Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Data assimilation

errors and sensitivity

“*All forecasts fail*” —— Prof. Dr. Nedjeljka Žagar

Presenter: Dong Jian

Advisors: Prof. Dr. Detlef Stammer Dr. Armin Köhl; Dr. Nuno Serra

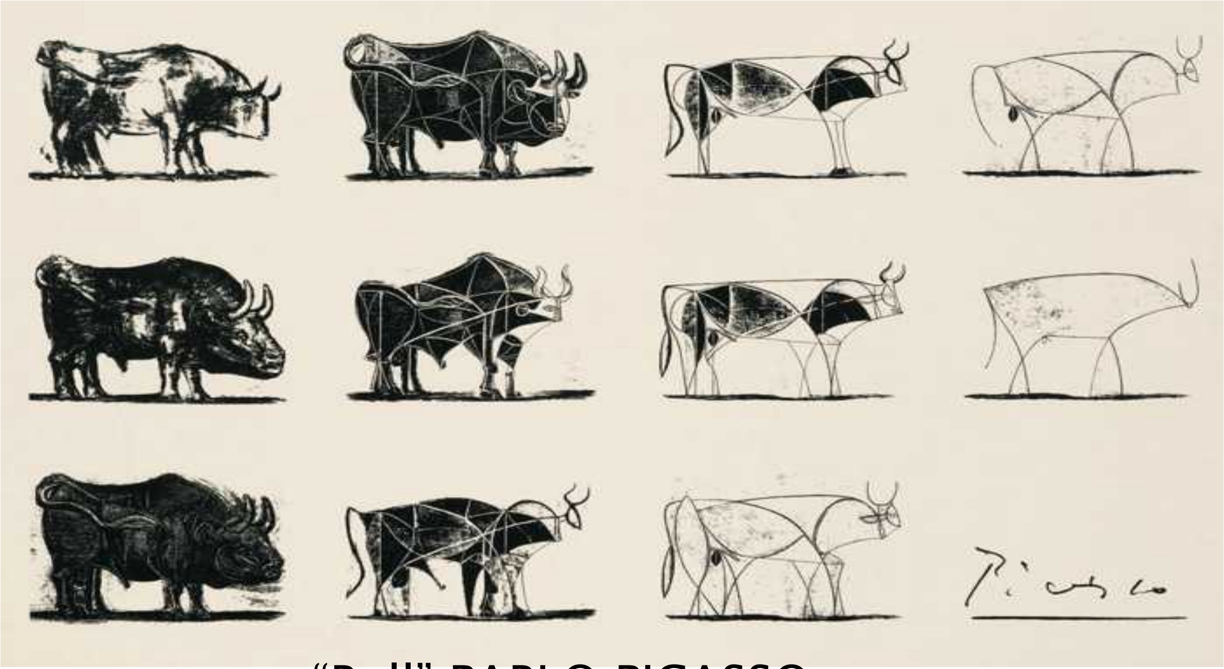
Institut für Meereskunde (IfM)
Centrum für Erdsystemforschung und Nachhaltigkeit
Universität Hamburg (UHH)

05,July,2021

Why all forecasts fail?

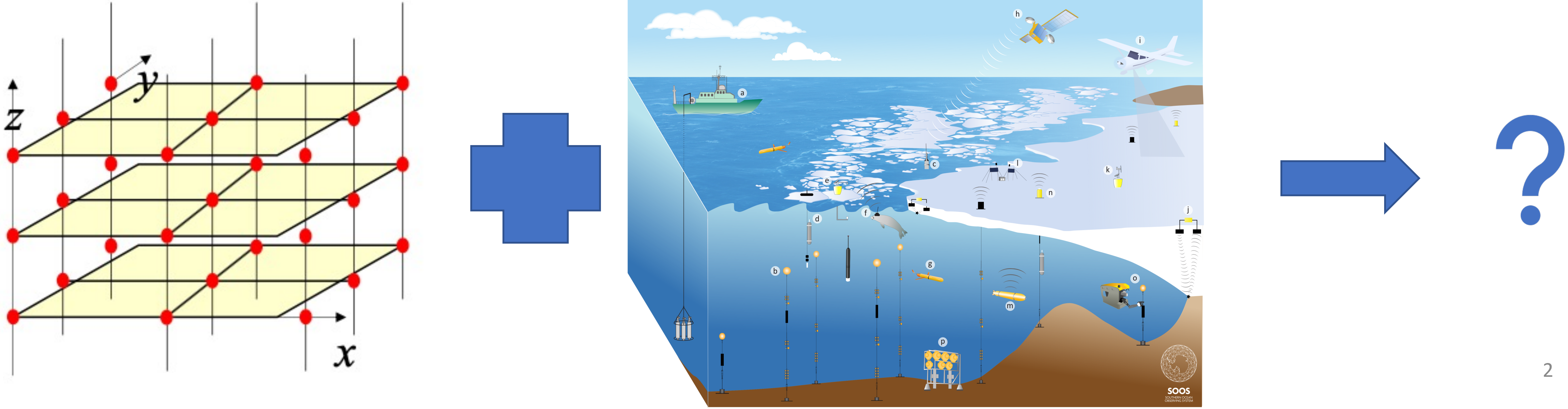
- **Observation**
- **First guess/ background**
- **Model**

combine model with observations make a better forecast?



"Bull" PABLO PICASSO

Model grid



least square estimation ⟷ Bayes theorem

Two independent observations with unknown errors:

$$\left.\begin{aligned} T_1 &= T_t + \varepsilon_1 \\ T_2 &= T_t + \varepsilon_2 \end{aligned}\right\}$$

Assume T_1 and T_2 are unbiased: $E\,(T_1 - T_t) = E\,(T_2 - T_t) = 0$

Assume we know the variances of the observational errors:

$$E(\varepsilon_1^2) = \sigma_1^2 \quad E(\varepsilon_2^2) = \sigma_2^2$$

Assume errors of two measurements uncorrelated: $E(\varepsilon_1\varepsilon_2) = 0$

Estimate T_t from a linear combination: $T_a = a_1 T_1 + a_2 T_2$

"Analysis" T_t should be unbiased: $E(T_a) = E(T_t)$ which implies

$$a_1 + a_2 = 1$$

Minimize the mean squared error of T_a:

$$\sigma_a^2 = E\left[(T_a - T_t)^2\right] = E\left[(a_1(T_1 - T_t) + a_2(T_2 - T_t))^2\right]$$

Optimal weight: $a_1 = \dfrac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad a_2 = \dfrac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$

Analysis formula: $\mathrm{T_a} = \mathrm{T_1} + \dfrac{{\sigma_1}^2}{\sigma_1^2 + \sigma_2^2}(\mathrm{T_2} - \mathrm{T_1})$

Errors formula: $\dfrac{1}{\sigma_a^2} = \dfrac{1}{\sigma_1^2} + \dfrac{1}{\sigma_2^2}$

$$\mathrm{p}(\mathbf{x} \mid \mathbf{y}) = \frac{\mathrm{p}(\mathbf{x}) \times \mathrm{p}(\mathbf{y} \mid \mathbf{x})}{\mathrm{p}(\mathbf{y})}$$

$$\text{posterior probability} = \frac{\text{prior probability} \times \text{likelihood}}{\text{normalizing constant}}$$

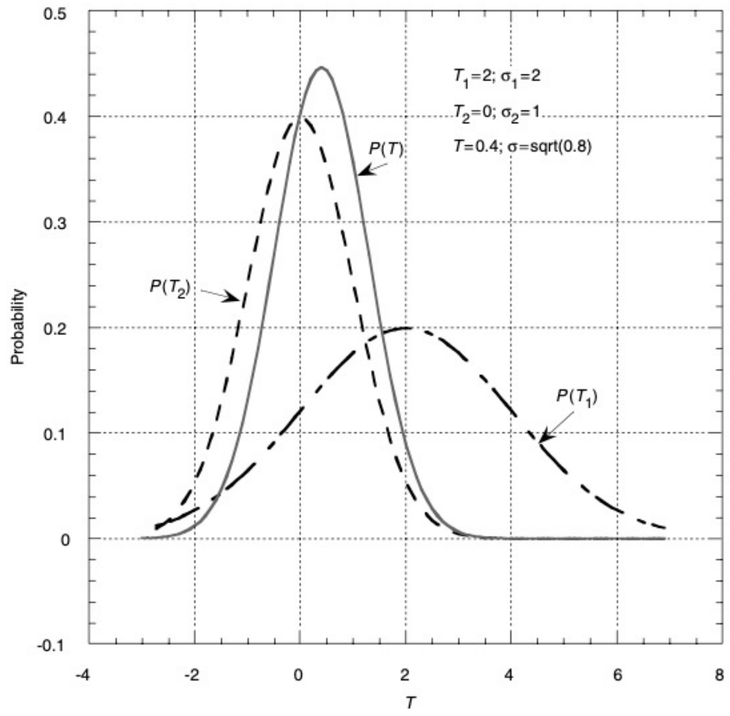
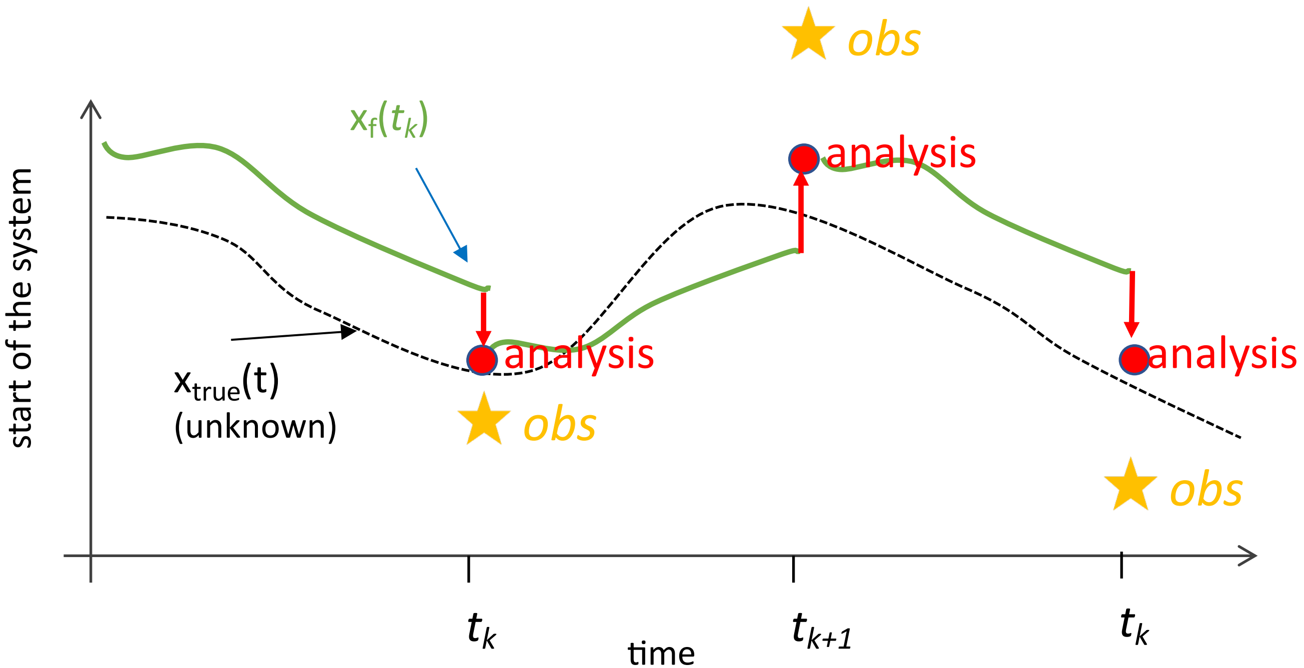


Figure 5.3.1: Illustration of the properties of the probability distribution of the analysis T, given observations T_1 and T_2, using either the least squares approach or the Bayesian approach (after Purser, 1984).

