Raster interpolation

steppe

2/6/2022

Contents

List of Figures

List of Tables

1 Spatial autocorrelation

1.1 Tobler's First Law

"I invoke the first law of geography: everything is related to everything else, but near things are more related than distant things." - Waldo Tobler

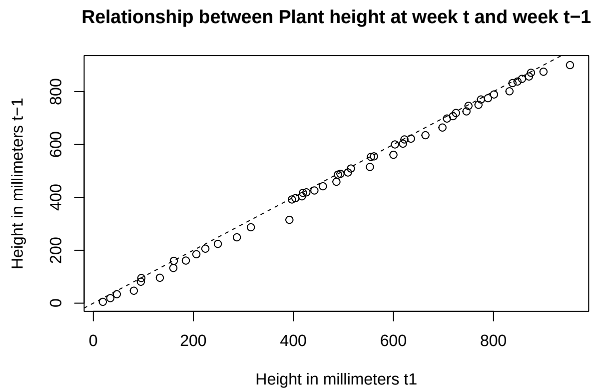
Accordingly, if we want to model processes in space, considerable information may be found from those locations in closet proximity.

1.2 Autocorrelation

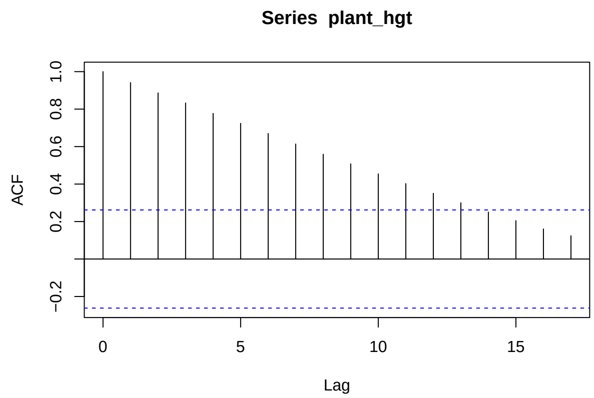
The correlation between two values of a variable as function of the 'lags' between them.

1.2.1 Temporal Autocorrelation

If we were to measure the height of a fast growing plant in a greenhouse every week for a year, each measured value would be strongly related to the value collected the week preceding it.




```
[1] 0.9985263
```



Until we go to around 15 lags away - or in other words 15 weeks back, our data set is significantly autocorrelated. In statistics this will adversely affect the results of our models. However, this information can also be used.

1.3 Spatial Autocorrelation

Spatial auto correlation is a 2 dimensional phenomenon, and is slightly harder to visualize.

