

Analysis of soil at India

Facundo Muñoz

October 7, 2015

Contents

1 Data description	1
2 Cost-based distances	4
3 Analysis of Calcium	5
4 Analysis of Copper	13
5 Conclusions	20

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 5 variables:
$ sample: Factor w/ 70 levels "JIN10","JIN100",...: 28 45 66 1 16 22 25 26 27 29 ...
$ x      : num  13 12.5 13.5 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 ...

      sample      x          y          Ca
JIN10  : 1  Min.   : 7.50  Min.   :-13.50  Min.   :0.660
JIN100 : 1  1st Qu.: 9.50  1st Qu.:-12.00  1st Qu.:2.490
JIN101 : 1  Median :11.50  Median :-11.00  Median :2.945
JIN102 : 1  Mean    :11.53  Mean   :-10.99  Mean   :3.106
JIN103 : 1  3rd Qu.:13.50  3rd Qu.:-10.00  3rd Qu.:3.697
JIN106 : 1  Max.    :16.00  Max.   :- 8.00  Max.   :5.700
(Other):64
      Cu
Min.   : 7.00
1st Qu.:11.00
Median :12.00
Mean   :12.39
3rd Qu.:14.00
Max.   :20.00
```

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

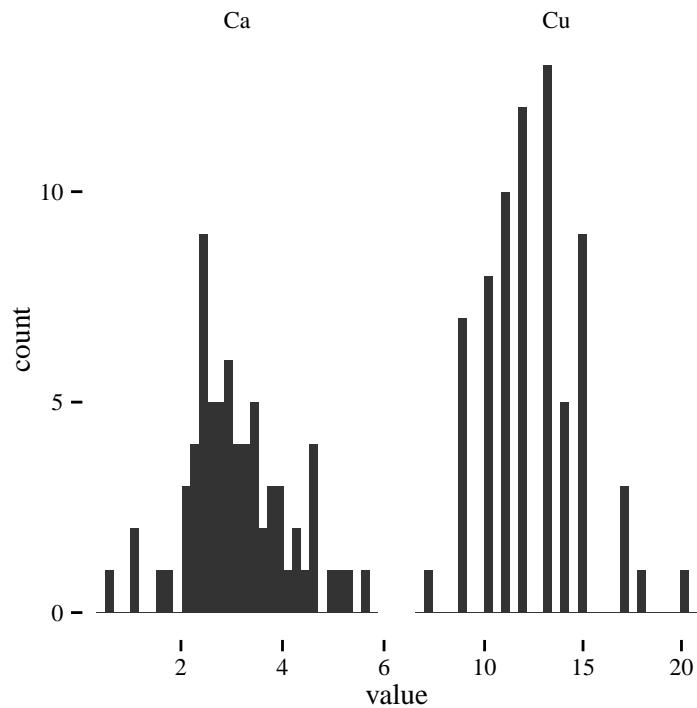


Figure 1:

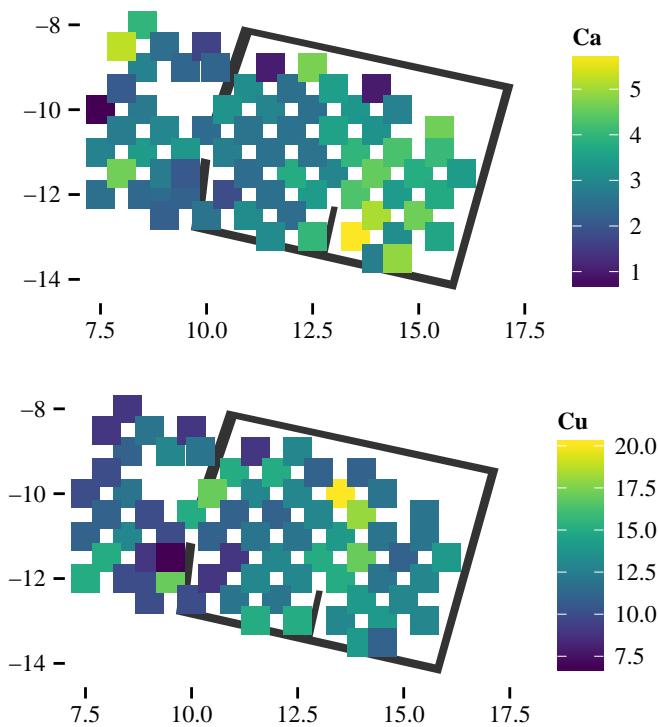


Figure 2: Measurement locations and observed values

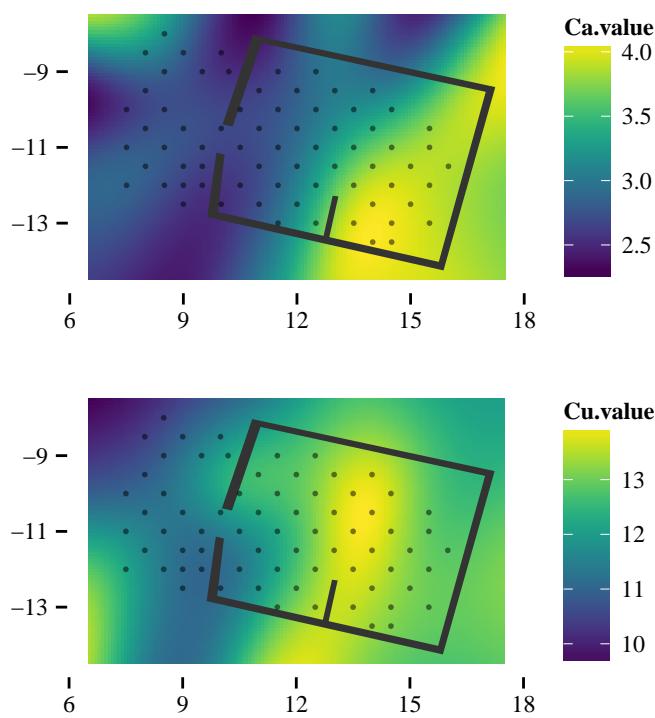


Figure 3: Exploratory kernel smoothing of the measurements

2 Cost-based distances

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

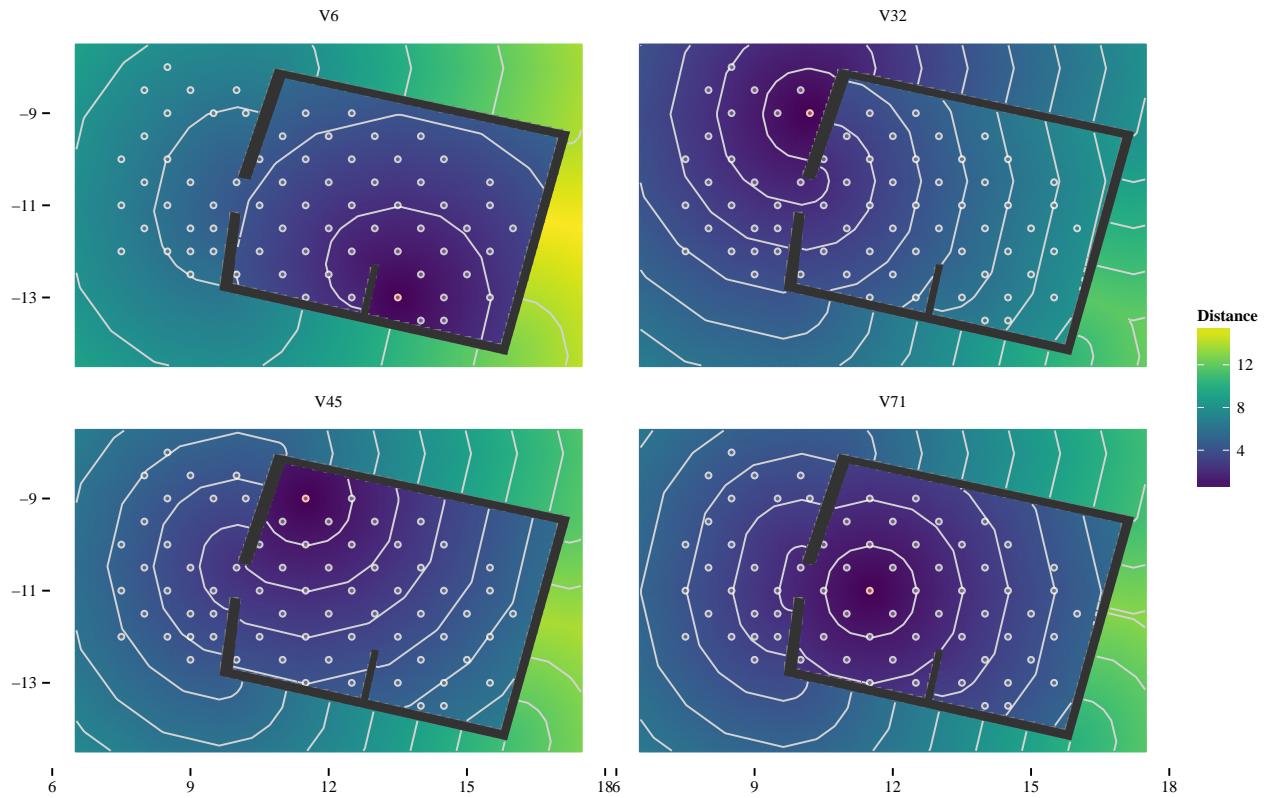


Figure 4: Some cost-based maps to selected observations.

