Evaluation of a Monte Carlo-based validation technique for data assimilated air quality assessments

Bino Maiheu[(1)], Laure Malherbe[(2)], Ana I. Miranda[(3)], Alexandra Monteiro[(3)], Claudio Carnevale[(4)] and **Stijn Janssen[(1)]**

[1]Flemish Institute for Technological Research (VITO), Belgium
[2]INERIS, France
[3]CESAM, Department of Environment and Planning, University of Aveiro, Portugal
[4]Department of Mechanical Engineering, University of Brescia, Italy

Outline

- Validation of data-assimilated models
- Proposed Monte Carlo methodology
- 3 independent evaluations (INERIS, UAVR, VITO)
- Conclusions

© 2014, VITO NV

Validation of data assimilated models

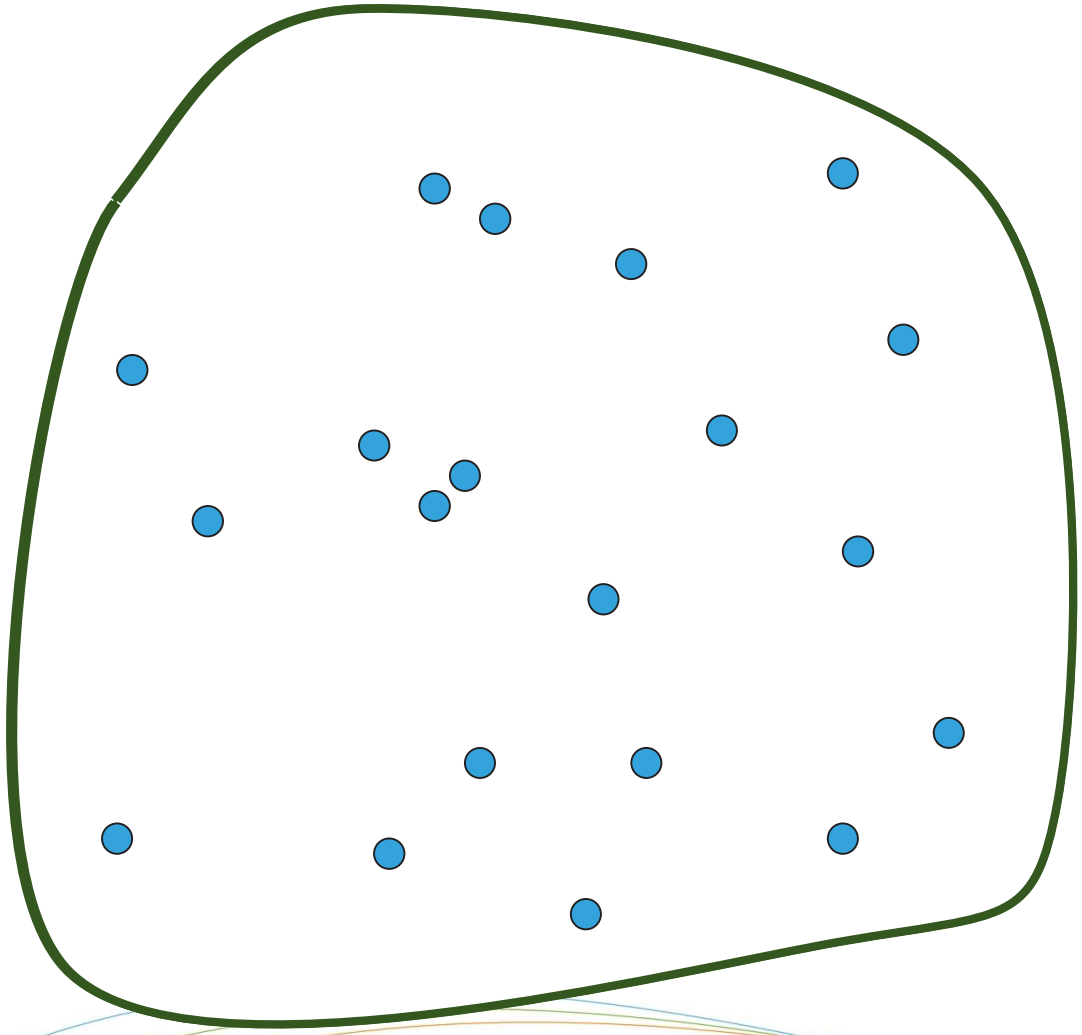
- AQD suggests integrated use of modelling & measurements for assessment
- Wide range of techniques : online vs. offline, kriging, optimal interpolation, 3DVar, 4DVar, etc..
- Validation usually :
  - Leave one out & compare : criticized as not independent enough
  - Leave n out & compare : how to perform selection of subset ?
- Idea ! use **Monte Carlo** technique to leave n out

