Introduction to Polyphemus

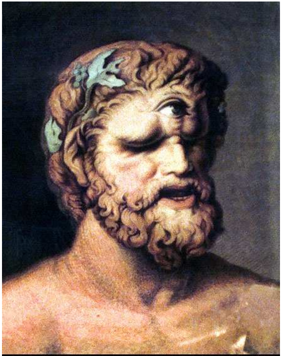
First SAEMC-IAI, STIC-AmSud Workshop

Meryem Ahmed de Biasi, Irène Korsakissok, Vivien Mallet and Lin Wu

5 June 2007

Outline

Polyphemus Images



Johann Heinrich Wilhelm Tischbein, 1751-1829: Polyphem. Photo ©Maicar Förlag-GML

Greek Mythology

- Polyphemus, cyclops in Odyssey

Why this name ?

- “Poly” : **multiple**
- “phemus” : speech

Multiple Goals

Multiple Models

- Scales : from local scale to continental scale
- Formulations : Gaussian, Eulerian, . . .

Multiple Pollutants

- Passive, radionuclides
- Photochemistry
- Aerosols
- Persistent organic pollutants, heavy metals, . . .

Multiple Inputs

- From meteorological models : MM5, ECMWF
- Ground data : GLCF, USGS

Multiple Methods

Data Assimilation

- Sequential
- Variational
- Inverse modeling (parameter estimation)

Ensemble Forecast

- Multimodels
- Monte Carlo
- Models combinations (“superensembles”, . . .)

Models Coupling

- Feedbacks
- Impact

Constraints

Perennial Code

- System maintenance
- Scalable, integration of new developments

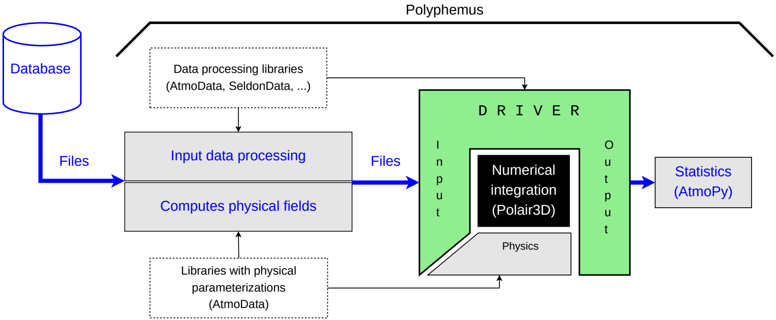
Open

- Availability, distribution
- Development or contributions from other teams

Field Context

- From research to operational use

Overall structure



Structure

Libraries

- Talos : operations on configuration files and management of dates
- SeldonData : input-output operations
- AtmoData : extension of SeldonData to atmospheric sciences, physical parameters
- AtmoPy : postprocessing

Preprocessing

- Meteorological fields from ECMWF or MM5
- Ground data : Land Use Cover, Emission, Deposition
- Boundary conditions from Mozart 2 or Inca for gaseous species and Gocart for aerosols
- Gaussian models.

Structure

Models

- Gaussian Models
  - Stationary Gaussian model and puff Gaussian model
  - Gas and aerosols
  - Several parameterizations to compute the dispersion
- Eulerian Models
  - Castor (clone of the **gas** version of Chimere)
  - Polair3D (passive, chemical, aerosol and adjoint [for gas] versions)
  - Modules for transport, chemistry and aerosols

Postprocessing

- Comparison to observations
- Water diagnosis in a plume
- Visualization

Programming Choices

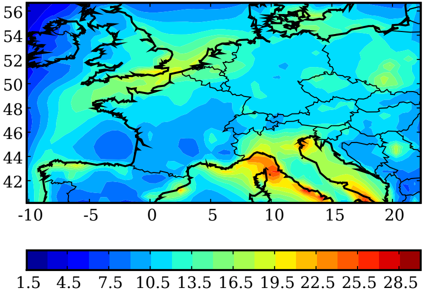
Main language : C++

- Efficient for scientific computing
- Advanced object design
  - Inheritance, genericity
  - Management of complex objects
  - Exceptions
- Widely used and perennial

Programming Choices

Complementary Language : Python

- Dynamic, interactive
- Visualization
- Scripts
- Increasingly used in scientific computing



Programming Choices

Historical Language : Fortran 77

- Automatic differentiation
- Continuity
- Calls from C++

Policy

- To avoid dealing with too many languages
- To use primarily languages with strong potential and productivity

Application and References

Main Reference

Main article for Polyphemus :
"Technical Note : The air quality modeling system Polyphemus", V. Mallet, D. Quélo, B. Sportisse, M. Ahmed de Biasi, É. Debry, I. Korsakissok, L. Wu, Y. Roustan, K. Sartelet, M. Tombette and H. Foudhil, Atmos. Chem. Phys. Discuss., 7 (3), 2007

