27/07/2015

Monte Carlo Validation

Application for RIO interpolation model
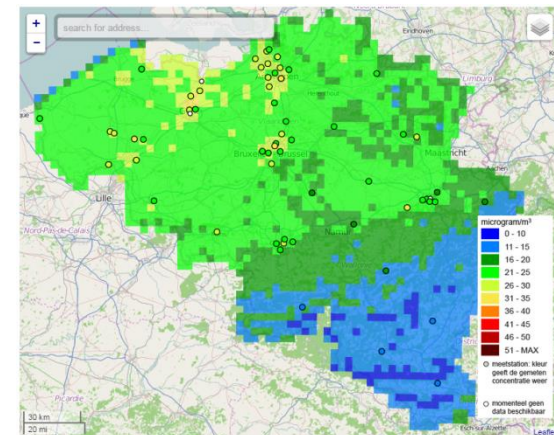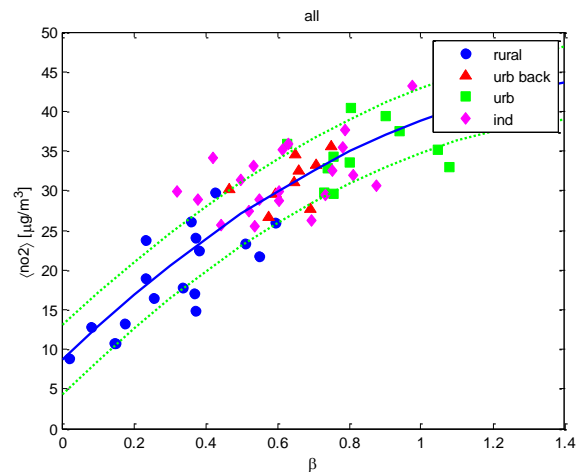
bino.maiheu@vito.be

Outline

» RIO in 1 slide

» RIO validation, what is "the model" ?

» Monte Carlo validation

    » Some analyses

    » Impact on target plot results for RIO

» Conclusions

The RIO model in 1 slide
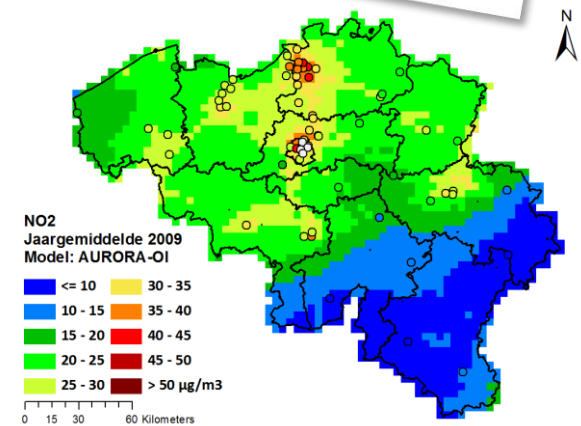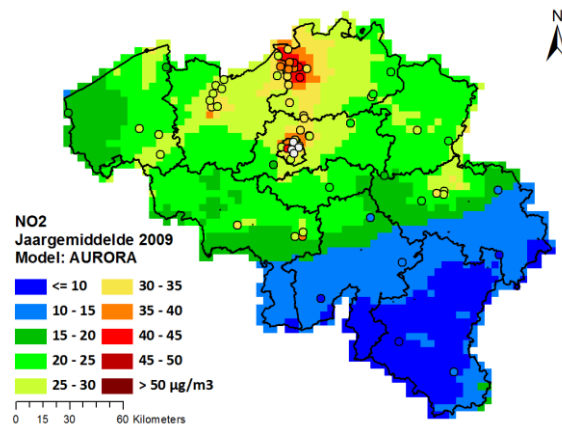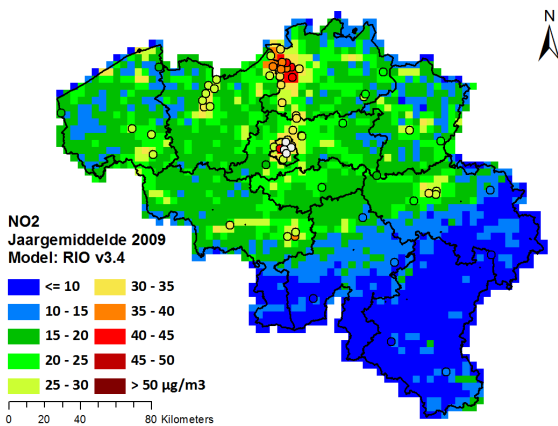
» Detrended Kriging interpolation model

» Spatial trend captured by trendfunctions expressed vs. land use regression parameter β (CORINE).

  » per hour of the day, week/weekend

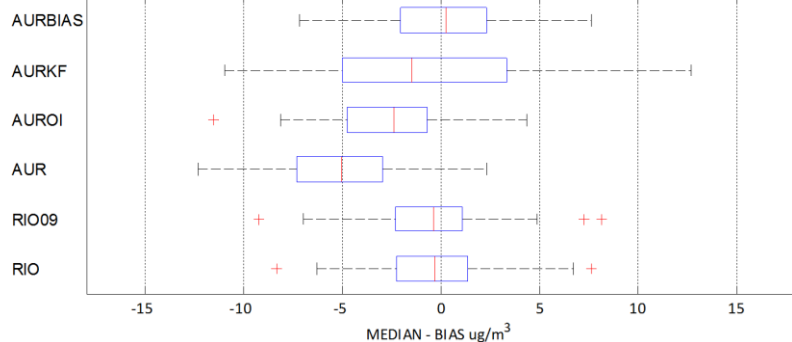» Operational mapping model in Belgium (IRCELINE) & Netherlands (RIVM)

vision on technology

RIO validation

» Elaborate model comparison (2012)
  » RIO (Interpolation model)
  » AURORA CTM (+ callibration/DA: OI)
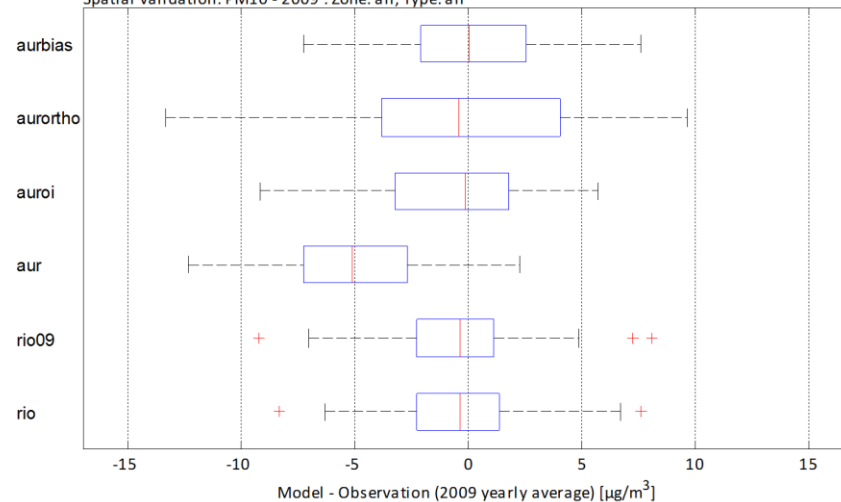 » OPS
» 2009, ~~Traffic stations~~
» Leaving one out

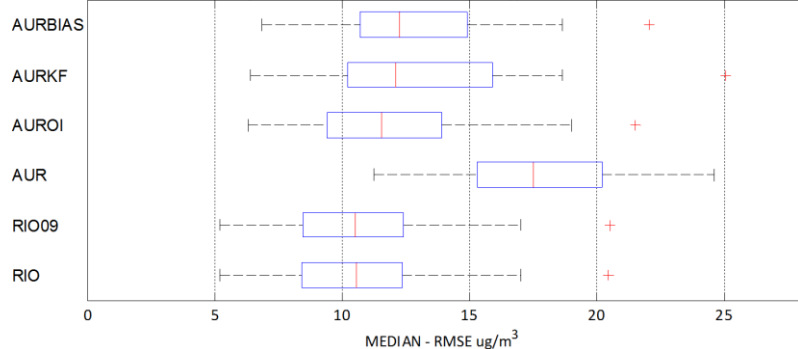Bepaling van de best beschikbare grootschalige concentratiekaarten luchtkwaliteit voor België

Bino Maiheu, Nele Veldeman, Peter Viaene, Koen De Ridder, Dirk Lauwaet, Felix Deutsch, Stijn Janssen

