

WG1 | CCA Modelling & Monitoring

Evaluation of the re-analysis validation methodology for Portugal

Alexandra Monteiro
Ana Patrícia Fernandes
Ana Isabel Miranda



universidade de aveiro

FAIRMODE Technical Meeting | Aveiro 24-25 June 2015

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

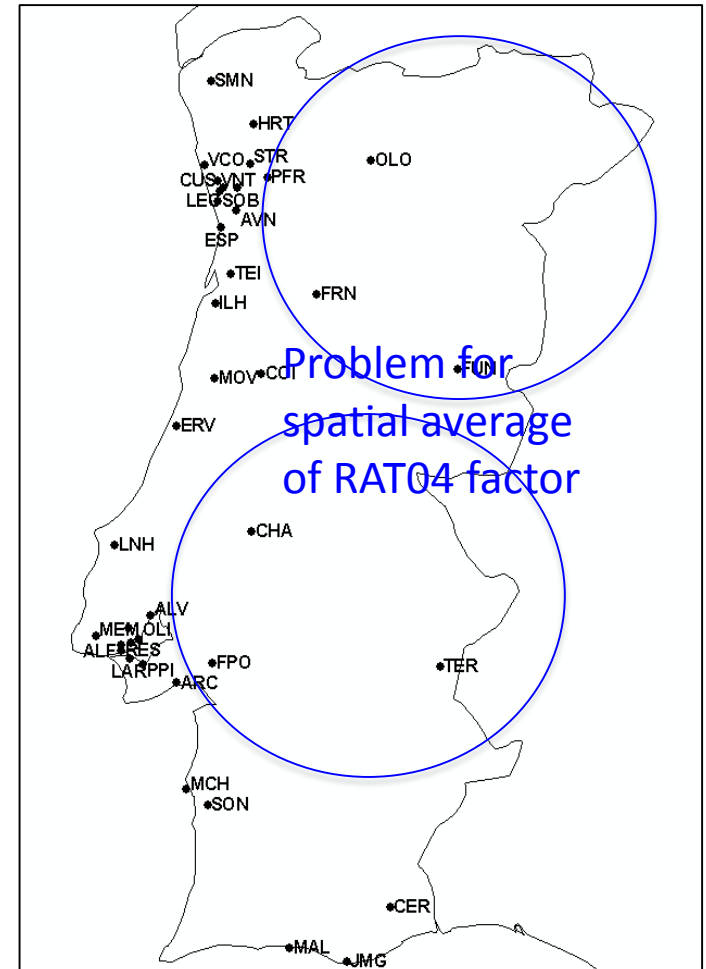
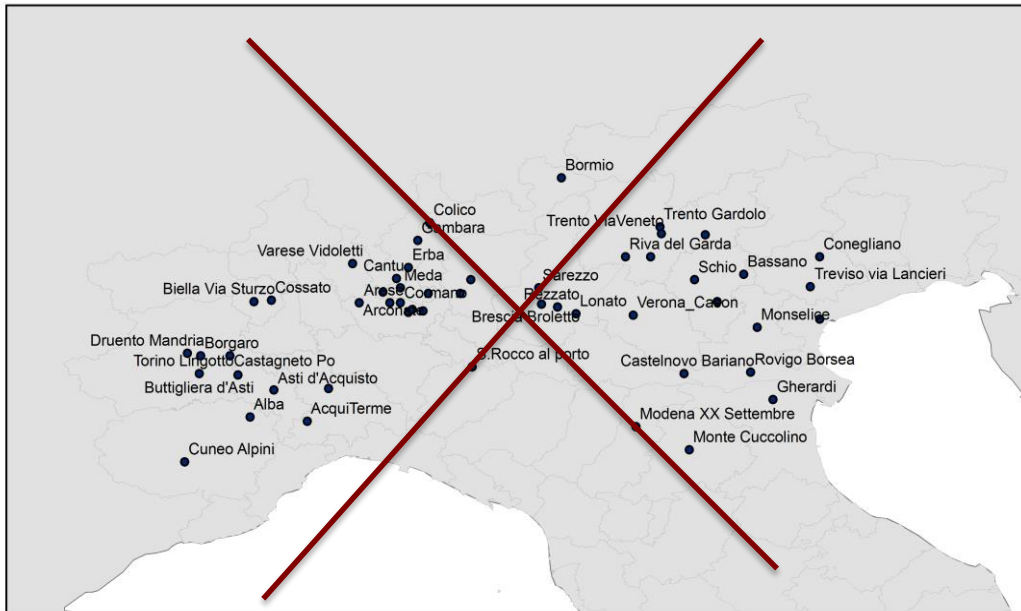
Calculate the RAT04 average factor (per hour) and apply it to all grid cells