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.

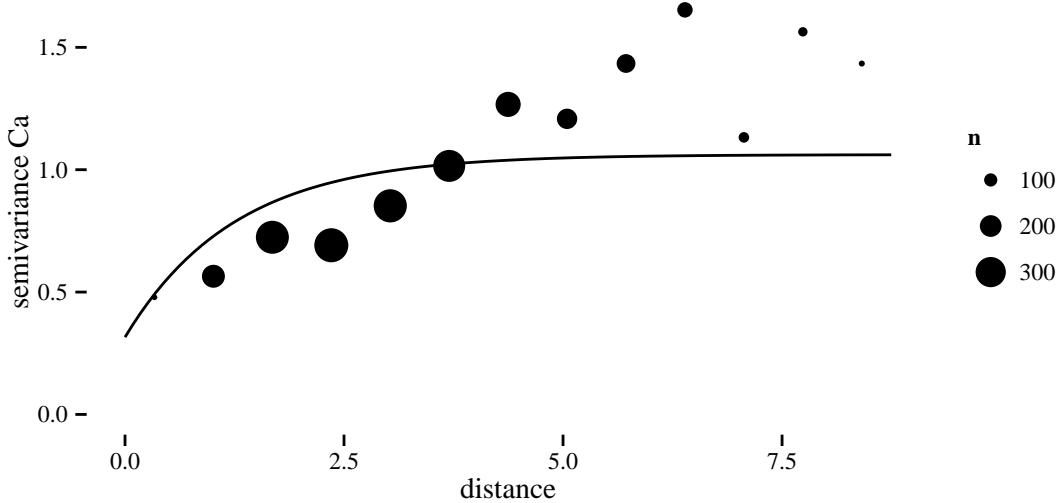


Figure 5: Empirical variogram and fitted model.

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)

method	rmse(error)
std	0.93

method	rmse(error)
cst	0.96

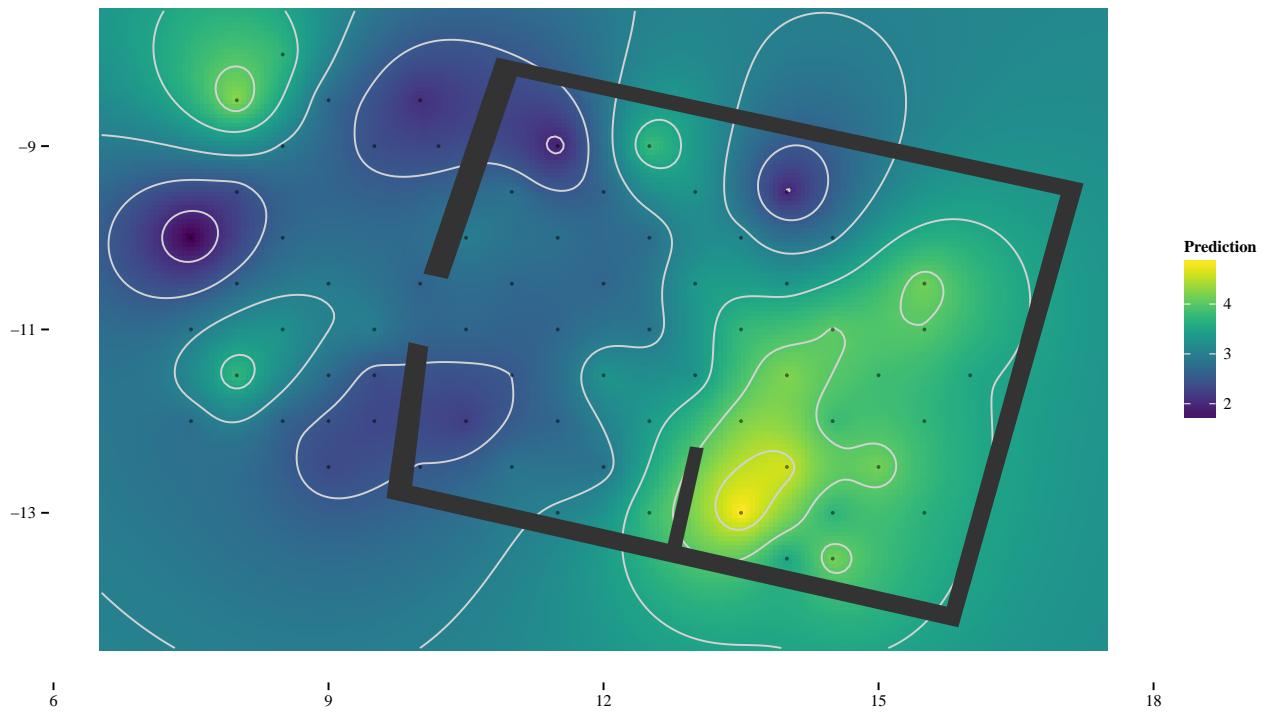


Figure 6: Euclidean kriging prediction

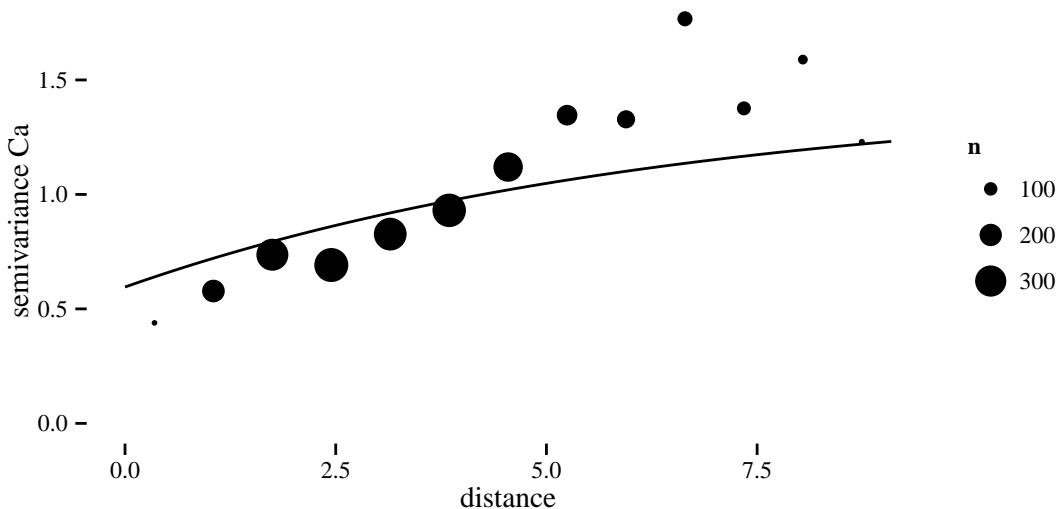


Figure 7: Empirical cost-based variogram and fitted model.

