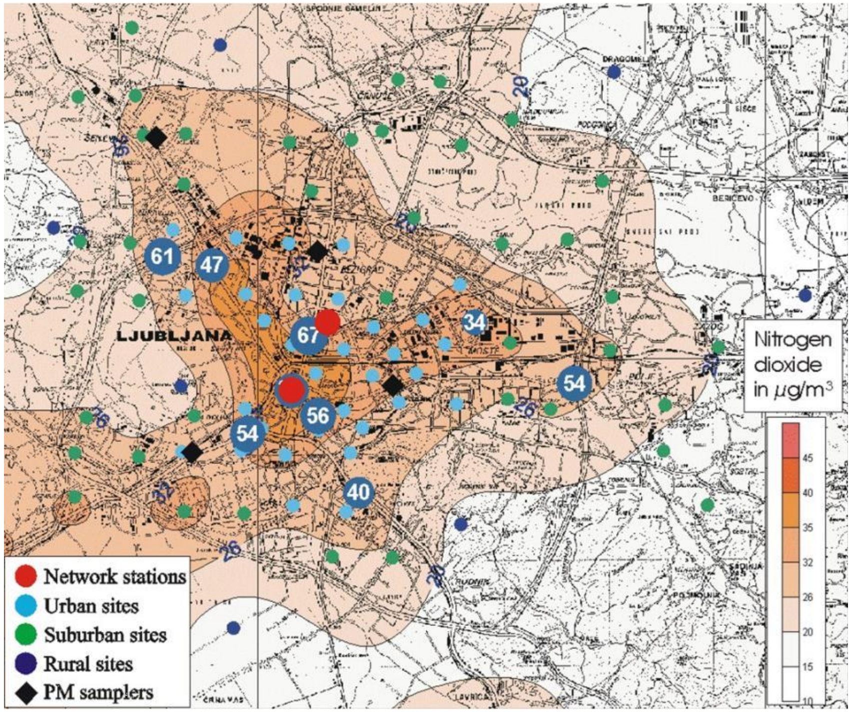
Cross-Cutting Activity on Spatial Representativeness



NO_2 annual average concentration levels in Ljubljana, from *Gerboles et al. (2007)*

Oliver Kracht

European Commission –
Joint Research Centre

I – 21026 Ispra (VA)
www.jrc.ec.europa.eu

FAIRMODE Technical Meeting

28th and 29th April 2014
Kjeller - Norway

Outline

1) Different topics and variety of objectives covered under the term spatial representativeness
2) WG 1 key questions for participants' contributions to the technical meeting
3) Own research activities in this context (JRC)
    - potential links to spatial representativeness topics in WG 1
    - caveats about porting our approaches to this different field
4) Perspectives for introducing spatial representativeness into benchmarking applications (DELTA tool)

European Commission

Spatial Representativeness in the Literature

"Representativeness is the extent to which a set of measurements taken in a space-time domain reflects the actual conditions in the same or different spacetime domain taken on a scale appropriate for a specific application."

(Nappo et al. 1982)

"[the area of representativeness] … is the area in which the concentration does not differ from the concentration measured at the station by more than a specified amount."

(Larssen et al. 1999)

"A monitoring station is representative of a location if the characteristic of the differences between concentrations over a specified time period at the station and at the location is less than a certain threshold value."

(Spangl et al. 2007)

European Commission

Possible definitions of Spatial Representativeness

The variety of definitions does also reflect the variety of objectives covered under the term of spatial representativeness:

Different definitions can be required to suit different purposes:

- Model calibration and model validation
- Detection of spatio-temporal outliers
- Design of monitoring networks
- Exposure assessment
- Area of representativeness vs. simplified mathematical definitions
- Statistical evaluations
- Regulatory purposes and legislation
- …

Key Questions to structure participants contribution

CCA Spatial Representativeness:

Q1: What kind of methodology do you use to assess the spatial representativeness of your monitoring stations?

Q2: How do you take into account this information in your model evaluation?

Q3: How do you define outliers in your monitoring network / measurement data set?

European Commission

Own research activities:

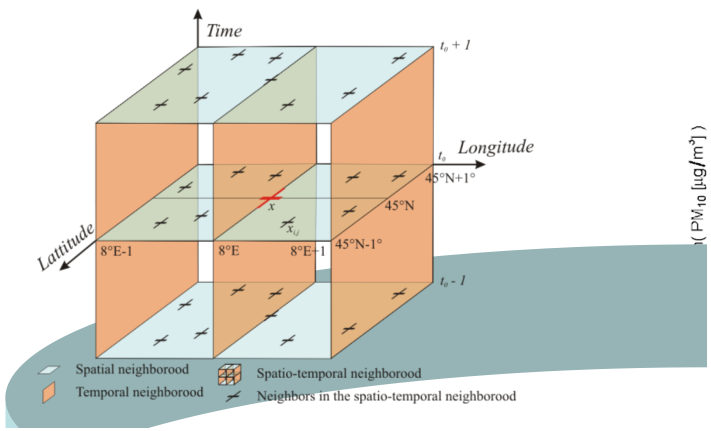
1) Automatic screening tools for the recognition of anomalies in AQ monitoring data based on attribute values and spatio-temporal relationships *("Automatic Outlier Detection")*

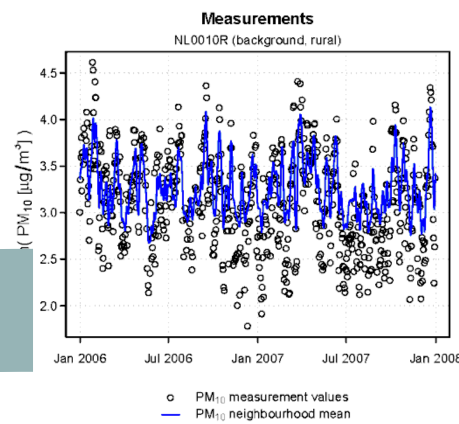
2) Uncertainty of Measurement evaluated by geostatistical tools (using estimated nugget variances)

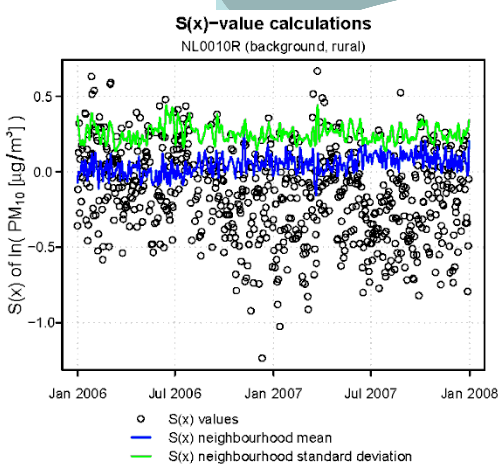
3) How can this support the consideration of spatial representativeness / spatial uncertainty in MQO and MPC ? (link to the DELTA tool?)

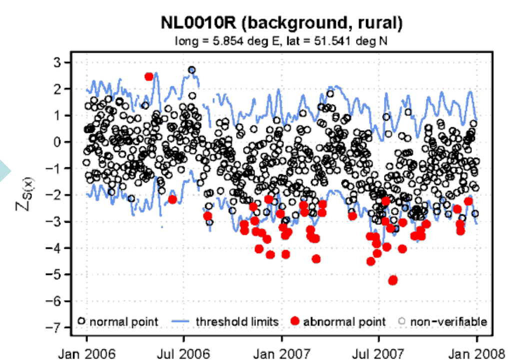
European Commission

1st method: Automatic screening tools for the recognition of anomalies in AQ monitoring data





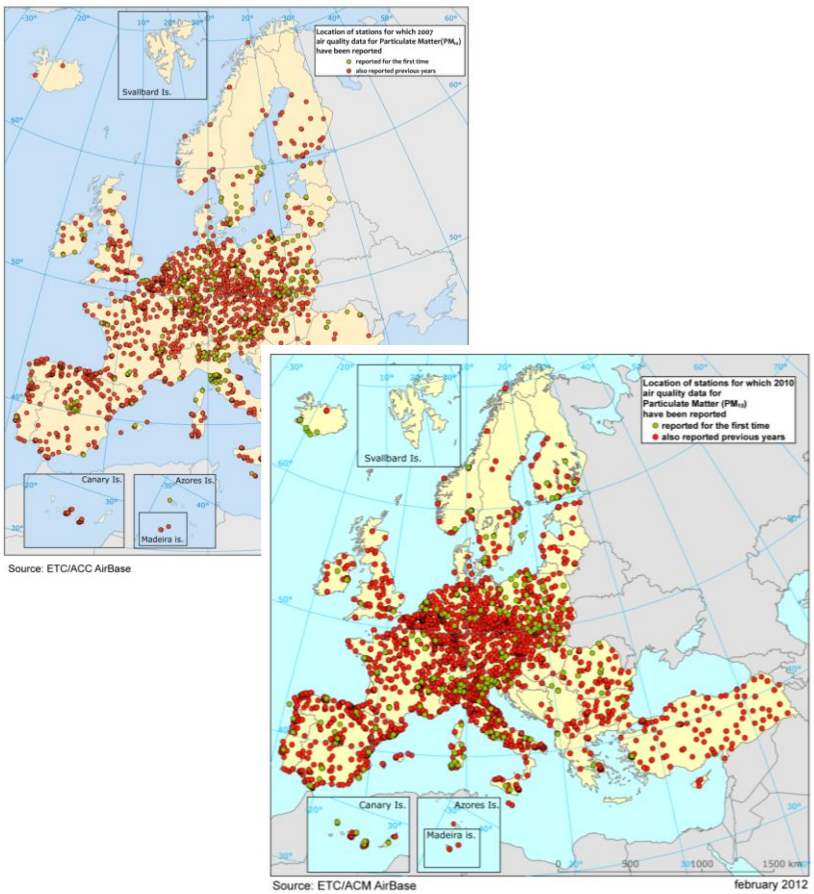




number of datapoints investigated for NL0010R: 685
identified outliers: 49 (outlier content = 7.15 %)
outlier counts station ranking = 3 within 24 stations from NL
non verifiable datapoints: 0
spatial neighbourhood: between 5 to 8 stations (between 20 to 40 records)