Unit Ruimtelijke Milieuaspecten
Vlaamse Instelling voor Technologisch Onderzoek

Studie uitgevoerd in opdracht van MIRA, Milieurapport Vlaanderen

MIRA/2013/01

December 2012

NO2
Jaargemiddelde 2009
Model: RIO v3.4

<= 10	30 - 35
10 - 15	35 - 40
15 - 20	40 - 45
20 - 25	45 - 50
25 - 30	> 50 µg/m3

0   20   40   80 Kilometers

NO2
Jaargemiddelde 2009
Model: AURORA

<= 10	30 - 35
10 - 15	35 - 40
15 - 20	40 - 45
20 - 25	45 - 50
25 - 30	> 50 µg/m3

0   15   30   60 Kilometers

NO2
Jaargemiddelde 2009
Model: AURORA-OI

<= 10	30 - 35
10 - 15	35 - 40
15 - 20	40 - 45
20 - 25	45 - 50
25 - 30	> 50 µg/m3

0   15   30   60 Kilometers

vito
vision on technology

© 2014, VITO NV
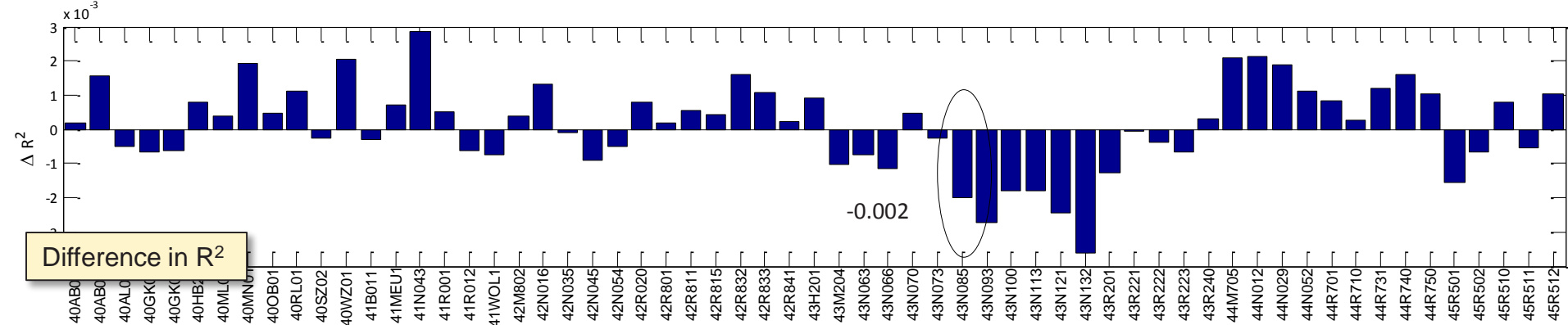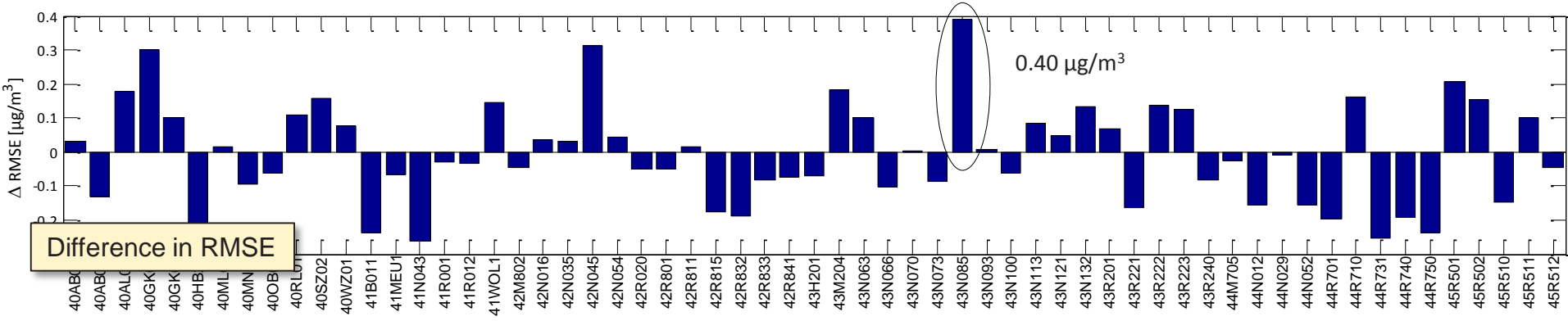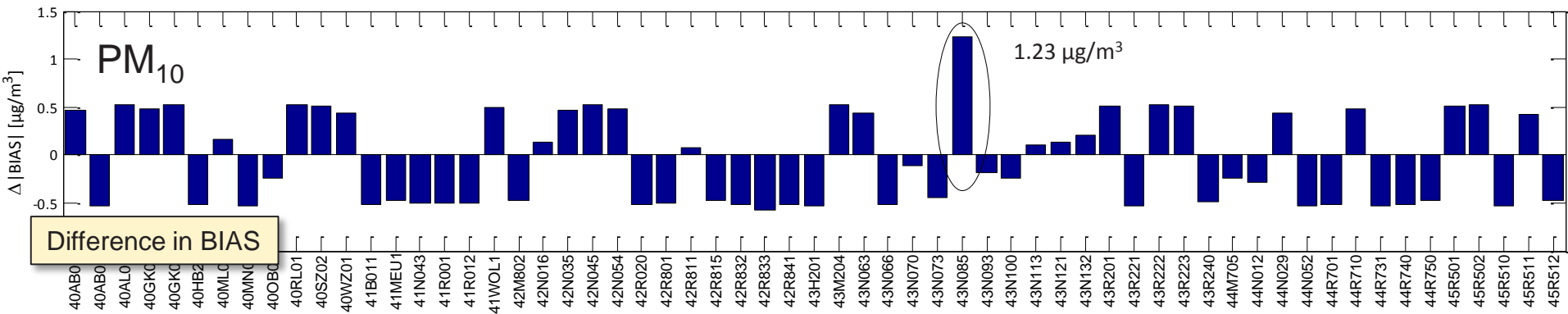
Model comparison exercise

© 2014, VITO NV
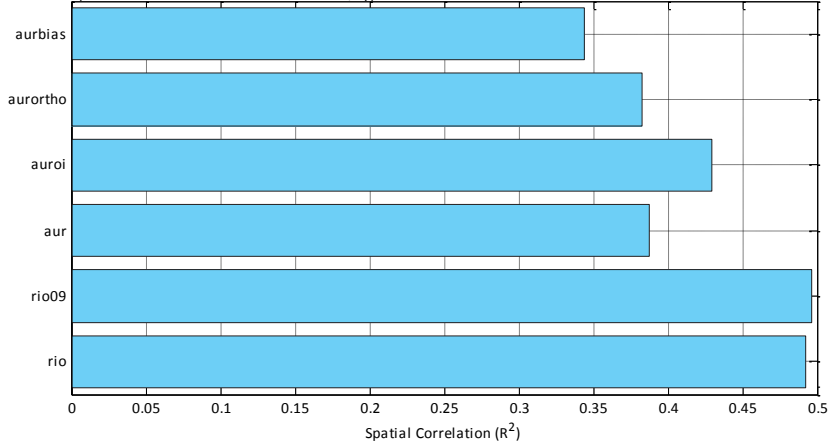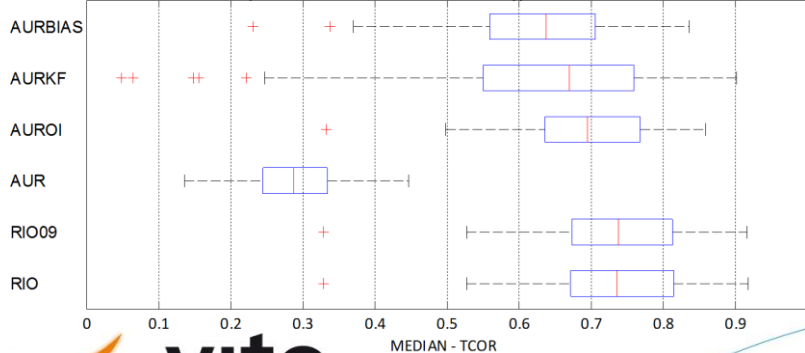
DA/interpolation model-calibration ?

» RIO uses measurements

  » Model configuration : trendfunctions, long term averages, spatial correlations, land use regression weights (β)

  » Interpolations : hourly measurements from available stations

» DA (e.g. OI) similar !

  » Model configuration : error covariance matrices…

» Leaving-one-out : what do we leave out ?

