FINNISH METEOROLOGICAL INSTITUTE

# Error statistics for Data Assimilation in NWP: perspectives for snow

**E. Kourzeneva[1], M. Choulga[2,3], L. Rontu[1]**

(1) Finnish Meteorological Institute,
(2) European Center for Medium-range Weather Forecasts,
(3) Russian State Hydrometeorological University

ESSEM COST Action ES1404 Worksop,
Budapest, October 30-31, 2018

# Outlines:

- Introduction
- What are errors in DA?
- Errors and variability on different scales
- Example of calculating errors
- Conclusions

# Background

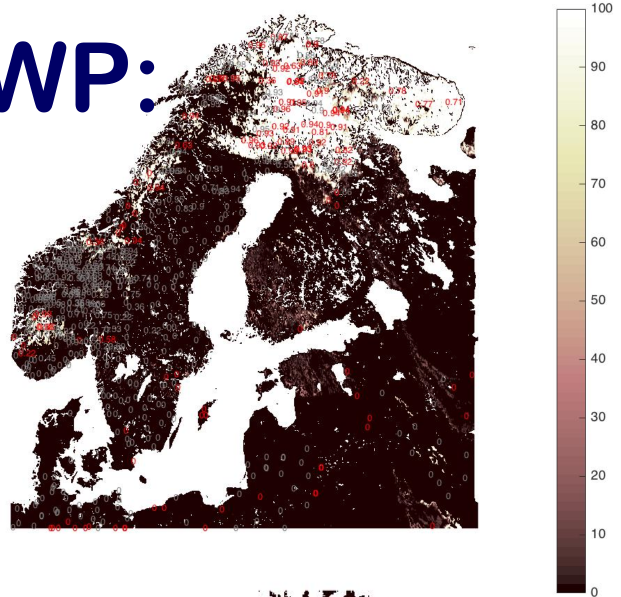
- Snow cover is very complicated in terms of parameters and scales

- Errors are in observations and in modeling

- We never know the truth! But we know something ...

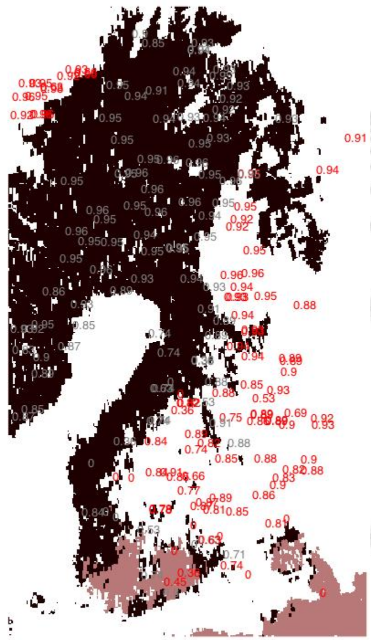
- What errors depend on?
  - on our methods?
  - on the complexity of a situation, i.e. on the variability of the parameter to measure?

# Snow Data Assimilation in NWP: what is important ?




- Physical aspect
  Snow variables: the snow depth, SWE, the snow density, the albedo, etc.

- Geographical aspect
  On global, regional, meso-scales (down to 1 km)



- In most cases, we understand sources of errors in modeling and in observations. But for DA, we need error statistics.

# What are errors in Data Assimilation?

$\mathbf{x}$ - the state vector

All model (prognostic) variables in all grid points.

Defines the model space.

$\mathbf{x}_t$ - the true state vector

Always unknown!

Not yet the whole truth, but a representation of the truth, because of discretization.

$\vec{\varepsilon} = \mathbf{x} - \mathbf{x}_t$ - the vector of errors

Always unknown!

But we consider it as a random variable and try to estimate its statistics.

For this, sometimes we use the ergodicity assumption: the averaging in realizations is substituted by the averaging in time or space.

# What are errors in Data Assimilation?

$\mathbf{x}_b$ - the background vector

Values of the state vector in the specific model run (or gridded climatology).

$\mathbf{y}$ - the vector of observations

All observations in all locations.

Defines the observational space

$H(\mathbf{x}_t)$ - the observational operator

Projection of the true state vector in the observational space.

Always unknown!