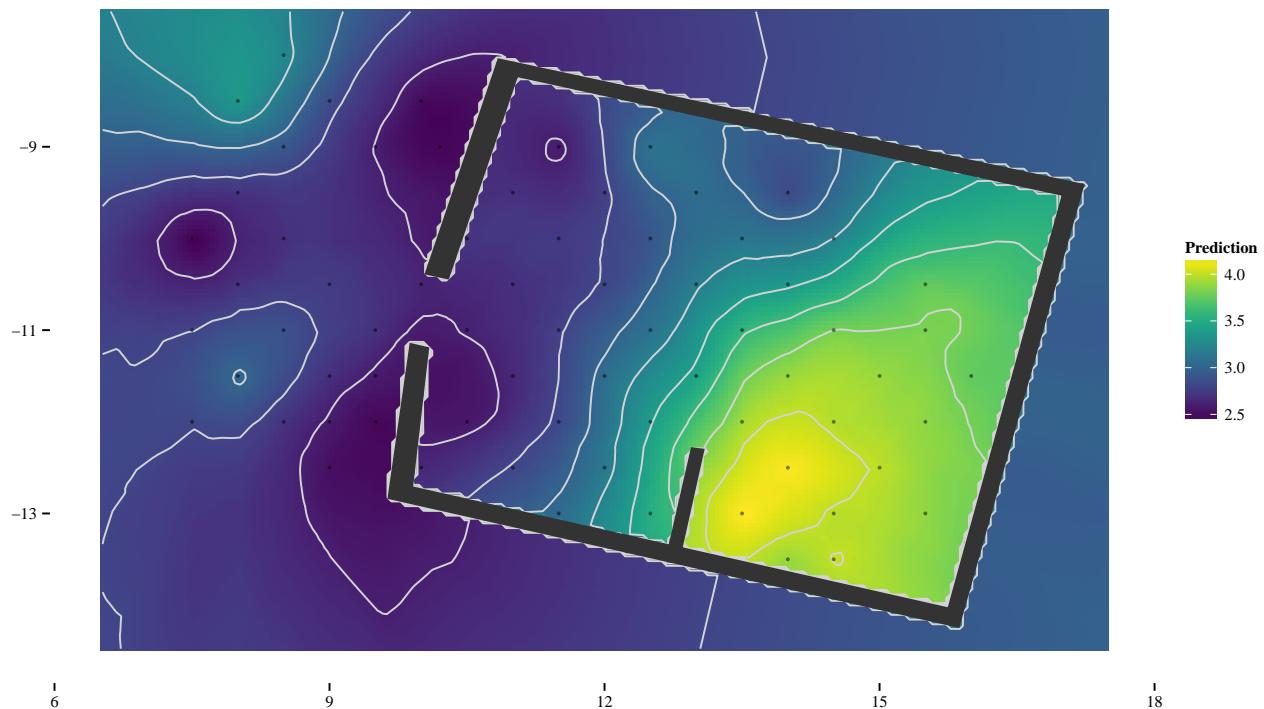


Figure 8: Cost-based kriging prediction

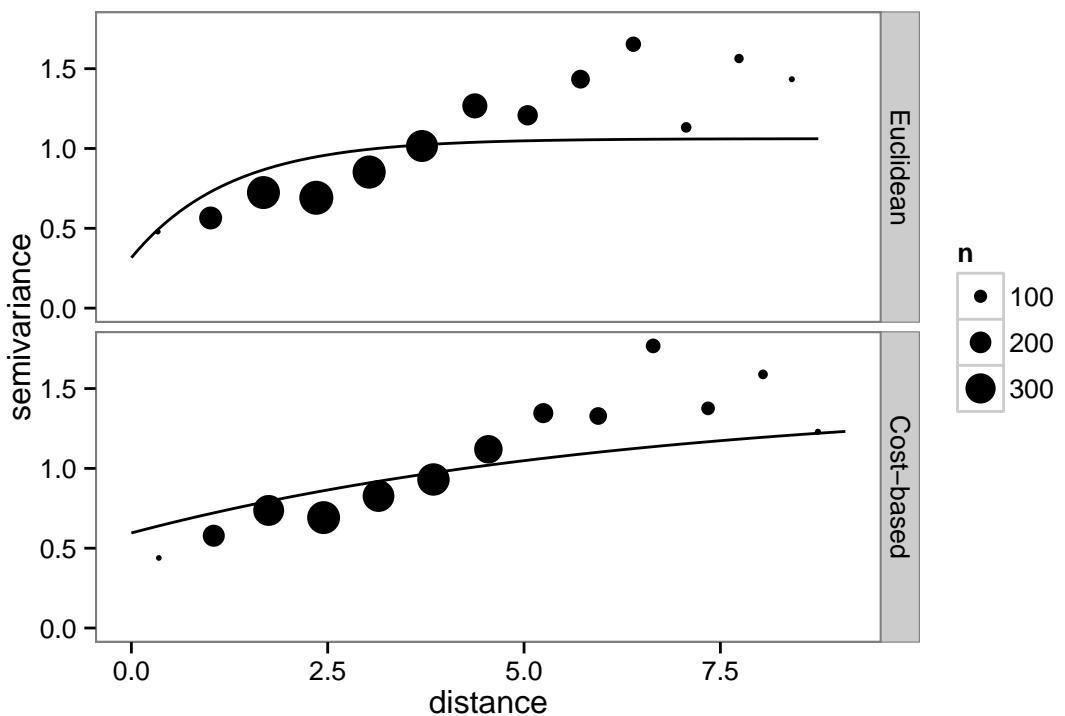


Figure 9: Empirical variogram and fitted models by method for Calcium.

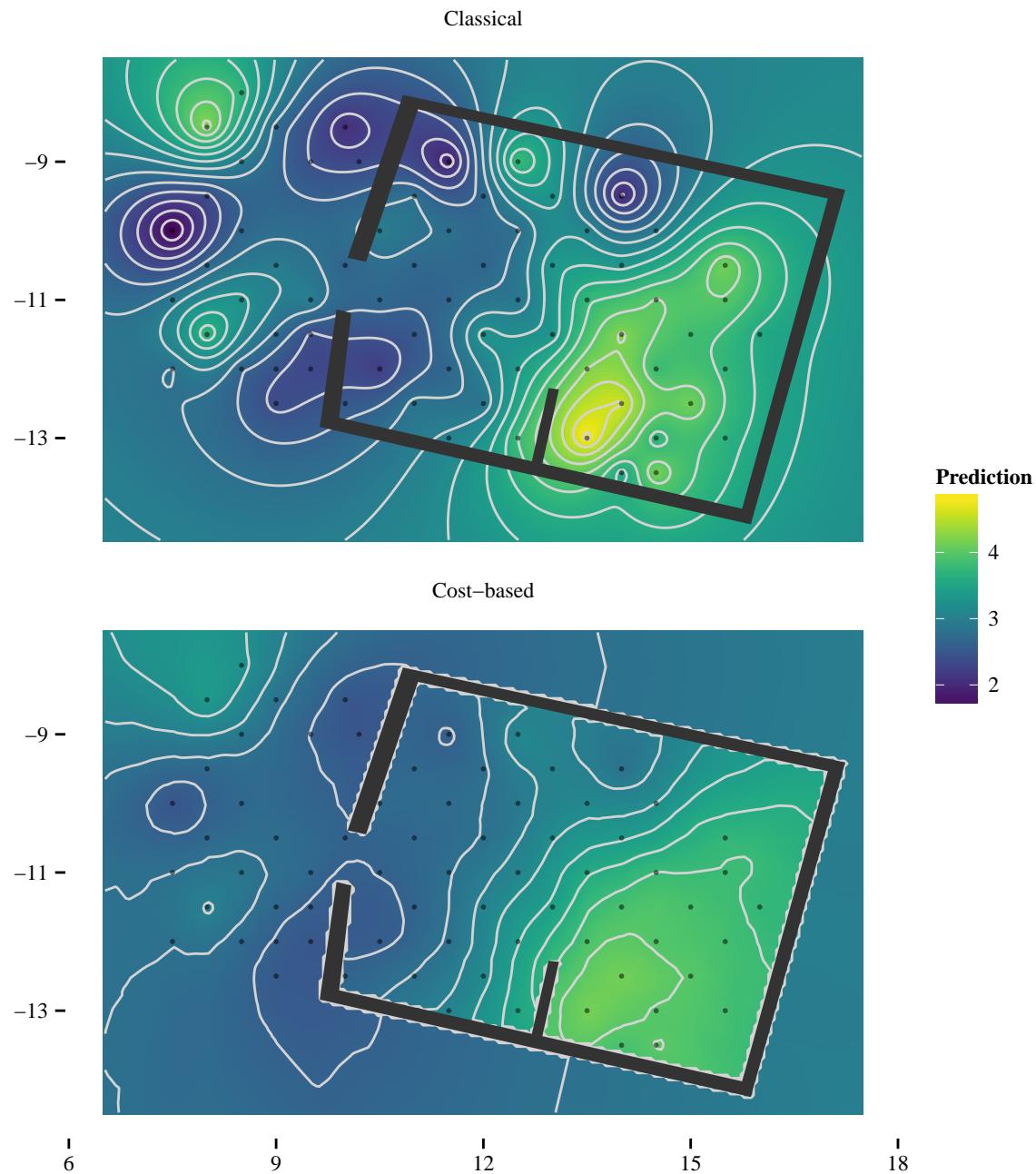


Figure 10: Comparison of Kriging estimates.