  » → RIO : influence of leaving out one station in trend fitting
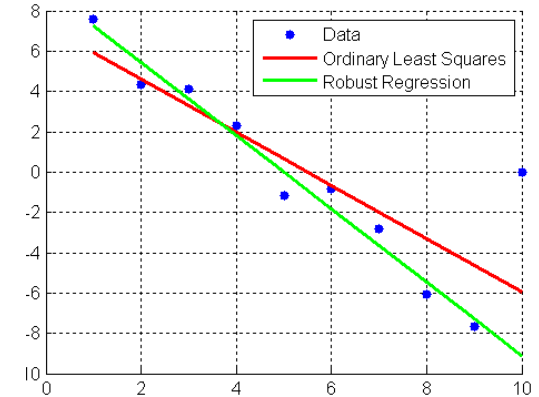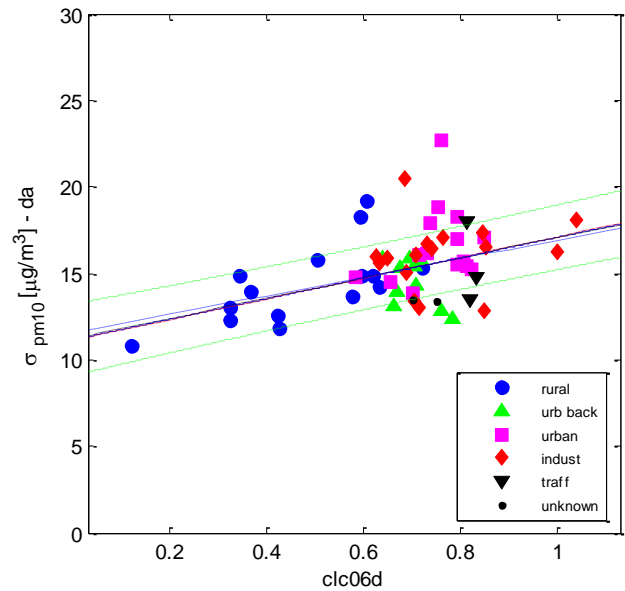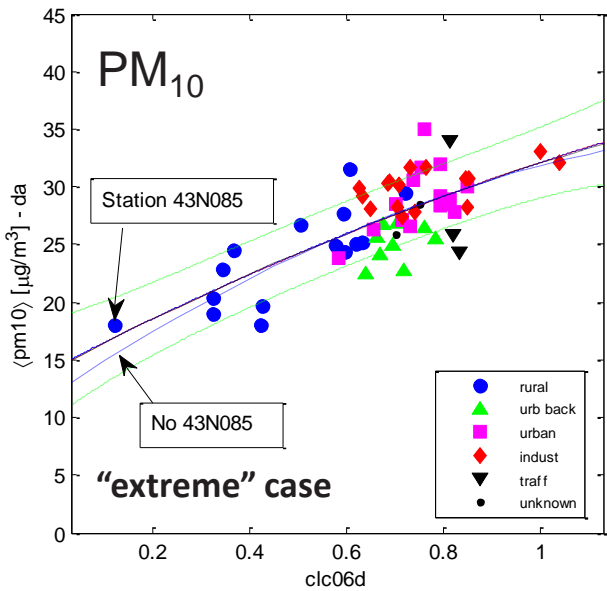
Impact on (temporal) validation statistics

Verschil in validatie statistieken: Geen 43N085 t.o.v. basis RIO v3.4

PM$_{10}$

Difference in BIAS

1.23 µg/m³

Difference in RMSE

0.40 µg/m³

Difference in R²

-0.002

Some thoughts

» Differences relatively small compared to differences in RMSE, BIAS & R^2 between models

» How do we define the "model" in this case ?

  » Trendfunctions, spatial correlations, long term averages considerd part of the model.

vito
vision on technology

Monte Carlo validation

» MC methodology (C. Carnevale)

» **repeat until** (each station is left out at least once (=n_{min})):
   » leave N stations out at random (typ. 20 %)
   » validate
   » compute RMSE for stations that were left out
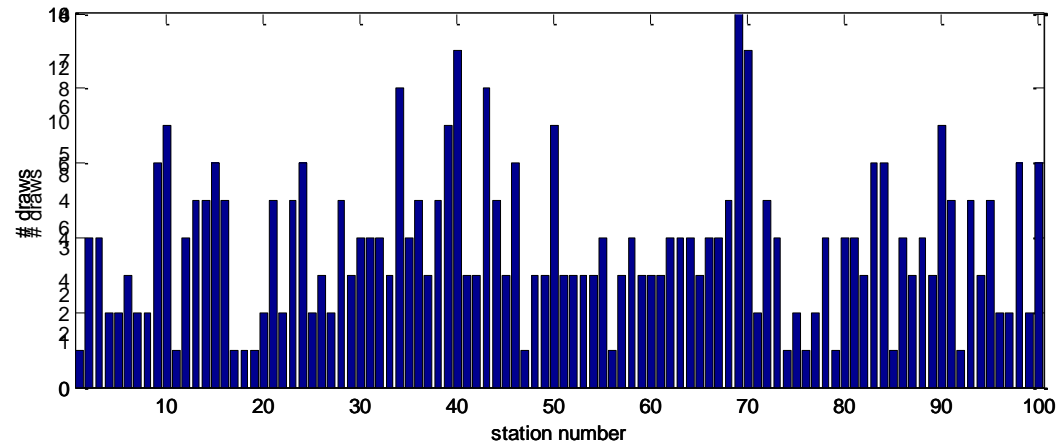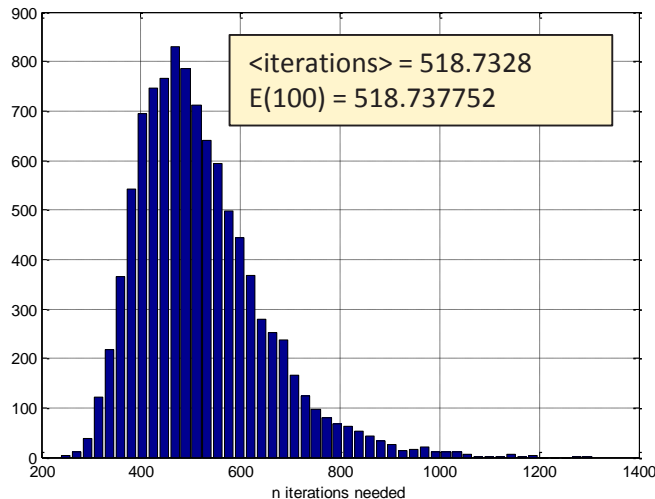» Use validation statistics of series which give worst RMSE in Δ-tool

"Coupon collector's problem"

» "Given *n* coupons, how many coupons X do you expect you need to draw with replacement before having drawn each coupon at least once?"

   » https://en.wikipedia.org/wiki/Coupon_collector%27s_problem

» $E(X) = n \sum_{i=1}^{n} \frac{1}{i}$

N=100, m=1, n_min=1
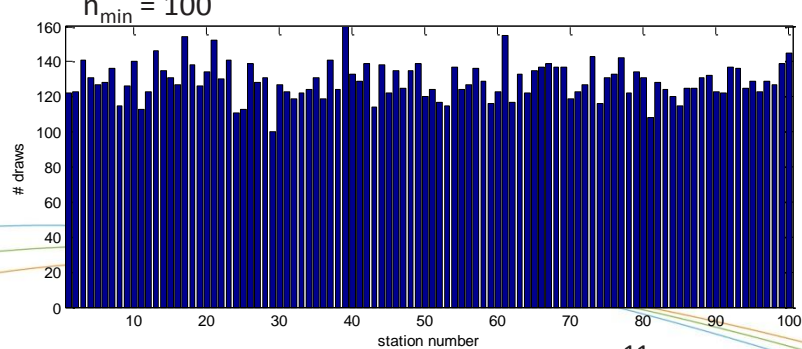

<iterations> = 518.7328
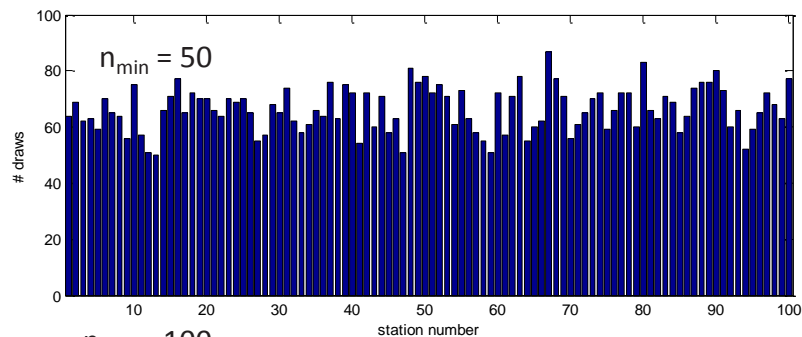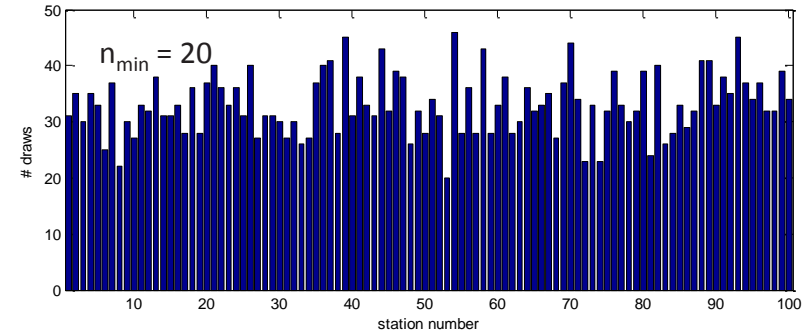E(100) = 518.737752

