


# CO2 inversion as a system and its uncertainty quantification

Lin Wu

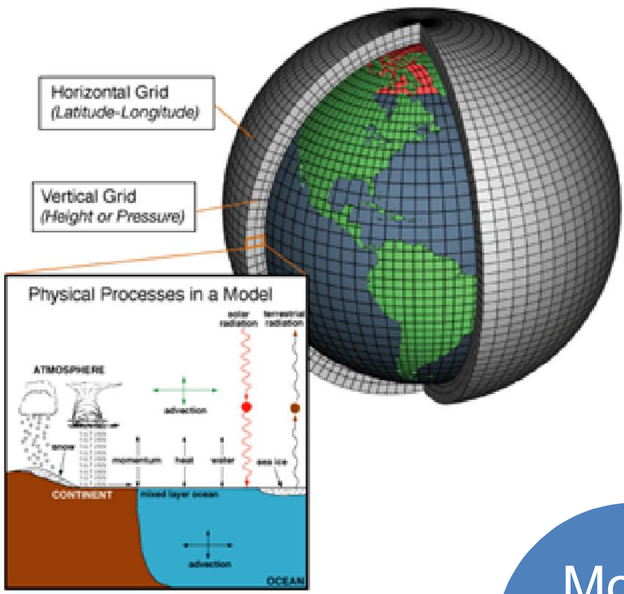
Marc Bocquet, Frédéric Chevallier, Thomas Lauvaux, Peter Rayer, Ken Davis

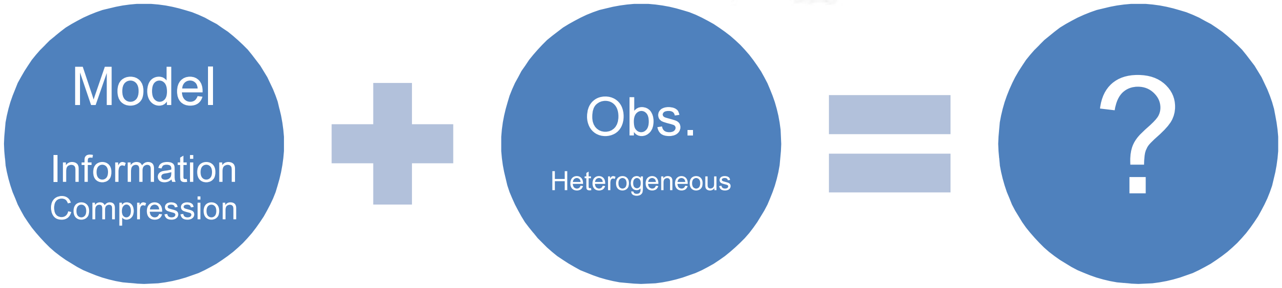
SOFIE Spring school

May 12 2014

BridGES VEOLIA ENVIRONNEMENT ThalesAlenia Space cea UNIVERSITÉ DE VERSAILLES ST-QUENTIN-EN-YVELINES LSCE

# CO2 story: one carbon cycle to understand

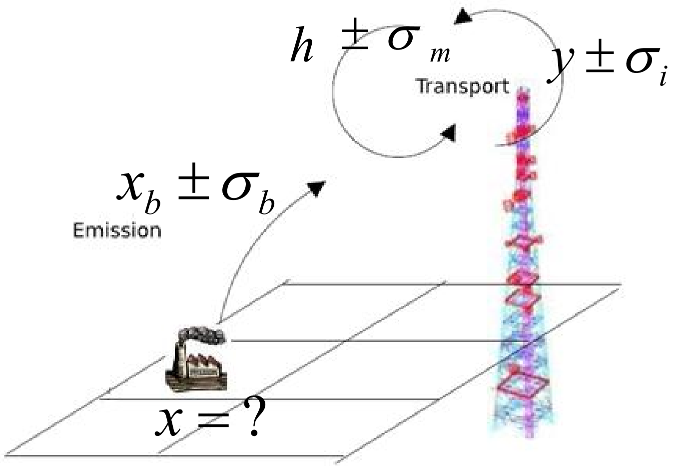



Fusion of informtion from diverse sources

# Simplest scalar case

Bouttier & Courtier 1999, Jacob 2007




Observation Eq.

$$y = hx + \varepsilon_o$$

$$\varepsilon_o = \varepsilon_i + \varepsilon_m$$

Information 1: first guess $x_b$

Information 2: observation $y$

Information? Probability distribution

First guess:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp[-\frac{1}{2\sigma_b^2}(x - x_b)^2]$$

Observation conditioned by emission:

$$p(y|x) = \frac{1}{\sqrt{2\pi}\sigma_o} \exp[-\frac{1}{2\sigma_o^2}(y - hx)^2]$$

Information fusion? Production rule of proba.

$$p(x, y) = p(x)p(y|x) = p(y)p(x|y)$$

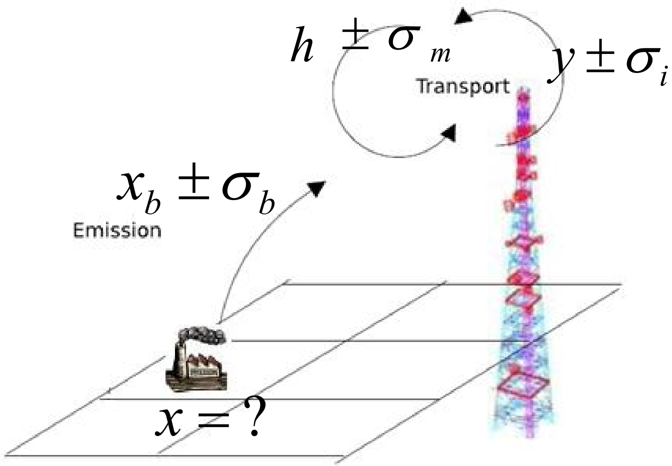
Inference? Bayes theorem/rule.

$$\underbrace{p(x|y)}_{\text{posterior}} = \frac{\overbrace{p(x)}^{\text{prior}}\,\overbrace{p(y|x)}^{\text{likelihood}}}{\underbrace{p(y)}_{\text{evidence}}}$$

*posterior ∝ likelihood × prior*

# Simplest scalar case

Bouttier & Courtier 1999, Jacob 2007




Bayesian calculus

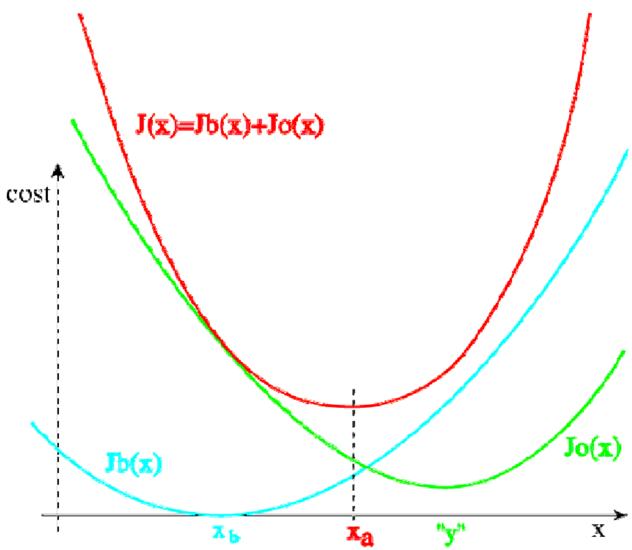
$$p(x|y) \propto \exp[-\frac{1}{2}(\frac{(x-x_b)^2}{\sigma_b^2} + \frac{(y-hx)^2}{\sigma_o^2})]$$

Estimation with posterior: find a criteria (MAP)

$$x_a = \arg\max_x p(x|y)$$

Calculus: minimization of a $\chi^2$ cost function

$$J(x) = \frac{1}{2}[\frac{(x-x_b)^2}{\sigma_b^2} + \frac{(y-hx)^2}{\sigma_o^2}]$$




Estimation:

$$x_a = x_b + k(y - hx_b)$$

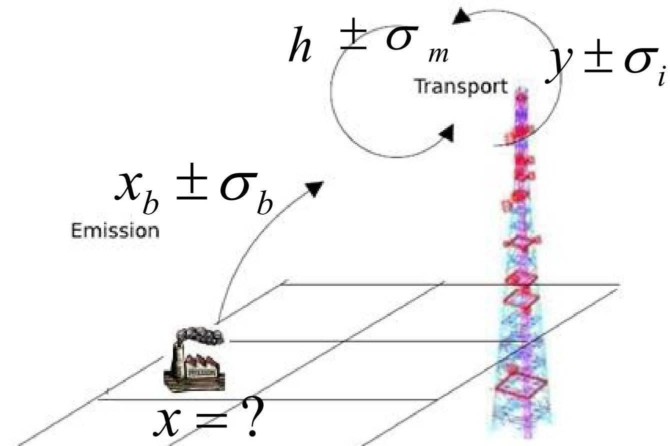
$$k = \sigma_b^2 h(h^2\sigma_b^2 + \sigma_o^2)^{-1}$$

(Kalman) gain $k = \frac{\partial x_a}{\partial y}$

- Sensitivity of analysis to obs
- Weighted by error statistics

# Simplest scalar case

Bouttier & Courtier 1999, Jacob 2007




Estimation: $$x_a = x_b + k(y - hx_b)$$

$$k = \sigma_b^2 h (h^2 \sigma_b^2 + \sigma_o^2)^{-1}$$

$$J(x) = \frac{1}{2}\left[\frac{(x - x_b)^2}{\sigma_b^2} + \frac{(y - hx)^2}{\sigma_o^2}\right]$$

