Using ensemble modeling to improve particulate matter forecasting in France

E. Debry[1], V. Mallet[2], F. Meleux[1], B. Bessagnet[1], L. Rouïl[1]

[1] Institut National de l'EnviRonnement industriel et des rISques (INERIS). Verneuil-en-Halatte France.

[2] Institut National de Recherche en Informatique et Automatique (INRIA). Le Chesnay, France.

International Workshop on Air Quality Forecasting Research
2-3 december 2009, Boulder, Colorado

INERIS
maîtriser le risque
pour un développement durable

INRIA

PREV'AIR

Outline

1. Ensemble modeling
2. Particle modeling
3. Applying ensemble to particles
   - Setup
   - Rank diagram
   - Weight computation
   - Results
4. A few words to conclude

Ensemble modeling

Why using ensemble modeling ?

Several dispersion models depending on

- Input fields (meteo, emissions) and parameters (deposition).
- Physical modeling (chemistry, phase transfer).
- Space and time resolution.
- Numerical schemes.

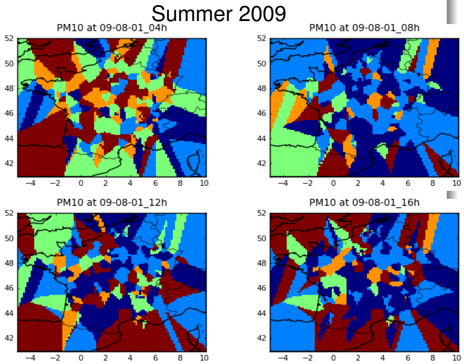
Because. . .

Each model may have a significant contribution, with best predictions over given areas, during given time periods, and for given pollutants.

Ensemble modeling

PM_{10} ensemble space and time variability.

- Ensemble composed with 5 models.
- Each model is assigned one color.
- Map painted at a given time and location with color of the model closest to observation (RMSE).



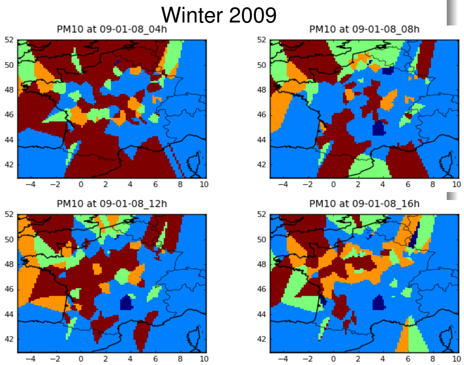
Ensemble modeling principle

PM_{10} predictions can be improved by sequential aggregation of models.

Ensemble modeling

PM_{10} ensemble space and time variability.

- Ensemble composed with 5 models.
- Each model is assigned one color.
- Map painted at a given time and location with color of the model closest to observation (RMSE).



Ensemble modeling principle

PM_{10} predictions can be improved by sequential aggregation of models.

Particle modeling

Uncertainties

Significant uncertainties remain in the particle module as opposed to gas chemistry.

Particle modules

Currently in modules :

- Internal mixing assumption.
- Size-resolved, at most 8 bins.
- Coagulation, condensation/evaporation.
- Nucleation H_2O-H_2SO_4.
- Inorganic chemistry (Isorropia).
- SOA scheme (surrogate components).

Not in modules :

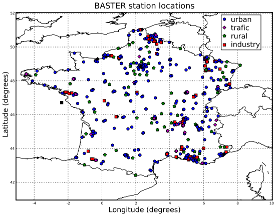
- External mixing.
- Detailed description of nano-particles.
- Nucleation pathways (organic, NH_3).
- Condensation/evaporation pathways.
- Inorganic/organic interactions.
- Organic precursors.

Models

- CHIMERE :
  - GFS & ARPEGE meteorological forcing.
  - French & European domains.
- Polair3d on European domain.

Observations

BASTER network, 620 stations



Periods

- Summer : from 2009-04-30 to 2009-09-15.
- Winter : from 2008-11-01 to 2009-02-28.

Sequential aggregation

Build a linear combination (EM) with several models M^m :

$$EM_{t,x} = \sum_{m=1}^{n} \alpha_t^m M_{t,x}^m$$

weighted (α_t^m) according to past predictions ($M_{t'<t,x}$) and observations ($O_{t'<t,x}$).

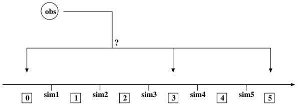
Applying ensemble to particles

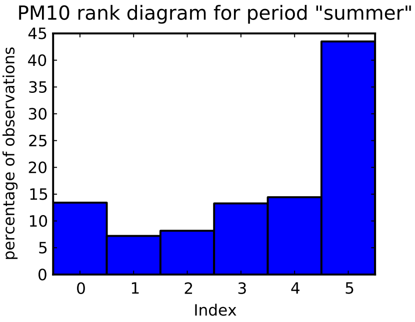
How well is my ensemble ?

- Prediction improvements depends more on the ensemble quality than on the “quality” of each models.
- Ensemble should be representative of the uncertainty range.
- Depend on your target (RMSE, threshold detection).

Rank diagram.

How observations are distributed with respect to simulations ?





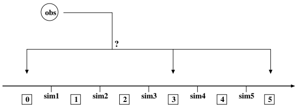
Applying ensemble to particles

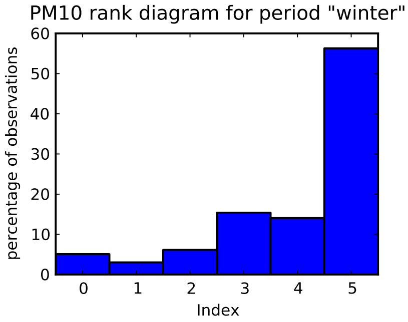
How well is my ensemble ?