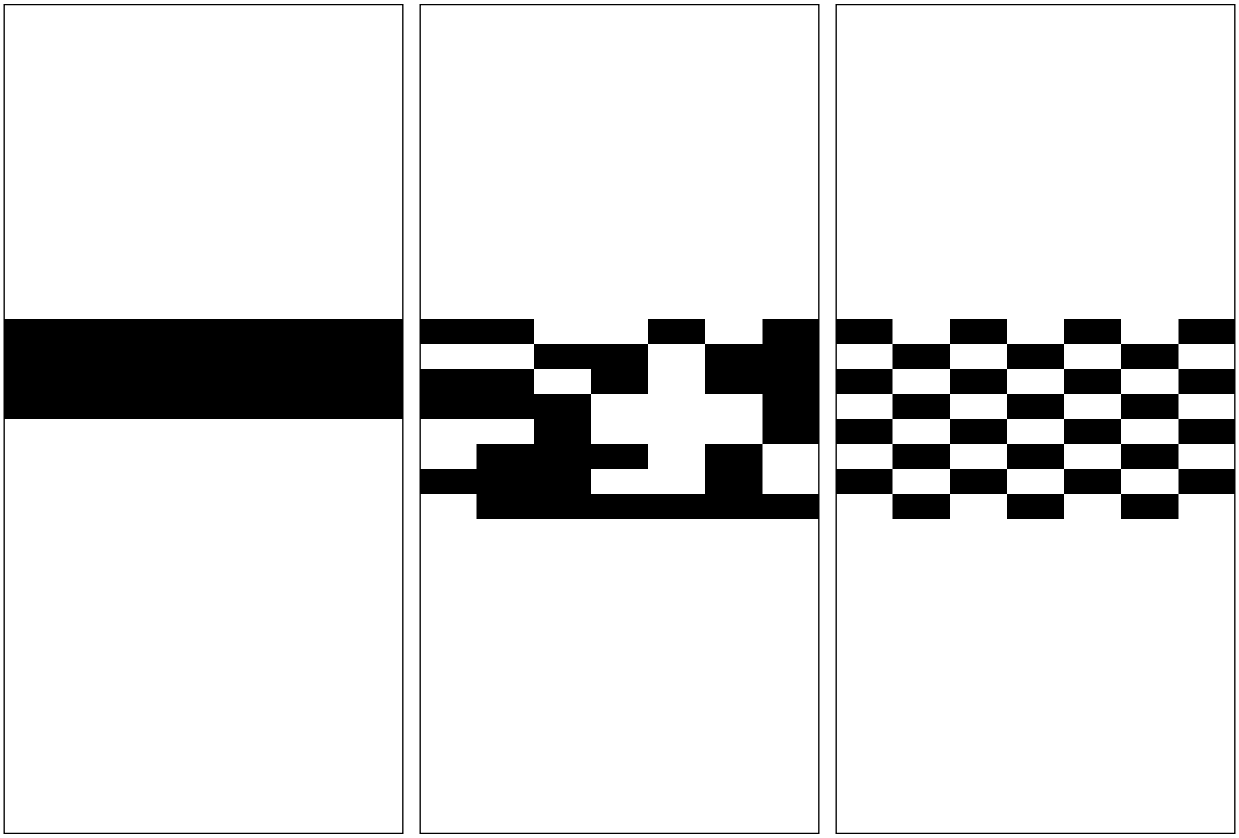


Figure 1: Example figure of spatial autocorrelation. From L to R, Positive, None, Negative

1.3.1 Calculate Morans I

Here we will see if there is a an spatial pattern in the presence of Park Areas in Chicago. First, we will intersect each park to the neighborhood they are in. Then we will determine what proportion of each neighborhood is Park.

```
Neighbour list object:
Number of regions: 98
Number of nonzero links: 504
Percentage nonzero weights: 5.247813
Average number of links: 5.142857
Link number distribution:

 1  2  3  4  5  6  7  8  9 10
 2  8 12 15 18 20 11  6  5  1
2 least connected regions:
Edison Park Boystown with 1 link
1 most connected region:
West Town with 10 links
```

```
    Monte-Carlo simulation of Moran I

data:  chi_neigh_sp$prop_park
weights: spatial_weights_list
```

```
number of simulations + 1: 1000

statistic = 0.17899, observed rank = 995, p-value = 0.005
alternative hypothesis: greater
```

The null hypothesis of this test is that the data are randomly distributed The statistic ranges from 1 (perfect clustering of similar values) to -1 (perfect dispersion; opposites values close).

This test tells us that there is some spatial structure in our data, but we do not know the geographic range to which this structuring is present.

If we recall the location of the Chicago Parks, the most noticeable cluster of them is along the Lake Shore. I presume that the Moran.I index is slightly above random due to the presence of these. It also points out an interesting point regarding the history of the development of Chicago. A focus on restricting development along the Shore. Chicago is quite unique in that it's coasts are publicly accessible, and can largely be owed to the contributions of Montgomery Ward.

This amount of spatial auto-correlation would not adversely affect analyses, not too mention, this is an real pattern of these data.

https://www.chicagotribune.com/news/ct-xpm-1995-10-19-9510190079-story.html 02.04.2022 By: Stephen Lee and Tribune Staff Writer. Chicago Tribune. 10.19.1995

https://www.fotp.org/lakefront-protection-and-public-trust.html 02.04.2022

2 Change Grain of Raster Cells.

- Spatial *Extent* of projects often constrained by areas funding (political boundaries via funds from agencies), or computation power.
- *Grain* of projects almost always constrained by computer power, or data sets.

Within an *extent* an analyst wants to find a *grain* relevant to the their study. *Grain* can be, somewhat readily, altered.