OI $\mathbf{x^a = x^b + BH^T (HBH^T + R)^{-1} (y - Hx^b)}$

Precision $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$

1) Forecast – Observations (radiosonde)



Filter: Assimilate every time observations are available

$$\mathbf{B} \cong \frac{\langle (B_k - T_k)(B_l - T_l) \rangle}{[\langle (B_k - T_k)^2 \rangle \langle (B_l - T_l)^2 \rangle]^{1/2}}$$

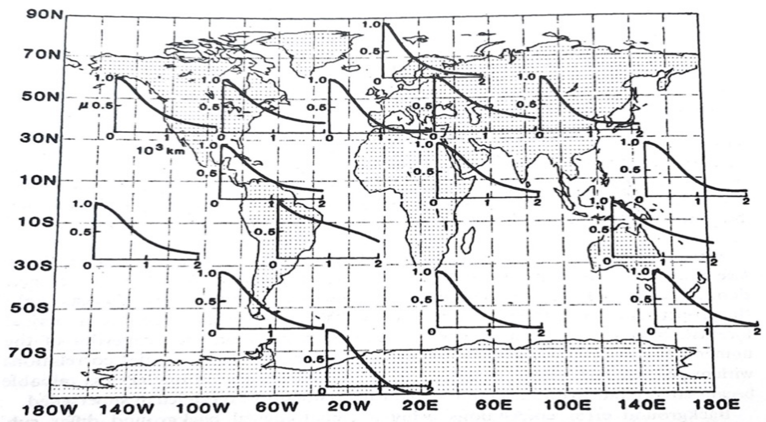


Figure 4.5 Isotropic component of 500 mb geopotential background (forecast) error correlation in different parts of the globe. (After Baker et al. *Mon. Wea. Rev.* **115**: 272, 1987. The American Meteorological Society.)

$$\mathbf{B} = \sigma_b^2 \begin{pmatrix} 1 & \gamma_{12} & \cdots & \gamma_{1N} \\ \gamma_{21} & 1 & \cdots & \gamma_{2N} \\ \vdots & & \ddots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \cdots & 1 \end{pmatrix}$$

$$\gamma_{ij} = \exp\left(-\left(r_{ij}/L\right)^2\right)$$

B assumed homogeneous and isotropic

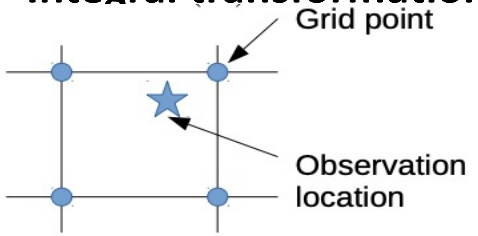
$$\mathbf{R} = \sigma_r^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

observation errors are not correlated

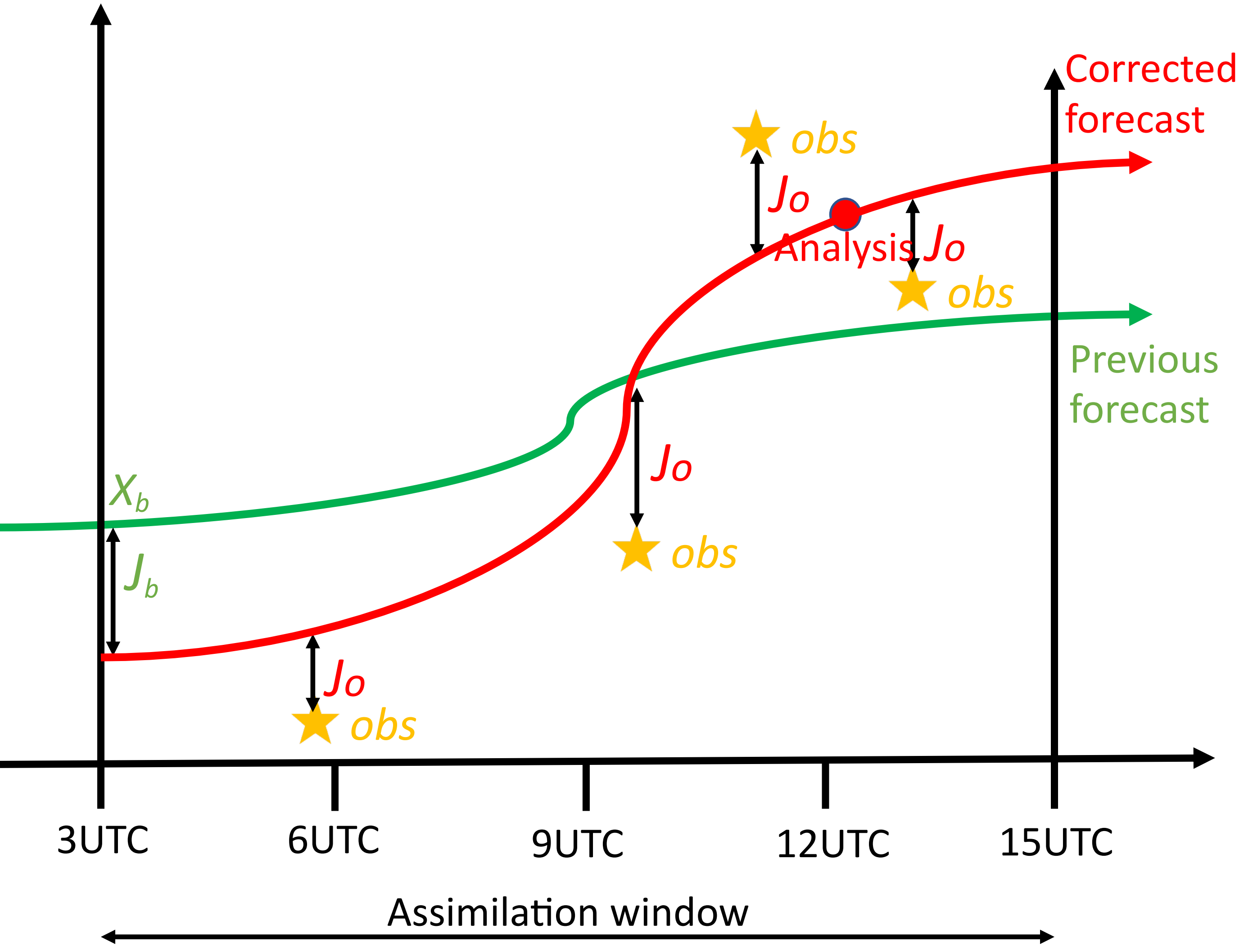
error variances are same for same observation type

H

- **Identity**
- **Interpolator**
- **Integral transformation**



4D-Var



Smoother: assimilate observations over a time window

minimize cost function $J(\mathbf{x})$

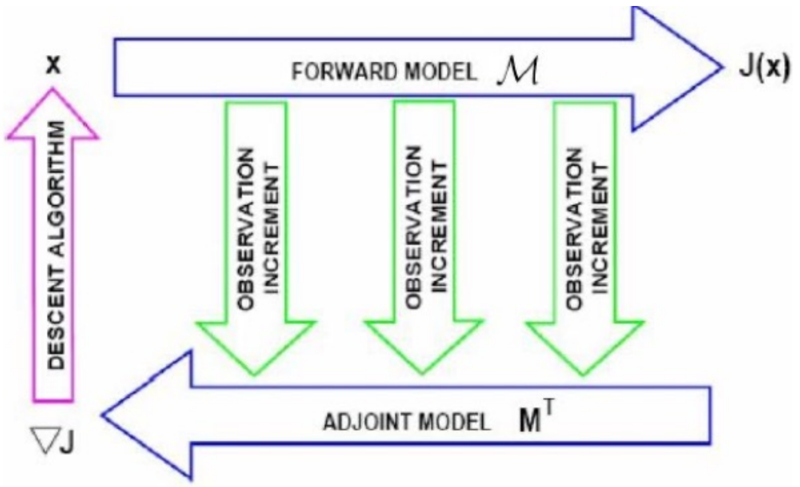
$$J(\mathbf{x}) = J_b(\mathbf{x}) + J_o(\mathbf{x})$$

$$= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_{0b})^T \mathbf{B}_0{}^{-1}(\mathbf{x}_0 - \mathbf{x}_{0b})$$

$$+ \frac{1}{2}\sum_{i=0}^{N}\left(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)\right)^T \mathbf{R}_i^{-1}\left(\mathbf{y}_i - \mathcal{H}_i(\mathbf{x}_i)\right)$$

$x_{i+1} = \mathcal{M}_i(x_i)$ assume model is perfect

$\mathbf{x}_0$ ------ analysis at t_0 to be solved
$\mathbf{x}_{ob}$ ----- background field
$\mathbf{x}_i$ ----- nonlinear model forecast at t_i or $\mathcal{M}_i(x_0)$
$\boldsymbol{y}_i$ ----- observation at t_i
$\mathcal{H}_i$ ----- *observation operator*



