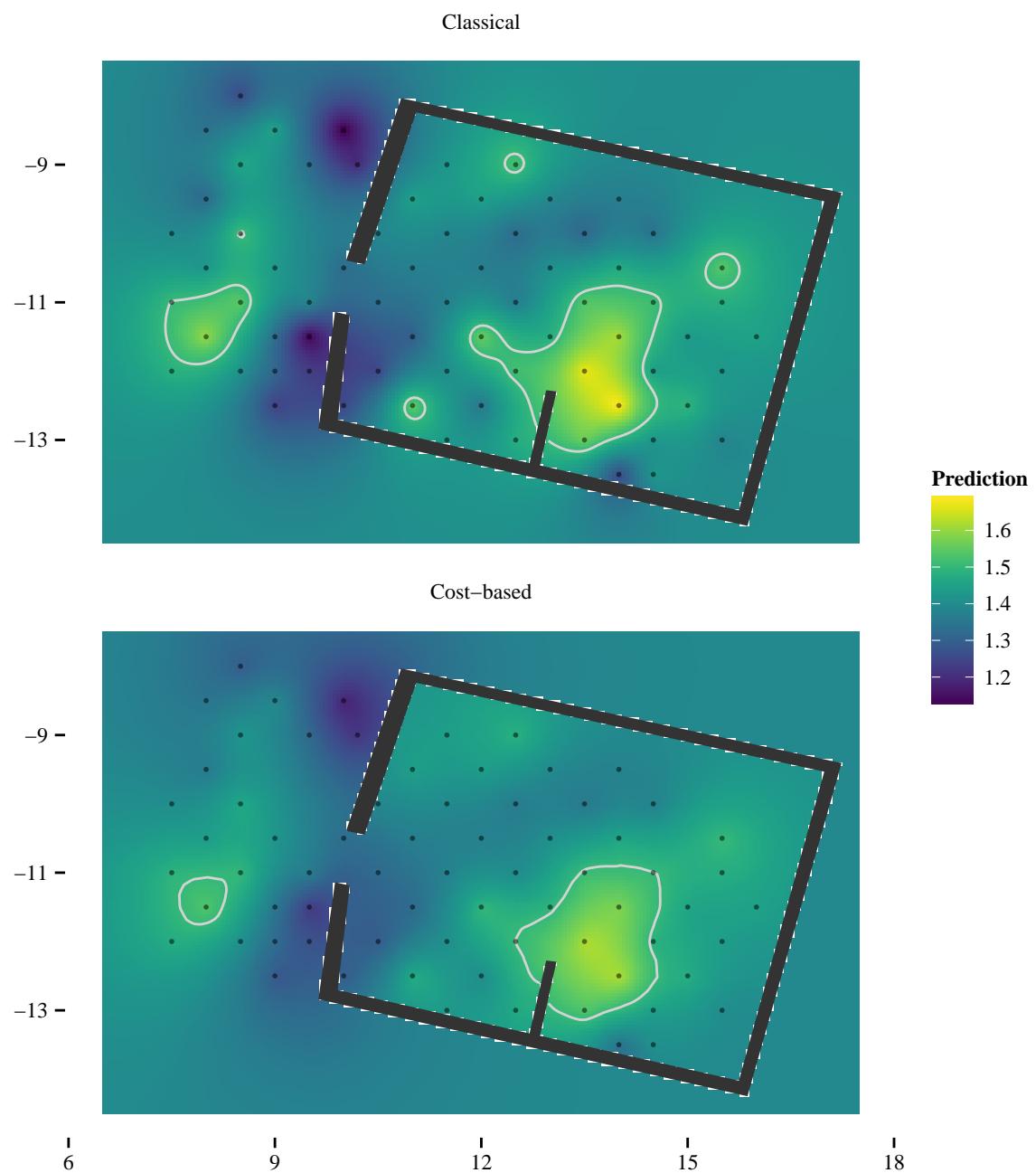


Figure 120: Comparison of Kriging estimates.