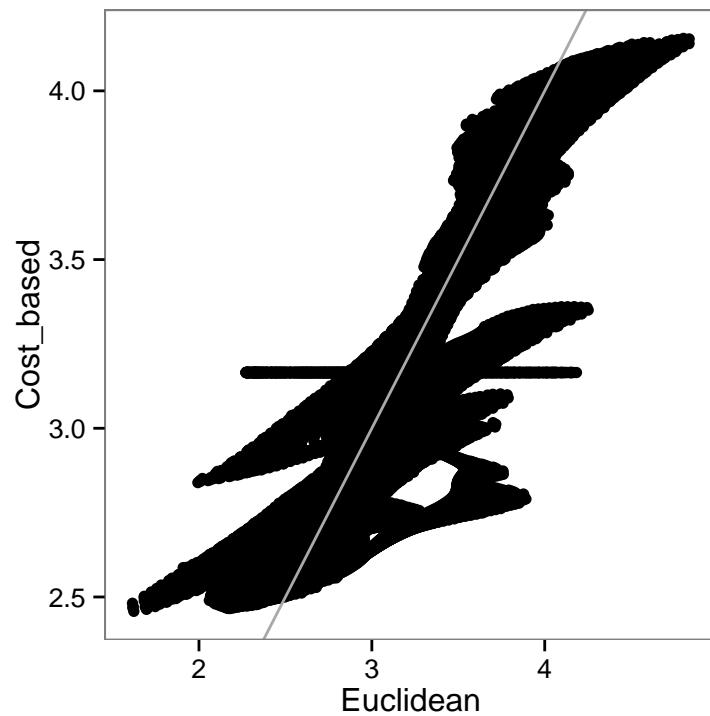


Figure 11: Pointwise comparison of predictions by method.

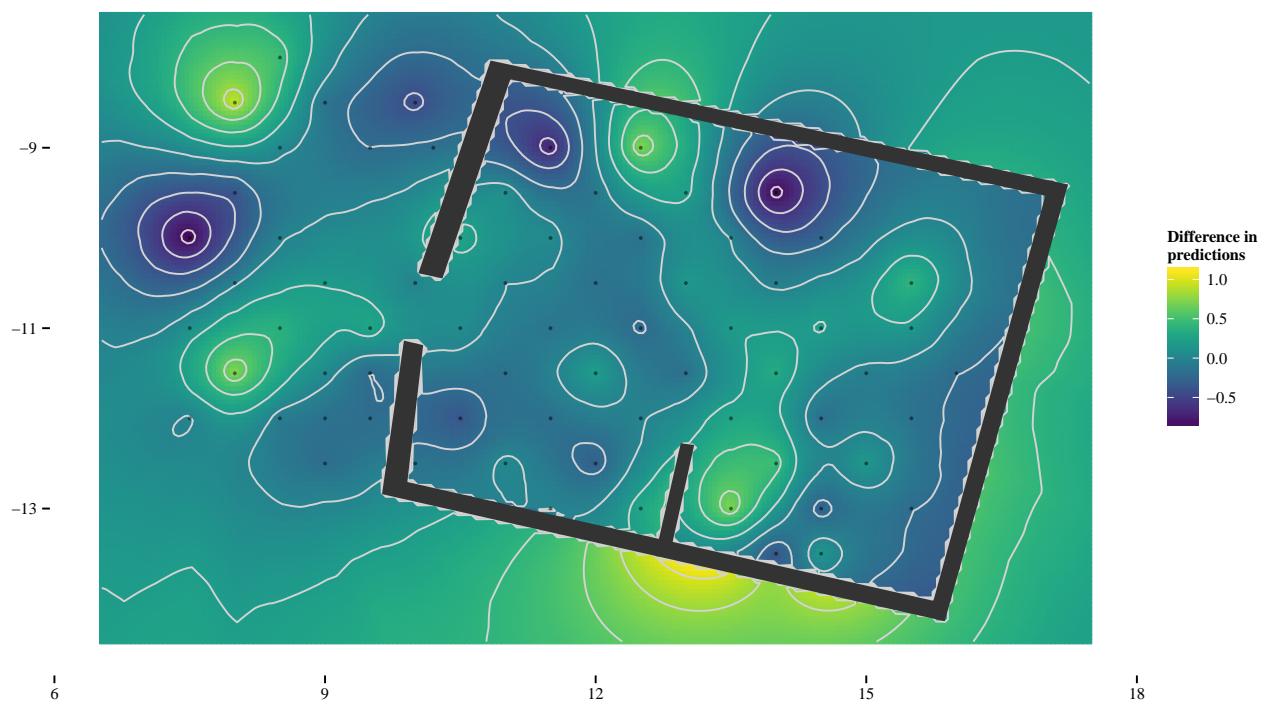


Figure 12: Difference between the Euclidean and the cost-based predictions.

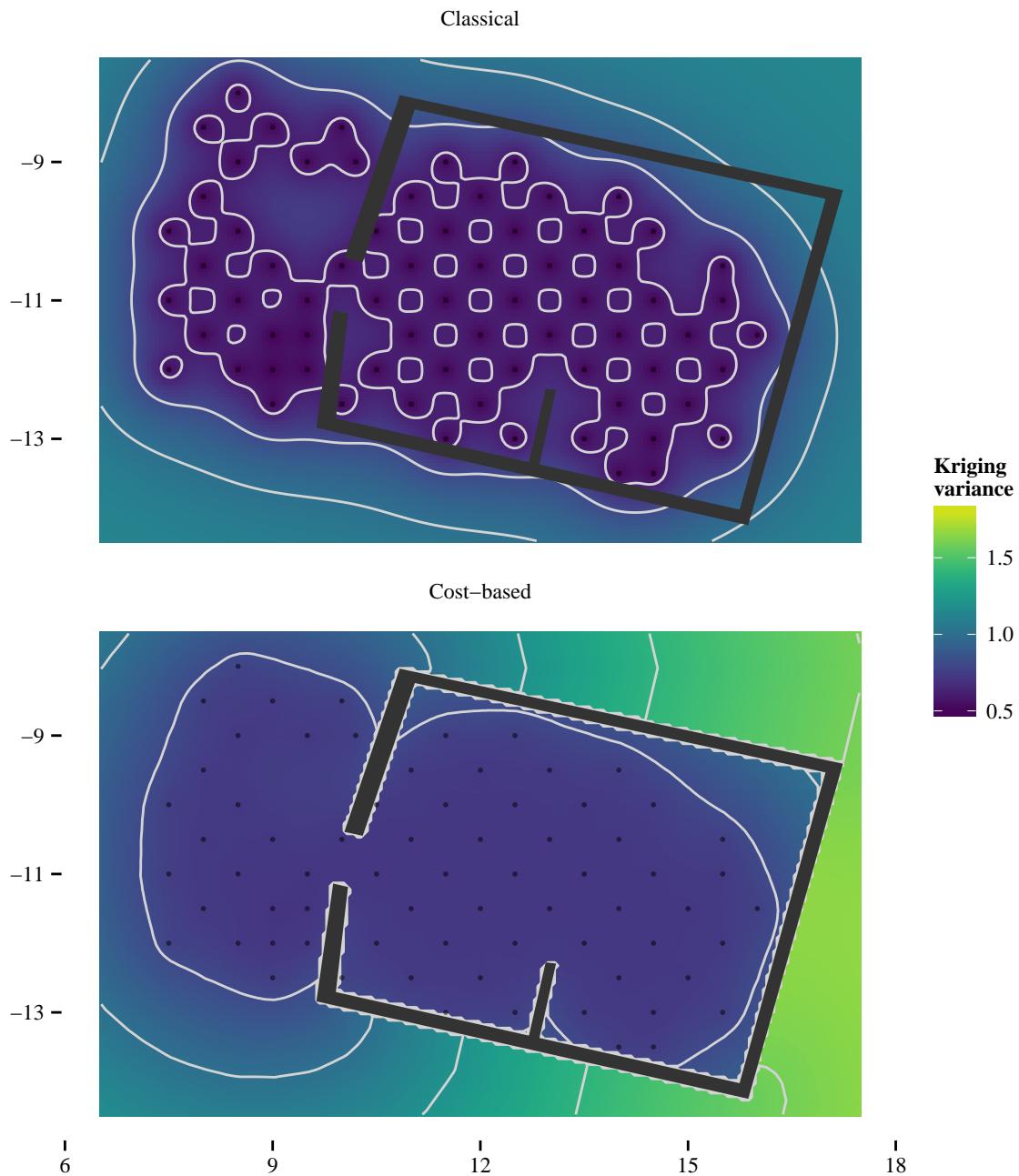


Figure 13: Comparison of prediction error by method.