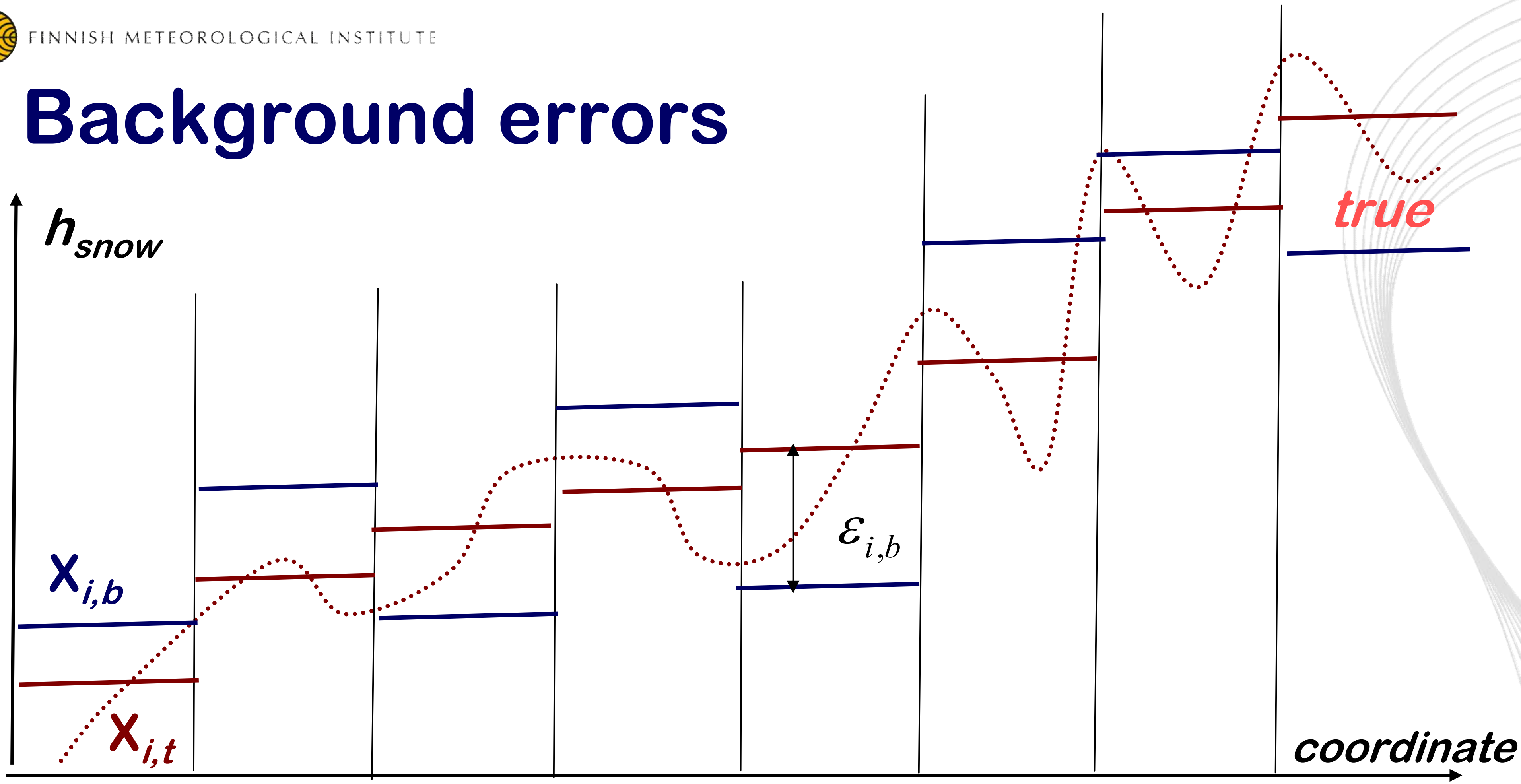
$\vec{\varepsilon}_b = \mathbf{x}_b - \mathbf{x}_t$ - the vector of background errors

$H(\mathbf{x})$ - the observational operator

Projects the model space into the observational space.

