Combining observations and ensemble air-quality forecasts

Vivien Mallet (speaker), Bruno Sportisse

Vivien.Mallet@cerea.enpc.fr

École Nationale des Ponts et Chaussées

CEREA (Teaching and Research Center in Atmospheric Environment, ENPC / EDF R&D)

CLIME project (INRIA)

Outline

Uncertainties in chemistry-transport models

- Numerical approximations
- Input data
- Model formulation

Ensemble forecast

Combining ensemble members

Forecasting linear combinations

Uncertainties in CTM

Chemistry-transport models

$$\frac{\partial c_i}{\partial t} = \underbrace{-\mathrm{div}(V c_i)}_{\text{advection}} + \underbrace{\mathrm{div}\left(\rho K \nabla \frac{c_i}{\rho}\right)}_{\text{diffusion}} + \underbrace{\chi_i(c)}_{\text{chemistry}} + \underbrace{S_i - L_i}_{\text{sources and losses}}$$

Uncertainty

- Statistical model $Y = F(X)$
- Given data x and given model f: $y = f(x)$
- Error: discrepancy between y and observations
- Uncertainty: spread of Y, e.g. σ_Y

Uncertainty sources

- Numerical schemes, model formulation, input data

Uncertainties in CTM

Numerical schemes

- "Agreement coefficient": relative differences below 5%

Comparison	$\Delta(\mathrm{O}_3)$
$\Delta t = 600\ \mathrm{s}$ / $\Delta t = 1800\ \mathrm{s}$	54.7
Reference / first order upwind (advection)	66.0
$K_h = 10\,000\ \mathrm{m}^2 \cdot \mathrm{s}^{-1}$ / $K_h = 50\,000\ \mathrm{m}^2 \cdot \mathrm{s}^{-1}$	80.0
$\Delta t = 600\ \mathrm{s}$ / $\Delta t = 30\ \mathrm{s}$	96.4

- Low sensitivity

Pourchet, A., Mallet, V., Quélo, D., and Sportisse, B. (2005). Some numerical issues in Chemistry-Transport Models – a comprehensive study with the Polyphemus/Polair3D platform. *In preparation for J. Comp. Phys.*

