Ozone data and Model output

combined using Kriging with External Drift .

H. Wackernagel[1], C. Lajaunie[1], C. Roth[2], N. Blond[3], R.Vautard[3]

http://cg.ensmp.fr/∼hans

[1]CG – Ecole des Mines de Paris, [2]Airparif, [3]LMD – Ecole Polytechnique

IMPACT Project

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. $\longrightarrow$ 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. $\longrightarrow$ 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. $\longrightarrow$ 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.

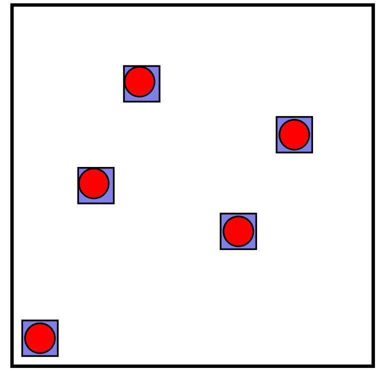
Neighborhood:

a subset of data used in cokriging.

Configurations: Iso- and Heterotopic Data

primary data   secondary data

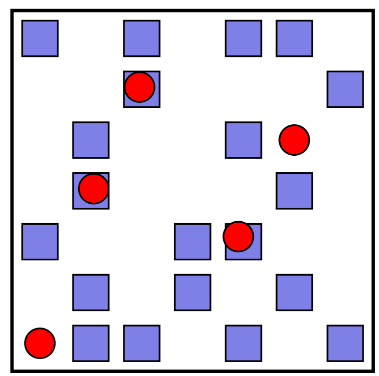
Isotopic data



Sample sites

are shared

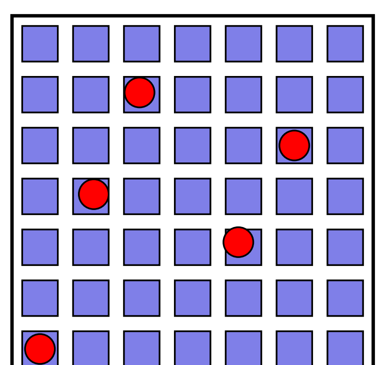
Heterotopic data



Sample sites

may be different

Dense auxiliary data



Secondary data

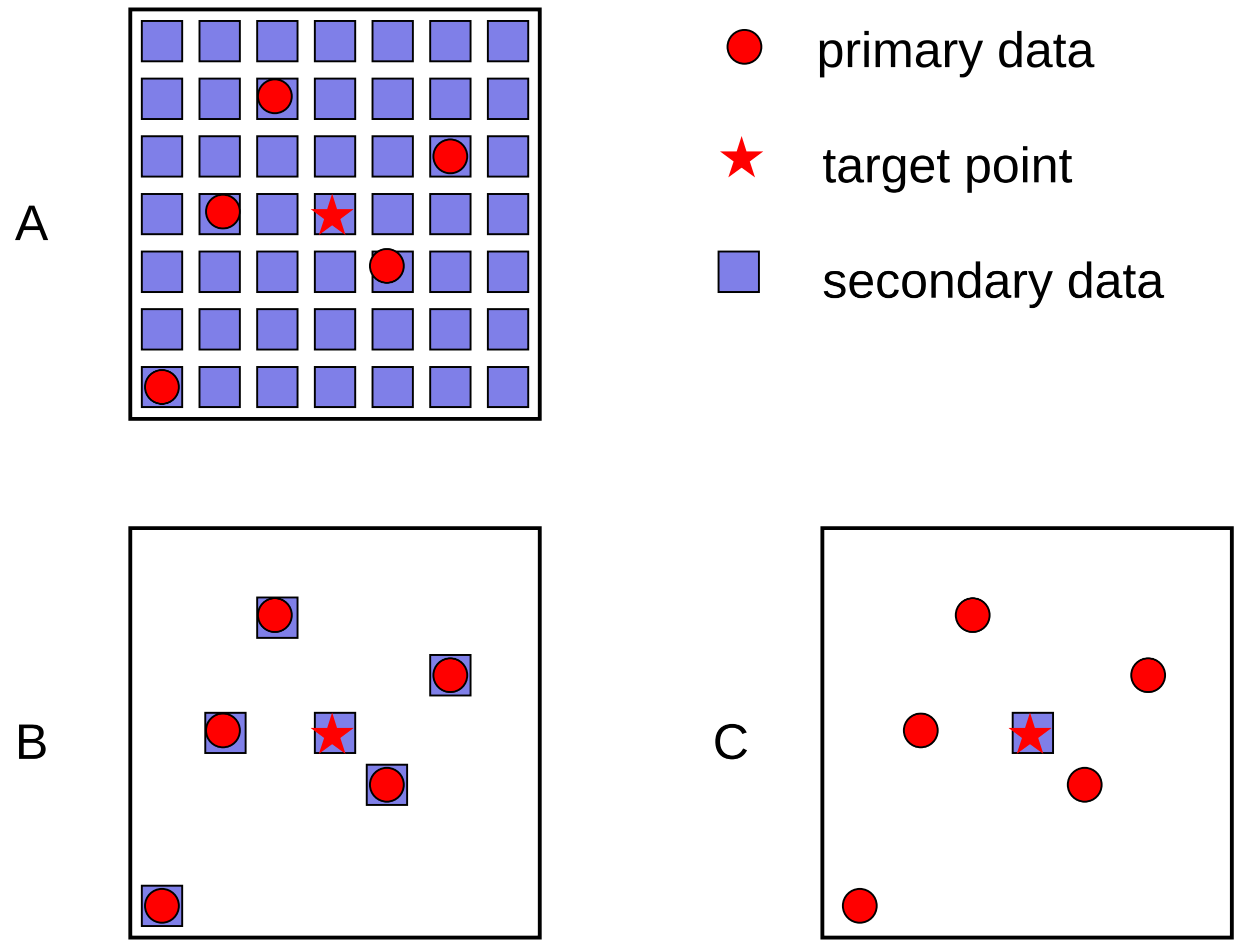
covers whole domain

Kriging with external drift

Estimator: $Z^{\star}(\mathbf{x}_0) \;=\; \sum_{\alpha=1}^{n} w_\alpha \, Z(\mathbf{x}_\alpha)$

- constrained weights: $\sum w_\alpha = 1$
- auxiliary variable $s(\mathbf{x})$ put as a constraint:
  $\sum_\alpha w_\alpha \, s(\mathbf{x}_\alpha) = s(\mathbf{x}_0)$
- dense auxiliary data,
- linear relation with primary variable.

Configuration: dense auxiliary data

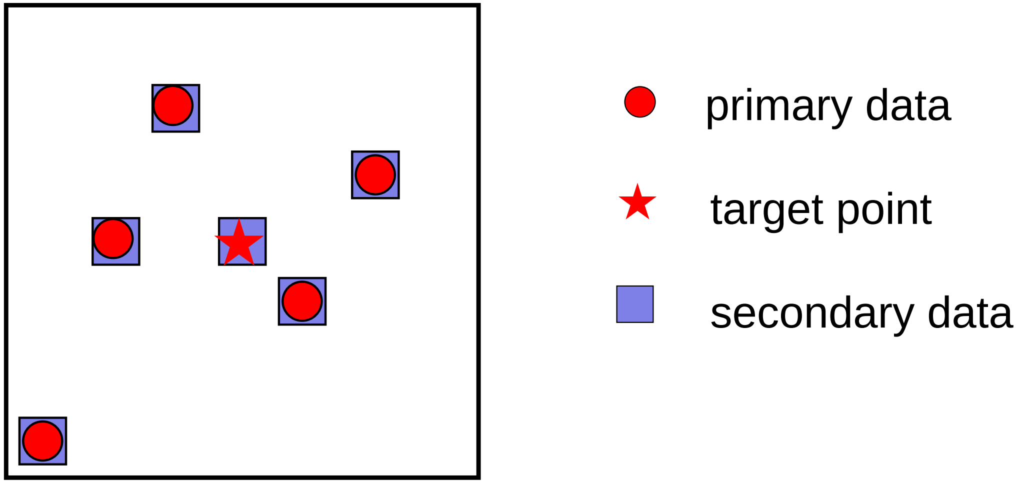


A: neighborhood using all data

B: multi-collocated neighborhood

C: collocated neighborhood

Neighborhood: multi-collocated



Chilès & Delfiner (1999), Rivoirard (2000)

- 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):

$$\begin{aligned} C_Z(\mathbf{h}) &= p^2\, C(\mathbf{h}) + C_R(\mathbf{h}) \\ C_S(\mathbf{h}) &= C(\mathbf{h}) \\ C_{ZS}(\mathbf{h}) &= p\, C(\mathbf{h}) \end{aligned}$$

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

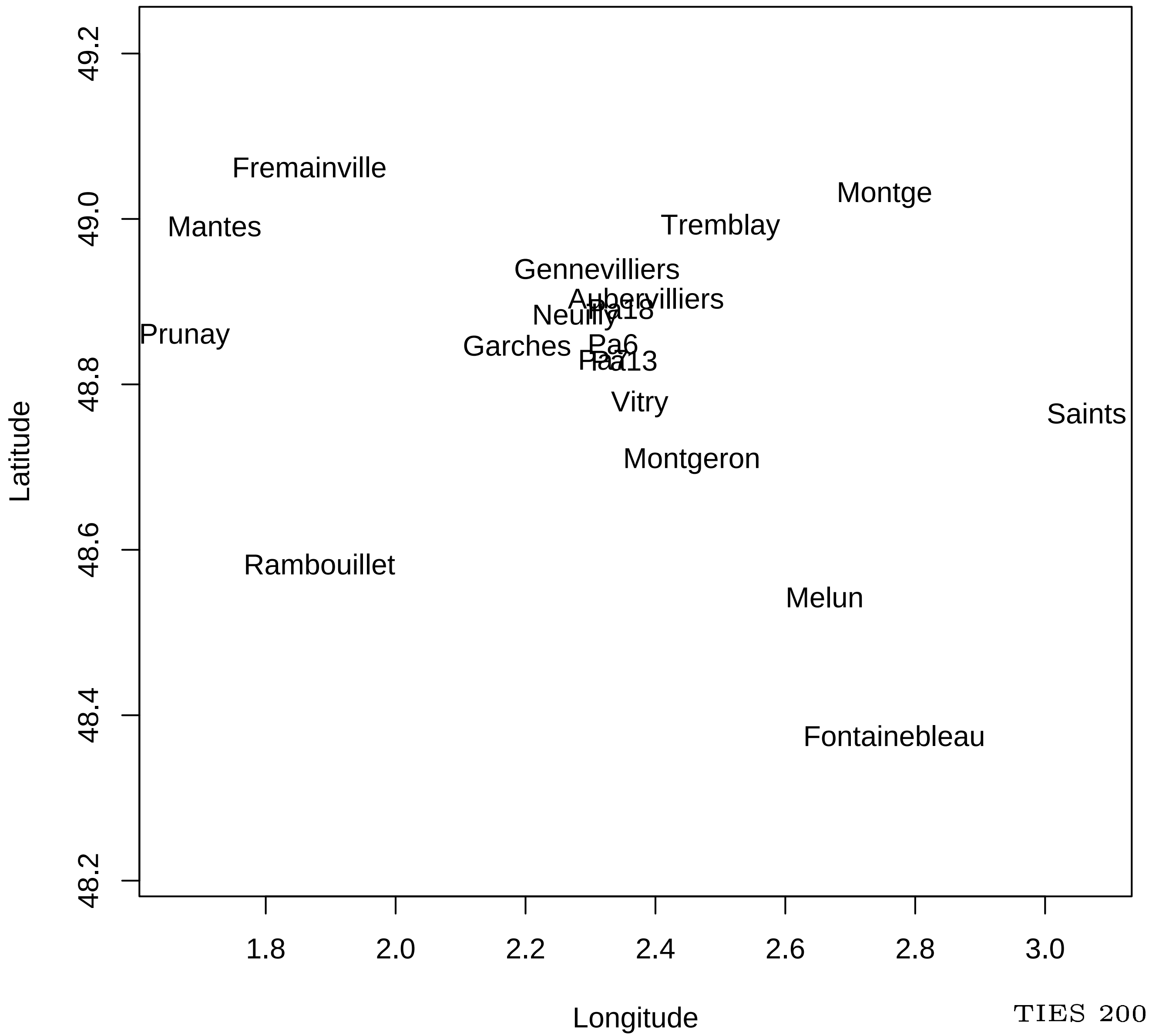
- $C_S(\mathbf{h})$ more regular than $C_Z(\mathbf{h})$,
  if $C_R(\mathbf{h})$ less regular than $C(\mathbf{h})$,
- S is smoother than Z.