$\vec{\varepsilon}_o = \mathbf{y} - H(\mathbf{x}_t)$ - the vector of observational errors

# Background errors

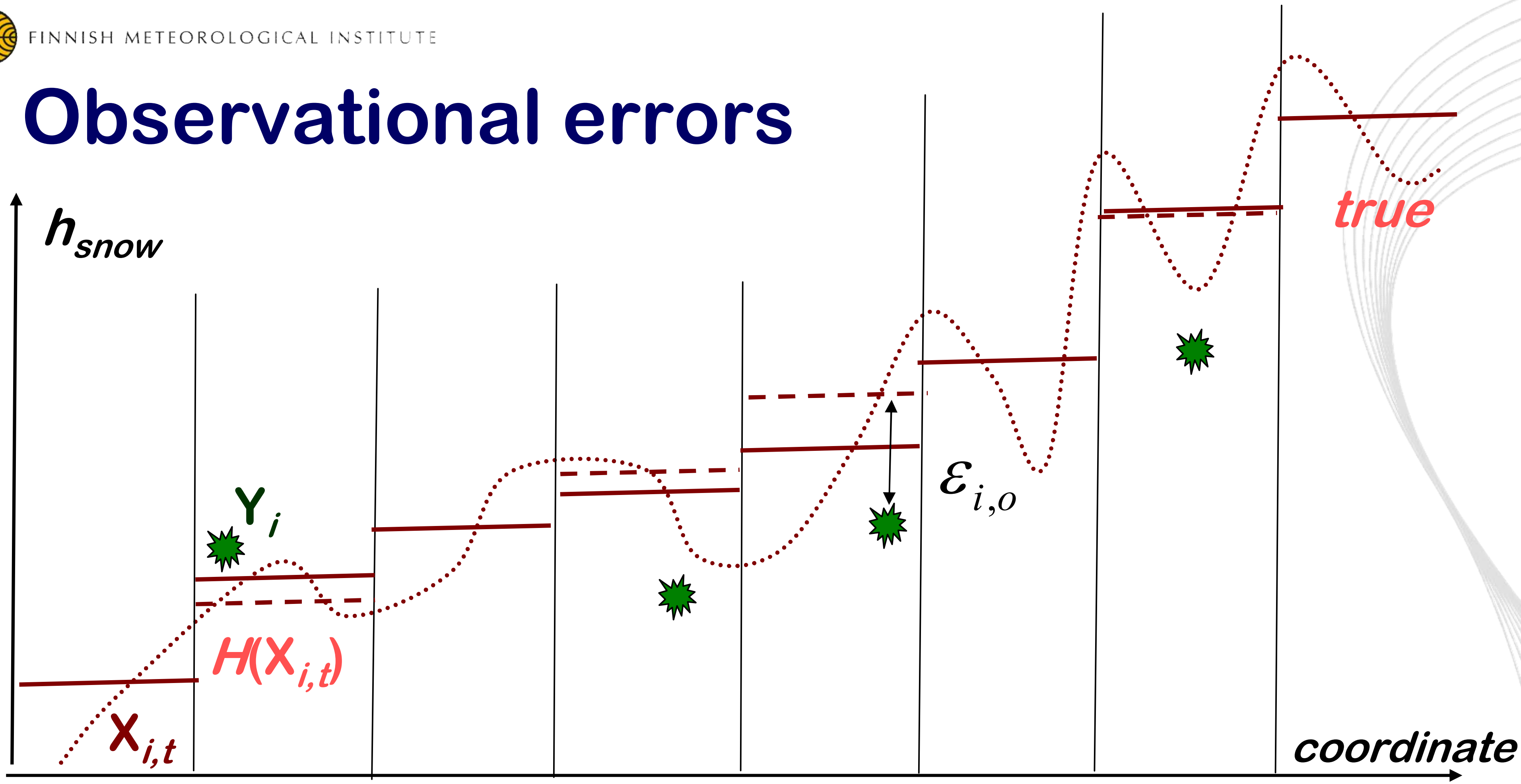



**Appear from the model assumptions, parameters and the initial state.**

**Not from discretization, because for $x_b$ and $x_t$ discretization is the same.