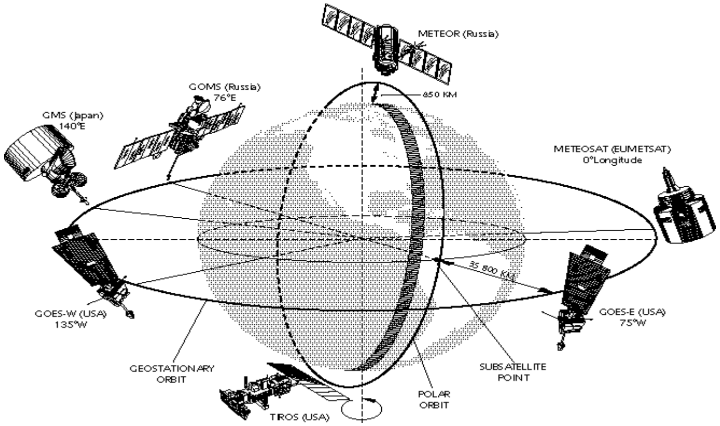
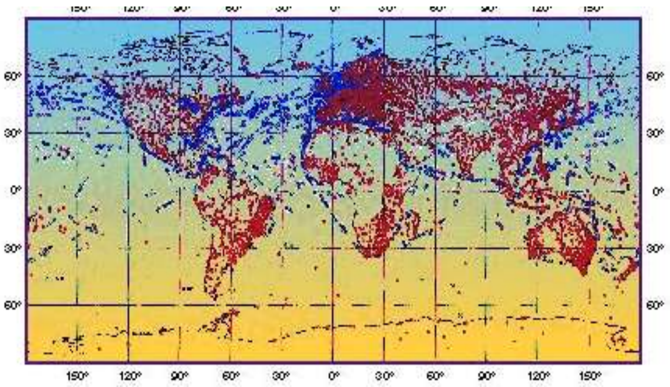


Klimova Ekaterina, *Kilanova Natalia,*
Zudin Andrey

Institute of Computational Technologies Russian Academy of Sciences, Siberian Branch

Methods of an environment assessment by the use of mathematical model and the observational data



klimova@ict.nsc.ru

Plan of report

<u>The report consists of three parts:</u>

I. Data-processing technologies in problems of an environment monitoring.

II. The problems of practical realisation of the data assimilation algorithms.

III. The results of some numerical experiments.

Introduction

- ***Data assimilation is a problem of an optimum estimation reception of investigated geophysical fields using the observational data and the mathematical model describing dynamics of fields on time.***

- ***Such problems are considered now for the description of processes in atmosphere, ocean, and also distributions of polluting substances in the environment.***

- ***The purpose of the data assimilation is the initial fields for the forecast, and more the general - the description of behaviour on time of investigated fields, climate studying etc.***

Data assimilation problem

- ***Two approaches are the most popular:***

  4DVAR and Kalman filter

- ***Advantages of Kalman filter:***

  algoritm allows to calculate simultaneously
  with estimated fields matrixes of estimation errors covariances

- ***The problem of the forecast error covariances calculation in the case of nonlinear model may be solved with the help of extended Kalman filter or ensemble Kalman filter.***
- ***In the first case for the forecast error covariances calculation linearised model is used, in the second case the matrix of the forecast error covariances is calculated on ensemble of forecasts.***

Ensemble Kalman Filter

In the ensemble approach we have an ensemble of N initial data and N forecasts:

$$x_i^f = M[x_i^a(t-1)] + q_i, \quad i = 1,\ldots,N$$
$$q_i \sim N(0,Q)$$

On the ensemble of the observations the ensemble of analyses is calculated:

$$y_i^0 = y^0 + r_i, i = 1,\ldots,N$$
$$r_i \sim N(0,R)$$
$$x_i^a = x_i^f + K(y_i^0 - Hx_i^f), \quad i = 1,\ldots,N$$
$$K = P^f H^T (HP^f H^T + R)^{-1}$$

Ensemble Kalman Filter

Covariance matrix P is calculated by the formulas:

$$P^f H^T = \frac{1}{N-1}\sum_{i=1}^{N}(x_i^f - \overline{x}_i^f)(Hx_i^f - H\overline{x}_i^f)^T,$$

$$HP^f H^T = \frac{1}{N-1}\sum_{i=1}^{N}(Hx_i^f - H\overline{x}_i^f)(Hx_i^f - H\overline{x}_i^f)^T.$$

<u>Let's note that:</u>
- it is supposed, that the deviation from an average behaves as a deviation from "true value";
- the rank of matrix P estimated on ensemble is equal to number of members of ensemble.