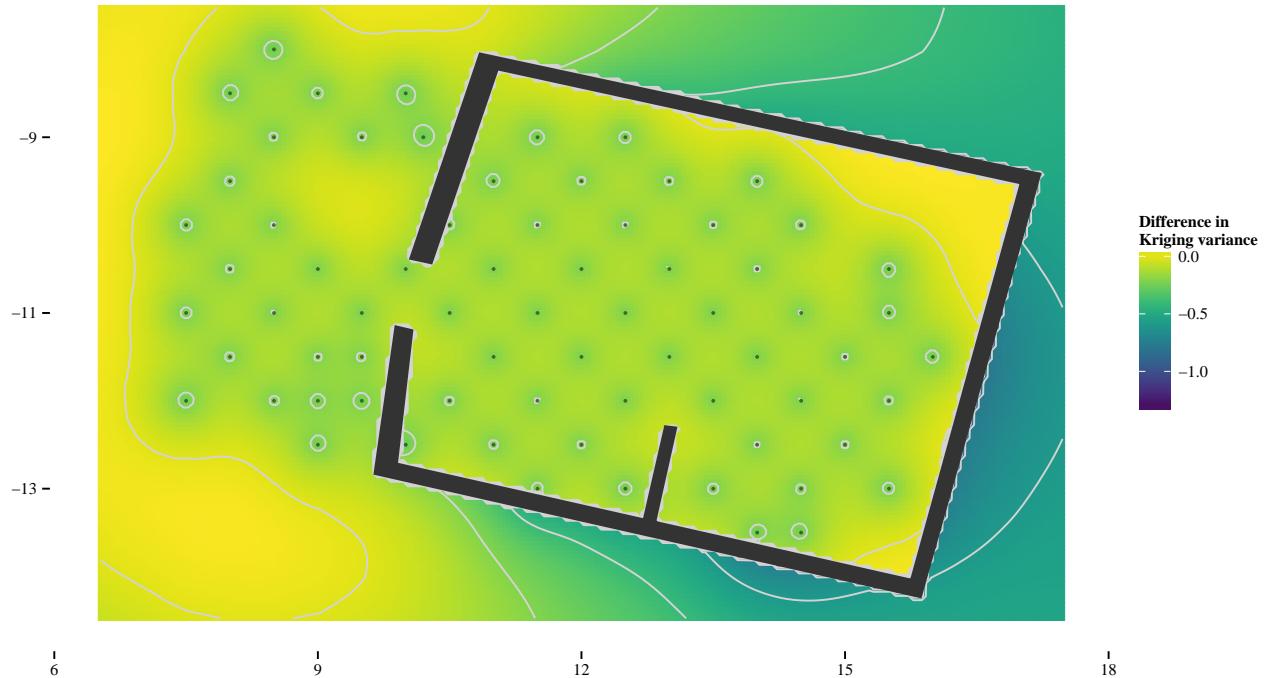


Figure 14: Difference between the Euclidean and the cost-based prediction errors

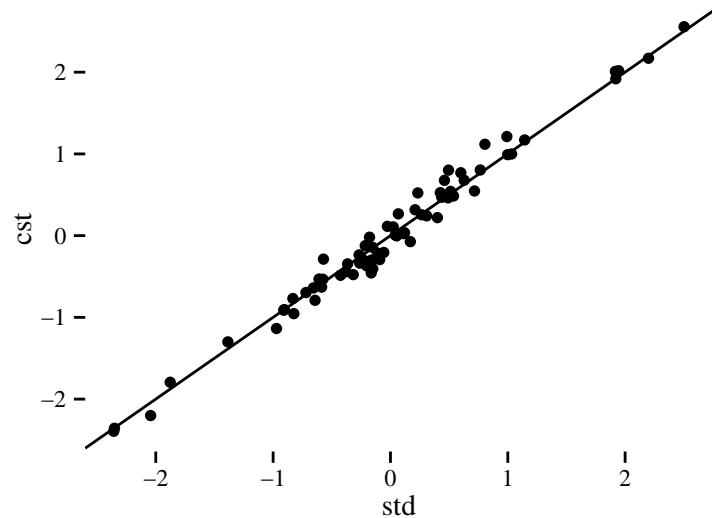


Figure 15: Pointwise leave-one-out prediction error by method.

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.

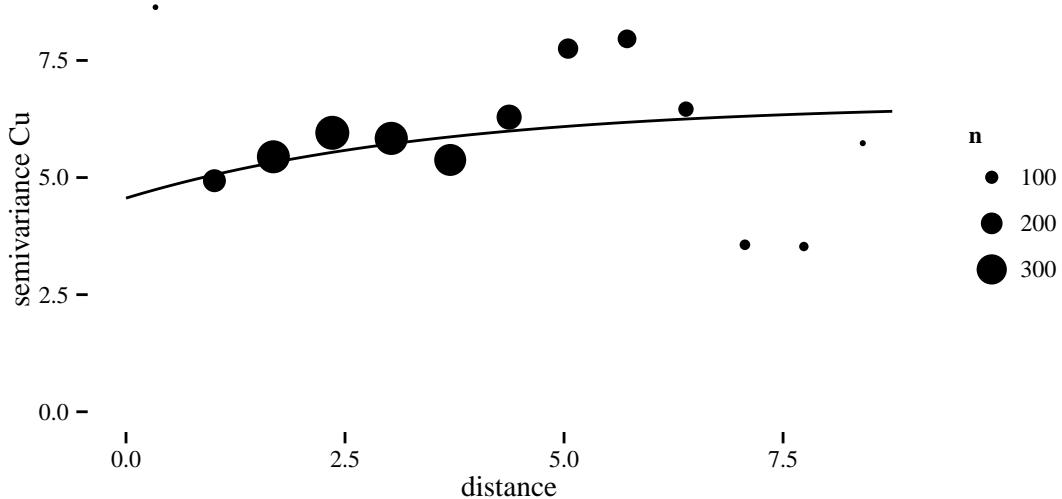


Figure 16: Empirical variogram and fitted model for Copper.

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.