Partners

Main Organizations

1. ENPC
2. EDF *Impact studies of thermal power plants*
3. INRIA *data assimilation*

French partners

1. IRSN (Nuclear safety) *Radionuclides*
2. INERIS (industrial risks) *Local scale, ozone forecasting, liquid water content diagnosis*
3. CETE (urban equipment) *Evaluation of Urban mobility plan*
4. DGA (Defense) *Local scale, network evaluation*

International collaborations

1. CRIEPI in Japan *Sensitivity studies over Europe, Paris and Tokyo*
2. EDF Polska *Impact studies, air quality modeling*
3. IER Stuttgart, European projects *Health impact evaluation*
4. CMM in Chile, STIC-AMSUD project *Air quality forecasting, data assimilation*

Some Applications

Aerosols at Regional and Continental Scales

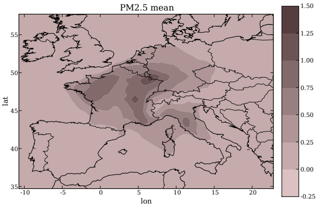
- Polair3D hosts an aerosol model (SIREAM).
- Validations over Europe and Asia, and over Paris, Marseille and Lille.
- Sensitivity studies over Europe, Paris and Tokyo.

- M. Tombette and B. Sportisse. Aerosol modeling at regional scale : Model-to-data comparison and sensitivity analysis over Greater Paris. *Atmos. Env.*, 2007. In press
- K. Sartelet, E. Debry, K. Fahey, M. Tombette, Y. Roustan, and B. Sportisse. Simulation of aerosols and gas phase species over Europe with the POLYPHEMUS system. Part I : model-to-data comparison for year 2001. *Atmos. Env.*, 2007. doi :10.1016/j.atmosenv.2007.04.024
- K. Sartelet, H. Hayami, and B. Sportisse. Dominant aerosol processes during high-pollution episodes over Greater Tokyo. *J. Geophys. Res.*, 2007. Accepted for publication

Some Applications

Impact Studies at Continental Scale

- Sensitivity study of ozone concentrations with respect to emissions
- Evaluation of air quality in Polska
- Evaluation of the emissions from French thermal power plants for the year 2005



- Vivien Mallet and Bruno Sportisse. A comprehensive study of ozone sensitivity with respect to emissions over Europe with a chemistry-transport model. *J. Geophys. Res.*, 110(D22), 2005

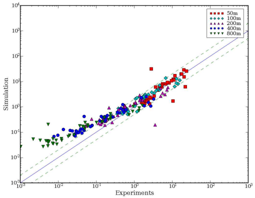
Some Applications

Local scale models

- Comparison of sigma parameterizations
- Network design for defense applications

Plume in Grid model

- Better modeling of major point source emissions
- Application to passive tracers (ETEX, Chernobyl)



- Irène Korsakissok. Performance evaluation of Polyphemus Gaussian models with Prairie Grass experiment. Technical Report 2007-15, CEREA, 2007

Some Applications

Air Quality Modeling over Lille

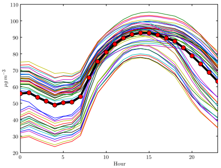
- Assessment of health impact of the Urban Mobility Plan (CETE)
- Coupling with emission and traffic model, and indicators of health impact
- Inverse modeling of emissions with 4D-var and second-order sensitivity

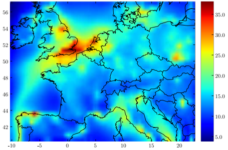
- Denis Quélo, Vivien Mallet, and Bruno Sportisse. Inverse modeling of NO_x emissions at regional scale over Northern France. Preliminary investigation of the second-order sensitivity. *J. Geophys. Res.*, 110(D24310), 2005
- D. Quélo, R. Lagache, and B. Sportisse. Etude de l'impact qualité de l'air des scénarios PDUs sur Lille à l'aide du modèle de Chimie-Transport Polair3D. Technical Report 2004-28, CEREA, 2004. Rapport PREDIT
- R. Lagache, C. Declercq, D. Quélo, B. Sportisse, P. Palmier, B. Quetelard, and F. Haziak. Evaluation du PDU de Lille-Métropole sur le trafic, les concentrations de polluants atmosphériques et la mortalité. In *Actes de la 15ième conférence sur les transports et la pollution de l'air*, 2006

Some Applications

Air Quality Ensemble Forecast

- Estimation of the uncertainties in the output
- Model sequential agregation with weights learned over the past
- Ozone forecasting over Europe (Prév'air operational platform).