Ensemble Kalman filter

The main problems of Ensemble Kalman Filter realisation are:

- ***generation of ensemble of initial fields with the given properties;***
- ***the limited number of ensemble members – so there is a problem of the description of all leading eigenvectiors of the model operator; therefore in a number of papers Ensemble Kalman Filter is offered to be realised locally;***
- ***at realisation of Ensemble Kalman Filter locally there are problems on borders of subareas and balance between analyzed variables;***

Ensemble Kalman Filter

- ***on the big distances «false covariances» arise which are necessary for suppressing;***

- ***Ensemble Kalman Filter does not solve a nonlinearity problem; on a forecast step estimation P on nonlinear model is calculated, but the analysis step corresponds to a linear estimation problem;***

- ***Ensemble Kalman Filter has no fundamental theoretical justification for nonlinear dynamic system with statistics which is not Gaussian.***

Ensemble Kalman Filter

Ensembles of initial fields and forecasts:

$$x_f^{0(i)} = \hat{x} + \Delta x_0^{(i)}, \quad i = 1,\ldots,N$$

$$x_{k+1}^{f(i)} = M_k(x_k^{a(i)}) + \xi_k^{(i)},$$

Estimation of covariance matrix:

$$P_k^f H_k^T = \frac{1}{N-1}\sum_{i=1}^{N}(x_k^{f(i)} - \overline{x_k^{f(i)}})(H_k x_k^{f(i)} - \overline{H_k x_k^{f(i)}})^T,$$

$$H_k P_k^f H_k^T = \frac{1}{N-1}\sum_{i=1}^{N}(H_k x_k^{f(i)} - \overline{H_k x_k^{f(i)}})(H_k x_k^{f(i)} - \overline{H_k x_k^{f(i)}})^T,$$

Ensemble of «analyses»:

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1},$$

$$x_k^{a(i)} = x_k^{f(i)} + K_k(y_k^{0(i)} - H_k x_k^{f(i)}),$$

$$y_k^{0(i)} = y_k^0 + r_k^{(i)}$$

$$\overline{\mathrm{x}_{\mathrm{k+1}}} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{x}_{\mathrm{k+1}}^{(\mathrm{i})}$$

Ensemble pi-algorithm

Let the estimation error be determined as

$$\mathbf{dx}^{k+1} = \mathbf{x}_t(t_{k+1}) - \mathbf{x}(t_{k+1})\,.$$

The error suffices the following equation:

$$\mathbf{dx}^{k+1} = M(\mathbf{x}_t(t_k)) - M(\mathbf{x}(t_k)) - \boldsymbol{\eta}(t_k) - \mathbf{P}_{k+1}\mathbf{H}^{T}\mathbf{R}_{k+1}^{-1}(H(M(\mathbf{x}_t(t_k)) + \boldsymbol{\varepsilon}^{k+1} - H(M(\mathbf{x}(t_k) + \boldsymbol{\eta}(t_k))).$$

If to estimate $\mathbf{P}_{k+1}$ using the formula (Yaglom, 1987)

$$\mathbf{P}_{k+1} = \overline{\mathbf{dx}_n^{k+1}\left(\mathbf{dx}_n^{k+1}\right)^T} \approx \frac{1}{N-1}\sum_{n=1}^{N}\mathbf{dx}_n^{k+1}\left(\mathbf{dx}_n^{k+1}\right)^T,$$

one obtains a version of the ensemble Kalman filter.

Taking this formula for $\mathbf{P}_{k+1}$ into account one obtains a system of equations relative to $\mathbf{dx}_n^{k+1}$

Ensemble pi-algorithm

Analyses step:

$$\mathbf{X}_n^{(k+1)T} = \mathbf{F}_2^T + \mathbf{\Pi}_2^T \mathbf{D}^T, \qquad \left(\mathbf{F}_2^T\right)_n = M(\mathbf{x}_n(t_k)) + \boldsymbol{\eta}_n(t_k),$$

$$\left(\mathbf{\Pi}_2^T\right)_m^n = \frac{1}{N-1}\left(\mathbf{dx}_m^{k+1}\right)^T \mathbf{H}^T \mathbf{R}_{k+1}^{-1}(\mathbf{y}_{t_{k+1}} + \boldsymbol{\varepsilon}_n^{k+1} - H(M(\mathbf{x}_k^n) + \boldsymbol{\eta}_n(t_k))).$$

Where

$\mathbf{F}$ is a matrix with columns $\{\mathbf{f}_n^k, n = 1, \ldots, N\}$:

$$\mathbf{f}_n^k = M(\mathbf{x}_n(t_k)) + \boldsymbol{\eta}_n(t_k) - \overline{M(\mathbf{x}_n(t_k))}$$

$\tilde{\mathbf{F}}$ is a matrix with columns $\{\tilde{\mathbf{f}}_n, n = 1, \ldots, N\}$:

$$\tilde{\mathbf{f}}_n = H(M(\mathbf{x}_n(t_k)) + \boldsymbol{\eta}_n(t_k)) + \boldsymbol{\varepsilon}_n^{k+1} - \overline{H(M(\mathbf{x}_n(t_k)) + \boldsymbol{\eta}_n(t_k))}$$

$$\mathrm{D} = \begin{pmatrix} \mathrm{dx}_1^1 & \cdots & \mathrm{dx}_1^{\mathrm{N}} \\ \vdots & \ddots & \vdots \\ \mathrm{dx}_{\mathrm{K}}^1 & \cdots & \mathrm{dx}_{\mathrm{K}}^{\mathrm{N}} \end{pmatrix}$$

$$\mathbf{D}^T = (\mathbf{I} + \mathbf{\Pi}^T)^{-1}\mathbf{F}^T,$$

$$\mathbf{\Pi}^T = (\mathbf{C} + 0.25\mathbf{I})^{\frac{1}{2}} - 0.5\mathbf{I}.$$

$$\mathbf{C} = \frac{1}{N-1}\mathbf{F}^T \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{HF} + \mathbf{E}) = \mathbf{C}_1 + \mathbf{C}_2.$$

«Local» ensemble pi-algorithm

The formula $P : \dfrac{1}{N-1}\sum_{i=1}^{N} dx^i \left(dx^i\right)^T$ is approximation, so, if dimension of sample N is small covariance function property is not carried out. In a number of works it is offered for repayment "false" covariances on the big distances to use formula $\tilde{P} = P \circ \Phi(\rho)$, where $\Phi(\rho)$ - function from distance between the points, usually looking like $e^{-\alpha\rho^2}$.
It is well known, that $\tilde{P}$ is also covariance matrix.

