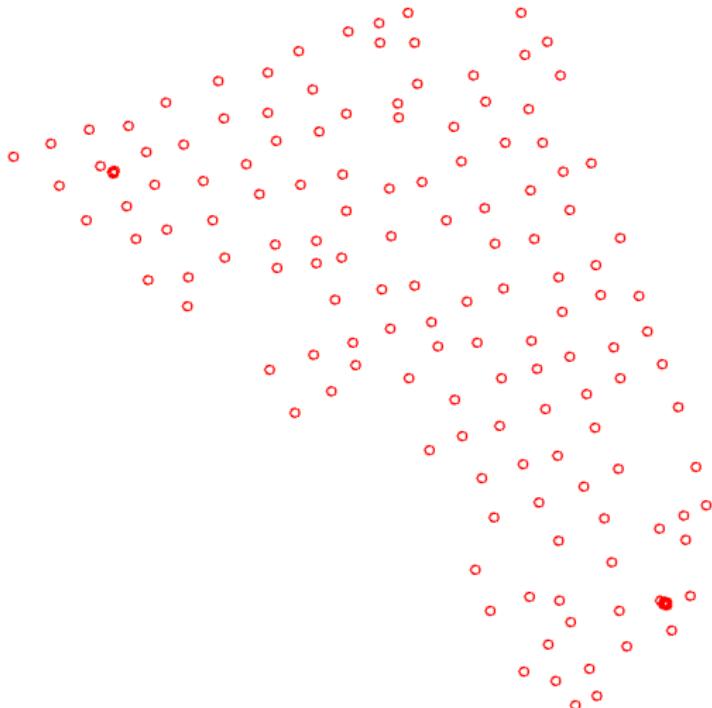
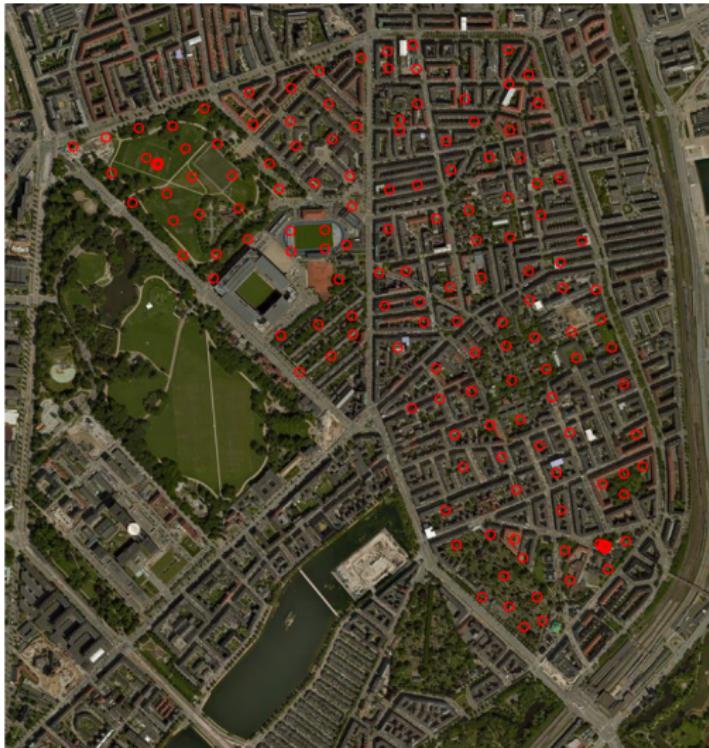

Analysis and Visualization of Spatial and Spatio-temporal Data

Søren Lophaven

Sampling locations - where are we?