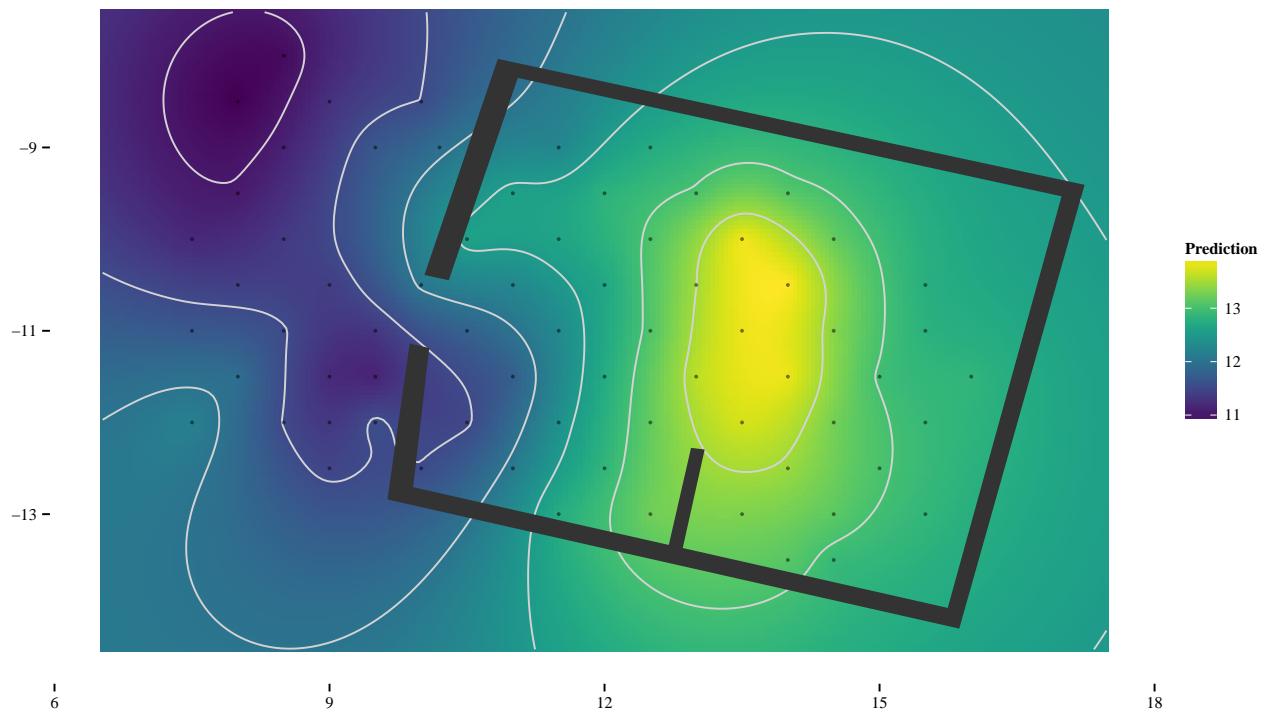


Figure 17: Euclidean kriging prediction for Copper.

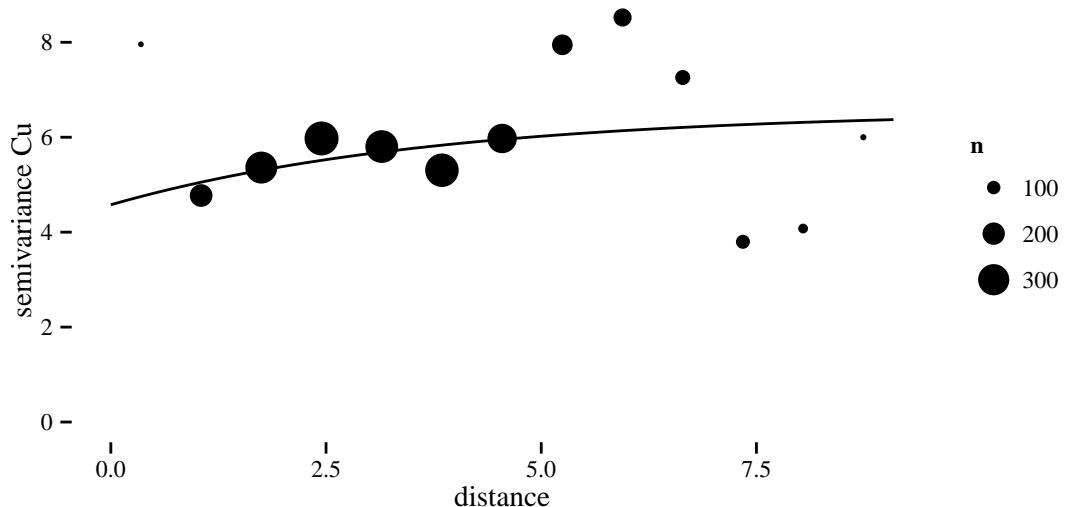


Figure 18: Empirical cost-based variogram and fitted model.

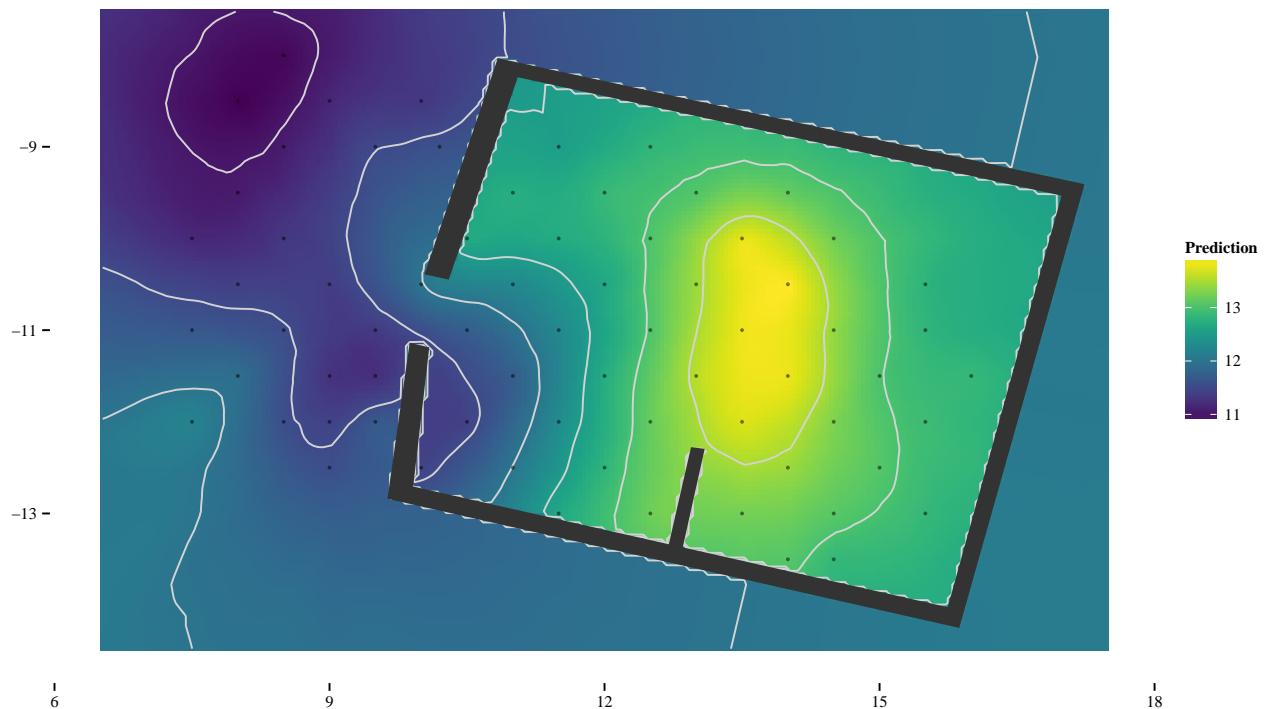


Figure 19: Cost-based kriging prediction

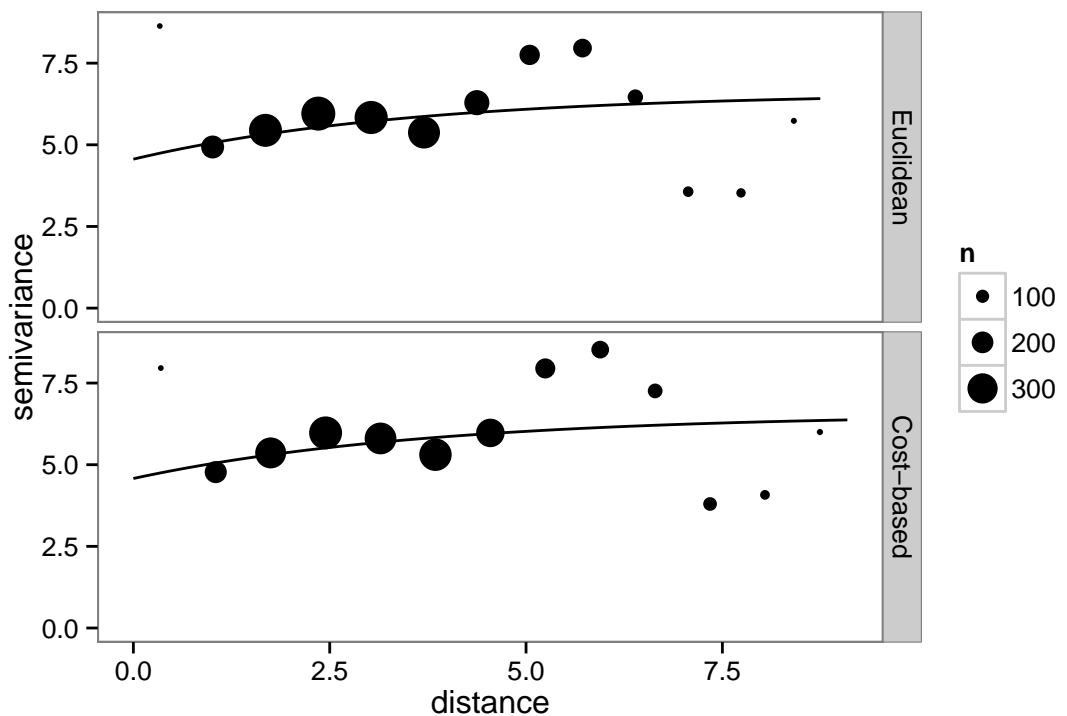


Figure 20: Empirical variogram and fitted models by method for Copper.