Model 1.5-layer reduced gravity shallow water model

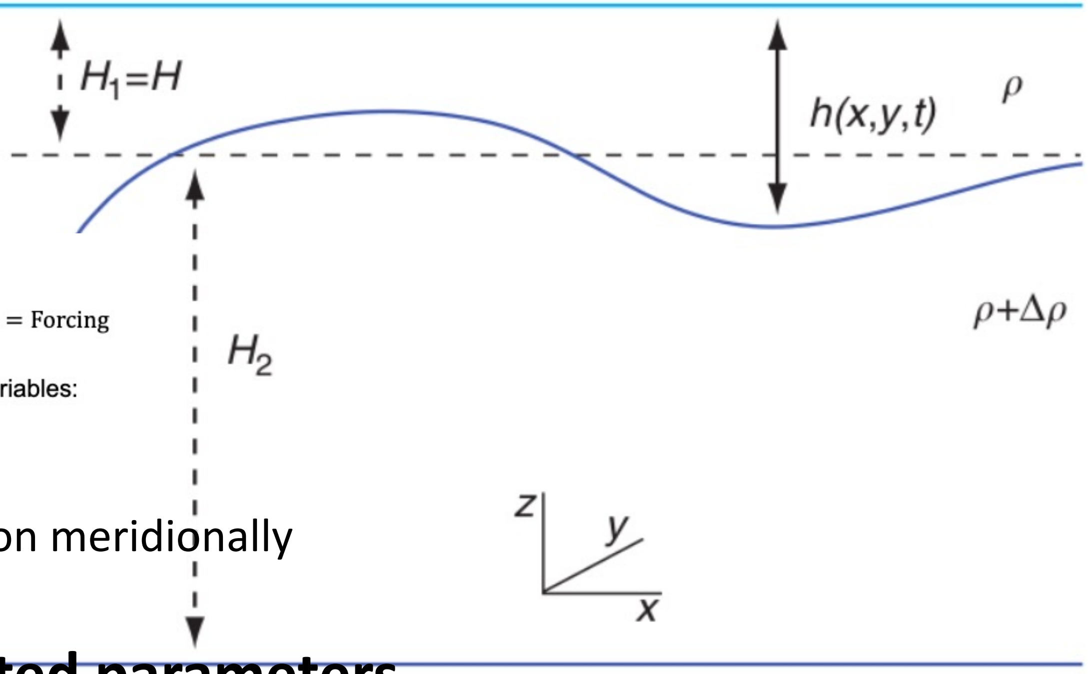
- **aim: simulate Kelvin/Rossby wave propagation forced by wind burst**
- Dynamic equation in matrix form
- Discretization: CTCS scheme
- Equatorial Pacific idealized configuration

- The equations are first cast in matrix form:

$$\frac{\partial W}{\partial t} + A\frac{\partial W}{\partial x} + B\frac{\partial W}{\partial y} + C\,y\,W + D\,\nabla^2 W = \text{Forcing}$$

where W is a 3x1 vector with the prognostic variables:

$$W = [\,h\;\; u\;\; v\,]'$$



Reflective boundary condition zonally, absorbing boundary condition meridionally

Model Error → get projected on the estimated parameters

$$\mathbf{x}^t = m^{t-1\rightarrow t}\left(\mathbf{x}^{t-1}\right) + \text{error}$$

$\mathbf{x}^t \in \mathcal{R}^{N_x}$

Evolution operator

Previous value of the variable.

The models are not perfect.
Errors come from:
a. Unknown physics
b. Numerical error in the time/space discretisation of continuous equations.
c. Subgrid processes that need to be parameterised.
...

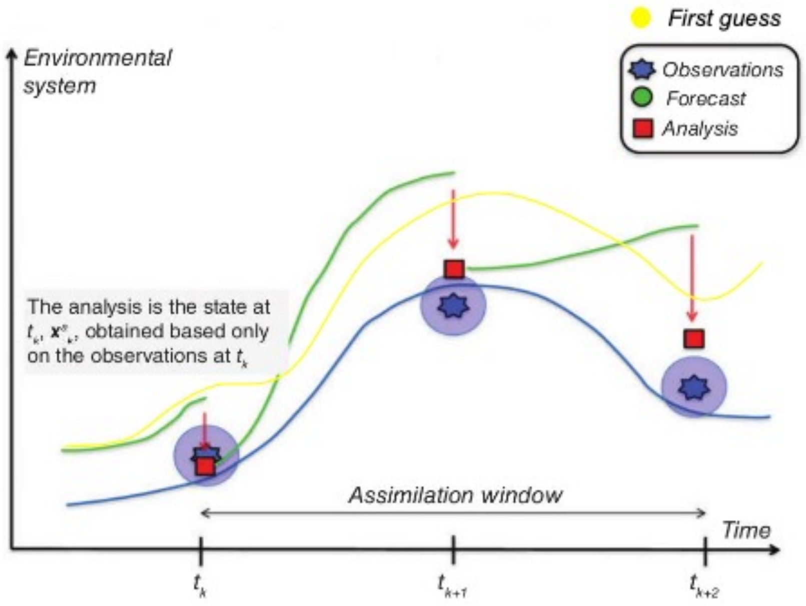
- parameters, eg. forcing unknown
- Resolution: unresolved processes
- imperfect model discretization

Lab1 OI configuration:

The Matlab main program is:

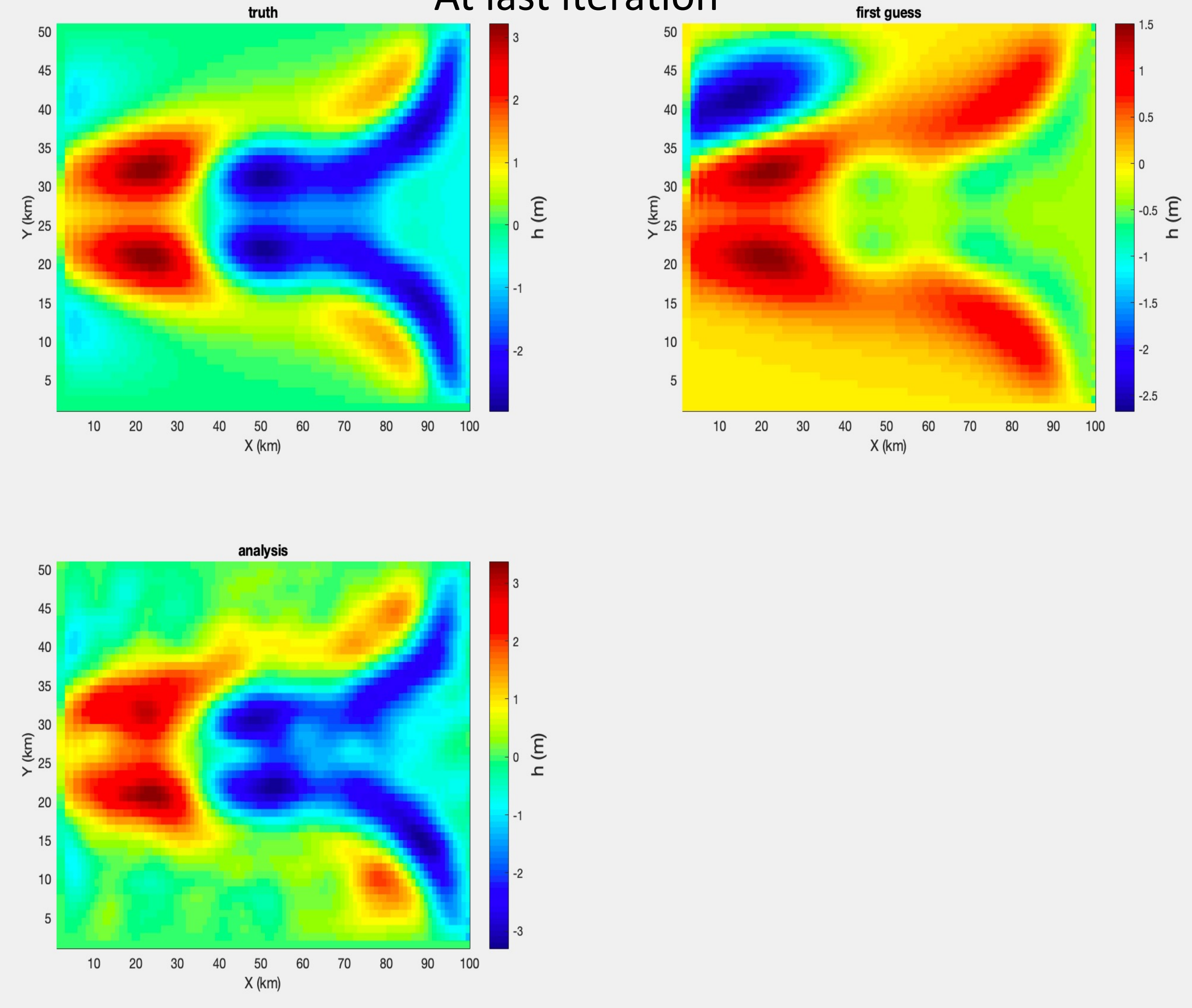
```
function OI

% set model and DA parameters
    model_params
    da_params
% initialize and run forward to get "truth" and synthetic "observations"
    initial('observ')
    forward('observ',1)
% do cycles of forecast + analysis
    initial('forec',1)
    forward('first_guess',1)
    optim_interp
% merge all forecasts into a "reanalysis"
    reanalysis
% compute and plot metric: root mean square difference between "reanalysis" and "truth"
    metric
```