Joint Research Centre

Scope of this application:

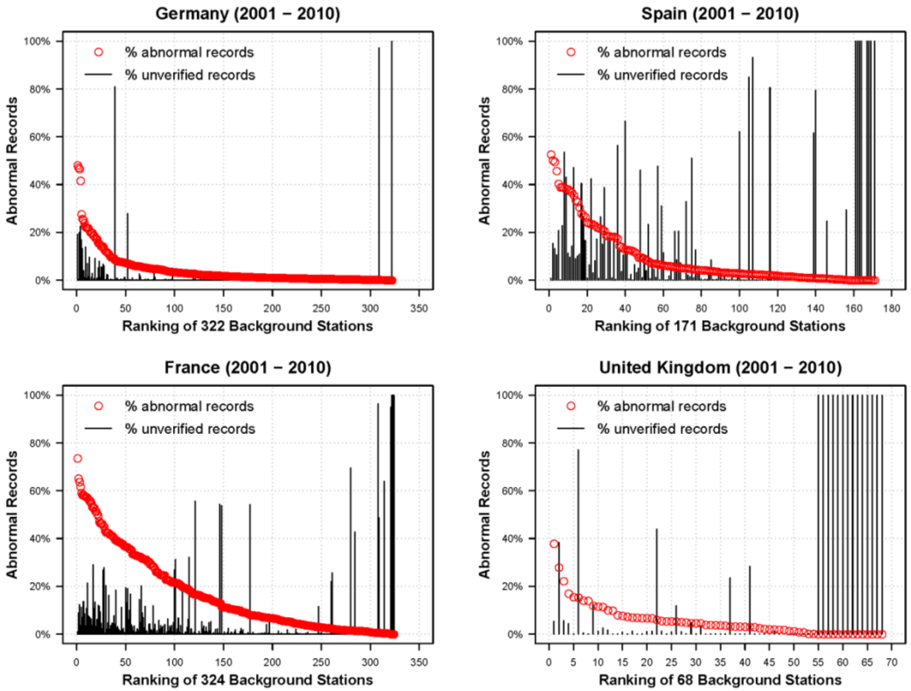


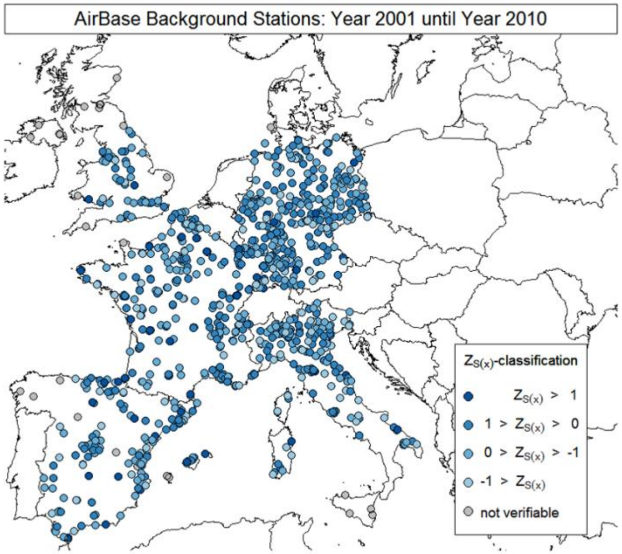
- Records with varying time-extend from AirBase versions 4 and 7
- Daily PM_{10} values
- station type "Background"
- all area types (urban, suburban and rural – to be discussed)

European Commission

Automatic screening tools for the recognition of anomalies in AQ monitoring data

- Identification of spatio-temporal anomalies
- Indicators for evaluating the consistency of station classifications





European Commission

"Smooth Spatial Attribute Method"

- Proposed for traffic sensors by Lu et al. 2003 & Shekhar et al. 2003

  Lu, CH.-T., D. Chen & Y. Kou, 2003: Detecting Spatial Outliers with Multiple Attributes. ICTAI'03, IEEE 2003.

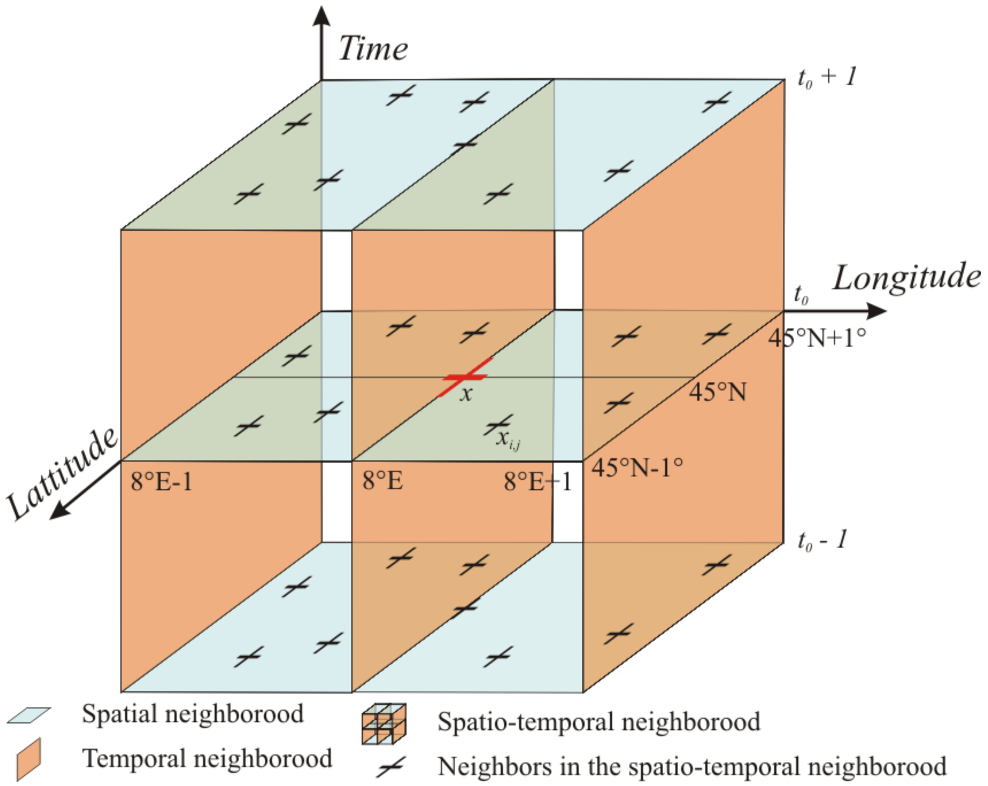
  Shekhar, S., CH.-T. Lu & P. Zhang, 2003: A Unified Approach to Detecting Spatial Outliers. GeoInformatica, 7(2), 139-166.

- 1st quantify how the measurement value of a station deviates from the corresponding values observed within its spatio-temporal neighbourhood (the 'Sx value')

- 2nd compare this Sx-deviation to the corresponding Sx-deviations observed for the station's neighbours

European Commission

Definition of Neigbourhood in 3 Dimensions:



- spatial domain limited to +/- 1 spherical degrees
- temporal domain limited to +/- 2 days
- temporal domain is automatically expanded if initial neighbourhood is too little

European Commission

"Smooth Spatial Attribute Method"

- Proposed for traffic sensors by Lu et al. 2003 & Shekhar et al. 2003)