© 2014, VITO NV

Proposed Monte Carlo methodology

- Claudio Carnevale (Uni. Brescia):

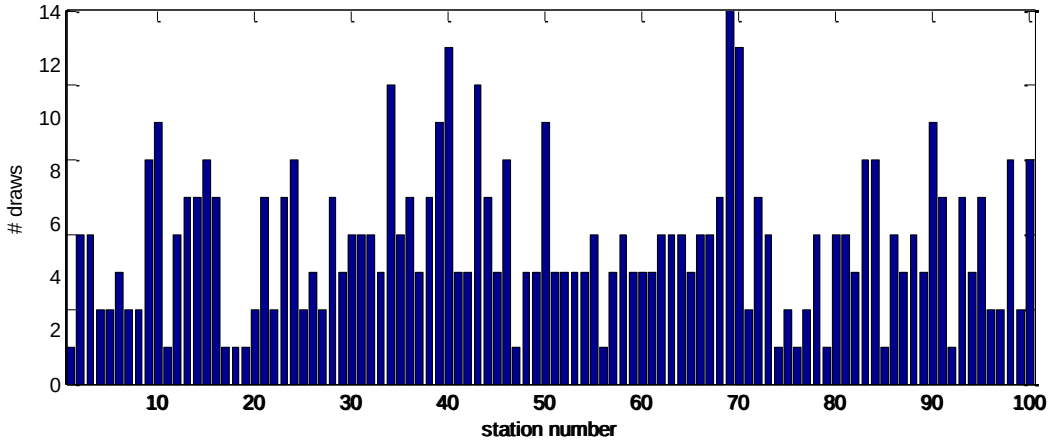
- Select 20% of the stations for validation
- Do it until every stations is selected at least once
- For most of the station you have more than one modelled time series
- Take the worst case (based on RMSE) for the final evaluation



© 2014, VITO NV

Monte Carlo distribution

- **Depending on chance** some stations are selected 1 times, some > 10 times



- **Preliminary consideration:**
  - How sensitive are results to n_{min} = 1? Alternatives?
  - Which time series to select final validation? Worst case?

© 2014, VITO NV

Methodology implementation/tests

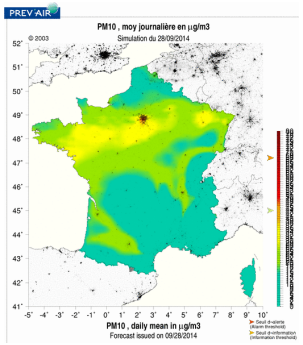
- 3 independent tests:
  - CHIMERE + Kriging (INERIS, L. Malherbe)
  - EURAD + bias correction (Uni. Aveiro, A. Monteiro)
  - RIO detrended kriging (VITO, B. Maiheu)

© 2014, VITO NV

Application : CHIMERE

-Meteorology
- Emissions
- Landuse
- Boundary conditions

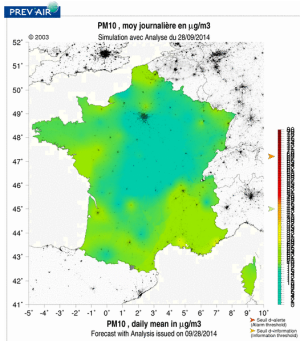
CHIMERE



Example, PM_{10}
Map produced on the 28th of September for the 27th

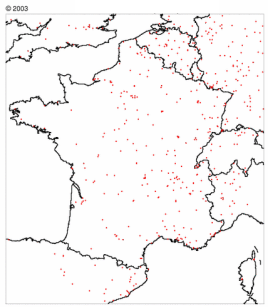
D-1, 27 September 2014, daily mean
0.1° x 0.15°

ANALYSIS



Analysed map

D-1, 27 September 2014, daily mean



Monitoring data (France + Europe)

Combination of background observations with CHIMERE data

Geostatistical approach: **external drift kriging**

The kriging is done for each hour (input data: hourly values) or each day (input data: average daily values).
It is implemented with R: *RGeostats* (Renard, 2010) and *gstat* (Pebesma, 2004) packages.

