Adjoint-based techniques for the analysis of large-scale uncertain systems

Dacian N. Daescu
Institute for Mathematics and its Applications
University of Minnesota

daescu@ima.umn.edu

Outline of Presentation

- Problem Formulation and Research Impact
- Variational Data Assimilation and Adjoint Modeling
  - ▷ The adjoint model approach for gradient evaluation
  - ▷ Study of chemical reactions models
  - ▷ Applications to sensitivity analysis and data assimilation
- The optimization process
  - ▷ Second order information via adjoint modeling
- Adjoint-based adaptive observations strategies
  - ▷ Gradient and singular vectors methods
  - ▷ Interaction between adaptive observations
- Future research directions

Data Assimilation & Prediction

- Advances in computing power, remote sensing of the atmosphere from satellites, and forecast models have led to considerable progress in numerical weather prediction (NWP)

- Facts:

  ▷ The forecast of some meteorological events (e.g. hurricane) is a difficult prediction problem even at short range (12 to 48 hours)

  ▷ Forecast failures may have dramatic consequences in terms of economical and societal impact

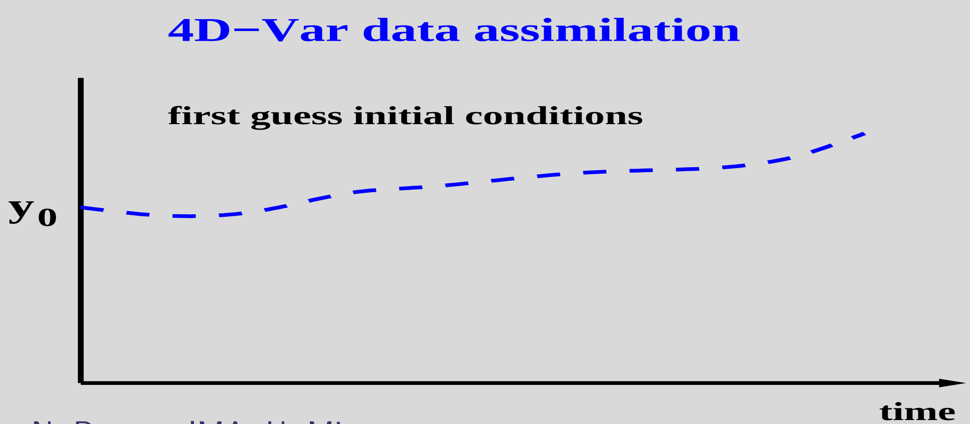
  ▷ Atmospheric models require the specification of a large number of parameters $\sim 10^6 - 10^7$

  ▷ Adaptive observational network design is necessary to improve short range forecasts

  ▷ Assimilation of atmospheric chemical data is a fast growing research field

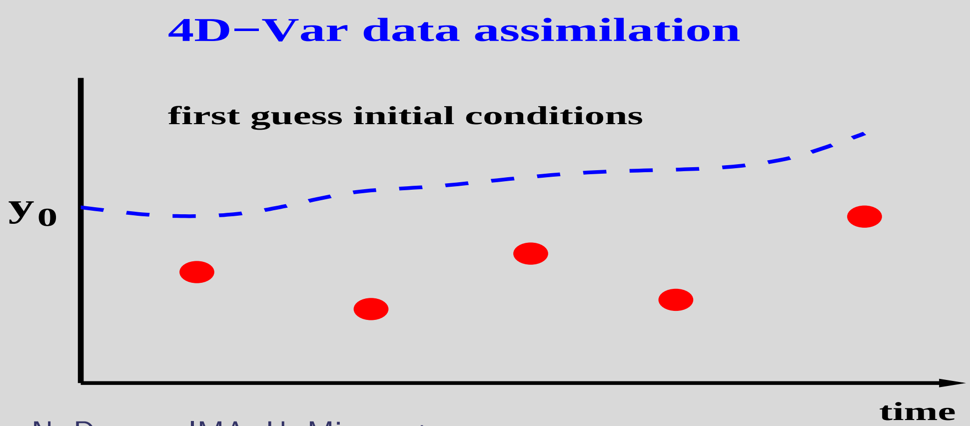
The Data Assimilation Process

- Aim:
  - Provide an optimal estimate of the evolving state of a dynamical system
- Ingredients:
  - a dynamical model of the system
  - observations of the system
  - a method for systematically combining the model with observations



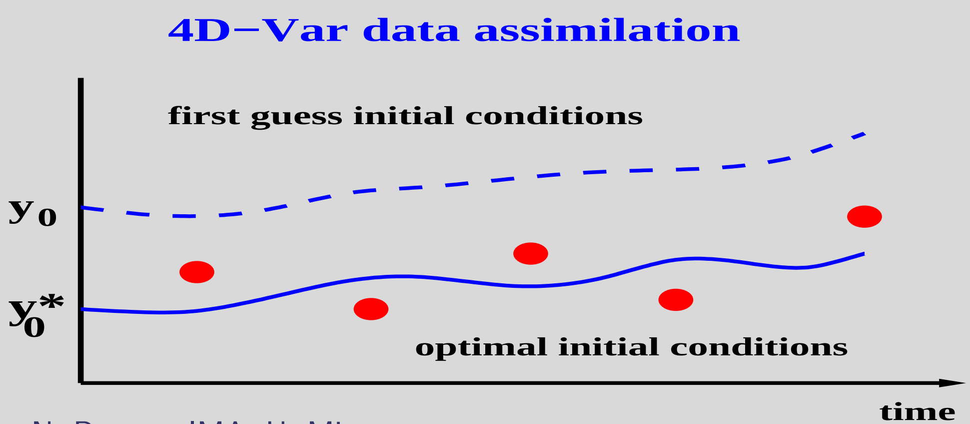
The Data Assimilation Process

- Aim:
  - ▷ Provide an optimal estimate of the evolving state of a dynamical system
- Ingredients:
  - ▷ a dynamical model of the system
  - ▷ observations of the system
  - ▷ a method for systematically combining the model with observations



The Data Assimilation Process

- Aim:
  - Provide an optimal estimate of the evolving state of a dynamical system
- Ingredients:
  - a dynamical model of the system
  - observations of the system
  - a method for systematically combining the model with observations



Four Dimensional Variational Data Assimilation

- Time-discrete deterministic model

$$\mathbf{y}_k = M_{k-1,k}(\mathbf{y}_{k-1}), \quad k = 1, 2, \ldots, n$$

- Consider the set of observations up to time t_n

$$\mathcal{O}_n = \{\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_n\}, \ \hat{\mathbf{y}}_k = \mathbf{H}_k \mathbf{y}_k + \epsilon_k^o, k = 1, 2, \ldots n$$

- Least-squares approach: minimize the cost functional

$$\mathcal{J} = \frac{1}{2}(\mathbf{y}_0 - \mathbf{y}^b)^T \mathbf{B}^{-1}(\mathbf{y}_0 - \mathbf{y}^b) + \frac{1}{2}\sum_{k=1}^{n}(\hat{\mathbf{y}}_k - \mathbf{H}_k\mathbf{y}_k)^T \mathbf{R}_k^{-1}(\hat{\mathbf{y}}_k - \mathbf{H}_k\mathbf{y}_k)$$

with respect to $\{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_n\}$ subject to the model constraints.

- Reduced problem

$$\mathbf{y}_k = M_{k-1,k}(M_{k-2,k-1}(\ldots\ldots M_{0,1}(\mathbf{y}_0))) \Rightarrow \min_{\mathbf{y}_0} \mathcal{J}(\mathbf{y}_0)$$

- Large-scale problems $\longrightarrow$ critical need for accurate and efficient gradient evaluation $\nabla_{\mathbf{y}_0}\mathcal{J}(\mathbf{y}_0)$

Parameter estimation

Problem statement:

Find an estimated value $\hat{\mathbf{u}}$ of the parameter $\mathbf{u}$ from knowledge of:

- observational data $\hat{\mathbf{y}}$
- parameter to output mapping $\mathbf{y} = \mathbf{\Phi}(\mathbf{u})$
- some *'a priori'* knowledge on the parameter (admissible parameters set U_{ad})

Least- squares formulation

$$\mathcal{J}(\mathbf{u}) = \|\mathbf{\Phi}(\mathbf{u}) - \hat{\mathbf{y}}\|_2^2 \ ; \ Find \ \hat{\mathbf{u}} \in U_{ad} : \ \mathcal{J}(\hat{\mathbf{u}}) \le \mathcal{J}(\mathbf{u}), \forall \mathbf{u} \in U_{ad}$$

A simple example: $NO_2 + h\nu \xrightarrow{k} NO + O$

Model

$$\frac{dc(t)}{dt} = -kc(t) + E;\;\; c(0) = c_0,\;\; c(t) = [NO_2(t)],\;\; E = E_{NO_2}$$

Solution $c(t) = c_0 e^{-kt} + (1 - e^{-kt})\frac{E}{k}$

Sensitivity to parameter variations

$$\begin{aligned} \delta c(0) \Rightarrow \delta c(t) &= \delta c(0) e^{-kt} \stackrel{t\to\infty}{\longrightarrow} 0 \\ \delta E \Rightarrow \delta c(t) &= \frac{1}{k}(1 - e^{-kt})\delta E \stackrel{t\to\infty}{\longrightarrow} \frac{1}{k}\delta E \end{aligned}$$

Parameter sensitivity to data

$$\begin{aligned} \delta c(t) \Rightarrow \delta c(0) &= e^{kt}\delta c(t) \stackrel{t\to\infty}{\longrightarrow} \infty \\ \delta c(t) \Rightarrow \delta E &= \frac{k}{1 - e^{-kt}}\delta c(t) \stackrel{t\to\infty}{\longrightarrow} k\delta c(t) \end{aligned}$$

Model Problem

$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(t, \mathbf{y}, \mathbf{p}), \quad \mathbf{y}(t_0) = \mathbf{y}_0$$

$$\mathbf{y} \in R^N, \mathbf{p} \in R^m, \; t_0 \leq t \leq t_F$$

Let $\mathbf{u} = (\mathbf{y}_0, \mathbf{p}) \in R^{N+m}$ be the vector of model parameters, $\mathbf{y} = \mathbf{y}(t, \mathbf{u})$.

Given a scalar response function $\mathcal{J} = \mathcal{J}\big(\mathbf{y}(t_F, \mathbf{u})\big)$

Aim: Estimate the rate of change in the output of a model with respect to changes in the model input: $\nabla_{\mathbf{u}} \mathcal{J}$

$$\frac{\partial \mathcal{J}}{\partial u_i} \approx \frac{\mathcal{J}(y(t_F, \mathbf{u} + \epsilon \mathbf{e}_i)) - \mathcal{J}(y(t_F, \mathbf{u}))}{\epsilon}, \quad i = \overline{1, N+m}$$

Direct methods

$(+)$ easy to implement, efficient for few in $\rightarrow$ many out.

$(-)$ inefficient when the number of parameters is large, many in $\rightarrow$ few out.

Adjoint sensitivity analysis

Introduce the adjoint variable $\lambda(t) \in R^N$ as the solution of the adjoint model problem

$$\frac{d\lambda}{dt} = -\mathbf{f}_{\mathbf{y}}^T(t, \mathbf{y}, \mathbf{p})\lambda, \;\; \lambda(t_F) = \nabla_{\mathbf{y}}\mathcal{J}(\mathbf{y}(t_F))$$

Then

$$\delta\mathcal{J} = \langle \lambda(t_0), \delta\mathbf{y}_0 \rangle_N + \int_{t_0}^{t_F} \langle \mathbf{f}_{\mathbf{p}}^T(t, \mathbf{y}, \mathbf{p})\lambda, \delta\mathbf{p} \rangle_m dt$$

$$\nabla_{\mathbf{y}_0}\mathcal{J} = \lambda(t_0), \;\; \nabla_{\mathbf{p}}\mathcal{J} = \mathbf{f}_{\mathbf{p}}^T(t, \mathbf{y}, \mathbf{p})\lambda \tag{1}$$

