Preliminary Results of Data Assimilation within the Modeling Platform Polyphemus

L.Wu, V.Mallet, M.Bocquet, and B.Sportisse

CLIME Project (INRIA / ENPC)
Research and Teaching Center in Atmospheric Environment
École Nationale des Ponts et Chaussées / EDF R&D

Les premières Journées Nationales des ARC, 17-18 octobre 2006

Outline

Polyphemus: Purpose

Purpose

Comprehensive and perennial platform for air quality modeling, with advanced forecasting methods such as data assimilation and ensemble forecast.

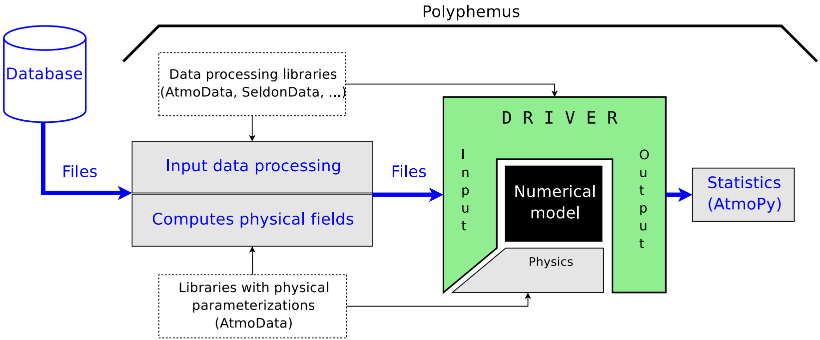
Highlights

- Designed to share developments and to host other models
- Wide range of applications

How?

- Modern programming (priority on C++ and Python)
- Open source (GNU GPL)
- Developed by CEREA and CLIME, supported by IRSN and INERIS

Polyphemus: Overall Structure



Quick Review of Polyphemus Content

Libraries: AtmoData, AtmoPy, etc.

Data post- and preprocessing, physical parameterizations

Models: Castor, Polair3D, etc.

3D chemistry-transport models, Gaussian models

Modules

Photochemistry, aerosols, radionuclides, transport

Drivers

Data assimilation (OI, EnKF, RRSRQT, 4D-Var), local-scale simulation

Introduction to Data Assimilation

Objective and content

- Universal model-experiment problem
- State estimation from diverse information

Components

- Model (physics)
- Data (observation)
- Assimilation algorithms

State-of-the-art

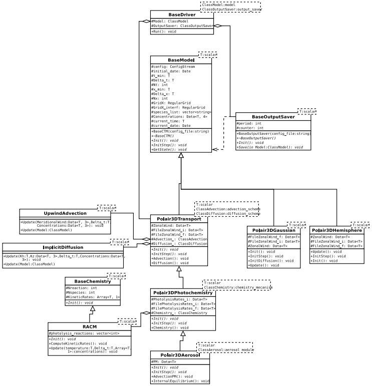
Components	*In general*	*ADOQA in action*
Model	Meteorology, oceanography hydrology, agronomy, ...	Photochemistry (Polair3D), aerosol, ...
Data	In situ, radar, Satellite, ...	Ground stations, Ozone column info (Berroir) Image sequences (Huot et al.)
Algorithms	Sequential and variational	Maximum entropy (Bocquet)

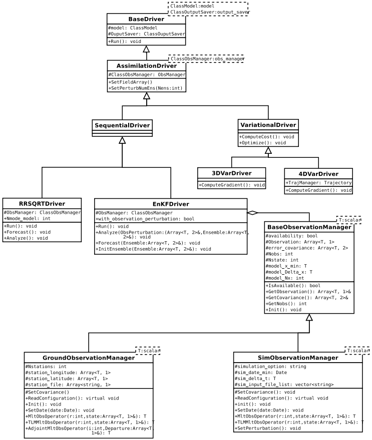
Difficulties

- Nonlinearity + high dimension
  - Constraint on data assimilation window; reduction on model or error covariance matrices
  - Highly nonlinear reaction item of chemistry; but stable because of the eigenvalues of the Jacobian are negative
- Error modeling
  - Needed because of the high uncertainties of the model
  - Multi-species, multivariate error covariance

Data Assimilation: System Design

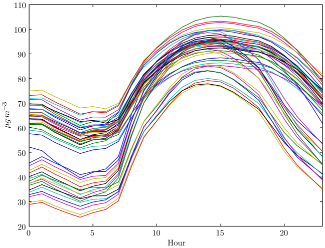
- Idea: Object-orient; data, model and algorithm components are independent of one another.