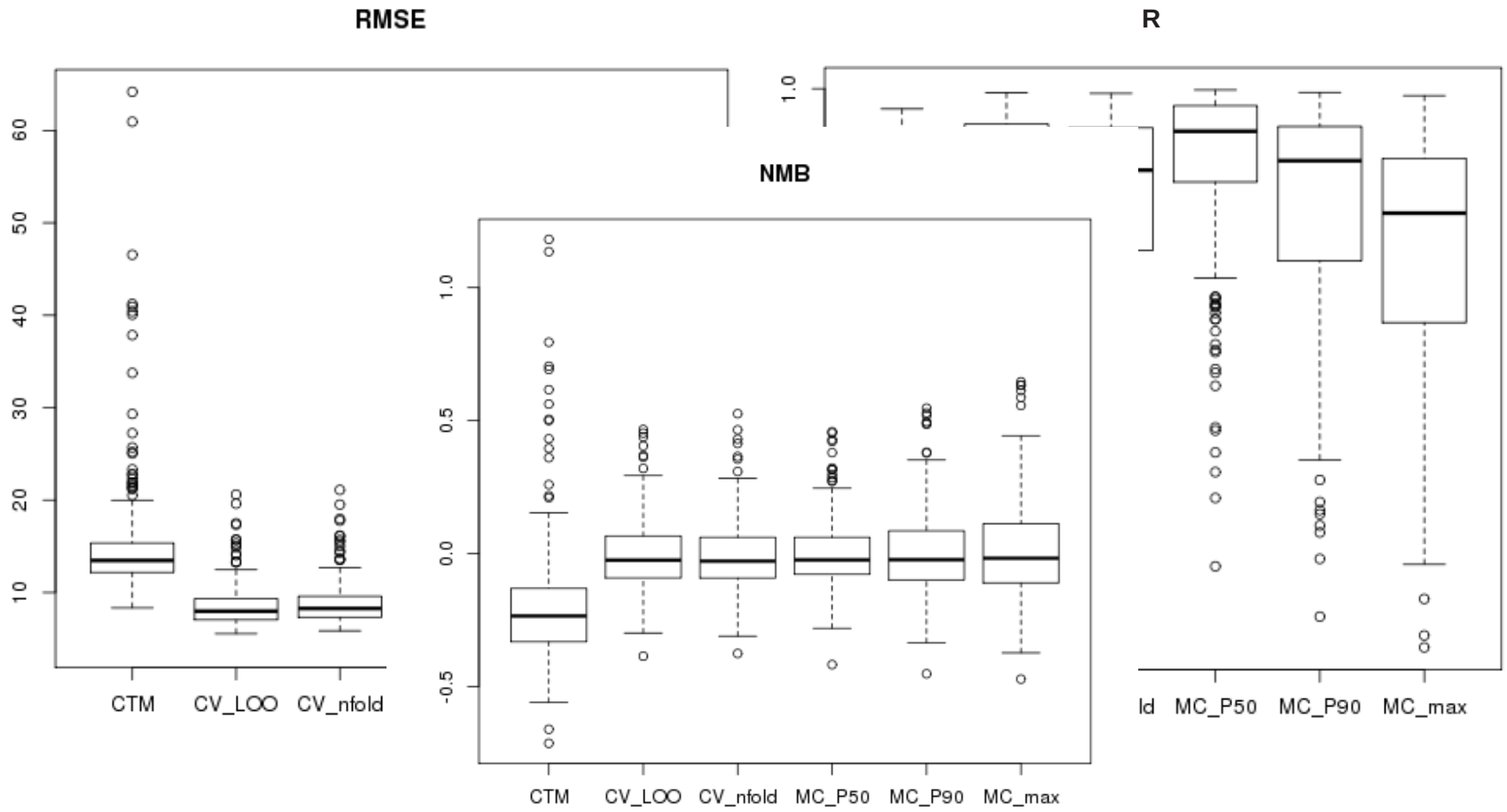
© 2014, VITO NV

Application : CHIMERE + Kriging

- The methodology has been tested for:
  - the French domain,
  - **PM_{10}** ,
  - the whole year **2012**, on an **hourly basis**.
- Input data:
  - hourly time series of PM_{10} concentrations measured at rural and suburban or urban background stations in France and surrounding countries (source: French national AQ database and Airbase v8)
  - hourly time series simulated by CHIMERE CTM with a spatial resolution of approximately 4km
- Monte-Carlo parameters:
  - **20% of stations removed for validation** at each random selection (function *sample* of R)
  - Number *n* of random selections: ***n* = 200**, ***n* = 300**, and ***n* = 500**
  - **Selection** of Monte Carlo member based on Max, P90, P50 (RMSE based)