(+) Very efficient for many in $\longrightarrow$ few out

(+) Computational cost of the same order as a forward model integration

(+) Provides backward sensitivity at no additional cost

(-) Inefficient for few in $\longrightarrow$ many out

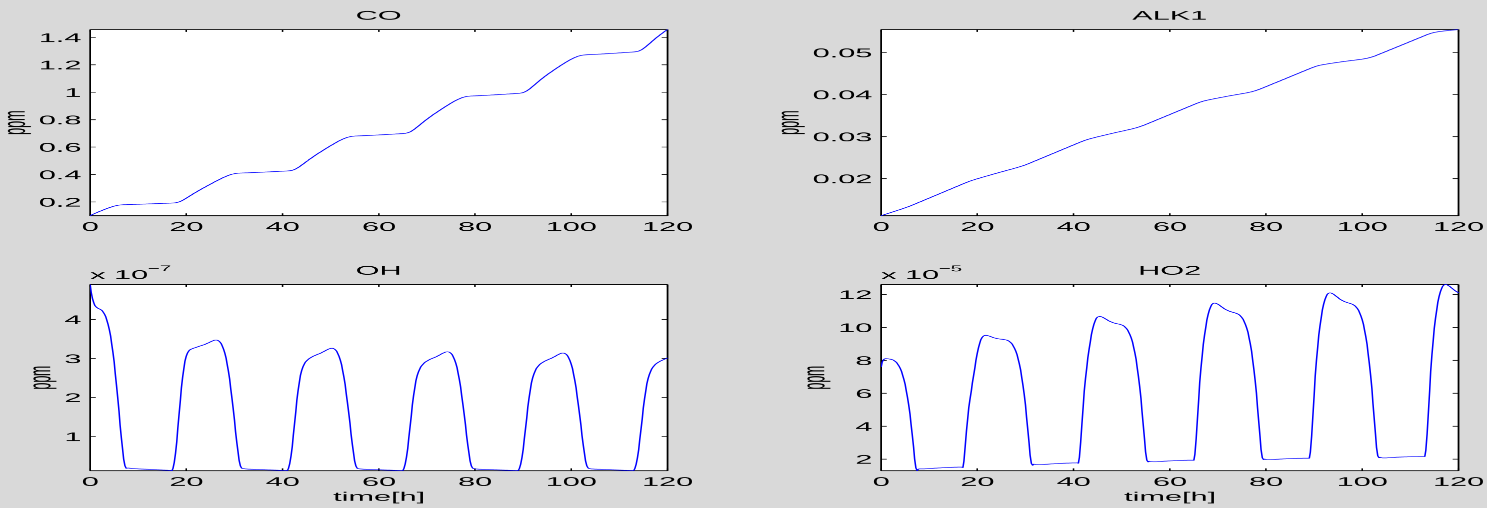
(-) High memory storage requirements $\Rightarrow$ trade-off CPU/memory

(-) Implementation may not be trivial

Case study: SAPRC-99 model

- Comprehensive gas-phase atmospheric reactions mechanism used to predict the effects of VOC and NOx emissions on tropospheric secondary pollutants formation
- 211 reactions among 74 variable chemical species
- Emissions are prescribed for 30 chemical species in the model.
- Currently incorporated into 3D regional models

Forward model integration: slow/fast varying species $\Rightarrow$ stiffness



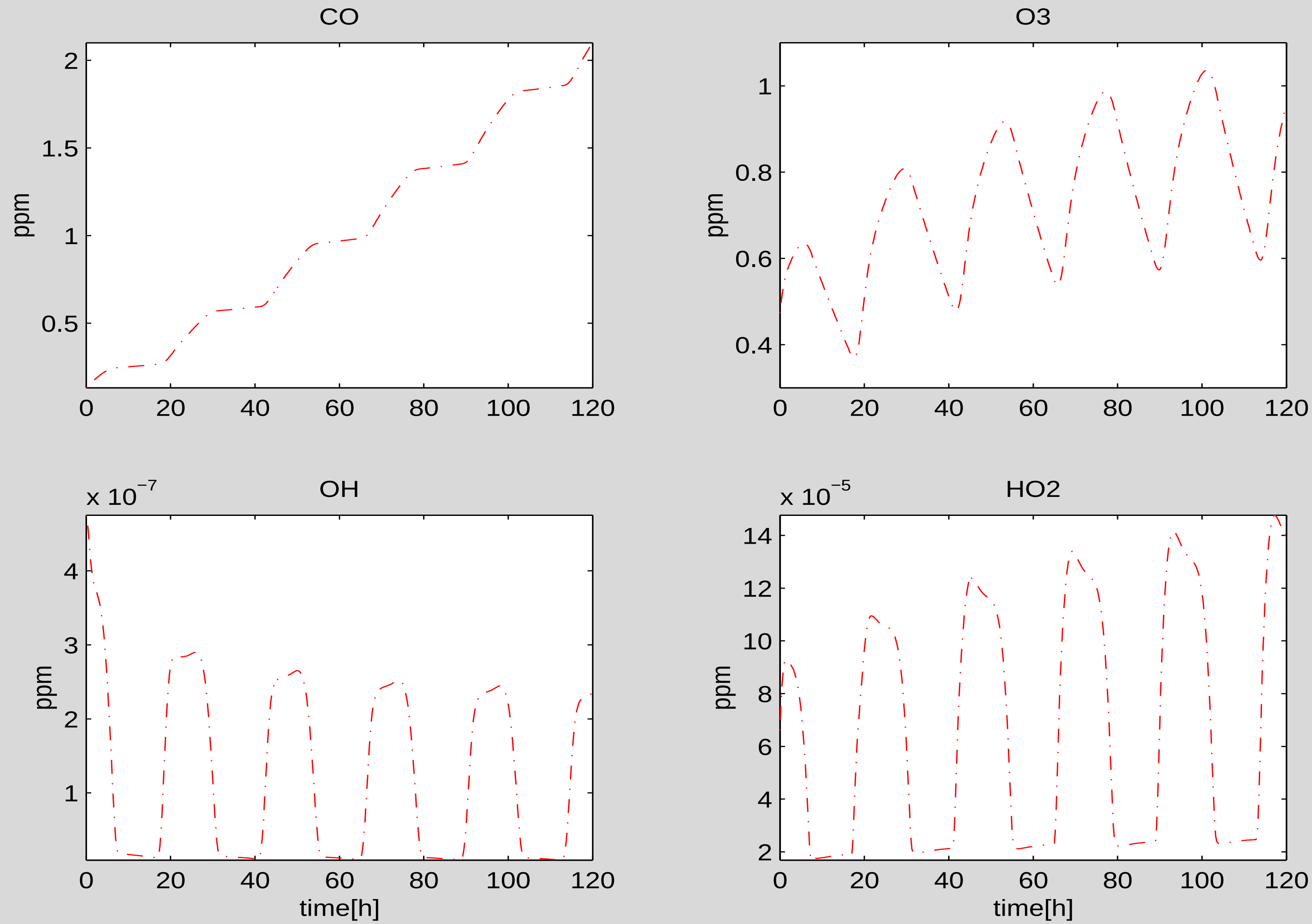
- Jacobian spectrum: $|Re\{\mu\}|_{max} \approx 10^{11}, |Re\{\mu\}|_{min} \approx 10^{-6}$
- Fast/slow subsystems $\mathbf{y} = (\mathbf{y}_s, \mathbf{y}_f), \quad \epsilon = |Re\,\mu_1|/|Re\,\mu_{m+1}|$

$$\frac{d\,\mathbf{y}_s}{d\,t} = \mathbf{f}_s(\mathbf{y}_s, \mathbf{y}_f) \qquad \epsilon\frac{d\,\mathbf{y}_f}{d\,t} = \mathbf{f}_f(\mathbf{y}_s, \mathbf{y}_f)$$

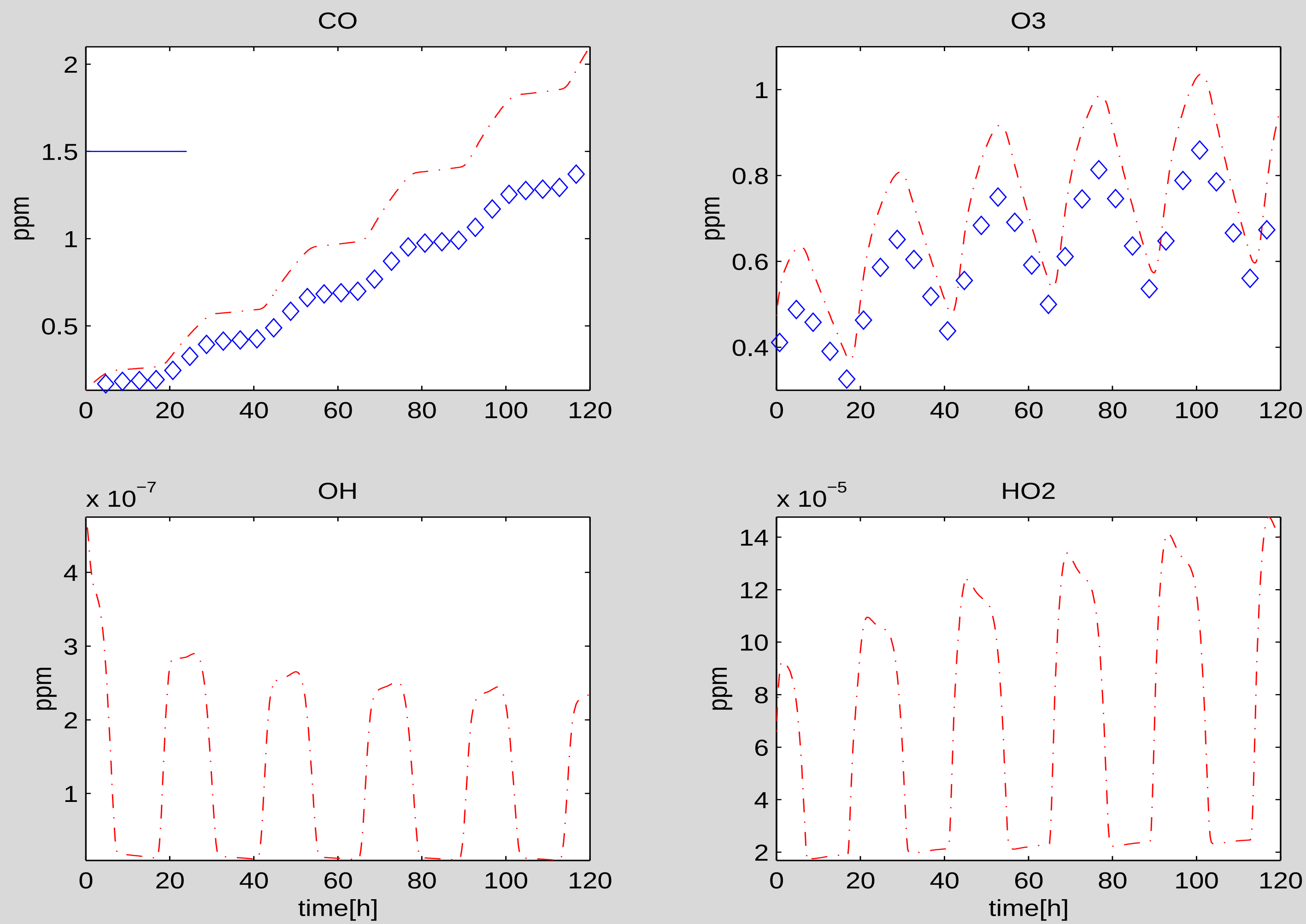
$$\frac{d\,\mathbf{y}_s}{d\,t} = \mathbf{f}_s(\mathbf{y}_s, \mathbf{g}(\mathbf{y}_s)) \qquad 0 = \mathbf{f}_f(\mathbf{y}_s, \mathbf{y}_f)$$