**

**Still, depend on a model resolution as a model parameter.**

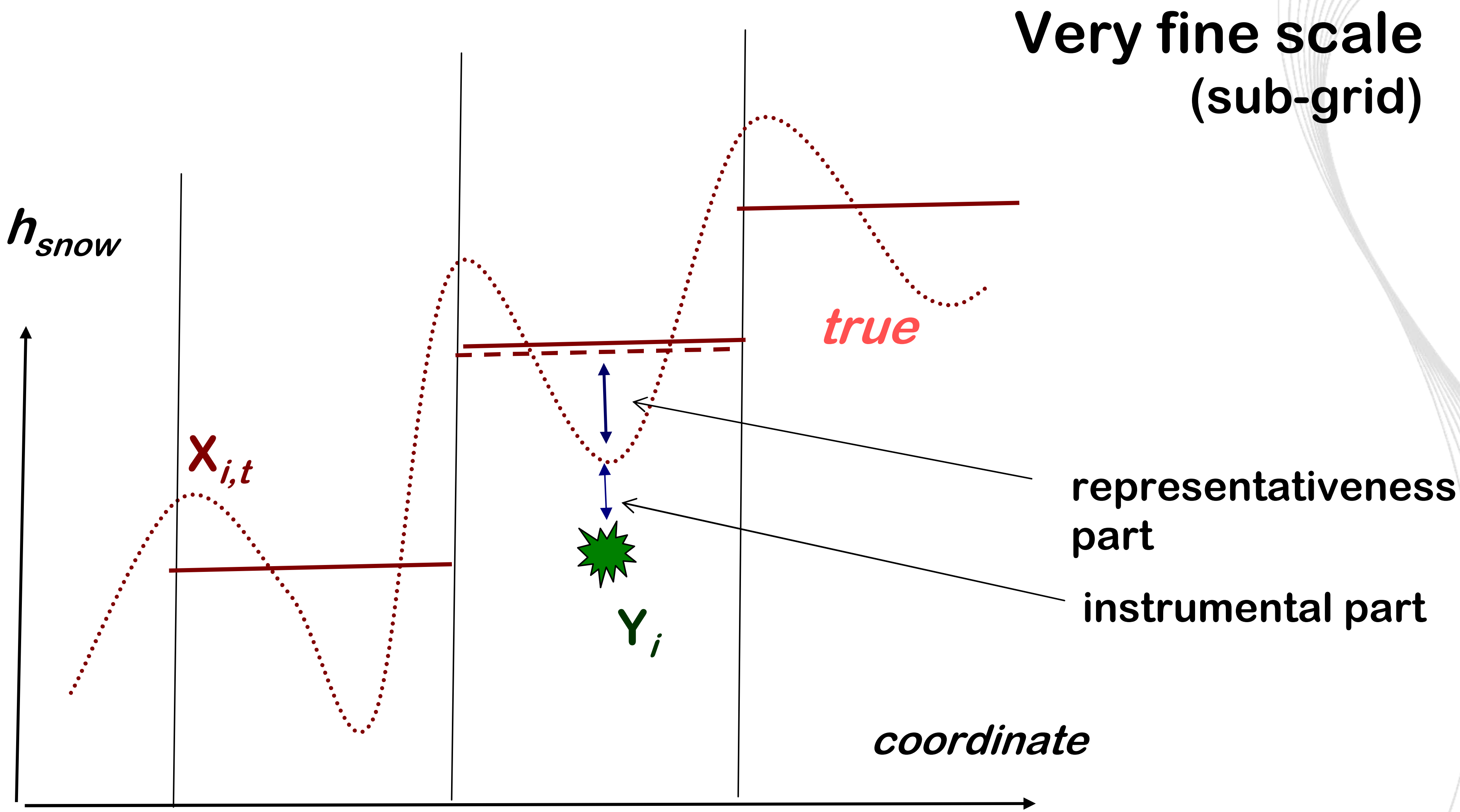
# Observational errors




**Appear from instrumentation and the observation operator (internal algorithms, interpolation).**

**Depend on the model resolution and on the field variability (not on the distance between observations).**

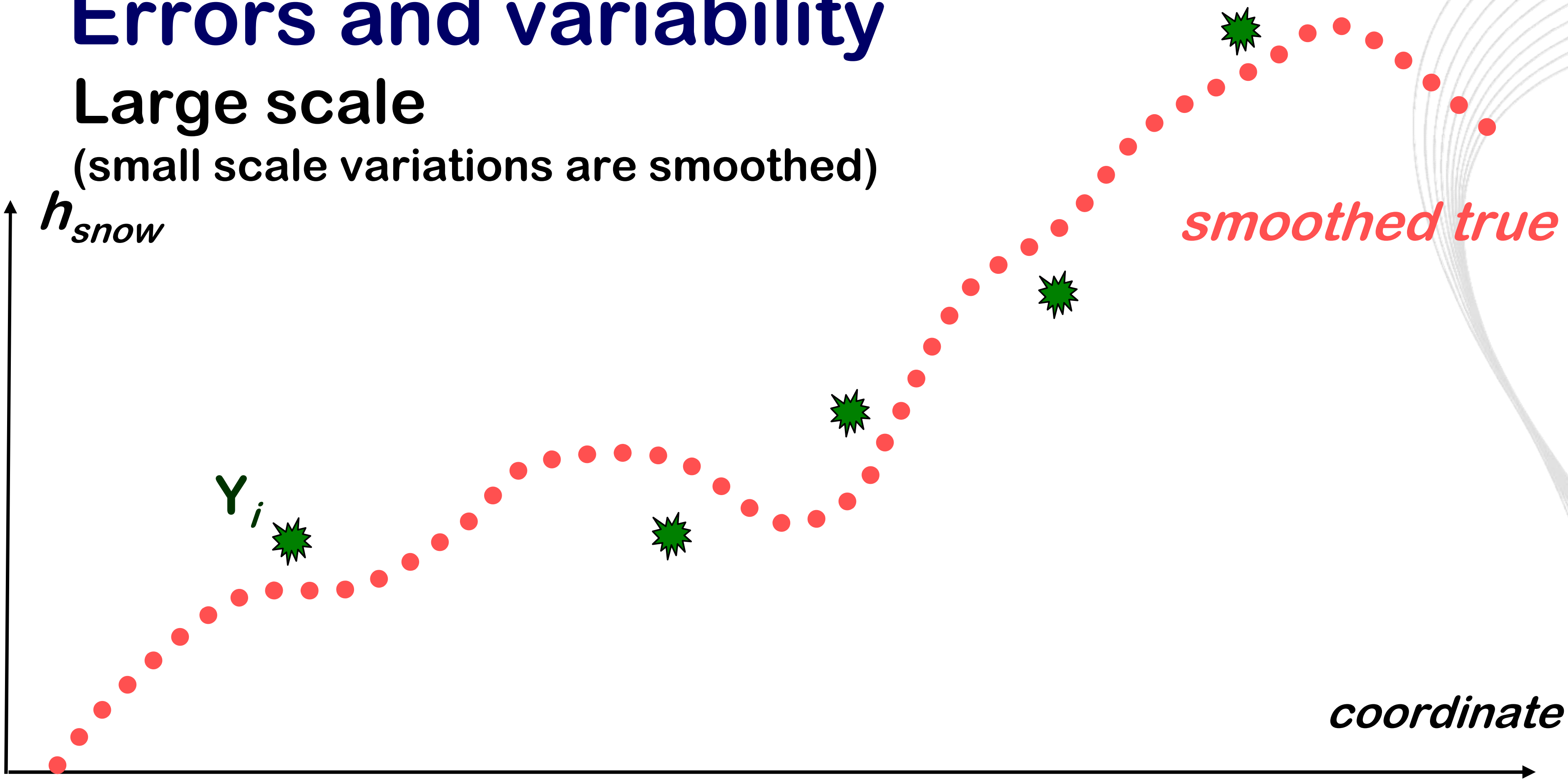
# Errors and variability



# Errors and variability

## Large scale

(small scale variations are smoothed)




The representativeness error is a matter of scales.
It is dependent on application.
If we have obs. dataset with a certain density (reflecting our scales of interest), we may calculate the obs. error statistics from it.

