Point-referenced level data

Borrowing Strength from Closeness

Audrey Yeo
audrey.yeo@uzh.ch

Master of Science
UZH Biostatistics

Motivation

What is point-reference level data?

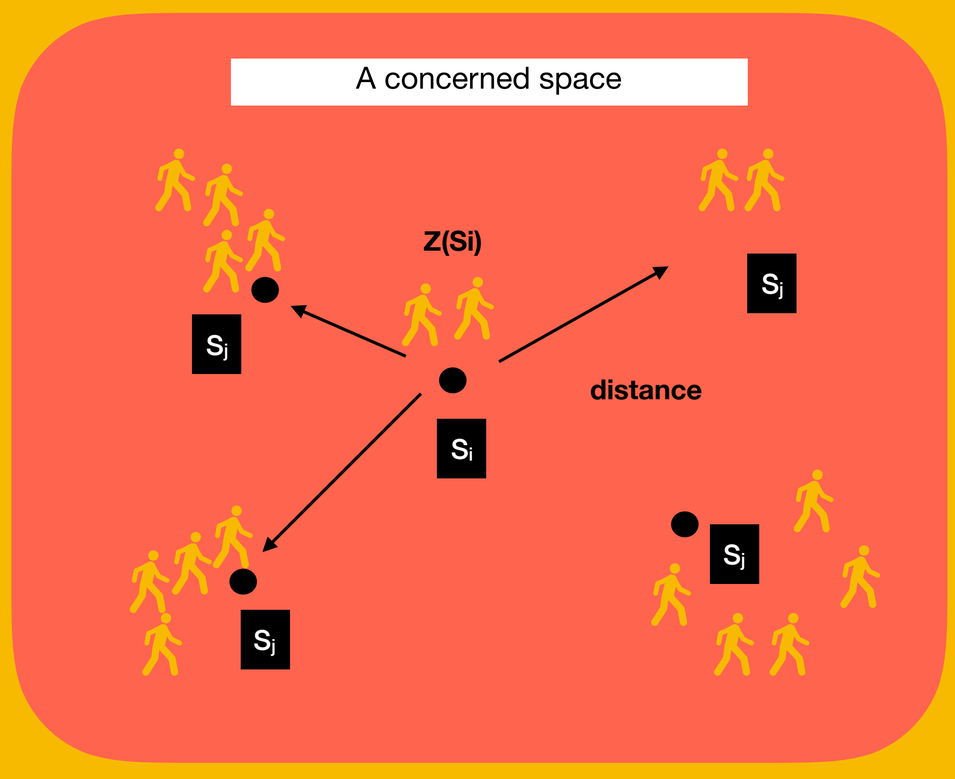
Manner of interpreting spatial data.
Aims to analyse and predict the nature of the points in a concerned space.

Some applications are in:

Geography:
Understanding the link between natural processes and spatial structures. This is under the assumption that two points that are closer are more similar to each other.

Public Health:
Mapping of diseases and predicting intensity of an outbreak, etc.

Economic policy and program:
allocation of the right resources at the right time and space.



Observations e.g. prevalence rate, incidence rate

Sj Point address of distal point

Si Point address of proximal point

Elements

The First Law of Geography
Two points closer together share more similar properties.

The observation of each point is a random process with dependence. For example prevalence rate is a random process.

Isotropy
The idea that predicting the nature of another point only depends on the Euclidean distance and not on the direction of these points

Stationarity
The degree to which two points' nature share the same mean and variance

Variograms
Describes the degree of location dependence of a stochastic process. It quantifies the First Law of Geography.