Uncertainties in CTM

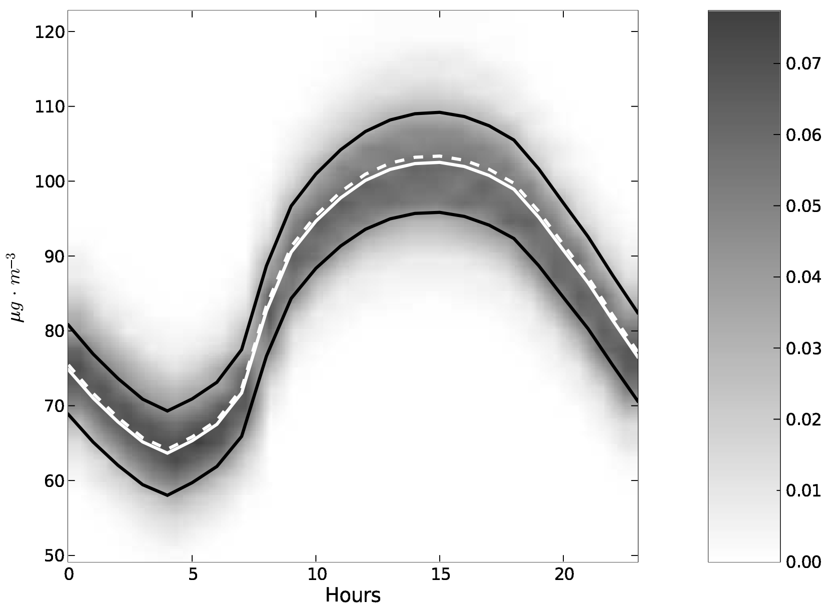
Input data

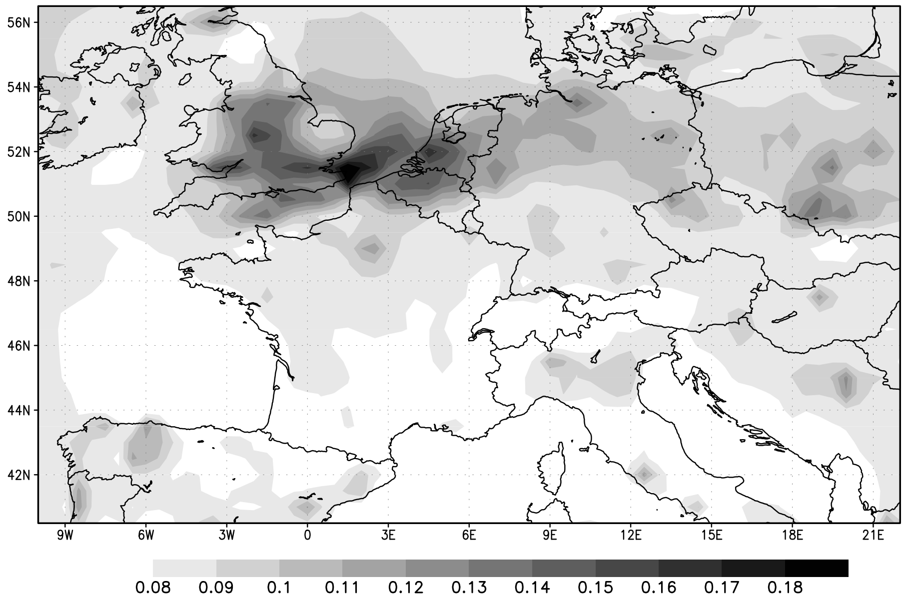
- Monte Carlo simulations (800 runs)
- Most input data except chemical reaction rates and meteorological fields

Input data	Uncertainty (LN)
Cloud attenuation	$\pm 30\%$
Deposition velocities (O_3 and NO_2)	$\pm 30\%$
Boundary conditions (O_3)	$\pm 20\%$
Anthropogenic emissions	$\pm 50\%$
Biogenic emissions	$\pm 100\%$
Photolysis rates	$\pm 30\%$

Uncertainties in CTM

Input data





- Uncertainty of about 7–8% (standard deviation, lower bound)

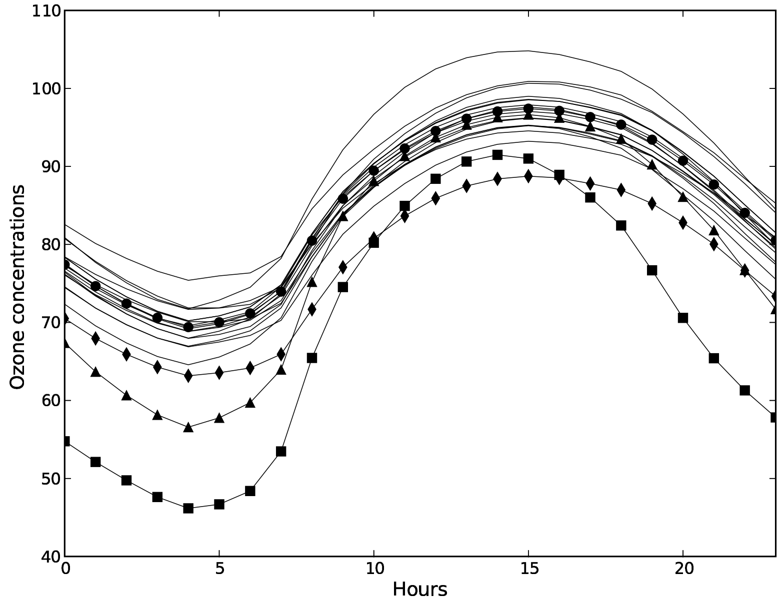
Uncertainties in CTM

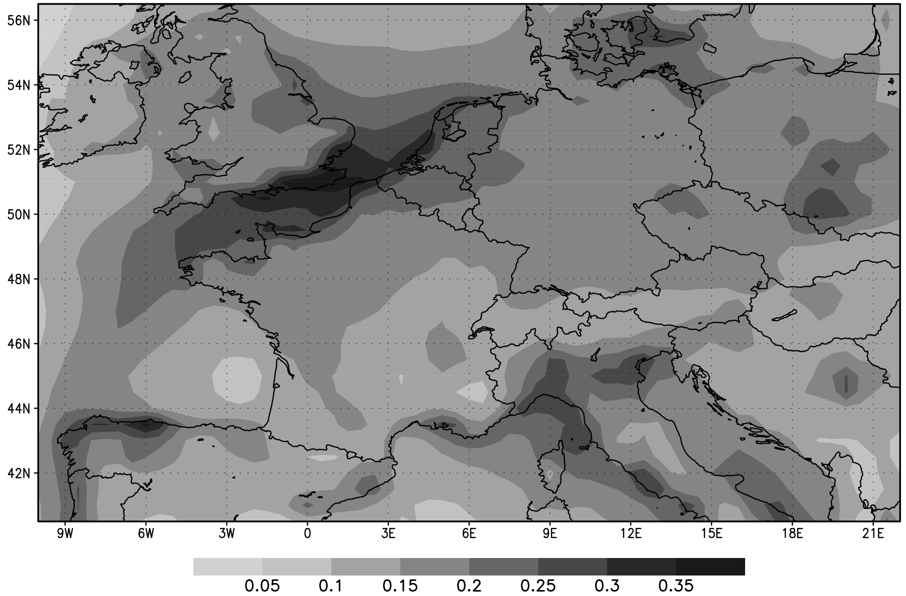
Multi-models approach (model formulation)