© 2014, VITO NV

Validation comparison



© 2014, VITO NV

Application : UAVR

» **A bias correction** data fusion technique

STEP 1. **RAT04**
a **multiplicative ratio** correction with 4 days (for each station)

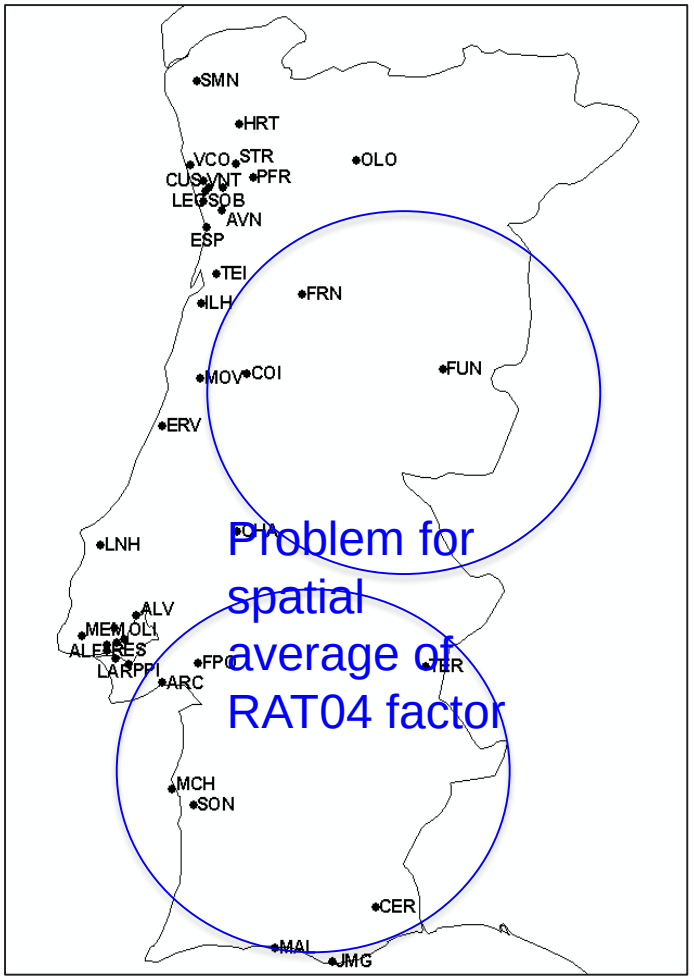
$$C^{corrected}(h,day) = \frac{\sum_{ndays} C^{obs}(h,day)}{\sum_{ndays} C^{model}(h,day)} \times C^{model}(h,day)$$

STEP 2. **Spatial** approach
Calculate the RAT04 average factor (per hour) and apply it to all grid cells

© 2014, VITO NV

Case study : modeling setup

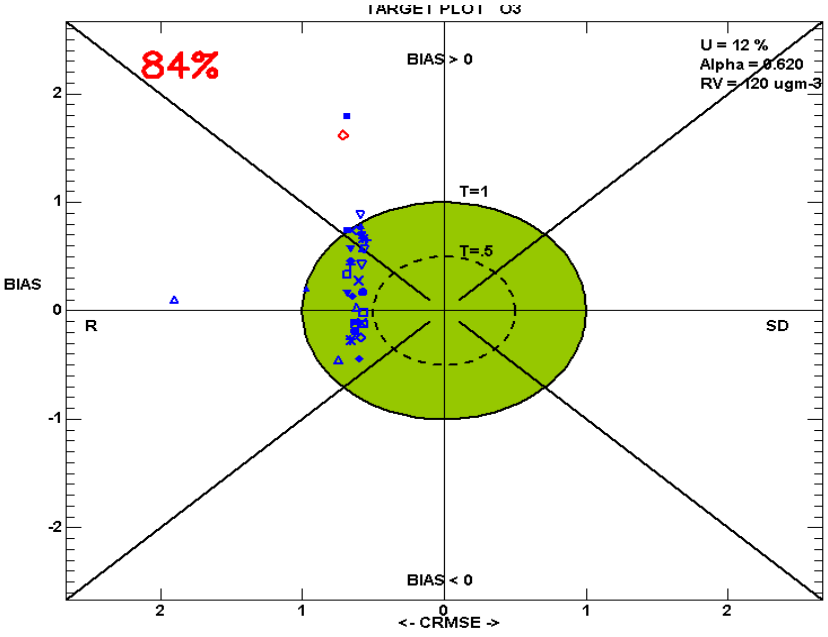
- 39 stations in Portugal
- Model : EURAD, 5x5 km^2 resolution
- Year : 2005
- O_3