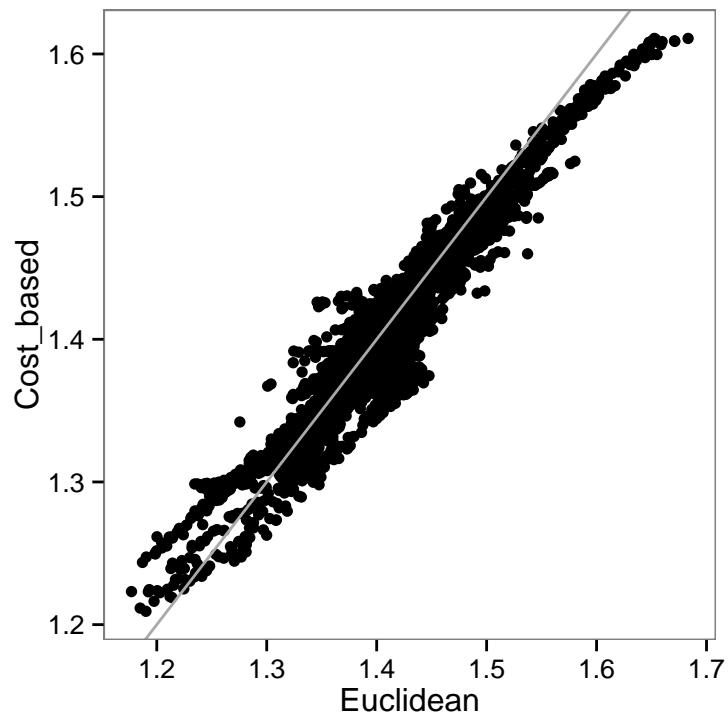


Figure 121: Pointwise comparison of predictions by method.

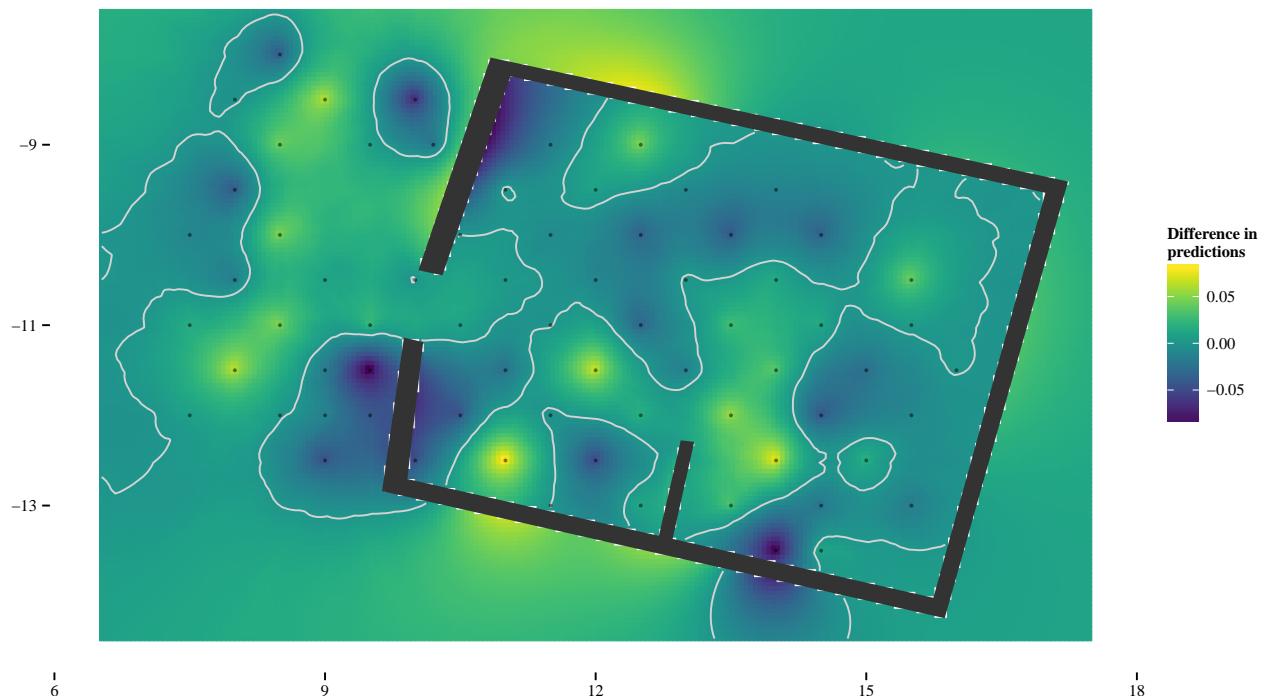


Figure 122: Difference between the Euclidean and the cost-based predictions.