- Definition of Sx-values (for each individual neighbourhood)

$$Sx = f \quad x - \overline{f \quad x_n}$$

$$\Rightarrow Sx = x - \overline{x_n}$$

- Z-transformation of Sx

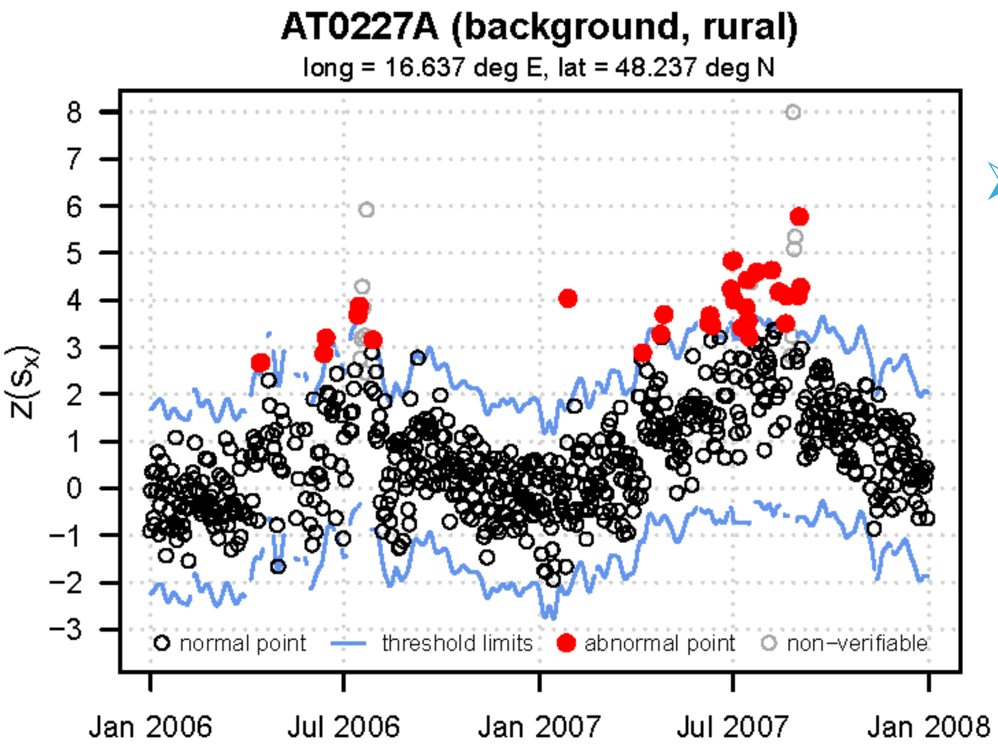
$$z = \frac{Sx - \overline{Sx_n}}{s_{Sx_n}}$$

- Test statistics for spatio-temporal outliers

$$\theta - 1.96 < z_i < \theta + 1.96$$

European Commission

Type of results:



$$\theta - 1.96 < z_i < \theta + 1.96$$

- final step:

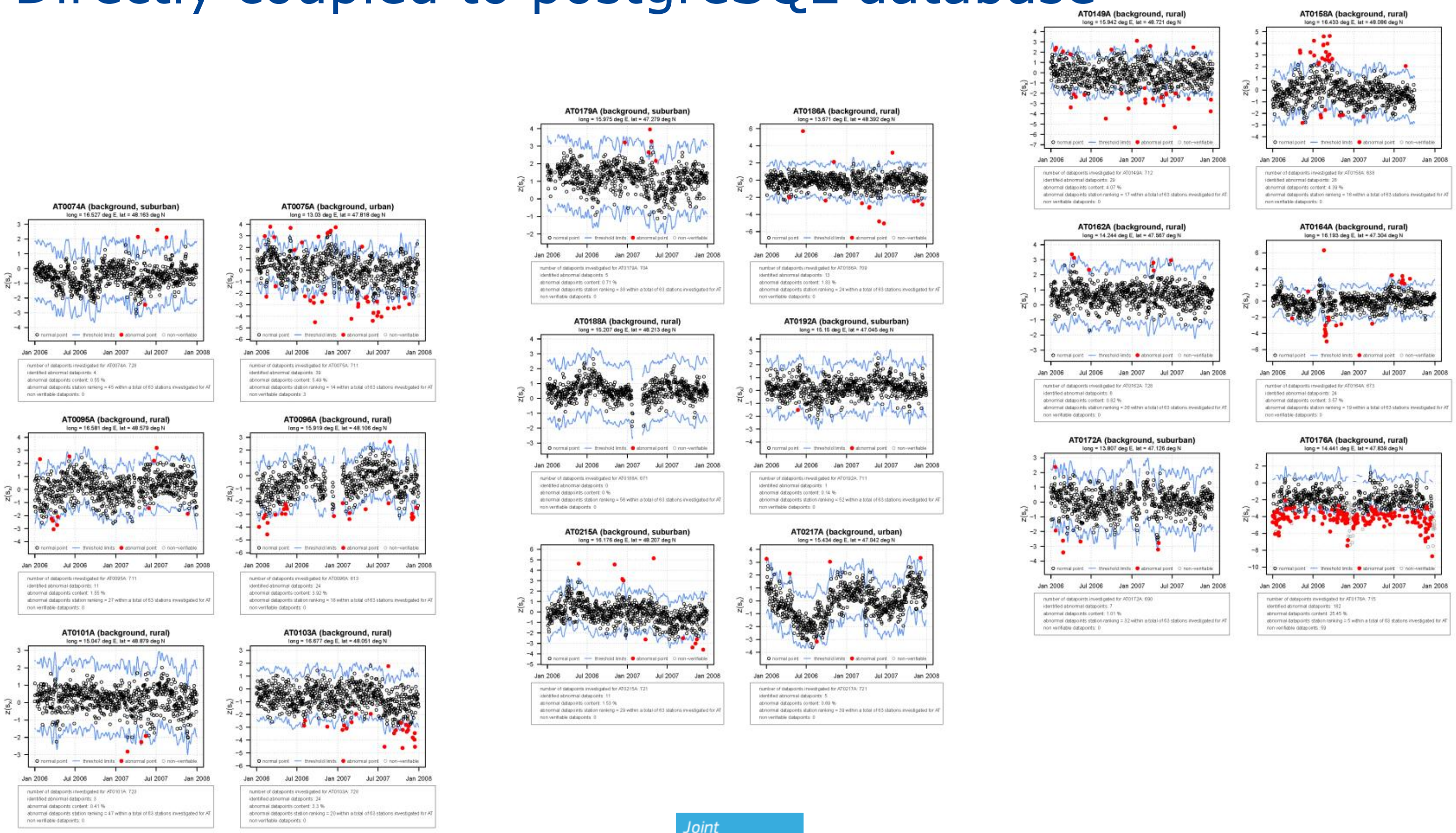
  Test statistics for abnormal values searches for z_i values exceeding the upper/lower limits chosen as a reference.

  (e.g., θ +/- a predefined threshold of 1.96)

European Commission

Automated Data Processing

- All codes prototyped in the R environment
- Directly coupled to postgreSQL database



European Commission

Input Data Requirements

- Sufficiently dense spatial and temporal data coverage.
- Time Series (e.g., daily values) of a spatially distributed random field.
- Limitations should be anticipated to be imposed from the network design (station density and spatial organisation of the monitoring network).

European Commission

Important properties of the method

- Able to detect local outliers which are characterized by a sharp spatial or temporal non-stationarity of a pollutant concentration.
- Values fulfilling the condition of weak-sense stationarity (even less strict, the condition of a smoothly varying spatio-temporal covariance structure) will not be miss-classified as outliers.
- This makes it a useful technique for the analysis of background type monitoring stations.
- The inherent spatial non-stationarity of other types of sites (traffic or industrial) makes it less suitable for such type of applications.

European Commission

Important properties of the method

- Results about abnormal data-points content are dependent on the parameter values chosen in the screening method.
- The effective confidence level for real world data depends on the spatiotemporal correlation of the data field.
- An absolute definition for outlying stations is not feasible, but depends on the intended objectives for using the method.
- Limitations can originate from the network design (station density and spatial organisation of the monitoring network).

Possible implementation into the DELTA tool?

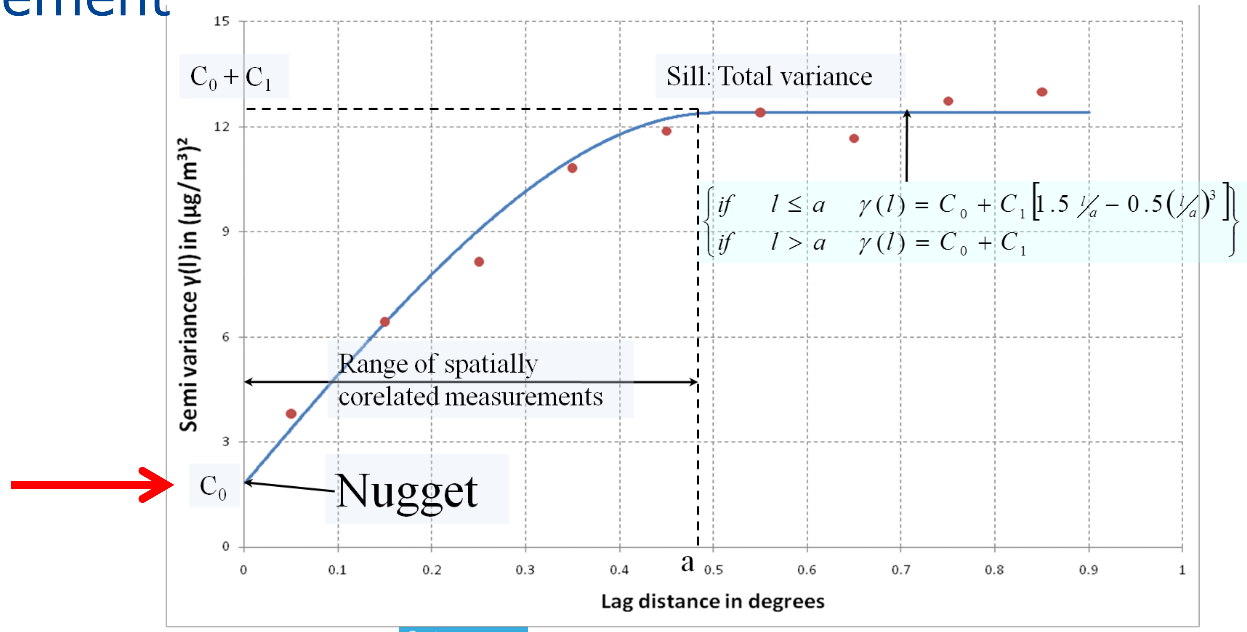
1) External pre-processor for the observation datasets.