CASE STUDY: Ozone in Paris area

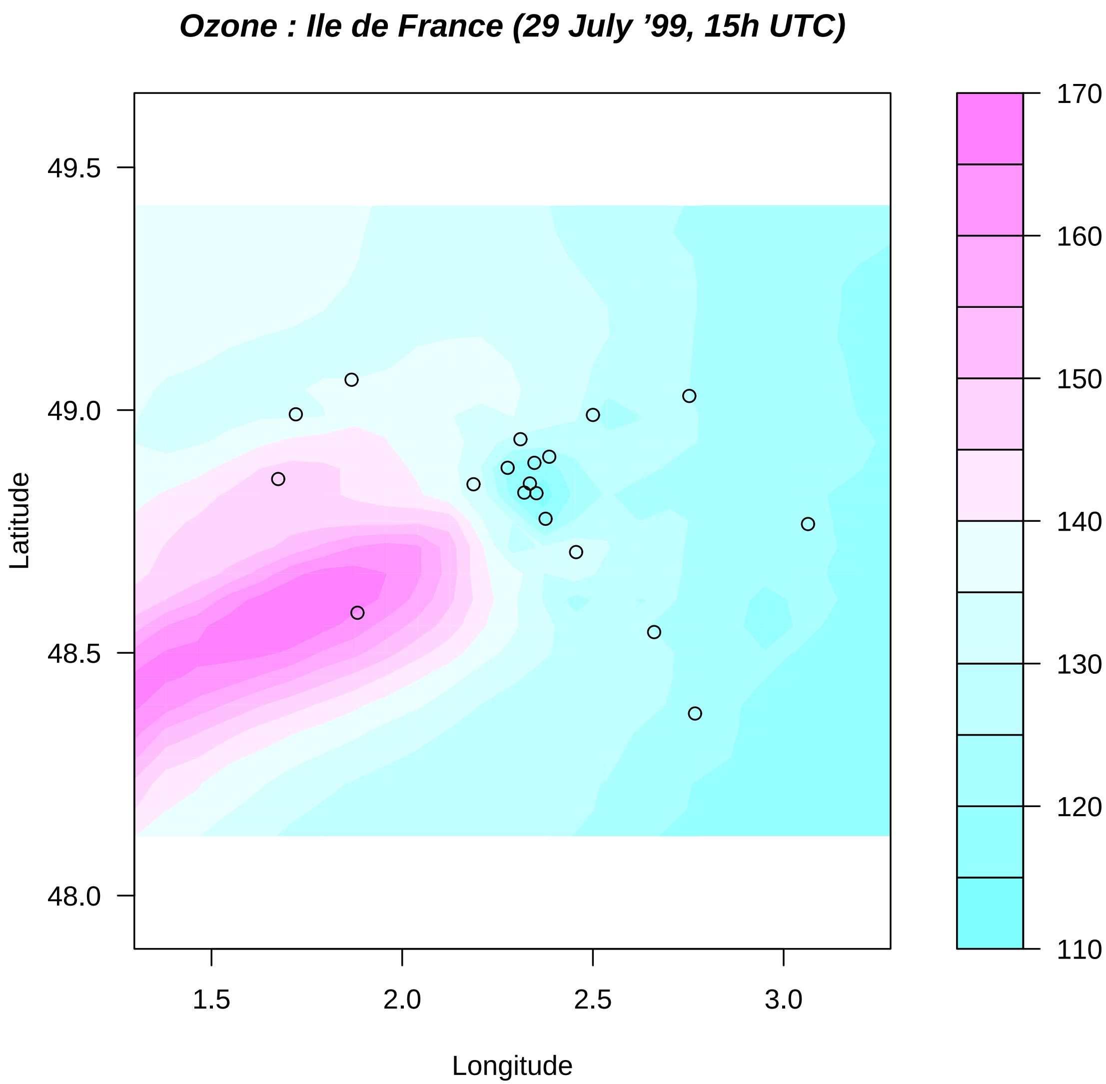
- Ozone hourly values from 19 AirParif stations
- CHIMERE model output (LMD/IPSL):
  - simplified chemistry transport, forced by ECMWF weather forecasts (80 gaseous substances, 200 processes)
  - 50×50 Km^2 large scale resolution
    with 6×6 Km^2 subgrid in 150×150 km^2 square (Ile-de-France)

We compare two pollution events on 29 and 17 July 1999.

Ozone values at 19 AirParif Stations

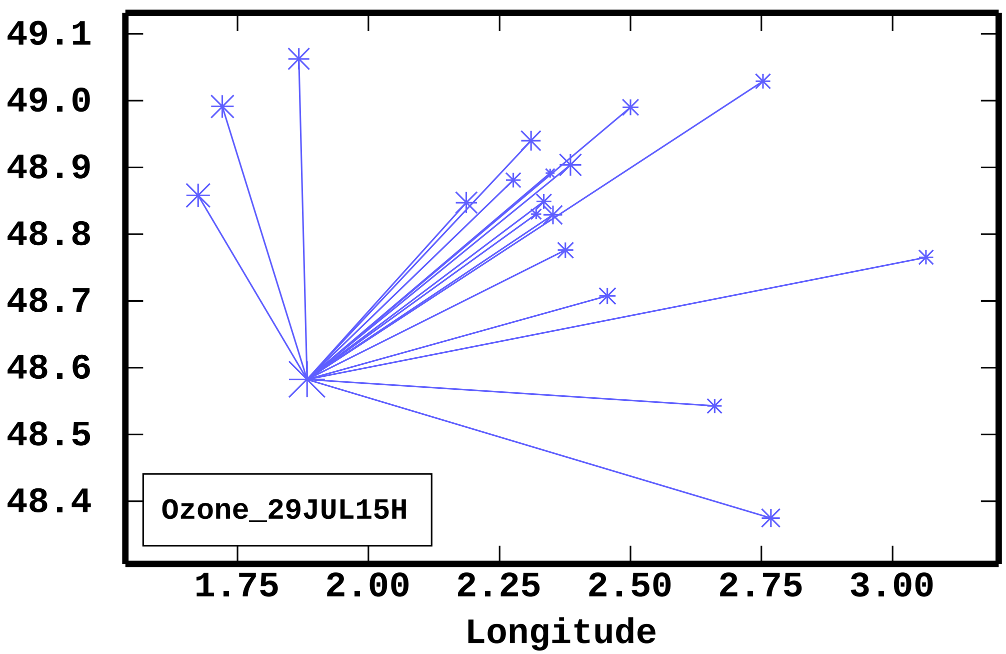


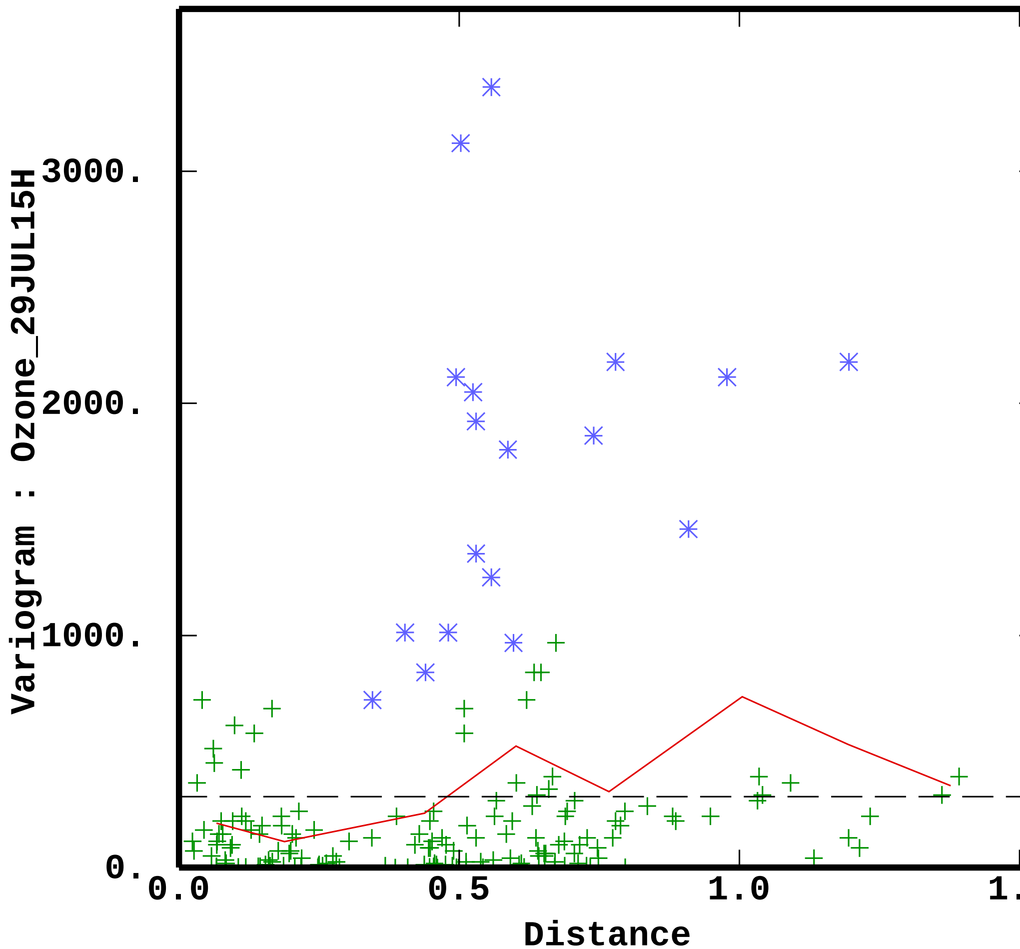
CHIMERE model: 29 July ’99, 15h UT



Scale of pollution: continental plume

Variogram of Ozone on 29 July '99 at 15h UT



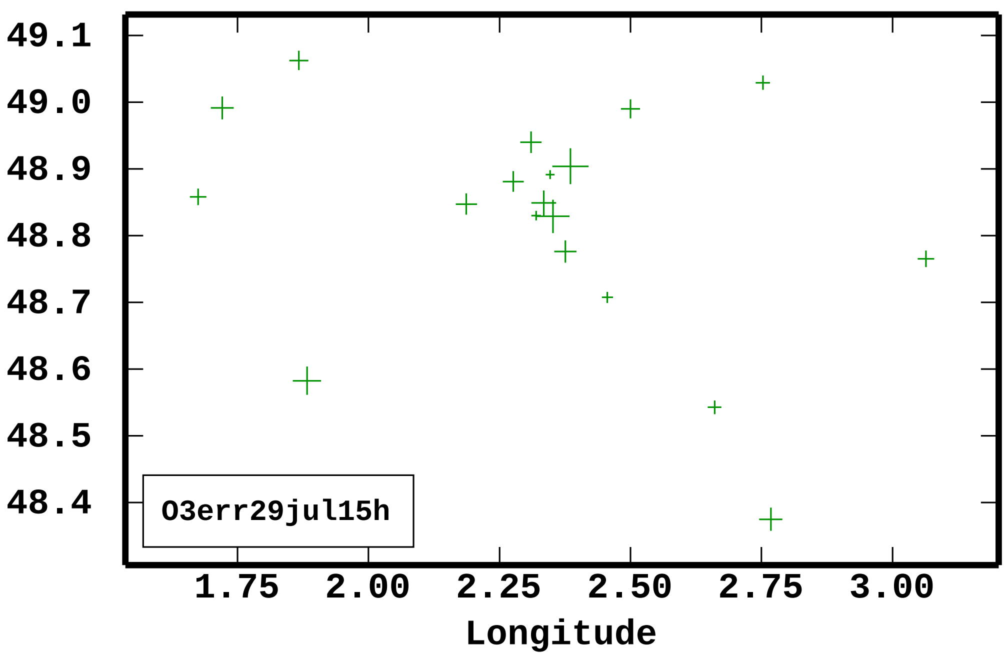


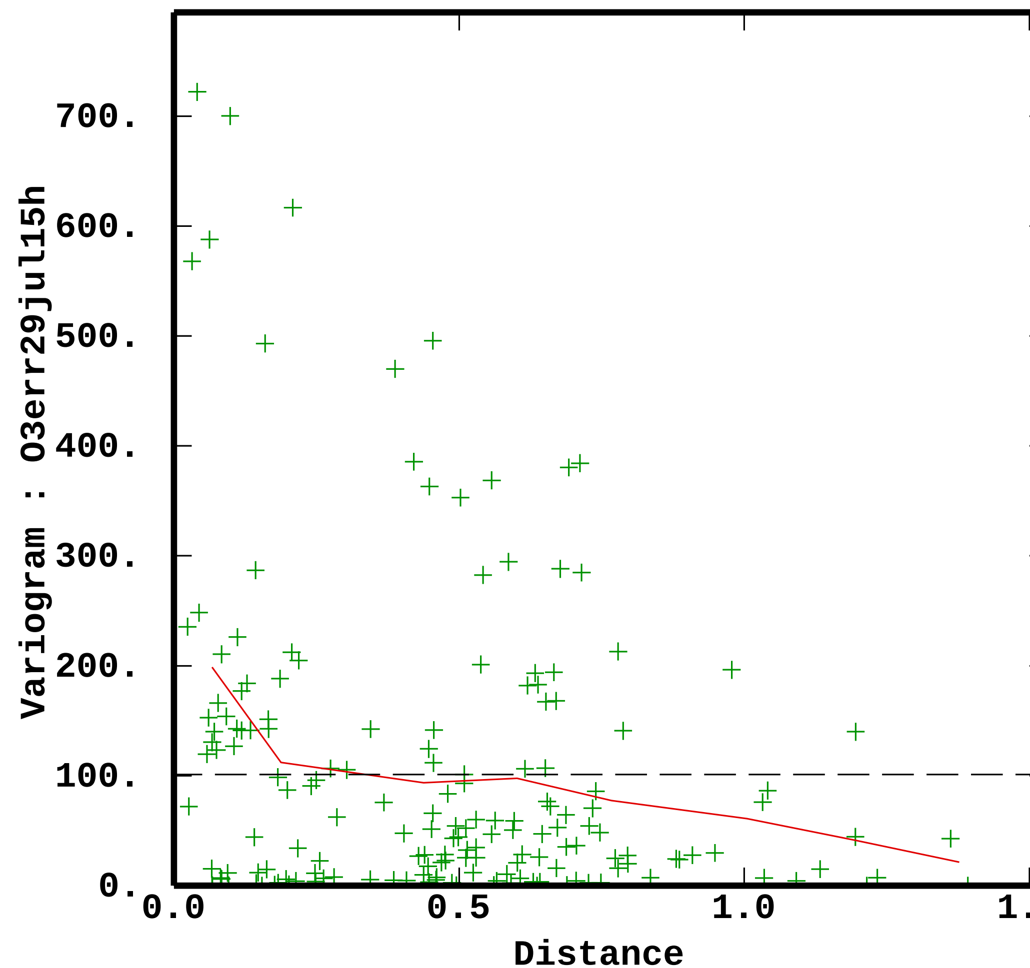
LEFT: Map of highlighted pairs of stations

