GEMS RAQ
WP4.3

Statistical post-treatment of
forecast

- In **WP 4.1** skill scores for error characterisation are defined

$\Rightarrow$ from the hourly to daily basis

$\Rightarrow$ site by site (where surface stations are located)

- In **WP 4.3** we are interested in spatio-temporal correlations in model errors

$\Rightarrow$ Produce **analysis**

$\Rightarrow$ Prepare **ensemble modelling (WP4.4)**

*Prepare **data assimilation** (WP2.3) ?*

example of the PREV'AIR analysis
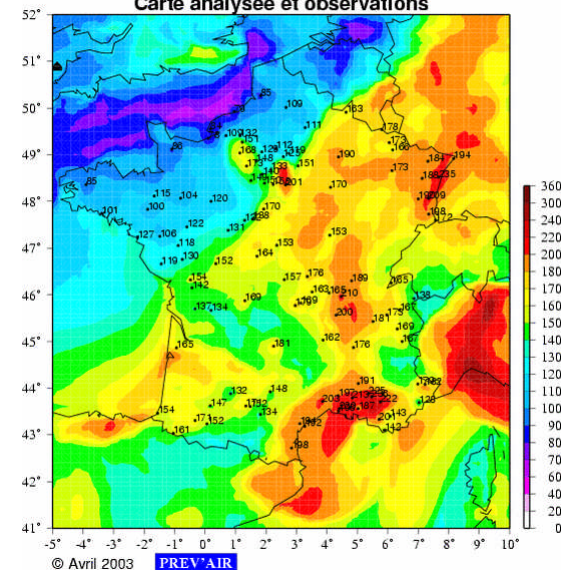
Analysis deals with D-1 and D+0 ozone peak

Model:

– CFM = France CTM, Gaseous

– Modelled peak ozone concentrations corrected with observations: **Kriging method** (daily updating, homogen, isotrope)
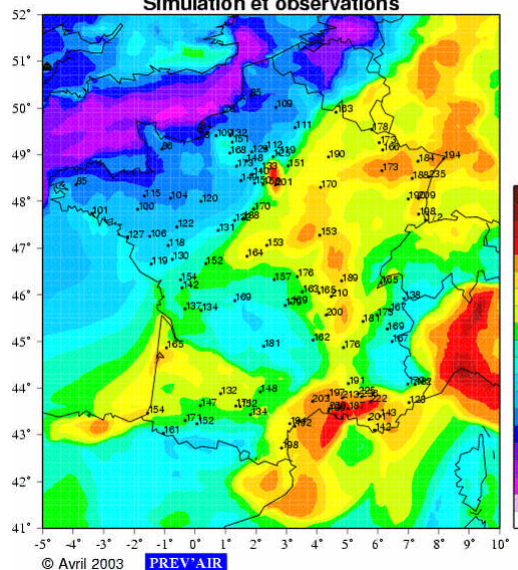
Outputs:

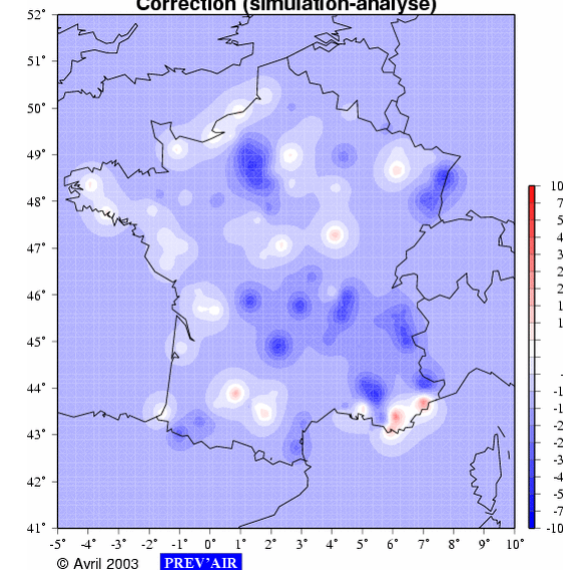– Ozone peak concentration
– Daily data

July 31, 2004

Ozone , pic en µg/m3, le 31/07/2004
Carte analysée et observations

© Avril 2003   PREV'AIR

Ozone , pic en µg/m3, le 31/07/2004
Simulation et observations

© Avril 2003   PREV'AIR

Ozone , pic en µg/m3, le 31/07/2004
Correction (simulation-analyse)