Degree of freedom for signal (DFS)

$\sigma_o \ll \sigma_b$ $\quad \frac{(y - hx)^2}{\sigma_o^2} \uparrow$ **in** $J(x)$

$k \to 1/h \qquad x_a \to y/h$

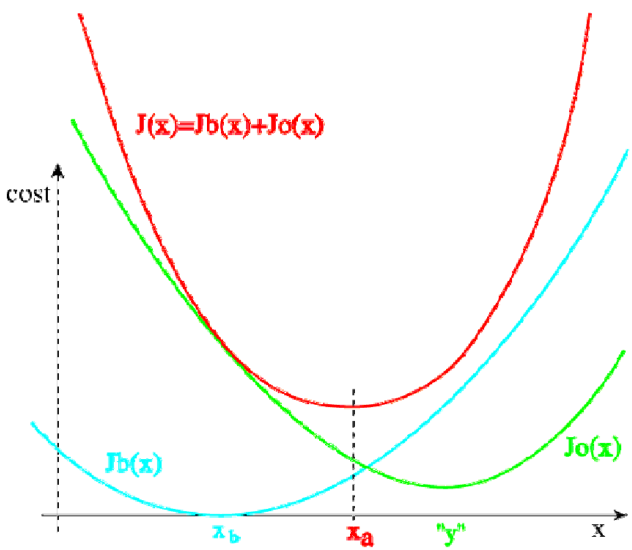
$y$ provides information on $x$

Degree of freedom for noise

$\sigma_o \gg \sigma_b$ $\quad \frac{(x - x_b)^2}{\sigma_b^2} \uparrow$ **in** $J(x)$

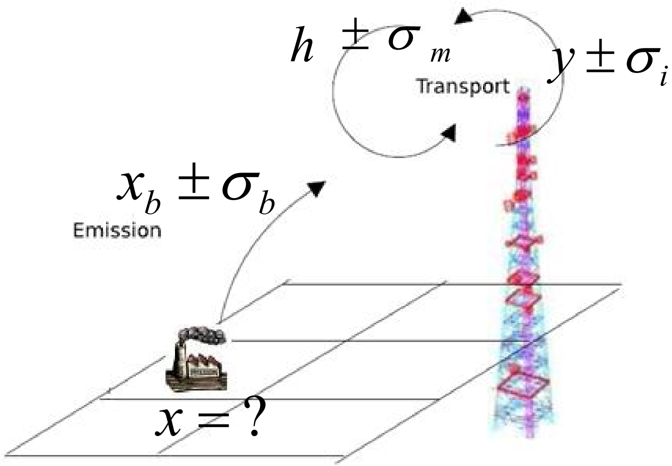
$k \to 0 \qquad x_a \to x_b$

$y$ provides only noise



# Simplest scalar case

Bouttier & Courtier 1999, Jacob 2007




Posterior uncertainty

$$p(x|y) \propto \exp[-\frac{1}{2}(\frac{(x-x_b)^2}{\sigma_b^2} + \frac{(y-hx)^2}{\sigma_o^2})]$$

$$= \exp(-\frac{1}{2}\frac{(x-x_a)^2}{\sigma_a^2})$$

$$(\sigma_a^2)^{-1} = (\sigma_b^2)^{-1} + h^2(\sigma_o^2)^{-1}$$

Sum of precision

Fisher info. matrix

Observation Eq.

$$y = hx + \varepsilon_o$$

$$\varepsilon_o = \varepsilon_i + \varepsilon_m$$

Information 1: first guess $x_b$

Information 2: observation $y$

Estimation:

$$y = hx \pm \sigma^o$$

$$x_a = x_b + k(y - hx_b)$$

$$x_a = ax + (1-a)x_b + k\varepsilon_o$$

Averaging kernel: $a = kh = \frac{\partial x_a}{\partial x}$

- Sensitivity of analysis to true emission
- Ideally 1

# A language of inversion: Bayesian synthesis

- Bayes' Theorem: uncertainty computation (information propagation) converting a prior probability to a posterior probability by assimilating Information from observations.

$$\underbrace{p(x|y)}_{\text{posterior}} = \frac{\overbrace{p(x)}^{\text{prior}}\,\overbrace{p(y|x)}^{\text{likelihood}}}{\underbrace{p(y)}_{\text{evidence}}}$$

  - y : observation
  - x : unknown parameter (source)

- Bayesian analysis in plain words

posterior ∝ likelihood × prior

# Bayesian inversion: vectorial case of linear dynamics and Gaussian error

Inverse modelling of sources $\mathbf{x}$ (2D+T); Gaussian assumption + linear observation operator.

- $\mathbf{H}$ Jacobian matrix of the problem (observation + model):

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \varepsilon$$

- $\mathbf{x} - \mathbf{x}_b \propto \mathrm{N}(\mathbf{0}, \mathbf{B})$ $\mathbf{x}_b$ prior fluxes, $\mathbf{B}$ background error covariance matrix.

- $\varepsilon \propto \mathrm{N}(\mathbf{0}, \mathbf{R})$ $\mathbf{R}$ observation error covariance matrix.

# Bayesian inversion: vectorial case of linear dynamics and Gaussian error

- Bayes' Theorem: $$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})\,p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} \quad \mathbf{x} \in \Re^n \quad \mathbf{y} \in \Re^d$$

  - Prior: $$p(\mathbf{x}) = \frac{\exp(-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b))}{2\pi^{\frac{n}{2}}|\mathbf{B}|^{\frac{1}{2}}}$$

  - Likelihood $$p(\mathbf{y}|\mathbf{x}) = p(\overbrace{\mathbf{y}-\mathbf{H}\mathbf{x}}^{\varepsilon}) = \frac{\exp(-\frac{1}{2}\varepsilon^T \mathbf{R}^{-1}\varepsilon)}{2\pi^{\frac{d}{2}}|\mathbf{R}|^{\frac{1}{2}}}$$

  - Evidence $$p(\mathbf{y}) = \int p(\mathbf{x})\,p(\mathbf{y}-\mathbf{H}\mathbf{x})\,d\mathbf{x}$$
$$= \frac{\exp(-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T (\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{x}-\mathbf{x}_b))}{2\pi^{\frac{d}{2}}|\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^T|^{\frac{1}{2}}}$$

  - Posterior: $$p(\mathbf{x}|\mathbf{y}) = \frac{\exp(-\frac{1}{2}(\mathbf{x}-\mathbf{x}_a)^T \mathbf{P}_a^{-1}(\mathbf{x}-\mathbf{x}_a))}{2\pi^{\frac{n}{2}}|\mathbf{P}_a|^{\frac{1}{2}}}$$

# Vectorial analog of the simplest scalar case

| | | |
|---|---|---|
| **Cost function** | $J(x)=\frac{1}{2}[\frac{(x-x_b)^2}{\sigma_b^2}+\frac{(y-hx)^2}{\sigma_o^2}]$ | $J(\mathbf{x})=\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)+(\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))$ |
| **Inversion** | $x_a=x_b+k(y-hx_b)$ | $\mathbf{x}_a=\mathbf{x}_b+\mathbf{K}(\mathbf{y}-\mathbf{H}\mathbf{x}_b)$ |
| **Kalman gain** | $k=\sigma_b^2h(h^2\sigma_b^2+\sigma_o^2)^{-1}$ | $\mathbf{K}=\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})^{-1}$ |
| | or equivalently | $\mathbf{K}=(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+\mathbf{B}^{-1})\mathbf{H}^T\mathbf{R}^{-1}$ |
| **Aver. Kernel** | $a=kh$ | $\mathbf{A}=\mathbf{K}\mathbf{H}$ |
| **DFS** | $\frac{(y-hx)^2}{\sigma_o^2}\uparrow$ **in** $J(x)$ | $E[(\mathbf{x}_a-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}_a-\mathbf{x}_b)]$ |
| **DoF Noise** | $\frac{(x-x_b)^2}{\sigma_b^2}\uparrow$ **in** $J(x)$ | $E(\varepsilon^T\mathbf{R}^{-1}\varepsilon)$ |
| **Fisher Info. Mat. (precision)** | $(\sigma_a^2)^{-1}=(\sigma_b^2)^{-1}+h^2(\sigma_o^2)^{-1}$ | $\mathbf{P}_a^{-1}=\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ |
| **Posterior Err. Cov. Mat.** | $\sigma_a^2=(1-kh)\sigma_b^2$ | $\mathbf{P}_a=(\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{B}$ |