RIGHT: Variogram cloud

Variogram of RESIDUALS

RESIDUALS: AirParif Stations minus CHIMERE Model

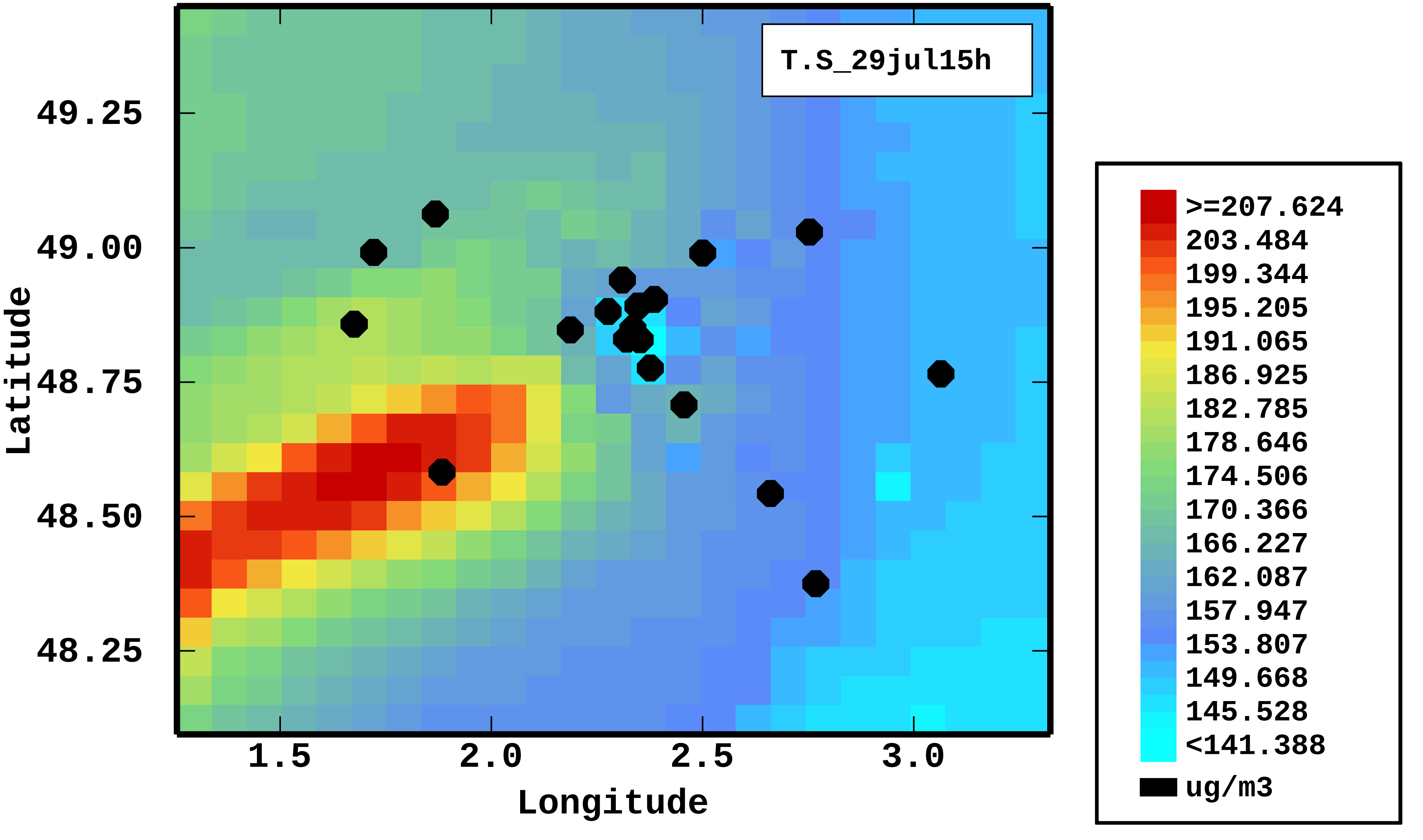




Variogram cloud (+) and experimental variogram (—)

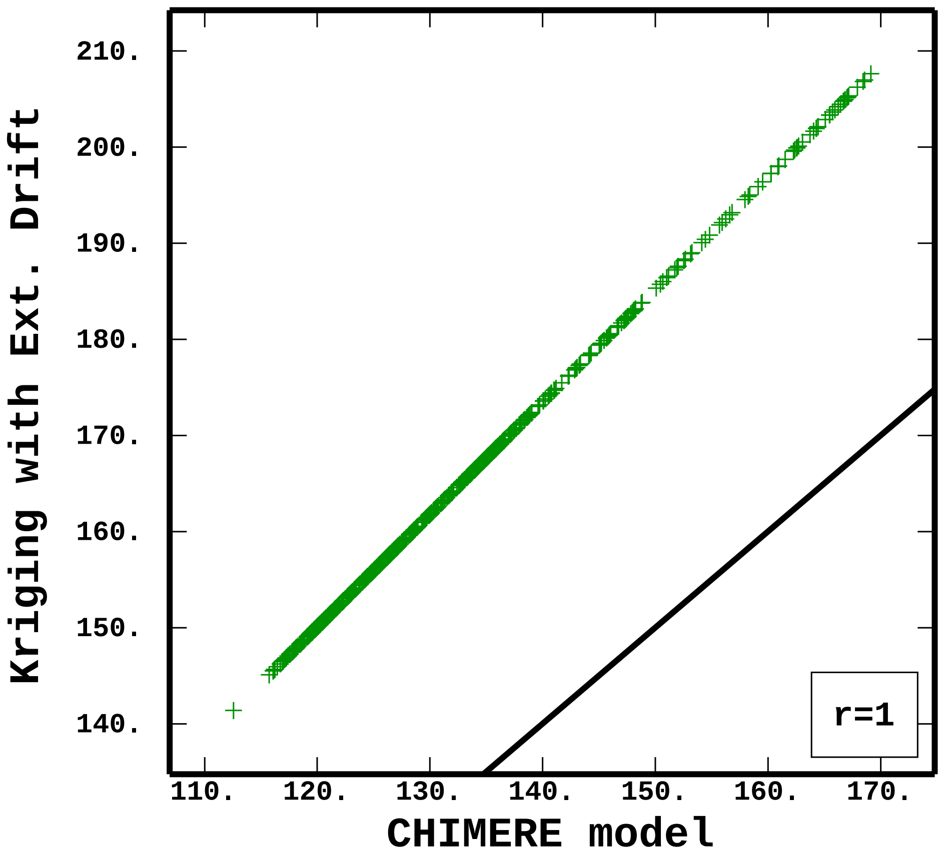
Interpretation: white noise, i.e. nugget effect model

Kriging with external drift: 29 July ’99 at 15h



KED model: no autocorrelation between residuals

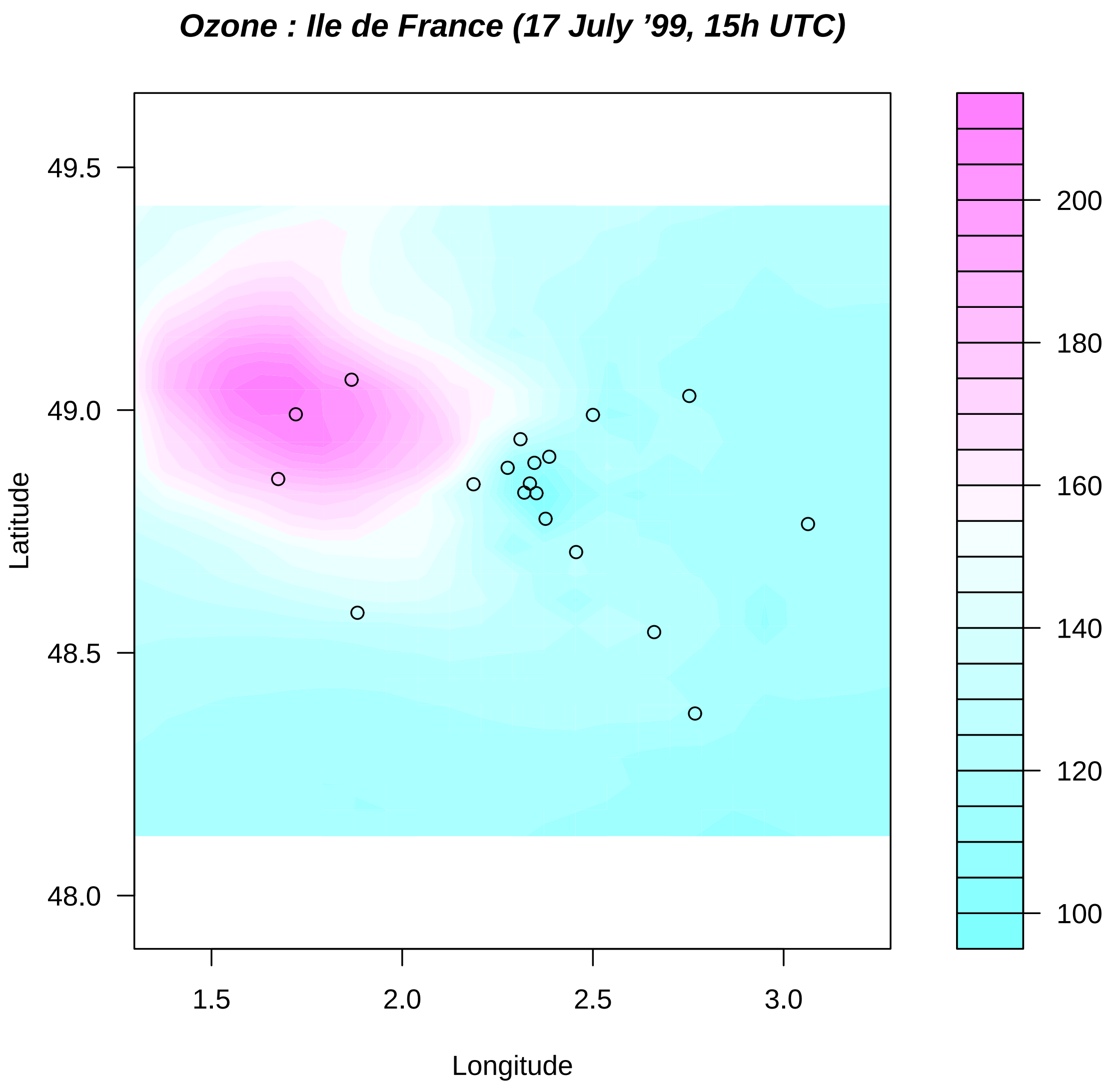
KED vs CHIMERE



The CHIMERE model is linearly transformed by KED:

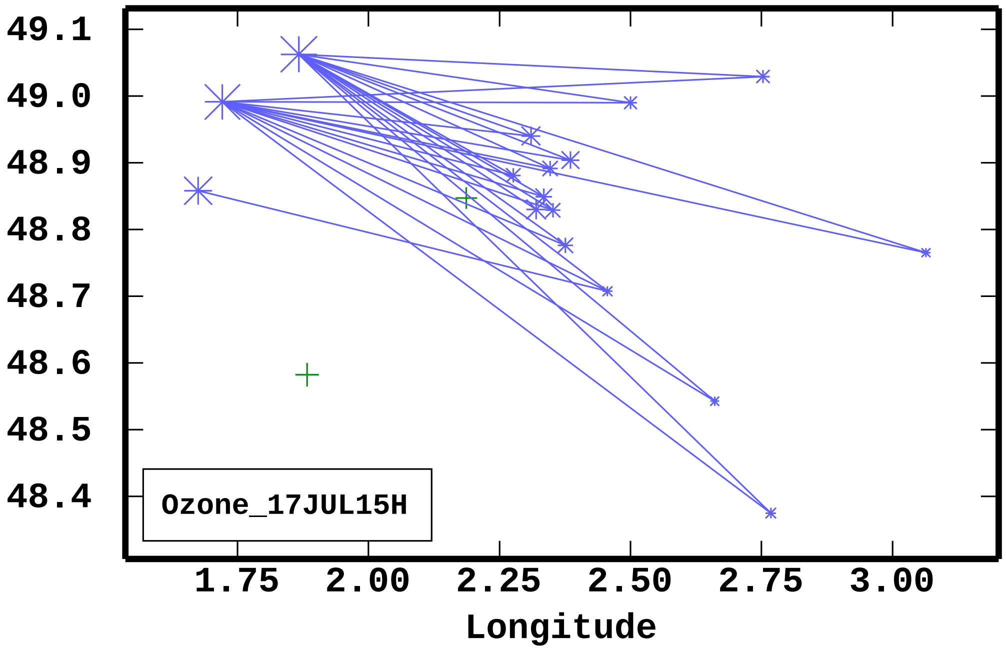
amounts mainly to adding a constant!