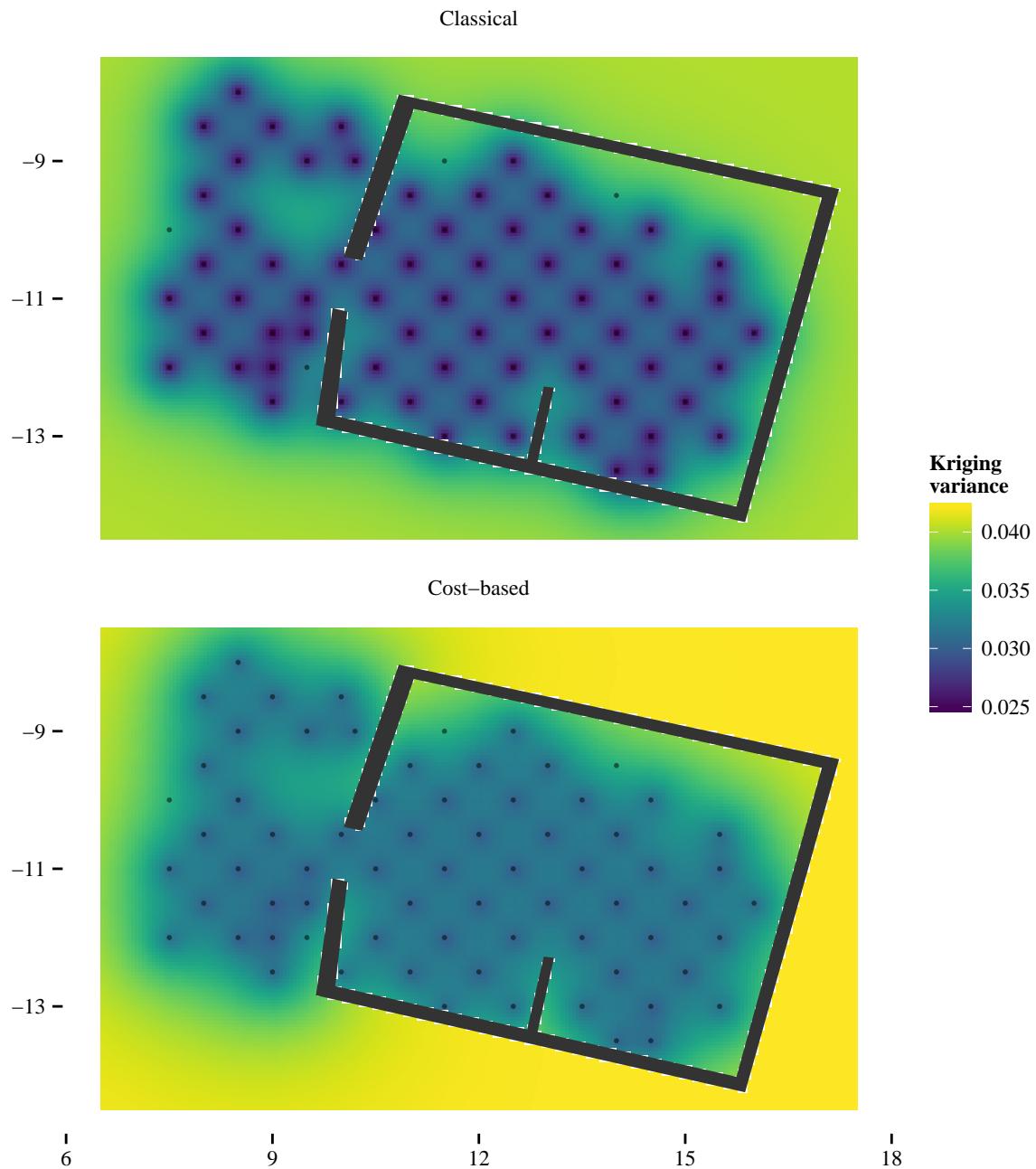


Figure 123: Comparison of prediction error by method.

