Ozone data and Model output
combined using Kriging with External Drift.

H. Wackernagel\textsuperscript{1}, C. Lajaunie\textsuperscript{1}, C. Roth\textsuperscript{2}, N. Blond\textsuperscript{3},
R. Vautard\textsuperscript{3}

\url{http://cg.ensmp.fr/~hans}

\textsuperscript{1}CG  Ecole des Mines de Paris,  \textsuperscript{2}Airparif,  \textsuperscript{3}LMD  Ecole Polytechnique

IMPACT Project

\url{http://www.mai.liu.se/impact}
Normalization of Model output with Station data

In many environmental applications:

- measurements at points of a region,
- gridded data covering the region.

**Example:** ground samples and remote sensing data.

This talk:

- combining **station data** and **model output**
  $\rightarrow$ multivariate geostatistics
- change of support problem
  between station data and model output
Multivariate Kriging

Kriging is a special type of optimal linear prediction applied to random functions in space or time with the particular requirement that their covariance structure should be known. → Cokriging
Multivariate Kriging

Kriging is a special type of optimal linear prediction applied to random functions in space or time with the particular requirement that their covariance structure should be known. → Cokriging

Covariance structure: variograms (or covariance functions) for a set of variables.
Multivariate Kriging

Kriging is a special type of optimal linear prediction applied to random functions in space or time with the particular requirement that their covariance structure should be known. \( \rightarrow \) Cokriging

Covariance structure: variograms (or covariance functions) for a set of variables.

Drift: translation-invariant polynomial drift, or external drift.
Data: Configuration & Neighborhood

Data configuration:
  sites of different types of measurements
  in a spatial/temporal domain.
Data: Configuration & Neighborhood

Data configuration:
sites of different types of measurements
in a spatial/temporal domain.

Neighborhood:
a subset of data used in cokriging.
Configurations: Iso- and Heterotopic Data

- **Primary data**
- **Secondary data**

### Isotopic data
- Sample sites are shared

### Heterotopic data
- Sample sites may be different

### Dense auxiliary data
- Secondary data covers whole domain
Kriging with external drift

Estimator: \[ Z^*(x_0) = \sum_{\alpha=1}^{n} w_{\alpha} Z(x_{\alpha}) \]

- constrained weights: \[ \sum w_{\alpha} = 1 \]
- auxiliary variable \( s(x) \) put as a constraint:
  \[ \sum_{\alpha} w_{\alpha} s(x_{\alpha}) = s(x_0) \]
- dense auxiliary data,
- linear relation with primary variable.
Configuration: dense auxiliary data

A: neighborhood using all data
B: multi-collocated neighborhood
C: collocated neighborhood
Neighborhood: multi-collocated


- all forms of cokriging: simple, ordinary, universal
- multi-collocated cokriging equivalent to full cokriging when proportionality in the cross-covariance model.

External drift: uses a multi-collocated neighborhood.
Neighborhood: multi-collocated

Cokriging with all data is equivalent to cokriging with multi-collocated neighborhood when (Rivoirard 2000):

\[ C_Z(h) = p^2 C(h) + C_R(h) \]
\[ C'_S(h) = C(h) \]
\[ C'_{ZS}(h) = p \, C(h) \]

where \( p \) is a proportionality coefficient, \( C_R(h) \) is the covariance of the residual, i.e. the difference between \( Z(x) \) and \( S(x) \).

- \( C_S(h) \) more regular than \( C_Z(h) \), if \( C_R(h) \) less regular than \( C(h) \),
- \( S \) is smoother than \( Z \).
CASE STUDY: Ozone in Paris area

- Ozone hourly values from 19 AirParif stations
- CHIMERE model output (LMD/IPSL):
  - simplified chemistry transport, forced by ECMWF weather forecasts (80 gaseous substances, 200 processes)
  - $50 \times 50$ Km$^2$ large scale resolution with $6 \times 6$ Km$^2$ subgrid in $150 \times 150$ km$^2$ square (Ile-de-France)

We compare two pollution events on 29 and 17 July 1999.
Ozone values at 19 AirParif Stations
CHIMERE model: 29 July ’99, 15h UT