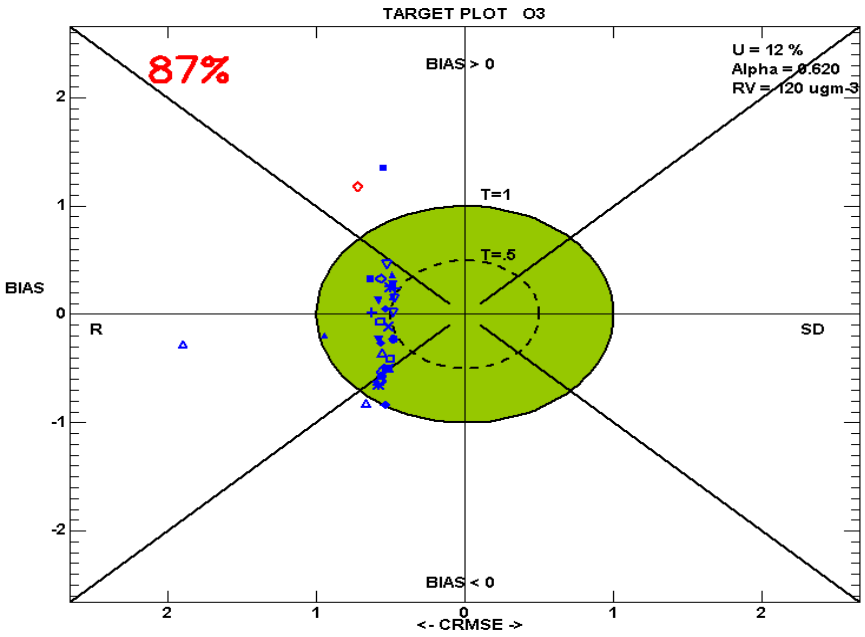
© 2014, VITO NV

Results UAVR

BEFORE RAT04



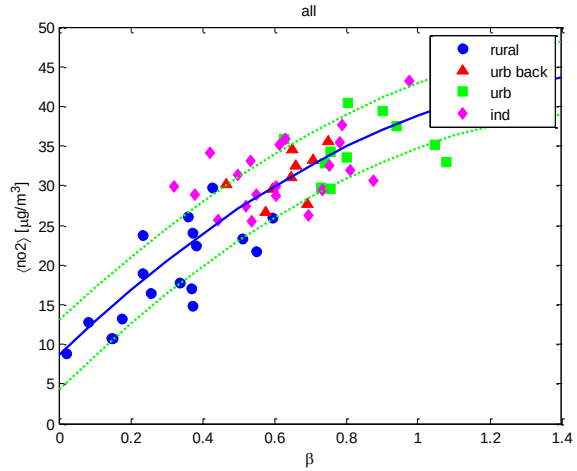
AFTER RAT04 and Monte Carlo approach

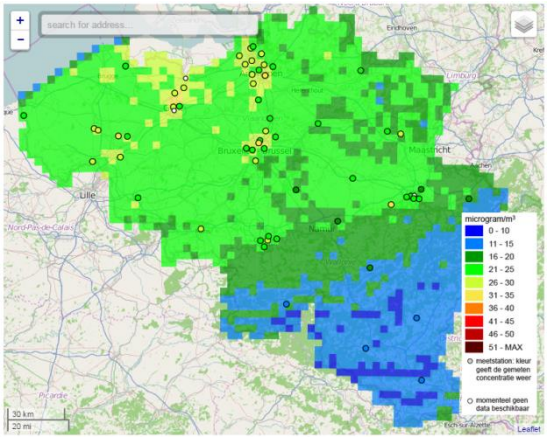


© 2014, VITO NV

The RIO model in 1 slide

- Detrended Kriging interpolation model
- Spatial trend captured by trendfunctions expressed vs. land use regression parameter β (CORINE).
  - per hour of the day, week/weekend