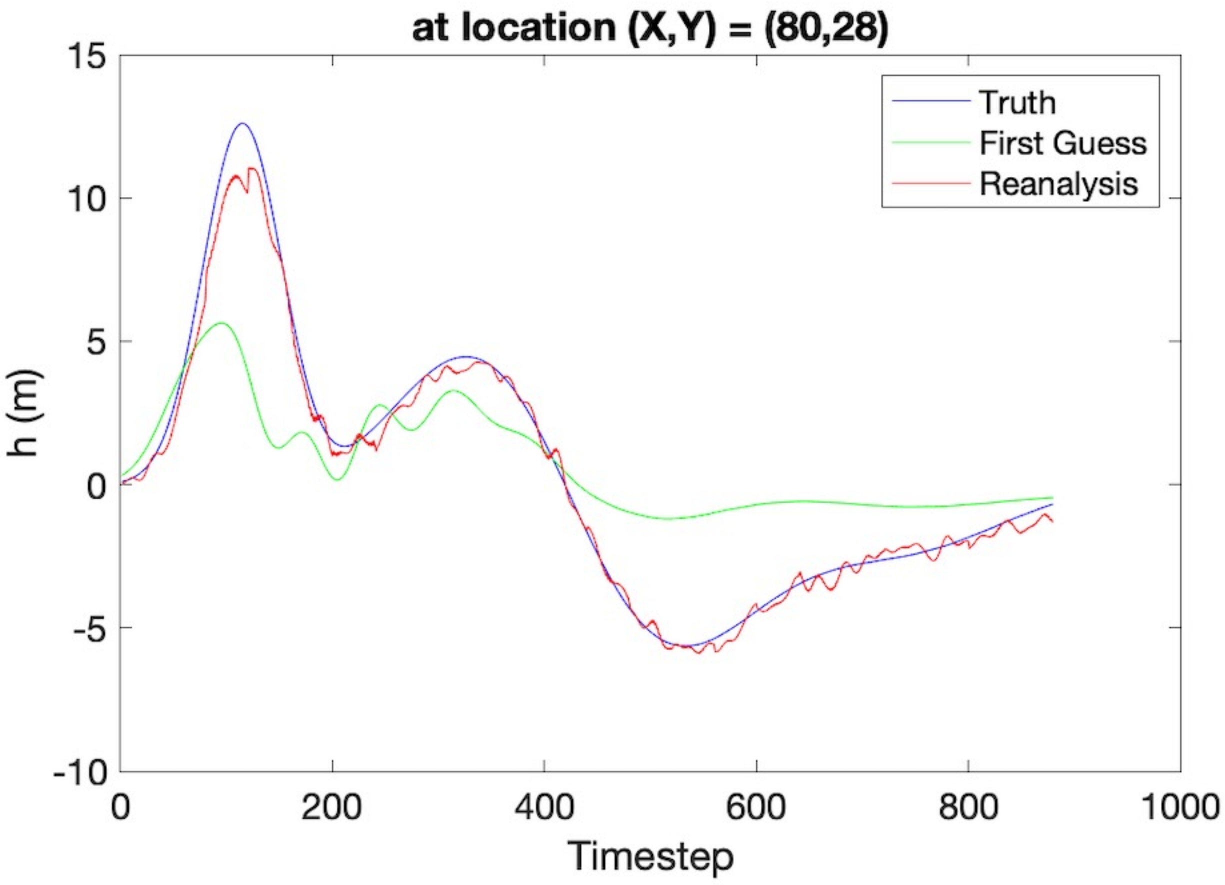


default experiment: comparation of analysis, truth and first guess

At last iteration



At random location



Lab1 OI task 3 sensitivity to decorrelation scale L

Qeustion1: which decorrelation scale L is reasonable to assume?(exp1,2,3)

experiment	decorrelation L	observation
1 (default)	250 km	5 %
2 (half L, same obs)	125 km	5%
3 (double L, same obs)	500 km	5%

Qeustion2: If only few observations exit, will adjust L make forecast better ? (4,5,6)

experiment	decorrelation L	observation
4 (double L, half obs)	500 km	2.5%
5(half L, half obs)	125 km	2.5%
6(default L, half obs)	250 km	2.5%
7 (increase L properly)	300 km	2.5 %

RMSD between analysis and " truth" normalized by the STD of the "truth"

rms

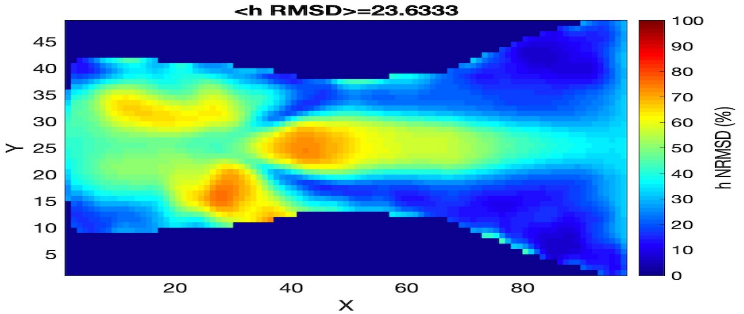
$$\frac{\sqrt{\sum_{i=1}^{nt}\frac{(rean_i - truth_i)^2}{nt}}}{nt^{-1}\cdot\sum_{i=1}^{nt}(truth_i - mean)^2}\cdot 100 \quad \text{(in percent)}$$

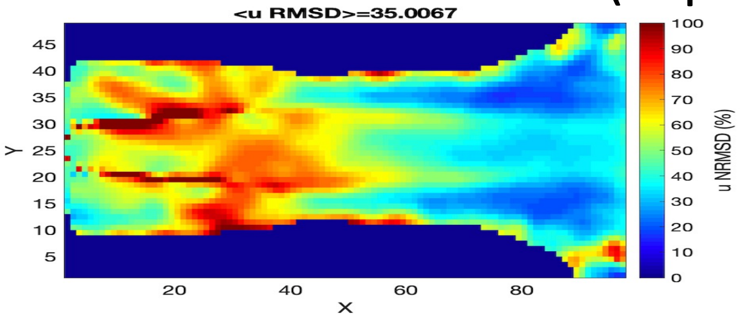
variance

note: due to random generator set rng('default')% to keep observation same for different L

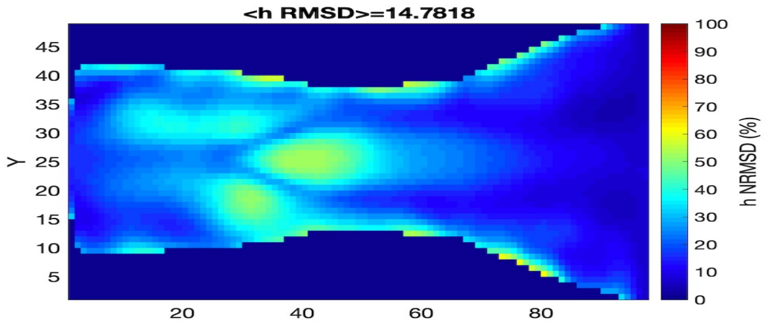
Qeustion1: which decorrelation scale L is reasonable to assume?(exp1,2,3)

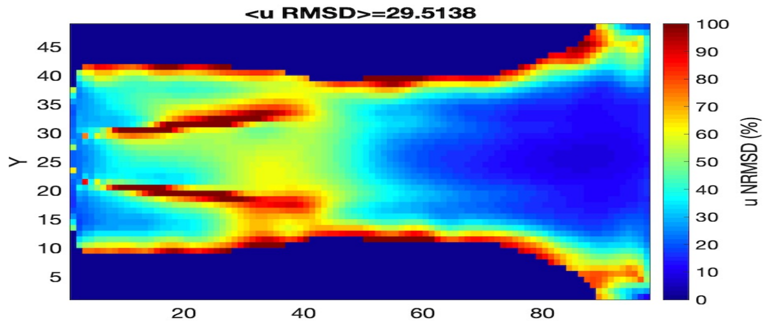
L=125 km





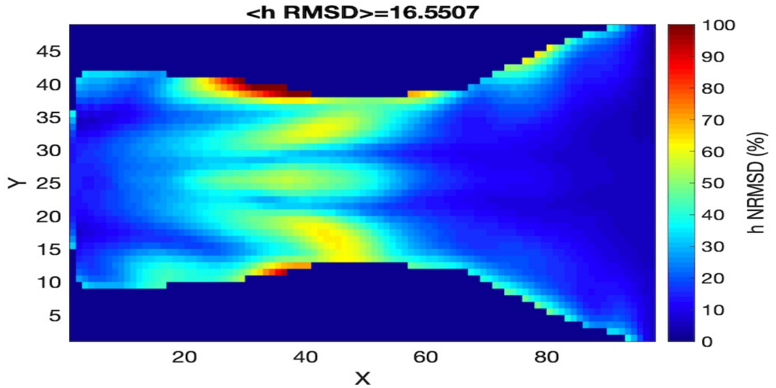
L=250 km

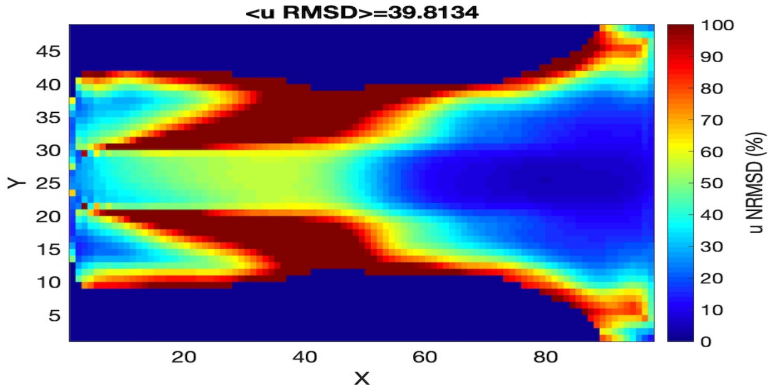




Default setting:smallest RMSD (reanalysis-"truth") normalized by STD of "truth"

L=500 km

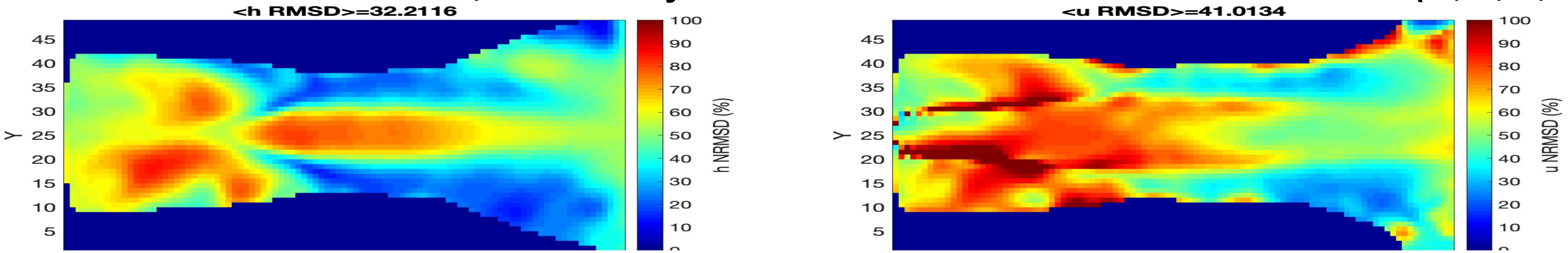




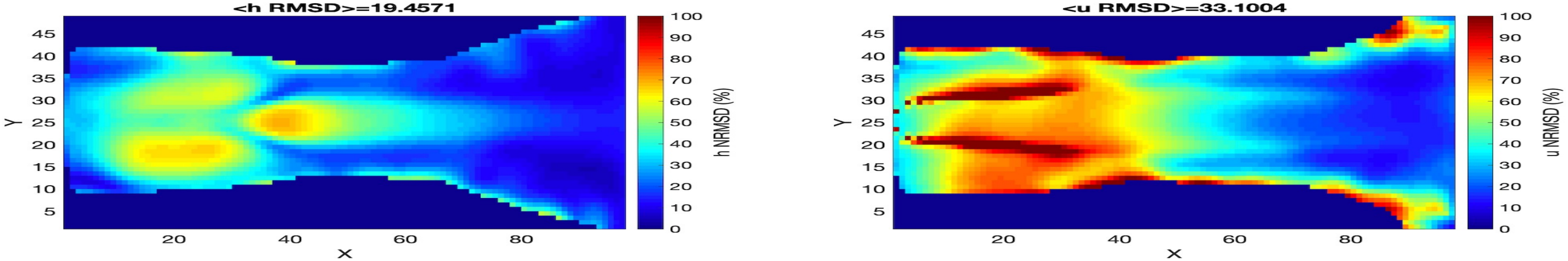
Answer 1: L should not be too small nor too large. Larger L, Larger influence range
one can consider the L which produces smallest RMSD is the most reasonable setting