Sampling locations on a map

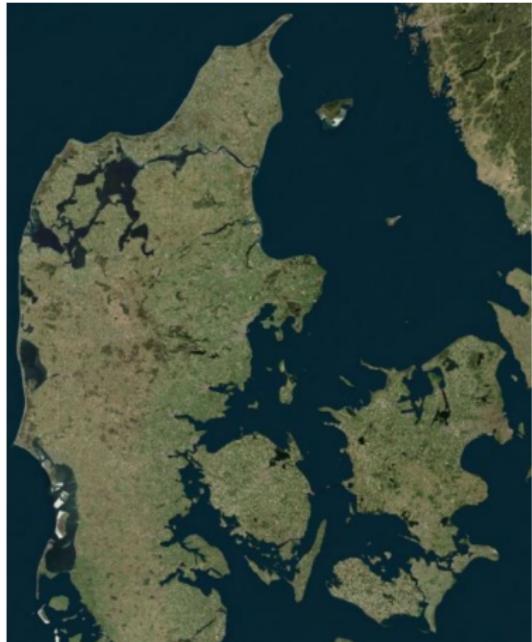


Creating a map in R

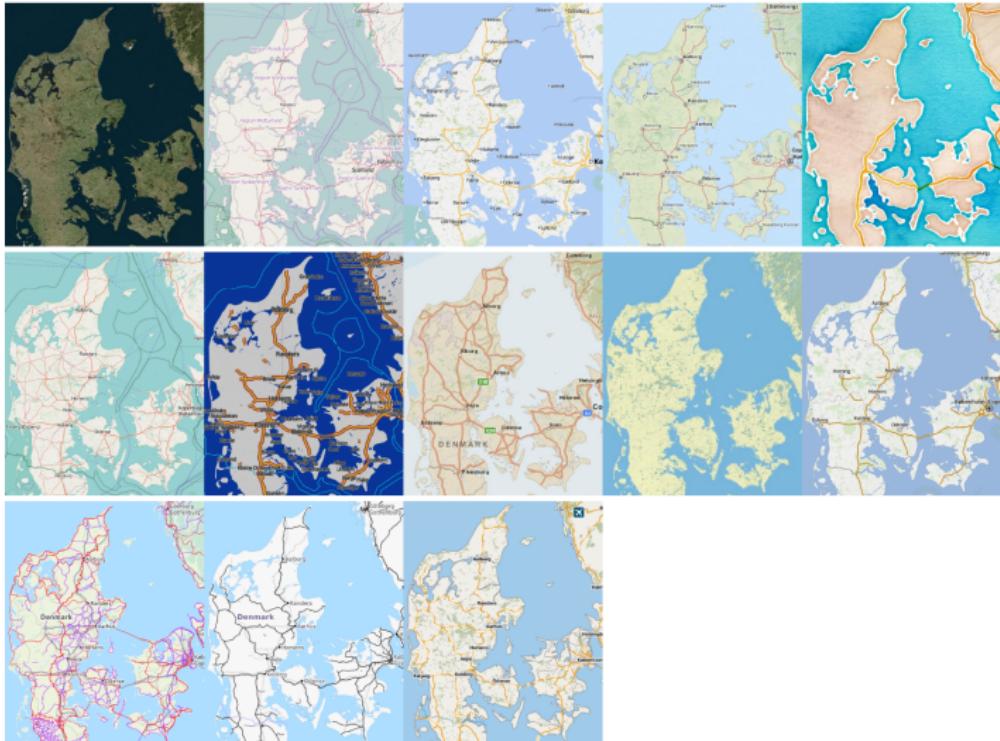
```
# crowd-sourced and open-licensed  
# world map (www.openstreetmap.org)  
install.packages("OpenStreetMap")  
library(OpenStreetMap)  
denmark<-openmap(  
c(lat=57.800,lon=8.020),  
c(lat=54.500,lon=12.876),type="bing")  
plot(denmark)
```