Let’s consider the analyses step:

$$x_a^i = x_f^i + D[(HD)^T R^{-1}(y^i - Hx_f^i)]$$

Rectangular matrix $D(HD)^T$ is a covariance matrix of forecast errors in a grid cells and points of observations. We will multiply the elements of matrix by $\Phi(\rho)$.

«Local» ensemble pi-algorithm

So we have the following formula for m-th grid point:

$$\left(dx^T\right)_m = \left(f^T\right)_m - \frac{1}{N-1}[DD^T H^T \Phi(\rho) R^{-1}(HF+E)]$$

where $\left(dx^T\right)_m = \begin{pmatrix} dx_m^{(1)} \\ \vdots \\ dx_m^{(N)} \end{pmatrix}$ - the ensemble of perturbations.

And, for individual perturbation in m-th grid point

$$\left(dx^T\right)_m = \left(f^T\right)_m - \left(\Pi_m\right)^T \left(dx^T\right)_m$$

$$\left(\Pi_m\right)_i^j = \sum_{i_0} dx_{i_0}^i (f_{i_0}^i + \varepsilon_{i_0}^i) e^{-\alpha\rho(i_0,m)}$$

$$D^T \equiv [dx_1^T, \ldots, dx_K^T] = F^T - [\Pi_1^T dx_1^T, \ldots, \Pi_K^T dx_K^T],$$

$$F^T \equiv [f_1^T, \ldots, f_K^T]$$

In the case $\Pi_m \equiv \Pi$ we will have the common variant of ensemble π - algorithm.

«Local» ensemble pi-algorithm

To realise the algorithm with localisation following variants can be considered:

- ***the errors of observations do not correlate;***
- ***observations can be separated with mutually not correlating errors;***
- ***observations represent the data file which errors correlate among themselves (satellite data).***

Data assimilation in environment modelling

1. ***The modelling of distribution in atmosphere:***
   - ***Passive gases ("greenhouse") (CO_2, CH_4).***
   - ***Reactive gases.***
   - ***Aerozols.***
2. ***Observations.***
3. ***The restoration of the time-space distribution of gases on the base of the data assimilation.***
4. ***The estimation of regions influencing on the distribution of pollution in the given territory.***
5. ***The estimation of gas fluxes using the results of data assimilation.***

Data assimilation in environment modelling

<u>The main problems:</u>

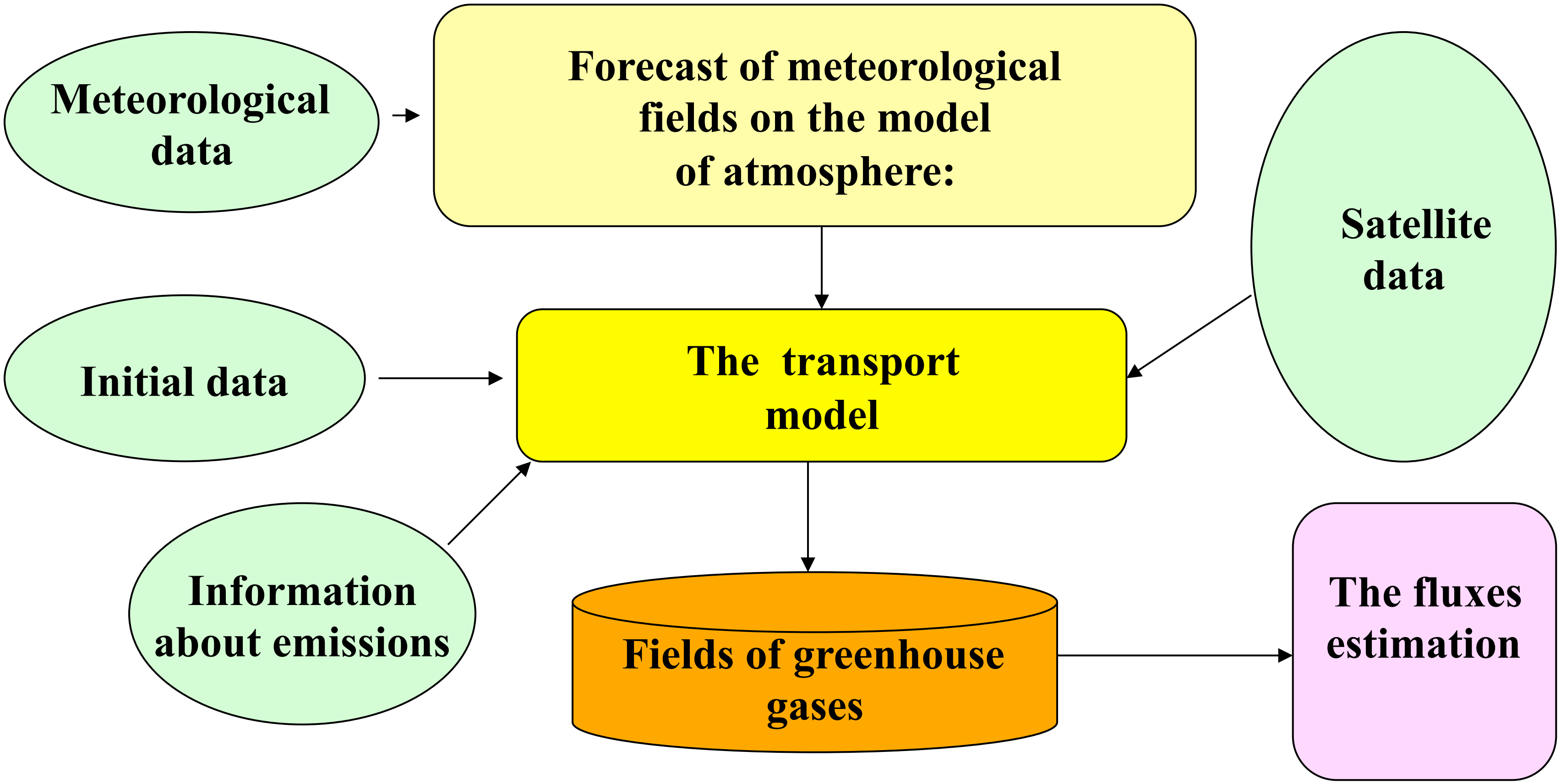
- ***There is no enough regular observations of high quality similar to meteorological observations;***