# More on DFS

$$\mathbf{DFS} = E[(\mathbf{x}_a - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b)]$$

$$= E\left\{\mathbf{tr}\left[(\mathbf{x}_a - \mathbf{x}_b)(\mathbf{x}_a - \mathbf{x}_b)^T\right]\mathbf{B}^{-1}\right\}$$

$$= \mathbf{tr}\left\{E\left[(\mathbf{x}_a - \mathbf{x}_b)(\mathbf{x}_a - \mathbf{x}_b)^T\right]\mathbf{B}^{-1}\right\}$$

$$= \mathbf{tr}\left\{\mathbf{K}E\left[(\mathbf{y} - \mathbf{H}\mathbf{x}_b)(\mathbf{y} - \mathbf{H}\mathbf{x}_b)^T\right]\mathbf{K}^T\mathbf{B}^{-1}\right\}$$

$$= \mathbf{tr}\left\{\mathbf{K}(\mathbf{HBH}^T + \mathbf{R})\mathbf{K}^T\mathbf{B}^{-1}\right\}$$

$$= \mathbf{tr}(\mathbf{KH}) = \mathbf{tr}(\mathbf{A})$$

Trace of averaging kernel

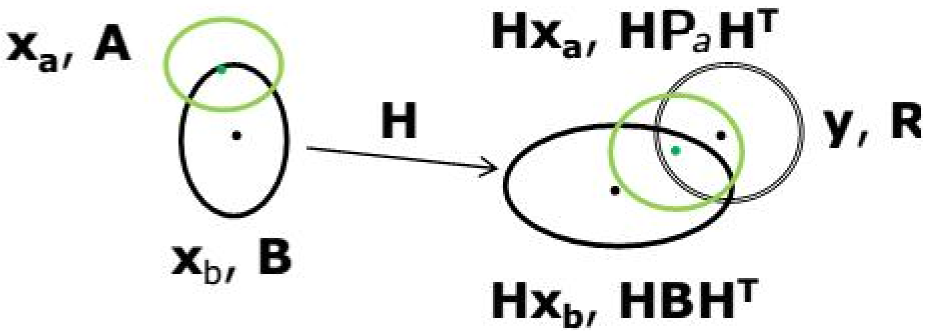
$$= \mathbf{tr}\left[(\mathbf{B} - \mathbf{P_a})\mathbf{B}^{-1}\right]$$

Reduction of uncertainty

$$= \mathbf{tr}\left(\overbrace{\mathbf{BH}^T\underbrace{(\mathbf{HBH}^T + \mathbf{R})^{-1}}_{\text{Info. from obs.}}}^{\mathbf{K}}\mathbf{H}\right)$$

Propagation of informatioin

# Inversion methods




Analytical inversion: linear algebra, maximal 5000-10000 parameters

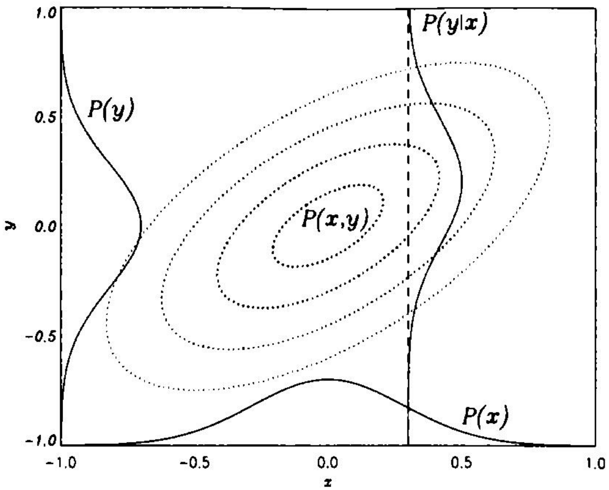
$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$
$$\mathbf{P}_a = (\mathbf{I} - \mathbf{KH})\mathbf{B}$$

Variational Analysis: Gaussian assumptions + MAP ⇒ least square errors (Gauss' result)
Numerical optimization; easily dealing with a million parameters; adjoint techniques

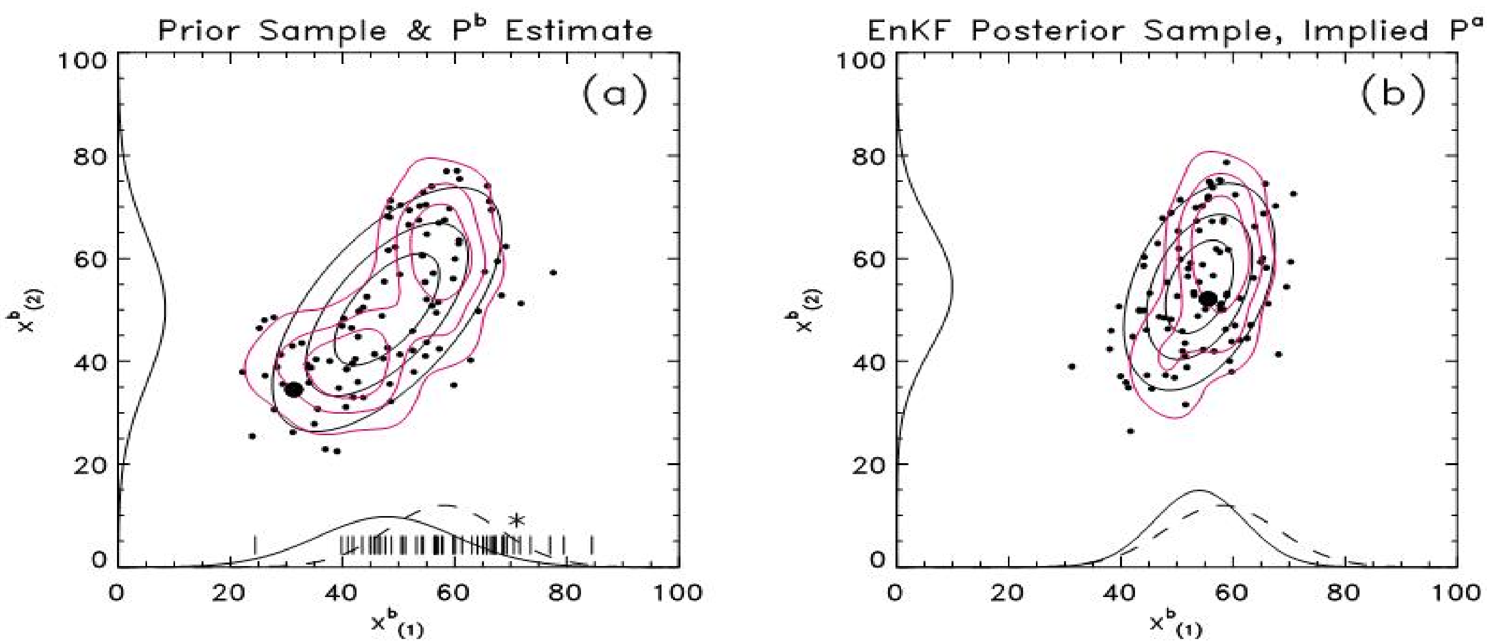
$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))$$
$$\mathbf{P}_a = \left(\frac{1}{2}J''\right)^{-1}$$

Ensemble approach: representing PDFs with samples of manageable size

# Sketch of Bayesian Synthesis




- Red: true prior and posterior
- Points: samples
- Contours: Gaussian prior and posterior
- Obs for $x_1$




Hamil 2006

# Imporatant roles of $\mathbf{B}$ and $\mathbf{R}$

Fundamental role of $\mathbf{B}$ : corrections only in the column space of $\mathbf{B}$ !

Kalnay 2003

$\mathbf{B}$ spanned by a single vector $\mathbf{b}$

$$\mathbf{B} = \mathbf{b}\mathbf{b}^T$$

Suppose $\mathbf{H} = \mathbf{I}, \quad \mathbf{R} = \alpha^2 \mathbf{I}$

$$\begin{aligned}\delta \mathbf{x}_a &= \mathbf{x}_a - \mathbf{x}_b \\ &= \mathbf{B}\mathbf{H}^T \, [\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1} \, [\mathbf{y}_o - H(\mathbf{x}_b)] \\ &= \mathbf{b} \, \mathbf{b}^T \delta \mathbf{y}_o / (\mathbf{b}^T \mathbf{b} + \alpha^2)\end{aligned}$$

Correction Direction; Information spreading

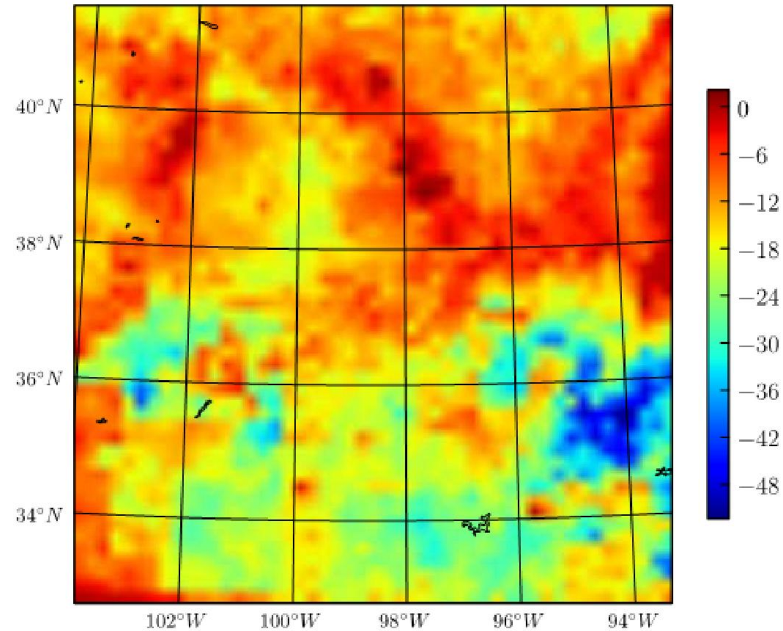
Projection of obs increment in (subspace) $\mathbf{b}$ info smoothing