How to validate?

based on a Monte Carlo approach

1. A set of **n Monte Carlo re-analyses** has to be done
 - a) For each one **randomly select 20% of the stations to be used as validation** stations (do not use them to perform the re-analysis)
 - b) Compute for each station i (at least) in each re-analysis j the **RMSE** (i,j)
2. Compute **for each station i the maximum of RMSE** (i,j). Let $vect_max(i)$ be the number of the re-analysis associated to the maximum RMSE for station i
3. Create a **CDF file to be used in the DELTAtool** by selecting for each station i the $vect_max(i)$
4. Use the Deltatool as if the CDF file was the CDF of a single model

Case study: modeling setup

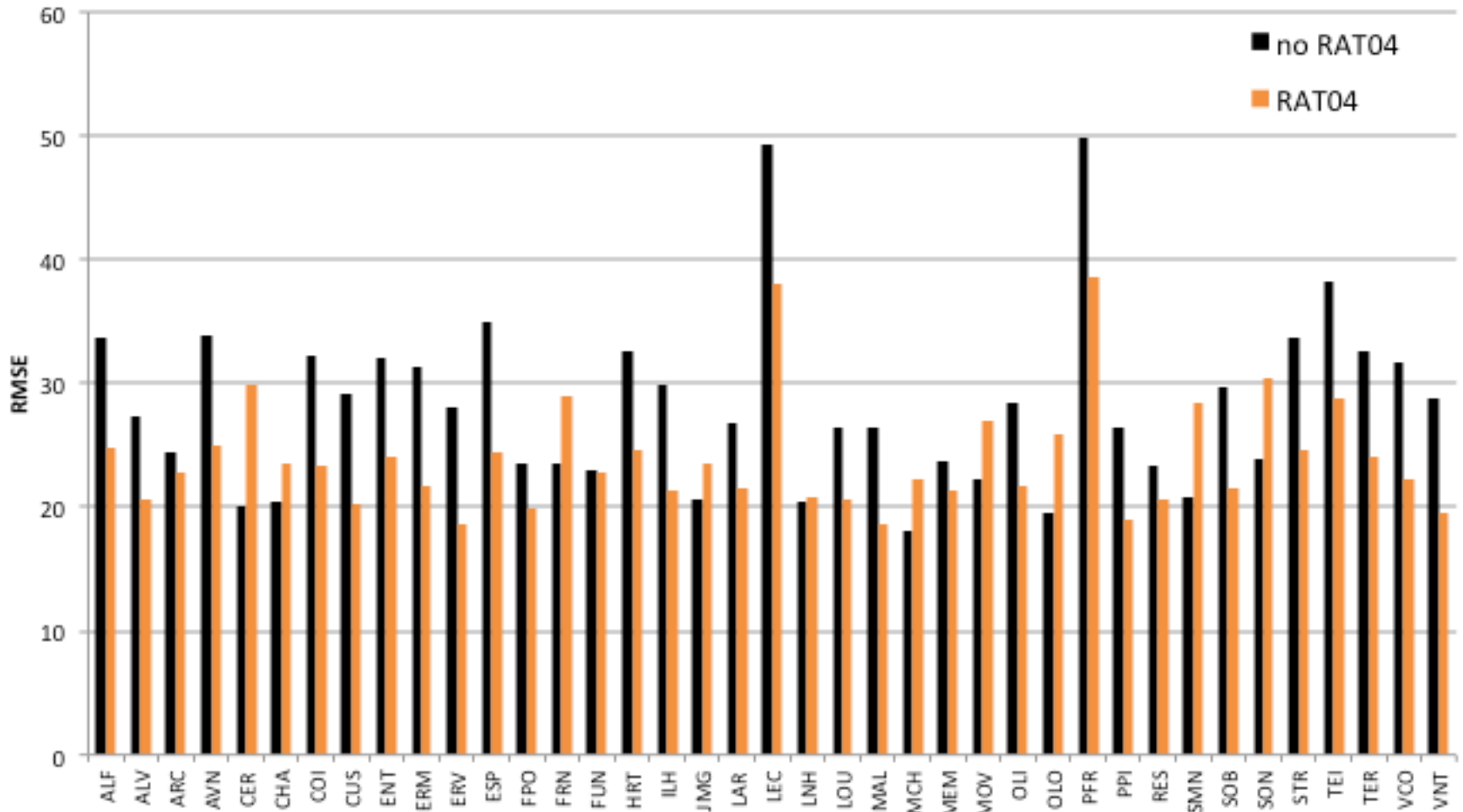


- Measures: ~~50~~³⁹ monitoring sites
- Model: ~~TCAM~~ EURAD
- Year: ~~2005~~ 2013
- Domain resolution: ~~6x6km²~~ (POMI exercise) 5x5 km²
- Pollutants: ~~PM10~~ O₃