Coordinates from:

```
launchMapHelper()
```

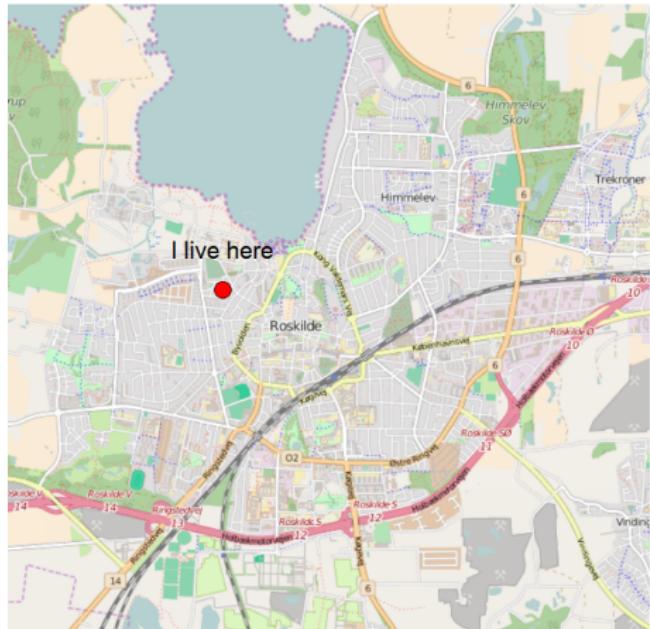


More maps of Denmark



Another map with points and text

```
roskilde<-openmap(  
  c(lat=55.673,lon=12.0350),  
  c(lat=55.615,lon=12.140),type="osm")  
  
plot(roskilde)  
  
lovparken21<-data.frame(  
  lat=55.64655,lon=12.07023)  
  
#Project to the same coordinate system  
lovparken21<-projectMercator(  
  lovparken21$lat,lovparken21$lon)  
  
points(lovparken21[1],lovparken21[2],  
  bg="red",cex=4,pch=21)  
  
text(lovparken21[1],  
  lovparken21[2]*1.0001,"I live here",  
  cex=2.5)
```



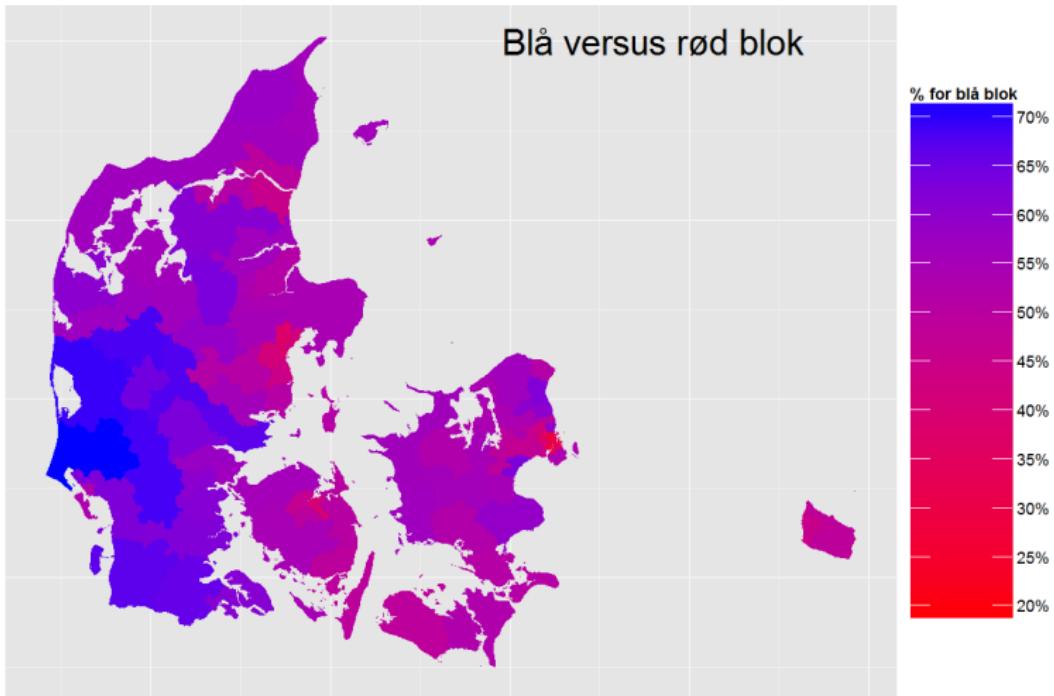
Spatial data types and correlation

- Referenced by a spatial location or an area
- Spatial correlation: Data that are close together in space are often more alike than those that are far apart