2) In a first step this might be a collection of R-Scripts and instructions for use made available on the DELTA tool homepage.

European Commission

2nd method: Uncertainty of measurement evaluated from estimated nugget variance

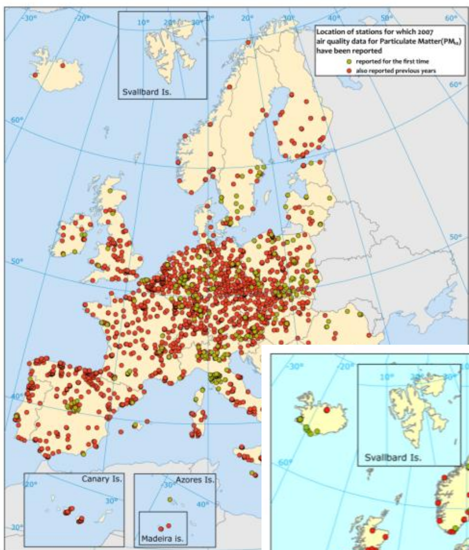
- Comparison with the data quality objectives
- Identify trends over time in the nugget variance to investigate improvement (or worsening) of the uncertainty of measurement



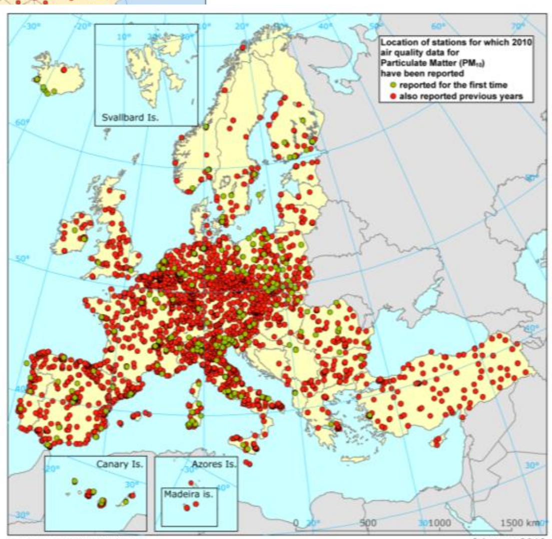
Centre

European Commission

Scope of this application:



Source: ETC/ACC AirBase

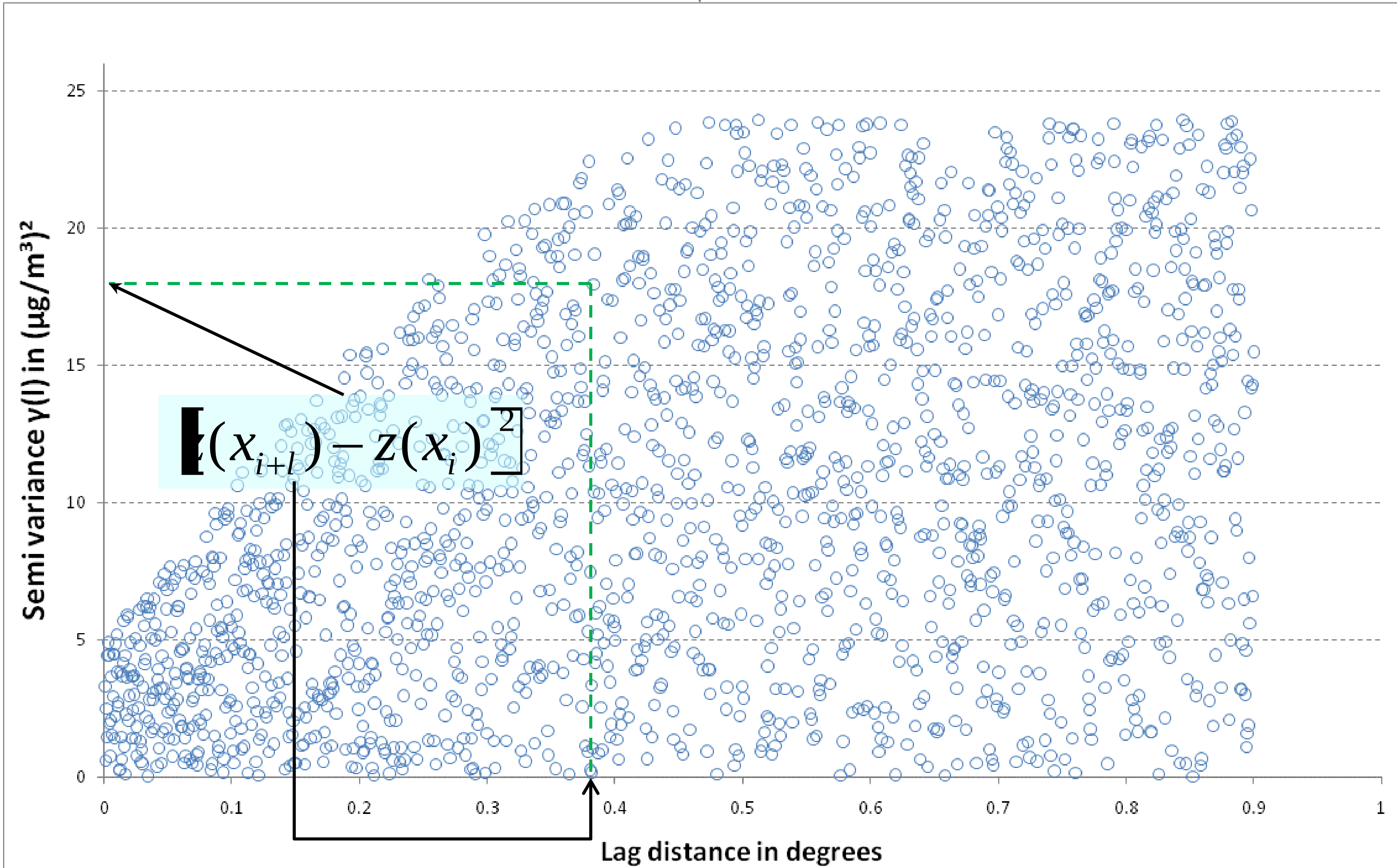


Source: ETC/ACM AirBase february 2012

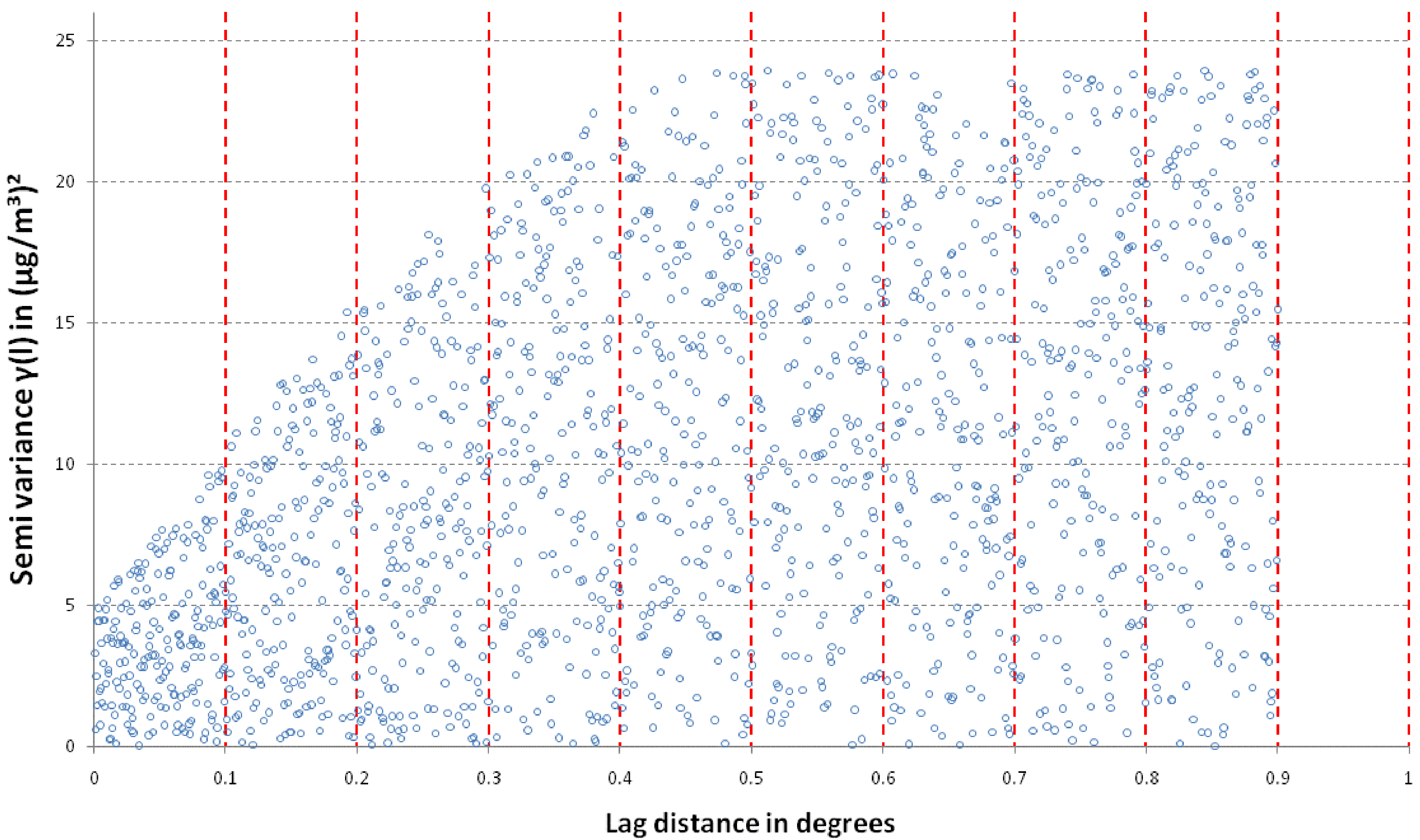
- Country wide records with varying time-extend from AirBase versions 4 and 7
- Daily PM_{10} values
- All station types
- All area types (urban, suburban and rural)