© Avril 2003   PREV'AIR

PM10 analysis : feasability study

- 15/06/2003-15/09/2003

- PM10 daily mean concentrations

- PM10 measurement stations:

  - Rural (green, 5)

  - Suburban (blue, 24 out of 34)

  - Urban (black, 40 out of 150)
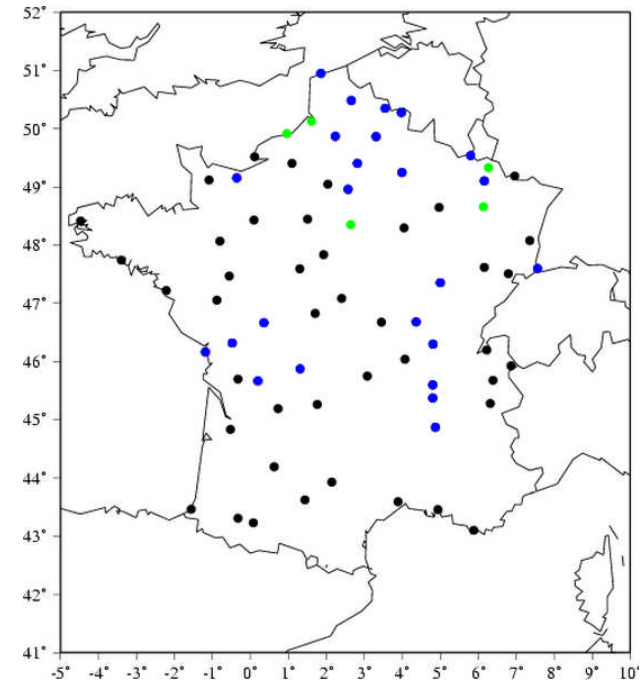
Station type	Mm	Nobs	BAa	BAm	RMSEa	RMSEm	Corra	Corrm	
Rural	14.1	365	-0.1	-8.3	9.2	12.9	0.65	0.57	Cross validation
Suburban	13.7	2148	1.2	-6.8	6.1	10.2	0.81	0.66	
Urban	12.0	3540	-0.8	-9.3	7.1	12.4	0.75	0.67	
Suburban	14.0	1714	1.6	-4.8	7.7	9.1	0.69	0.62	Verification
Urban	13.9	9682	0.0	-9.0	5.9	12.0	0.84	0.68	

What is our goal in GEMS framework for WP 4.3?

→ To build a common analysis tool ?

How to build analysis ?

Merge observations and simulations using **Krigging** methods :

→ Correction of simulation by innovations

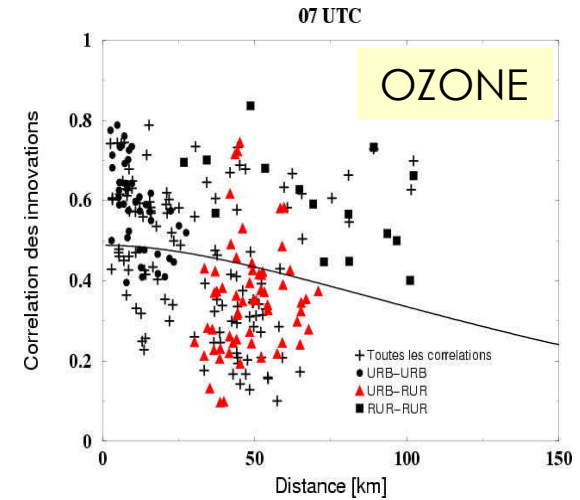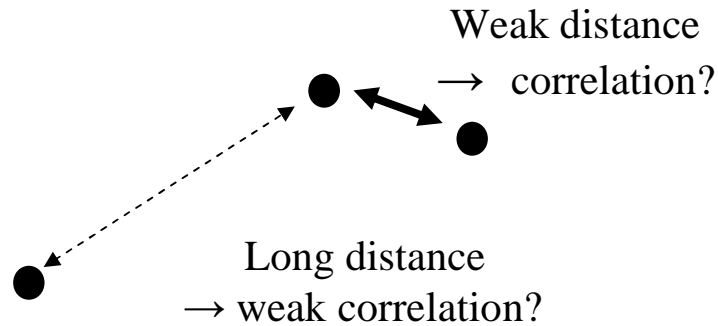→ need to know the **model error covariance matrix (B)**