Spatial data types:

- Geostatistical data (point referenced, soil pollution)
- Lattice data (area referenced)
- Spatial point process data (locational data)

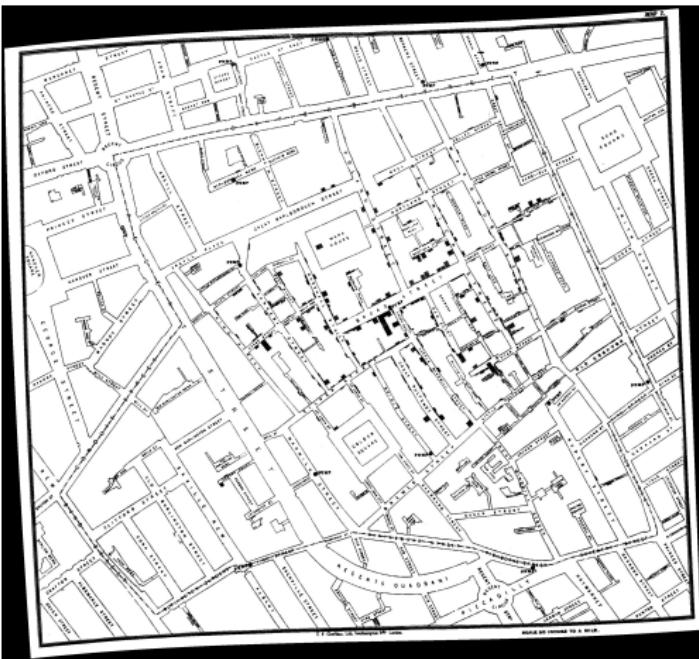
Lattice data - Danish election June 2015



Spatial data types

- Geostatistical data (point referenced)
- Lattice data (area referenced)
- Spatial point process data (locational data)

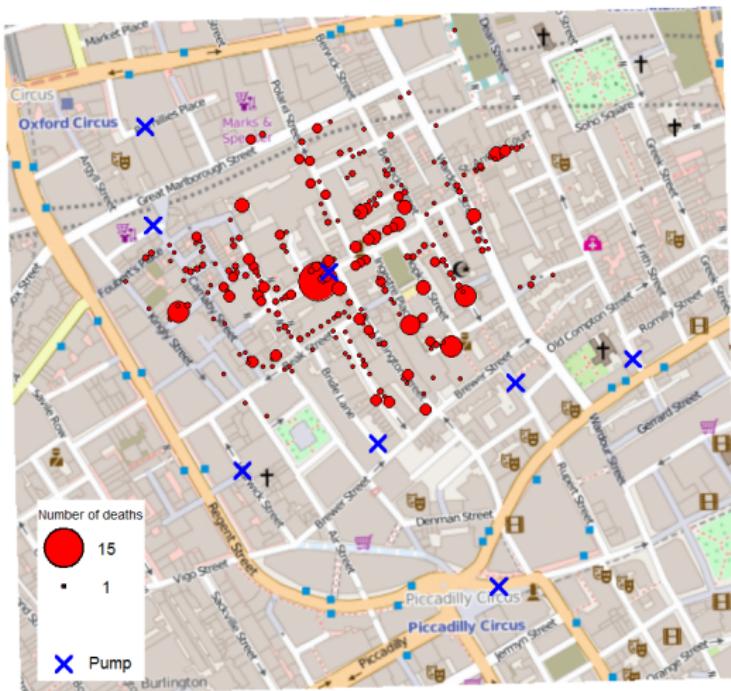
Spatial point process data - cholera outbreak



- The Broad Street cholera outbreak in Soho, London in 1854
- John Snow:
Contaminated water, not air, spread cholera
- The waterworks company supplying water to Broad Street pump was taking water from the sewage polluted area of the Thames river