Qeustion2: If only few observations exit, how adjust L will make forecast better ? (4,5,6,7)

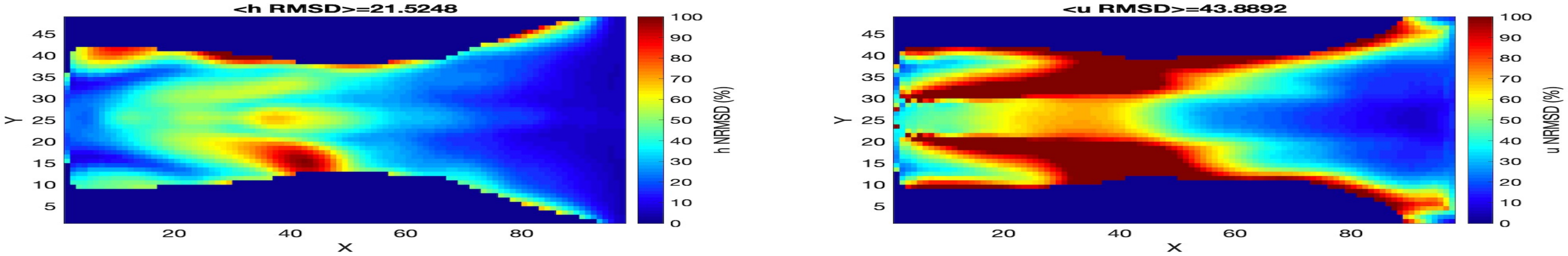
L = 125, observation 2.5%



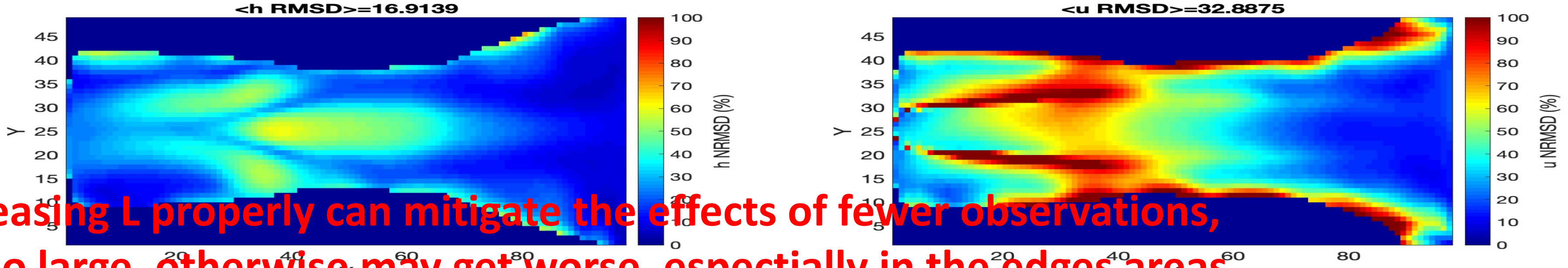
L=250, observation 2.5%



L=500, observation 2.5%



L=300, observation 2.5%



Answer 2: increasing L properly can mitigate the effects of fewer observations, but L cannot too large, otherwise may get worse, espectially in the edges areas

Lab1 OI task 4 sensitivity to sparseness and noise

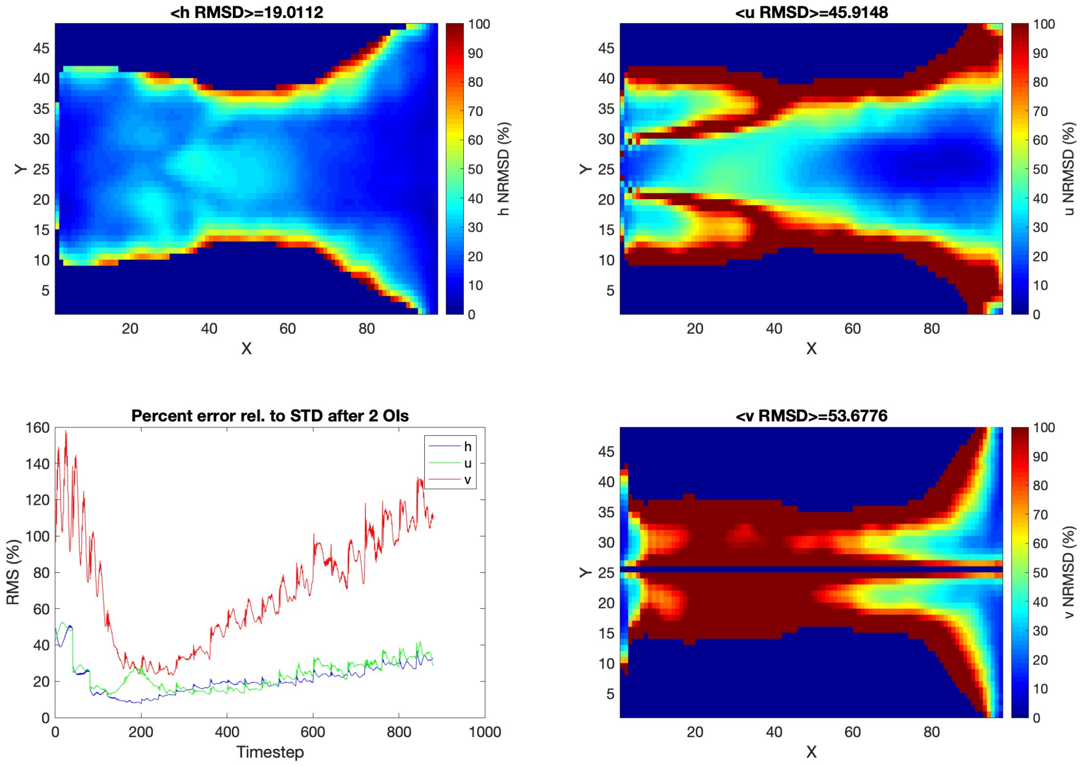
Question3: more observations but less accurate, less observation but more accurate, which one is better?

experiment	noise	& observation
1 default	[1.0 0.005 0.005].*1	5%
6 (less precise, more obs)	[1.0 0.005 0.005].*2	10%
7 (more precise, less obs)	[1.0 0.005 0.005].*0.5	2.5%

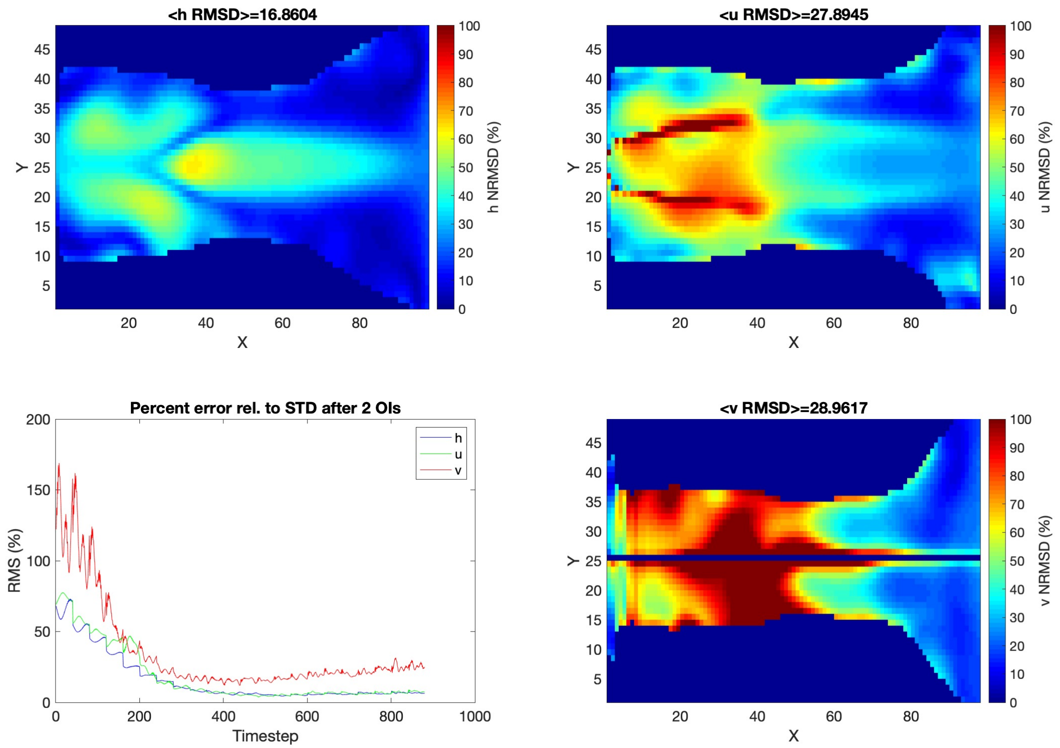
Question4: what if observations are restricted in equatorial regions?

experiment	noise	& observation
8 (restrict obs in equatorial)	[1.0 0.005 0.005].*1	reduced

Question3: more observations but less accurate, less observation but more accurate, which one is better?