Ozone: Ile de France (29 July ’99, 15h UTC)

Scale of pollution: continental plume
Variogram of Ozone on 29 July ’99 at 15h UT

LEFT: Map of highlighted pairs of stations

RIGHT: Variogram cloud
Variogram of RESIDUALS

RESIDUALS: AirParif Stations minus CHIMERE Model

Variogram cloud (+) and experimental variogram (—)

Interpretation: white noise, i.e. nugget effect model
Kriging with external drift: 29 July ’99 at 15h

KED model: no autocorrelation between residuals
The CHIMERE model is linearly transformed by KED: amounts mainly to adding a constant!
Ozone: Ile de France (17 July ’99, 15h UTC)

Scale of pollution: local plume (i.e. within domain)
Variogram of Ozone on 17 July ’99 at 15h UT

LEFT: Map of highlighted pairs of stations

RIGHT: Variogram cloud
Variogram of RESIDUALS

Variogram cloud (+) and experimental variogram (—)

Interpretation: autocorrelated residuals
Spherical or Cubic variogram model?

LEFT Model: 285 spherical(h, range=.4) + 10 nugget(h)
RIGHT Model: 273 cubic(h, range=.33)
Kriging with external drift: 17 July ’99 at 15h

**LEFT:** Spherical + Nugget variogram model

**RIGHT:** Cubic variogram model
KED model: autocorrelated residuals

→ deformation of CHIMERE surface.
KED vs CHIMERE

KED

CHIMERE model

Kriging with Ext. Drift

r = 0.94

TIES 2002/Genova p.23/37
AirParif Stations vs CHIMERE

AirParif Stations

CHIMERE model

Kriging with Ext. Drift

CHIMERE model

r = 0.83

r = 0.94
External drift and Change of Support

Basic unsolved question

The measurements from stations and the numerical model output are not on the same support:

- model output on $6 \times 6$ km$^2$ cells,
- measurements on a smaller support (which size?).

The variability of cell values should be lower than that of point values:

$\rightarrow$ change-of-support problem?
CASE STUDY: Geostatistical simulation of $O_3$

Simulation of the exponential of a Gaussian variable

Region $600 \times 800 \text{ Km}^2$

Cells $1 \times 1 \text{ Km}^2$

Variogram with a range of 50 Km
Simulation of Ozone: 1 × 1 Km² support

Km

Km

O3: 1x1km2

>=96
90
84
78
72
66
60
54
48
42
36
30
24
18
12
6
<0

ug/m³

TIES 2002/Genova  p.27/37
Simulation of Ozone: $10 \times 10$ Km$^2$ support
Simulation of Ozone: $20 \times 20 \text{ Km}^2$ support
Simulation of Ozone

Increasing the support: the means are equal, but the extremes and the variance are reduced.
Simulation of Ozone

Increasing the support:
the range increases
Simulation of Ozone

black = 1x1Km
blue = 10x10Km
red = 20x20Km
Simulation of Ozone

![Graph showing the proportion of area above a certain O3 cutoff in different areas.]

- Black line: 1x1Km²
- Blue line: 10x10Km²
- Red line: 20x20Km²

Proportion above cutoff (%) vs. O3 cutoff (ug/m³)
CASE STUDY: Ozone in Paris area

What about the support effect for KED and CHIMERE?
In this example, at least, KED does not alter the distribution of the CHIMERE model except for a shift in the mean.
CONCLUSION

Kriging with External Drift (KED) is a simple method for normalizing model output with station data.
CONCLUSION

- Kriging with External Drift (KED) is a simple method for normalizing model output with station data.

- In our example:
CONCLUSION

- Kriging with External Drift (KED) is a simple method for normalizing model output with station data.

- In our example:
  - KED recalibrates the mean of the model output,
CONCLUSION

- Kriging with External Drift (KED) is a simple method for normalizing model output with station data.

- In our example:
  - KED recalibrates the mean of the model output,
  - the shape of the histogram is preserved: no support effect for the normalized model output.
References

Selected references

Air pollution case study
http://cg.ensmp.fr/~hans/projects.html

Ebro estuary case study