Cholera data on an OpenStreetMap

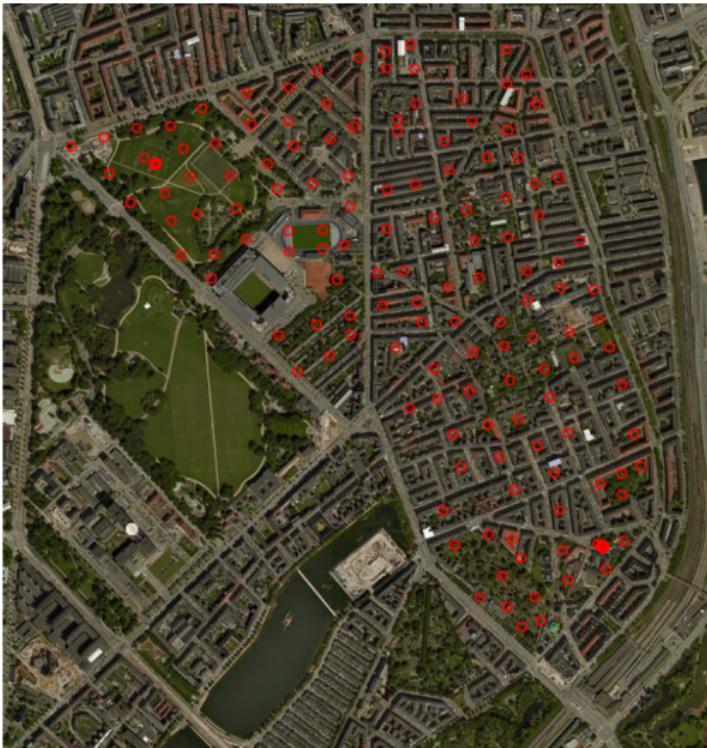


Soil pollution at Østerbro

Soil pollution at Østerbro

- Project about diffuse soil pollution
- Measurements of the concentration of lead and PAH in soil samples
- $138+33=171$ sampling locations

Sampling locations at Østerbro



Two clusters of sampling locations

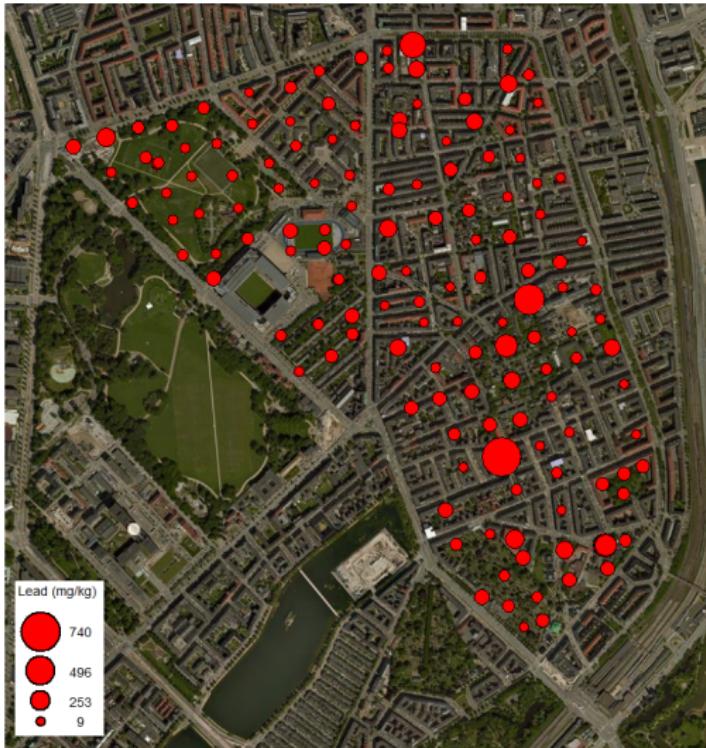
Fælledparken



By Kristianiagade



Observations of lead at Østerbro



Geostatistics (kriging)