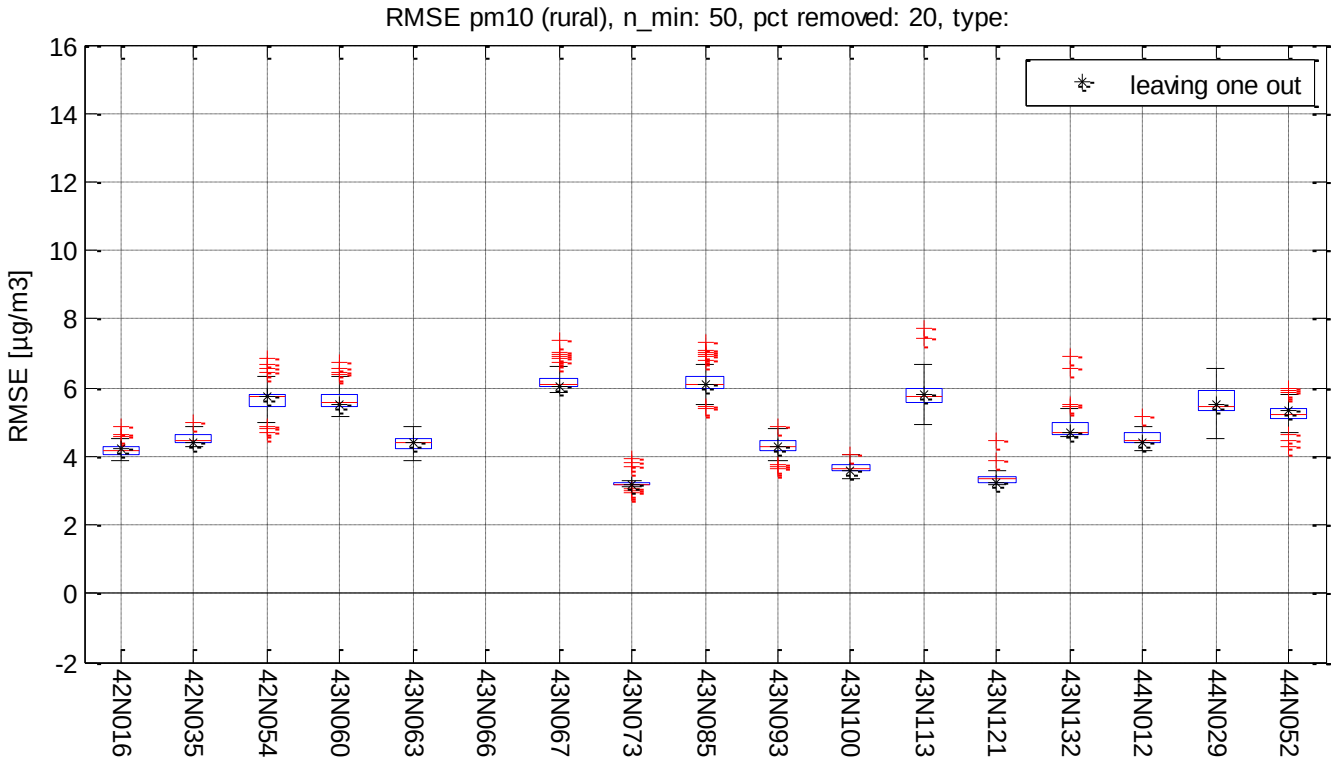




- Operational mapping model in Belgium (IRCELINE) & Netherlands (RIVM)