Total error budget

Realistic $\mathbf{B}$ : difficult issue

Physical consistency: balanced $\mathbf{B}$

Sum of prior SiBcrop fluxes
Over 1-15 June 2007
Center USA 980km x 980 km




Balgovind correlation model

$$C(h) = \kappa^2 (1 + \frac{h}{L}) \exp(-\frac{h}{L})$$

# Diagnostics of error

Desroziers et al 2005

$$\mathbf{d}_b^o = \mathbf{y}^o - H(\mathbf{x}^b)$$
$$= \mathbf{y}^o - H(\mathbf{x}^t) + H(\mathbf{x}^t) - H(\mathbf{x}^b)$$
$$\simeq \boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^b$$

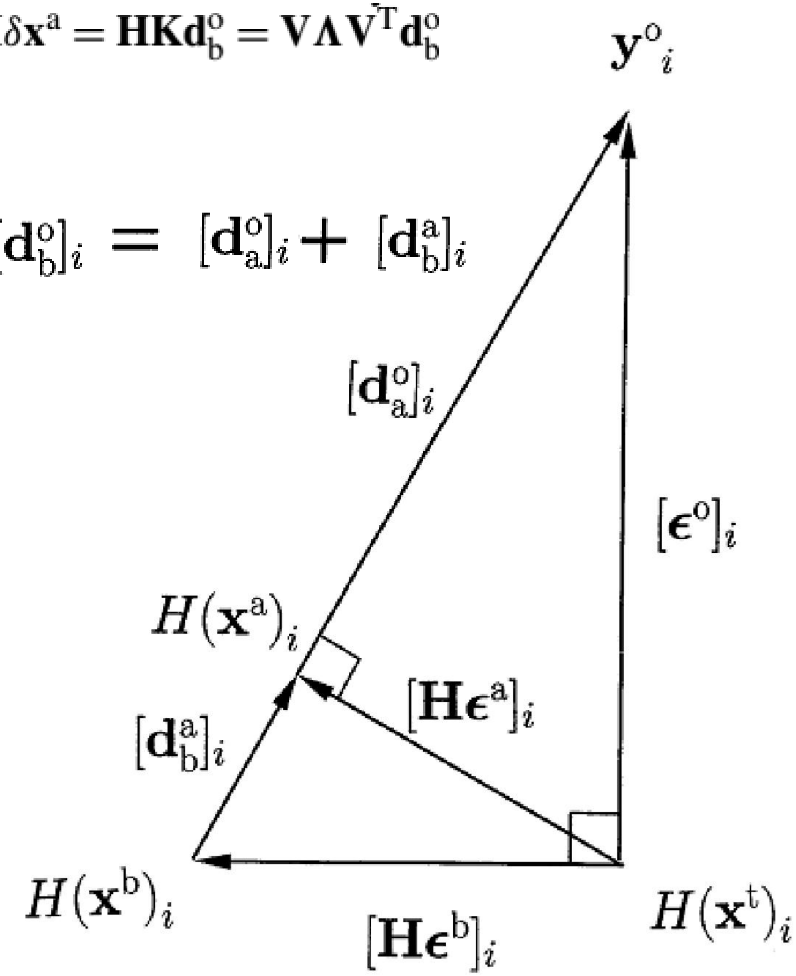
$$\mathbf{d}_b^a = \mathbf{H}\delta\mathbf{x}^a = \mathbf{HK}\mathbf{d}_b^o = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{d}_b^o$$

$$E[\mathbf{d}_b^o(\mathbf{d}_b^o)^T]$$
$$= E[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^T] + \mathbf{H}E[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^T]\mathbf{H}^T$$
$$= \mathbf{R} + \mathbf{HBH}^T$$

$$[\mathbf{d}_b^o]_i = [\mathbf{d}_a^o]_i + [\mathbf{d}_b^a]_i$$

$$\mathbf{d}_a^o = \mathbf{y}^o - H(\mathbf{x}^b + \delta\mathbf{x}^a)$$
$$\simeq \mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{HK}\mathbf{d}_b^o$$
$$= (\mathbf{I} - \mathbf{HK})\mathbf{d}_b^o$$
$$= \mathbf{R}(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{d}_b^o,$$

$$E[\mathbf{d}_a^o(\mathbf{d}_a^o)^T] = \mathbf{R} + \mathbf{H}\mathrm{P}_a^{-1}\mathbf{H}^T$$




$$E[\mathbf{d}_b^a(\mathbf{d}_b^o)^T] = \mathbf{HBH}^T$$
$$E[\mathbf{d}_a^o(\mathbf{d}_b^o)^T] = \mathbf{R}$$
$$E[\mathbf{d}_b^a(\mathbf{d}_a^o)^T] = \mathbf{H}\mathrm{P}_a^{-1}\mathbf{H}^T$$

# Optimality System (O.S. Le Dimet 90s) and SOI

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + (\mathbf{y}-H(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))$$

Dynamic context

Control theory for high-dimensional system

$$\begin{cases} \frac{dX}{dt} = F(X) + B.V \\ X(0) = U \\ \frac{dP}{dt} + \left[\frac{\partial F}{\partial X}\right]^t P = C^t(CX - X_{obs}) \\ P(T) = 0 \\ \nabla_U J = -P(0) + (U - U_0) = 0 \\ \nabla_V J = -B^t.P = 0 \end{cases}$$

Backward propagation

obs impulse

Adjoint variable: sensitivity to obs impulse

O.S. as a general model

All information contained in O.S.

Optimization based on O.S.

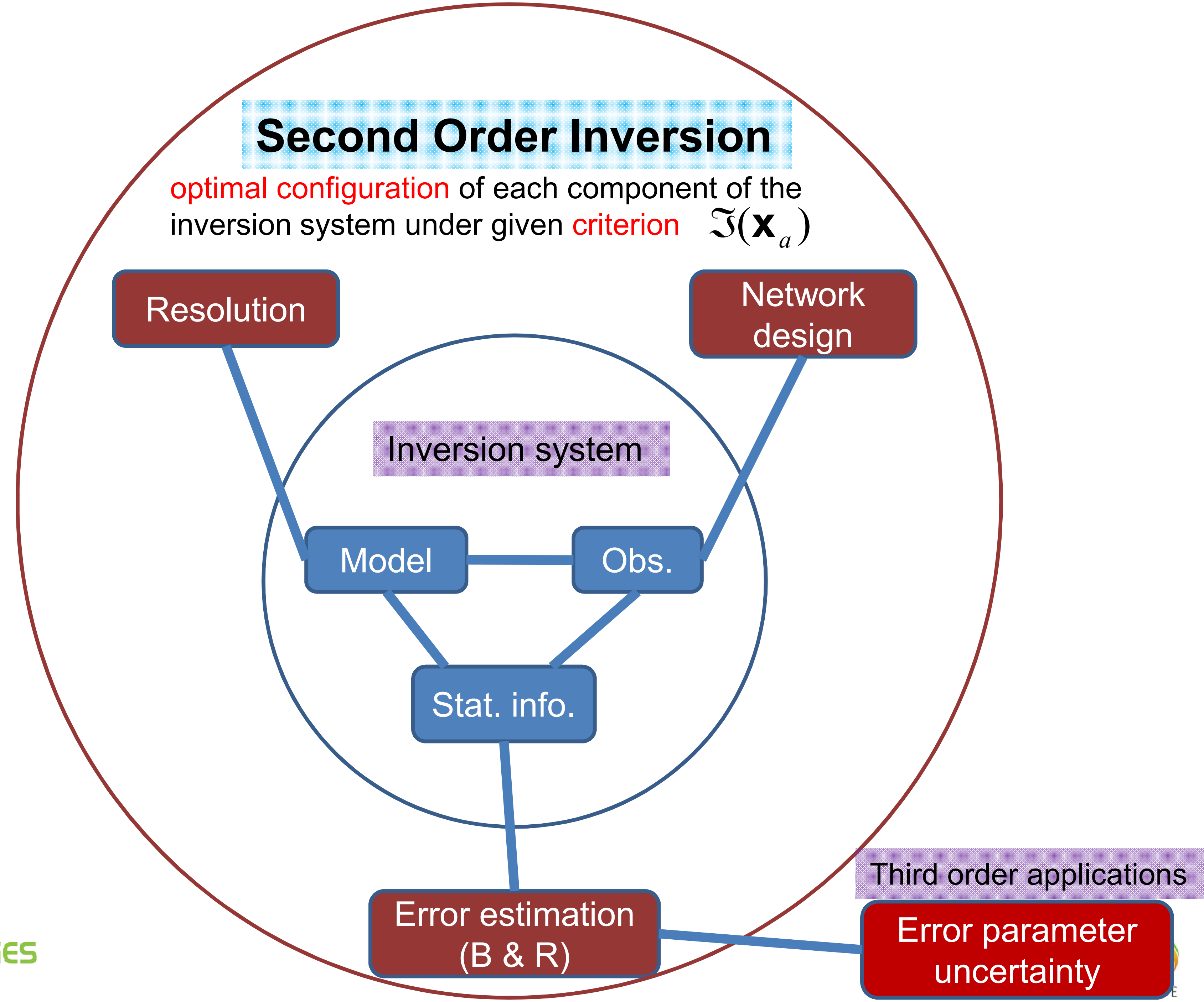
**Second order inversion (SOI)**

$$\Im(\mathbf{x}_a)$$

$\Im$ Performance of inversion system; not necessarily RMSE

$\mathbf{x}_a$ Solution given by O.S.

| | |
|---|---|
| Direct modeling | Cost 1 |
| Inversion/assim. | Cost 10 |
| SOI | Cost 100 |


Second Order Inversion
optimal configuration of each component of the inversion system under given criterion $\Im(\mathbf{x}_a)$
Resolution
Network design
Inversion system
Model
Obs.
Stat. info.
Error estimation (B & R)
Third order applications
Error parameter uncertainty

# Bayesian inversion: vectorial case of linear dynamics and Gaussian error

- Context: Inverse modelling of sources σ (2D+T); Gaussian assumption + linear observation operator.