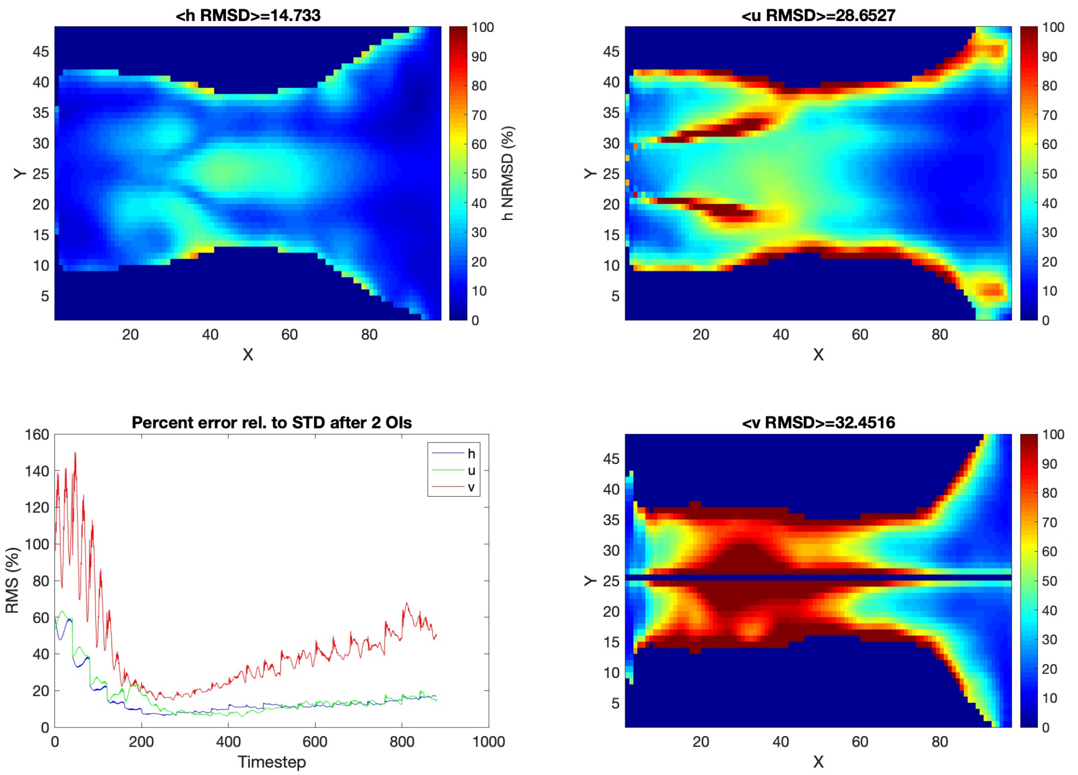
double observations double noise



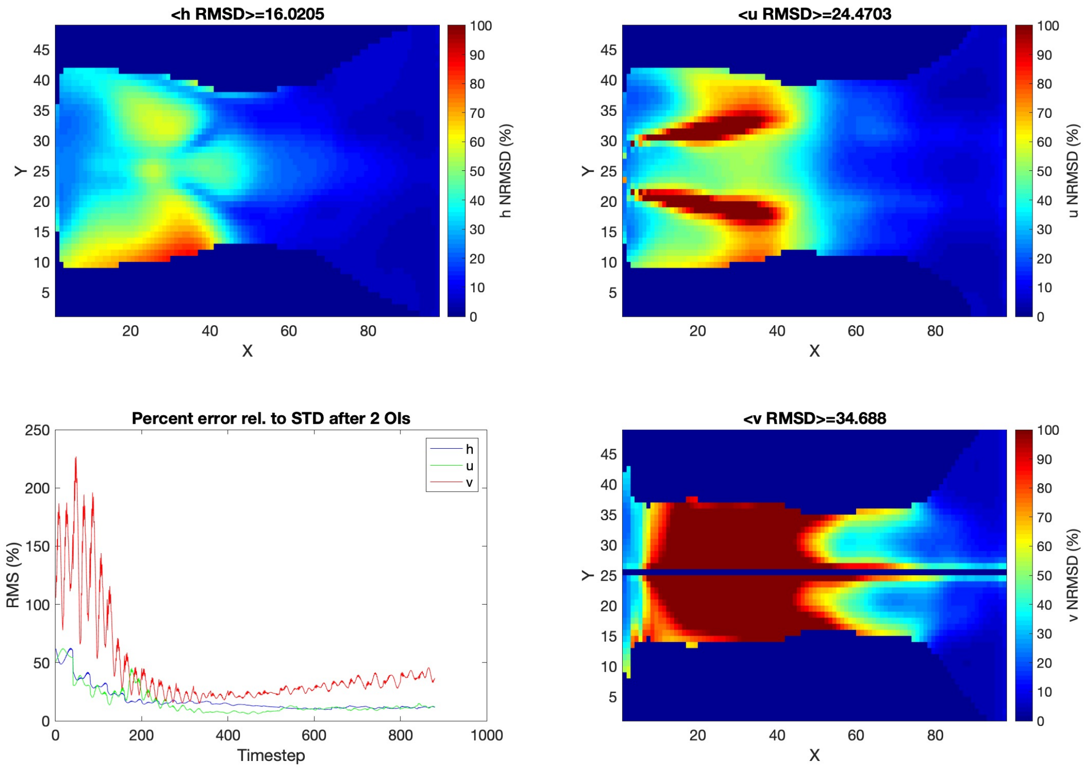
half observations half noise

Anwser3: less observation but more accurate is better, quality is more important than quantity

Question3: Question4: what if observations are restricted in equatorial regions?



default



observations restricted in equatorial

Anwser3: off-equtorials still have signal, observation can be spread to off-equatorial region, RMSD a bit worse for h and v, but u is better

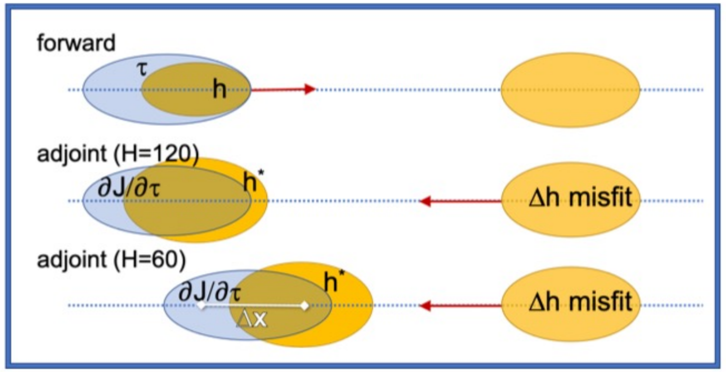
Lab2 4Dvar task3 imperfect model: layer thickness

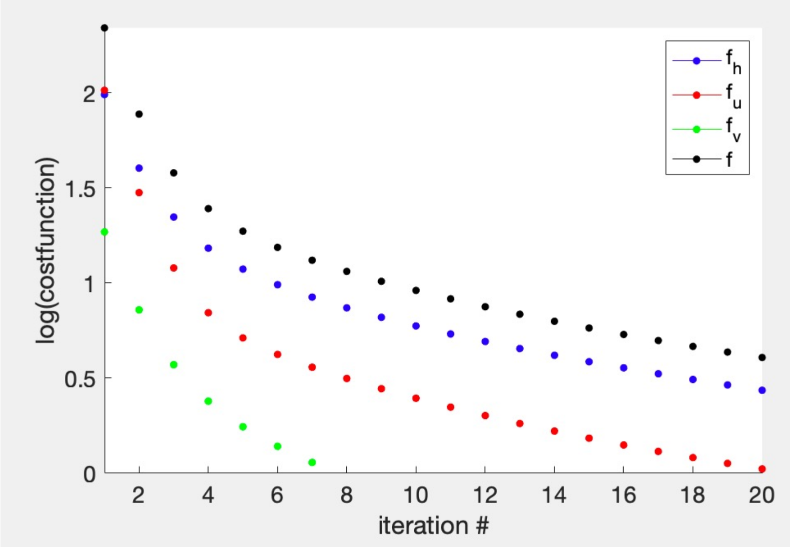
Question 4: how is the influence of layer thickness error?

experiments	thickness H
1 default	120 m
2 imperfect	60 m in the adjoint but with observations created with H=120

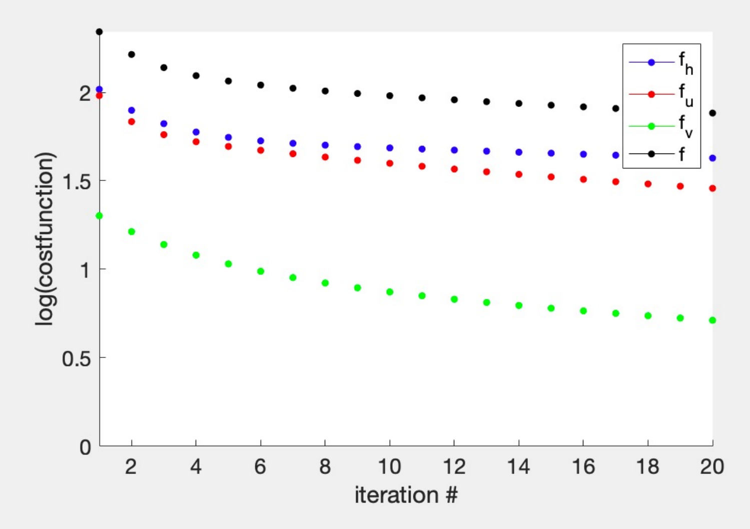
Influence of layer thickness error

- Default experiment: H=120m, g' =0.006; c = $\sqrt{g'H}$ = 0.85 m/s
- H=60m ⟹ c=0.60 m/s