Results of Monte Carlo approach

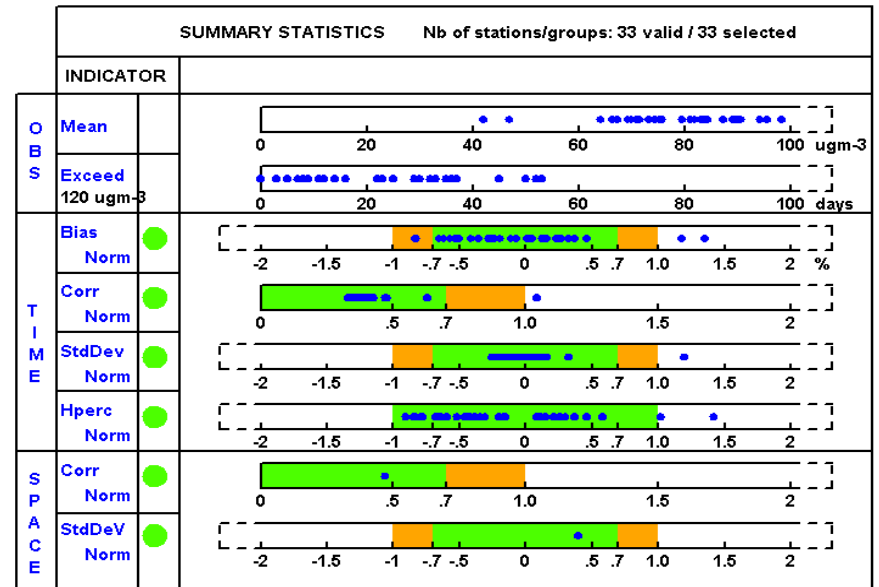
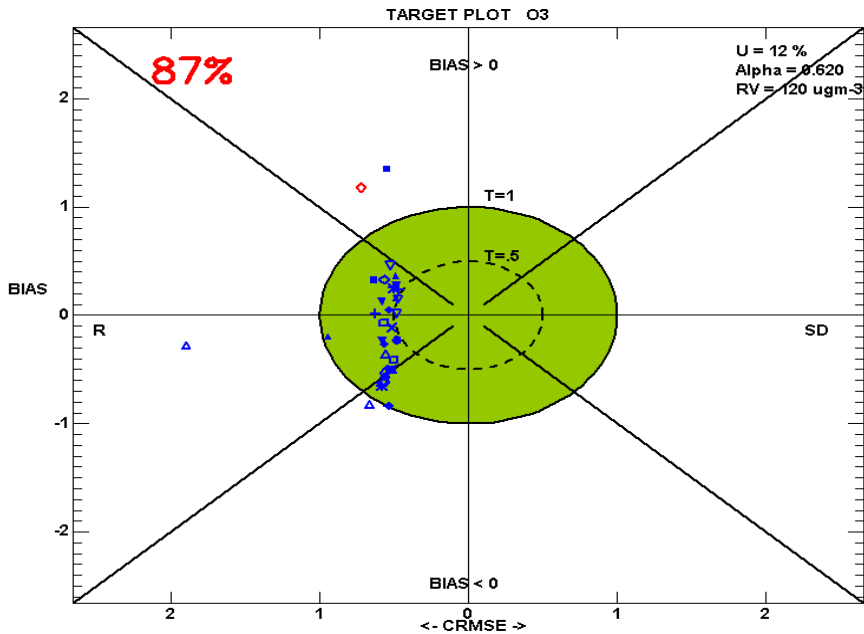
Stations/Re-analysis	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
ALF						24,8							24,1		24,0		22,7
ALV					20,6			20,5								20,7	
ARC			22,3			21,6			22,1	22,8		22,6		22,4		22,3	22,5
AVN						24,9	23,8							22,6		23,0	
CER				28,7					29,0							29,9	
CHA		23,1		21,9	23,4		22,0				23,6			23,2			
COI						23,3										22,0	
CUS			19,9					20,2									
ENT					23,3				24,0		23,3	23,5		23,4			
ERM	21,7		20,9					21,3									20,7
ERV												18,6					
ESP	23,2	22,3		23,6	21,8	24,4			23,1					22,2			
FPO					19,8			19,6		19,8		19,8		19,6			
FRN								27,8		29,0				28,4			
FUN		22,6	22,6		22,8											22,3	
HRT			24,0	24,6									24,6				
ILH	21,4							21,2				20,8					
JMG		22,7			23,0		22,1			23,5	23,3						
LAR							21,6										
LEC				38,0					37,7						38,1		
LNH		20,3	20,2	19,4						20,8							
LOU	20,7												20,7				
MAL	18,4						18,6			17,9					18,5		
MCH			21,2								22,3		20,3	21,5			21,7
MEM								20,6				21,2					21,3
MOV																	27,0
OLI		21,3									21,5		21,7		21,7		
OLO			26,0												24,7		
PFR							38,4						38,6				
PPI	19,1															18,9	
RES		20,5					20,4		20,3			20,7					
SMN	26,6										28,4	27,6	26,4		26,5		27,9
SOB				21,6			21,5										20,8
SON					29,8			28,8		30,4	30,2						
STR				24,6													
TEI		26,8				28,8		27,4	27,6		26,1				27,9		
TER	23,3					24,1										22,9	
VCO						22,3							21,3				
VNT										18,1					19,5		