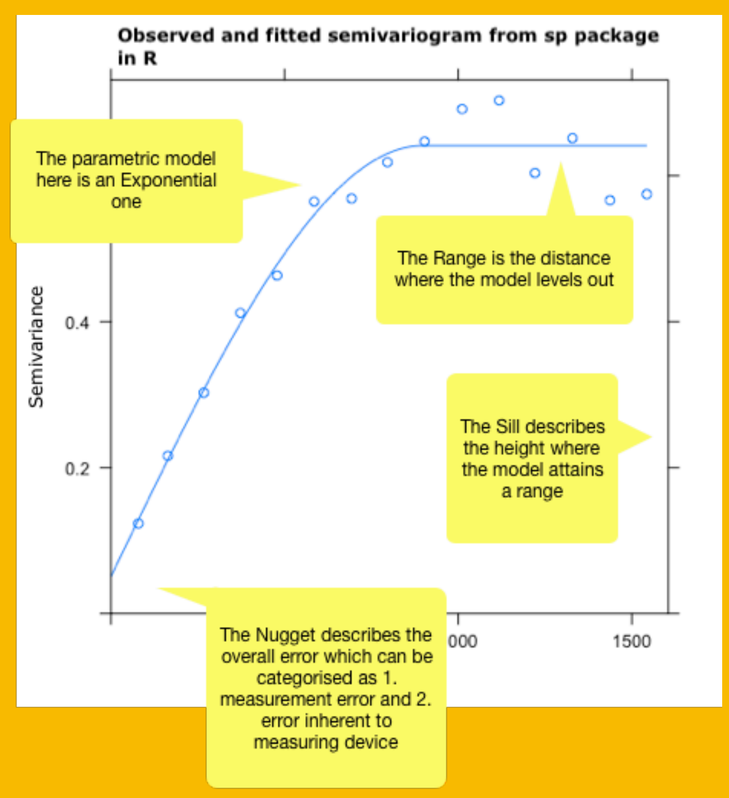
Semivariograms

The semivariogram plot is taken from the Meuse study in package *sp in R*

Model choice

Parametric models available include Spherical and Linear.

The choice of which model to use depends contextually. A theoretical variogram can be selected by plotting the empirical variogram model and compare to which theoretical setting it fits best.



Malaria Risk

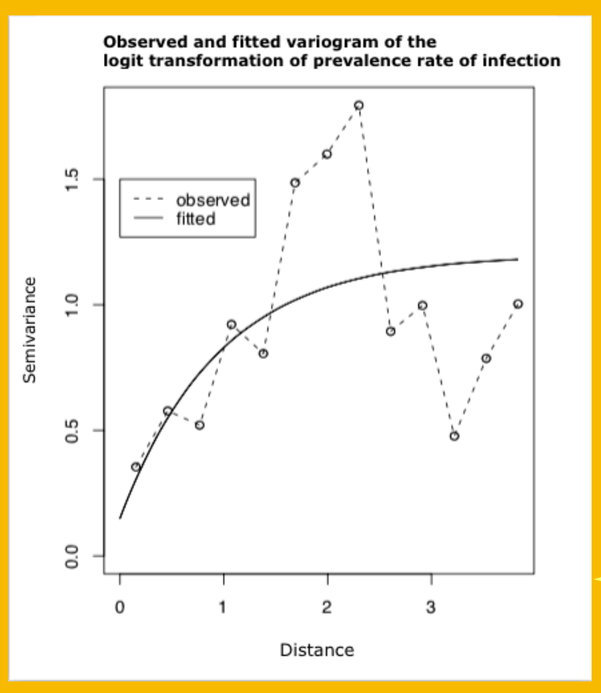


Figure from Kazembe & Holtz, 2006 (Page 4)
Empirical and fitted variogram of the logit transformed prevalence rate of infection.
Separation distance is given in degrees latitude.