  - H Jacobian matrix of the problem (observation + model):

$$\mu = \mathbf{H}\sigma + \varepsilon$$

  - $\sigma^b - \sigma \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$; $\sigma^b$ prior fluxes, B background error covariance matrix.
  - $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$; $\mathbf{R}$ observation error covariance matrix.

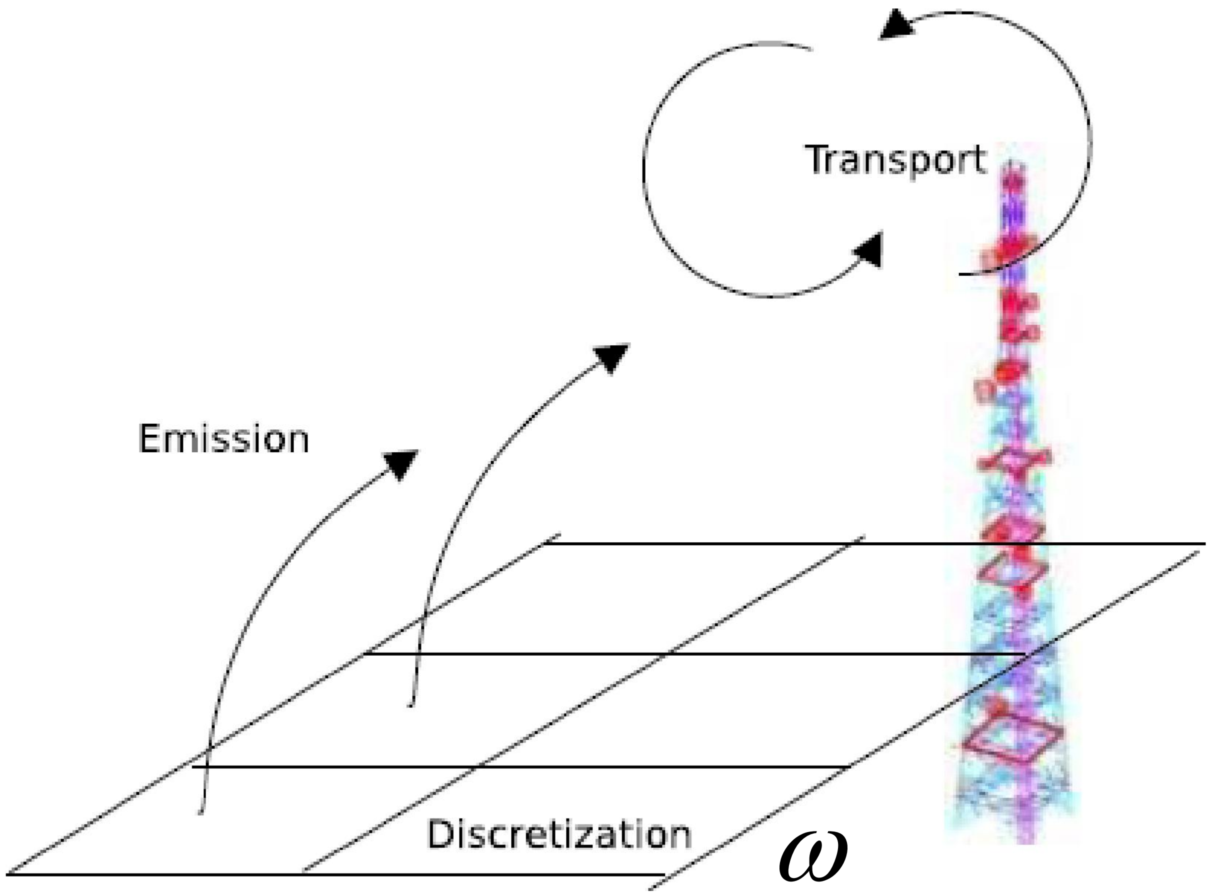
- BLUE analysis:

$$\sigma^a = \sigma^b + \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\left(\mu - \mathbf{H}\sigma^b\right),$$

$$\mathbf{P}^a = \mathbf{B} - \mathbf{B}\mathbf{H}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}\right)^{-1}\mathbf{H}\mathbf{B}.$$

- A representation ω is a discretization of the space-time domain of control (parameter) space Ω.

# Bayesian inversion: vectorial case of linear dynamics and Gaussian error




Decomposition of observation error: $\varepsilon_{\omega} = \varepsilon + \varepsilon_{\omega}^{c} + \varepsilon_{\omega}^{m}$

# $CO_2$ flux inversion

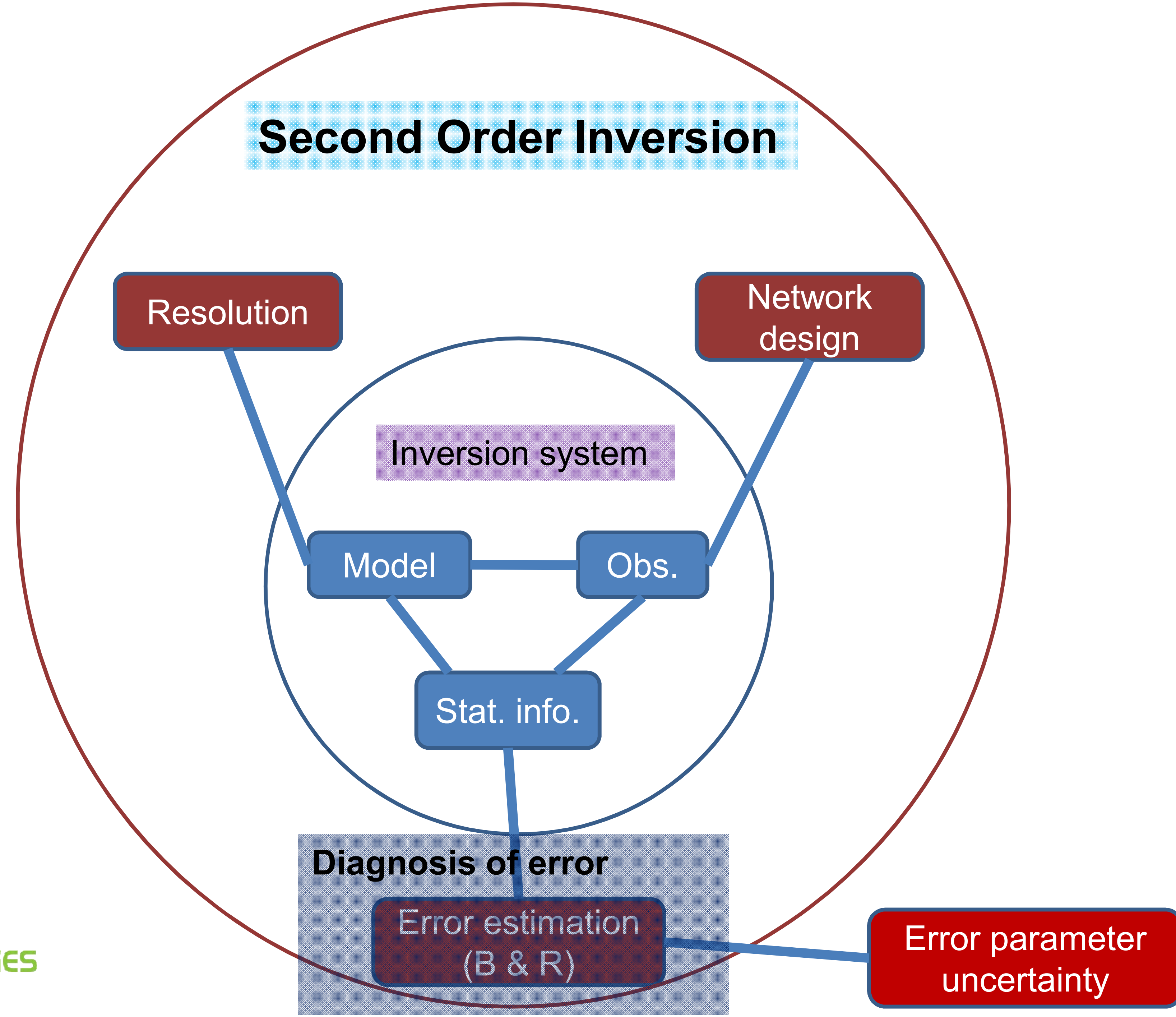
- $CO_2$ Inversion: Using concentration observations to retrieve surface $CO_2$ fluxes.