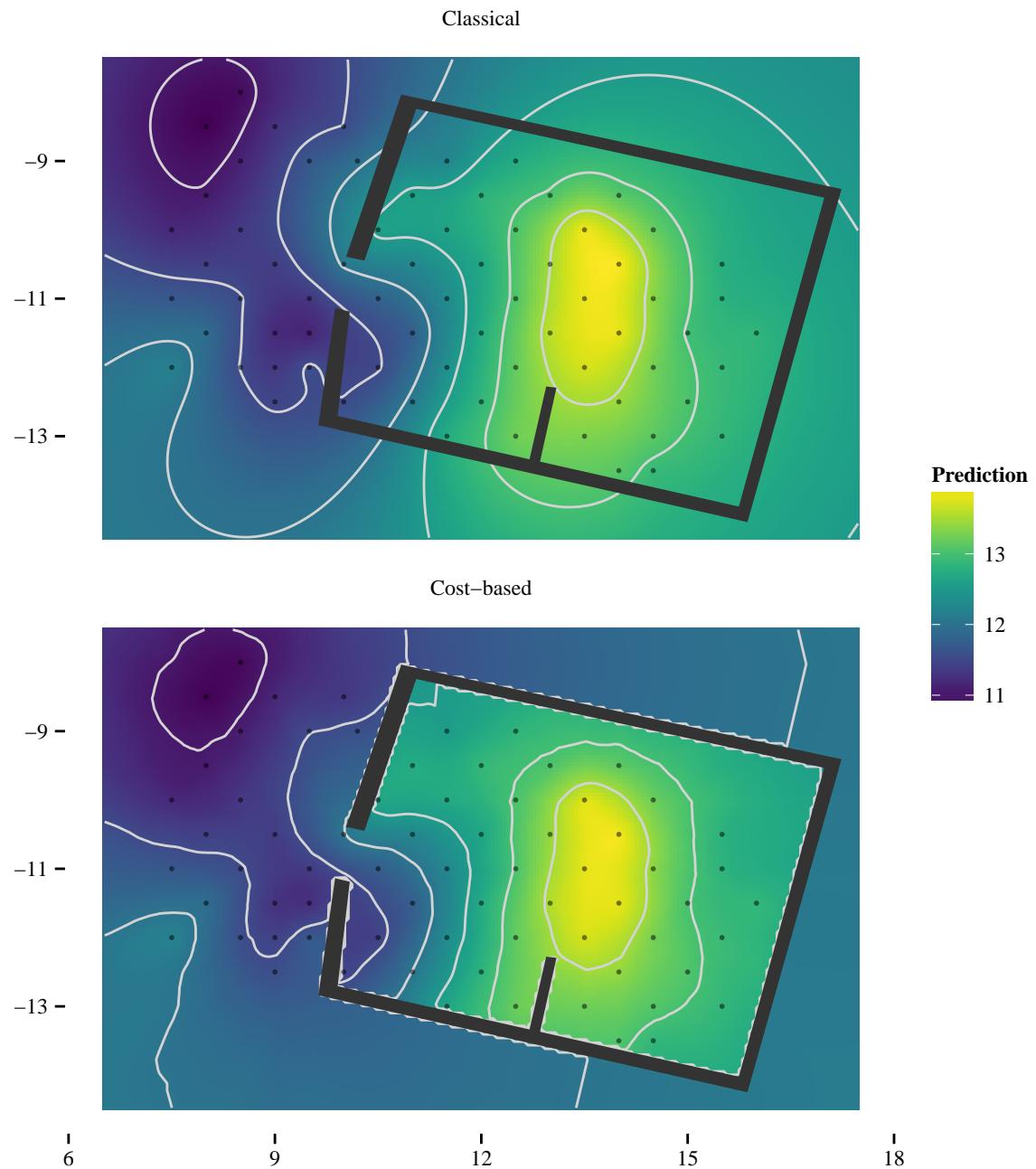


Figure 21: Comparison of Kriging estimates.

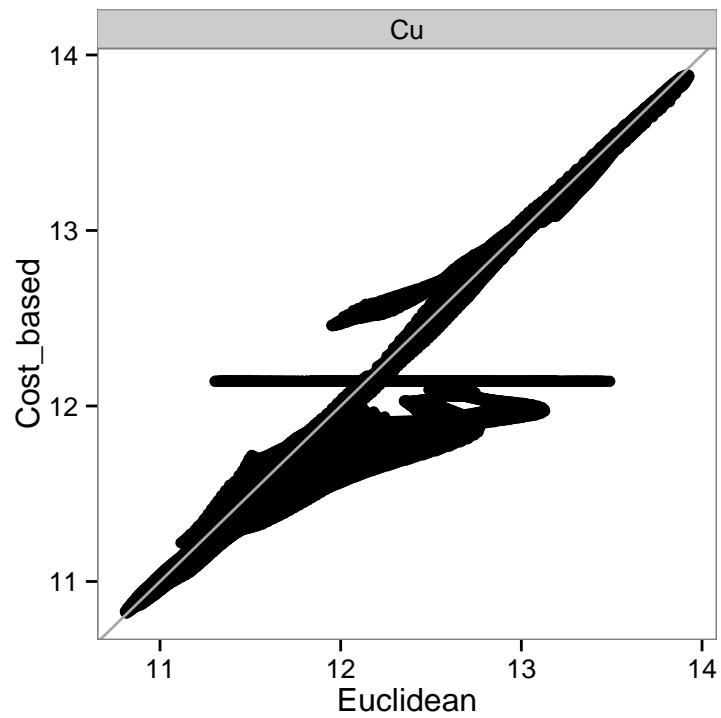


Figure 22: Pointwise comparison of predictions by method.

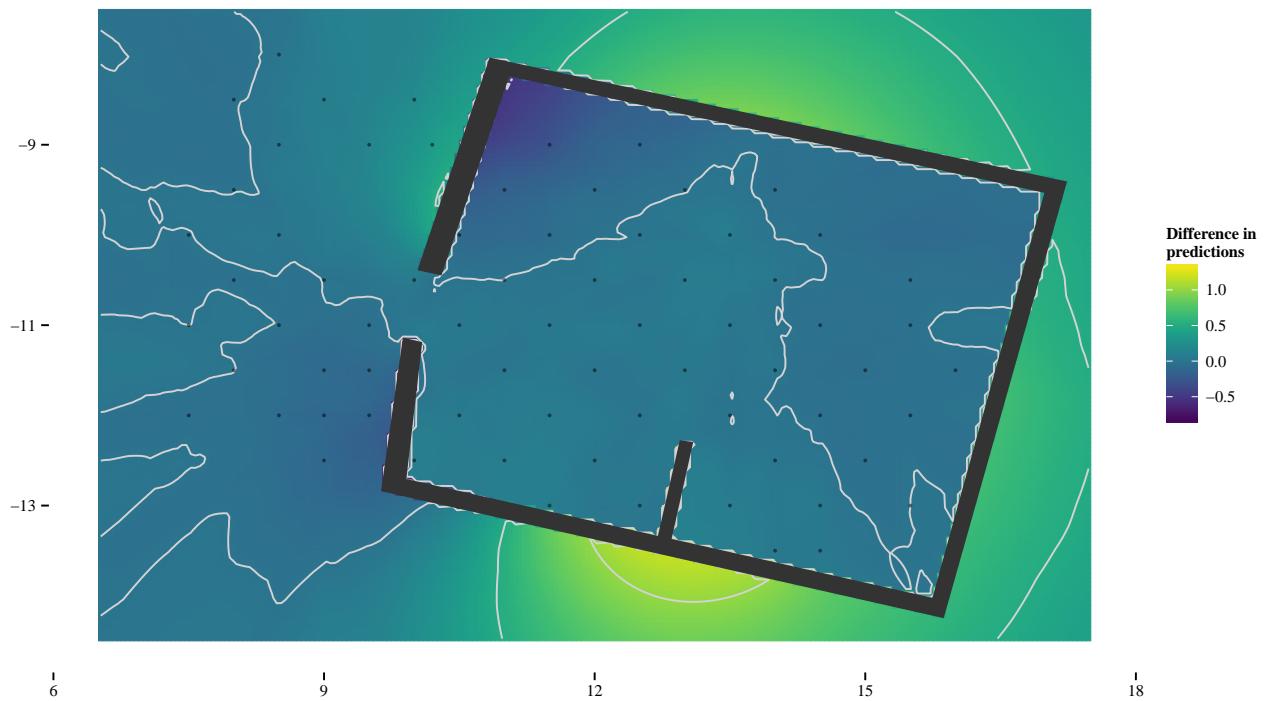


Figure 23: Difference between the Euclidean and the cost-based predictions.