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

vito
vision on technology

Monte Carlo validation

» Select each station at least n_{min} times

» How many times a station is drawn
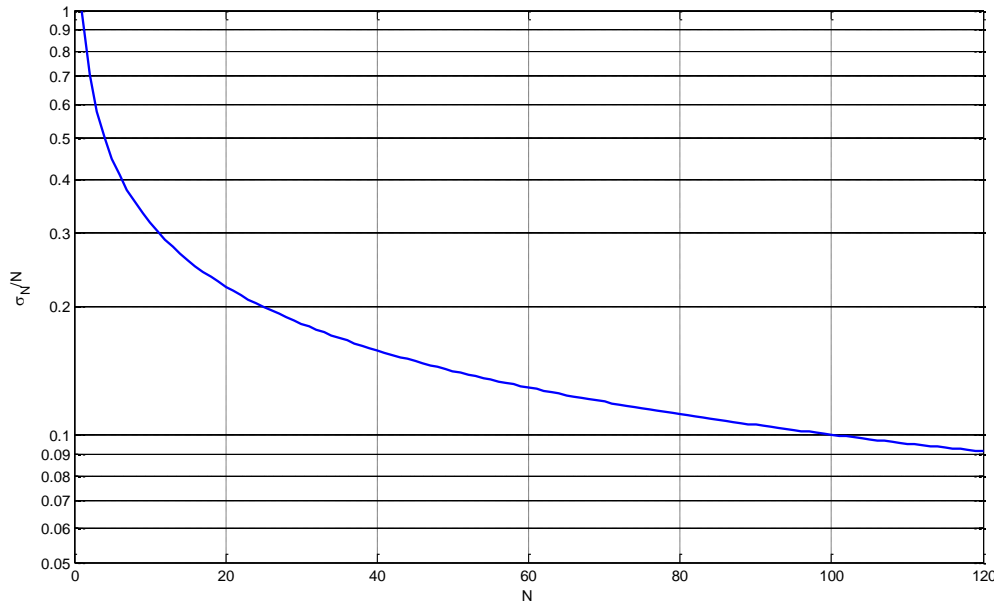
» How robust is the validation ?
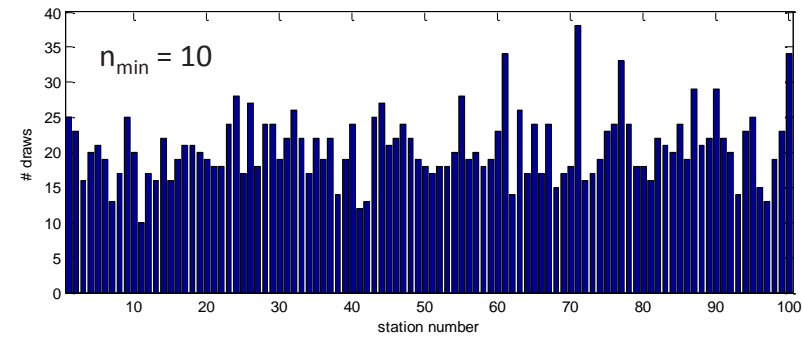
RIO Monte Carlo Validation
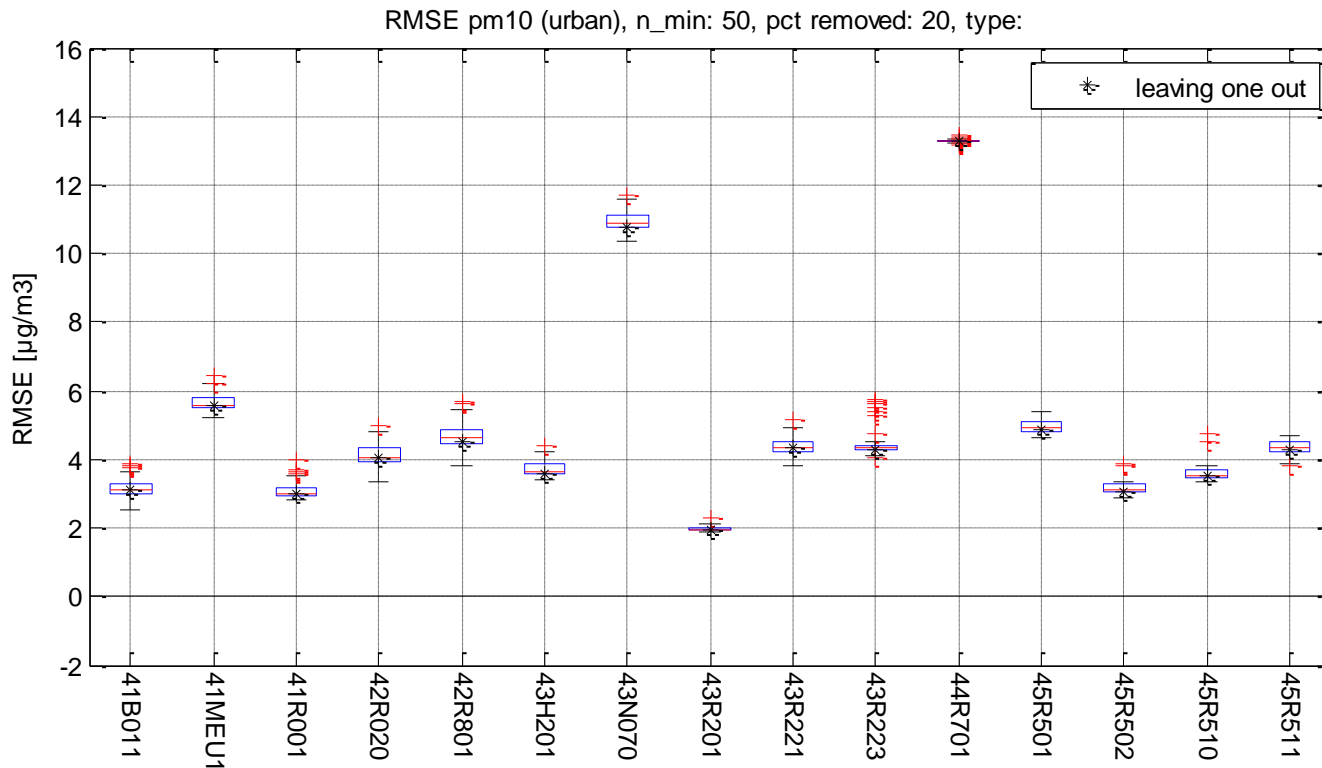# RMSE – rural vs. n_min