2.1 Raster Cell Aggregation (Coarser Grain)

Rasters may be of a resolution which is to fine to warrant use due to computational limits.

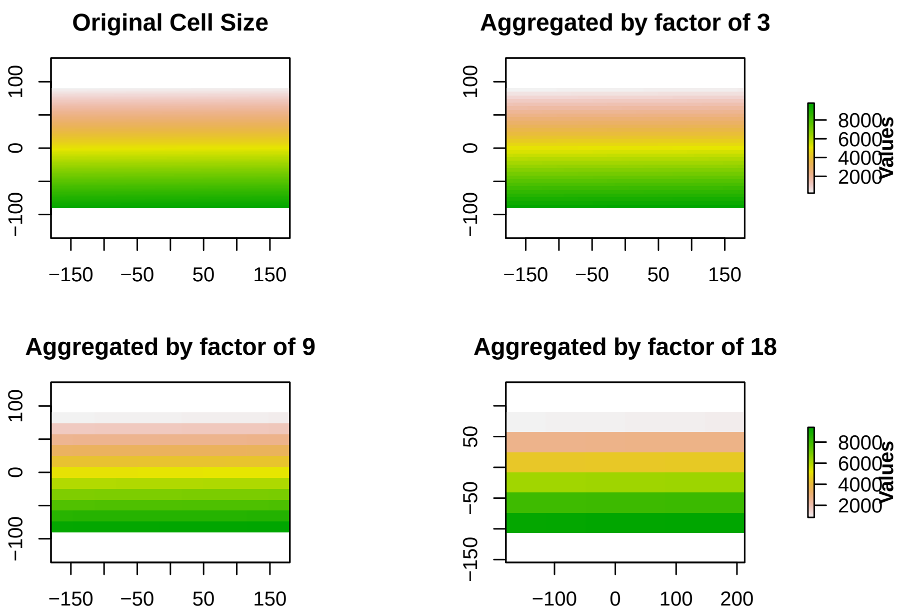


Figure 2: Raster Aggregation

2.2 Raster Interpolation (Finer Grain)

- Predict the value of a variable at an un-sampled location
- Using the values of this variable at sampled locations
- Utilizes the property of spatial auto-correlation

While a great number of raster data sets are developed form satellite imagery and represent the classification of observed values at each location in space, other raster data sets have values which are predicted in space. The most evident example of these are raster products of climate variables. All rasters of climate variables are based on measurements taken from meteorological stations, and then the values between these stations are predicted using spatial interpolation.

Figure 3: Surveying in the White Cloud Mtns. Idaho, by Hubert Szycygiel

Two of the most commonly used spatial interpolation techniques are 'Inverse Distance Weighting' and 'Kriging' interpolation. Our interaction with these processes will be brief, but we will illustrate their uses so that you may recognize the source of spatial data products in the future.

We will leverage simple versions of these techniques to create Rasters of finer resolution than they are currently.

2.2.1 Inverse Distance Weighting Theory

- First Interpolation Method
- Values at points further in space have less weight
- Scale of weight generally decreases linearly or linear^2

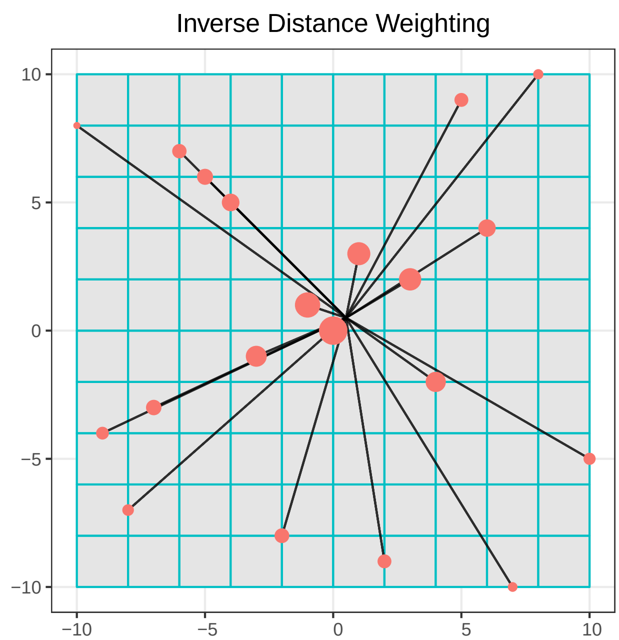


Figure 4: Inverse Distance Weighing Interpolation Theory

```
Warning:
Grid searches over lambda (nugget and sill variances) with  minima at the endpoints:
  (GCV) Generalized Cross-Validation
   minimum at  right endpoint  lambda  =  1.173416e-06 (eff. df= 728.65 )
```