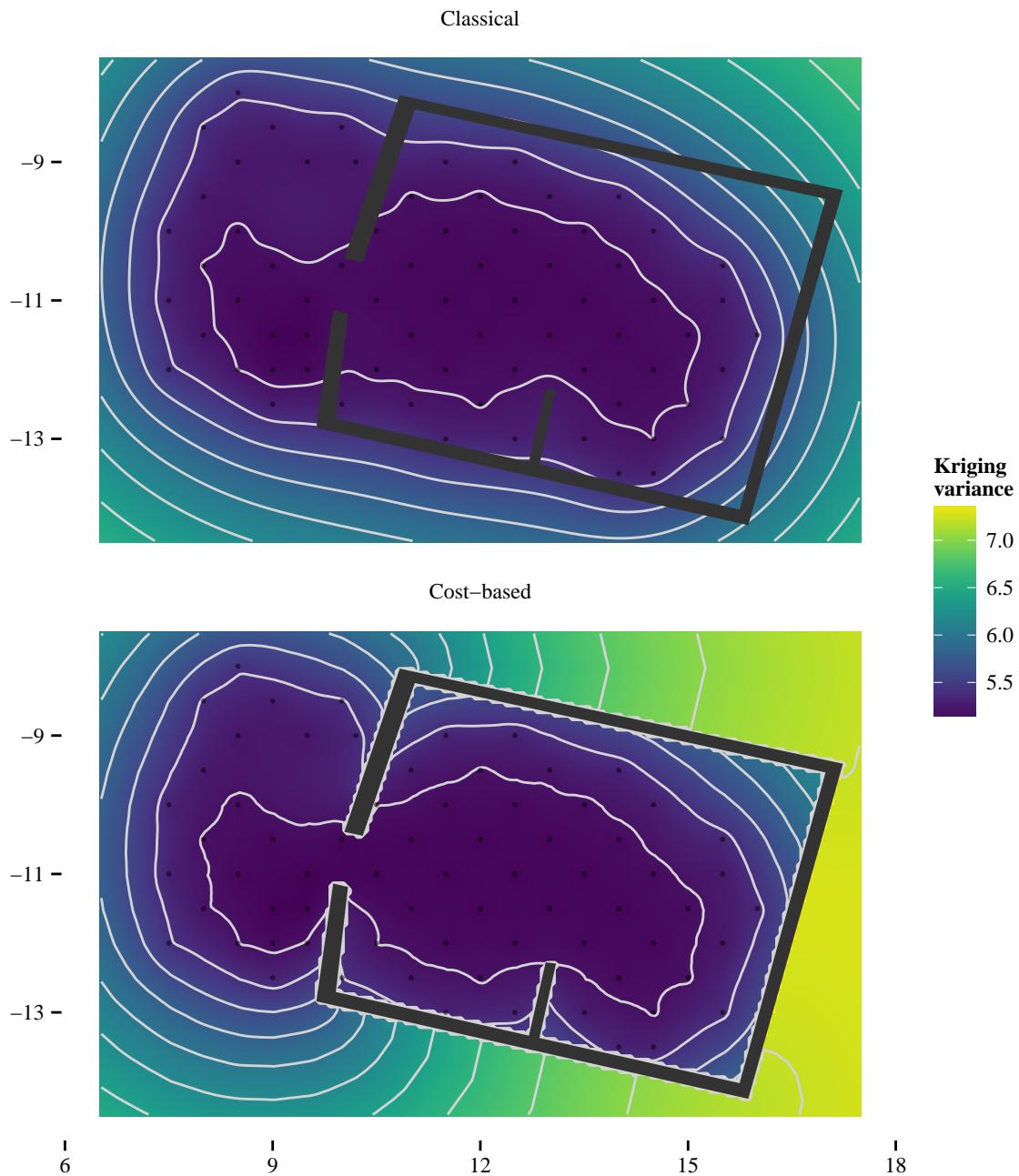


Figure 24: Comparison of prediction error by method.

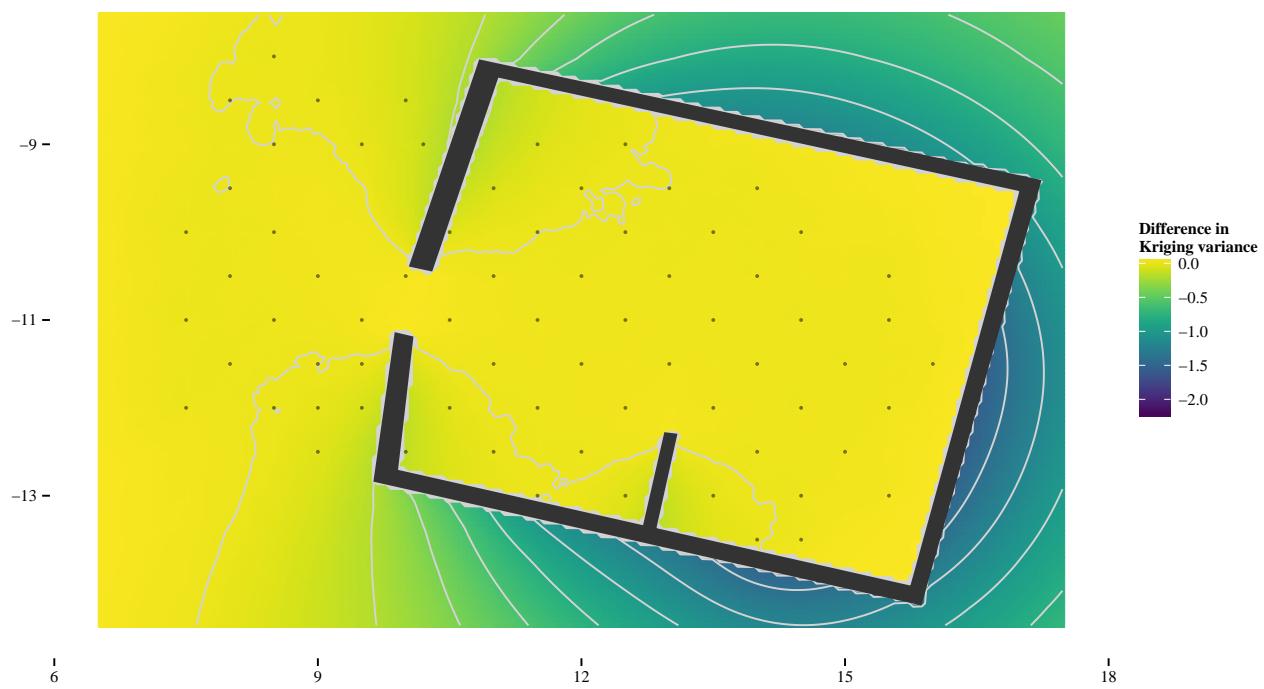


Figure 25: Difference between the Euclidean and the cost-based prediction errors

5 Conclusions

- Actually, the kriging model 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.