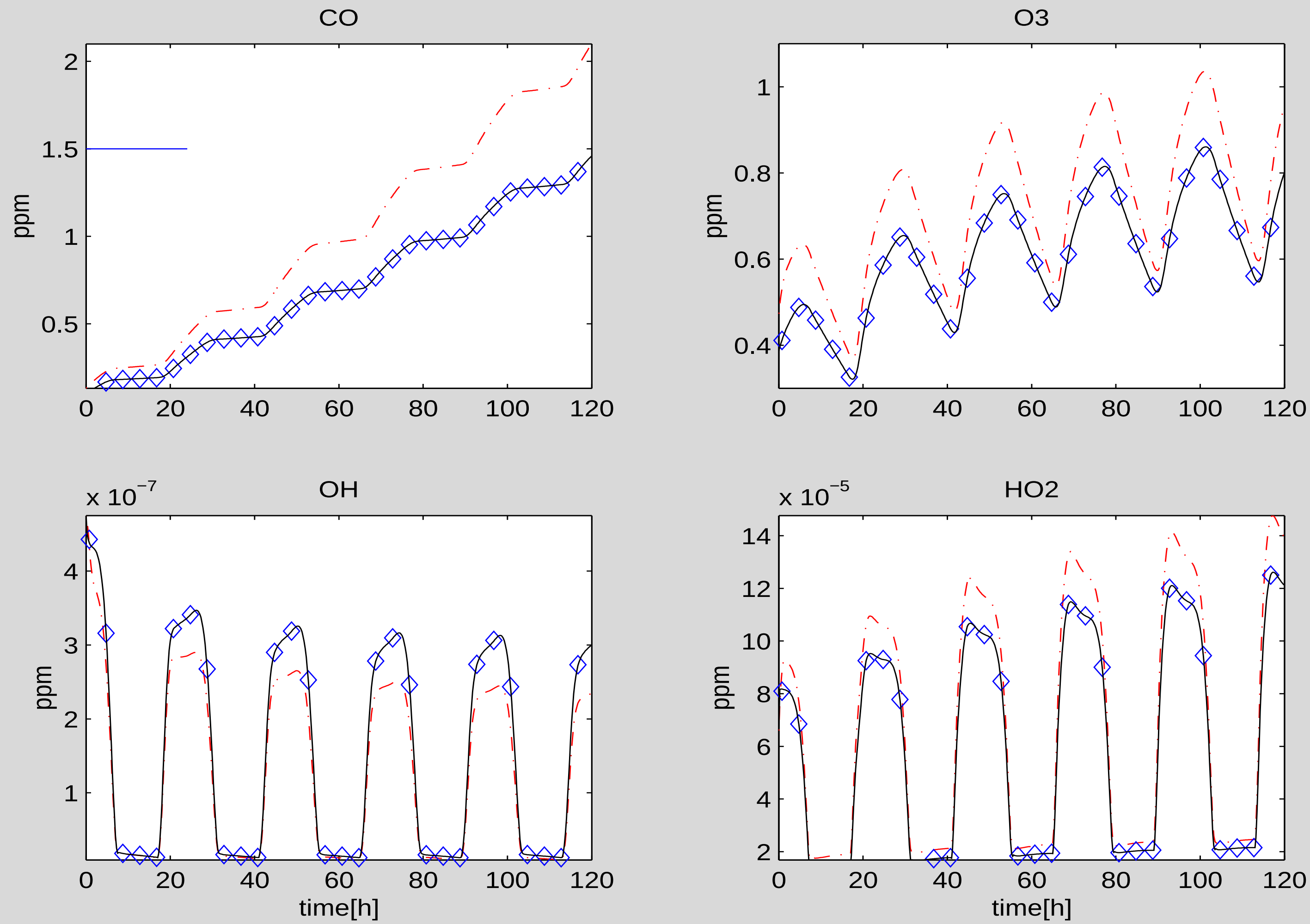
Data assimilation twin experiments

- $[0 - 24]$h assimilation window, followed by a 4 days forecast
- Parameters: initial state, emission rates
- Observations: hourly from 1h to 24h for 59 species
- Optimization algorithm: L-BFGS



CO
O3
OH
HO2
ppm
x 10^−7
x 10^−5
time[h]


CO
ppm
2
1.5
1
0.5
0
20
40
60
80
100
120
O3
ppm
1
0.8
0.6
0.4
OH
x 10^-7
ppm
4
3
2
1
time[h]
HO2
x 10^-5
ppm
14
12
10
8
6
4
2
time[h]



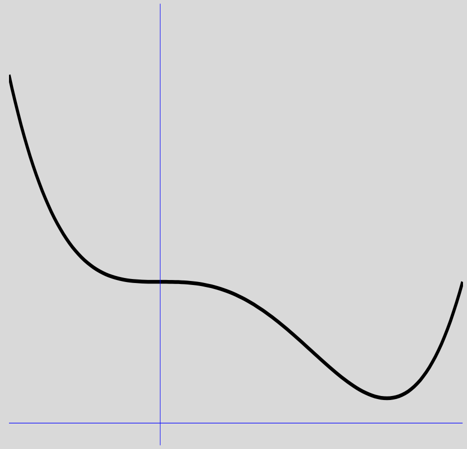
CO
O3
OH
HO2
ppm
x 10^−7
x 10^−5
time[h]

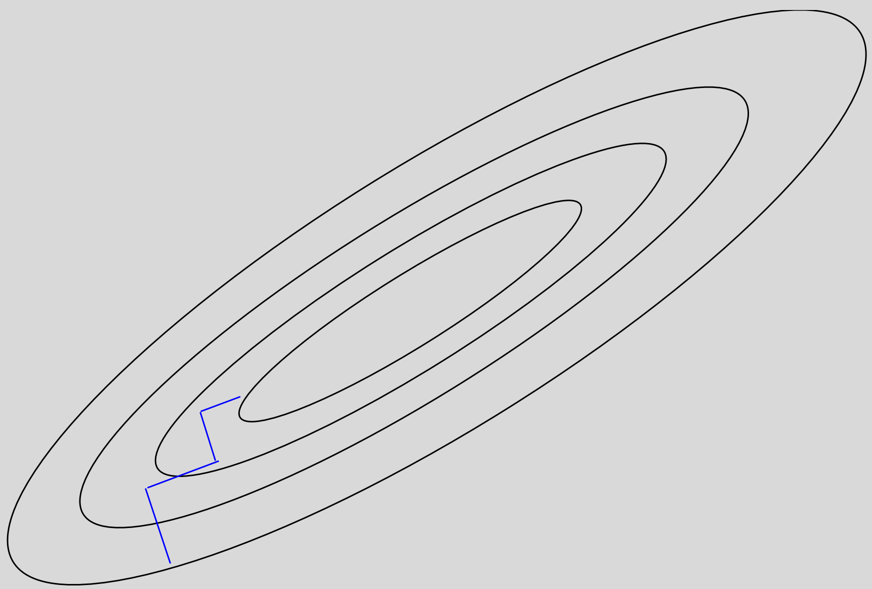
Emissions estimate

Chem species	Emission rate (ppm/min)				
	E_i^{ref}	E_i^{guess}	rel error	E_i^{assim}	rel error
NO	6.94E−5	10.4E−5	0.5	6.93E−5	0.0013
NO2	3.47E−5	5.20E−5	0.5	3.47E−5	0.0007
SO2	3.47E−5	5.20E−5	0.5	3.46E−5	0.0026
HONO	6.94E−7	1.04E−6	0.5	6.94E−7	0.0005
ALK1	8.10E−6	1.21E−5	0.5	8.17E−6	0.0093
ALK2	1.30E−5	1.95E−5	0.5	1.30E−5	0.0004

Numerical method	CPU time (sec)				
	FWD	CADJ	CADJ/FWD	DADJ	DADJ/FWD
RODAS3	1.6	1.9	1.2	3.7	2.3
ROS2	1.1	1.2	1.1	2.1	1.9

The optimization process





- Second order optimality conditions
  $\mathbf{z}^T \nabla^2 \mathcal{J}(\mathbf{y}_0^*)\mathbf{z} > 0, \mathbf{z} \neq 0$
- Hessian large condition-number $\Rightarrow$ ill-conditioned problem
- Hessian-free Newton-CG methods $\nabla^2 \mathcal{J}(\mathbf{y}_0^k)\mathbf{d}^k = -\nabla \mathcal{J}(\mathbf{y}_0^k)$
- Iterative solvers require only Hessian/vector product

Second order information via adjoint modeling

- Finite difference approximation

$$\nabla^2 \mathcal{J}(\mathbf{y}_0)\mathbf{v} \approx \frac{\nabla \mathcal{J}(\mathbf{y}_0 + \epsilon \mathbf{v}) - \nabla \mathcal{J}(\mathbf{y}_0)}{\epsilon}$$

- Second order adjoint model (exact)
  - ▷ Forward over reverse procedure

$$\mathbf{y}_0 \xrightarrow{\mathbf{M}} \mathcal{J}(\mathbf{y}_0) \xrightarrow{\mathbf{M}^*} \nabla \mathcal{J}(\mathbf{y}_0) \tag{2}$$

$$\mathbf{g} : R^N \rightarrow R^N, \mathbf{g}(\mathbf{y}_0) = \nabla \mathcal{J}(\mathbf{y}_0) \tag{3}$$

$$\nabla^2 \mathcal{J}(\mathbf{y}_0)\mathbf{v} = \left(\frac{\partial \mathbf{g}}{\partial \mathbf{y}_0} \right) \mathbf{v} \tag{4}$$

  - ▷ $CPU(SOA) \sim 7 - 12 \times CPU(FWD)$
  - ▷ Automatic differentiation packages: ADIFOR, TAF (TAMC), ODYSSEE
  - ▷ Continuous equations for the second order adjoint may be derived

Case study: Shallow-Water Equations

- Model the flow of an incompressible fluid whose depth is small with respect to the horizontal dimension

- Widely used in oceanographic and atmospheric studies

- The SW equations in a Cartesian system

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv + \frac{\partial \phi}{\partial x} = 0 \quad (5)$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu + \frac{\partial \phi}{\partial y} = 0 \quad (6)$$

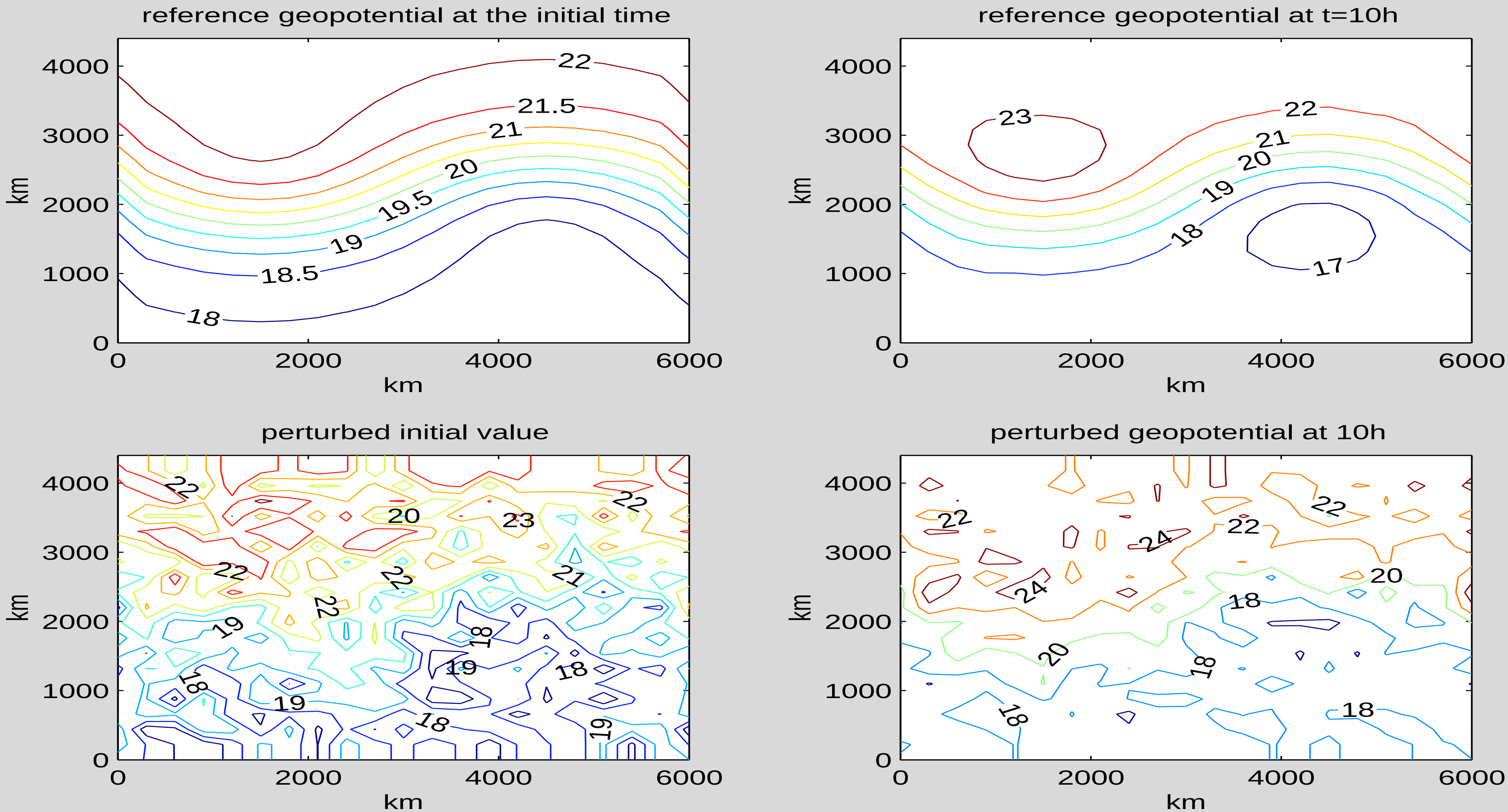
$$\frac{\partial \phi}{\partial t} + \frac{\partial u\phi}{\partial x} + \frac{\partial v\phi}{\partial y} = 0 \quad (7)$$

- The state vector is $\mathbf{y} = (u, v, \phi)$, where (u, v) are the components of the horizontal velocity, $\phi = gh$ is the geopotential and f is the Coriolis parameter.