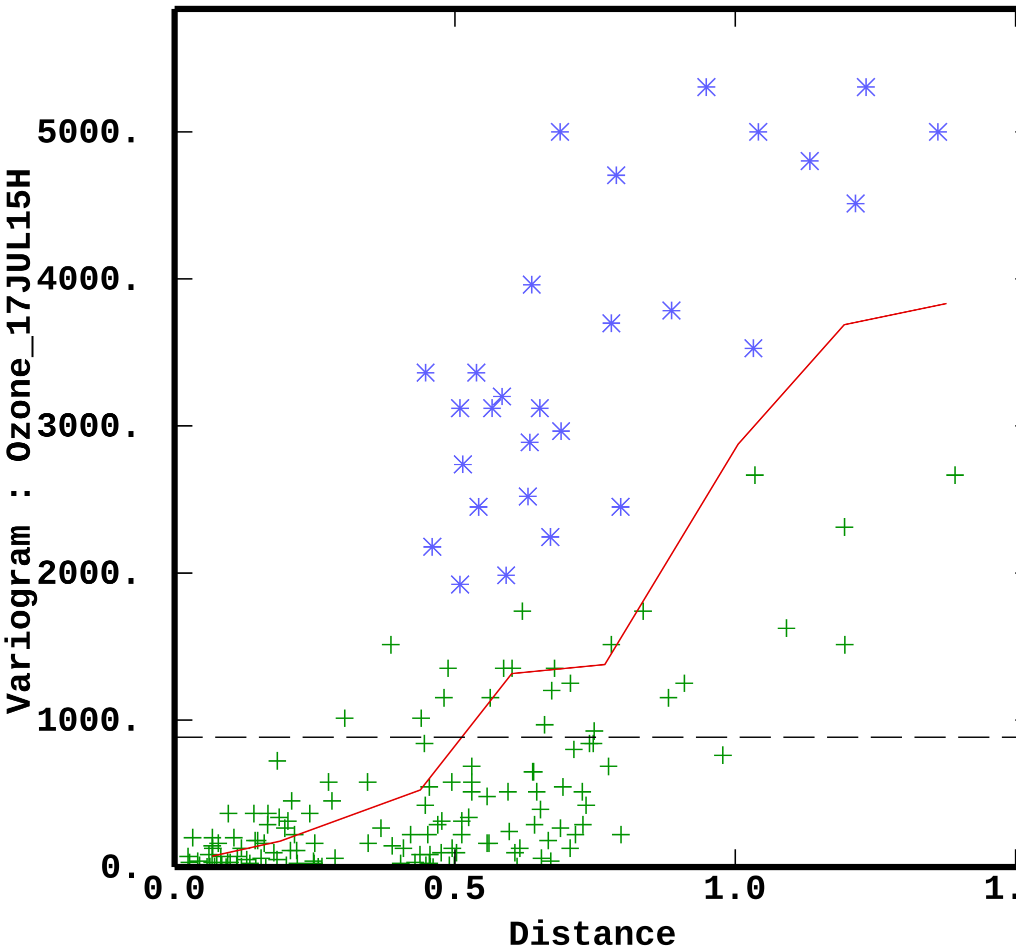
CHIMERE model: 17 July ’99, 15h UT



Scale of pollution: local plume (i.e. within domain)

Variogram of Ozone on 17 July '99 at 15h UT

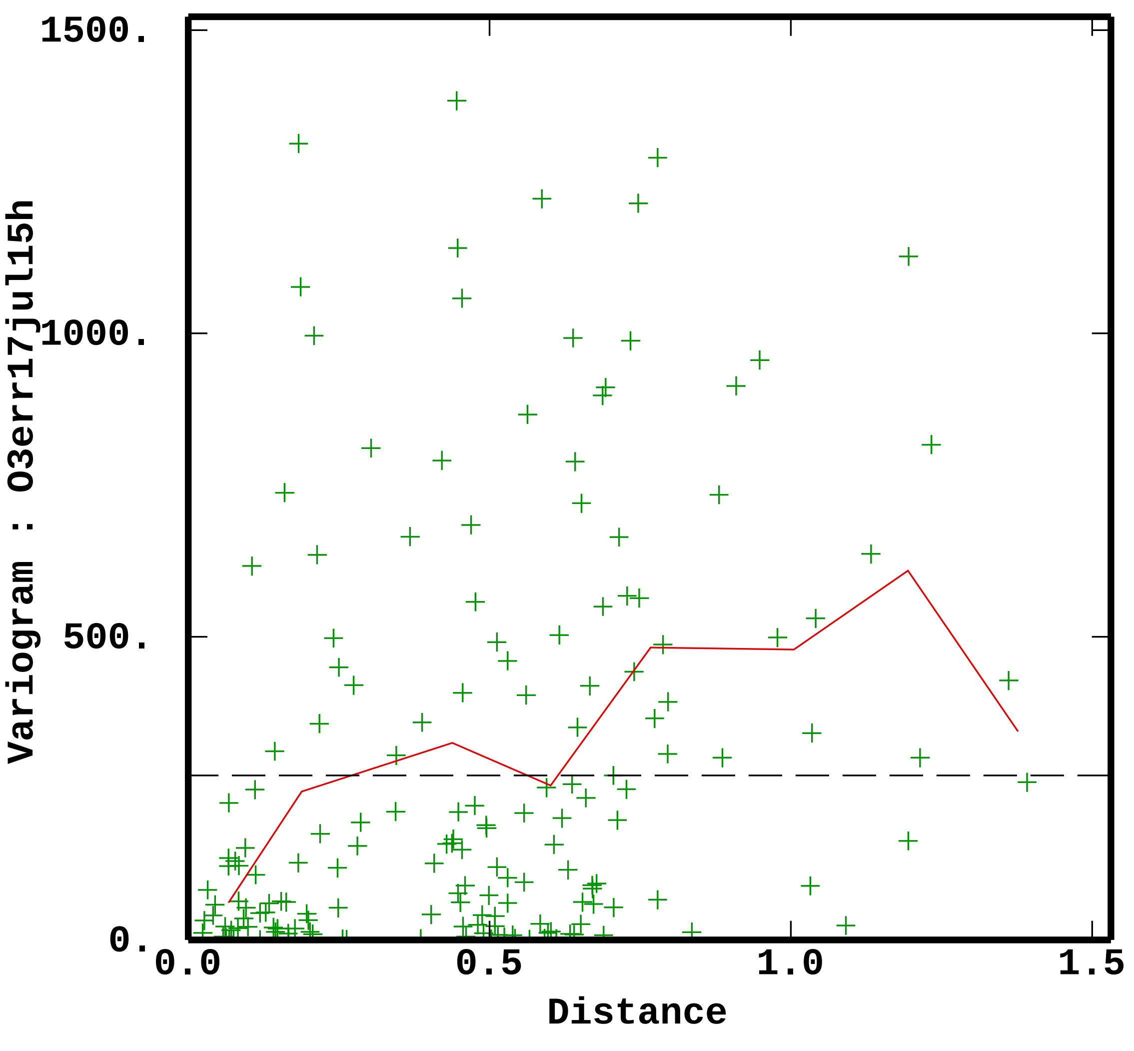




LEFT: Map of highlighted pairs of stations

RIGHT: Variogram cloud

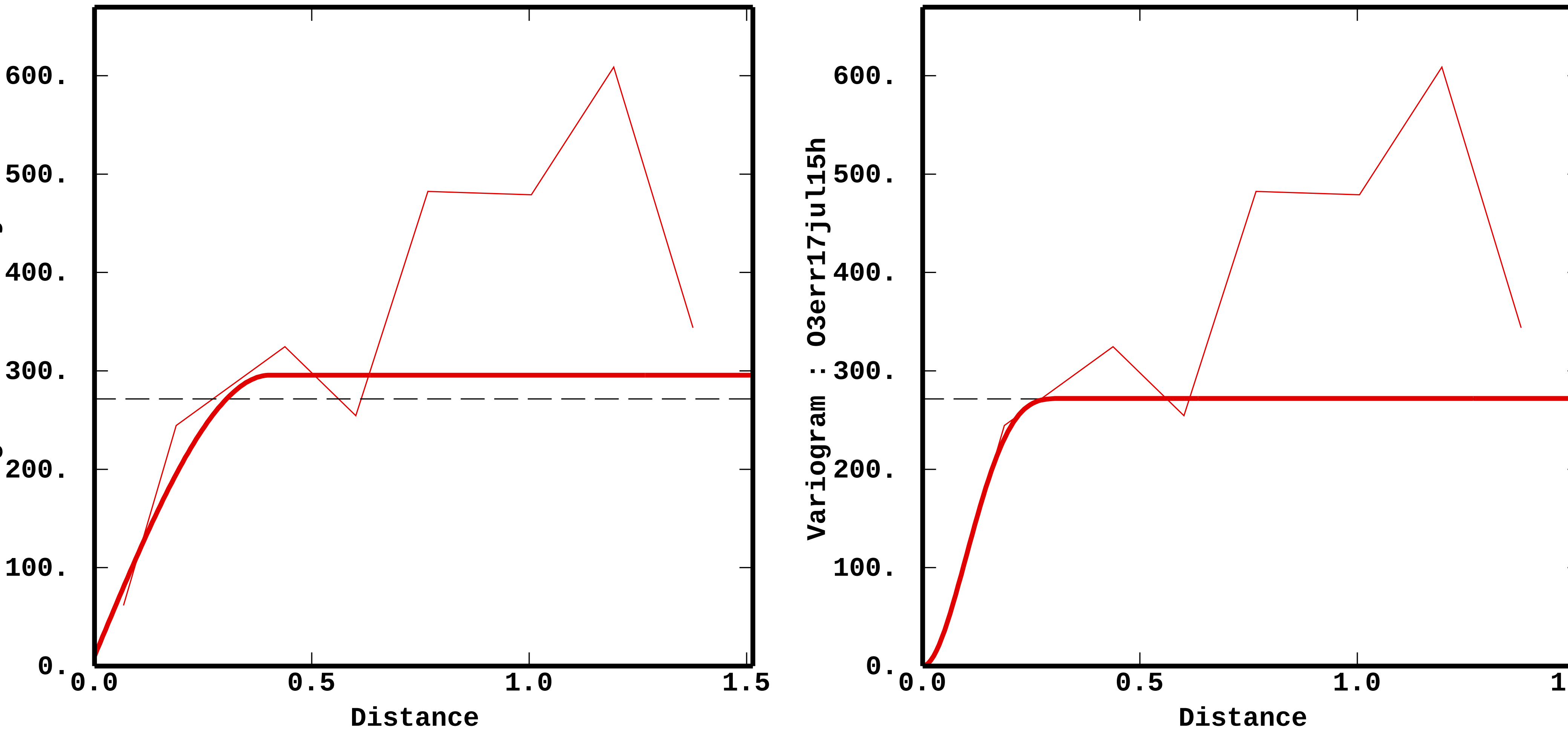
Variogram of RESIDUALS



Variogram cloud (+) and experimental variogram (—)

Interpretation: autocorrelated residuals

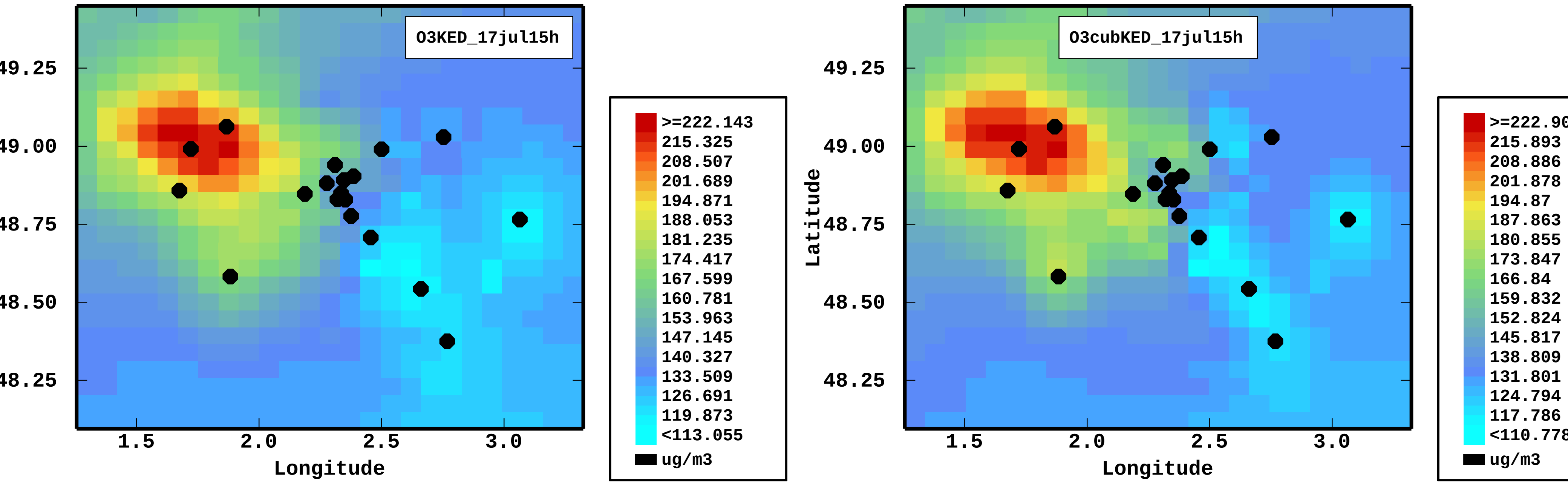
Spherical or Cubic variogram model?



