Big Geospatial Data - Home assignment 1

Prof. Dr. Philipp Otto

> **Note**
>
> issue date :
> Submission date :
> Name, matriculation number:
> Evaluation:

1. **Problem: Definition of statistical properties**

   Define the following mathematical/statistical terms (see Cressie 1993, chapters 1 and 2 https://rongxie.files.wordpress.com/2011/01/statistics-for-spatial-data-revised-version-1993.pdf) !

   a)Spatial data

   b)Space-time process

   c)Random process

   d)First-order stationarity

   e)Second-order stationarity

   f)Isotropy

   g)Ergodicity

   h)Variogram, semivariogram

2. **problem: Empirical examples**

   Discuss for the empirical examples given below, if they could be considered as weakly stationary and isotropic random process!

   - Annual average temperature in all European NUTS-2 region (Figure 1).
   - Number of vehicles passing crossroads in a road network, e.g.(Figure 2).

Figure 1: European NUTS-2 region

Figure 2: road network

- Particulate matter concentrations (PM10) measured at all stations in Niedersachsen https://www.umwelt.niedersachsen.de/themen/luft/luen/aktuelle_messwerte

Furthermore, find real-world examples of spatial random processes, which can be seen as :

a) weakly stationary and isotropic

b) weakly stationary, but not isotropic (meaning anisotropic)

c) non-stationary

3. **problem: Nugget effect**

   For real data, variograms typically do not approach zero as the distance between the locations approaches zero.This is commonly known as nugget effect. Try to find an explanation for this effect!

4. **problem: Big data**

   Read the paper "The Parable of Google Flu: Traps in Big Data Analysis" (Lazer, David, et al., Science 2014:pp. 1203-1205) and discuss the following statements! https://science.sciencemag.org/content/343/6176/1203

   a) Big data are a substitute for traditional data collection and analysis.

   b) Big data were overfitting the small number of cases.

   c) Replication is a growing concern across the academy.

   d) There is a tendency for big data research and more traditional applied statistics to live in two different realms.