- Ill-posed problem due to the flux-observation mismatch (e.g. diffusive atmospheric transport that links fluxes with observations)

  - Aggregation of flux variables, e.g. eco-regions or coarser regular grid => aggregation error
  - Bayesian inversion: regularized by prior information (correlation in prior flux errors)

- Plan

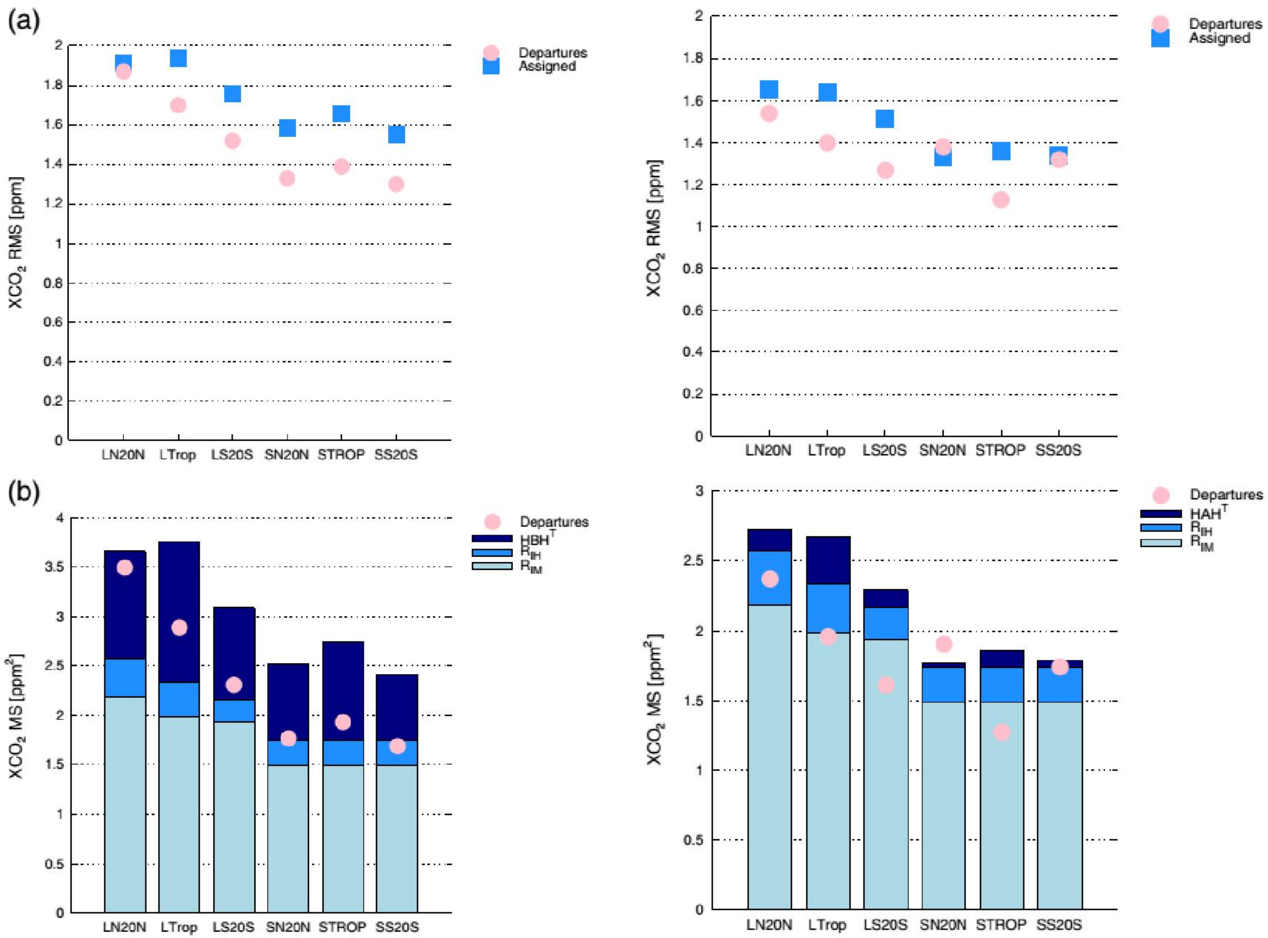
  - Error diagnosis
  - Aggregation error: multiscale inversion (resolution optimization) & direct aggregation.
  - Estimates parameters of the prior and observation errors (hyper-parameter estimation)


Second Order Inversion
Resolution
Network design
Inversion system
Model
Obs.
Stat. info.
Diagnosis of error
Error estimation (B & R)
Error parameter uncertainty

# Diagnosis of error

Chevallier & O'Dell 2013

- Variational inversion, Monte Carlo simulations for error statistics
- Compare  with GOSAT data

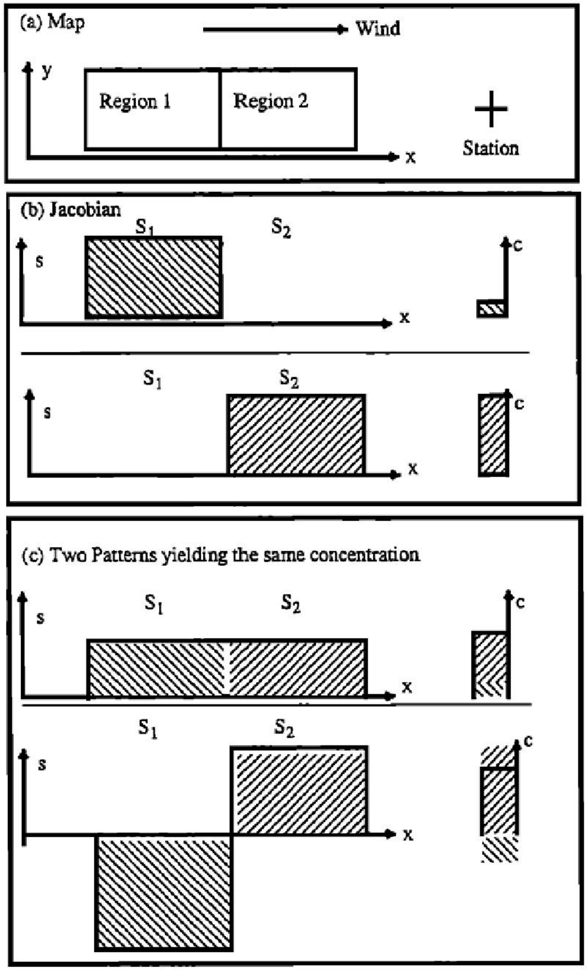



$$E\left[(\mathbf{Hx}_b - \mathbf{y})(\mathbf{Hx}_b - \mathbf{y})^T\right] = \mathbf{HBH}^T + \mathbf{R}$$

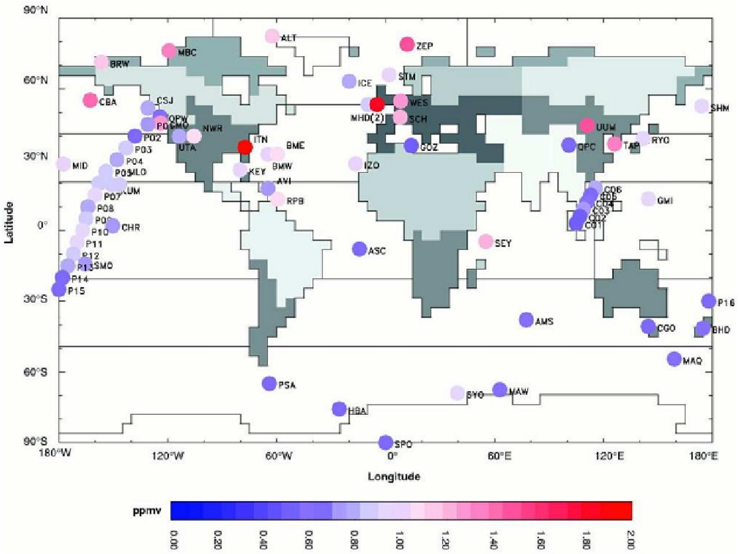
$$E\left[(\mathbf{Hx}_a - \mathbf{y})(\mathbf{Hx}_a - \mathbf{y})^T\right] = \mathbf{H}\,\mathbf{P}_a^{-1}\,\mathbf{H}^T + \mathbf{R}$$

# Aggregation error

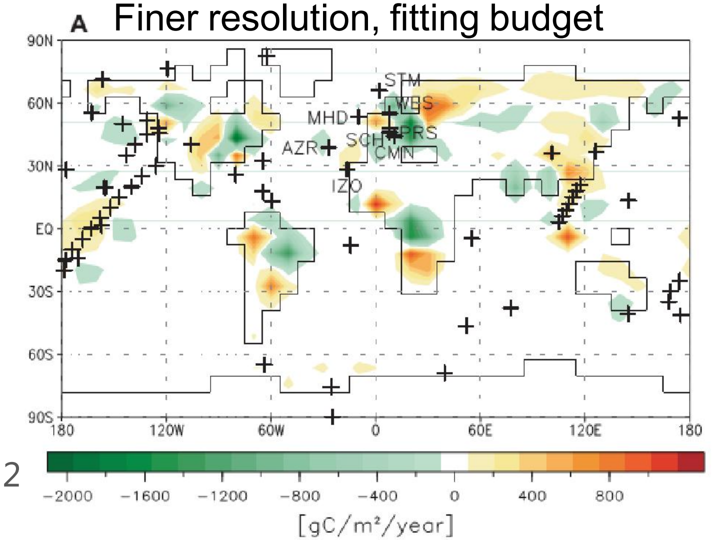
Kaminski et al 2011
Missing small scale details
(high frequency )