LEFT Model: 285 spherical(h, range=.4) + 10 nugget(h)

RIGHT Model: 273 cubic(h, range=.33)

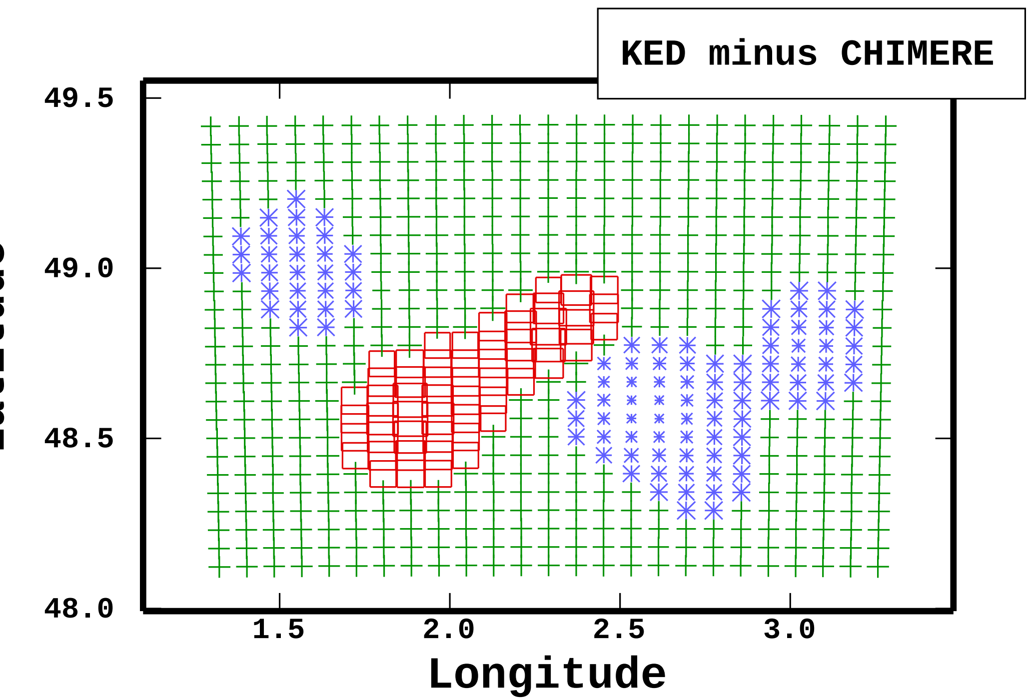
Kriging with external drift: 17 July '99 at 15h

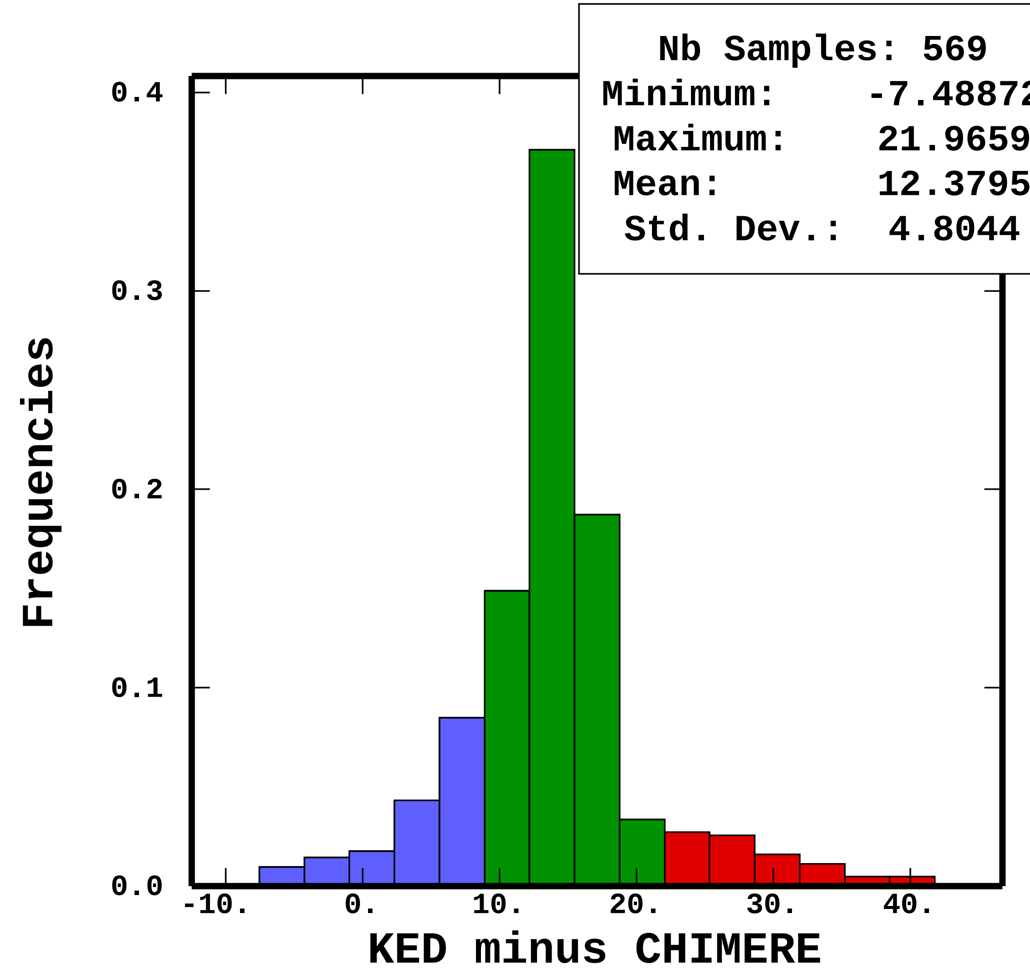


LEFT: Spherical + Nugget variogram model

RIGHT: Cubic variogram model

KED minus CHIMERE

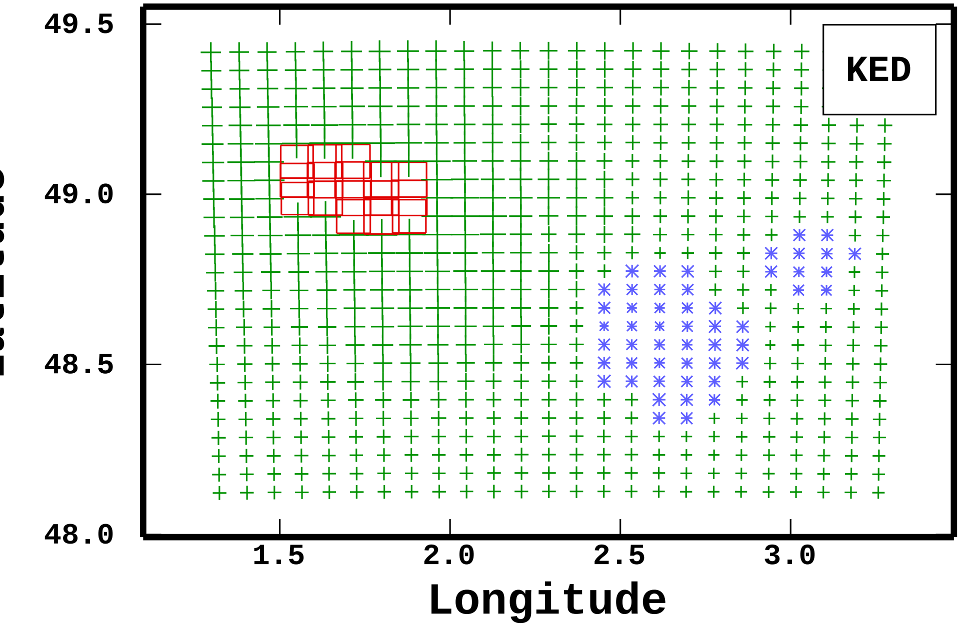


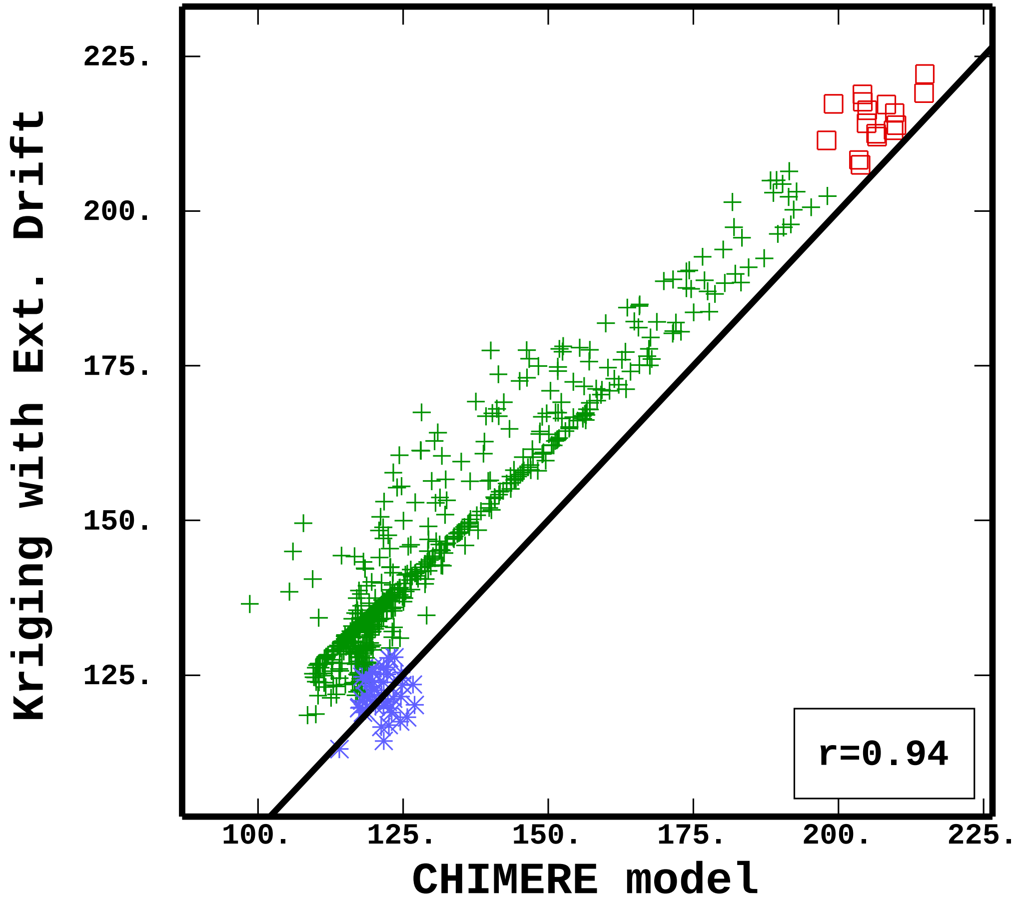


KED model: autocorrelated residuals

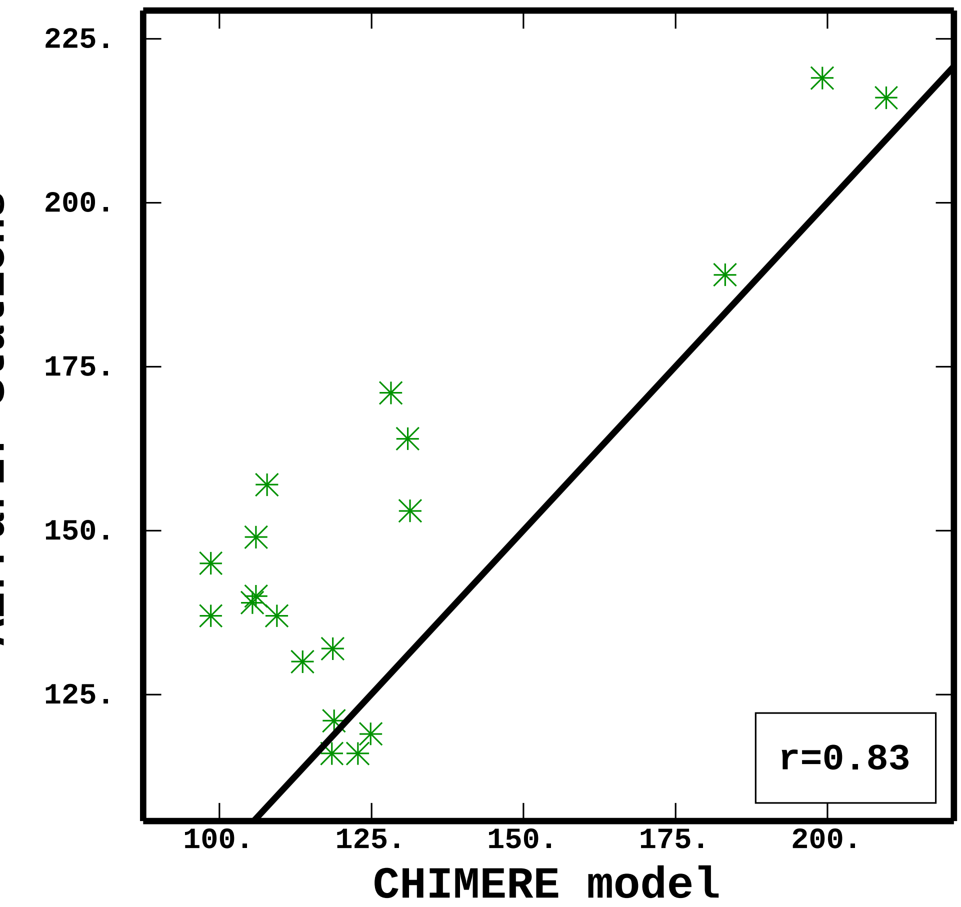
→ deformation of CHIMERE surface.