- Control parameters: initial conditions, $dim(\mathbf{y}) = 1083$

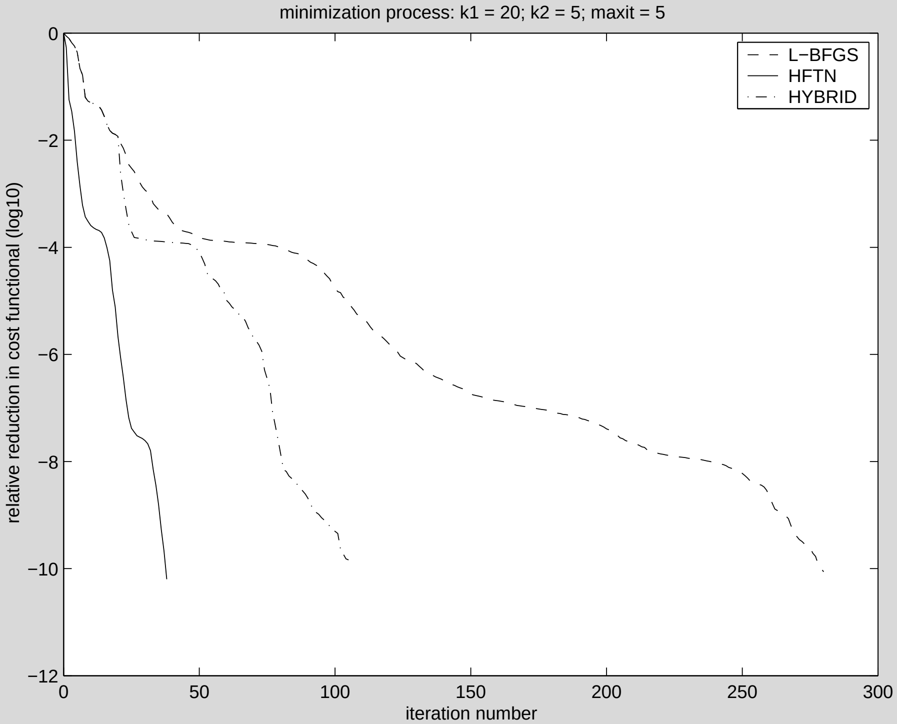
- Assimilation window 10 hours

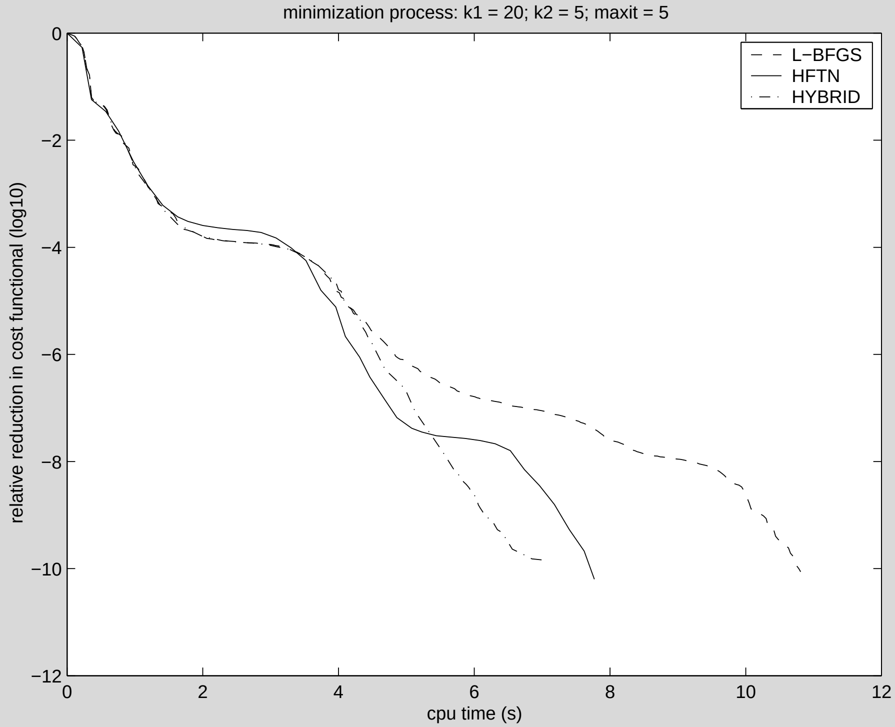
Twin experiments data assimilation



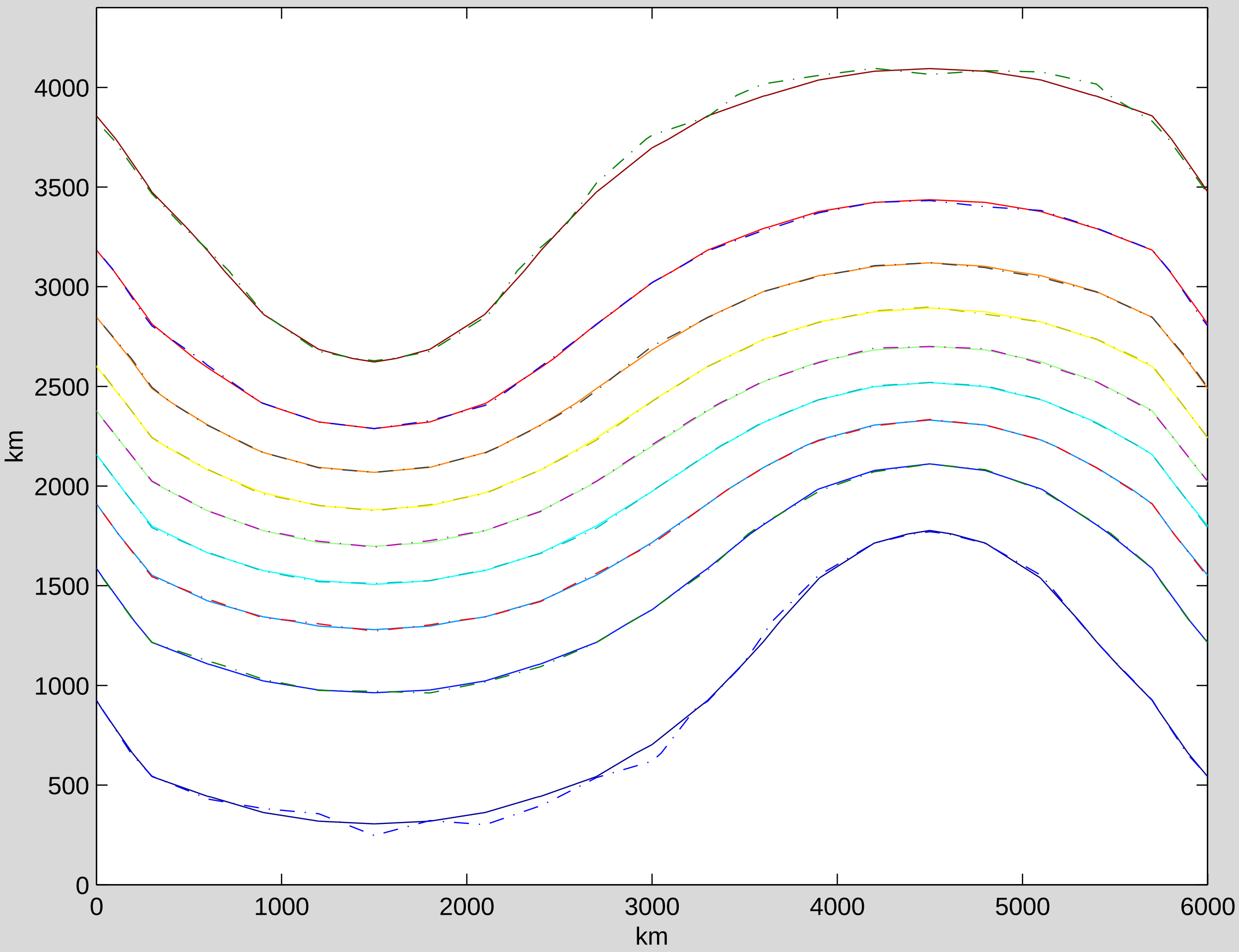
Hessian Condition Number $\sim O(10^5)$ (Lanczos-ARPACK)

H-F Newton vs L-BFGS vs HYBRID





reference (solid line) and retrieved (dotted line) geopotential at the initial time



”We interrupt this program to bring you an emergency alert from the National Broadcast Emergency Center. Hurricane Andrew has unexpectedly shifted five degrees south. Andrew is expected to strike South Dade within minutes. All South Dade residents should take immediate cover! This is an emergency alert!”

”We interrupt this program to bring you an emergency alert from the National Broadcast Emergency Center. Hurricane Andrew has unexpectedly shifted five degrees south. Andrew is expected to strike South Dade within minutes. All South Dade residents should take immediate cover! This is an emergency alert!”

THE EMERGENCY ALERT THAT CAME TOO LATE

The Adaptive Observations Problem

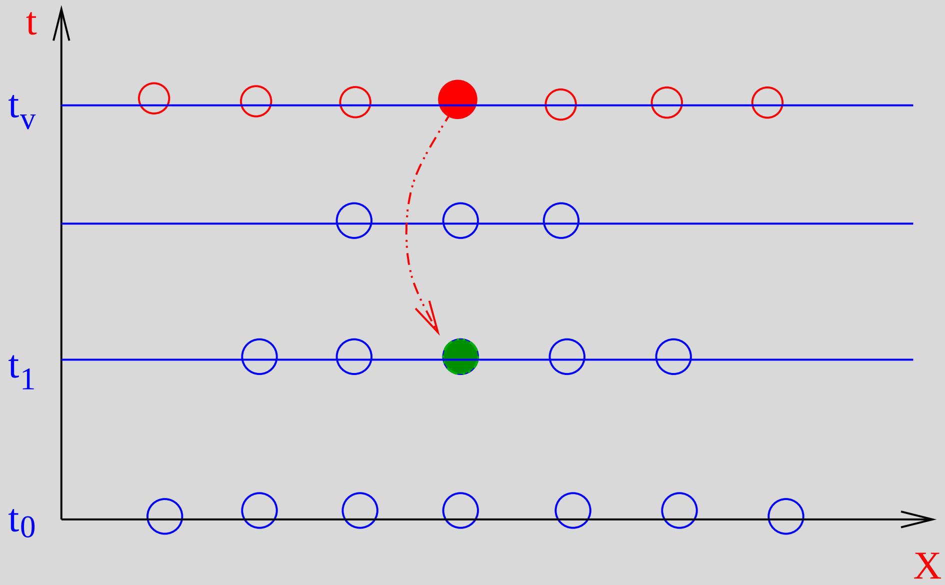
A severe weather event is predicted over area $\mathcal{D}_v$ at a future time $t_v > t_0$

Question: find optimal locations where additional observational resources must be deployed prior to the event, $t_0 < t^o < t_v$, to improve the forecast

4D-Var adaptive observational problem

- Assume that adaptive observations may be taken at instants $t_i, i = \overline{1, I}$
- Question: Find an optimal observational path $\mathcal{O}^a = \{\mathcal{O}^a_1, \mathcal{O}^a_2, \ldots \mathcal{O}^a_I\}$

Targeting strategies



- Field experiments (FASTEX, NORPEX, WSRP) revealed that the efficiency of adaptive observations relies on the identification of the areas where the errors in the initial conditions are large and/or are rapidly growing as well as the characteristics of the data assimilation system

- Current adjoint-based targeting strategies include

  ▷ *Singular vectors techniques*

  ▷ *Gradient (sensitivity) techniques*

Singular vectors approach

Tangent-linear model

$$\frac{d\delta\mathbf{y}}{dt} = \mathbf{M}_\mathbf{y}(\mathbf{y})\delta\mathbf{y} \Rightarrow \mathbf{P}\delta\mathbf{y}(t) = \mathbf{P}\mathbf{L}(t, t_i)\delta\mathbf{y}(t_i)$$

where $\mathbf{P}$ denotes the projection operator on $\mathcal{D}_v$

The directions characterized by maximum growth $\|\mathbf{P}\delta\mathbf{y}(t)\|/\|\delta\mathbf{y}(t_i)\|$ are the singular vectors $\nu_j(t_i)$ associated with the largest singular values σ_j

$$[\mathbf{PL}]^* [\mathbf{PL}] \nu_j = \sigma_j^2 \nu_j \tag{8}$$

This eigenvalue problem must be solved in the optimization interval $t_v - t_i$ for each $t_i, i = 1, 2, \ldots I$.