© 2014, VITO NV

RIO Monte Carlo Validation
RMSE – rural vs. n_min

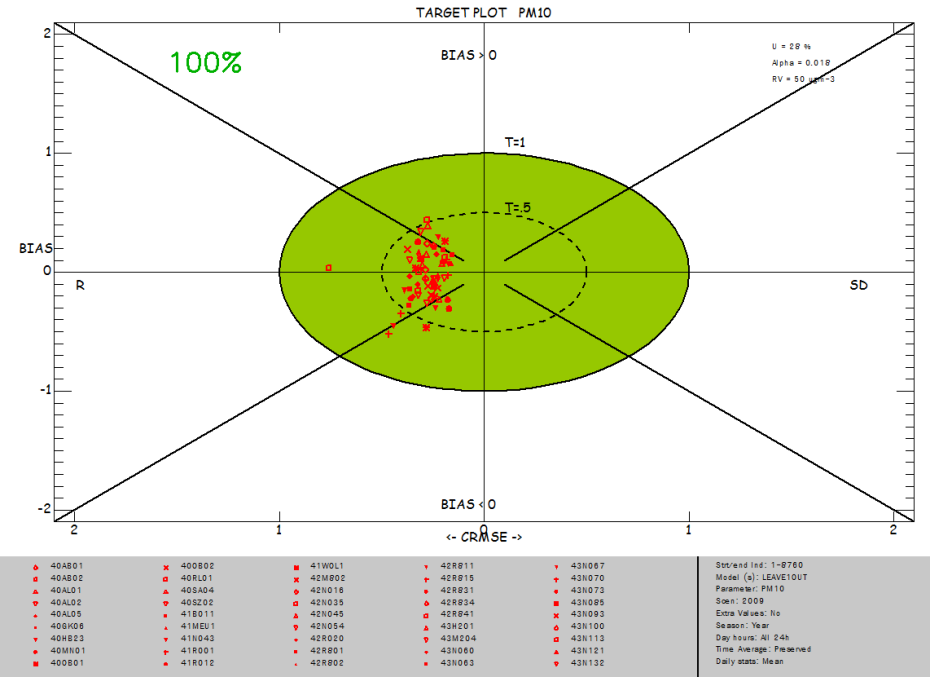


© 2014, VITO NV

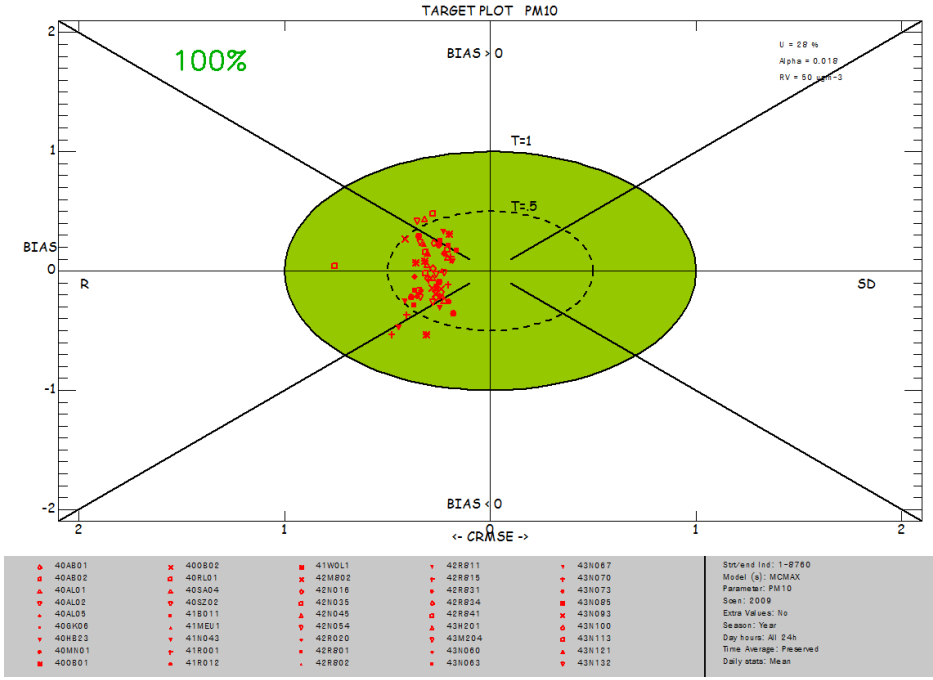
Comparison with LOO

» PM10 daily averages 2009

Leave one out



MC method - worst RMSE, n_min=1



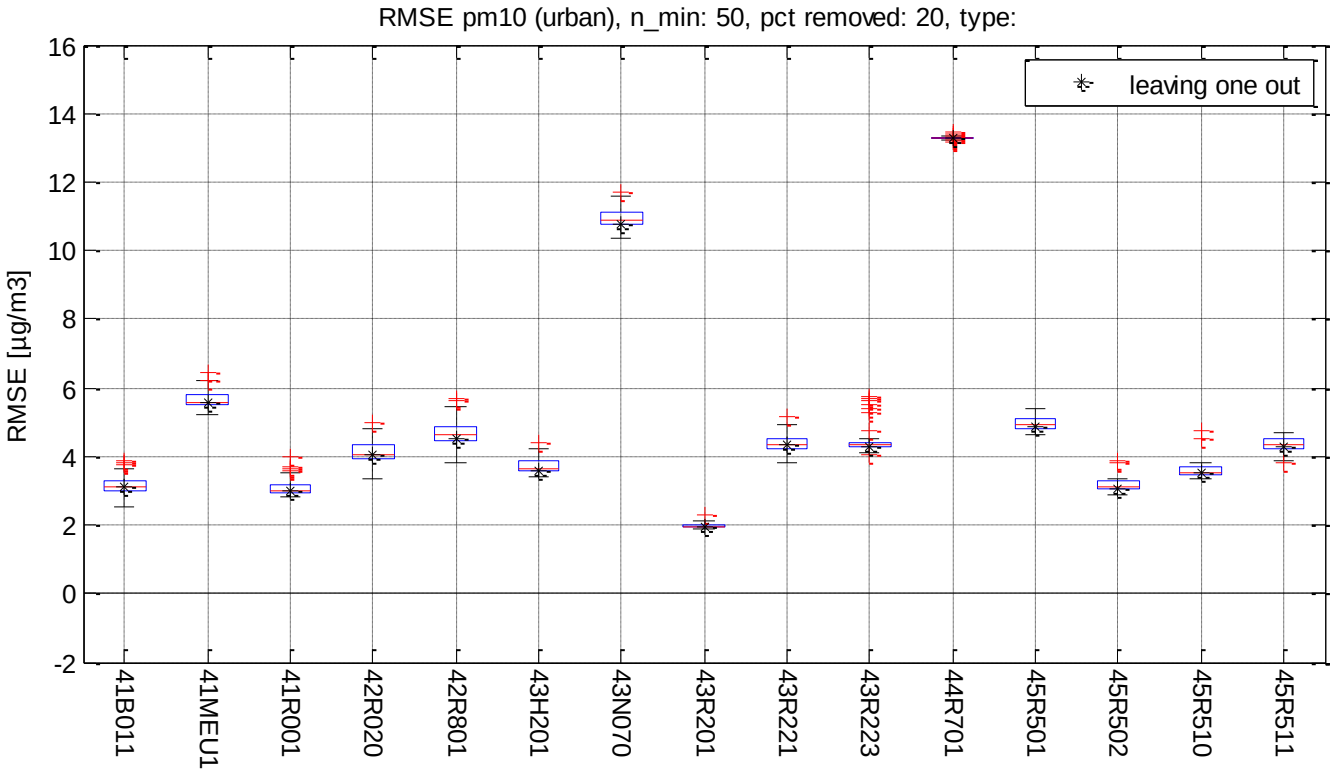
© 2014, VITO NV

General conclusions

- Rather significant effort to implement the Monte Carlo approach
- Still some points to clear out:
  - Is the selection of the worst RMSE the best way?
  - $n_{min} = 1$ → introduces randomness in the validation (especially when selecting the worst RMSE).
  - Checks needed for different pollutants, different situations
- Leaving-one-out is much easier to implement and seems to give similar results
- What about validation of more complex data assimilation schemes (Ens. Kalman filter, 4DVAR…)

© 2014, VITO NV

RIO Monte Carlo Validation
RMSE – urban vs. n_min



© 2014, VITO NV

Methodology output

Date	Obs	CTM	CV_LOO	CV_Nfold	MC_P50	MC_P90	MC_max
2012010101	15	7.6	20.0	24.0	20.0	27.1	33.1
2012010101	12	7.9	16.0	23.2	18.8	20.8	22.5
…	…	…	…	…	…	…	…
2013010100	…	…	…	…	…	…	…

Obs	**Measured value**	
CTM	CHIMERE (interpolation at the station)	
CV_LOO	Leave-one-out cross-validation	
CV_Nfold	5-fold cross-validation	
MC_P50	Monte-Carlo validation, estimated value corresponding to the median square error	added for comparison
MC_P90	Monte-Carlo validation, estimated value corresponding to the 90th percentile of the square error	added for comparison
MC_max	Monte-Carlo validation, estimated value with maximum square error (worst case)	