Joint Research Centre

3

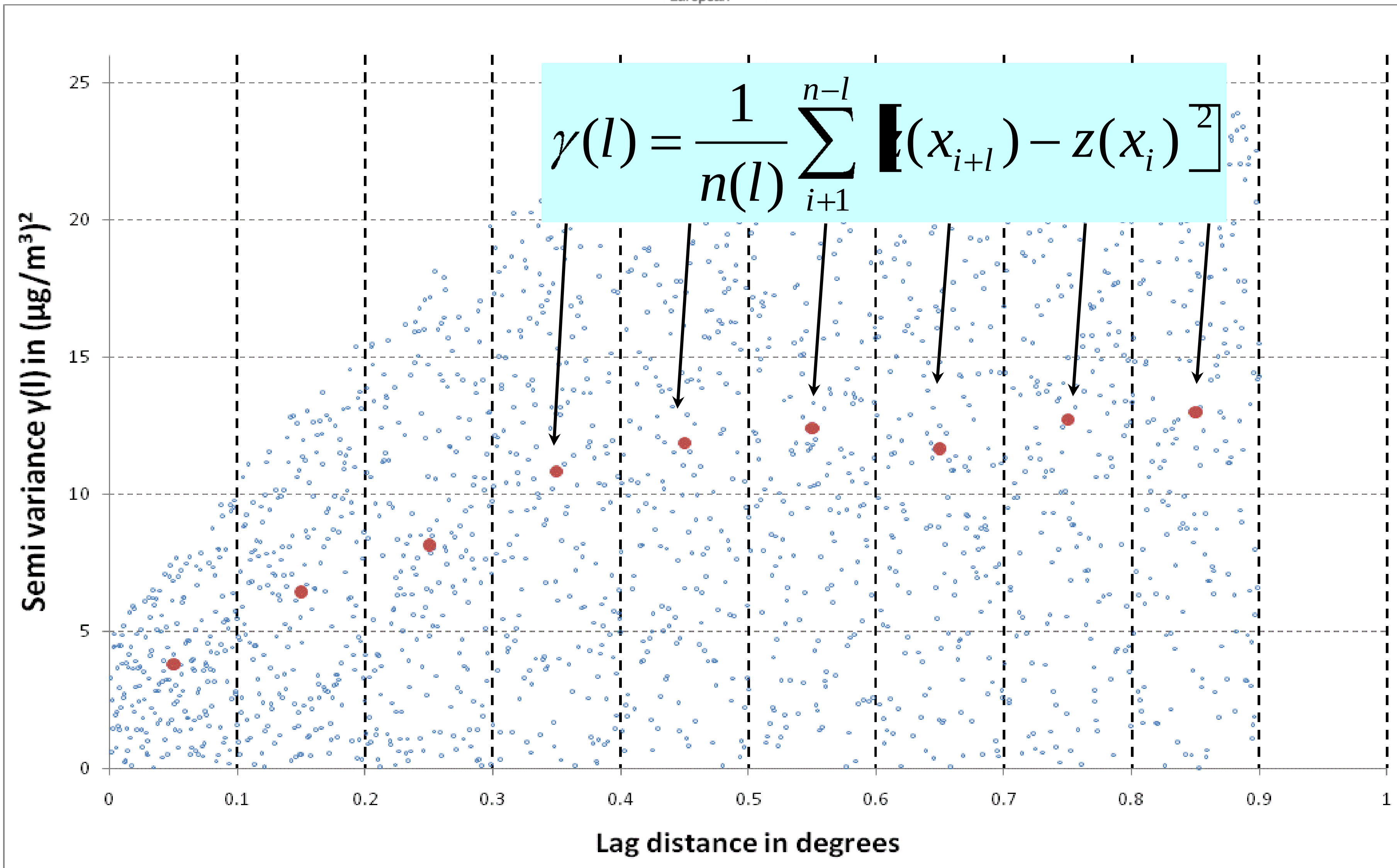


source: explanation of variography techniques, from M. Gerboles (2007): AQUILA Workshop presentation



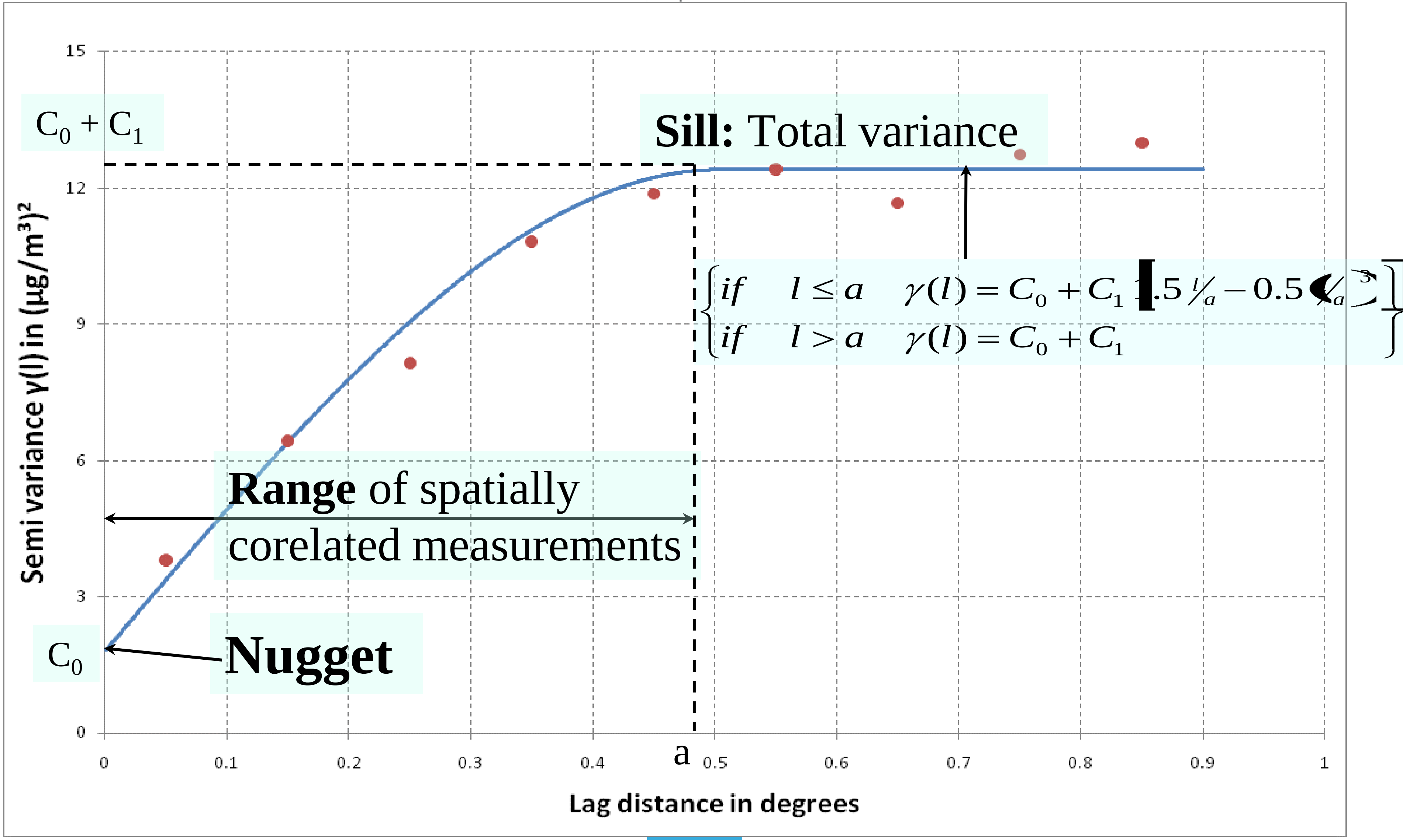
source: explanation of variography techniques, from M. Gerboles (2007): AQUILA Workshop presentation