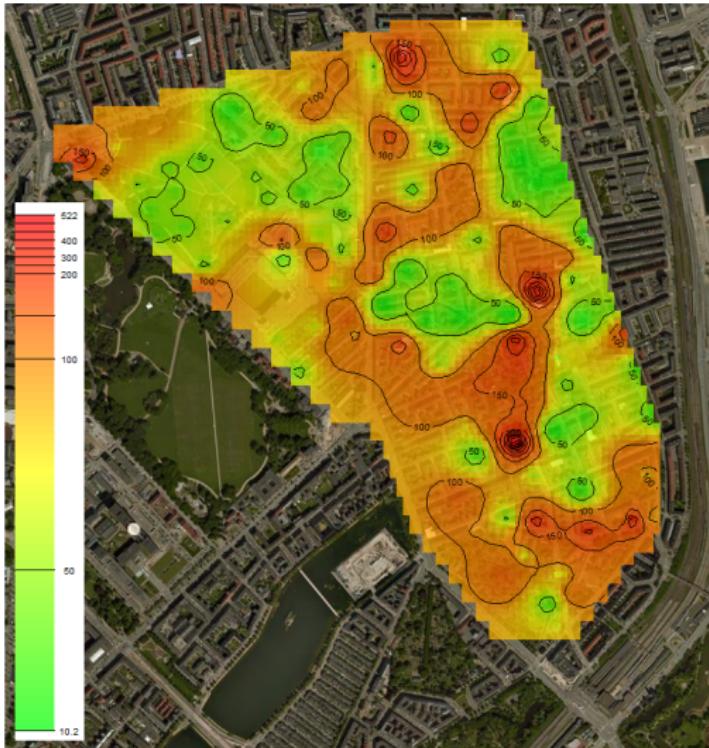
The Gaussian random field Z over a finite set of locations s_1, \dots, s_n is given by:

$$Z \sim \text{Gau}(\boldsymbol{\mu}, \Sigma)$$

with $\Sigma = (\sigma_{ij}^2)_{ij}$, $\sigma_{ij}^2 = \sigma^2 - \gamma(s_i - s_j)$, $\sigma^2 = \text{Var}(Z(s))$,
 $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$

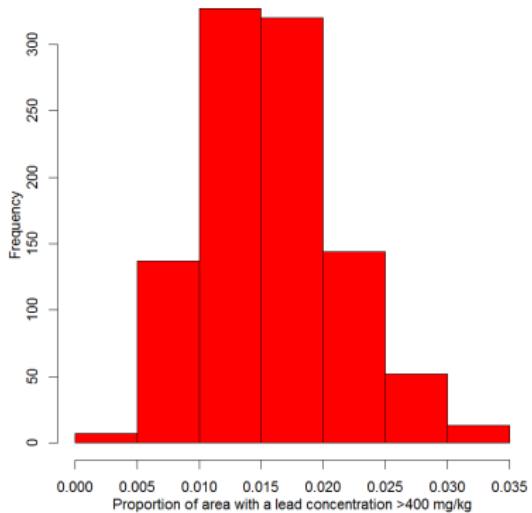
Spatial predictions can be computed using matrix inversion
and matrix multiplications.

Spatial prediction

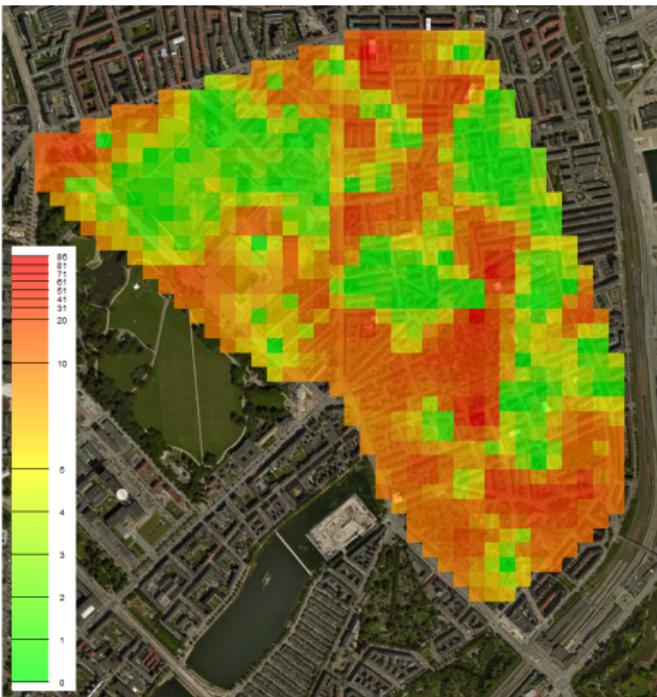


Other prediction targets - The size of the area with a concentration of lead above 400 mg/kg