#	Parameterization	Reference	Alternative(s)
		Physical parameterizations	
1.	Chemistry	RACM	RADM 2
2.	Vertical diffusion	Troen & Mahrt	Louis
3.			Louis in stable conditions
4.	Deposition velocities	Zhang	Wesely
5.	Surface flux	Heat flux	Momentum flux
6.	Cloud attenuation	RADM method	Esquif
7.	Critical relative humidity	Depends on σ	Two layers
		Input data	
8.	Emissions vertical distribution	All in the first cell	All in the two first cells
9.	Land use coverage (dep.)	USGS	GLCF
10.	Land use coverage (bio.)	USGS	GLCF
11.	Exponent p in Troen & Mahrt	2	3
12.	Photolytic constants	JPROC	Depends on zenith angle
		Numerical issues	
13.	Time Step	600 s	100 s
14.			1800 s
15.	Splitting method	First order	Strang splitting
16.	Horizontal resolution	0.5°	0.1°
17.			1.0°
18.	Vertical resolution	5 layers	9 layers
19.	First layer height	50 m	40 m

Uncertainties in CTM

Physical parameterizations, approximations, input data





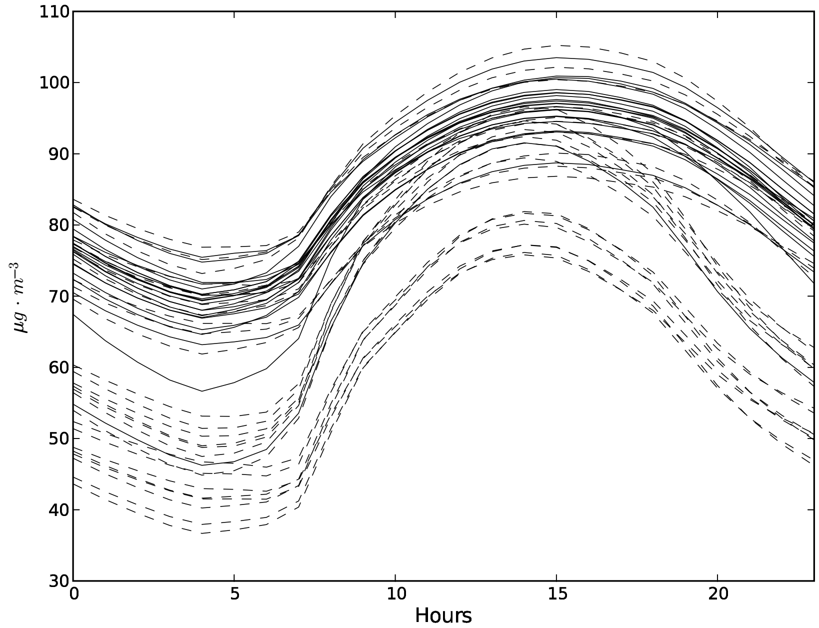
- Uncertainty of about 6–7% (single changes)
- Uncertainty of about 16–17% (multiple changes)

Mallet, V. and Sportisse, B. (2005b). Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: an ensemble approach applied to ozone modeling. *To appear in J. Geophys. Res.*

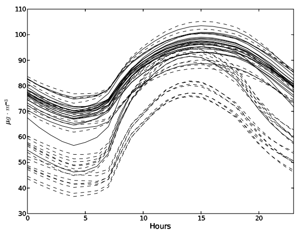
Combining models

Ensemble forecast

- First ensemble: 22 members, single changes
- Second ensemble: 48 simulations, multiple changes



Combining models



Ensemble forecast

- First ensemble: 22 members, single changes
- Second ensemble: 48 simulations, multiple changes

Purpose

- Minimize the root mean square error (RMSE)
$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - o_i\right)^2}$$
- Beat the best (tuned) model for forecasts, with a decrease by 10% of RMSE on ozone concentrations
- Experiment: 4 months in summer 2001, over Europe, $\sim$2100 cells (first layer) and about 100 stations
- Based on about 240,000 hourly observations

Combining models

Notations

- Ensemble $\mathcal{E}$
- Model output $M_{t,x}$ or $M_{m,t,x}$ (model #m)
- Time average $\overline{M}_x^t$; spatial average $\overline{M}_t^x$; average $\overline{M}^{t,x}$
- Observations $O_{t,x}$
- Cardinal: $|\cdot|$

