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

10th October 2006

Outline

Polyphemus: Purpose

Purpose

Achieve a comprehensible and perennial platform for air quality modeling, primarily for risk assessment and impact studies

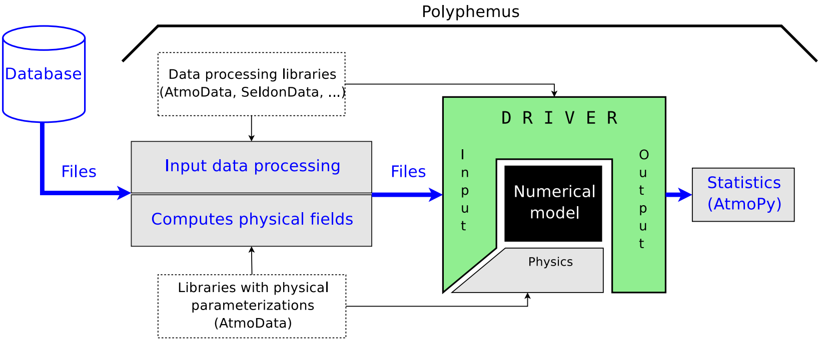
Highlights

- 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

Start-of-art

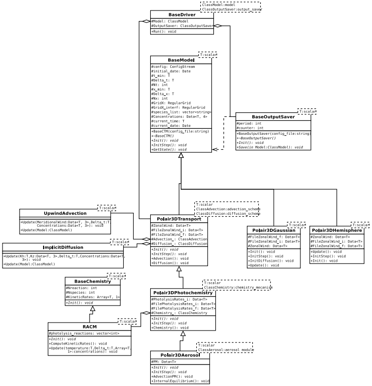
Components	*In general*	*ADOQA in action*
Model	Meteorology, oceanography hydrology, agronomy, ...	Photochemistry (Polair3D), aerosol, ...
Data	In situ, radar, Satellite, ...	Ground stations, Ozone column info (Berroir) Image sequences (Huot et al.)
Algorithms	Sequential and variational	Maximum entropy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

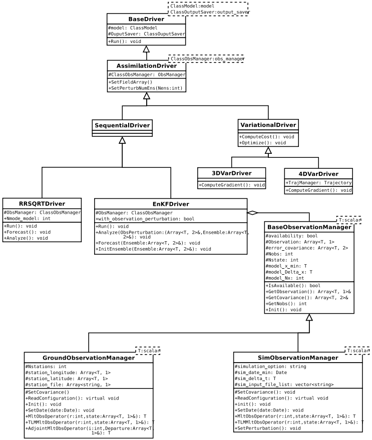
Difficulties

- 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 system stiffness.
- Error modeling
  - As difficult as physics parameterization
  - Multi-species, multivariate error covariance

Data Assimilation: System Design

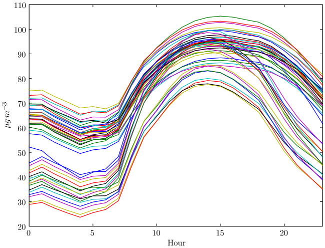
- Idea: Object-orient; data, model and algorithm componets are independent of one another.



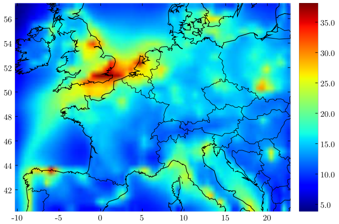


Model facts

- $\frac{\partial c_i}{\partial t} = -\mathrm{div}(Vc_i) + \mathrm{div}(K\nabla c_i) + \chi_i(c) + S_i - P_i$
  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).



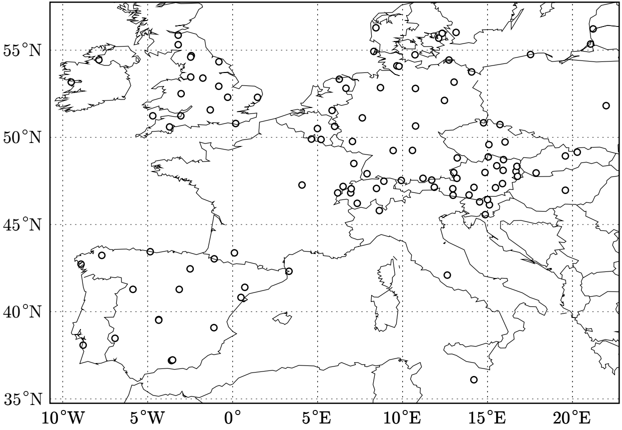
Ozone daily profiles from 48 members



Relative standard deviation over 15% over Europe

Observation managements

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



Algorithms

- Data assimilation: determining $\mathbf{x}_a$ given background $\mathbf{x}_b$, observation $\mathbf{y}^o$, 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.
- $\mathbf{y}^o$: 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 : $\mathbf{y}^o - H(\mathbf{x}_t)$ 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 $\mathbf{y}^o$: optimal interpolation