# Error statistics used in DA

**Basic assumption: background errors and obs. errors are uncorrelated.**
**Usual assumption: the distribution of errors is Gaussian.**

$m_b = 0$ **- the mean background error**

**Assumed to be zero. If not, bias correction is needed.**

$\sigma_b^2$ **- the background error variance**

**B** **- the matrix of the background error covariance's (correlations) between state vector members**

$m_o = 0$ **- the mean obs. error (systematic error)**

**Assumed to be zero. If not, bias correction is needed.**

$\sigma_o^2$ **- the obs. error variance (random error)**

**Gross errors in obs. should be removed by the quality control !**

**R** **- the matrix of the obs. error covariance's (correlations), e.g. between obs. locations**

# Error statistics used in DA

## Usual assumptions:

Often the **B** - matrix is calculated assuming that the normalized error correlations μ depend only on the distance ρ:

$$\mu(\rho) = \exp\left(-\frac{\rho^2}{2L^2}\right)$$

$L$ - the Gaussian scale.

Often the **R** - matrix is calculated assuming that the obs. error are uncorrelated:

$$\mathbf{R} = \sigma_o^2 \cdot \mathbf{I}$$

In the land-surface DA, for snow in particular, $\sigma_b^2$, $\sigma_o^2$ and $L$ are purely tuning parameters, depending on the model resolution and obs. density.
For example, in different limited area models, for the snow depth:

$$L = (30 \div 300)\ km$$

This is not the case in upper air DA!

# Methods to calculate error statistics:

- From observations (after Gandin, 1963)
- From obs. - background statistics (after Hollingsworth and Lönnberg, 1986)
- From NMC method
- From the model ensemble

… etc.

# Example of calculating error statistics for the lake surface temperature

.. after Gandin, 1963.

Statistics: $\sigma_o^2$

Parameters: $L$


## LSWT observations :

• 27 lakes
• 5 summers (JJA) of 2010–2014
totally, 12 227 observations

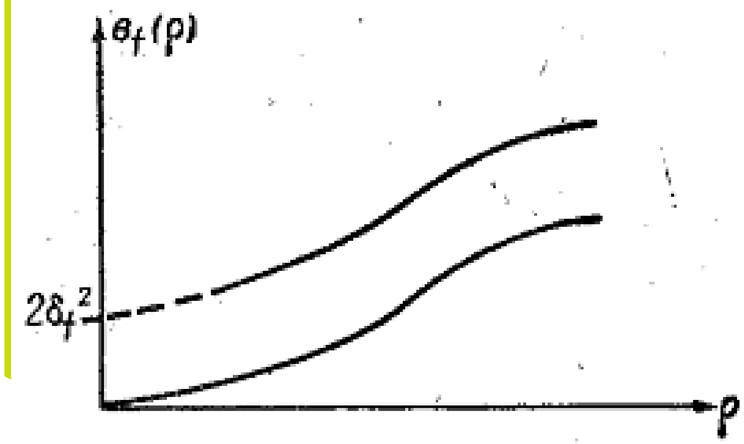
# Example of calculating error statistics of the lake surface temperature

**Basic idea of the method**: to use the fact, that obs. error increases the value of the structure function of the meteorological field calculated from the obs. dataset by $2 \cdot \sigma_o^2$ (shown by L. Gandin, 1963).

**Algorithm:**

- For each obs. location $r$, calculate the climatological norm of the LWST $\overline{f(r)}$
- For each obs. location, calculate deviations from the norm:

$$f'(r) = f(r) - \overline{f(r)}$$



# Example of calculating error statistics for the lake surface temperature

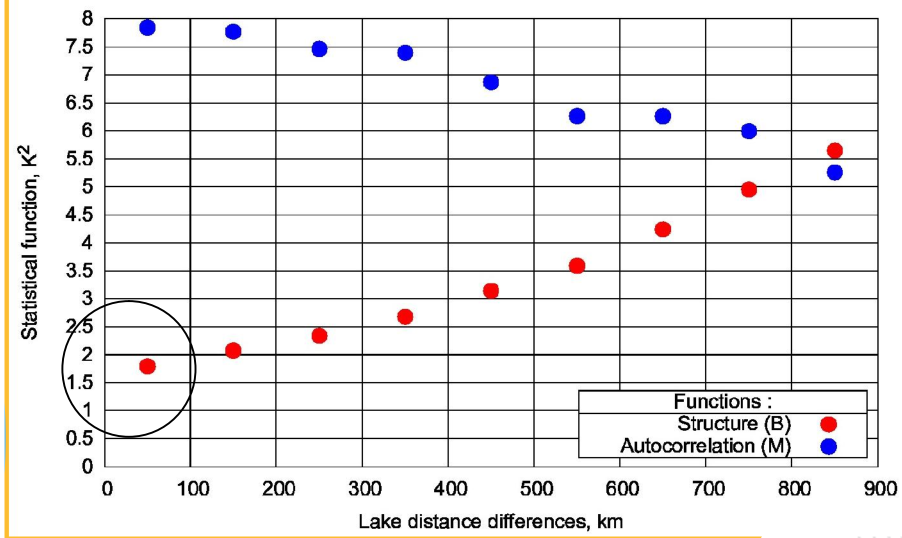
- Categorize pairs of lakes $r_1$ and $r_2$ according to the distance $\rho$ between them:
  [0-100], [100-200], … till 1600 km.

- Calculate the structure function within each category:

$$b(\rho) = \overline{\left(f'(r_1) - f'(r_2)\right)^2}$$

- Obtain the $b(0) = 2 \cdot \sigma_o^2$ value from the graph by extrapolation on the $y$ - axis.

- Calculate $\sigma_o^2$ .




$$2 \cdot \sigma_o^2 = 1.8\,\mathrm{K}^2 \qquad \sigma_o^2 = 0.9\,\mathrm{K}^2$$

- To obtain $L$ , we also calculate the autocorrelation function within each category:

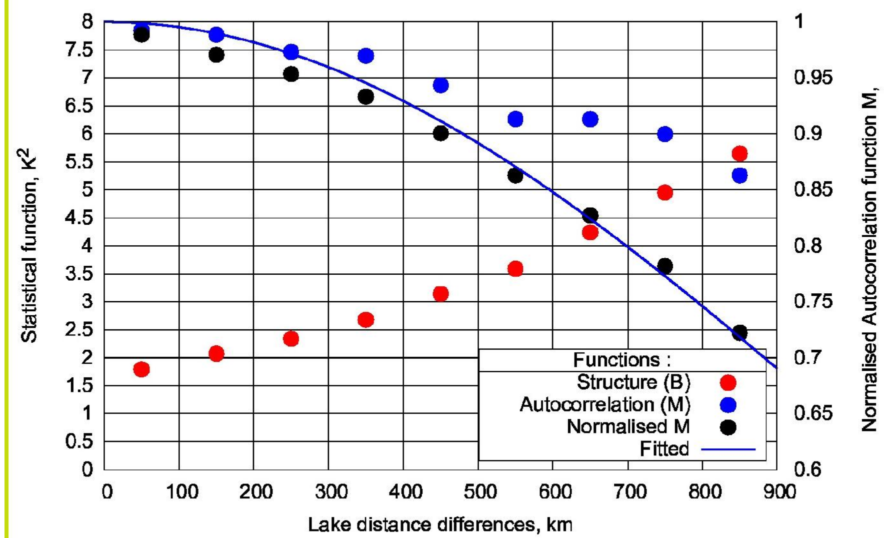
$$m(\rho) = \overline{f'(r_2) \cdot f'(r_2)}$$

# Example of calculating error statistics for the lake surface temperature

- **Calculate the normalized autocorrelation function for each category as:**

$$\mu(\rho) = \frac{m(\rho)}{\overline{f'^2} - \sigma_o^2}$$

- **Approximate $\mu$ by the Gaussian function and find $L$ .**




$$L = 1050\ km$$

**- 10 times larger then the value from tuning!**

**Kheyrollah Pour, H. et al., 2017: Towards improved objective analysis of lake surface water temperature in a NWP model: preliminary assesment of statistical properties, *Tellus A*, 69(1), http://dx.doi.org/10.1080/16000870.2017.1313025 .**

# Conclusion:

- **Error statistics for assimilation of snow observations in NWP are not tuning parameters, but they should be obtained objectively and studied.**

- **Error statistics may be calculated from observational datasets, forecast archives, ensembles, etc.**

# Thank you for your attention!

Questions?