Ensemble mean and median

$$\mathrm{EM}_{t,x} = \frac{1}{|\mathcal{E}|} \sum_{M \in \mathcal{E}} M_{t,x}$$

$$\mathrm{EMD}_{t,x} = \mathrm{median}(\{M_{t,x}\}_{M \in \mathcal{E}})$$

Combining models

Combinations based on least squares

- $\mathrm{ELS}_{t,x} = \sum_m \alpha_m M_{m,t,x}$
  where α minimizes $\sum_{t,x} \left[O_{t,x} - \sum_m \alpha_m M_{m,t,x}\right]^2$
- $\mathrm{EULS}_{t,x} = \overline{O}^{t,x} + \sum_m \alpha_m \left(M_{m,t,x} - \overline{M}_m^{t,x}\right)$
  where α minimizes
  $\sum_{t,x} \left[O_{t,x} - \overline{O}^{t,x} - \sum_m \alpha_m \left(M_{m,t,x} - \overline{M}_m^{t,x}\right)\right]^2$
  also called superensemble in Krishnamurti et al. (2000)
- $\mathrm{EULS}^{\mathrm{s}}_{t,x} = \overline{O}^{t}_x + \sum_m \alpha^{\mathrm{s}}_{m,x} \left(M_{m,t,x} - \overline{M}^{t}_{m,x}\right)$ where
  $\alpha^{\mathrm{s}}_x = (\alpha^{\mathrm{s}}_{1,x}, \alpha^{\mathrm{s}}_{2,x}, \alpha^{\mathrm{s}}_{3,x}, \ldots)$ minimizes
  $\sum_t \left[O_{t,x} - \overline{O}^{x}_t - \sum_m \alpha^{s}_{m,x} \left(M_{m,t,x} - \overline{M}^{t}_{m,x}\right)\right]^2$