- ***Chemical reactions are described by nonlinear operators.***

- ***Data assimilation algorithms should be adapted for features of the observations.***

- ***It is necessary to distinguish problems of long-term environment monitoring and a problem of short-term monitoring.***

Data assimilation of passive gases :



Practical realisation of data assimilation algorithm: satellite data

- ***Assumptions:***

  a. The observational errors are «unbiased».

  b. The random fields of observational errors are normally distributed.
- ***«Retrievals» or «radiation».***
- ***The problem of the quality control.***
- ***The definition of covariance matrix of observational errors.***
- ***The huge quantity of observations not always yields the best result.***

Practical realisation of data assimilation algorithm: satellite data

- ***A huge array of data which are close located to each other: "superobservation";***
- ***AIRS "total column value": average kernel;***
- ***observations errors correlate;***
- ***problem of "bias" estimation.***

Estimation of parameters in data assimilation algorithm, based on the Ensemble Kalman Filter

---

The forecast step:

$$\mathbf{x}_f^{n+1} = A(\mathbf{x}_f^n) + \mathbf{b}_f^n$$

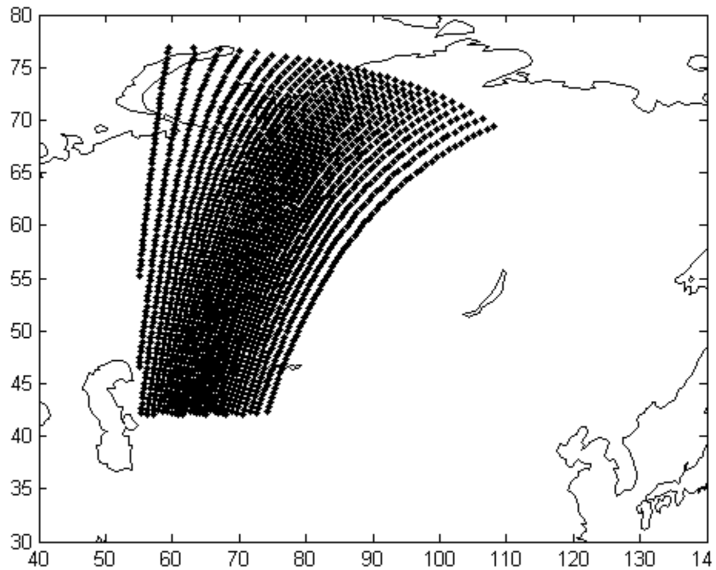
$$\mathbf{b}_f^{n+1} = B(\mathbf{b}_f^n)$$

The analyses step:

$$\mathbf{x}_a^n = \mathbf{x}_f^n + \mathbf{P}_{xx}(\mathbf{H}\mathbf{P}_{xx}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}_0 - \mathbf{H}\mathbf{x}_f^n)$$