RMSE pm10 (rural), n_min: 50, pct removed: 20, type:

vito
vision on technology
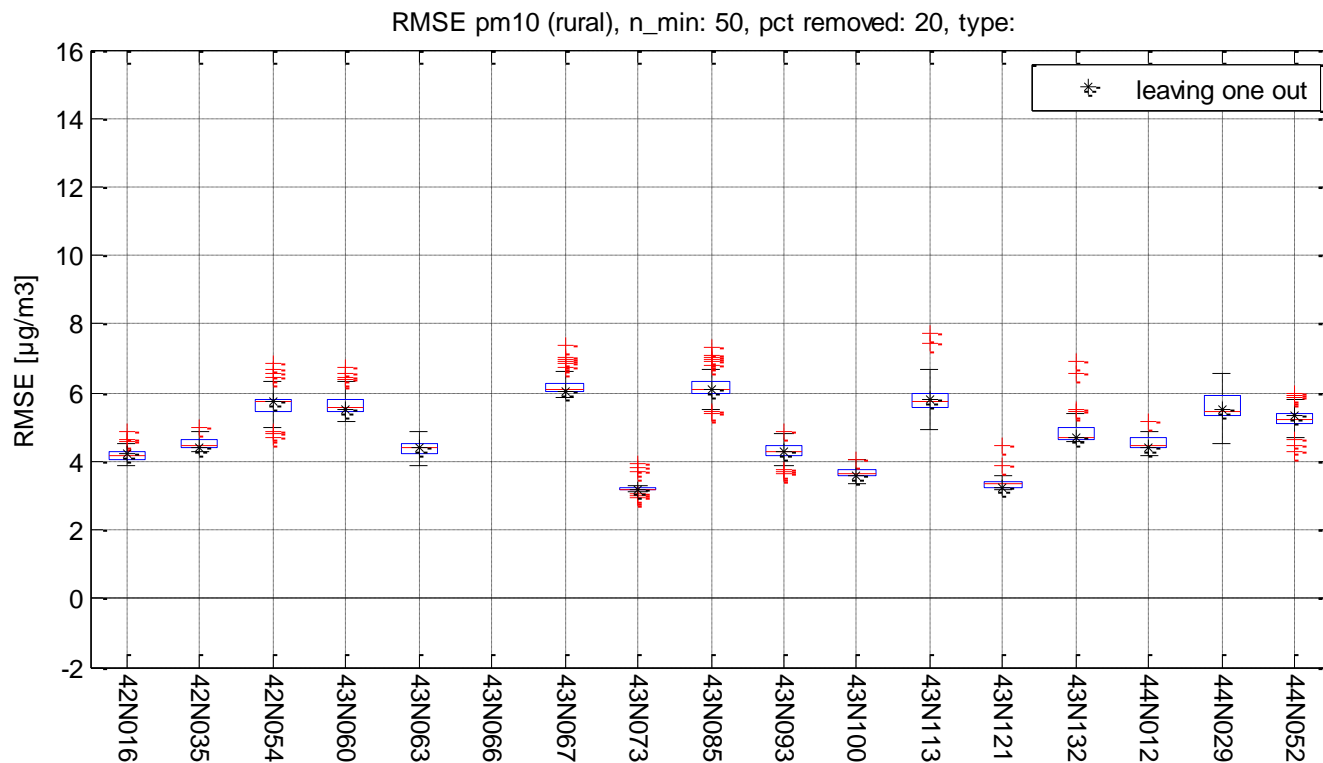
RIO Monte Carlo Validation
RMSE – urban vs. n_min

RMSE pm10 (urban), n_min: 50, pct removed: 20, type:

vito
vision on technology

RIO Monte Carlo Validation
R2 – rural vs. n_min

R^2 pm10 (rural), n_min: 50, pct removed: 20, type:

vito
vision on technology
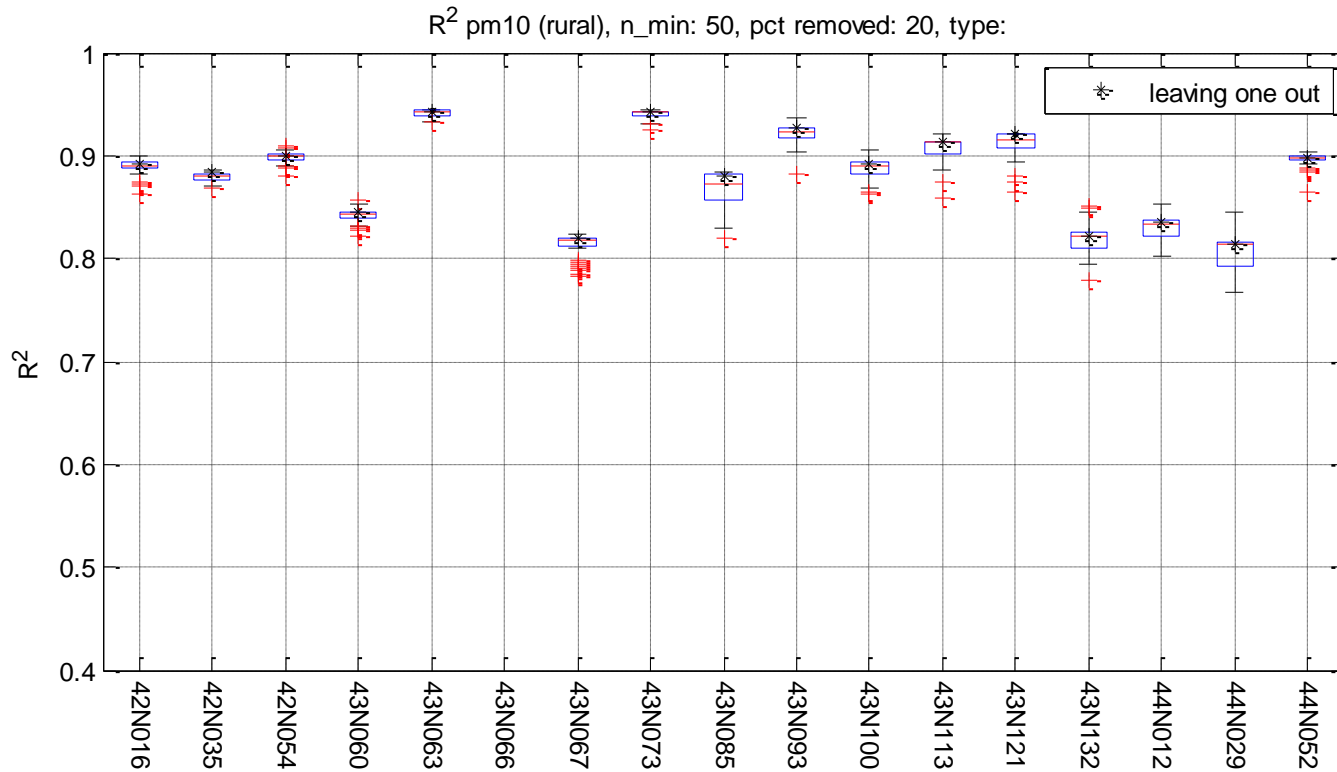
RIO Monte Carlo Validation
R2 – urban vs. n_min