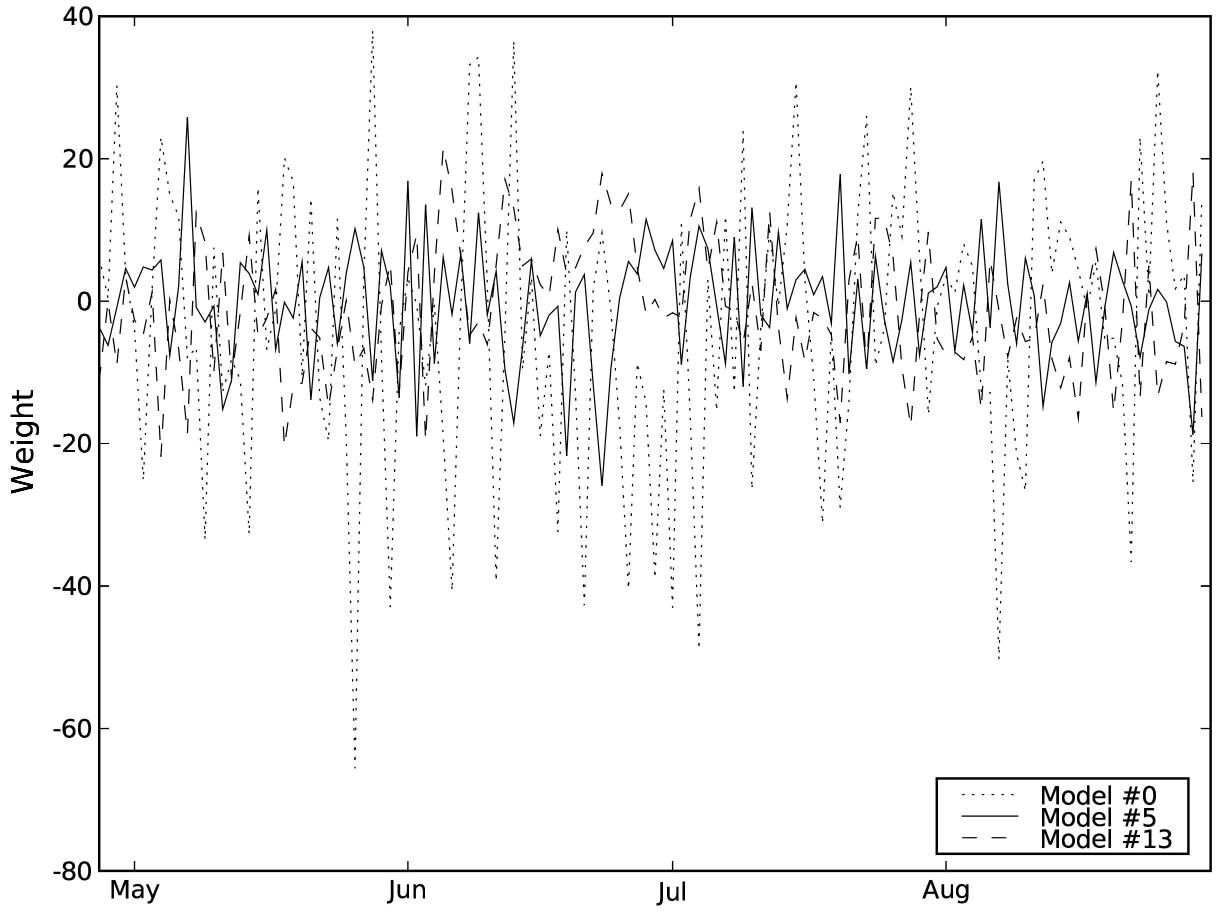
Combining models

Potentials (RMSE)

Combination	Hourly	Peak
Best model	25.7	21.5
EM	25.9	22.0
EMD	26.4	22.1
ELS	23.7	18.7
EULS	23.4	18.5
ELS$^{\mathrm{s}}$	16.4	12.9
EULS$^{\mathrm{s}}$	16.0	12.5
ELS$^{\mathrm{d}}$	17.1	12.5
EULS$^{\mathrm{d}}$	16.7	12.1

Combining models