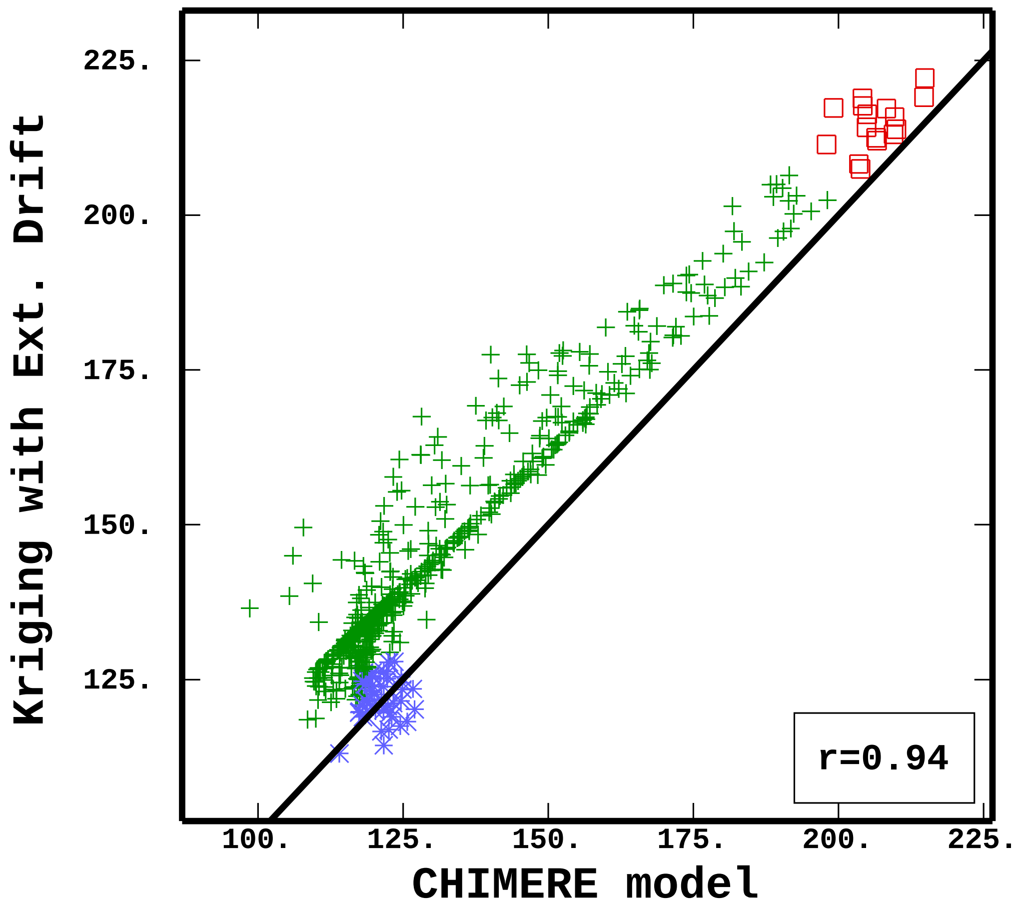
KED vs CHIMERE





AirParif Stations vs CHIMERE





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 6×6 km^2 cells,
- *measurements* on a smaller support (which size?).

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

$\longrightarrow$ change-of-support problem?

CASE STUDY: Geostatistical simulation of O_3

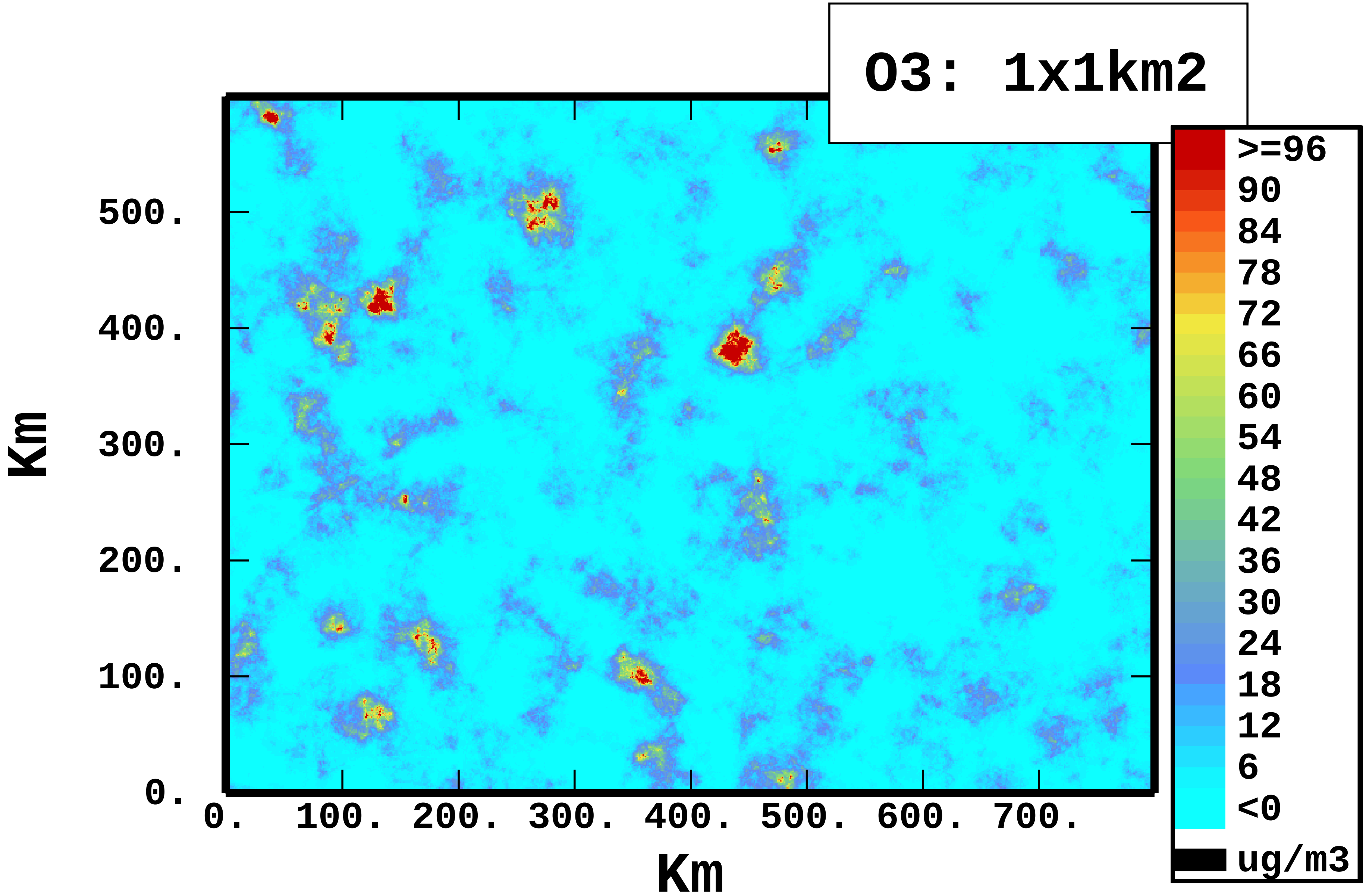
Simulation of the exponential of a Gaussian variable