main problem: how to assess it ?

→ need to model the **error covariance matrix**

Modelling spatial error correlation: an example

• Spatial error correlation= f(distance between points) ?

Weak distance
→ correlation?

Long distance
→ weak correlation?

• Spatial error correlation = f(concentration's correlation) ?

Weak correlation

High correlation

emission Source

Valley
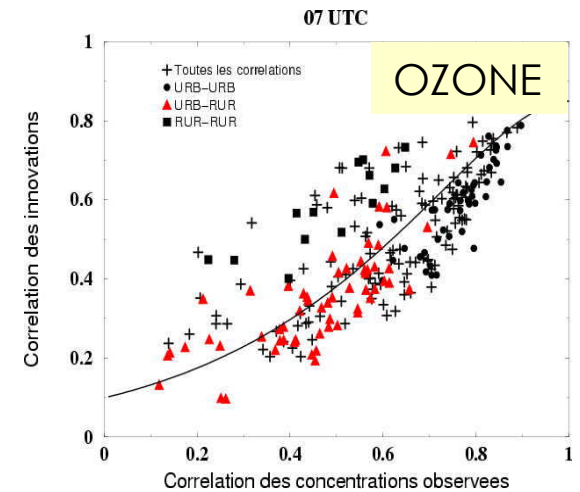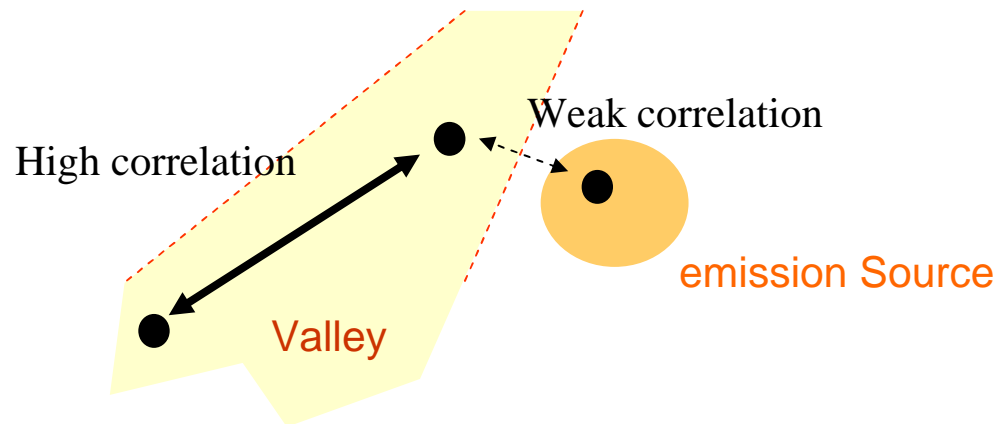
Some thoughts about error characterisation in regional air quality models

How are model errors on surface ozone spatially (2D) correlated (other species)?

How are they correlated in the vertical ?

How are they correlated temporally, i.e. is there a correlation length in time

How are they correlated between different models ?
→ Ensemble modelling (**WP4.4**)

How to assess it?