Weights α *for* $\mathrm{ELS^d}$



Combining models

Weights computed over a 30-day learning period

At each date, weights are computed on the basis of observations at all stations and during the 30 previous days:

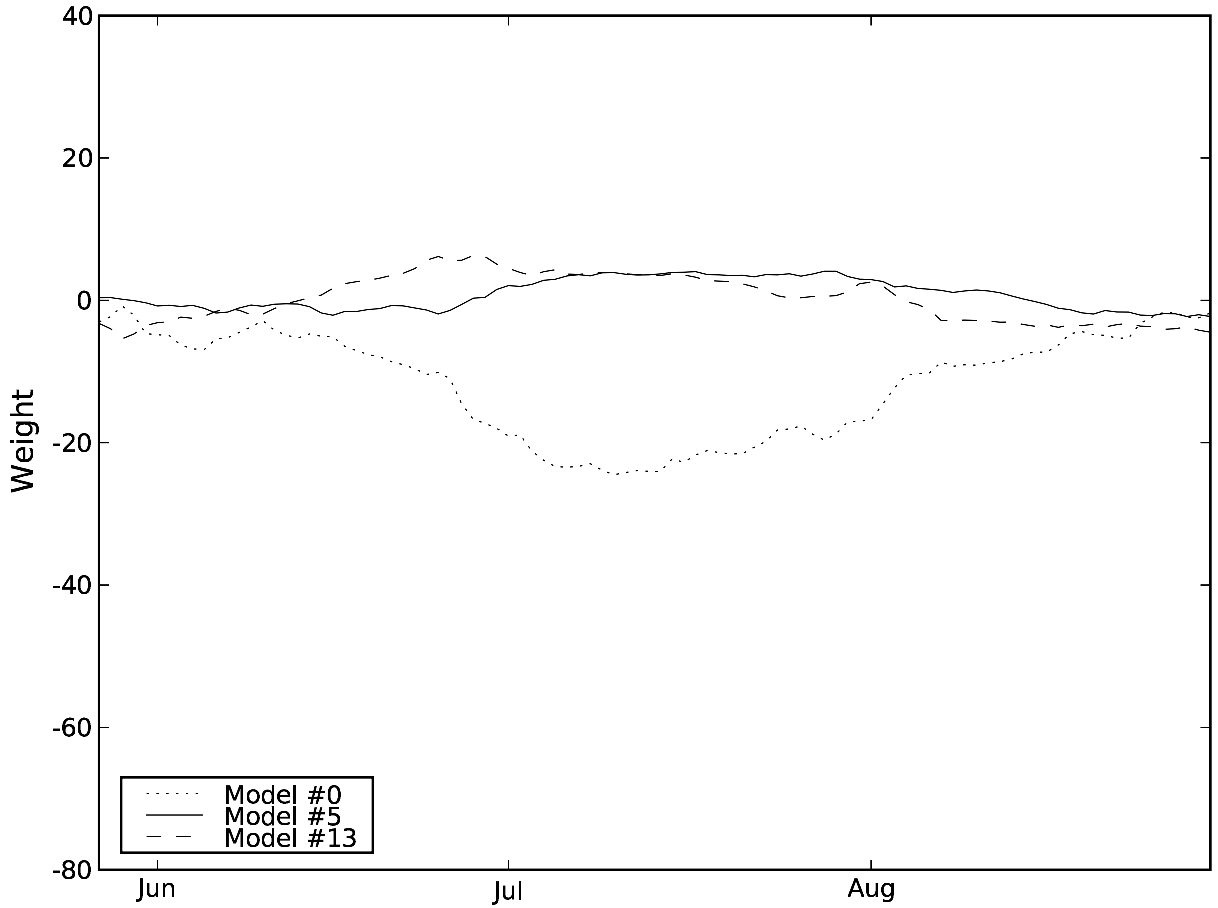
$$\mathrm{ELS}_{T,x}^{\mathrm{d},30} = \sum_m \alpha_{m,T}^{30} M_{m,T,x}$$