R^2 pm10 (urban), n_min: 50, pct removed: 20, type:
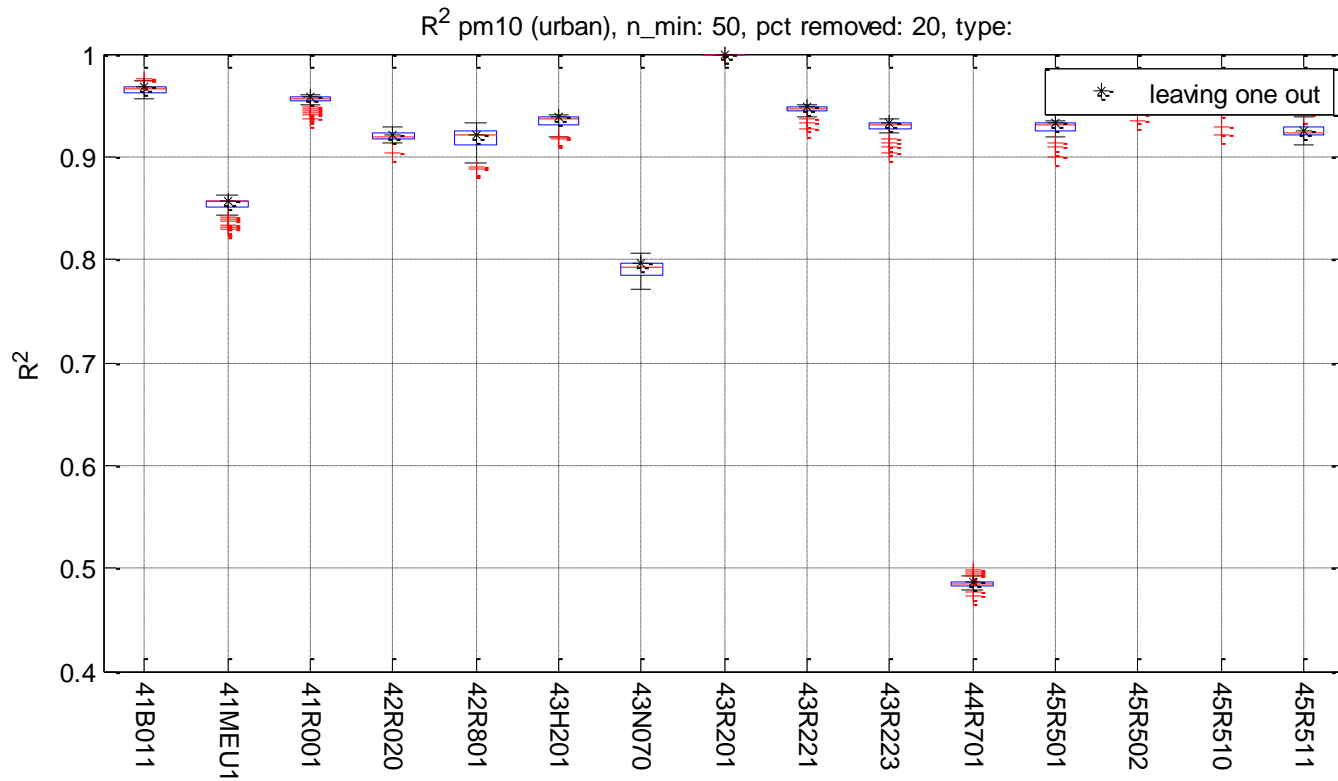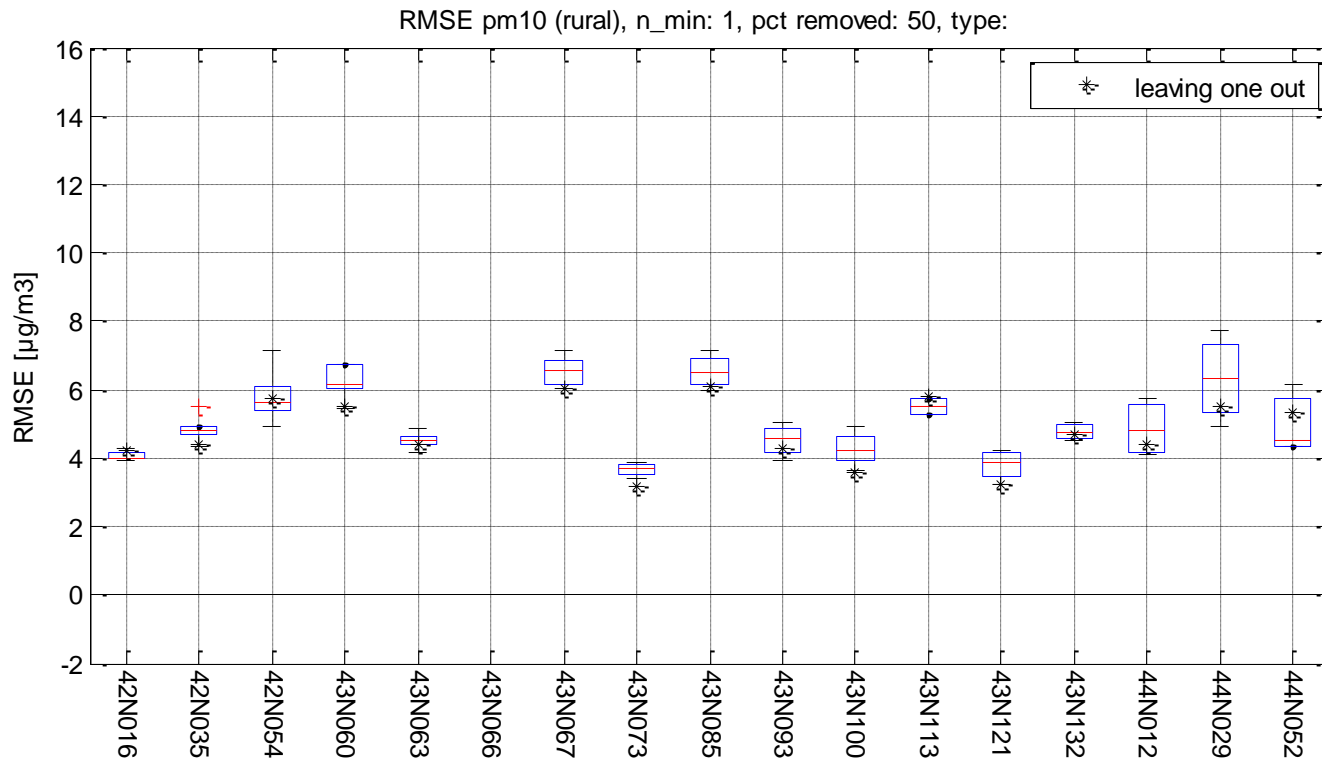
Observations

» No dramatic differences w.r.t. the percentage of stations removed

» Leaving one out does not necissarily yield best RMSE !

» Clustering of stations

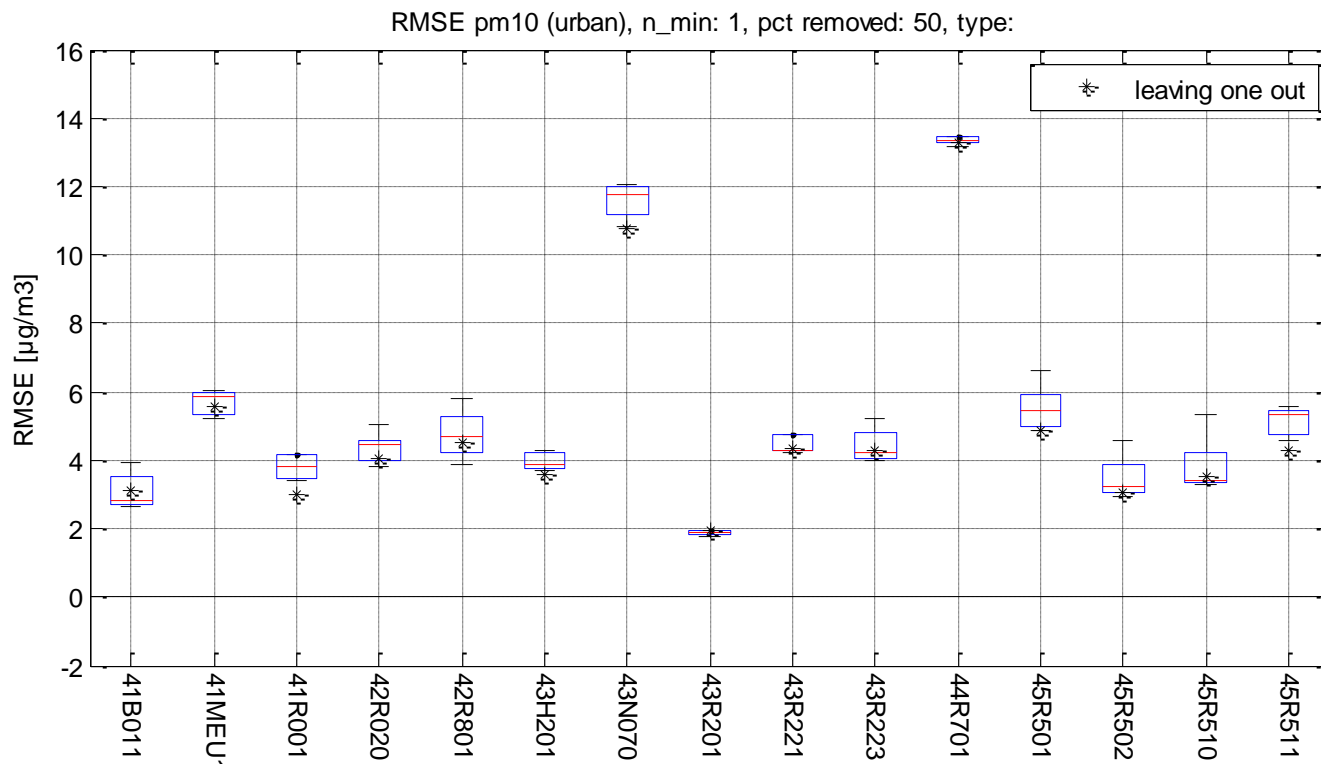
NO2
Jaargemiddelde 2009
Model: RIO v3.4

<= 10	30 - 35
10 - 15	35 - 40
15 - 20	40 - 45
20 - 25	45 - 50
25 - 30	> 50 µg/m3

0   20   40   80 Kilometers

vito
vision on technology
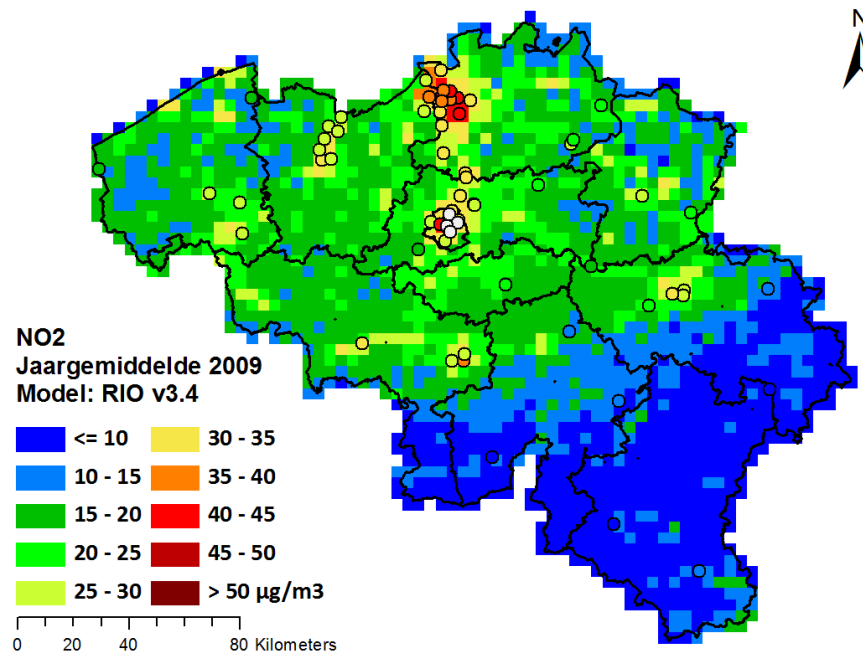
RIO Monte Carlo Validation
PM10 - rural vs. pct. removed