© 2014, VITO NV

Conclusions INERIS

- No significant difference according to the number of subset selections (n=200, n=300, n=500)
- In the present tests, performance criteria were satisfied. However, could the « worst case » be too penalizing? Consider a high percentile of the error instead of the maximum?
- The implementation requires attention but does not pose any particular problem
- The added value of the Monte-Carlo approach in relation to the usual leave-one-out or n-fold cross-validation will be further examined.

© 2014, VITO NV

Conclusions UAVR

- Very exhaustive methodology (mainly when the data fusion technique do not bring significant improvements)
- Only operational/automated is feasible
- A group of Matlab/Python programs was developed by UAVR and can be available for FAIRMODE community
- Maximum RMSE per station or per iteration (re-analysis) should be reviewed
- Results still need a deep analysis: too “fresh”!!

© 2014, VITO NV

Conclusions VITO

- Monte Carlo method seems to be quite robust for RIO w.r.t. leaving-one-out (at least for PM_{10})… at first sight.
  - Clustering of stations in urban area's
  - PM_{10} more regional pollutant → rural stations
- Look at other pollutants (NO_2, O_3) to confirm/reject
- Monte Carlo method not always yield worse statistics when looking at median
- Using worst RMSE is sensitive to N_{min}
  - Need to check what is happening with the distributions : increase in outliers
- A the moment : using daily averages
  - Computation time could become issue (for a "simple validation")

© 2014, VITO NV