Region 600×800 Km^2

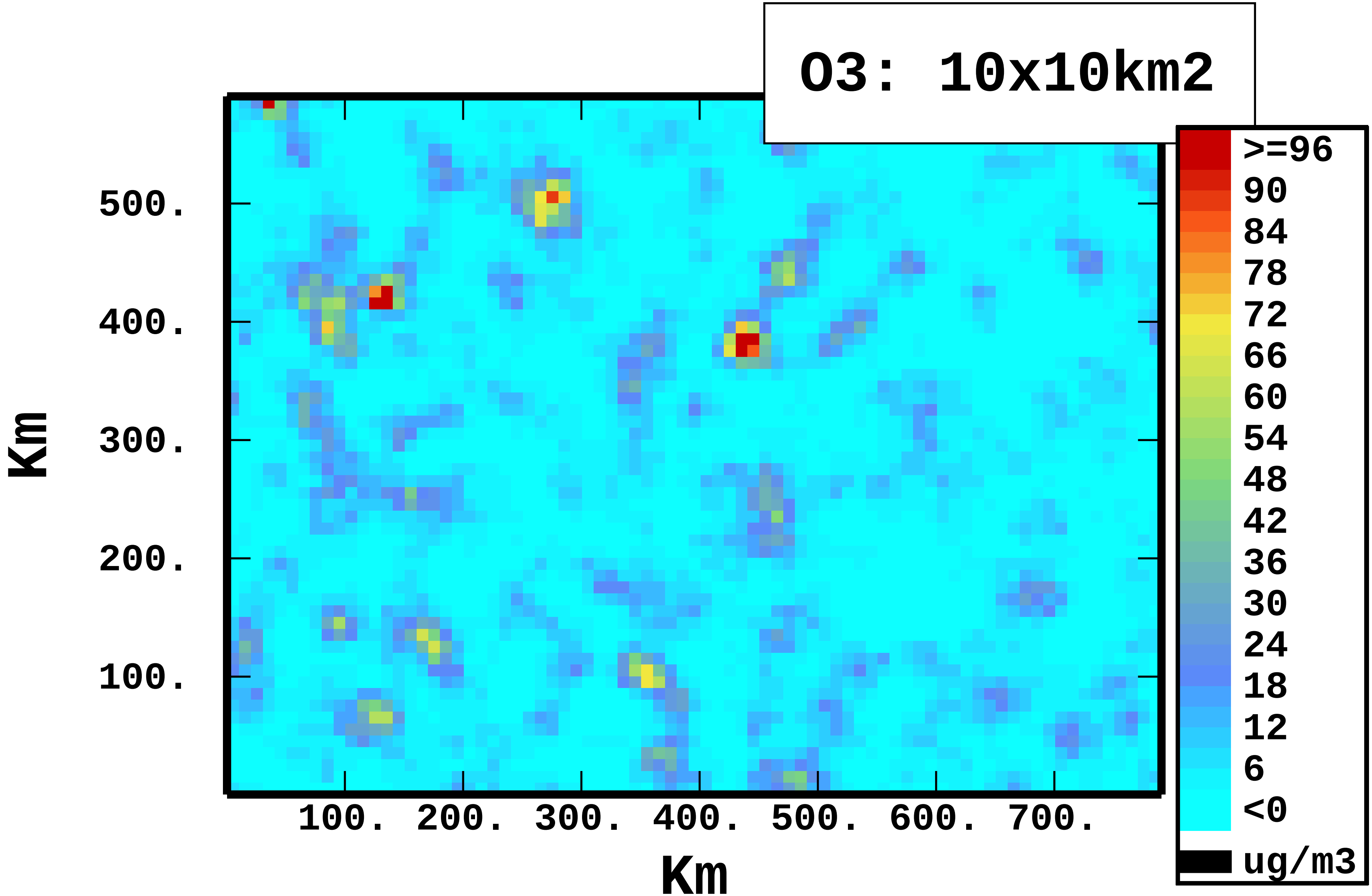
Cells 1×1 Km^2

Variogram with a range of 50 Km

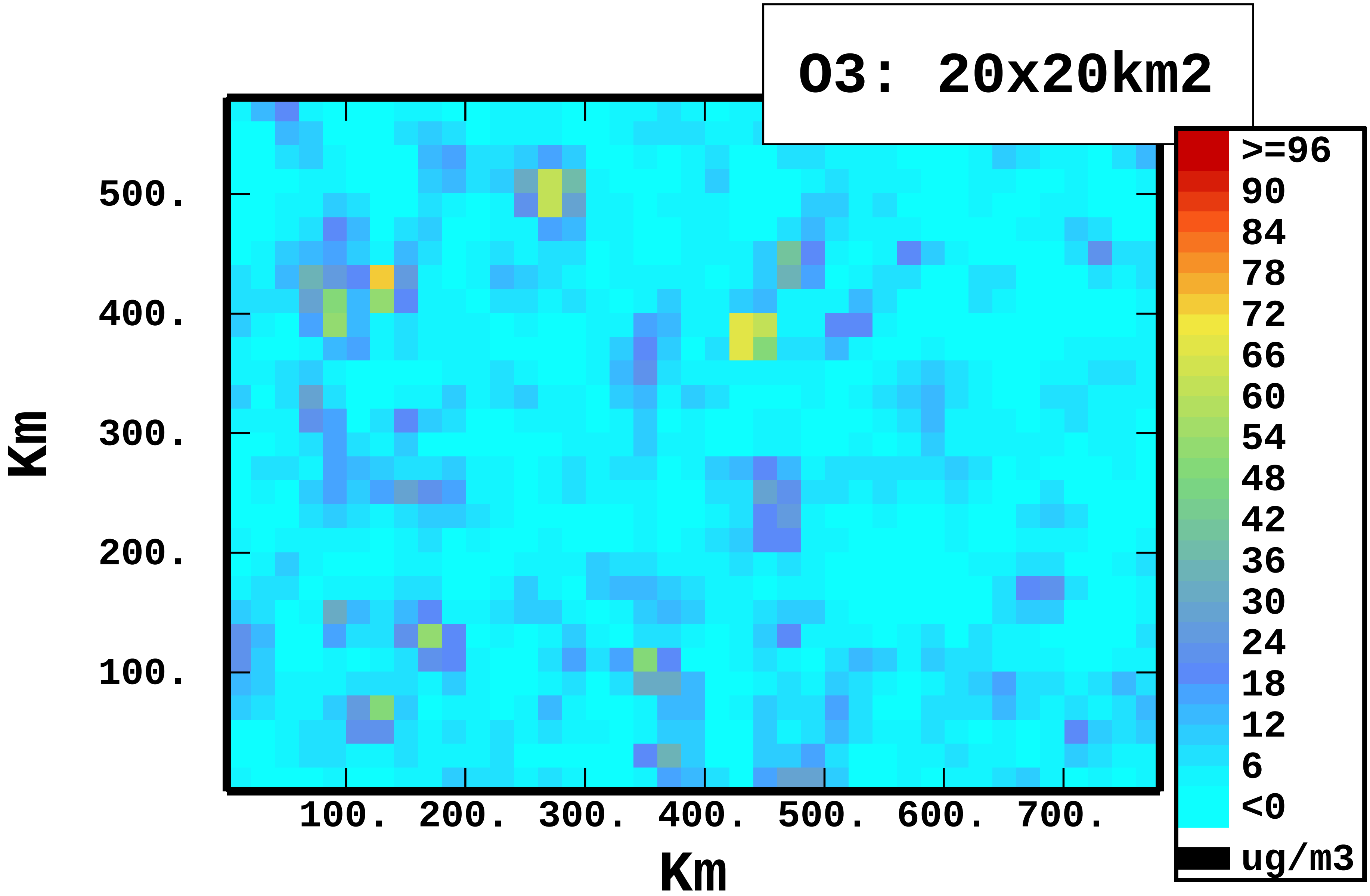
Simulation of Ozone: 1×1 Km^2 support



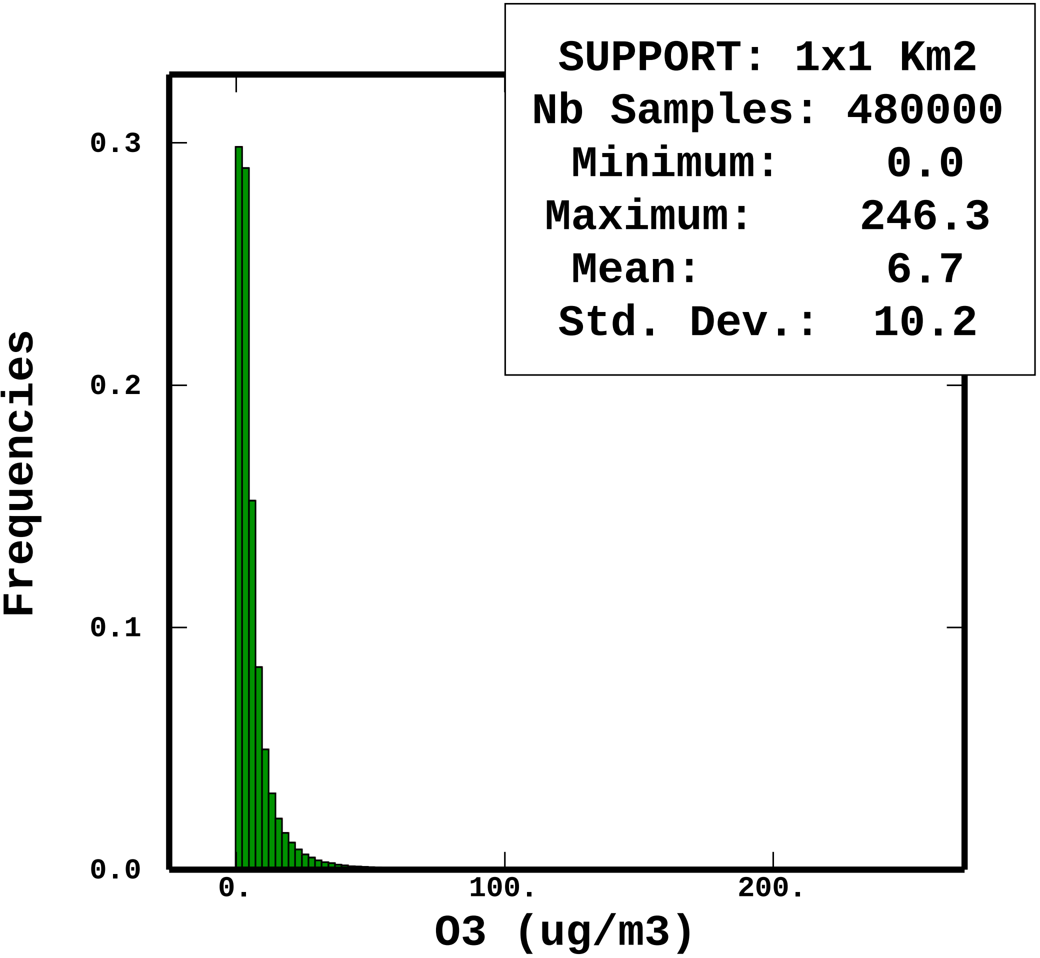
Simulation of Ozone: 10×10 Km^2 support

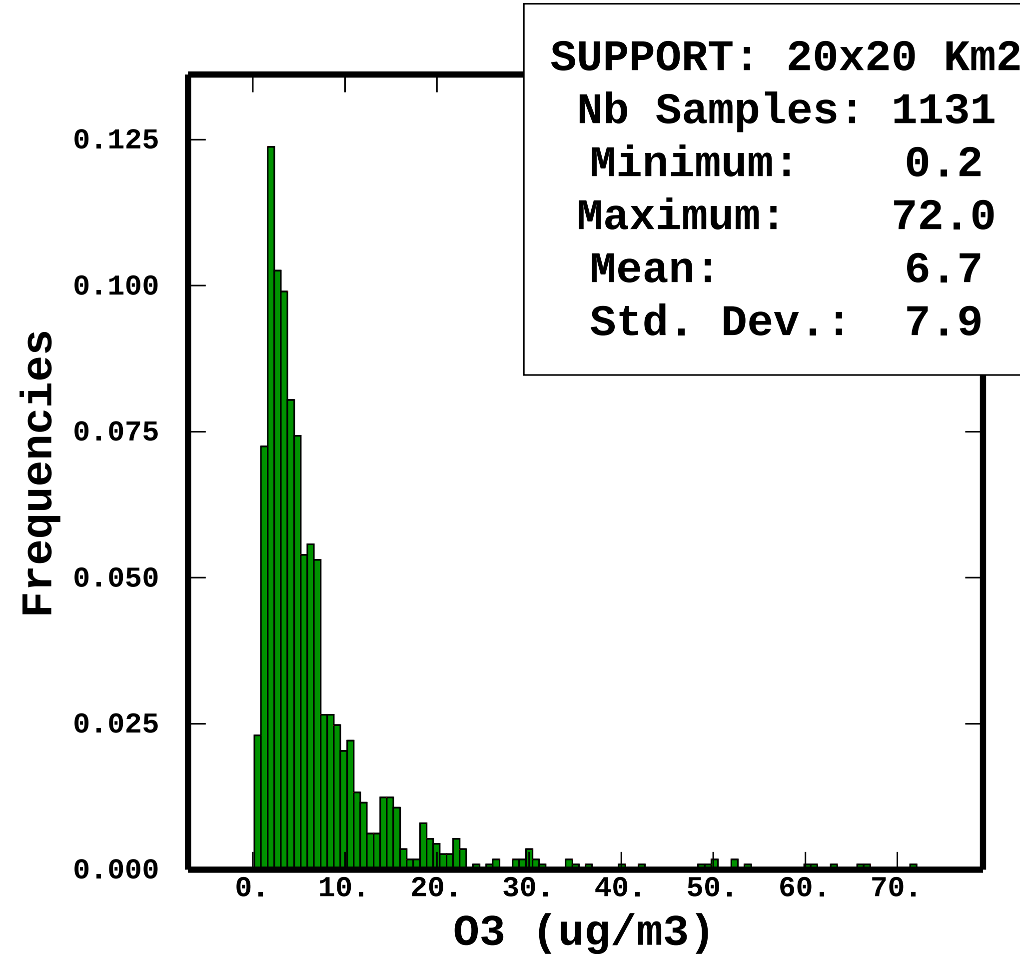


Simulation of Ozone: 20×20 Km^2 support



Simulation of Ozone

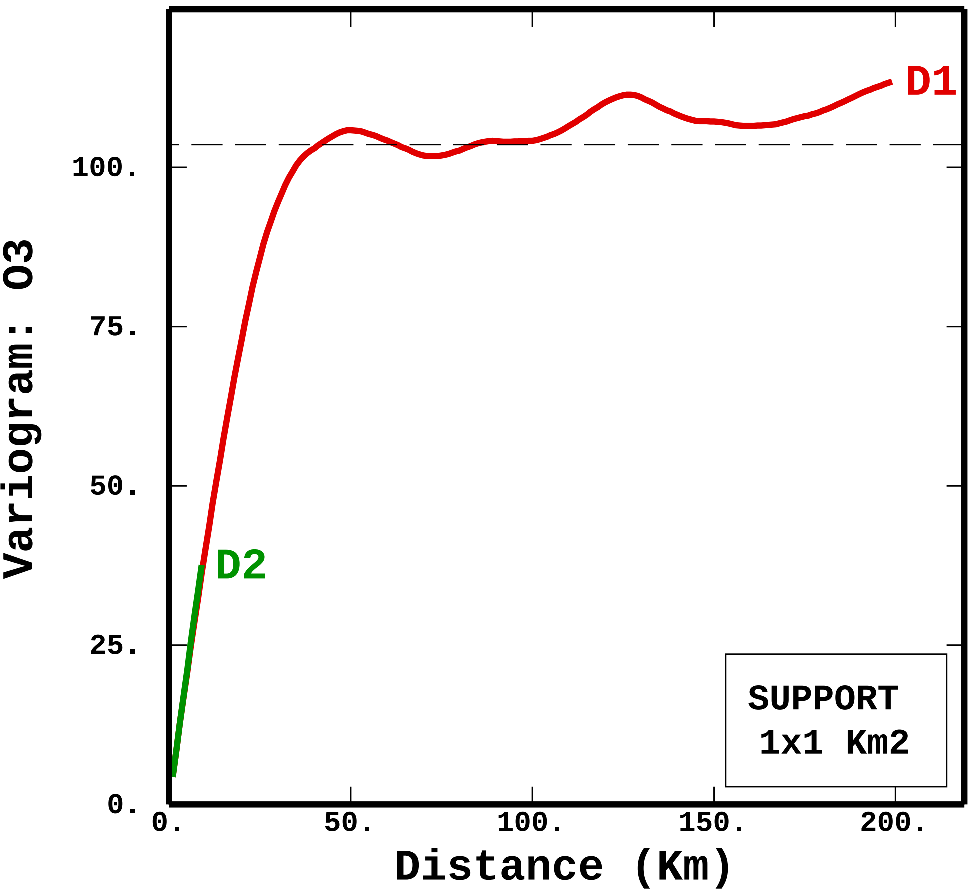


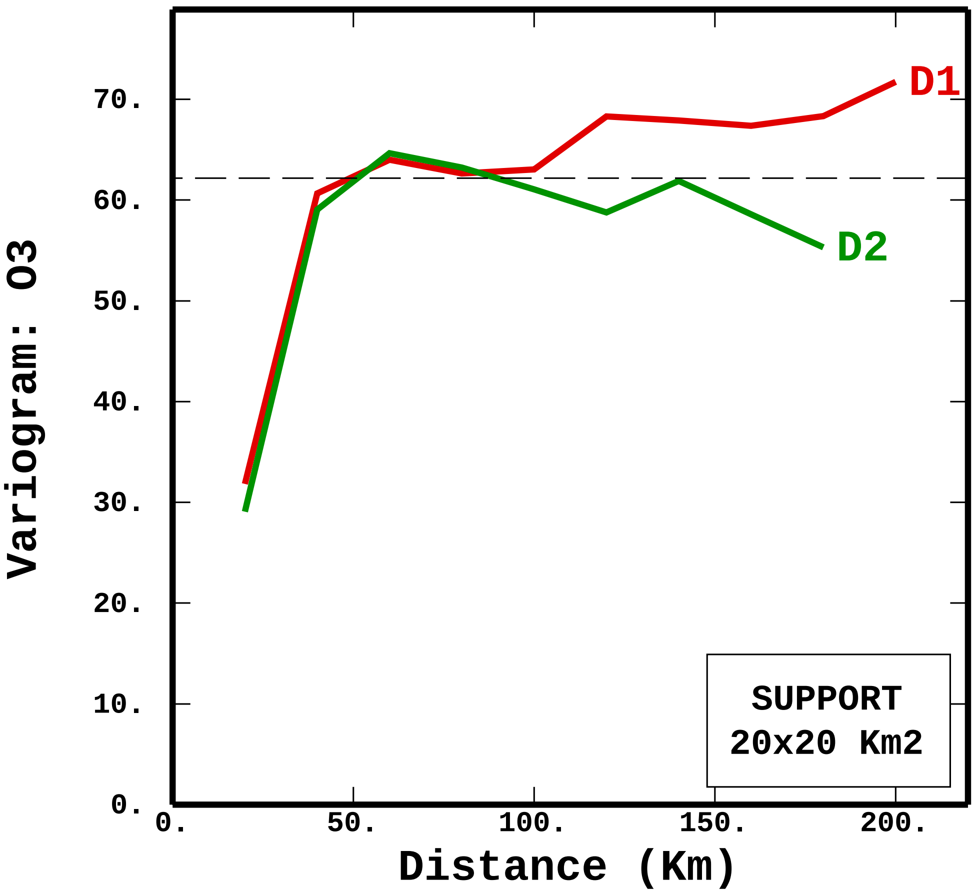


Increasing the support: the means are equal,

but the extremes and the variance are reduced

Simulation of Ozone

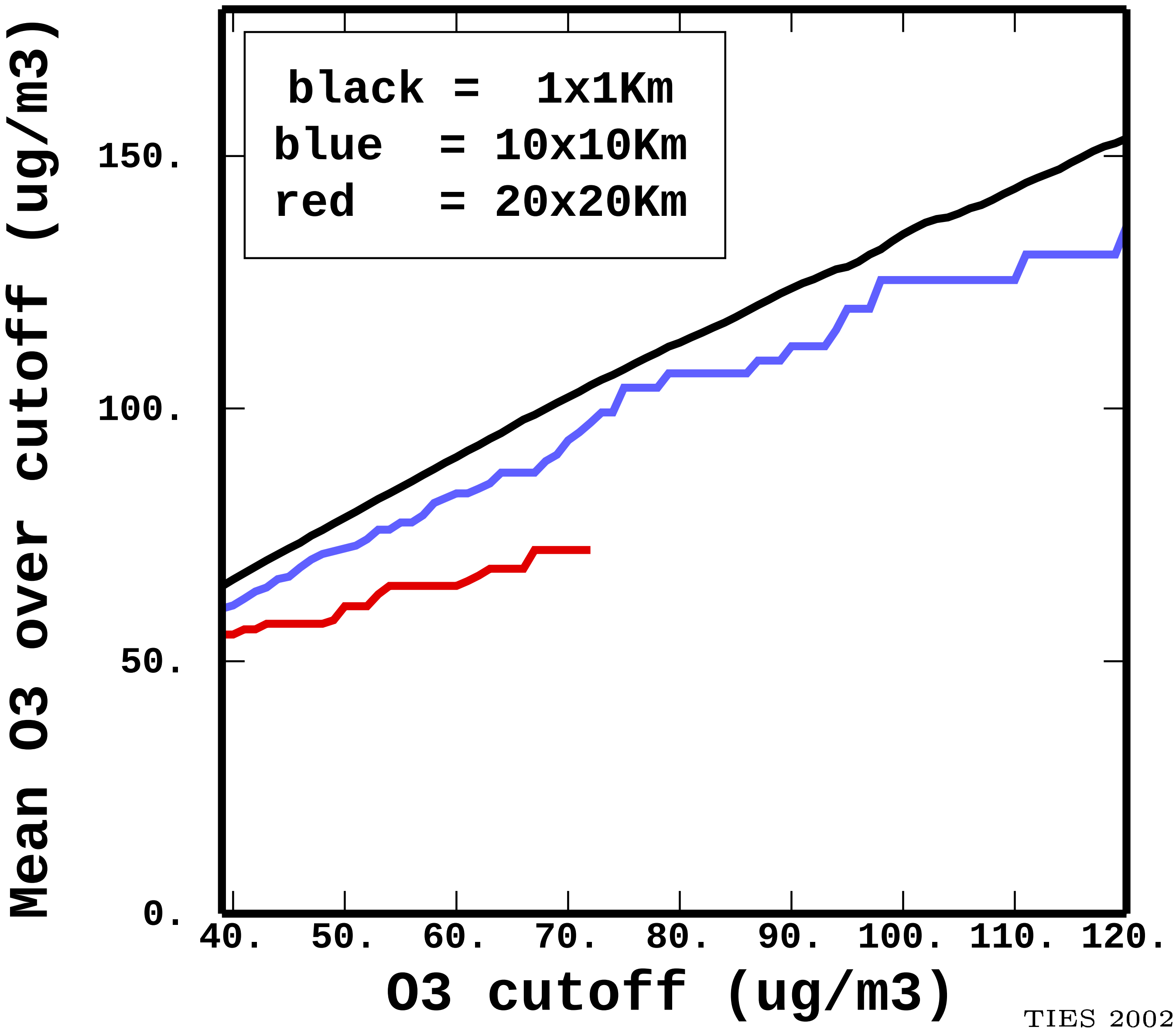




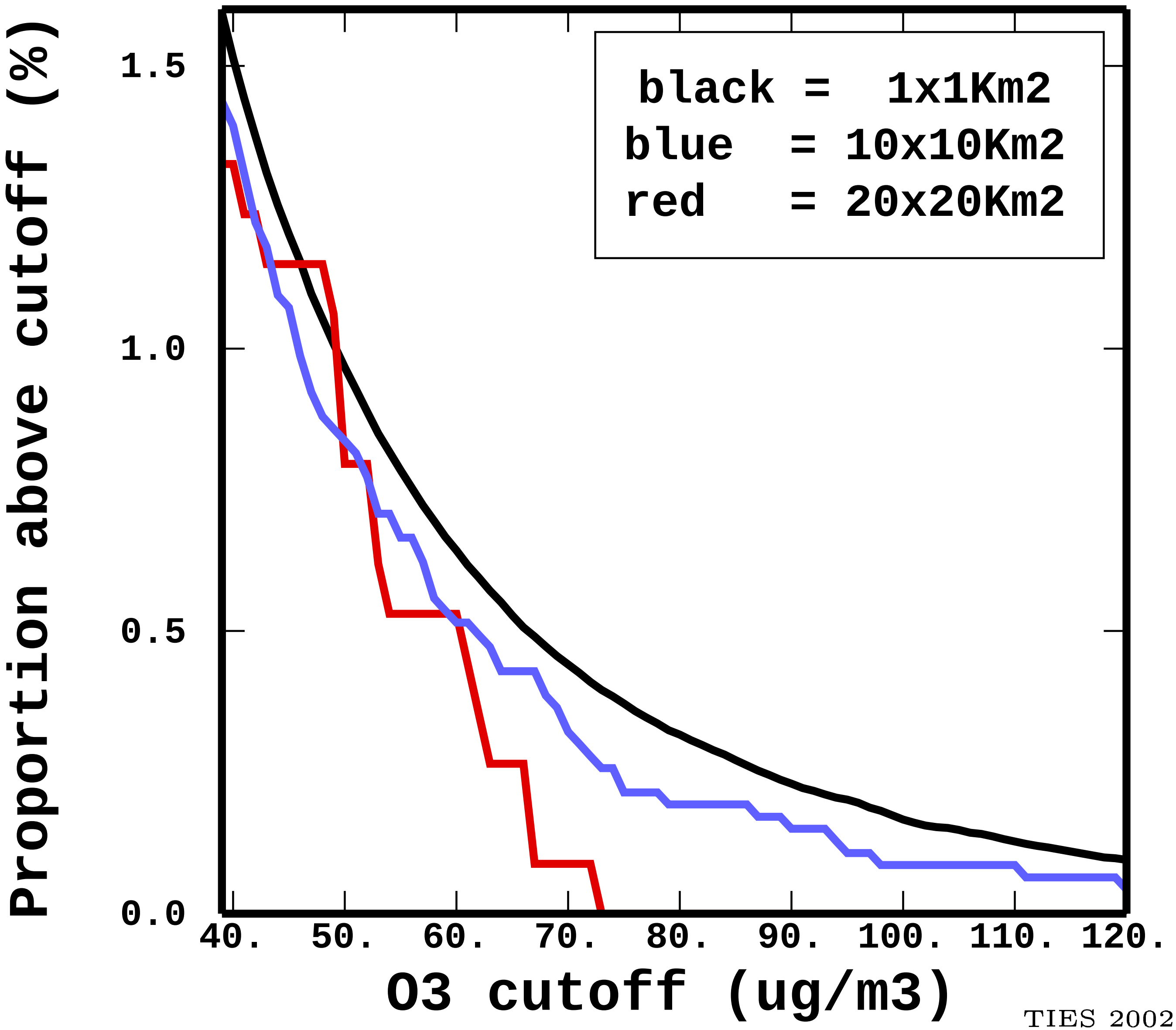
Increasing the support:

the range increases

Simulation of Ozone



Simulation of Ozone

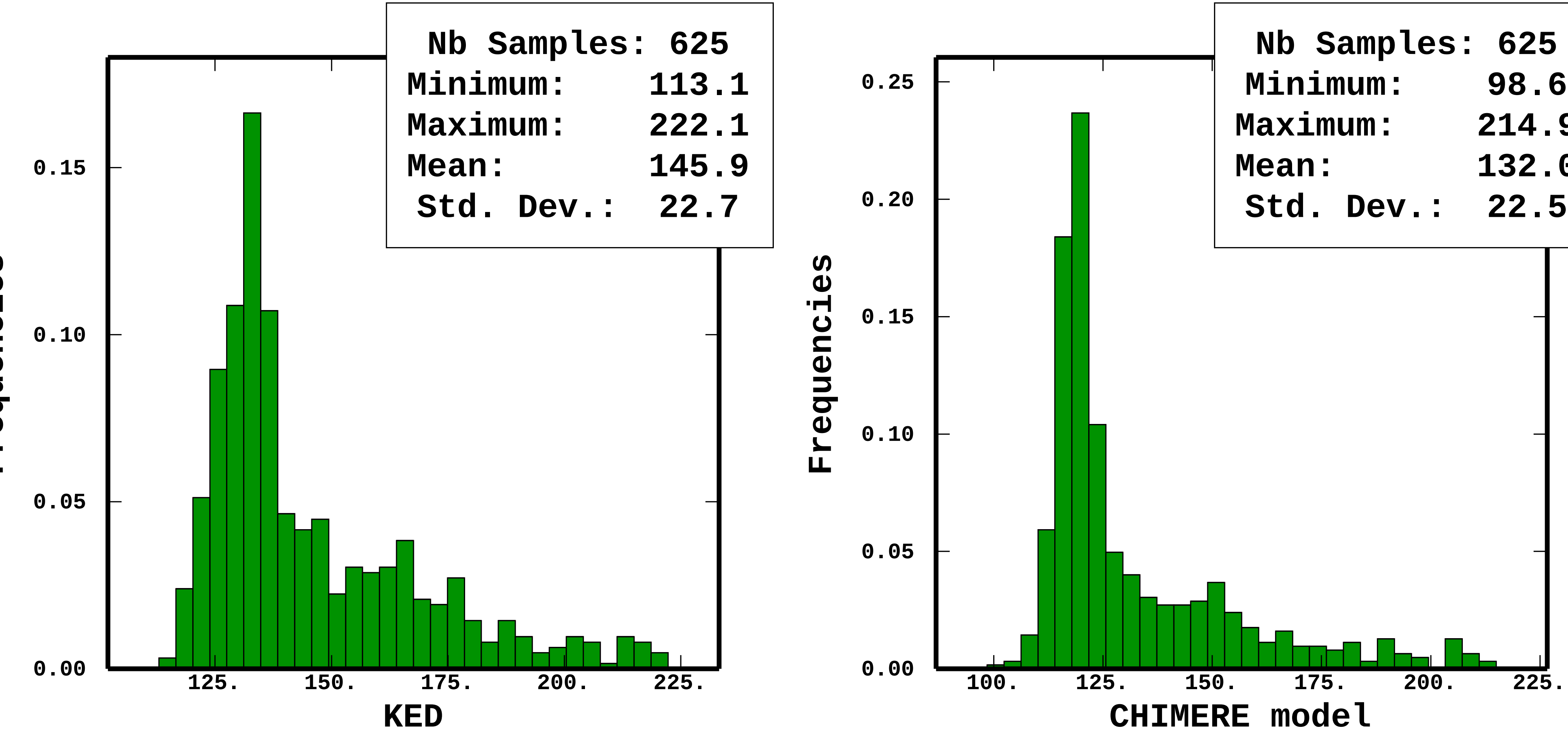


What about

the support effect

for KED and CHIMERE ?

KED vs CHIMERE (17 July '99, 15hUT)



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

Chilès J.P., Delfiner P. (1999) *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York

Hudson, G. & Wackernagel, H. (1994) Mapping temperature using kriging with external drift: theory and an example from Scotland. *International Journal of Climatology*, 14, 77–91

Rivoirard, J. (2000) Which models for collocated cokriging? Mathematical Geology, 33 (2).

Air pollution case study

Lajaunie, C. & Wackernagel, H. (2000) *Geostatistical normalization: Case Studies*. Tech. Rep. N–31/01/G, Centre de Géostatistique, Fontainebleau. Download: `http://cg.ensmp.fr/~hans/projects.html`

Ebro estuary case study

Wackernagel, H., Bertino, L., Sierra, J.P. & Gonzalez del Rio, J. (2002) Multivariate kriging for interpolating data from different sources. In: Anderson CW, Barnett V, Chatwin P & El Shaarawi AH (eds) Quantitative Methods for Current Environmental Issues, 57-75, Springer-Verlag, London.