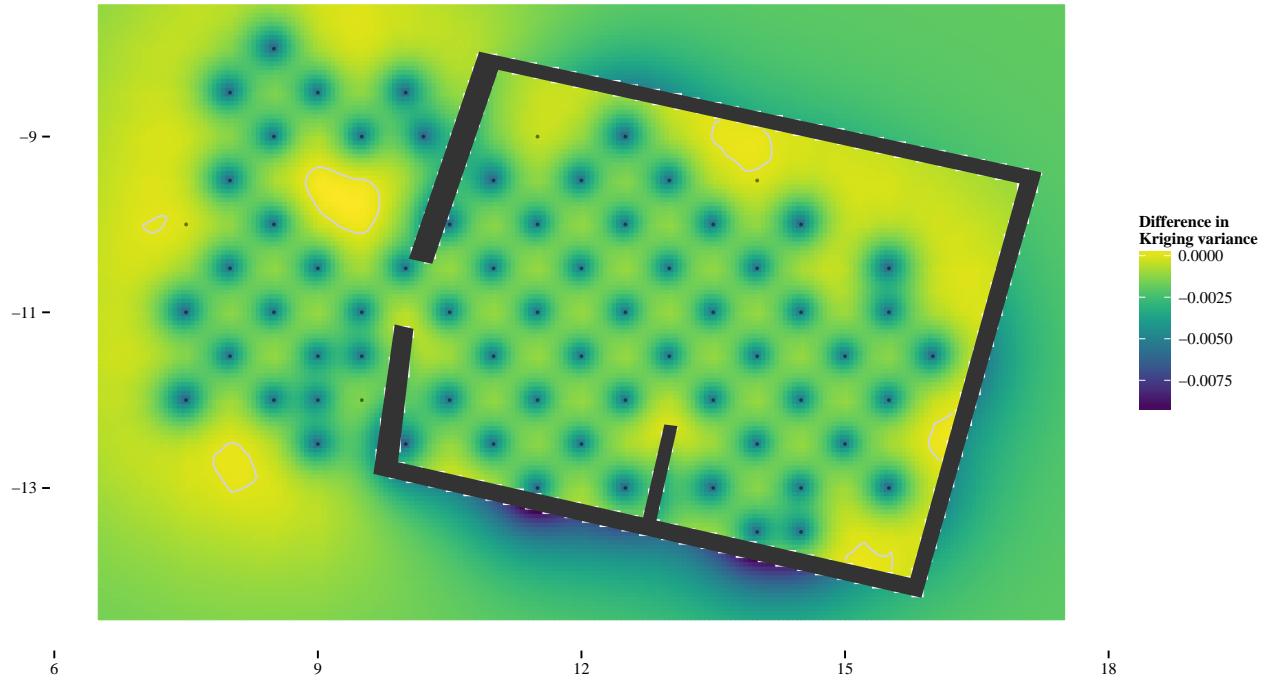


Figure 124: Difference between the Euclidean and the cost-based prediction errors

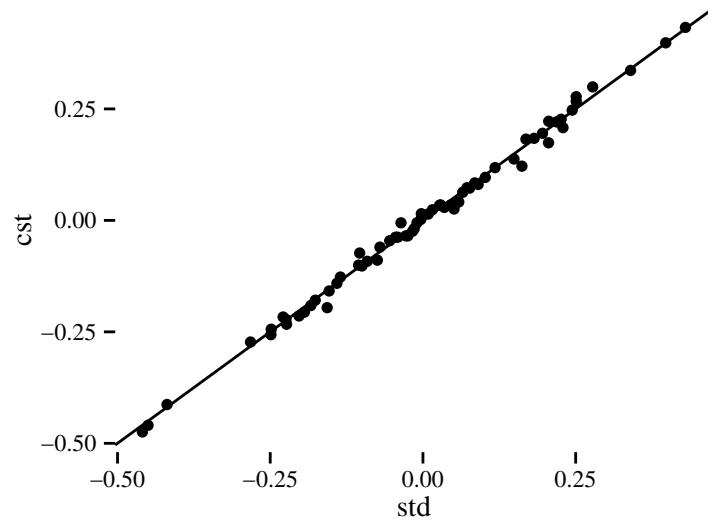


Figure 125: Pointwise leave-one-out prediction error by method.

method	rmse(error)
std	0.19
cst	0.19

## 14 Conclusions

- Actually, the kriging model for Calcium is not adjusting very well the tails of the data, which are heavier than expected. This happens both for the Euclidean and cost-based models. This means that none of both approaches will be really good predictors anyways.