- Prediction improvements depends more on the ensemble quality than on the "quality" of each models.
- Ensemble should be representative of the uncertainty range.
- Depend on your target (RMSE, threshold detection).

Rank diagram.

How observations are distributed with respect to simulations ?





Various methods

- Median : weights all set to $1/N$.
- Ridge Regression method (RRD) (Mallet et al. [2009]) :

$$\alpha_{\mathbf{t}} = \underset{\alpha \in \mathbb{R}^n}{\arg\min} \left[\lambda ||\alpha||_2^2 + \sum_{t'=1}^{t-1} \beta_{t-t'} \sum_{s \in \mathcal{N}_{\sqcup'}} \left(\alpha.\mathbf{M}_{\mathbf{t}',\mathbf{s}} - \mathbf{O}_{\mathbf{t}',\mathbf{s}} \right)^2 \right]$$

  - Past learning period windowed or discounted by $\beta_{t-t'}$.
  - Comes to a Least Square (ELS) method if $\lambda = 0$ and $\beta = 1$.
  - Theoretical guarantee to compete with best linear combination constant in time (Const ELS) of models.
- Exponential gradient method (EG) (Mallet et al. [2009]) :

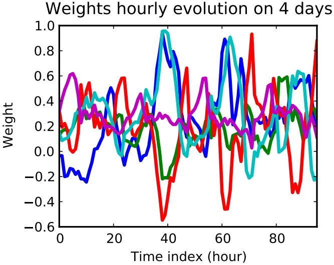
$$\alpha_{\mathbf{t}}^{\mathbf{m}} = \frac{\exp(-\eta \sum_{t'=1}^{t-1} L_m^{t'})}{\sum_{j=1}^{N} \exp(-\eta \sum_{t'=1}^{t-1} L_j^{t'})}, \quad L_m^{t'} = \sum_{s \in \mathcal{N}_{\sqcup'}} 2(\alpha_{t'}.\mathbf{M}_{\mathbf{t}',\mathbf{s}} - \mathbf{O}_{\mathbf{t}',\mathbf{s}})\mathbf{M}_{\mathbf{t}',\mathbf{s}}^{\mathbf{m}}$$

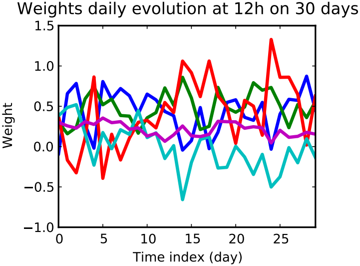
  - Learning rate η.
  - Weights $\in [0,1]$ and $\sum = 1$. Convex combinations.

Applying ensemble to particles

Weight computation

Weight time evolution for RRD method on summer period





Applying ensemble to particles

Results

Summer

model	RMSE ($\mu g/m^3$)	Corr	FB
AFM	14.9	0.3	0.83
AFMA	16.2	0.2	1.13
AWM	15.4	0.29	0.75
AWMA	15.9	0.18	1.05
Polair3d	15.0	0.28	0.94
MEDIAN	14.3	0.29	0.94
Const ELS	13.0	0.37	1.07
EG	13.7	0.34	0.96
RRD	12.7	0.39	1.11

Winter

model	RMSE ($\mu g/m^3$)	Corr	FB
AFM	28.4	0.39	0.56
AFMA	27.2	0.28	0.92
AWM	28.6	0.41	0.53
AWMA	27.5	0.28	0.86
Polair3d	28.1	0.23	0.78
MEDIAN	26.4	0.38	0.73
Const ELS	23.6	0.41	1.07
EG	26.2	0.34	0.83
RRD	22.9	0.46	1.09

RRD improvements with respect to best model

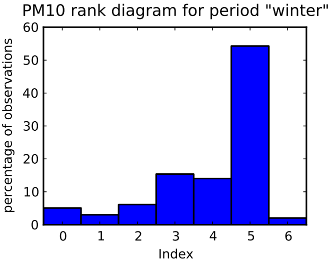
	RMSE ($\mu g/m^3$)	percentage
Summer	-2.2	-14.8%
Winter	-4.3	-15.7%

Correlation also improved, but Bias factor may be inversed.

How to improve the ensemble ?

- Adding a constant and uniform model at $100\mu g.m^{-3}$!!

model	RMSE ($\mu g/m^3$)	Corr	FB
AFM	28.4	0.39	0.56
AFMA	27.2	0.28	0.92
Const	72.3	nan	5.2
AWM	28.6	0.41	0.53
AWMA	27.5	0.28	0.86
Polair3d	28.1	0.23	0.78
MEDIAN	22.4	0.38	1.47
Const ELS	21.8	0.44	1.4
RRD	20.9	0.51	1.37



RMSE improved by 23.3% ($-6.3\mu g/m^3$).

- Compute the ensemble at regional scale and per station.

A few words to conclude

Result summary

- Improved RMSE and correlation with respect to observations.
- Reversed bias factor depending of the ensemble method.
- Operational on the French Prev'Air system.

Questions and future works

- How reliable is the observation ?
- Network optimization : do we need all stations ?
- Suitable ensemble methods for adressing threshold detection.

Thank you for your attention.

Any questions ?