(!) Computationally expensive ... ! ... but ...

the implementation is feasible using the adjoint modeling and iterative eigenvalue algorithms (e.g. Lanczos → ARPACK package)

Target area definition using N leading singular vectors

Define the sensitivity function

$$F_N(t_i, \mathbf{x}) = \sum_{j=1}^{N} \left(\frac{\sigma_j}{\sigma_1} \right) \|\nu_j(t_i, \mathbf{x})\|$$

Observations at t_i are taken at locations $\mathbf{x}$ where $F_N(t_i, \mathbf{x})$ has largest values

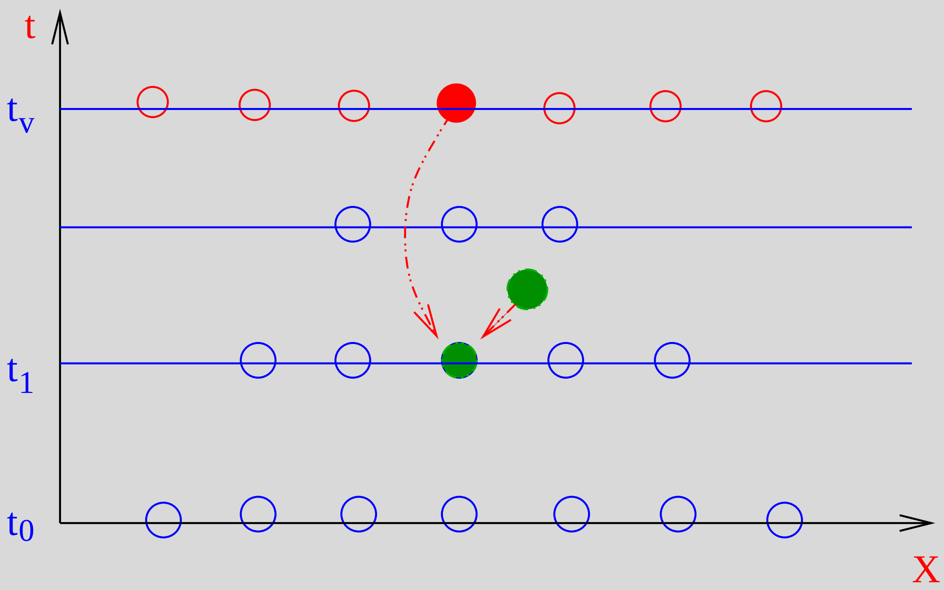
Gradient Approach

- A functional $\mathcal{J}_v(\mathbf{y}_v)$ is defined in terms of the forecast at t_v in $\mathcal{D}_v$.
- Define the sensitivity function

$$F_v(t_i, \mathbf{x}) = \|\nabla_{\mathbf{y}_i(\mathbf{x})} \mathcal{J}_v(\mathbf{y}_v)\|$$

- Observation locations at t_i: first n_i locations $\mathbf{x}$ where $F_v(t_i, \mathbf{x})$ has largest values.
- Target area identification proceeds backward in time from t_I to t_1
- An efficient evaluation of all $\nabla_{\mathbf{y}_i} \mathcal{J}_v(\mathbf{y}_v), i = 1, 2, \ldots I$ is obtained through a single adjoint model integration.

Practical Issues



- the 'a priori' evaluation of the sensitivity vector is completely ignorant of any existing observations.
- 4D-Var data assimilation takes into consideration all available observations in the assimilation window
- Optimal adaptive observations at t_i have been determined assuming that these are the only available observations.

Interaction between observations must be considered!

Interaction between adaptive observations

- *Daescu and Carmichael (JAS 2003), Daescu and Navon (MAP 2003)*: our new sensitivity approach takes into consideration the interaction between adaptive observations at distinct instants in time and the interaction with routine observations (a priori known locations).

- A periodic update of the adjoint sensitivity field at t_i is performed which takes into consideration all observations already located at t_{i+1}

$$\mathcal{O}_{i+1} = \mathcal{O}^f \cup \mathcal{O}_I^a \cup \ldots \cup \mathcal{O}_{i+1}^a$$

- We introduce a new sensitivity field associated to the observations set $\mathcal{O}_{i+1}$

$$F_i(\mathbf{x}) = \|\nabla_{\mathbf{y}_i(\mathbf{x})} \mathcal{J}_{\mathcal{O}_{i+1}}\|$$

and periodically update the sensitivity field F_v at each targeting instant

$$F_v(\mathbf{x}) = F_v(\mathbf{x}) \left(1 + \alpha \frac{F_i(\mathbf{x})}{F_v(\mathbf{x})}\right)^{-1}$$

- The updated sensitivity field is inversely proportional to the relative value of the sensitivity field provided by the set of observations that have been already located.

- New observations at t_i are located in regions where the sensitivity of $\mathcal{J}_v$ to the model state is large *and* little additional information may be obtained from previously located observations.

- The additional computational cost to evaluate the sensitivity F_i is given by the computational cost of a backward integration in the assimilation window $[t_0\ T]$.

The 2D-global shallow water model

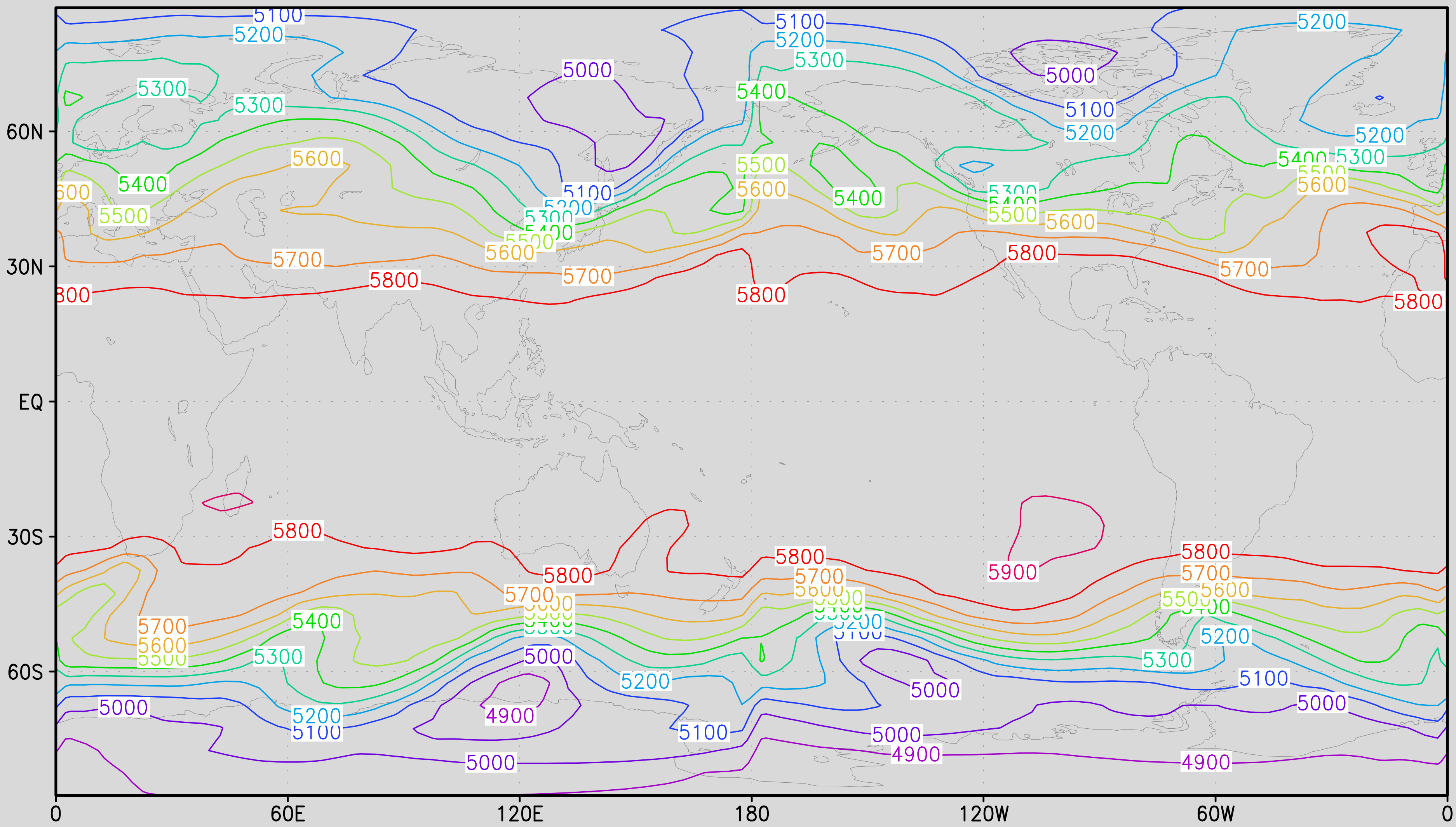
Spatial discretization: $72 \times 36\,(5^\circ \times 5^\circ)$ grid, $\mathbf{y} = (\mathbf{h}, \mathbf{u}, \mathbf{v}), dim(\mathbf{y}) = 7776$

Time integration 24h, explicit Turkel-Zwas finite difference scheme

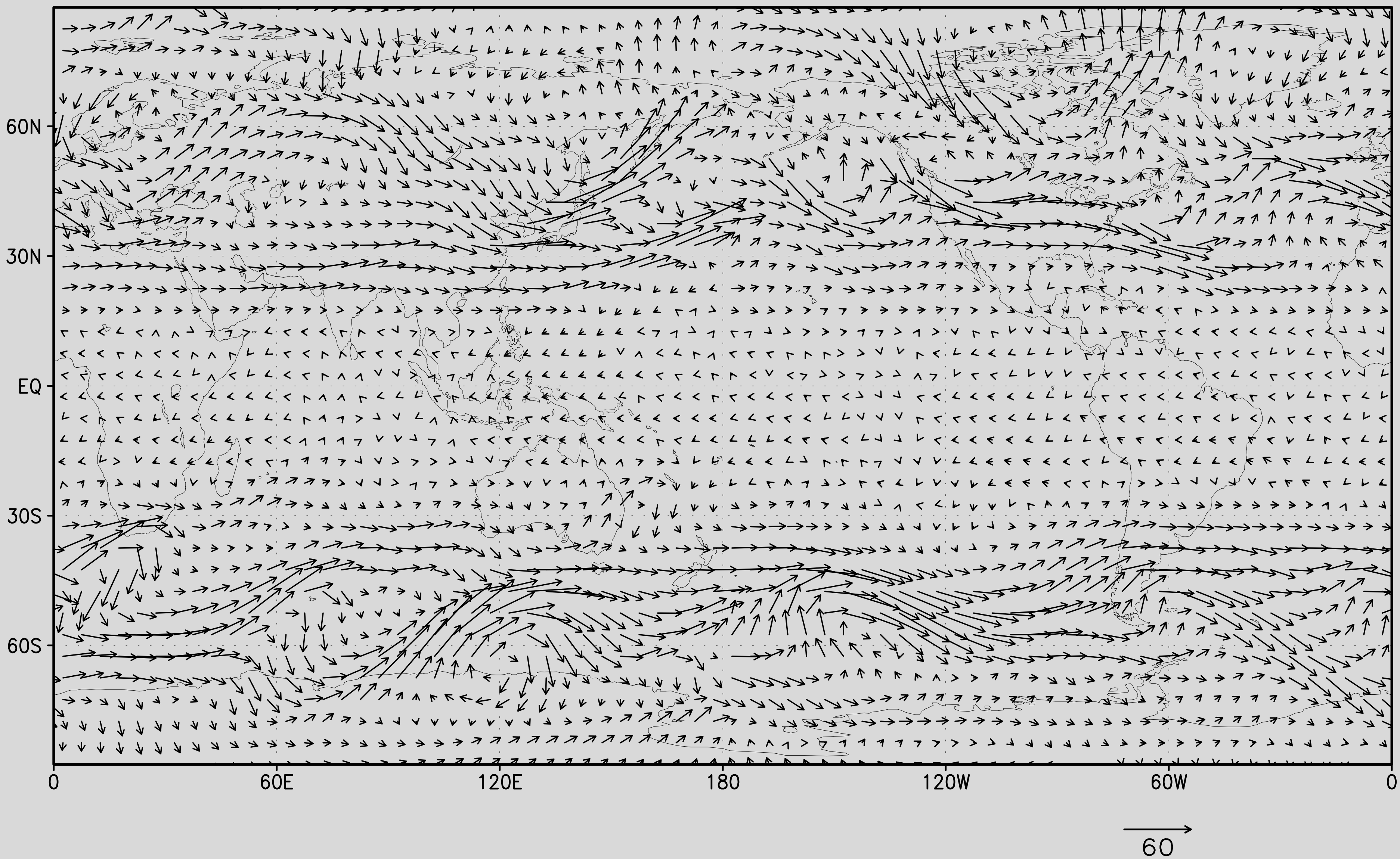
Data from NASA-DAO *Goddard Earth Observing System (GEOS-3)* model $(1.25^\circ \times 1^\circ)$ is used to specify the initial conditions.

- 4D-Var data assimilation is performed in the assimilation window $[0-6]$h in the twin experiments framework.
- A background field is prescribed using a 5% *random perturbation* in the reference velocities field.
- Observations at fixed location are prescribed at $t = 6$h *only*, on a $20^\circ \times 20^\circ$ *grid subdomain*.