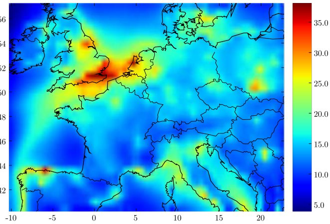


Model facts

- $$\frac{\partial c_i}{\partial t} = -\mathrm{div}(V c_i) + \mathrm{div}(K\nabla c_i) + \chi_i(c) + S_i - P_i$$
  Ground BC: $K\nabla c_i \cdot n = E_i - D_i$
- Dimension of state variables $10^5 - 10^7$
- Uncertainties due to physics parameterization and numerical approximations (Mallet and Sportisse 2006).



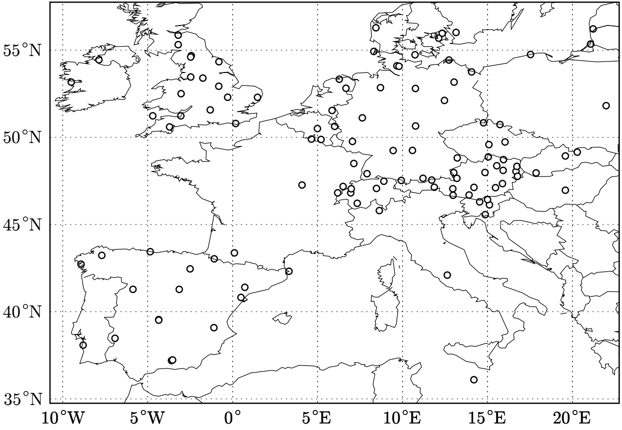
Ozone daily profiles from 48 members



Relative standard deviation over 15% over Europe

Observation managements

- EMEP observation network supported, to be extended to Pioneer, BDQA; real and synthetic observations.



Algorithms: Sequential and Variational

- Data assimilation: determining $\mathbf{x}^a$ given background $\mathbf{x}^b$, observation $\mathbf{y}$, and statistics information $\mathbf{R}, \mathbf{B}$.

Notations

- $\mathbf{x}$: model state vector.
- $\mathbf{x}^t$: true state.
- $\mathbf{x}^b$: background estimates.
- $\mathbf{x}^a$: analysis.
- $\mathbf{y}$: observation vector.
- ϵ^b : $\mathbf{x}^b - \mathbf{x}^t$ background errors.
- ϵ^a : $\mathbf{x}^a - \mathbf{x}^t$ analysis errors.
- H: Observation operator
- $\mathbf{B}$: $\overline{(\epsilon^b - \bar{\epsilon}^b)(\epsilon^b - \bar{\epsilon}^b)^T}$ background error covariance matrix.
- $\mathbf{A}$: $\overline{(\epsilon^a - \bar{\epsilon}^a)(\epsilon^a - \bar{\epsilon}^a)^T}$ analysis error covariance matrix.
- ϵ^o : $\mathbf{y} - H(\mathbf{x}_t)$ observation errors.
- $\mathbf{R}$: $\overline{(\epsilon^o - \bar{\epsilon}^o)(\epsilon^o - \bar{\epsilon}^o)^T}$ observation error covariance matrix.

Linear combination of $\mathbf{x}^b$ and $\mathbf{y}$: optimal interpolation

Formulae

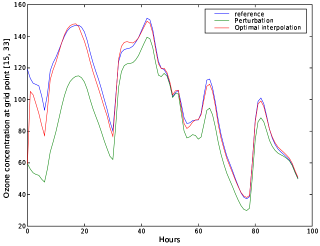
$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - H(\mathbf{x}^b)),$$
$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}.$$

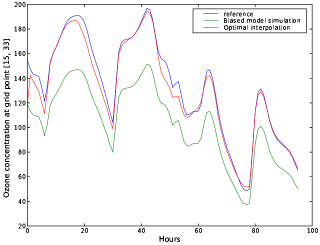
Study

B: using Balgovind correlation function.

$$f(r) = \left(1 + \frac{r}{L}\right)\exp\left(-\frac{r}{L}\right)v^b$$

- Perturbs ICs (upper).
- Biased model (lower).





$\mathbf{x}^b$ generated by model dynamics M: (extended) Kalman filter

Formulae

- Model error:

$$\epsilon_{m,k-1} = M_{k-1\to k}(\mathbf{x}_{t,k-1}) - \mathbf{x}_{t,k}$$

$\mathbf{Q}_{k-1}$: model error covariance matrix.