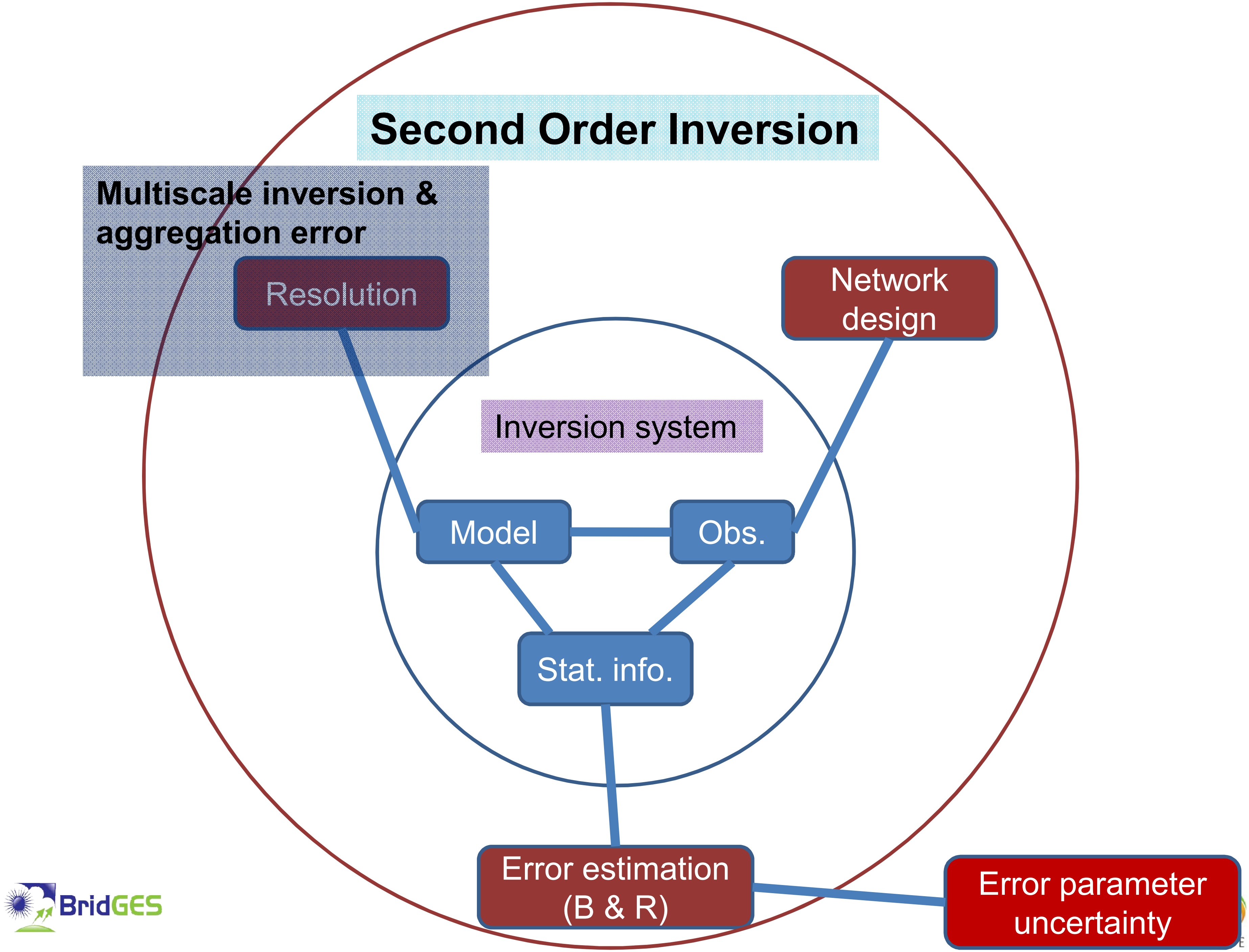



Bousquet et al 2000

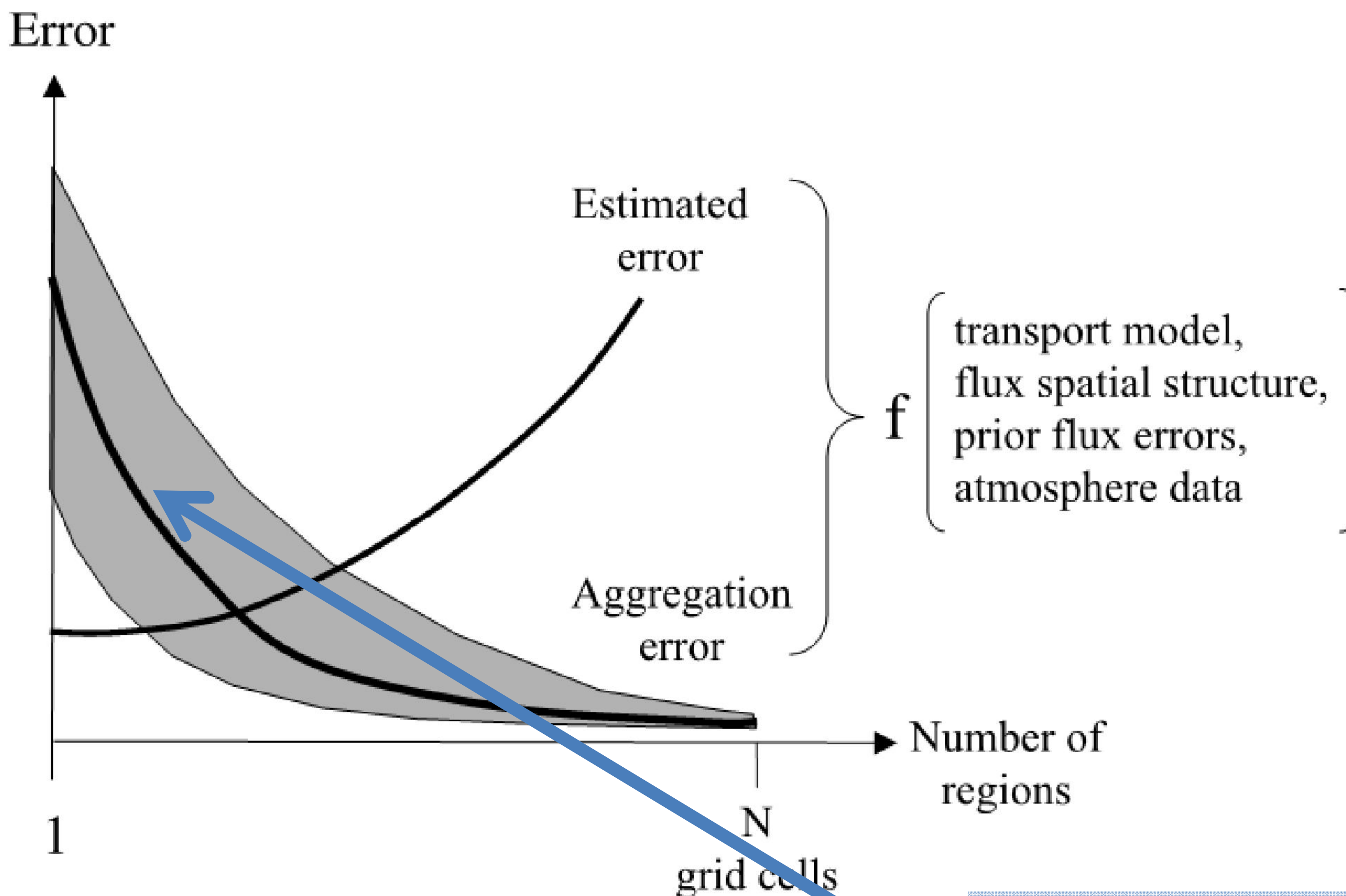



Kaminski & Heimann 2001
Finer resolution, fitting budget




Second Order Inversion
Multiscale inversion &
aggregation error
Resolution
Network
design
Inversion system
Model
Obs.
Stat. info.
Error estimation
(B & R)
Error parameter
uncertainty
BridGES