where $\alpha_{m,T}^{30} = (\alpha_{1,T}^{30}, \alpha_{2,T}^{30}, \alpha_{3,T}^{30}, \ldots)$ minimizes

$$\sum_{t=T-30}^{t=T-1} \sum_x \left[O_{t,x} - \sum_m \alpha_{m,x}^{30} M_{m,t,x} \right]^2$$

Combining models

Weights computed over a 30-day learning period



Combining models

Results: 22 members

Combination	Hourly	Peak
Best model	25.9	21.9
$\mathrm{ELS}^{\mathrm{d},30}$	23.6	19.2
ELS	23.9	18.7
$\mathrm{ELS}^{\mathrm{d}}$	17.3	12.8

Results: 48 members, BDQA monitoring network

Best model	28.5	23.9
$\mathrm{ELS}^{\mathrm{d},30}$	22.8	21.2
ELS	22.9	20.2
$\mathrm{ELS}^{\mathrm{d}}$	15.3	12.4

Combining models

Learning algorithm: gradient descent (Cesa-Bianchi et al., 1996)

$$L_t(\alpha_t) = \left(\sum_m \alpha_{m,t} M_{m,t} - O_t \right)^2$$

Weights $\alpha_{t-1} = (\alpha_{1,t-1}, \alpha_{2,t-1}, \alpha_{3,t-1}, \ldots)$ update:

$$\alpha_t = \alpha_{t-1} - \eta L'_{t-1}(\alpha_{t-1})$$

Network	Best model	$\mathrm{ELS}^{\mathrm{d},30}$	G.D.	$\mathrm{ELS}^{\mathrm{d}}$
Network 1	22.4	20.0	19.6	11.2
Network 2	21.8	18.8	18.2	10.6
Network 3	24.1	21.3	21.0	12.6

Conclusion

High potential of ensemble methods, promising first results for air quality

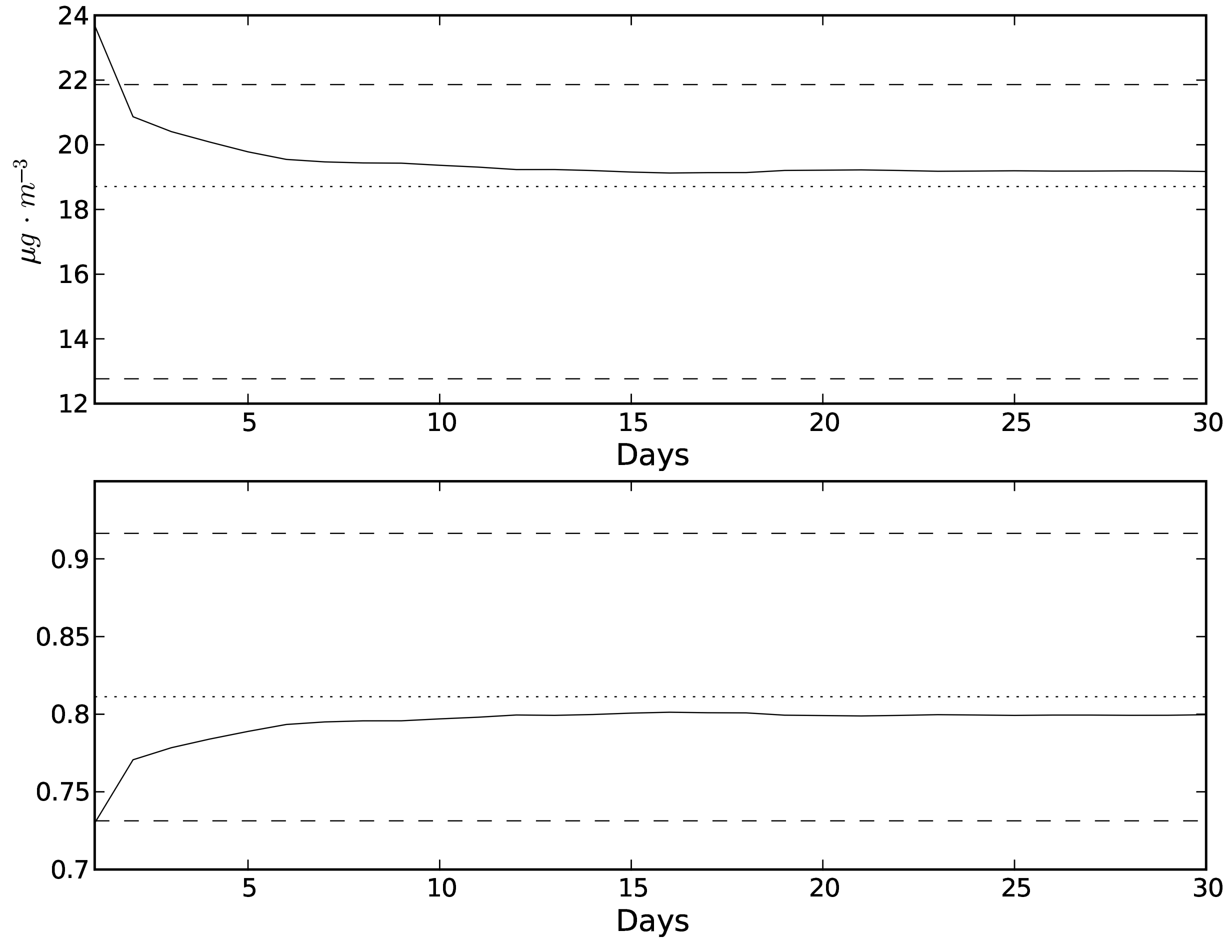
Mallet, V. and Sportisse, B. (2005a). Toward ensemble-based air-quality forecasts. *Submitted to J. Geophys. Res.*