European



Joint

source: explanation of variography techniques, from M. Gerboles (2007): AQUILA Workshop presentation



source: explanation of variography techniques, from M. Gerboles (2007): AQUILA Workshop presentation

European Commission

Primarily interest in our own research:

The nugget variance is reflecting fluctuations of the measurements at very short distance (towards 0).

$$s^2_{nugget} = s^2_{meas} + s^2_{sc}$$

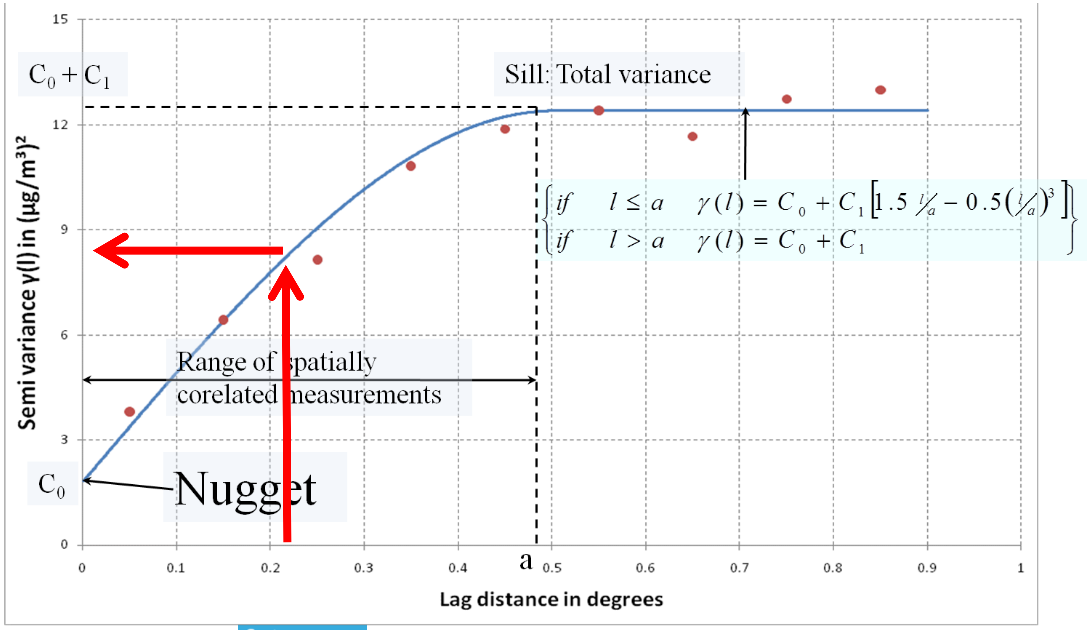
uncertainty of measurement
variance associated with the sampling and analytical variability

micro-scale variance
variability that occurs at distances lower than the shortest sampling distance (continuity).

European Commission

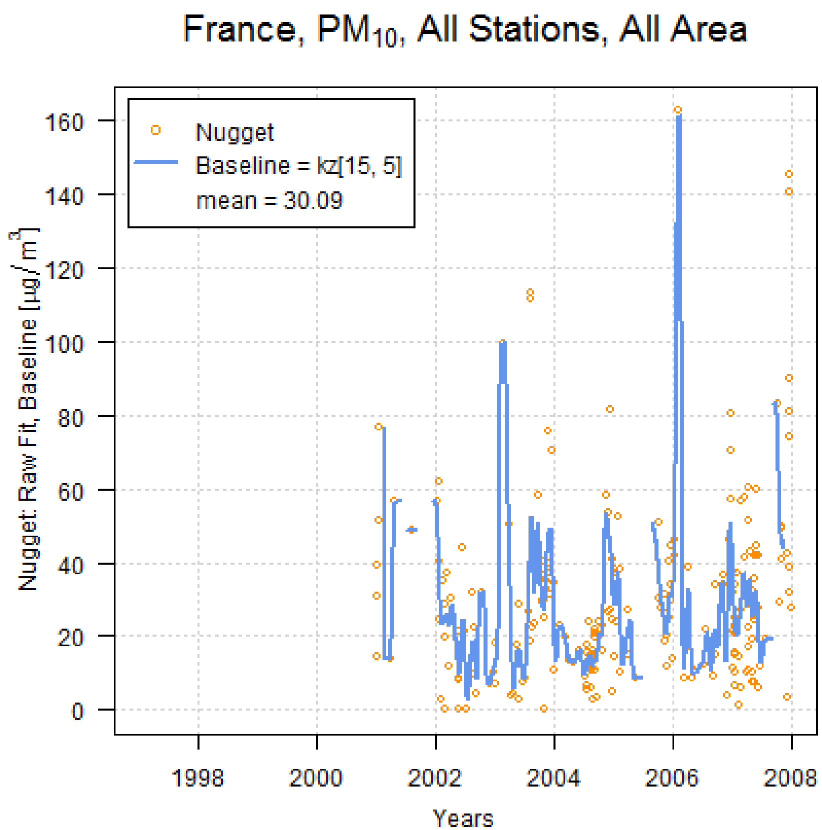
How can this knowledge contribute to introduce spatial representativeness into the model validation step?

- Variogram based specification of spatial uncertainty.
- Spatial Uncertainty as a weighting factor of data points for model validation.

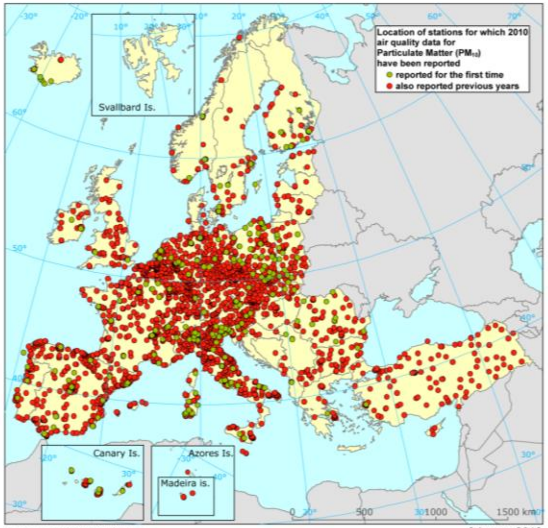


European Commission

A first trial: Variogram parameters estimated from AirBase daily PM_{10} data



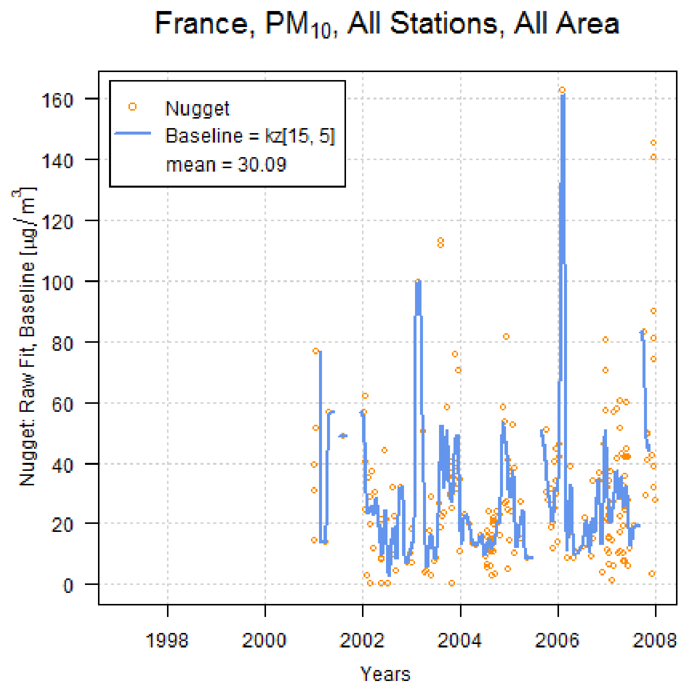
dataset contained 2555 initial cases, 217 accepted cases