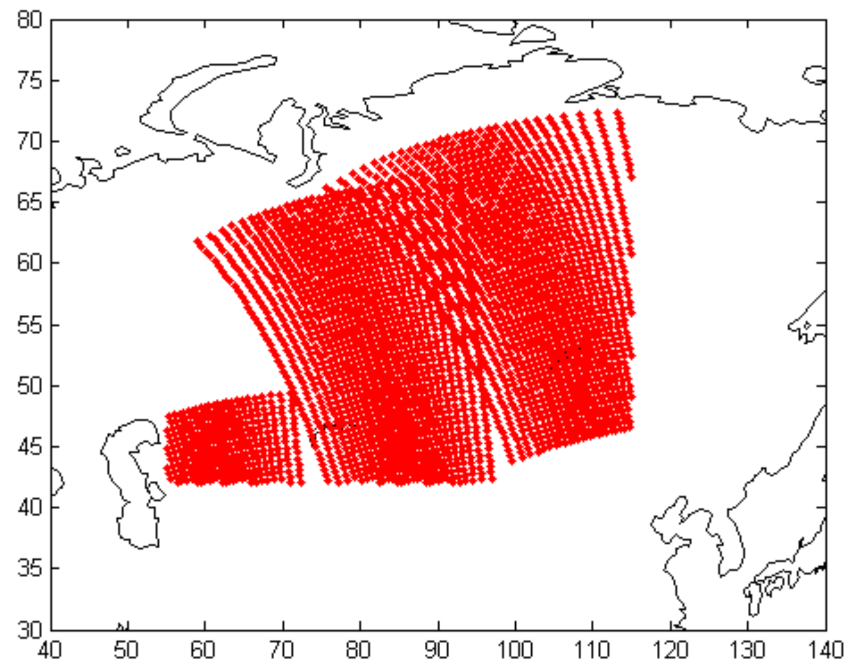
$$\mathbf{b}_a^n = \mathbf{b}_f^n + \mathbf{P}_{xb}(\mathbf{H}\mathbf{P}_{xx}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y}_0 - \mathbf{H}\mathbf{x}_f^n)$$

Practical realisation of data assimilation algorithm: satellite data



Space distribution of AIRS data 2011.06.07 (06 hours)

Space distribution of AIRS data 2011.06.07 (12 hours)



Variational approach: (2DVAR /3DVAR)

$$2J(x) = (x - x_b)^T P_b^{-1}(x - x_b) + (y_0 - H(x))^T R^{-1}(y_0 - H(x))$$

$$\downarrow$$

$$\nabla_x J(x_a) = 0$$

$$\downarrow$$

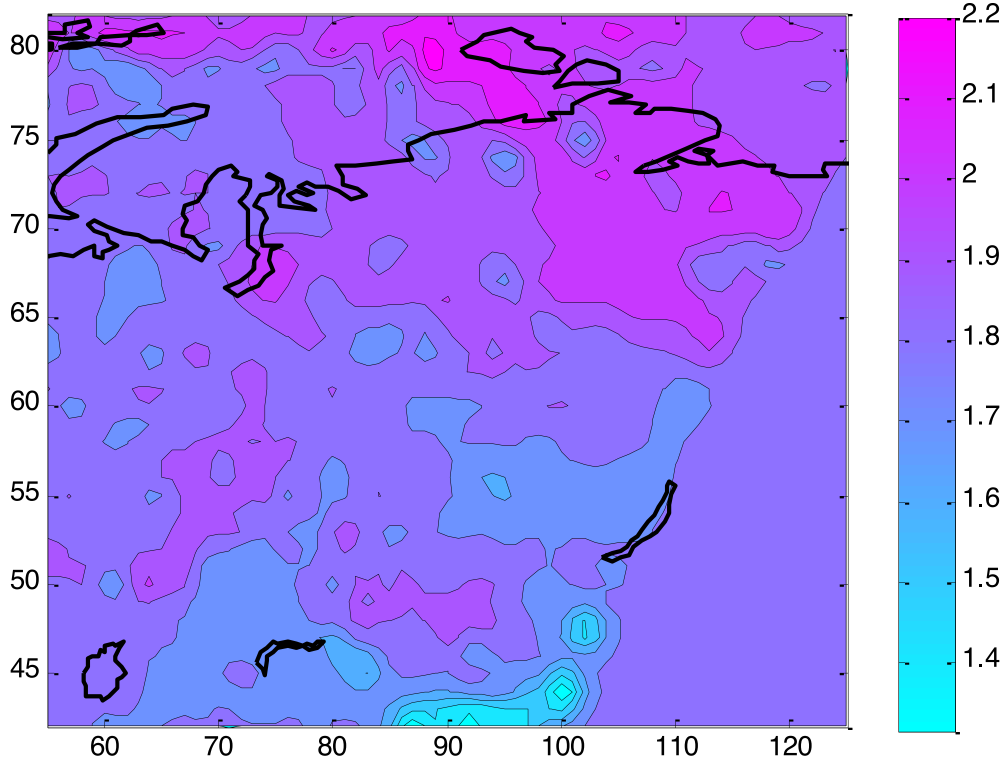
$$(P_b^{-1} + H^T R^{-1} H)(x_a - x_b) = H^T R^{-1}(y_0 - H(x_b))$$

$$\updownarrow$$

$$x_a = x_b + (P_b^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1}(y_0 - H(x_b))$$