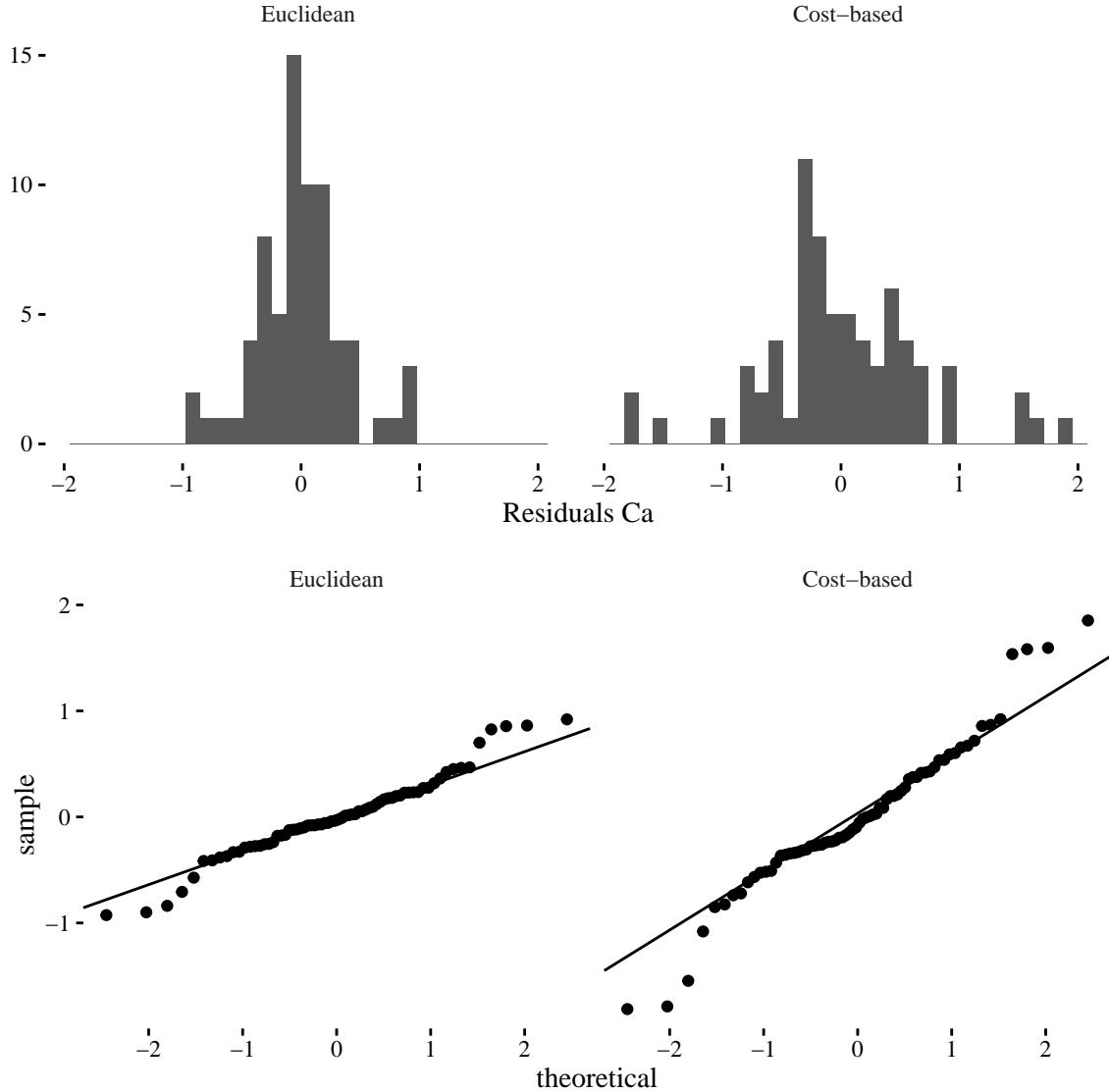


Figure 126: Diagnostics of residuals for Calcium.

- The rest of the variables do not show a clear discontinuity as a consequence of the walls. The Euclidean and cost-based predictions are very similar. Furthermore, the Euclidean and cost-based empirical variograms display practically the same shape in all cases. This suggests that the solid structures are not really affecting the spatial distribution of chemicals.
- Many variables display an initial drop in the semivariance. Even this is based on only 10 or 12 pairs of observations in the first lag, the drop is dramatic for several variables like Copper, Potassium or Magnesium. This may be due to the presence of some extreme values which contrast heavily with neighbouring values. The impact in higher lags is absorbed by the high number of pairs.