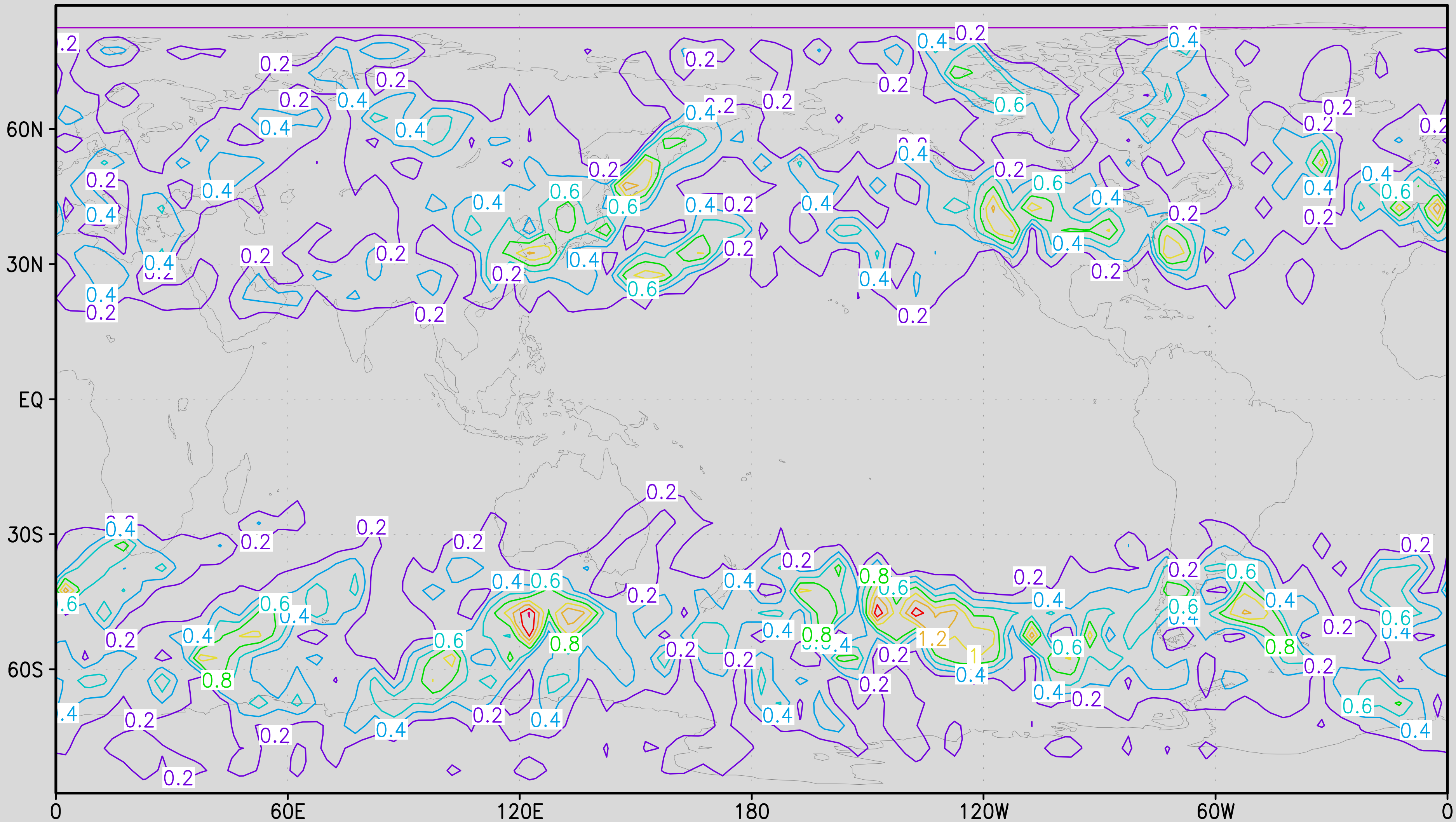
500hPa geopotential height at t=0h



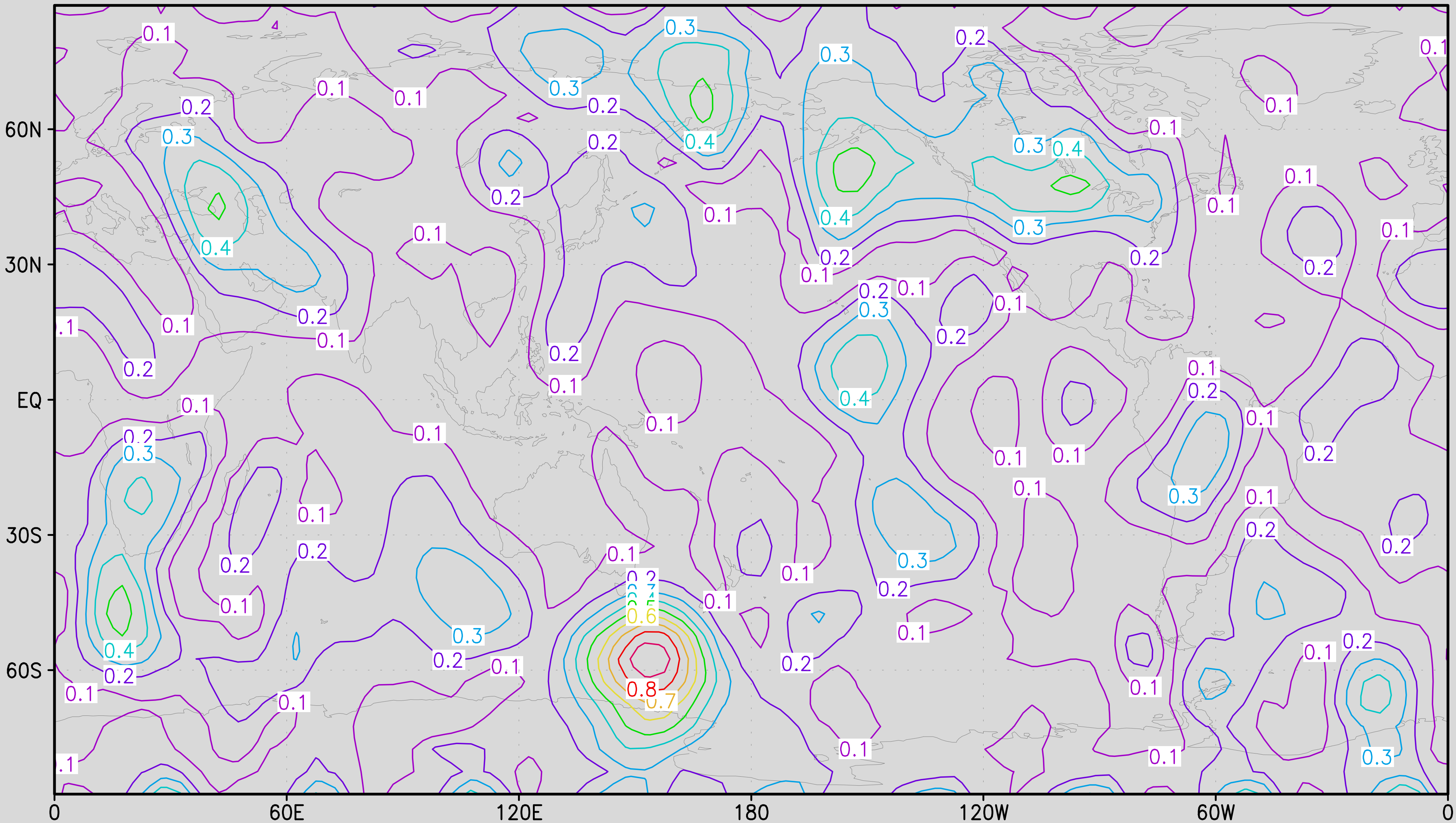
velocities field at initial time t=0h



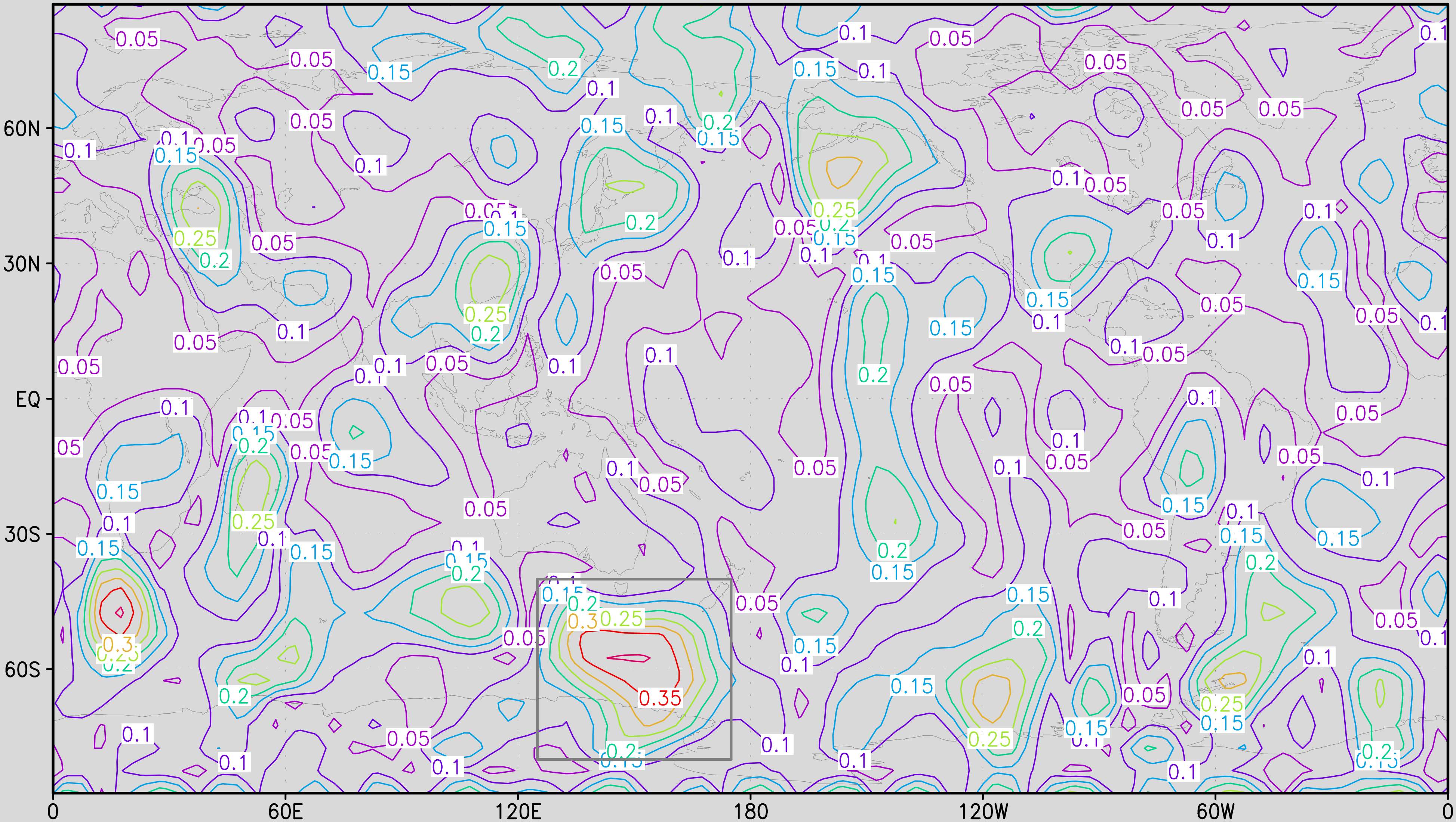
initial guess error in total energy norm at t=0h



initial guess error in total energy norm at t=24h



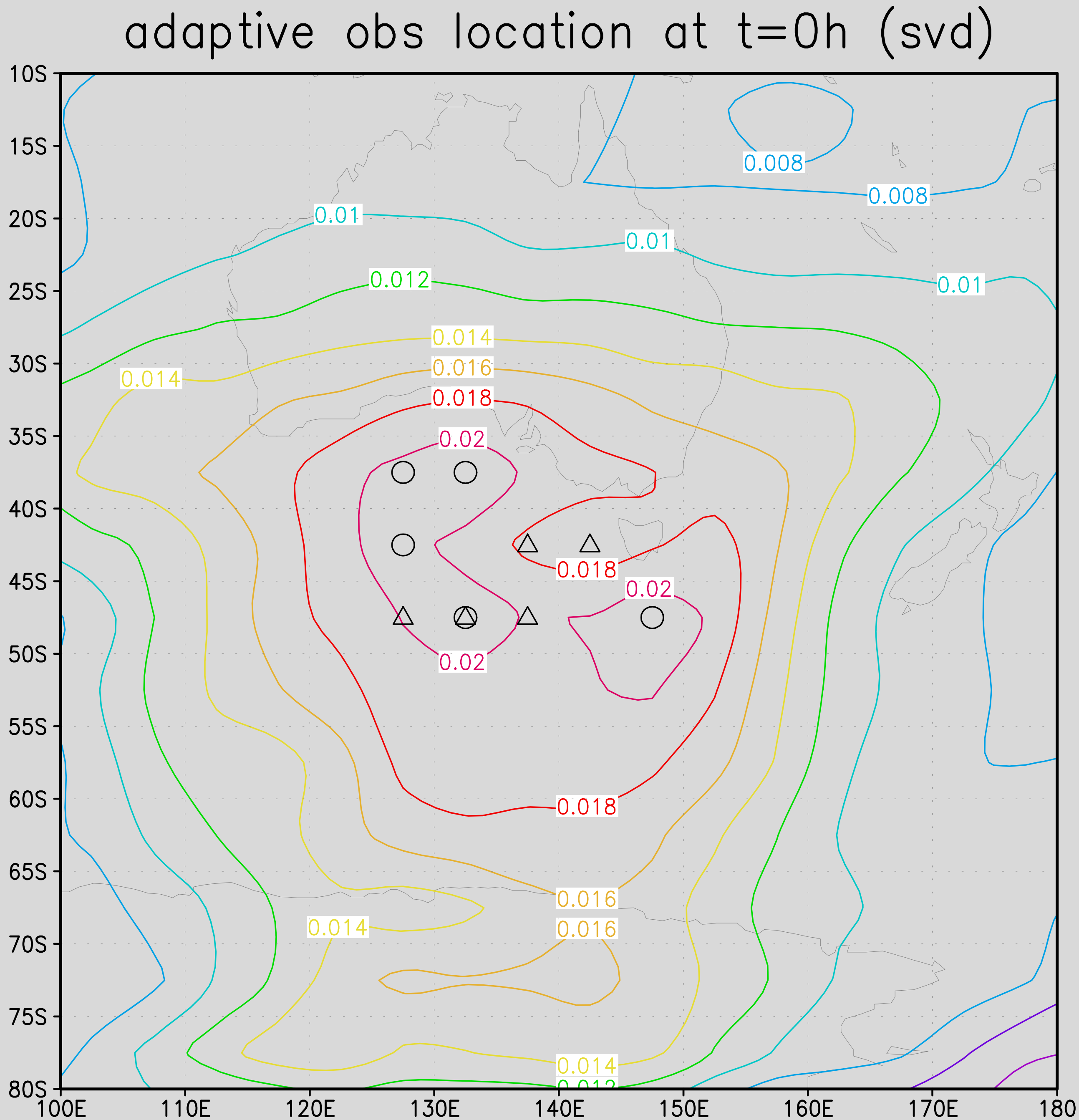
error (t.e. norm) of 24h forecast with fixed obs only