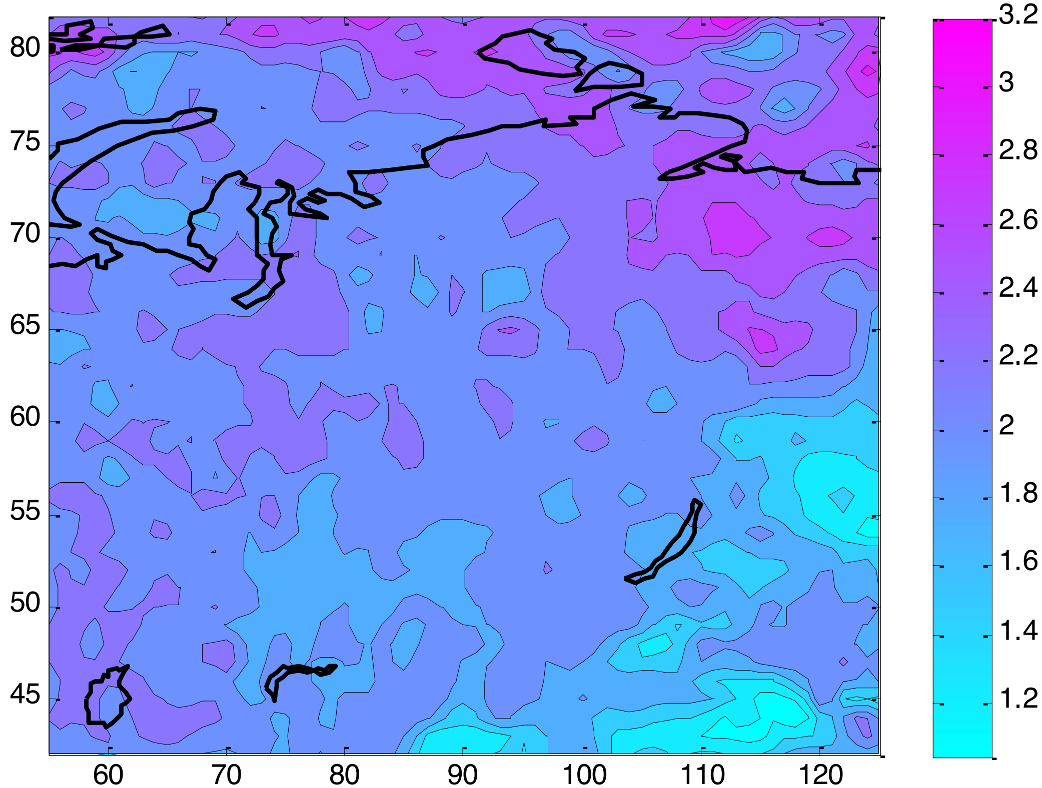
Estimation of CO concentration (AIRS)

Concentration of CO (mol/cm^2 10^{18} , total column volume, t=00 h), (55 º east long., 125 º east.long.), (42 º north lat., 82 º north lat.)



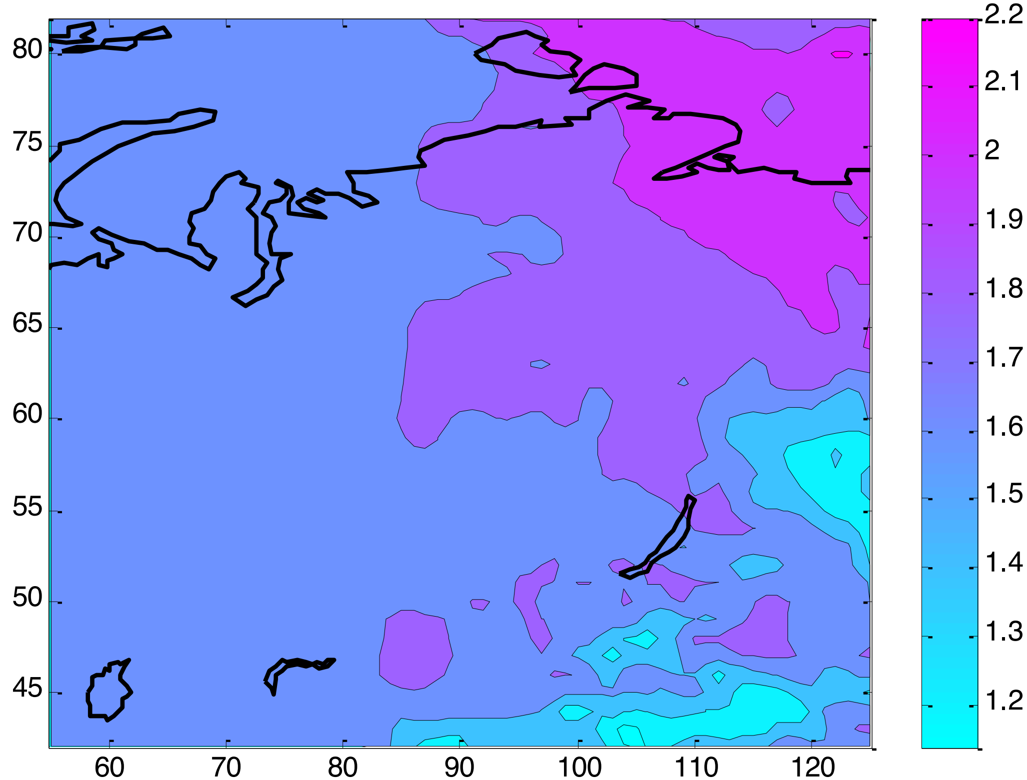
Estimation of CO concentration (AIRS)

Concentration of CO (mol/cm^2 10^{18} , total column volume, t=06 h),
(55 º east long., 125 º east.long.), (42 º north lat., 82 º north lat.)