- Forecast formula:

$$\mathbf{x}_k^f = M_{k-1\to k}(\mathbf{x}_{k-1}^a)$$
$$\mathbf{P}_k^f = \mathbf{M}_{k-1\to k}\mathbf{P}_{k-1}^a\mathbf{M}_{k-1\to k}^T + \mathbf{Q}_{k-1}$$

- Analysis formula:

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k - H(\mathbf{x}_k^f)),$$
$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^T + \mathbf{R}_k)^{-1},$$
$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f$$

Reduced rank square root Kalman filter: RRSQRT

Heemink et al. 2001

Formulae

- Initialization: $\mathbf{x}_0^a, \mathbf{L}_0^a = [\mathbf{l}_0^{a,1}, \dots, \mathbf{l}_0^{a,q}]$
- Forecast formula:

$$\mathbf{x}_k^f = M_{k-1\to k}(\mathbf{x}_{k-1}^a)$$

$$\mathbf{l}_k^{f,i} = \frac{1}{\epsilon}\{M_{k-1\to k}(\mathbf{x}_{k-1}^a + \epsilon \mathbf{l}_{k-1}^{a,i}) - M_{k-1\to k}(\mathbf{x}_{k-1}^a)\}, \quad \epsilon = 1$$

$$\tilde{\mathbf{L}}_k^f = [\mathbf{l}_k^{f,1}, \dots, \mathbf{l}_k^{f,q}, \mathbf{Q}_{k-1}^{\frac{1}{2}}], \qquad \mathbf{L}_k^f = \Pi_k^f \tilde{\mathbf{L}}_k^f$$

- Analysis formula:

$$\mathbf{P}_k^f = \mathbf{L}_k^f \mathbf{L}_k^{f,T}$$

$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1},$$

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k - H(\mathbf{x}_k^f)),$$

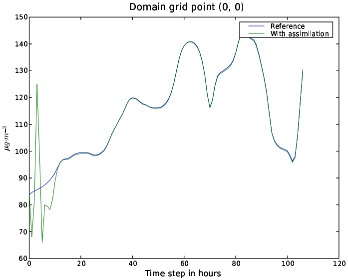
$$\tilde{\mathbf{L}}_k^a = [(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{L}_k^f, \mathbf{K}_k \mathbf{R}_k^{\frac{1}{2}}], \qquad \mathbf{L}_k^a = \Pi_k^a \tilde{\mathbf{L}}_k^a$$

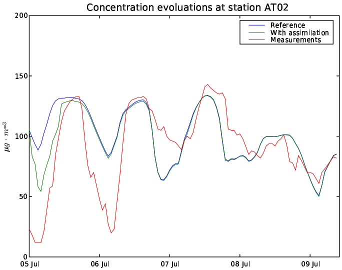
Reduced rank square root Kalman filter: RRSQRT

Study

- The column $\left\{\mathbf{Q}_{k-1}^{\frac{1}{2}}\right\}_i = \left(M_{k-1\to k}(\mathbf{x}_{k-1}^a, \mathbf{d} + \varepsilon \cdot \mathbf{w}_i) - M_{k-1\to k}(\mathbf{x}_{k-1}^a, \mathbf{d})\right)/\varepsilon, \varepsilon = 1$
- Experiment settings
  - Assimilation: 00H00 - 10H00, July 5
  - Prediction: 11H00, July 5 - 10H00, July 9
  - Number of mode $q = 10$, column number of $\mathbf{Q}_k^{\frac{1}{2}}$ set to 10, column number of $\mathbf{R}_k^{\frac{1}{2}}$ set to 10.
  - Perturbed fields are attenuation, deposition velocities, photolysis rates, surface emissions, and boundary conditions for O_3.

RRSQRT: Preliminary results





Ensemble Kalman filter: EnKF

Evensen 1994

Formulae

- Initialization: given initial pdf $p(\mathbf{x}_0^t)$, an ensemble of r members are generated randomly,

$$\{\mathbf{x}_0^{a,i}, \quad , i = 1, \ldots, r\}$$

$$\bar{\mathbf{x}}_0^a = \frac{1}{r} \sum_{i=1}^{r} \mathbf{x}_0^{a,i}$$

$$\tilde{\mathbf{P}}_0^a = \frac{1}{r-1} \sum_{i=1}^{r} \left(\mathbf{x}_0^{a,i} - \bar{\mathbf{x}}_0^a\right) \left(\mathbf{x}_0^{a,i} - \bar{\mathbf{x}}_0^a\right)^T$$

- Forecast formula:

$$\mathbf{x}_k^{f,i} = M_{k-1}[\mathbf{x}_{k-1}^{a,i}] + \eta_{k-1}^i$$

$$\tilde{\mathbf{P}}_k^f = \frac{1}{r-1} \sum_{i=1}^{r} \left(\mathbf{x}_k^{f,i} - \bar{\mathbf{x}}_k^f\right) \left(\mathbf{x}_k^{f,i} - \bar{\mathbf{x}}_k^f\right)^T$$

where $\bar{\mathbf{x}}_k^f$ is the mean of ensemble $\{\mathbf{x}_k^{f,i}, i = 1, .., r\}$