To read more about Inverse Distance Weighting, and view worked out calculations of the process please visit: https://www.geo.fu-berlin.de/en/v/soga/Geodata-analysis/geostatistics/Inverse-Distance-Weighting/index.html

Inverse Distance Weighting is the original spatial interpolation technique and makes great use of Tobler's First Law. While using this technique, we want to predict the value of a variable, such as rainfall, in a location.

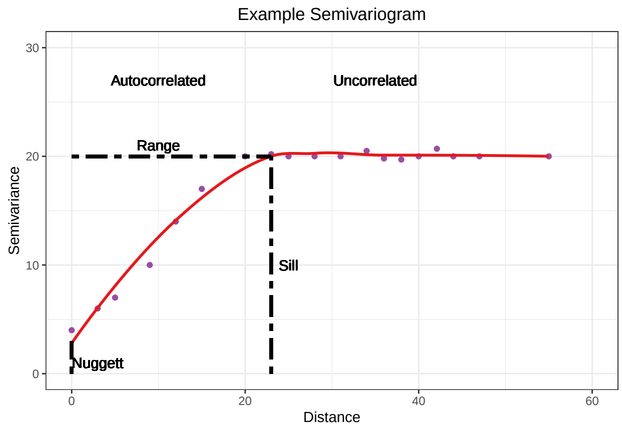
2.3 Kriging Interpolation

https://mgimond.github.io/Spatial/interpolation-in-r.html

2.3.1 Semivariogram

A semivariogram is a method for measuring the correlation between observations at sets of distance.

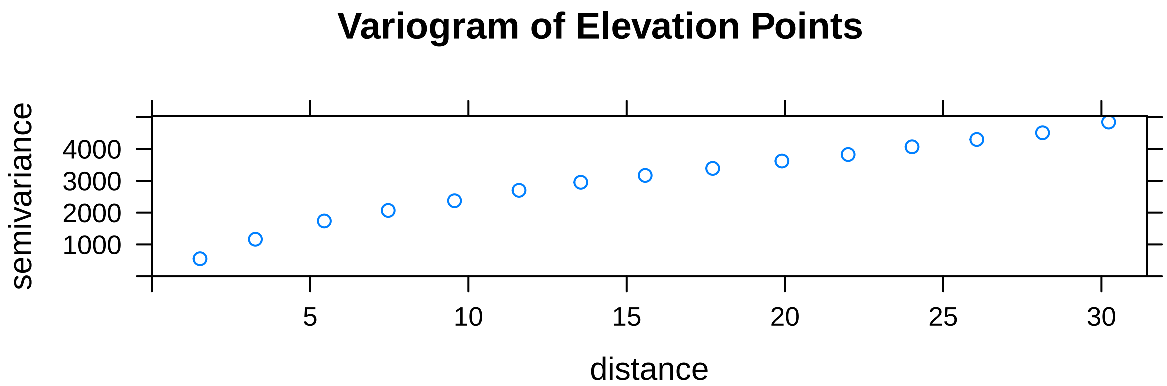
```
## `geom_smooth()` using formula 'y ~ x'
```