Which datas for which methodology?

(2) Spatial correlation of errors in models: horizontally

Ground stations: (**Ozone, others …??**)

Which stations?   EMEP, Air quality networks, others… (link to WP4.1)

- **availibility** (autorisation, timing: real-time or not)

- **number** of stations and horizontal **repartition** of stations
  - → some regions could be over-represented some other under-represented
    (clustering effects …)

- **quality** of datas (differences between validated and near real time datas)

- **representativity** of stations: problems of status (urban, peri-urban, …), model
  resolution and representativity.

(3) Spatial correlation of errors in models: horizontally

Satellite : (**Ozone**)

What is available? *GOME, SCHIAMACHY, TESS, IASI, OMI ?*

What are the problems? No sensivity to surface concentrations → dead-end?

Other species: **NO$_2$**   →   Useful for source inversion

Probably to early → more direct applications for inversion and assimilation and for large-scale models.

(4) Spatial correlation of errors in models: horizontally

Methodological concerns :

How to build error correlation (technically speaking with available datas)?

Identification (determination) of most important correlation criterium

$\rightarrow$ distances? concentrations?

(c.f. example *Blond et Vautard, JGR,2004*)

what are appropriate technics: Krigging (others) ? , variograms?

(4) Spatial correlation of errors in models: vertically

Purpose: *caracterisation of fully 3D model error*

Which datas?

- Use MOZAIC vertical profiles (O3, CO) to assess vertical correlations of model errors
  - → is ozone error at different altitudes in free troposphere , in the boundary layer correlated (i.e. PBL height should be included in analysis?)

Planned work:
- Evaluate possibility of error correlation at different sounding sites, but only few are available ….
  Definition of a well documented (does it exist?) area for such study.

  → *Important to prepare **assimilation** of lower tropospheric ozone data*

  What is the lowest information we could get from satellite (for O_3)?

   Is it relevant? → study the impact of free trop. O3 on the budget of

   surface concentrations

Temporal correlation of errors in models

What is governing it?