RMSE pm10 (rural), n_min: 1, pct removed: 50, type:

© 2014, VITO NV

vito
vision on technology

RIO Monte Carlo Validation
PM10 – urban vs. percentage removed

RMSE pm10 (urban), n_min: 1, pct removed: 50, type:

vito
vision on technology

Validation in Δ tool

» PM10 da 2009

» Leave 1 out

vito
vision on technology
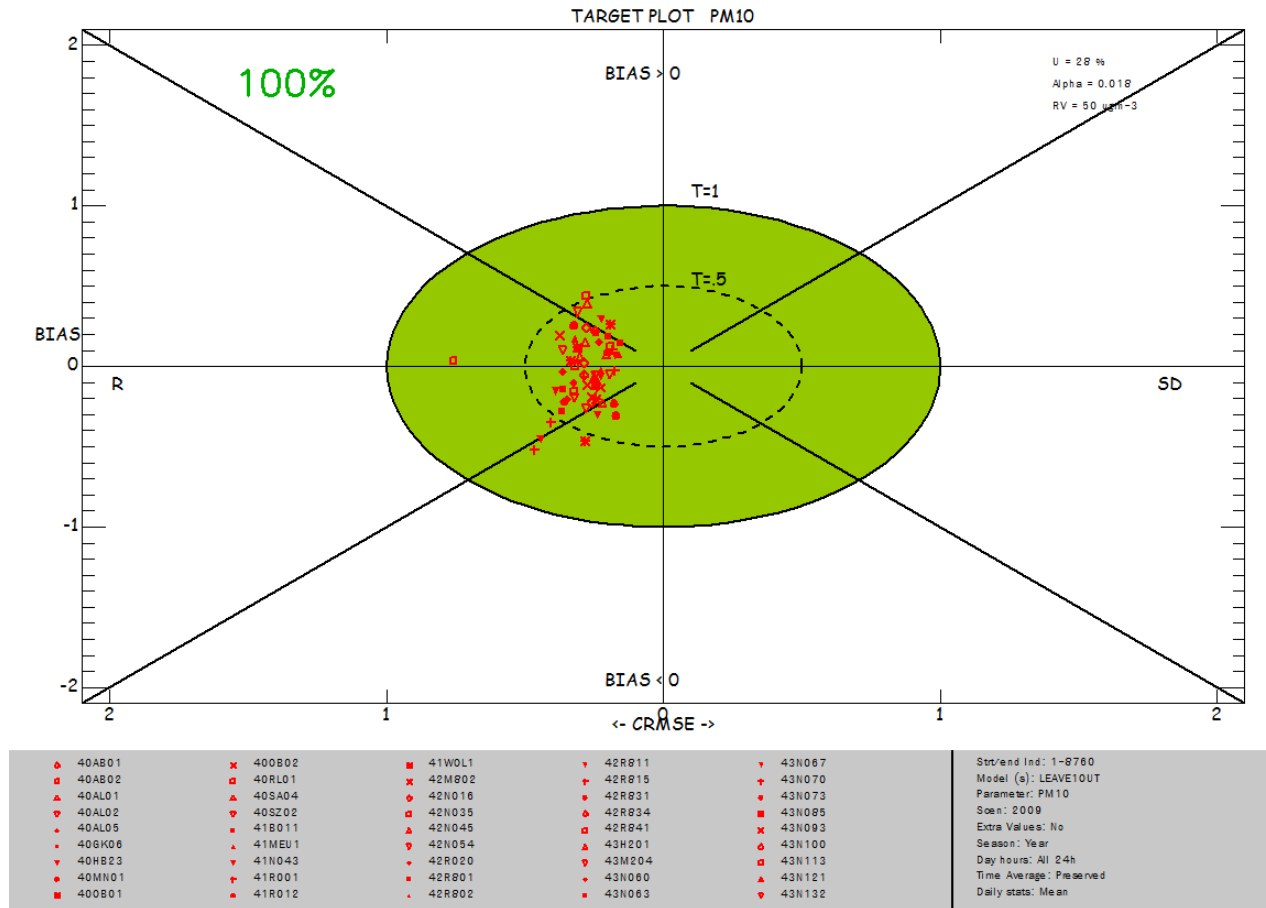
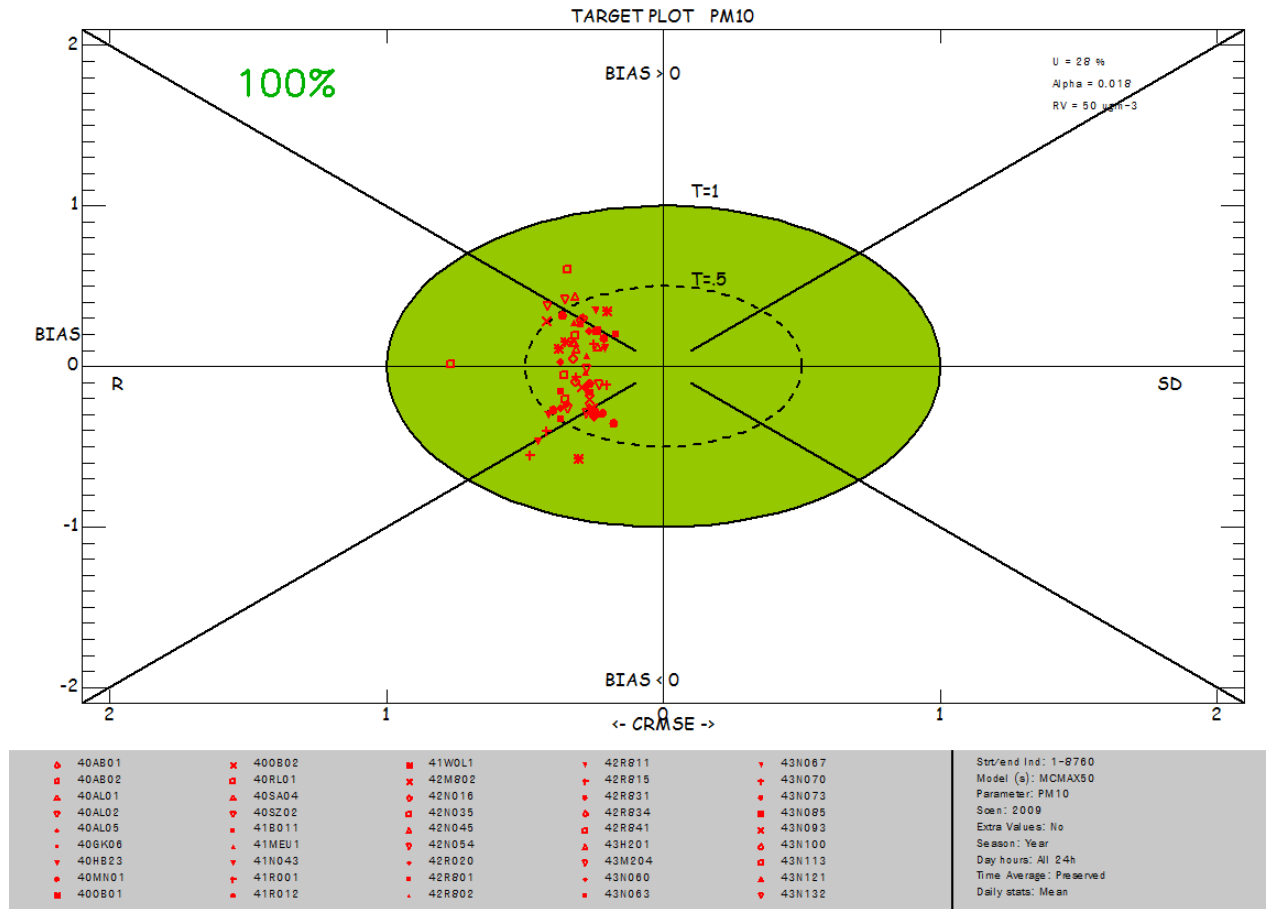# Validation in Δ tool