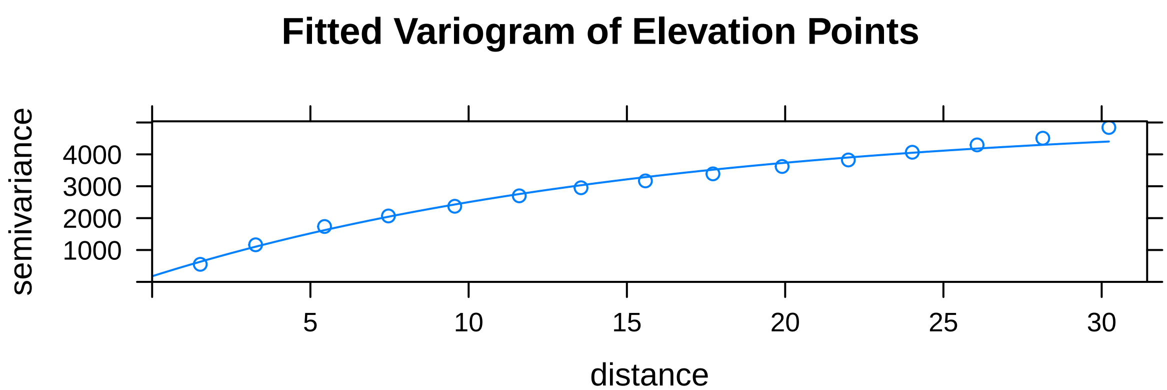


- Nugget: Variation due to errors in the observed measurements.
- Sill: Distance at which the maximum amount of variance in the data is reached.
- Range: Distance at which observations are independent.
- Model: The red line indicates a model which has been fitted to the points (observations), in this example with fictitious data the model is closet to a spherical model.

To determine the range of spatial auto-correlation we can use a Semivariogram

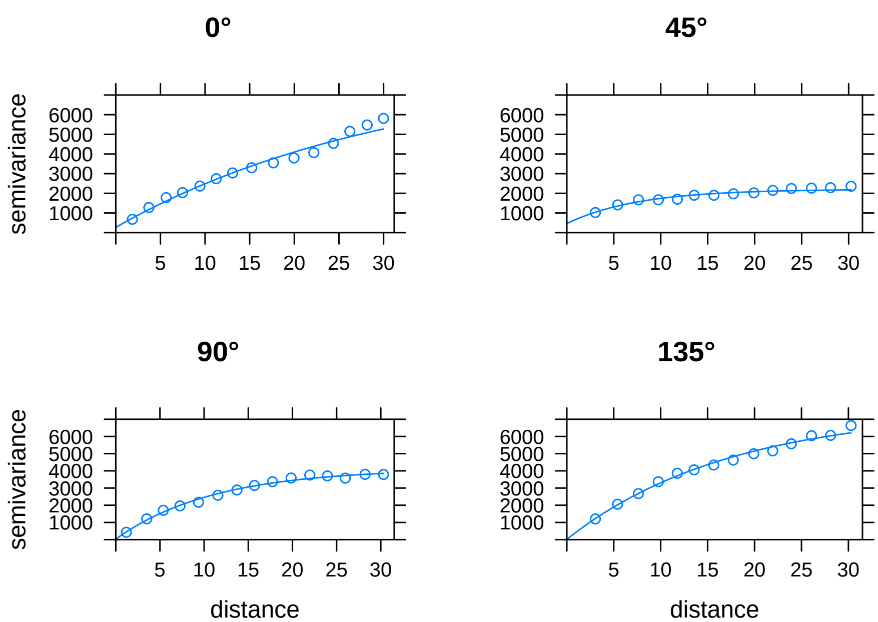
```
variogram_elevation <- variogram(rast_vals~1, data = points)

FittedModel <- fit.variogram(variogram_elevation, vgm(c( "Exp", "Sph", "Gau", "Mat"),
                                 fit.kappa = TRUE))

a <- plot(variogram_elevation, main = "Variogram of Elevation Points")
b <- plot(variogram_elevation, model=FittedModel, main = "Fitted Variogram of Elevation Points", yaxt="r
cowplot::plot_grid(a, b,  ncol = 1)
```






```
rm(variogram_elevation)
```

We can also view directions with predominant auto-correlation.



We see that heading in both 45 and 90 degree angles the effect of spatial auto-correlation deteriorates

2.4 Kriging Inteprolation applied

```
## [using ordinary kriging]
```

2.5 Compare Interpolated Values

We can visually assess the results of our two interpolations using maps.