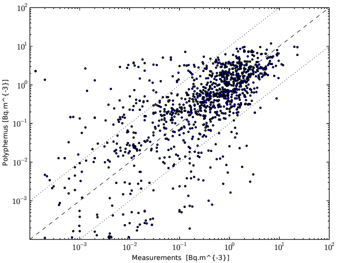


- Vivien Mallet and Bruno Sportisse. Ensemble-based air quality forecasts : a multi-model approach applied to ozone. *J. Geophys. Res.*, 111(D18) :18302, 2006
- Vivien Mallet and Bruno Sportisse. Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations : an ensemble approach applied to ozone modeling. *J. Geophys. Res.*, 111(D01302), 2006

Some Applications

Dispersion of Radionuclides

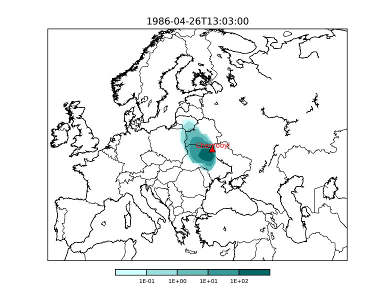
- Polyphemus is the support of the new operational forecast system at the Emergency Center of IRSN
- Simulation of the Chernobyl accident, sensitivity studies
- Inverse modeling applied to simulations of the ETEX exercise and Algeciras accident.



Some Applications

Chernobyl Accidental Release, 25 April-5 May 1986

Polyphemus run, Forecast Emergency Center IRSN/CEREA



- D. Quélo, M. Krysta, M. Bocquet, O. Isnard, Y. Minier, and B. Sportisse. Validation of the POLYPHEMUS system : the ETEX, Chernobyl and Algeciras cases. *Atmos. Env.*, 2007. doi :10.1016/j.atmosenv.2007.02.035
- B. Sportisse. A review of parameterizations for modeling dry deposition and scavenging of radionuclides. *Atmos. Env.*, (41) :2683–2698, 2007

Data Assimilation

Why ?

To improve the results (*analysis*) using several sources of information (model, observations).

How ?

Data assimilation involves three component :

- A model (currently only possible with Polair3D model and gaseous species)
- Observations
- An assimilation algorithm which combines both (implemented as a driver in Polyphemus)

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
- $\mathbf{B}$ (also noted $\mathbf{P}^f$) : $\overline{(\epsilon^b - \overline{\epsilon}^b)(\epsilon^b - \overline{\epsilon}^b)^T}$ background error covariance matrix.
- $\mathbf{A}$ (also noted $\mathbf{P}^a$) : $\overline{(\epsilon^a - \overline{\epsilon}^a)(\epsilon^a - \overline{\epsilon}^a)^T}$ analysis error covariance matrix.
- ϵ^o : $\mathbf{y} - H(\mathbf{x}_t)$ observation errors.
- $\mathbf{R}$: $\overline{(\epsilon^o - \overline{\epsilon}^o)(\epsilon^o - \overline{\epsilon}^o)^T}$ observation error covariance matrix.

Optimal Interpolation

Principle

Combination of background and background departure.

Formulae

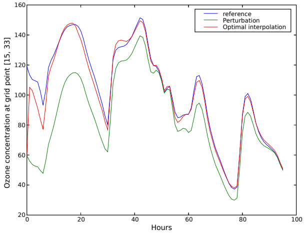
$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - H(\mathbf{x}^b)),$$
$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}.$$

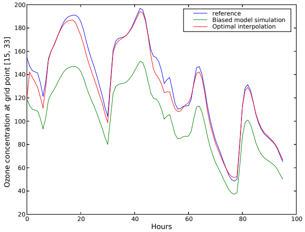
Study

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).





Ensemble Kalman filter (EnKF)

Principle

Time-dependent error covariance matrices. Ensemble statistics are used to approximate $\mathbf{P}^f$ and $\mathbf{P}^a$ at each time.

Formulae

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

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

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

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

EnKF

Formulae

- 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 $\{\mathbf{x}_k^{f,i}, i = 1, .., r\}$

- Analysis formula :

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

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

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

Reduced rank square root Kalman filter : RRSQRT

Heemink et al. 2001

Principle

Also approximate error covariance matrices, but using a low-rank representation $\mathbf{L}\mathbf{L}^T$.
L is the *Mode Matrix* whose q columns represent the dominant directions of the forecast.
In addition, square root of **Q** (the error model) is approximated using ensemble statistics.