» MC Validation

» N_{min} 1

» Worst RMSE

Validation in Δ tool

» MC Validation

» N_{min} 50

» Worst RMSE

© 2014, VITO NV

vision on technology

Validation in Δ tool

» PM10 da 2009

» Leave 1 out
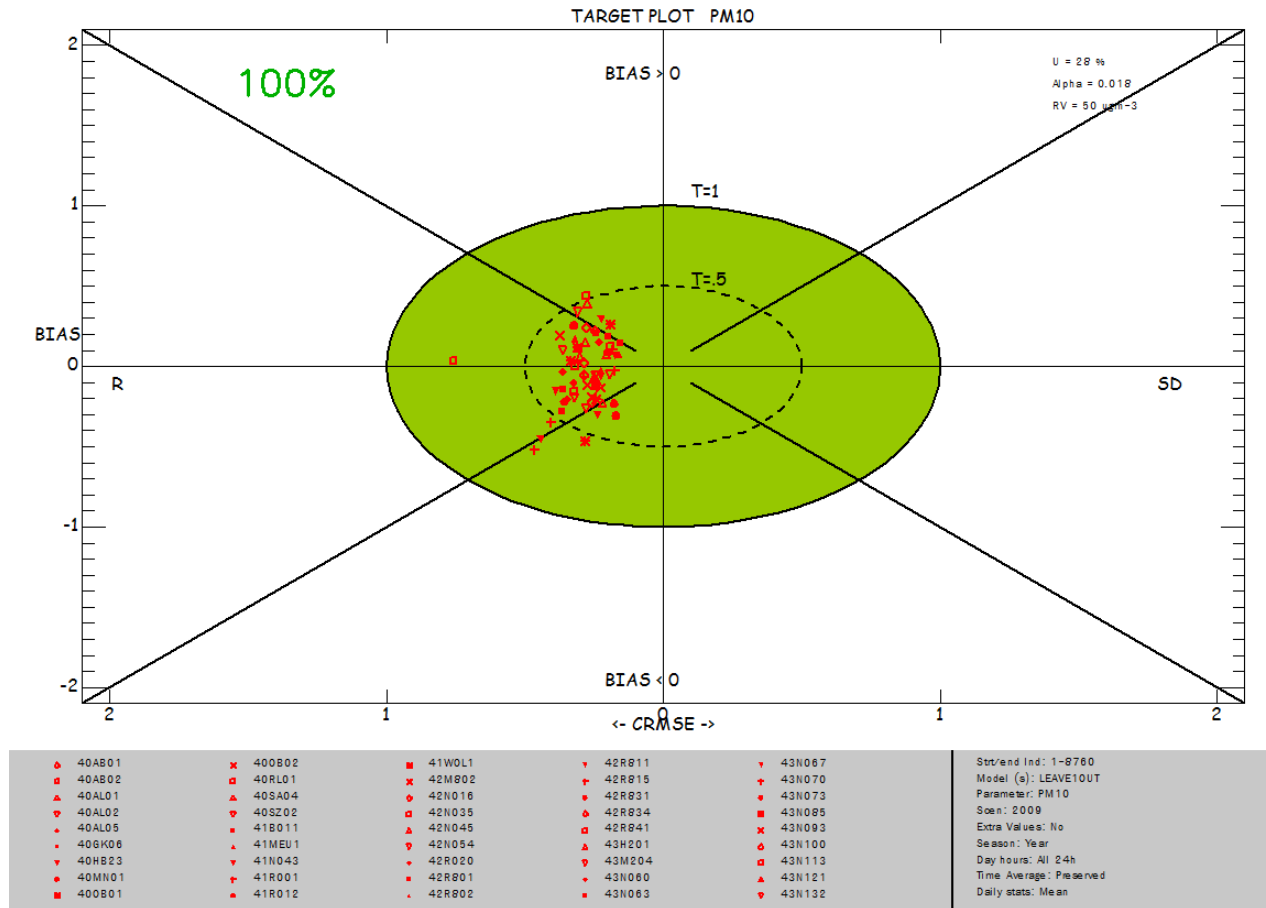
Validation in Δ tool

» MC Validation
» N$_{min}$ 50
» Median RMSE

Conclusions & discussion
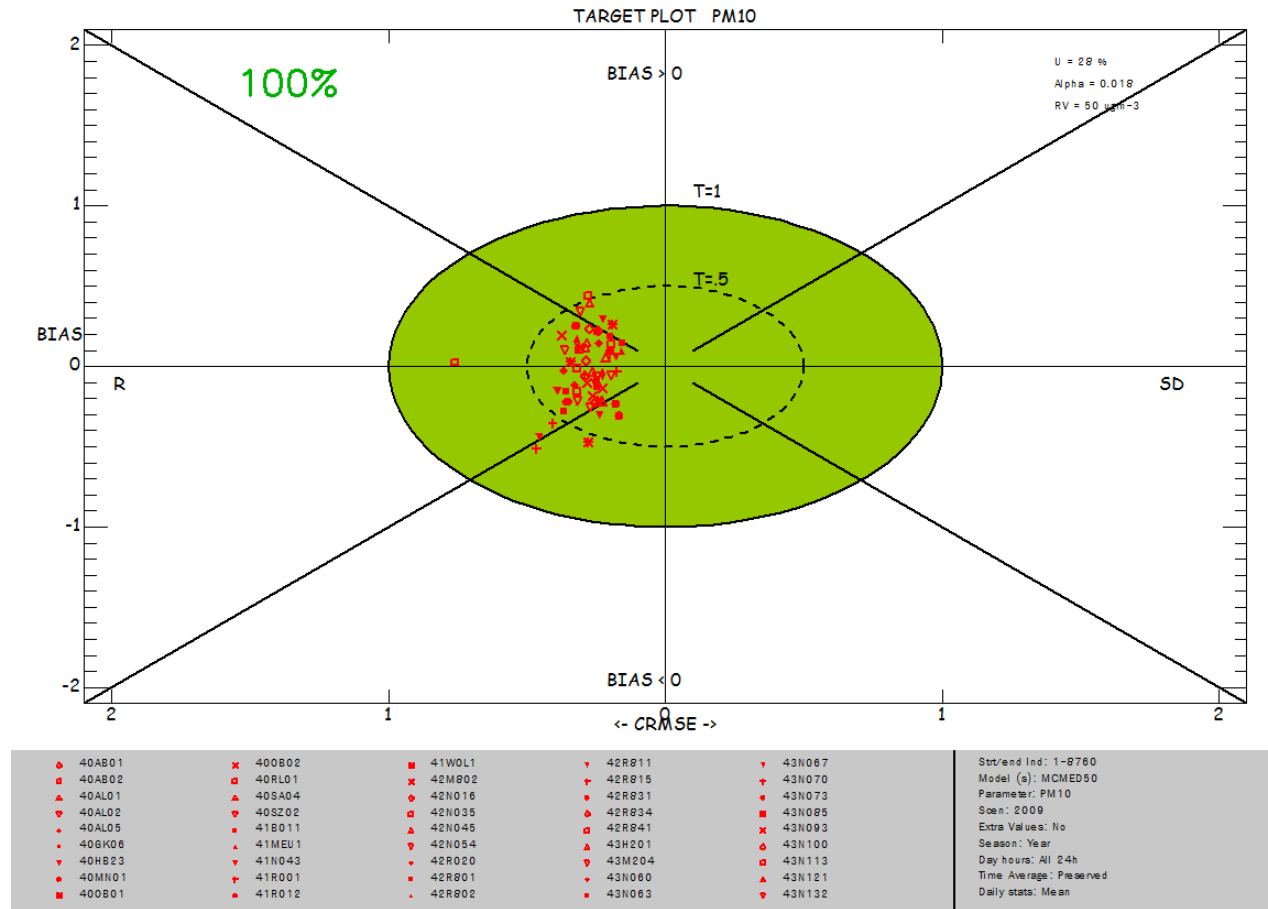
- » What is "the model" ?
- » Monte Carlo method seems to be quite robust for RIO w.r.t. leaving-one-out (at least for PM_{10})... at first sight.
  - » Clustering of stations in urban area's
  - » PM_{10} more regional pollutant → rural stations
- » Look at other pollutants (NO_2, O_3) to confirm/reject
- » 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")