Ensemble structure and network design

Link with classical data assimilation

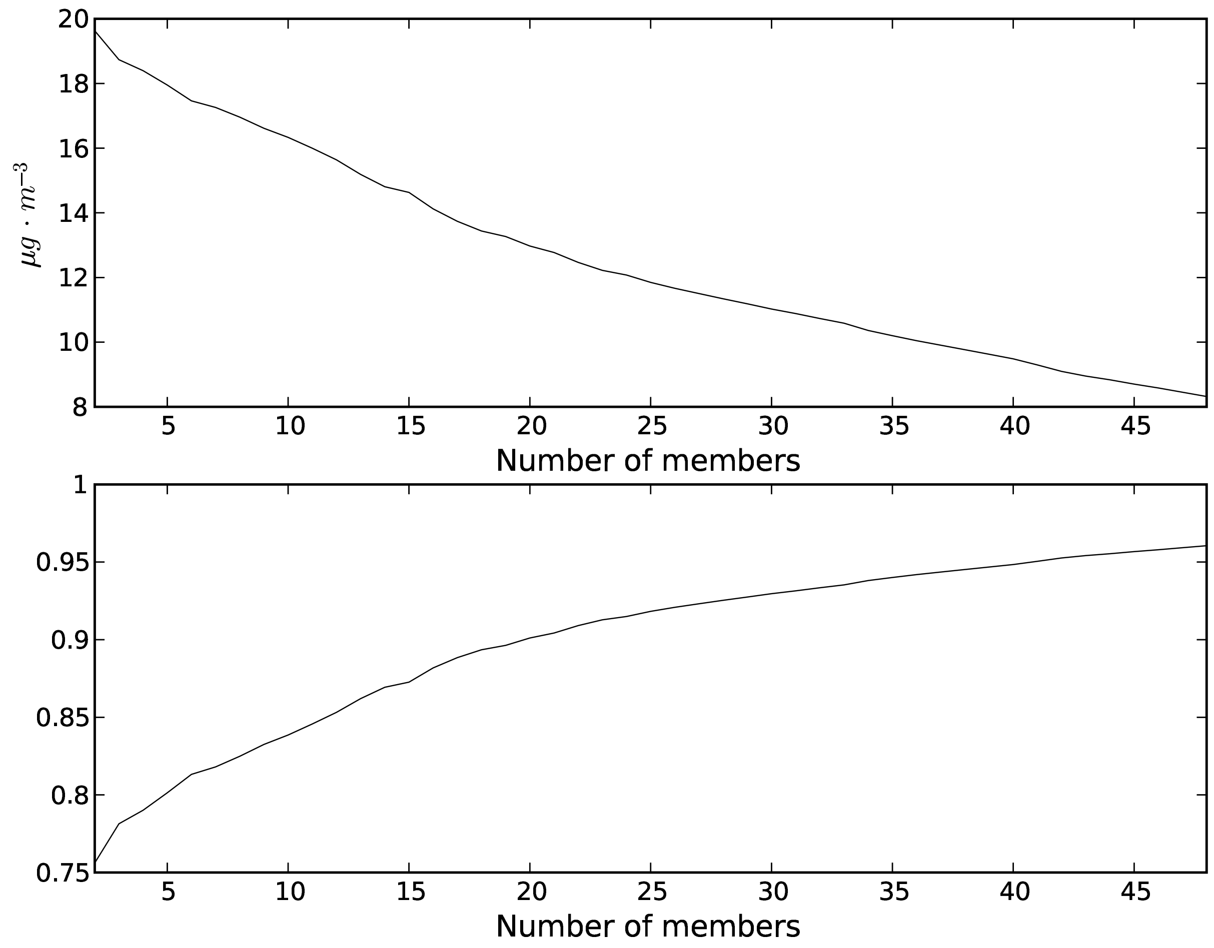
Further details

Performances versus the learning-period length



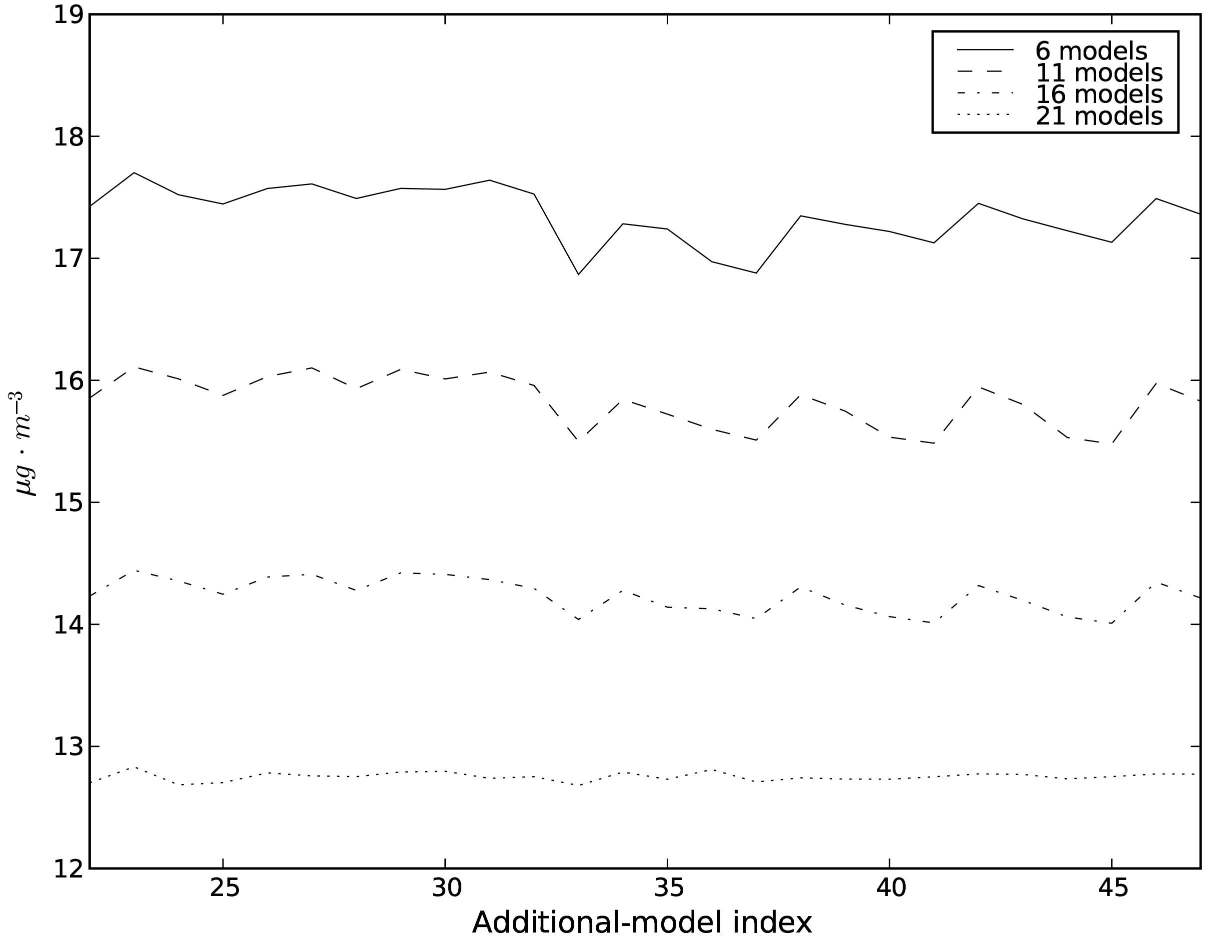
Further details

Performances versus the number of members



Further details

Contribution of models



Further details

Weights over all stations – $\mathrm{ELS}^{\mathrm{s}}$