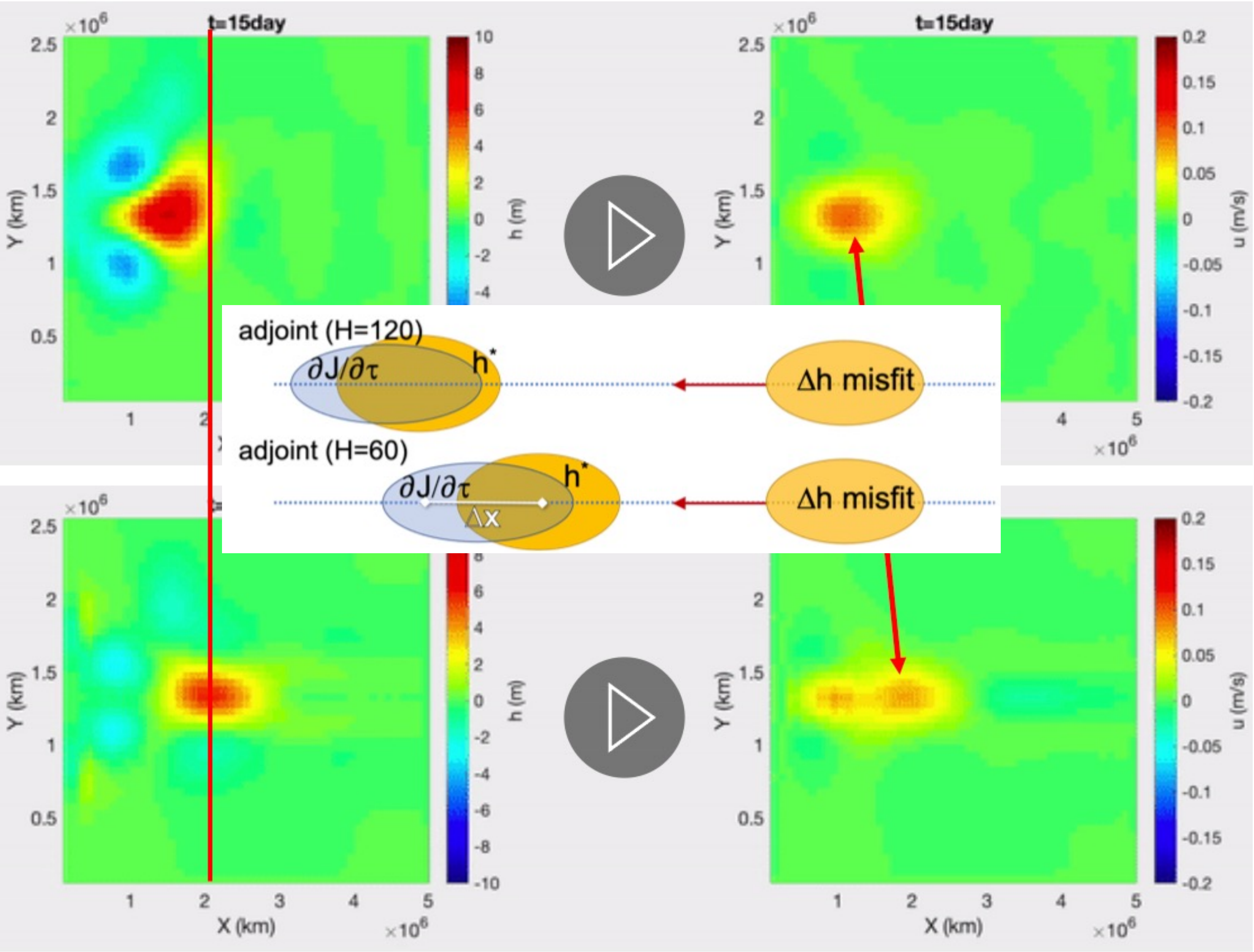
Default: adjoint H=120



Imperfect model: adjoint H=60

fail of converge of const function

Comparison speed of Kelvin wave at last iteration



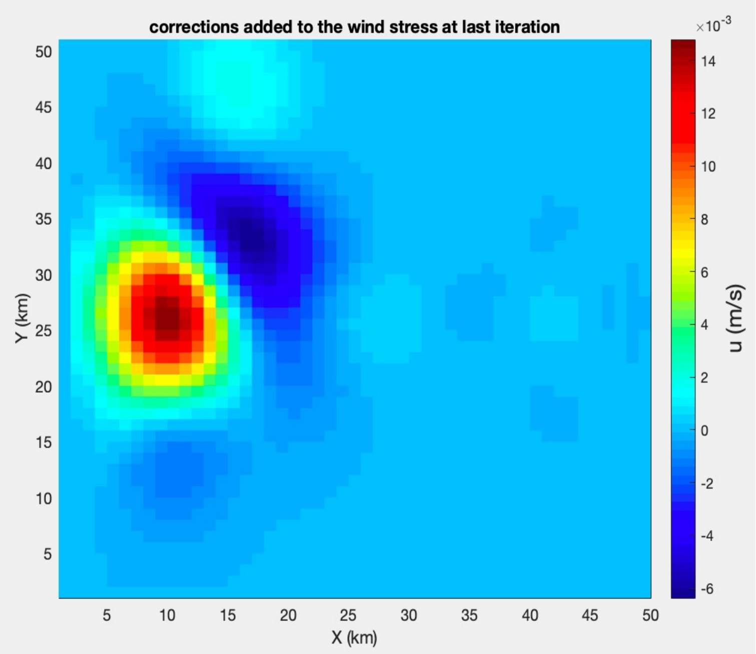
Default

Imperfect model

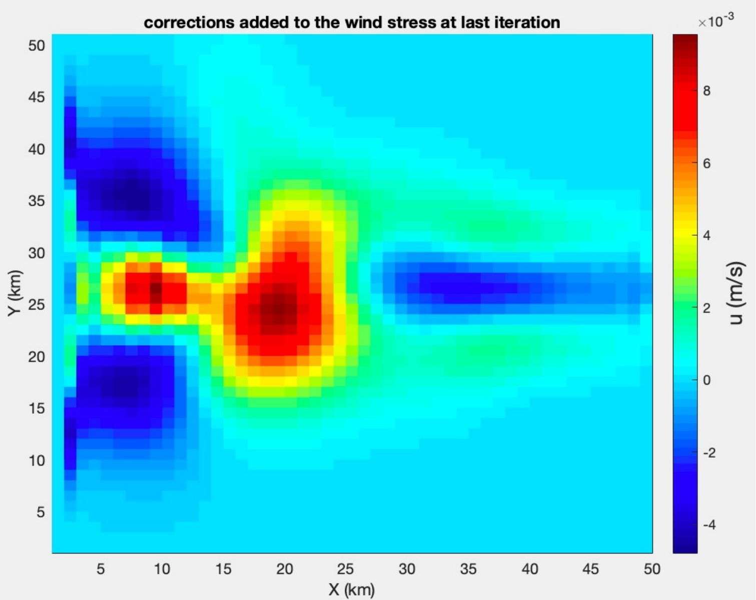
thermocline h shallower
Kelvin wave speed smaller

correction added to the wind stress at last iteration

Default H=120



Imperfect model
H=60



Fail to recover, a mismatch of distance

Lab2 4Dvar task4 Sparse data, B-Matrix transformation

Question 5: If we only have sparse observation data, can B-matrix transformation improve result?(compare 1 &2)

experiment	configurataion
1	sparse data in u,v,h
2	sparse data in u,v,h with B-matrx transformation(L=1.5)
3	sparse data in u,v,h with B-matrx transformation (L=3.0)

Original Image (Left) Vs. Gaussian Filtered Image (Right)

imgaussfilt(A,sigma)

```
%-----------------------------------------------------
% B-matrix transformation
%-----------------------------------------------------
grad = B_matrix(grad);
tau=tau+B_matrix(correc);
```

B-Matrix

$\boldsymbol{B}^{bg}$ is part of the background term of the costfunction which is not (and will not be) implemented. But we will transform the variable such that in the new space $\boldsymbol{B}^{bg}$ becomes unity:

$$\widetilde{W} = \boldsymbol{E}^{bg^{-\frac{1}{2}}} W$$

$$\mathbf{B} = \sigma_b^2 \begin{pmatrix} \gamma_{21} & 1 & \cdots & \gamma_{2N} \\ \vdots & & \ddots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \cdots & 1 \end{pmatrix}$$

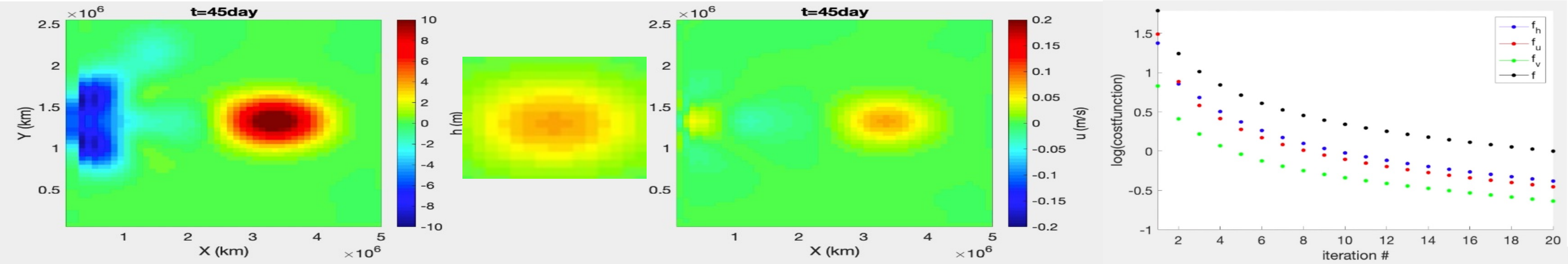
$$\gamma_{ij} = \exp\left(-\left(r_{ij}/L\right)^2\right)$$

Question 6: which L is reasonable to assume? (compare 2 & 3)

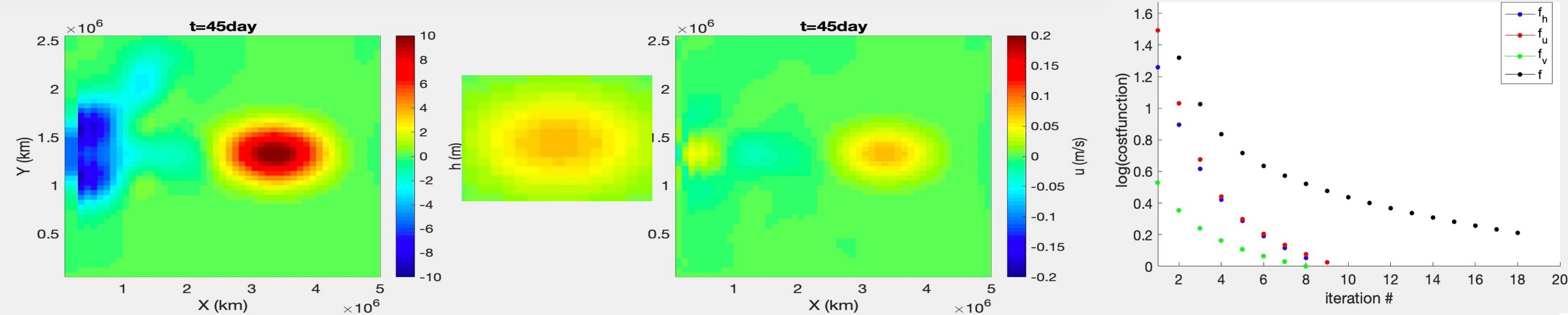
Question 5: If we only have sparse observation data, will B-matrix transformation improve result?(compare 1 &2)