Ensemble Kalman filter: EnKF

Formulae

- Analysis formula:

$$\mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} + \tilde{\mathbf{K}}_k \left(\mathbf{y}_k^i - H_k[\mathbf{x}_k^{f,i}] \right)$$

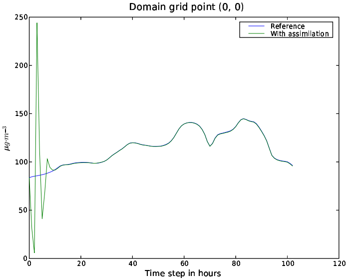
$$\mathbf{x}_k^a = \frac{1}{r} \sum_{i=1}^{r} \mathbf{x}_k^{a,i}$$

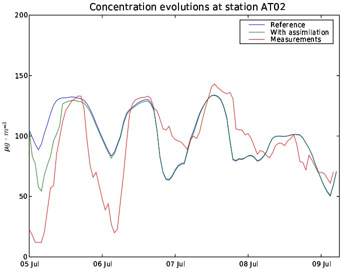
$$\tilde{\mathbf{P}}_k^a = \frac{1}{r-1} \sum_{i=1}^{r} \left(\mathbf{x}_k^{a,i} - \mathbf{x}_k^a \right) \left(\mathbf{x}_k^{a,i} - \mathbf{x}_k^a \right)^T$$

Study

- Assimilation: 00H00 - 06H00, July 5
- Prediction: 07H00, July 5 - 06H00, July 9
- Number of ensemble members $r = 30$.
- Perturbed fi elds are same as those of RRSQRT.

EnKF: Preliminary results





Four-Dimensional Variational Assimilation: 4D-Var
Bouttier and Courtier 1999

- A cost function $J(\mathbf{x})$ that deals with a set of obs.:

$$\underbrace{\frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b)}_{J_b}+\underbrace{\frac{1}{2}\sum_{k=0}^{N}\overbrace{(\mathbf{y}_k-H_k(\mathbf{x}_k))^T\,\mathbf{R}_k^{-1}\,(\mathbf{y}_k-H_k(\mathbf{x}_k))}^{J_{oi}}}_{J_o}$$

where the assimilation window is $0-N$,

$$\mathbf{x}_k=M_{0\to k}(\mathbf{x})=M_kM_{k-1}\ldots M_1\mathbf{x}$$

- The gradient is calculated by the backward intergration of adjoint model
  - $\tilde{\mathbf{x}}_N=0$
  - For $k=N,\ldots,1$, calculates $\tilde{\mathbf{x}}_{k-1}=\mathbf{M}_k^T\left(\tilde{\mathbf{x}}_k-\mathbf{H}_k^Td_k\right)$, where $d_k=\mathbf{R}_k^{-1}(\mathbf{y}_k-H_k(\mathbf{x}_k))$
  - $\tilde{\mathbf{x}}_0:=\tilde{\mathbf{x}}_0-\mathbf{H}_0^T(d_0)$ gives the gradient of J_o with respect to $\mathbf{x}$

Implementation: Adjoint Coding and Gradient Checking

- Adjoint model obtained by automatic differentiation (O∂yssee, TROPICS team, INRIA Sophia-Antipolis)
- Checks gradient calculation by fi nite difference
$$\varphi(\alpha) = \frac{J_o(\mathbf{x} + \alpha h) - J_o(\mathbf{x})}{\alpha \langle \nabla_{\mathbf{x}} J_o, h \rangle}, \qquad \alpha \to 0$$
- Checking case: synthetic observations, $\mathbf{R}_k, H_k$ identity matrix

α	$\varphi(\alpha)$	α	$\varphi(\alpha)$
10^{-1}	0.97169086073122756808	10^{-6}	1.0000210386562053966
10^{-2}	0.99714851925111402942	10^{-7}	0.99992712020685792229
10^{-3}	0.99971484771238313	10^{-8}	0.99873139920821940585
10^{-4}	0.99997151406647999394	10^{-9}	0.98614371814982537678
10^{-5}	0.99999664594783310712	10^{-10}	0.87396334809574471869

Conclusion

- Platform ready
- Easily extended to new features
  - Aerosol model
  - Advanced data assimilation methods for ozone column observations
  - Nonlinear data assimilation methods, i.e. Maximum entropy filter (M.Bocquet), particle filter (Project ASPI)
- Open scientifi c issues
  - Ensemble initialization
  - Background and model error modeling and its parameterization
- Applications
  - Long-run objective - operational platform (ensemble prediction already operational)
  - Open to other models / teams (GNU GPL), education purpose