Kazembe & Holtz, 2006
Geo-referenced prevalence data for children aged 1–10 years was used to analyze and predict malaria risk in areas where data were not observed .
Elevation, annual maximal temperature, rainfall and potential evapo-transpiration showed significant association with malaria risk, which includes those than optimize mosquito breeding conditions.
The variation of risk in the country described by the resulting map were broadly aligned to expert

References for this poster:

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical Modelling and Analysis for Spatial Data,* Chapman & Hall. 2, 6, 7, 9, 13, 14, 15, 16, 17, 18

Esri (2017). Modelling a semivariogram. http://desktop.arcgis.com/de/arcmap/latest/ extensions/geostatistical-analyst/modeling-a-semivariogram.htm. Accessed 2018- 04-01. 8

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, **58**, 1246. 8
Moore D.A., Carpenter, T. (1999,). Spatial analytical methods and geographic information systems: Use in health research and epidemiology. Epidemiologic Reviews, 21, 143{161. 1, 2, 4, 39

Pebesma E., Graeler B., (2018) Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation (R Package). https://github.com/edzer/gstat/

Rossi, R.E., Mulla, D, J., Journal, A. G., and Franz, E.H. (1992) Geostatistical tools for modelling and interpreting ecological spatial dependence. *Ecological Monographs.* **62**, 277-314. 6

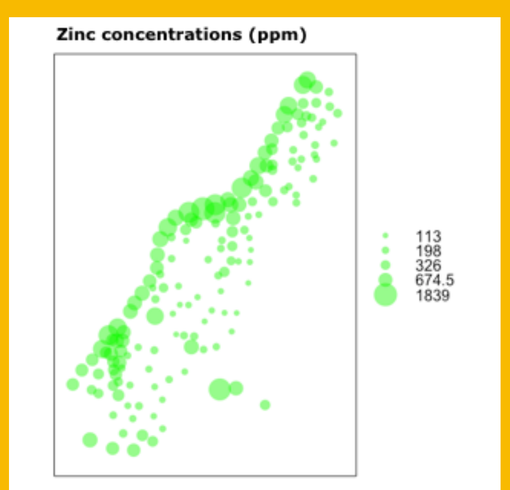
Kazembe, L., Kleinschmidt, I., Holtz, T., and Sharp, B. (2006). Spatial Analysis and Mapping of Malaria risk in Malawi using Point-Referenced Prevalence Infection Data. *International Journal of Health Geographics., **5,** **41.** 2*

Williams, C (2015). Geostatistics - basic definitions, application and general info. Accessed 24.4.2018. 4

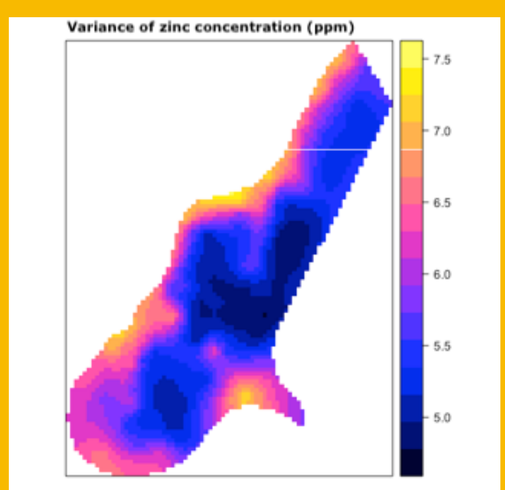
Williams, B.L., Pennock-Roman, M., Suen, H.K., Magsumbol, M. S., and Ozdenerol, E. (2006). Assessing the impact of the local environment on birth outcomes: a case for HLM. *Journal of Exposure Science and Environmental Epidemiology,* **17**, 445 EP. 84-92

Pollution

Pebesma & Graeler (2018)
The plots here are from an empirical study in the sp package in R.
The data was collected in a flood plain of the river Meuse which is located close to the village of Stein in the Netherlands. Heavy metal concentrations are from composite samples of an area 15 m x 15 m. There are multiple different metals whereas only zinc is of interest.
The zinc concentration is measured in mg per 1 kg soil, mostly referred to as part per million or ppm.



A bubble-plot of measured zinc



Modeled variance of zinc concentration. The variance of the measured zinc concentration as seen in bubble-plot is greater as points are further.