Last iteration

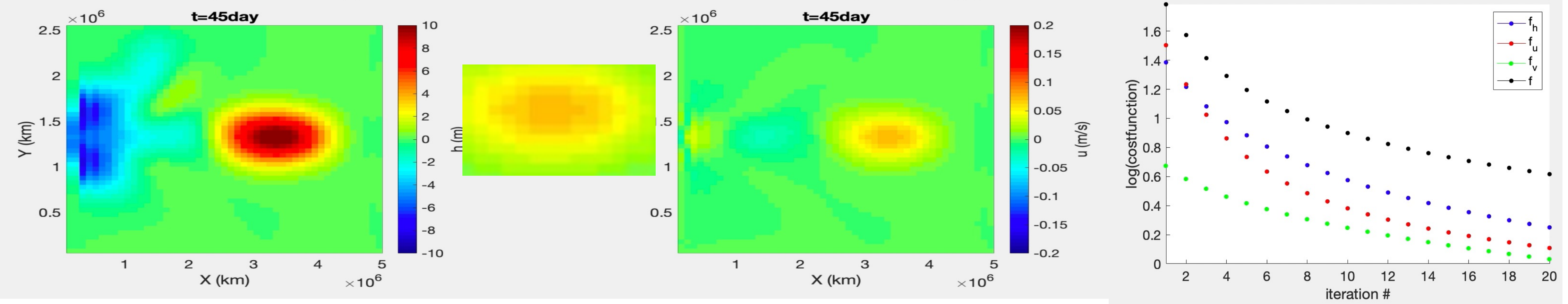
Without B-matrix



With B L=1.5



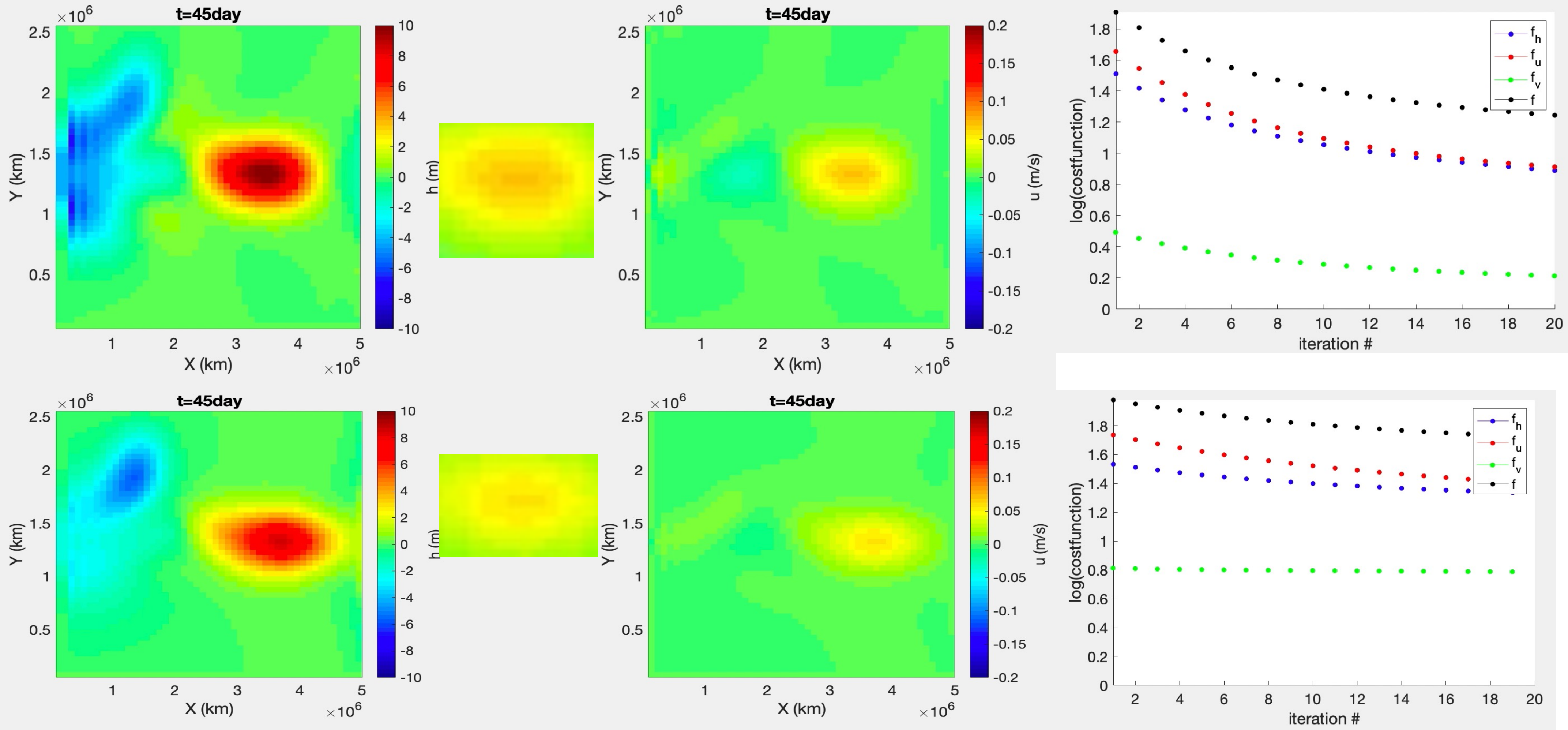
With B L=3.0



Question 5: If we only have sparse observation data, will B-matrix transformation improve result?(compare 1 &2)

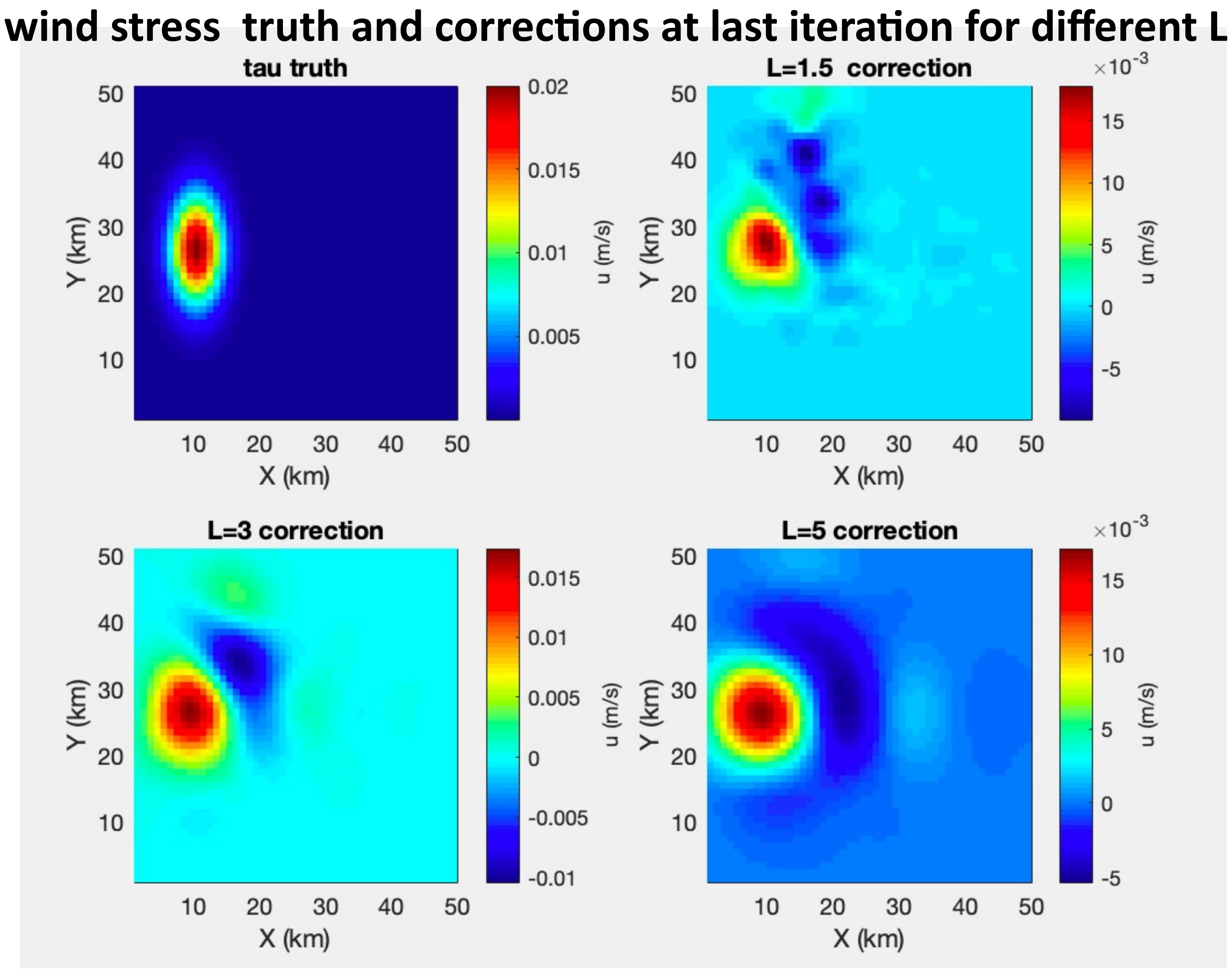
With B
L=5

With B
L=10



Answer 5: implementing B-matrix transformation will smooth the result
If L is too large, cost function may fail

Question 6: which L is reasonable to assume?



Answer 6:larger L produces larger correction, both 1.5 and 3.0 are okay, but L=3 is better due to coherence

Summary

	method	solution	observation	analsyis time	uncertainty
OI	least square	best linear estimator	sequential	instant	fixed in time
4DVar	variational	cost minimization	smoother	continuous	evolve in time

Role of B-matrix

- information spreading(E.g. in data-sparse region)
- information smoothing(E.g. Gaussian filtering)
- … …

Role of observations

- resources of errors
- interpolation error acts at observation
- … …

Good results require accurately estimating the necessary statistics (R and B) and careful quality control of the observations （y).

Thank you !