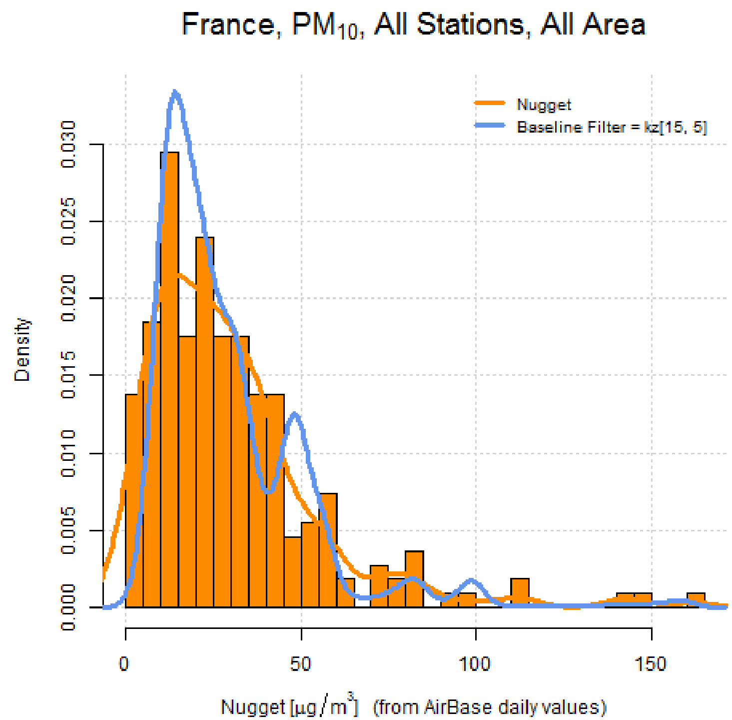
Source: ETC/ACM AirBase february 2012

European Commission

A first trial: Variogram parameters estimated from AirBase daily PM_{10} data

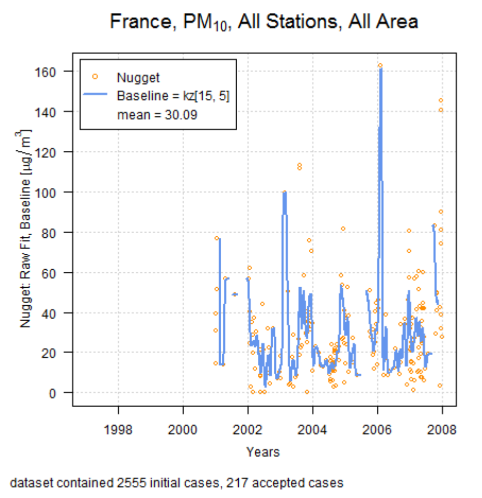


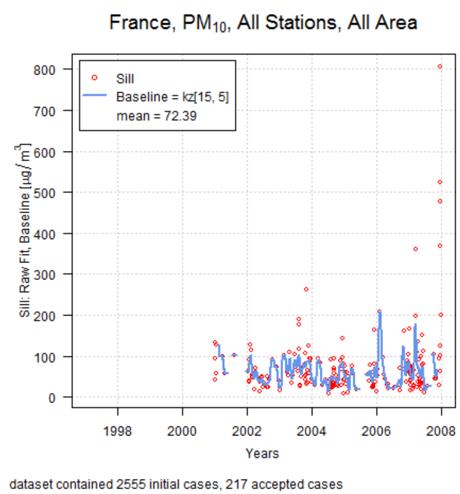
dataset contained 2555 initial cases, 217 accepted cases

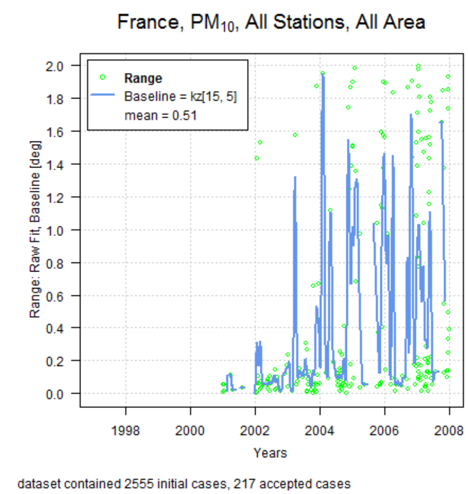


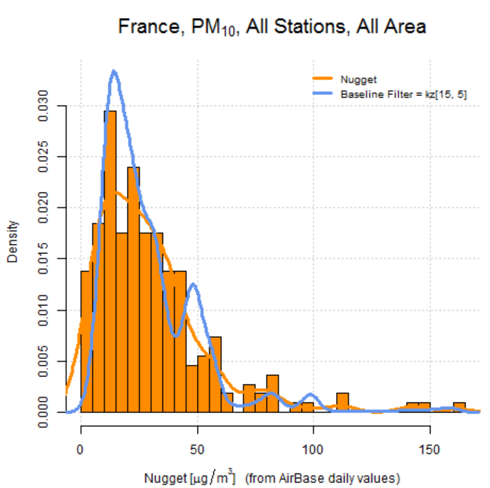
European Commission

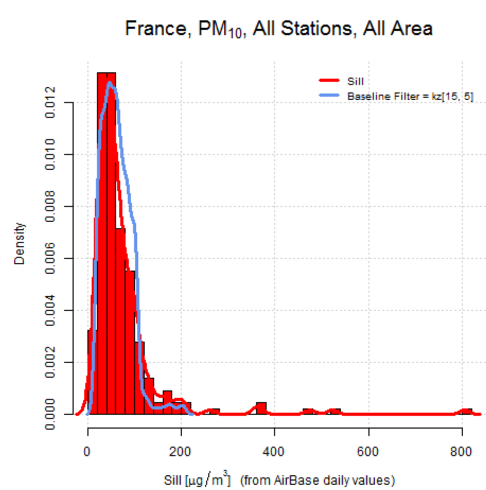
Variogram parameters estimated from daily PM_{10} data

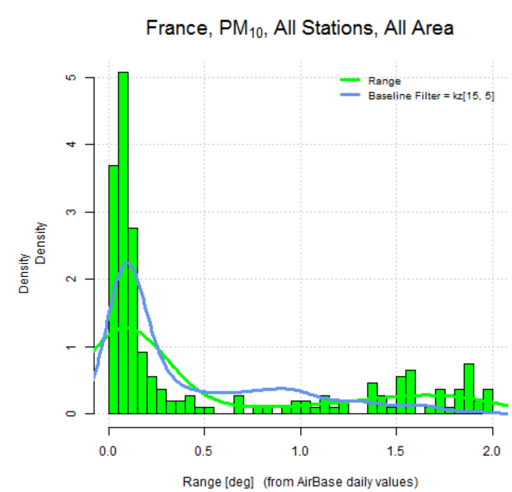






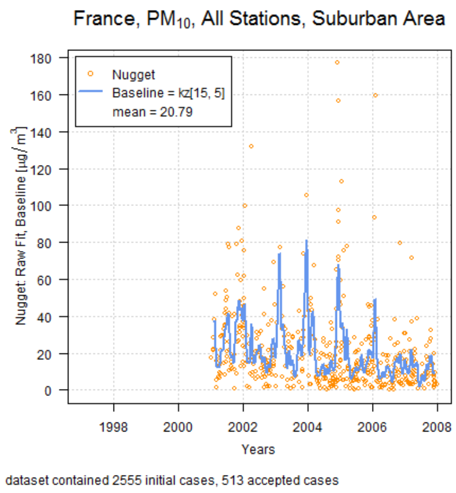


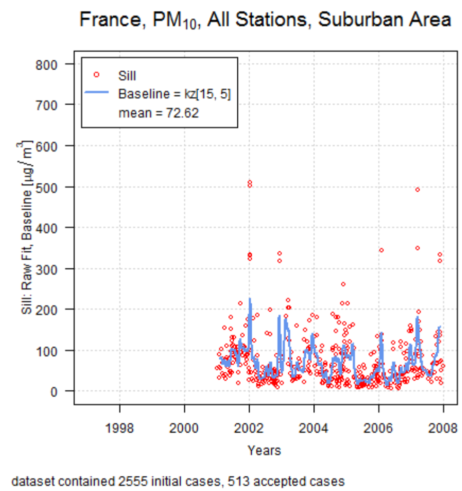


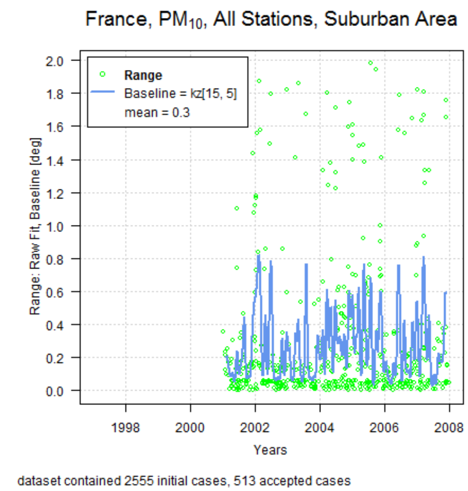


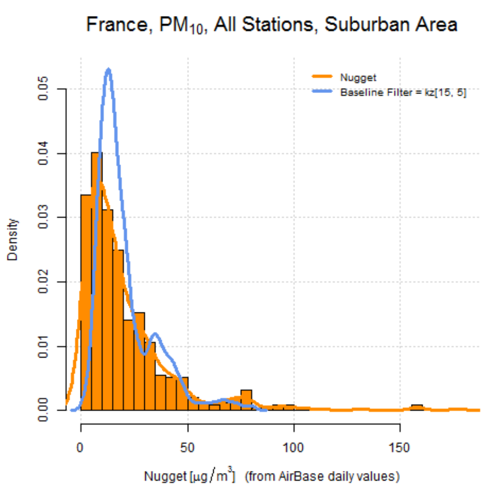
European Commission

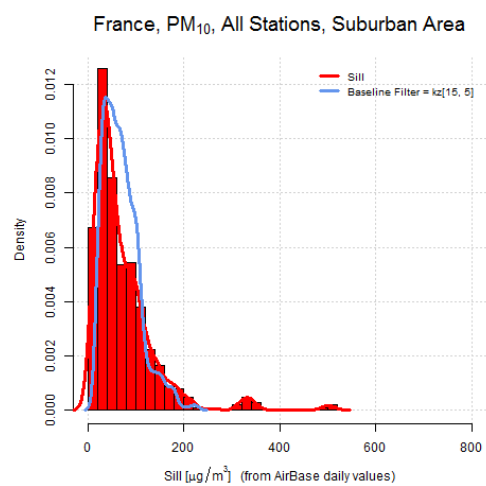
Variogram parameters estimated from daily PM_{10} data