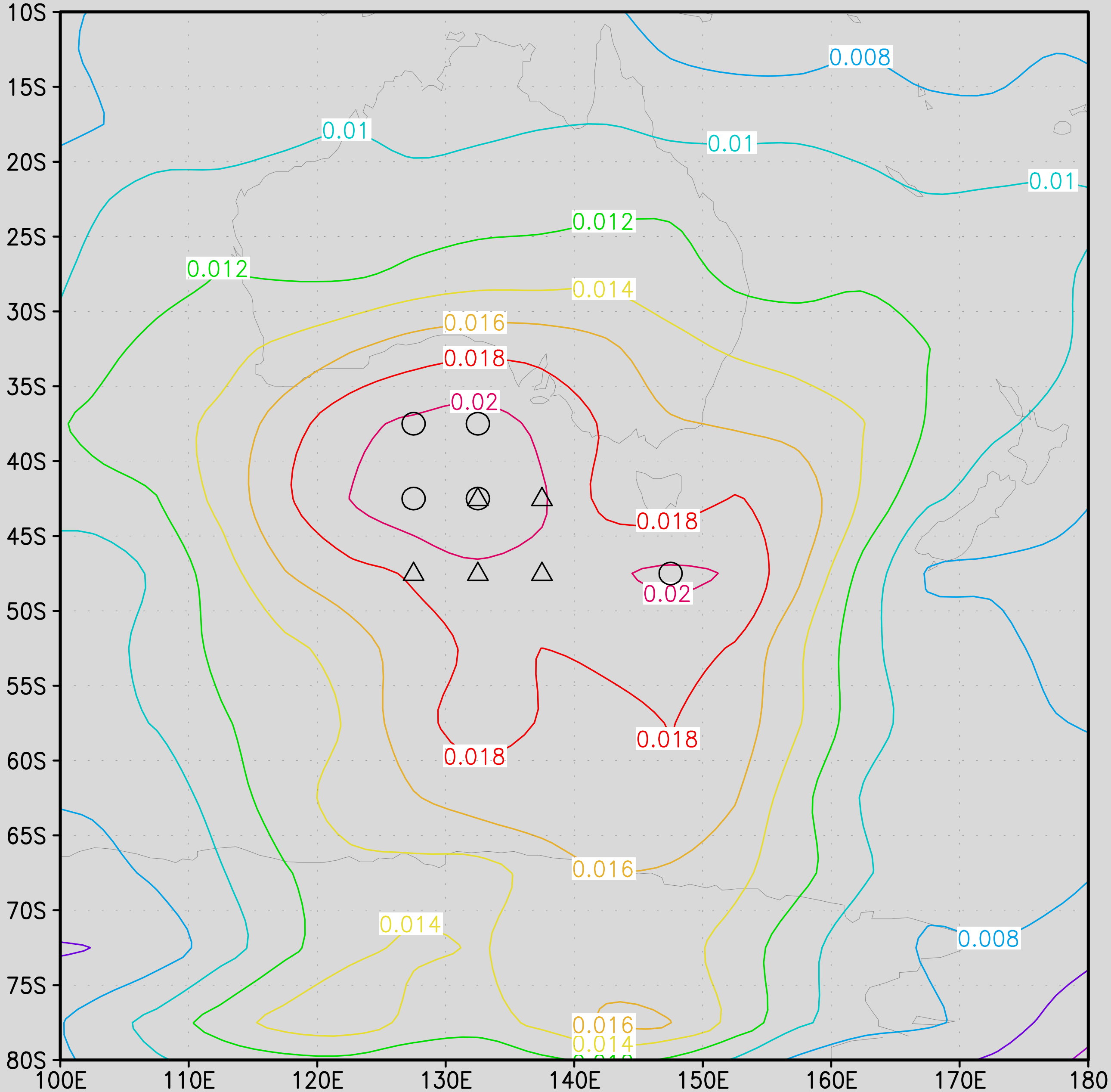
Numerical results

- A comparative analysis of targeting strategies using the leading singular vectors (first 10), gradient sensitivity, and gradient sensitivity with interaction between adaptive observations is performed when only the adaptive observations are included in the data assimilation process.

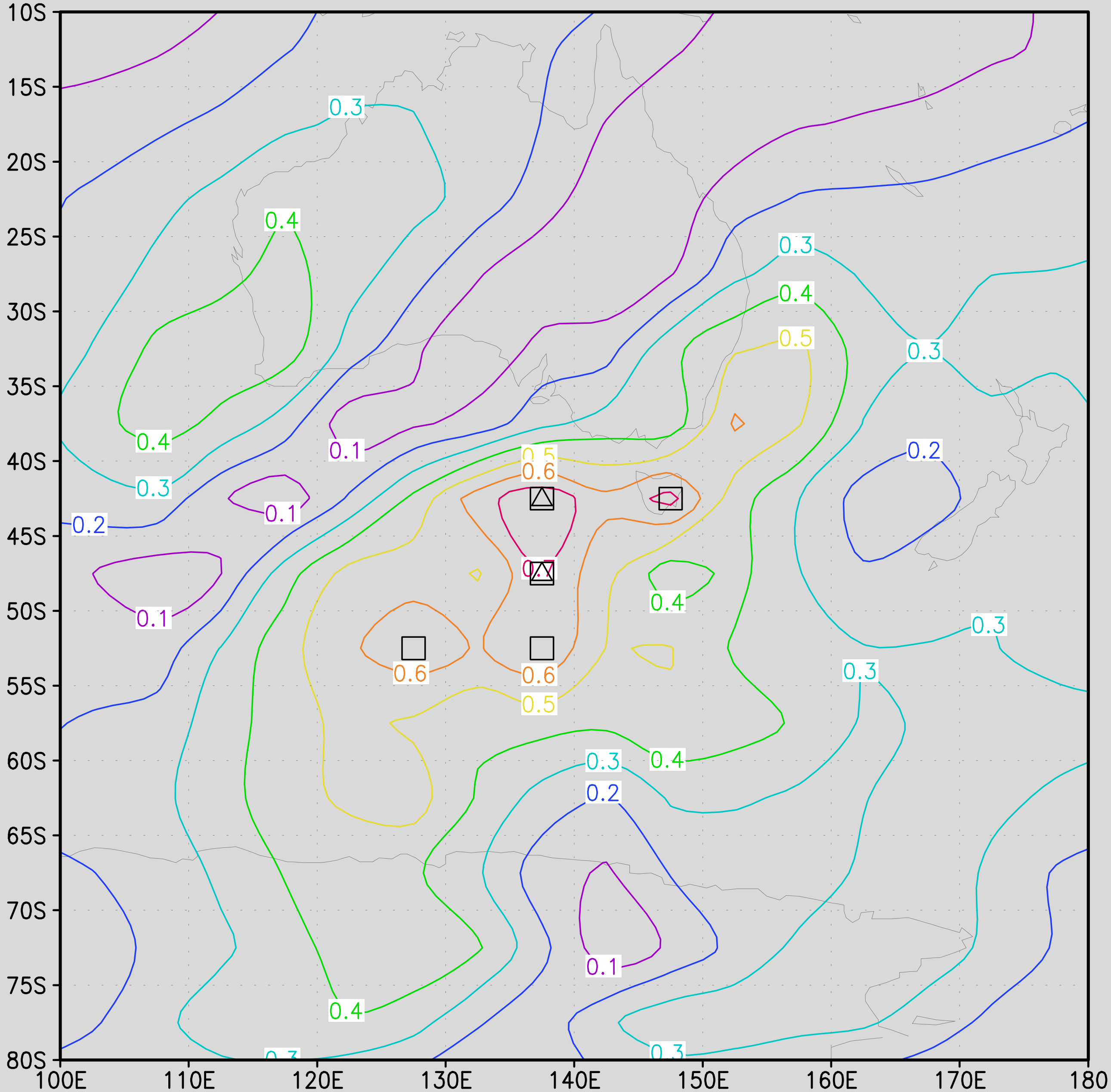
- For each method, during the data assimilation process to minimize the cost functional $\mathcal{J}$ (*over the entire domain*) we also *monitor the evolution* of the forecast error reduction $\mathcal{J}_v(\mathbf{y}_0^\star)/\mathcal{J}_v(\mathbf{y}_0)$ at t_v over $\mathcal{D}_v$.


adaptive obs location at t=0h (svd)
10S
15S
20S
25S
30S
35S
40S
45S
50S
55S
60S
65S
70S
75S
80S
100E
110E
120E
130E
140E
150E
160E
170E
180
0.008
0.008
0.01
0.01
0.01
0.012
0.014
0.014
0.016
0.018
0.02
0.018
0.02
0.02
0.018
0.016
0.014
0.016
0.014

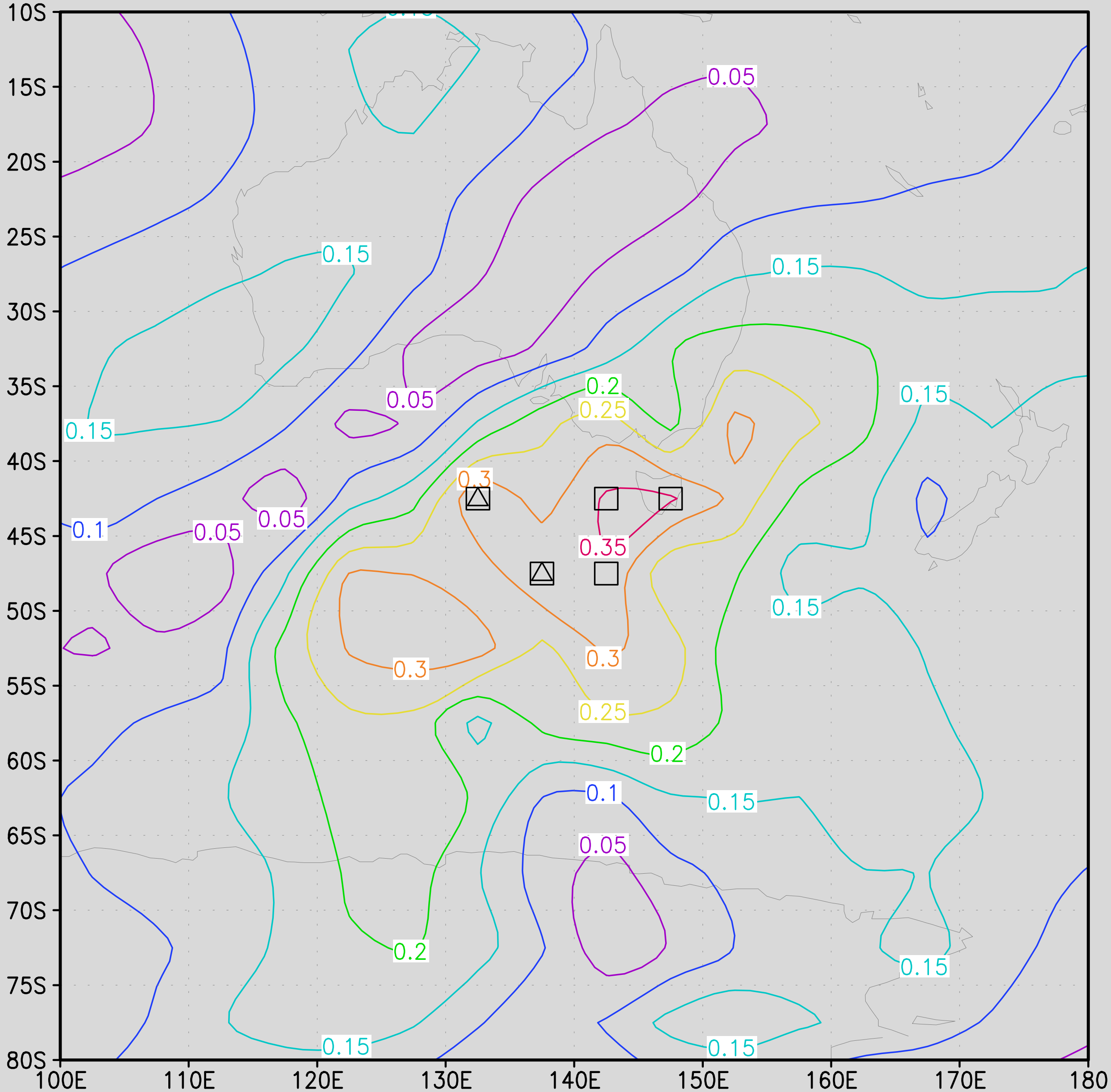
adaptive obs location at t=2h (svd)