→ depending on time scale of synoptical situation

→ depending on seasonal characteristics
(boundary conditions, surface characteristics ….)
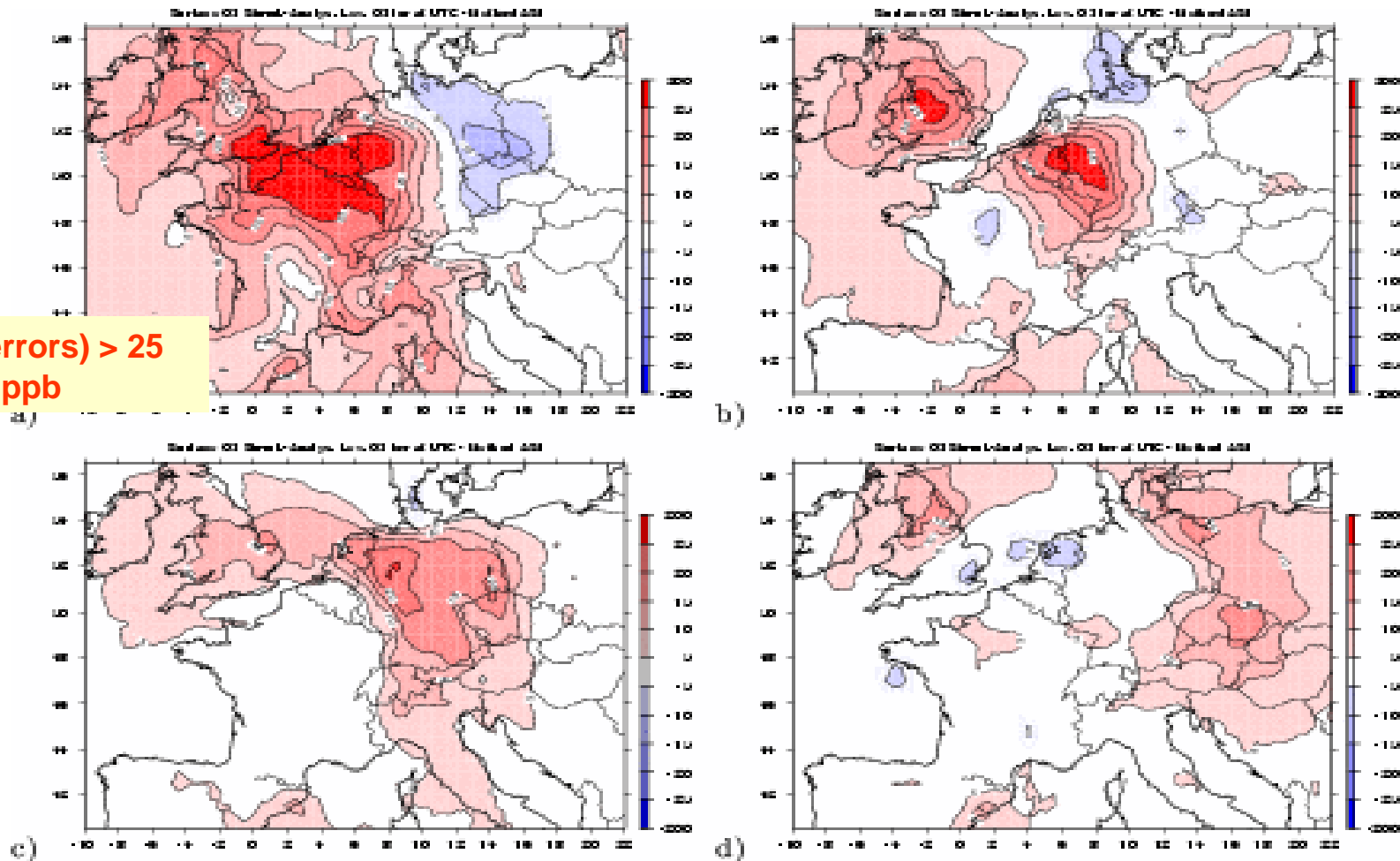
Is it necessary ?

Importance of the error propagation with time ?

→ Prblms of the technics → OI, Kalman, ensemble

Example of forecast error propagation

Chimere forecast for an episode running from the 28 of July to the 1st of August 2002

max (errors) > 25 ppb

Courtesy : Nadège Blond (LVI, Strasbourg)

Model error cross correlation (1)
(introduction to ensemble modelling WP4.4)

What could be sources of model error cross correlation ?
Can we identify sources? Is it meaningfull?

Errors mostly driven by forcings ($\neq$ meteorological chaotic system)
 $\rightarrow$ are differences between models = differences between
Emissions? meteorological forcing? chemical forcings? others …

Is there appropriate experiences to address these questions?
$\rightarrow$ experiences of passiv tracers
$\rightarrow$ experiences with common emissions
$\rightarrow$ experiences with same climatological forcings

What can be drawn from that?
$\rightarrow$ identification of common biases
$\rightarrow$ building « mean » or « best » forecast? $\rightarrow$ *Ensemble modelling*

Some more thoughts about ensemble modelling …

Different model forecasts are merged by giving them a weight $\sim 1 / \sigma^2$.
But, how to assess σ^2 ? … *back to observations …*

- ***on which timescale : for past season or daily update ?***

$\rightarrow$ there, study on error correlation length (in time) is needed.
Is error advected with time ? $\rightarrow$ need the use of Kalman filtering technics
(i.e transport of error covariance matrix)? or can the problem be simplified after
analysis time scale length of the error?

- ***on which spatial scale : same prblms previously mentionned***

$\rightarrow$ model errors will vary in space (and moreover with the horizontal resolution) but
limited and non-homogeneous informations are provided by surface stations
and satellite are of limited use for the surface. Can we, at least define a well
documented area of study?

how to formally assess this question ?

these are only some thoughts
to trigger discussion ….