Formulae

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

RRSQRT

Formulae

- 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 :

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

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

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

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

Four-Dimensional Variational Assimilation : 4D-Var

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

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

where the assimilation window is $0-N$,
$\mathbf{x}_k = M_{0\rightarrow k}(\mathbf{x}) = M_k M_{k-1}\dots M_1\mathbf{x}$ The gradient is calculated by the backward integration of adjoint model

- $\tilde{\mathbf{x}}_N = 0$
- For $k = N, \dots, 1$, calculates $\tilde{\mathbf{x}}_{k-1} = \mathbf{M}_k^T\left(\tilde{\mathbf{x}}_k - \mathbf{H}_k^T d_k\right)$, 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 $\mathbf{x}$

Results

Introduction to Polyphemus

Goals

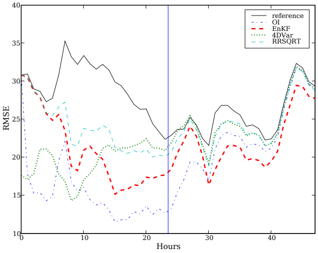
Structure

Applications and References

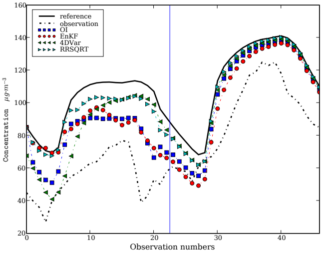
Data Assimilation

Resources

Next...



Root mean square errors.



Results at Montandon station.

Resources

Current Developers

1	Meryem Ahmed de Biasi (INRIA)	diffusion
2	Édouard Debry (ENPC)	aerosols
3	Karine Kata-Sartelet (ENPC)	aerosols
4	Irène Korsakissok (ENPC)	local, plume-in-grid
5	Vivien Mallet (ENPC)	ensemble
6	Denis Quélo (IRSN)	passive
7	Yelva Roustan (ENPC)	impact
8	Bruno Sportisse (ENPC)	aerosols
9	Marilyne Tombette (ENPC)	aerosols
10	Lin Wu (INRIA)	assimilation

Resources

Documentations for users

- User's guide (150 pages)
- Scientific documentation for AtmoData
- Reference documentation for AtmoPy

Documentations for developers

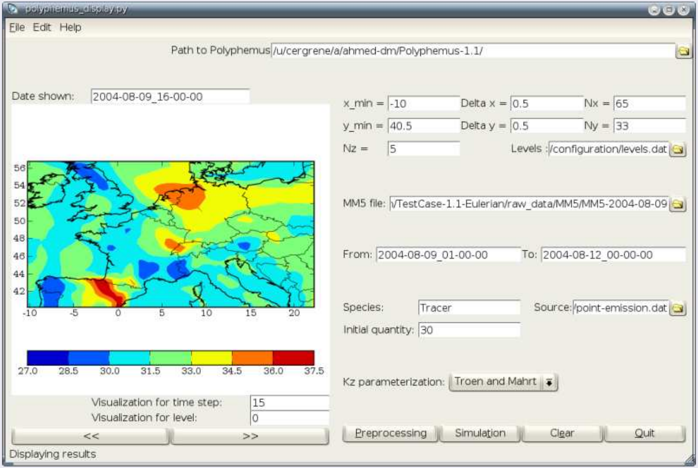
- Guide and reference documentation for SeldonData
- Reference documentation for AtmoData
- Guide and reference documentation for Talos

Examples

- Test cases (Eulerian and Gaussian)
- Practical sessions (primarily for courses at ENSTA and ENPC)

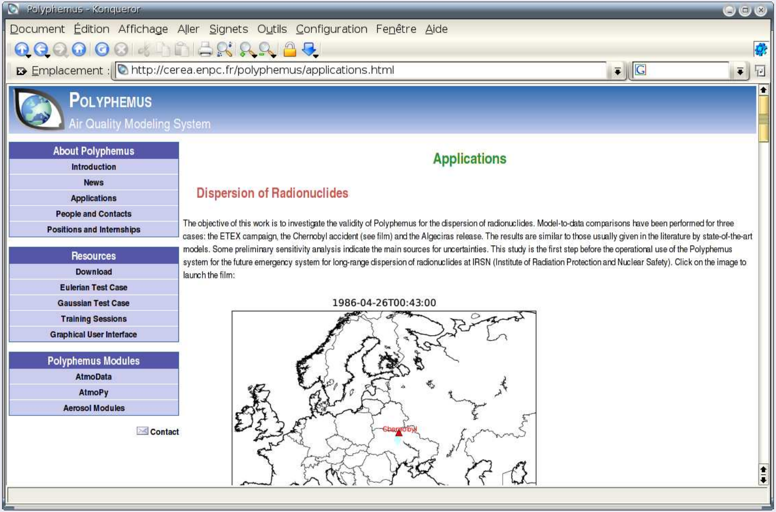
Resources

A Graphical User Interface



Resources

http://cerea.enpc.fr/polyphemus/



Next...

Models (mostly implemented, but not yet included in Polyphemus)

- Secondary organic aerosols
- Modal aerosol module (MAM)
- Heavy metals, mercury
- Persistent organic pollutants (POP)
- Passive hemisphere model

Next Steps

- Parallelization
- Lagrangian particle model
- Drivers for assimilation and coupling