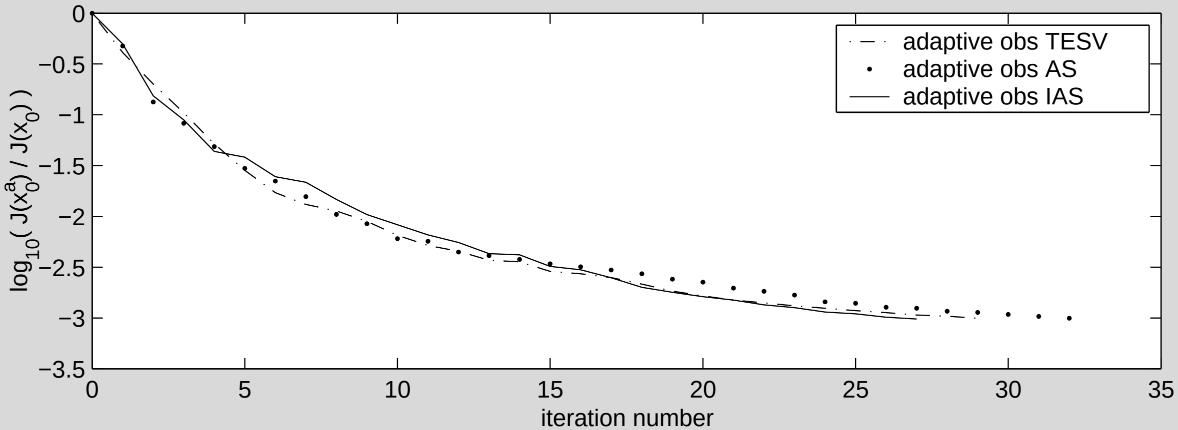
adaptive obs location at t=0h



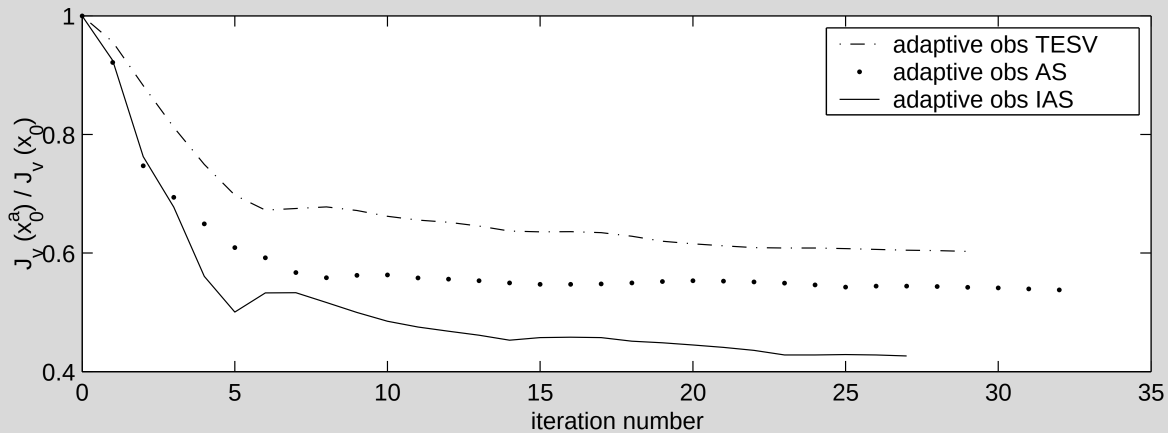
adaptive obs location at t=2h



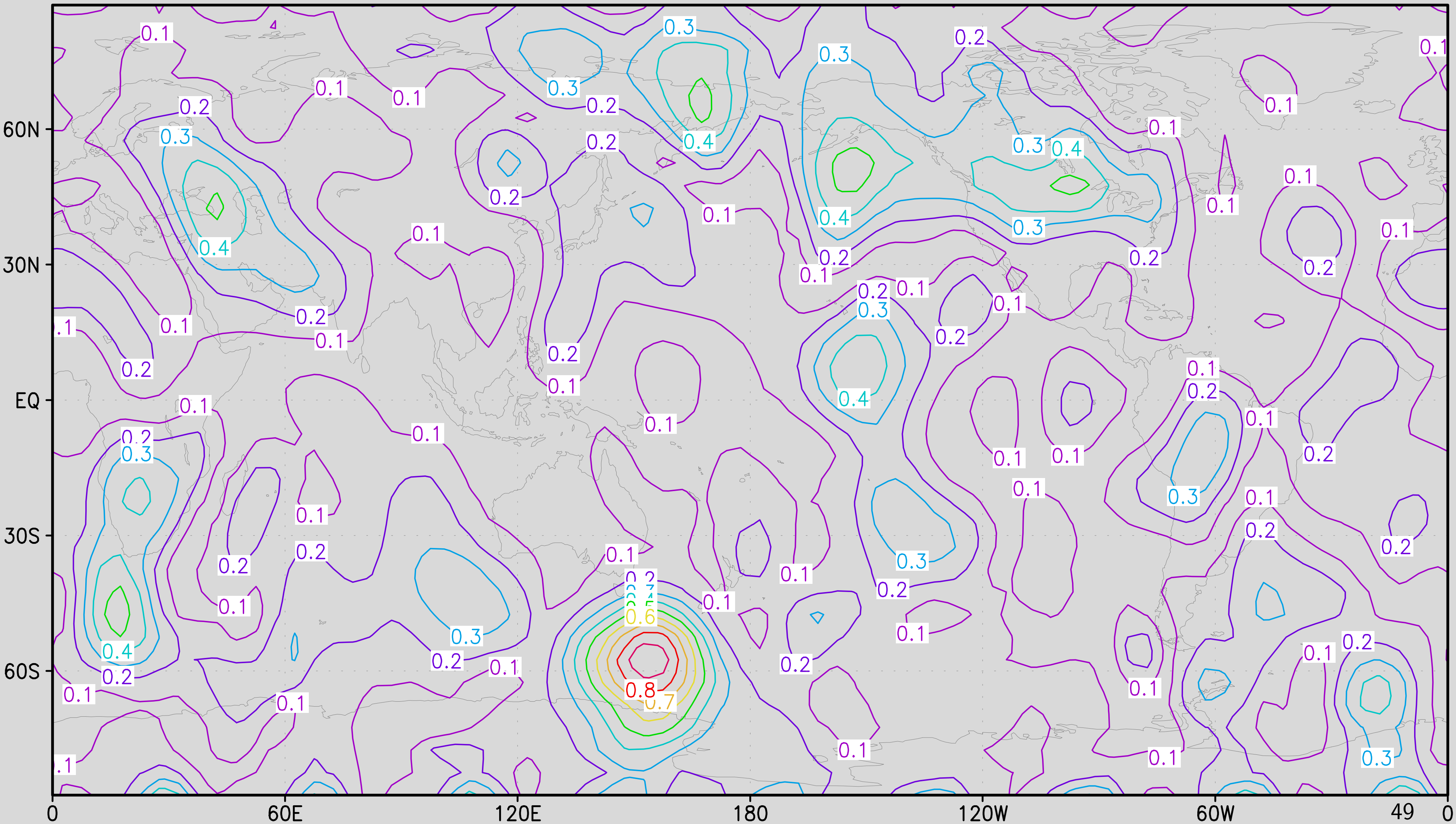
the minimization process of the cost functional J



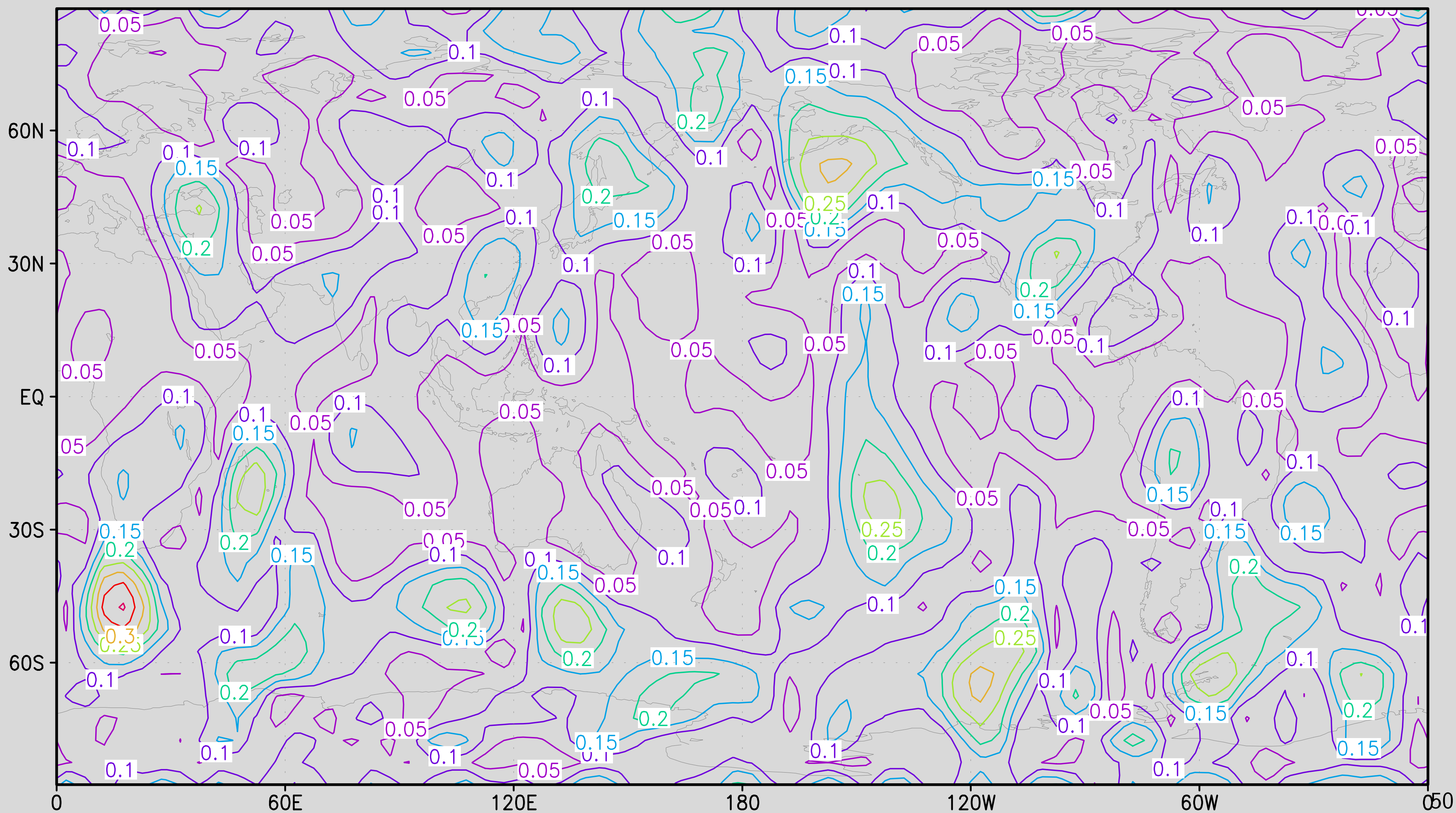
forecast error reduction $J_v(x_0^a) / J_v(x_0)$ during the minimization process of J



initial guess error in total energy norm at t=24h



error of 24h forecast with fixed and adaptive obs (inter-adj)



Future research directions

- 4D-Var data assimilation for comprehensive air pollution models
- Extended field of applications (e.g. shape design optimization)
- Robust optimization algorithms with approximate gradients
- Model reduction techniques
- Observational network design for optimal prediction