Formulae

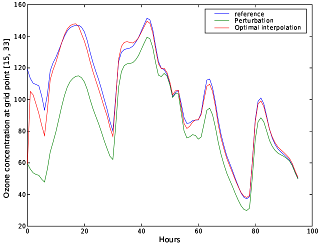
$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y}^o - H(\mathbf{x}_b)),$$
$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}.$$

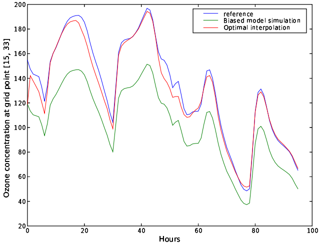
Study

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:

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

Reduced rank square root Kalman filter: RRSQRT

Formulae

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

$$\mathbf{x}_k^f = M_{k-1\to k}(\mathbf{x}_{k-1}^a)$$
$$\mathbf{l}_k^{f,i} = \frac{1}{\epsilon}\{M_{k-1\to k}(\mathbf{x}_{k-1}^a + \epsilon \mathbf{l}_{k-1}^{a,i}) - M_{k-1\to k}(\mathbf{x}_{k-1}^a)\}, \quad \epsilon = 1$$
$$\tilde{\mathbf{L}}_k^f = [\mathbf{l}_k^{f,1}, \ldots, \mathbf{l}_k^{f,q}, \mathbf{Q}_{k-1}^{\frac{1}{2}}], \qquad \mathbf{L}_k^f = \Pi_k^f \tilde{\mathbf{L}}_k^f$$

- Analysis formula:

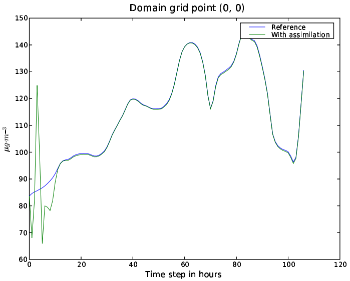
$$\mathbf{P}_k^f = \mathbf{L}_k^f \mathbf{L}_k^{f,T}$$
$$\mathbf{K}_k = \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}_k)^{-1},$$
$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^o - H(\mathbf{x}_k^f)),$$
$$\tilde{\mathbf{L}}_k^a = [(\mathbf{I} - \mathbf{K}_k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

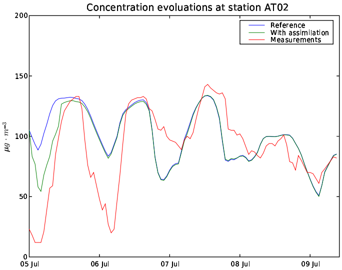
Reduced rank square root Kalman filter: RRSQRT

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, July 5, 2001 - 10H00, July 5, 2001
  - Prediction: 11H00, July 5, 2001 - 10H00, July 9, 2001
  - Number of mode $q = 10$, columns of $\mathbf{Q}_k^{\frac{1}{2}}$ set to 10, columns 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

Formulae

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

$$\{\mathbf{x}_0^{a(\alpha)}, \quad , \alpha = 1, \ldots, r\}$$

$$\bar{\mathbf{x}}_0^a = \frac{1}{r} \sum_{\alpha=1}^{r} \mathbf{x}_0^{a(\alpha)}$$

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

- Forecast formula:

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

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

where $\bar{\mathbf{x}}_k^f$ is the mean of ensemble $\{\mathbf{x}_k^{f(\alpha)}, \alpha = 1, .., r\}$ defi ned similar to$\bar{\mathbf{x}}_0^a$

Ensemble Kalman filter: EnKF

Formulae

- Analysis formula:

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

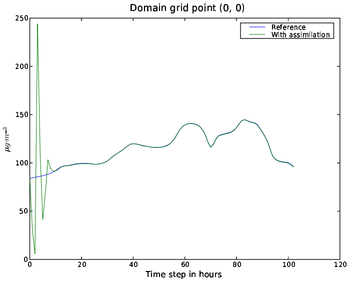
$$\mathbf{x}_k^{a} = \frac{1}{r} \sum_{\alpha=1}^{r} \mathbf{x}_k^{a(\alpha)}$$

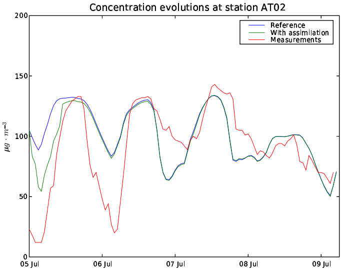
$$\tilde{\mathbf{P}}_k^{a} = \frac{1}{r-1} \sum_{\alpha=1}^{r} \left(\mathbf{x}_k^{a(\alpha)} - \mathbf{x}_k^{a} \right) \left(\mathbf{x}_k^{a(\alpha)} - \mathbf{x}_k^{a} \right)^T$$

Study

- Assimilation: 00H00, July 5, 2001 - 06H00, July 5, 2001
- Prediction: 07H00, July 5, 2001 - 06H00, July 9, 2001
- Number of ensemble members $r = 30$.
- Perturbed fi elds are same as those of RRSORT.

EnKF: Preliminary results





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 initilizations
  - Background and model error modeling and its parameterization
- Applications
  - Long-run objective - operational platform (ensemble prediction allready operational)
  - Education purpose (GNU GPL)