- 1000 simulations from the model in the prediction locations conditional on the observations of lead
- For each simulation find the proportion of predictions above 400 mg/kg
- Summary for this proportion: min: 0.003601, mean: 0.015840, max: 0.036010
- Total area: 1761795 m²
- Area with lead concentration>400 mg/kg: min: 6344, mean: 27907 m², max: 63442 m²

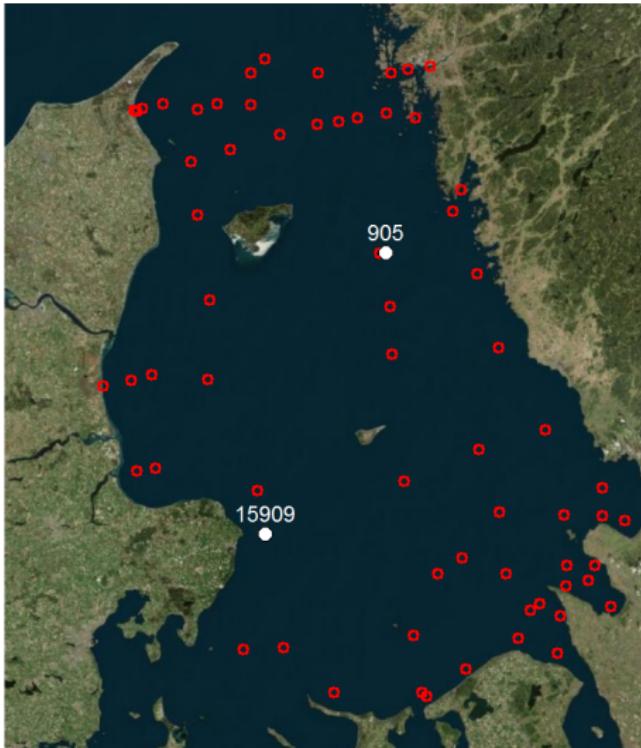


Other prediction targets - The probability of a lead concentration above 250 mg/kg

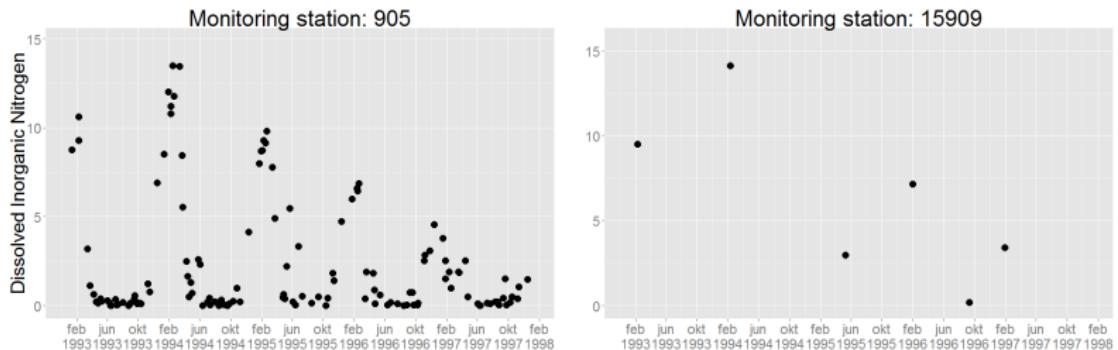


Dissolved Inorganic Nitrogen in Kattegat

Monitoring stations



Time series at two monitoring stations



Maps of DIN for two different days