Estimation of CO concentration (AIRS)

Concentration of CO (mol/cm^2 10^{18} , total column volume, t=18 h),
(55 º east long., 125 º east.long.), (42 º north lat., 82 º north lat.)



Passive pollution transport and diffusion model

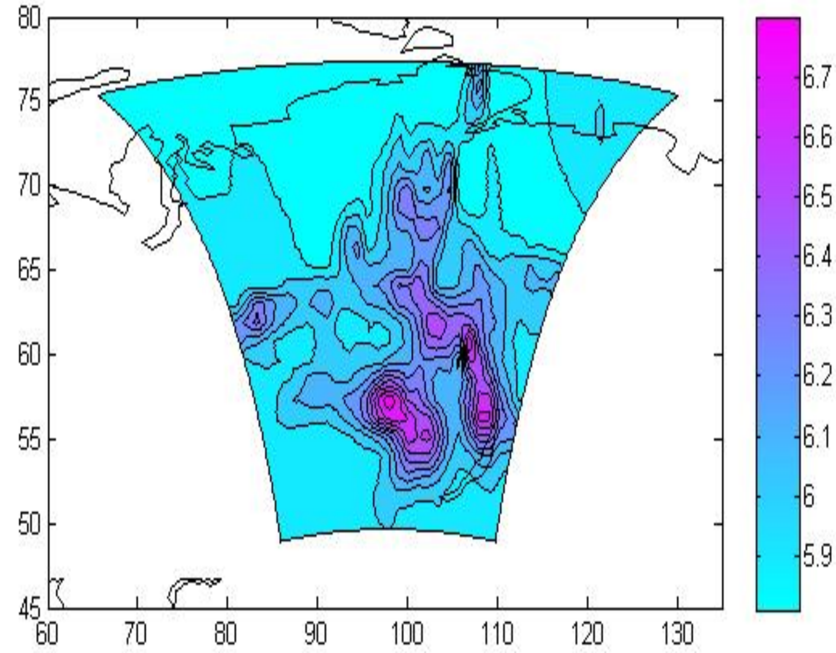
$$\frac{\partial s}{\partial t}+u\frac{\partial s}{\partial x}+v\frac{\partial s}{\partial y}+(w-wg)\frac{\partial s}{\partial z}=\frac{\partial}{\partial x}\left(k_1\frac{\partial s}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_2\frac{\partial s}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_3\frac{\partial s}{\partial z}\right)+\eta$$

The problem is considered in $Dt=G\times[0,T]$
$G=S\times[h,H]$; $S=\{0\le x\le X;0\le y\le Y\}$.

Boundary conditions: $s=Sg$

$$v\frac{\partial s}{\partial z}=\alpha(s-s_0)$$

$$v\frac{\partial s}{\partial z}=0$$



The suboptimal Kalman filter algorithm

Passive pollution transport and diffusion model *:*

(s – passive pollution concentration, η- emission)

$$s_{k+1} = As_k + \eta_k$$

$$\eta_0 = \tilde{\eta} \qquad \eta_{k+1} = \eta_k$$

Forecast errors covariance matrix:

$$P_k^f = \overline{\Delta s_k (\Delta s_k)^T} \cong \frac{1}{N-1} \sum_{i=1}^{N} \Delta s_i (\Delta s_i)^T$$

$$\Delta s_k = s_k^f - s_k^t$$

Emission evaluation:

$$\eta_k = \tilde{\eta}(1 + \delta\eta_k)$$

$$\delta\eta_{k+1} = \alpha\delta\eta_k + (\sqrt{1-\alpha^2})\chi_k$$

A.W.Heemink, A.J.Segers "Modeling and prediction of environmental data in space and time using Kalman filtering", Stochastic Environmental Research and Risk Assessment 16 (2002), 225-240 ,Springer-Verlag 2002.

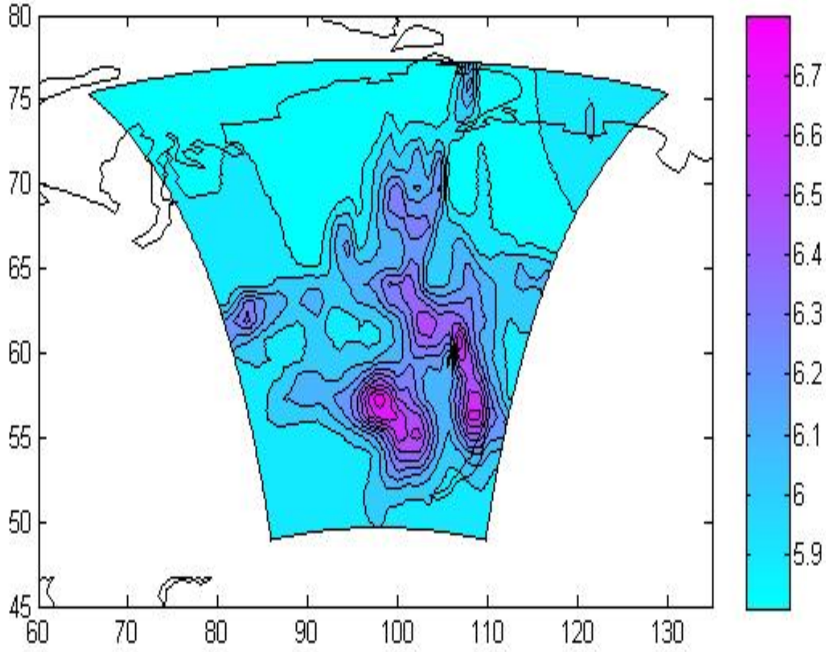
Numerical experiments with the modeled data