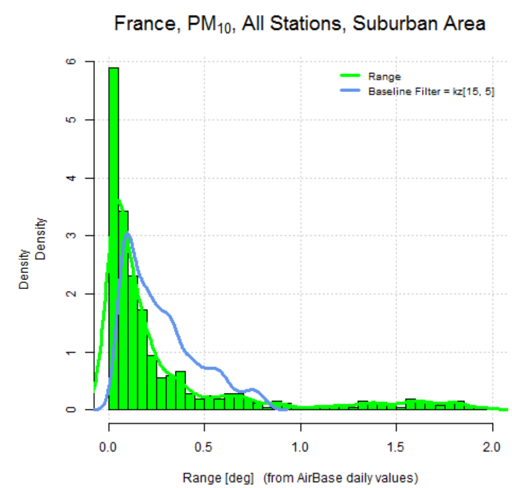








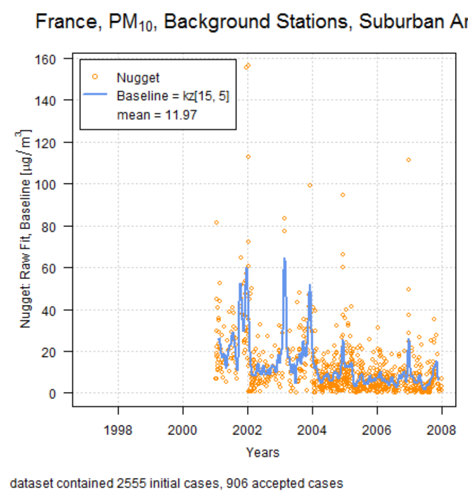


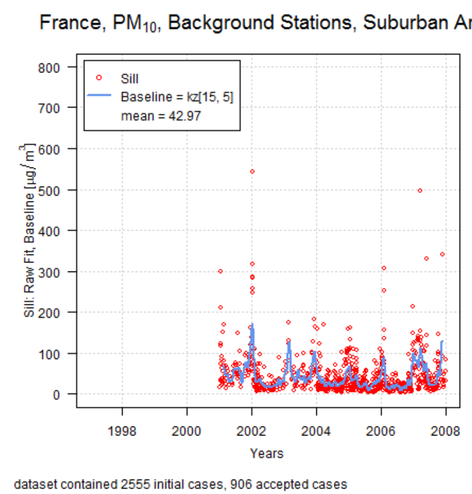


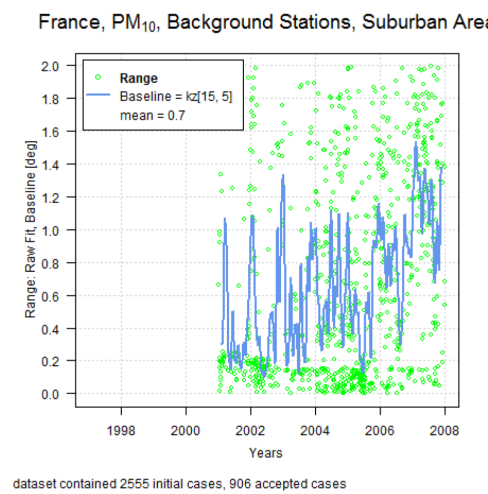
Research Centre

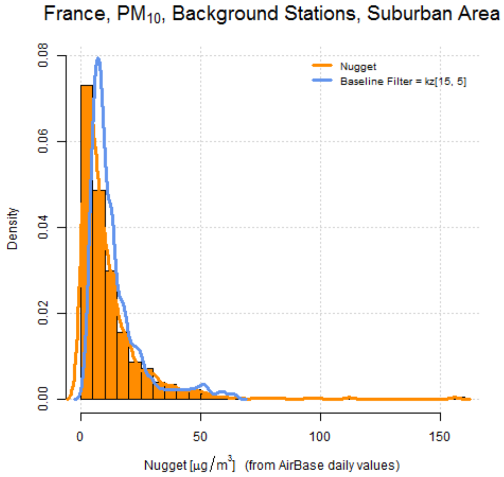
European Commission

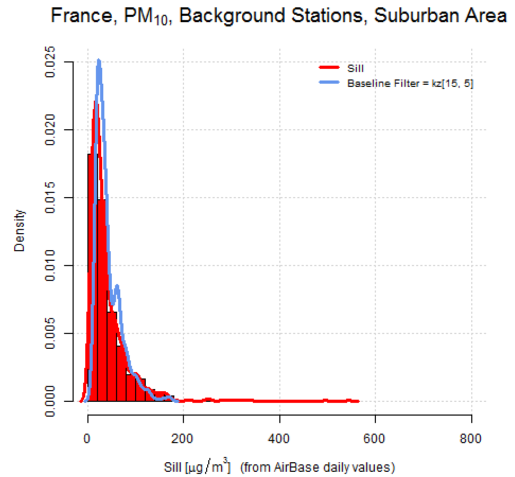
Variogram parameters estimated from daily PM_{10} data

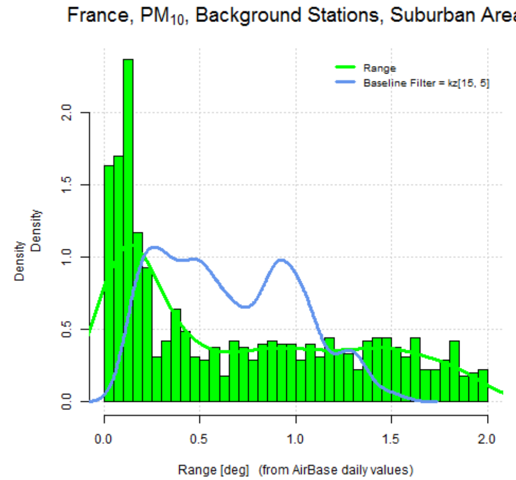












Research Centre

This is ongoing work …

next steps:

- Improve the automated variogram fits to obtain better stability of the algorithms and quality of fit control.
- Investigate the temporal behaviour of estimated variogram parameters (seasonality, trends over time, etc.).
- In addition to AirBase, the use of information from more densely spaced datasets might be needed.
- However, in the unfavourable case it can be that variogram parameter uncertainties cannot be narrowed down to reasonable intervals for practical applications.

European Commission

Possible consideration of spatial uncertainty in the MQO and in the MPC ?

- Variogram based description of spatial uncertainty
- Might be done in similar manner to the implementation of the measurement uncertainty?

$$MQO = \frac{1}{2}\frac{RMSE}{RMS_U} = \frac{1}{2}\frac{\sqrt{\sum_{i=1}^{N}(m_i - x_i)^2}}{\sqrt{\sum_{i=1}^{N}U^2(x_i)}} \leq 1$$

(Thunis et. al, 2013)

European Commission

Possible consideration of spatial uncertainty in the MQO and in the MPC ?

- Variogram based description of spatial uncertainty
- Analogy to measurement uncertainty?

$$MQO = \frac{1}{2}\frac{RMSE}{RMS_U} = \frac{1}{2}\frac{\sqrt{\sum_{i=1}^{N}(m_i - x_i)^2}}{\sqrt{\sum_{i=1}^{N}U^2(x_i)}} \leq 1$$

(Thunis et. al, 2013)

- Caveat: a distance based uncertainty measure introduces unfavourable dependencies of MQO from model configuration (grid spacing)
- Caveat 2: uncertainties in variogram parameter estimates can be large (note the different objective of our original approach)

Possible consideration of spatial uncertainty in the MQO and in the MPC ?

- Variogram based description of spatial uncertainty
- Analogy to measurement uncertainty?
- Continuous description of spatial uncertainty vs. definition of area of representativeness (include / exclude approach)?

European Commission

Implementations / DELTA tool pre-processor ?

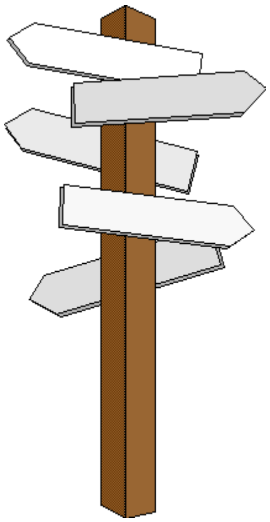
- Variogram based description of spatial uncertainty in the MQO.
- Default sets of variogram parameters might be provided within a DELTA tool extension or DELTA tool pre-processor?
- This could serve a common means for the characterisation of spatial uncertainty influences on model validation.
- It’s clearly a simplified approach. It not aims to replace or conflict with more detailed evaluations on spatial uncertainty performed by modellers in individual applications outside the DELTA tool.

4

European Commission

Aims of the Cross-Cutting Activity Session

Objectives for the discussions:



- Identify common needs and objectives on introducing spatial representativeness into model validation.
- Identify the interests in collaborations in this direction.