Before and after RAT04



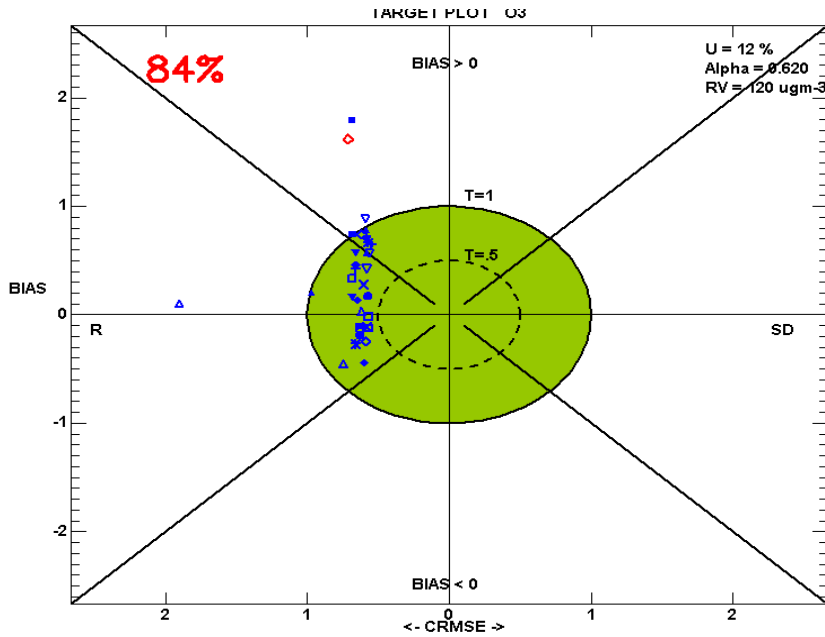
Using the DELTA Tool

Selecting for each station the worst possible timeseries

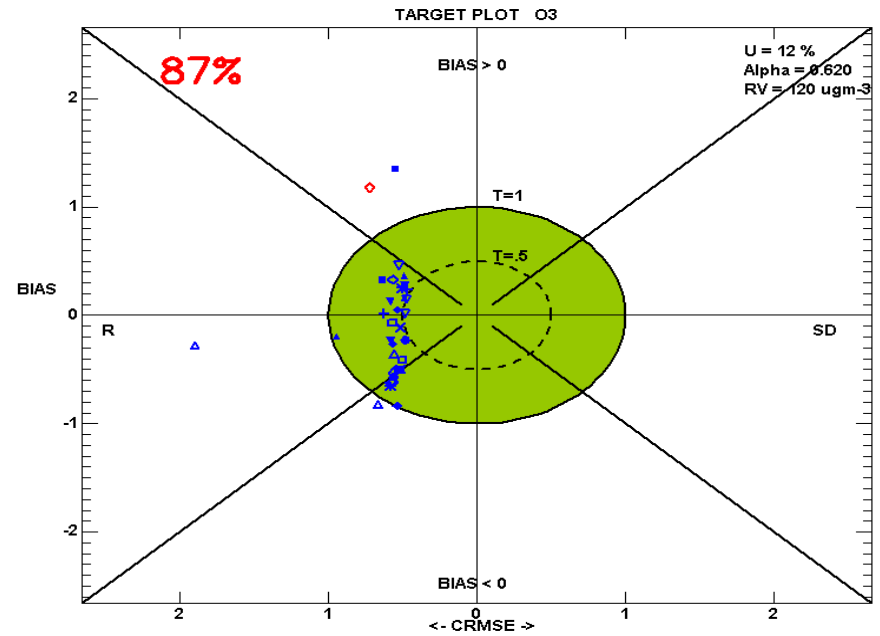


Using the DELTA Tool

BEFORE RAT04

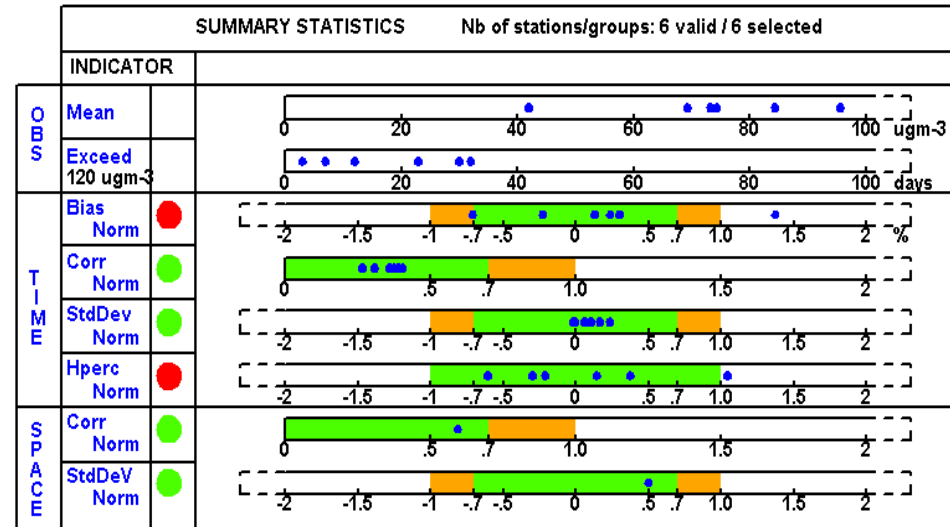