«True value»: $x_0^t = \widetilde{x}, \qquad x_k^t = A x_{k-1}^t + \eta_{k-1}^t$

Forecast: $x_0^f = \widetilde{x} + \varepsilon_0, \quad x_k^f = A x_{k-1}^f + \eta_{k-1}^f, \quad \varepsilon_0 = N(0, \sigma_0^2)$

Observations: $y_k^o = M_k x_k^t + \varepsilon_1 \qquad \varepsilon_1 = N(0, \sigma_0^2)$

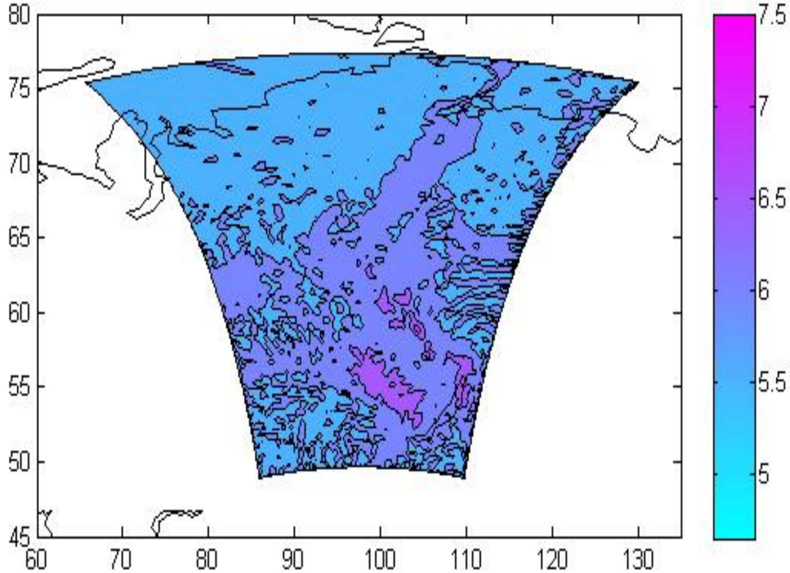
The initial CO2 field



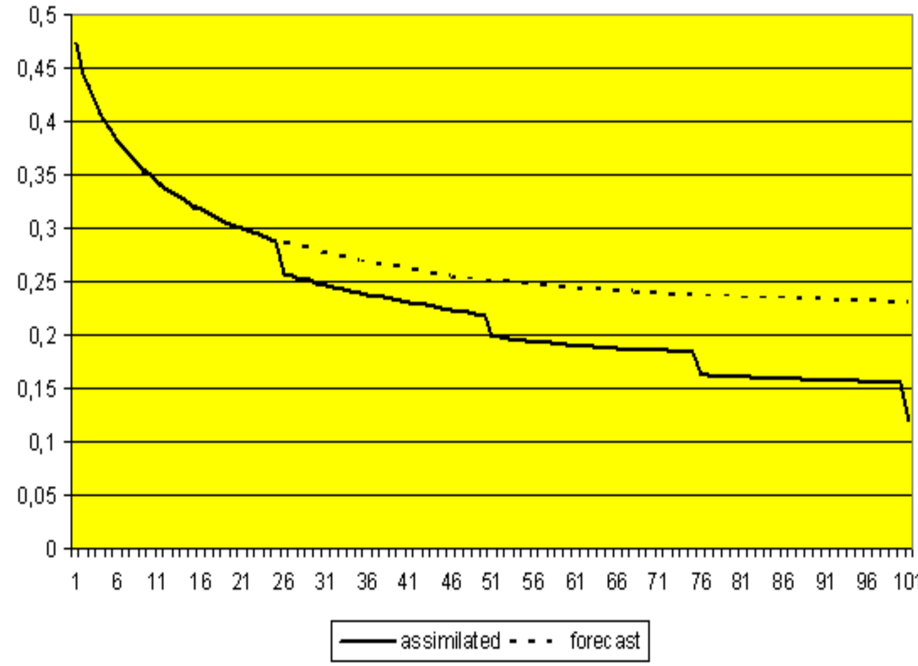
The region (48.9º N, 85.9º E; 76.9º N, 109º E)

[***http://www.gmes-atmosphere.eu/data***](http://www.gmes-atmosphere.eu/data)

Data assimilation for CO2 concentration (24 hours period)



Concentration field, ((kg / kg)*10^{-4}, height of 1200 m)



Root mean square error for concentration, ((kg / kg)*10^{-4})

Summary

- ***Observations play the major role at the estimation of an ecological situation in the given region.***

- ***Data assimilation algorithms should be adapted for features of the observations.***

- ***It is necessary to distinguish problems of long-term environment monitoring and a problem of short-term monitoring.***

Thank you for attention !