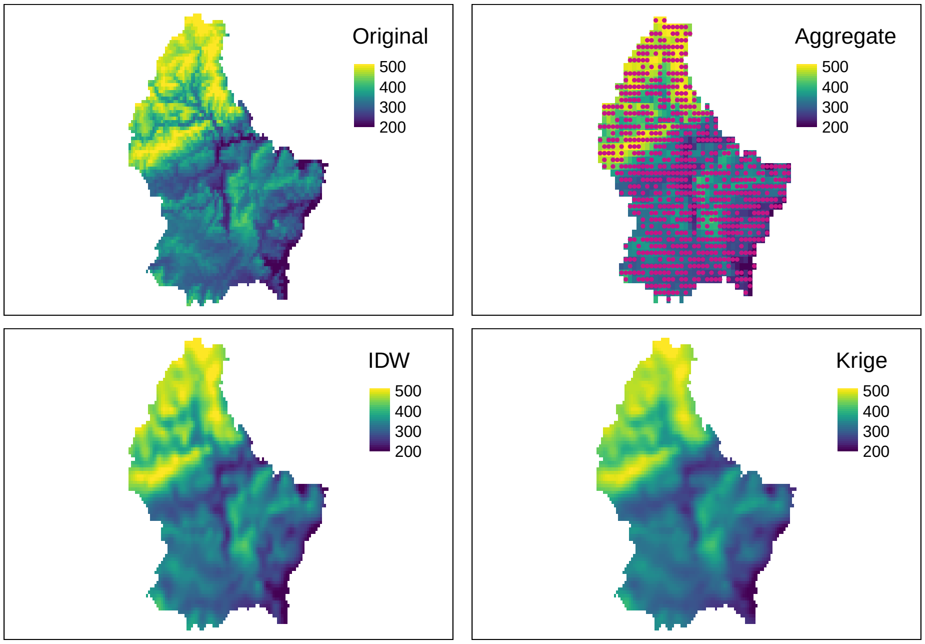
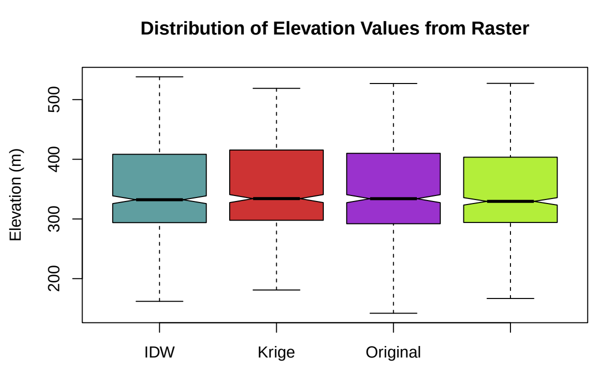


Figure 5: Comparision of Interpolated Values

Remember after importing the *Original* dataset, we merged four cells into one to create our *Aggregate* dataset. We then sampled 1/4 of the aggregated cells, to compute both the *IDW* and *Krige* rasters. we can see that both of them 'smooth' the Aggregated raster. Based on visual inspection, both of our interpolated surfaces appear to match up with the Original raster quite well - for many applications.

```
    Kruskal-Wallis rank sum test

data:  compare$elevation and compare$Type
Kruskal-Wallis chi-squared = 1.2142, df = 3, p-value = 0.7496
```



We can actually check whether the values are similar to the Original raster, and they are.

Kruskal Wallis null hypothesis: the means of each group is the same, we cannot reject the null hypothesis. We see that Krige may ‘smooth’ the results at either end more than IDW, this is a known parameter of it.

3 Works Cited

https://chicagoreader.com/news-politics/cityscape-how-the-lakefront-was-won/

Lee, S., and Tribune Staff Writer https://www.chicagotribune.com/news/ct-xpm-1995-10-19-9510190079-story.html 02.04.2022. Chicago Tribune. 10.19.1995

https://www.fotp.org/lakefront-protection-and-public-trust.html 02.04.2022

https://desktop.arcgis.com/en/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.html Accessed 1.17.2021

Tobler W., (1970) “A computer movie simulating urban growth in the Detroit region”. Economic Geography, 46(Supplement): 234–240.