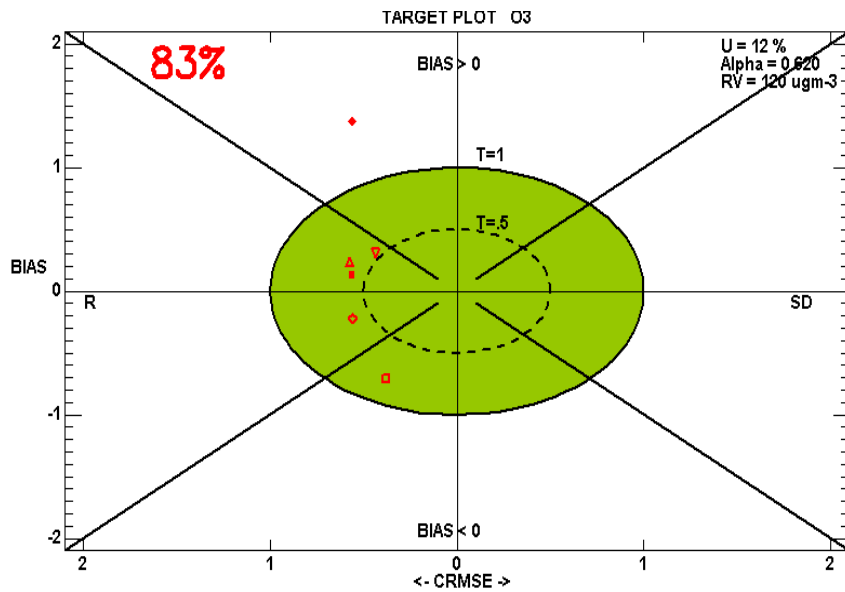


**AFTER RAT04 and
Monte Carlo approach**



Using the DELTA Tool

Selecting the interaction with worst average RMSE



Some comments

- 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”!!



**THANK
YOU
AND
ENJOY
AVEIRO**