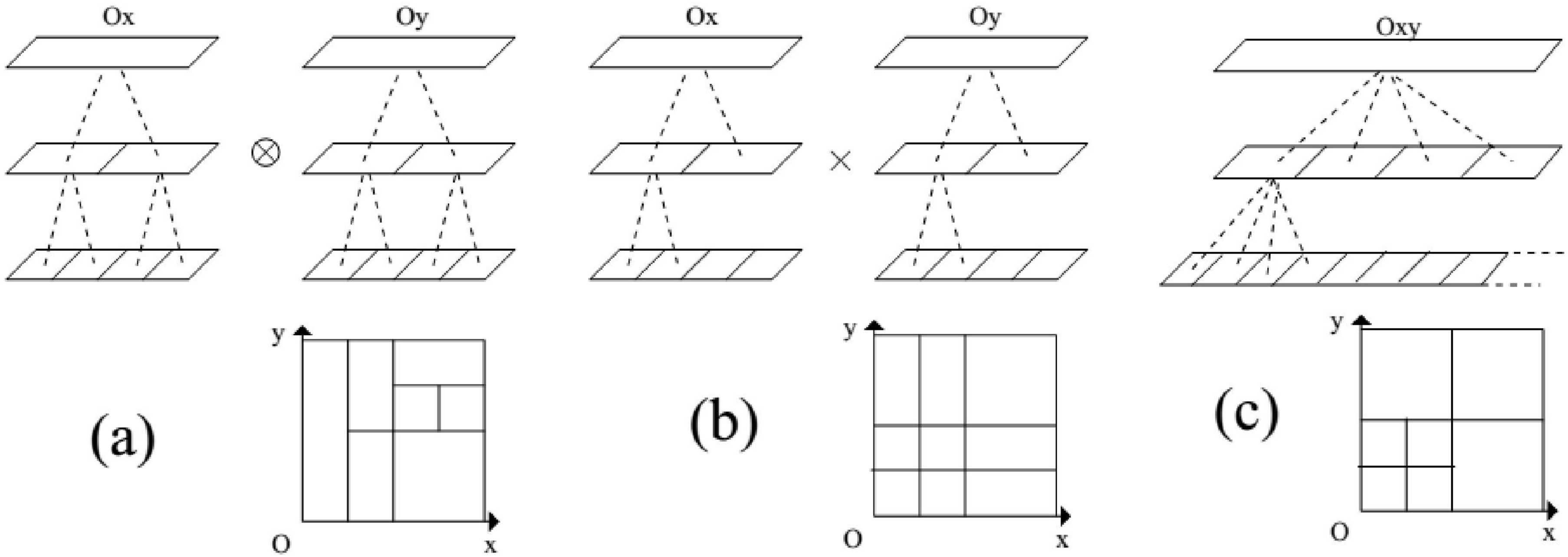
# Aggregation error




Peylin et al 2011

Ideal case:
- No model error
- Finest resolution offordable

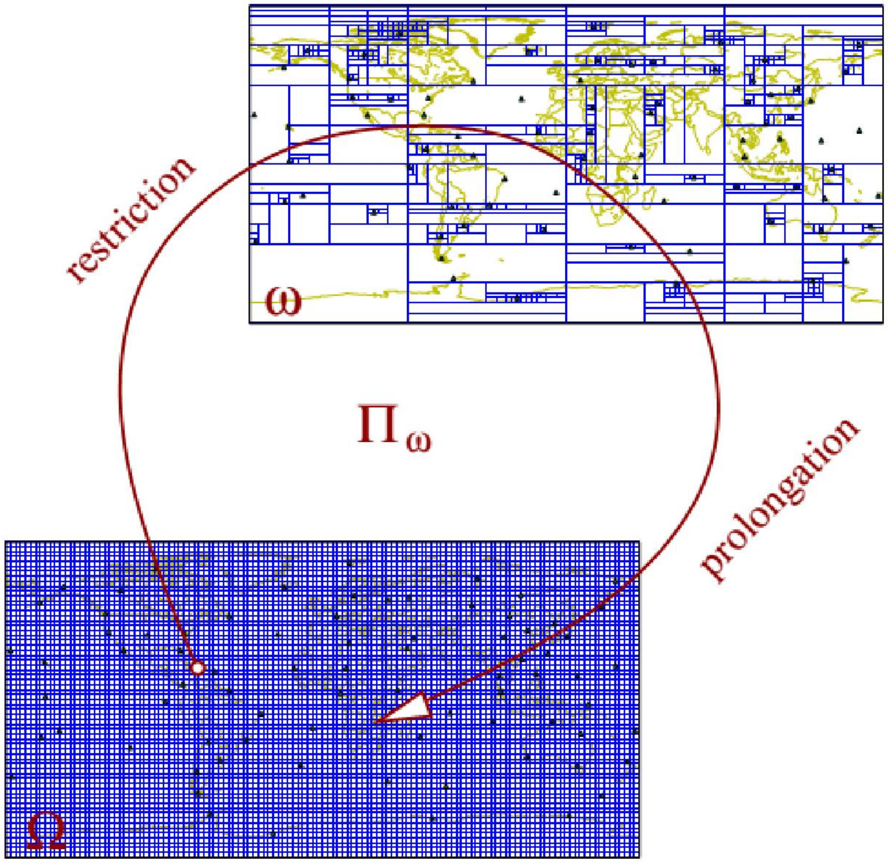
# Multiscale structure




- Memory costs for a 2D+T control space
  - Tilings: up to 8 times the size of the finest grid Jacobian
  - Qtrees: up to 8/3 times the size of the finest grid Jacobian.
- Empirically, optimisation on the qtrees is twice faster than on the tilings.

# Multiscale inversion

- The source variables (vector $\sigma$) can be discretised on an adaptive grid $\omega$.
- Restriction ($\mathbf{\Gamma}_\omega$) and prolongation ($\mathbf{\Gamma}^*_\omega$) operators can help to transfer $\sigma$ from the finest regular grid cell $\Omega$ to $\omega$.
- The composition of a restriction and a prolongation gives a projection operator $\mathbf{\Pi}_\omega$ which depends on the geometry of $\omega$.



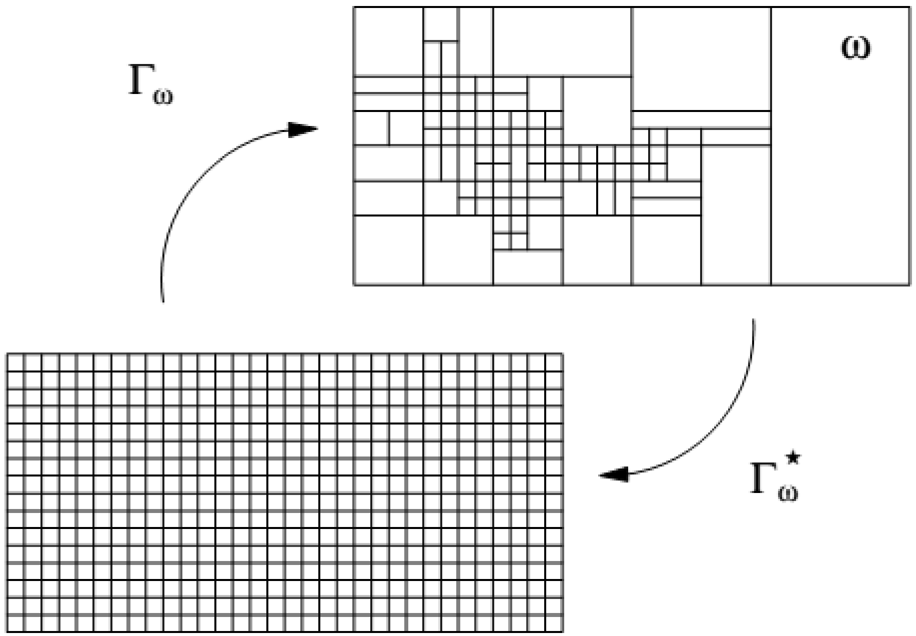
# Up and down the scale ladder (1/4)

## Restriction and prolongation

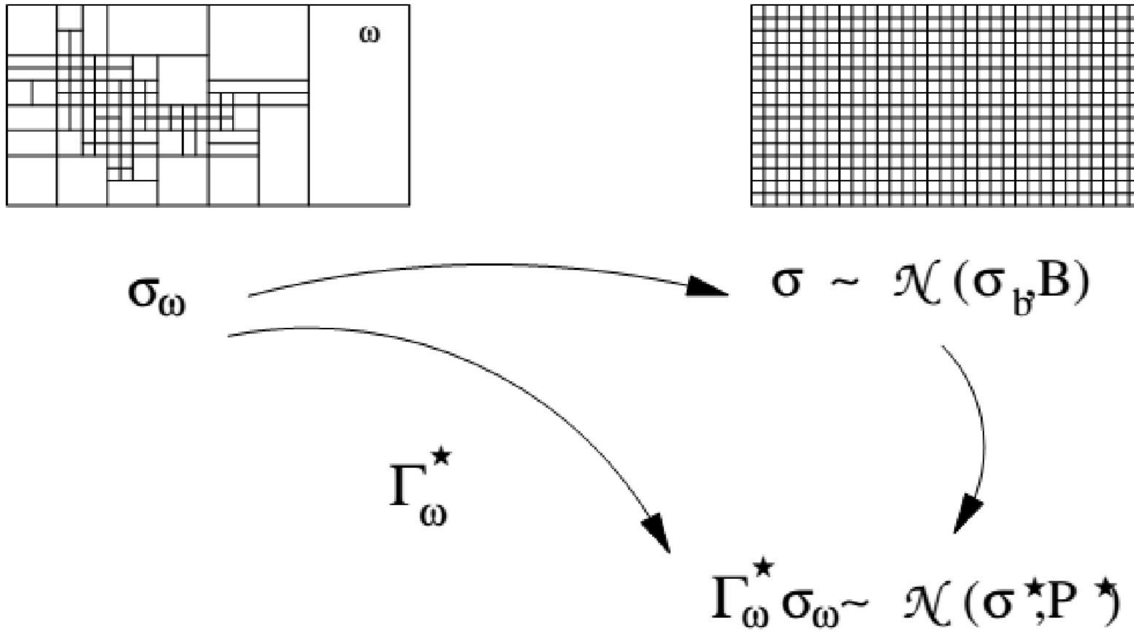
- Restriction operator : $\sigma \underset{\text{coarse graining}}{\longrightarrow} \sigma_\omega = \mathbf{\Gamma}_\omega \sigma$, where $\mathbf{\Gamma}_\omega : \mathbb{R}^{N_{\text{fg}}} \to \mathbb{R}^N$ defines the coarse graining operator (non-ambiguous).
- Prolongation operator : $\mathbf{\Gamma}^\star_\omega : \mathbb{R}^N \to \mathbb{R}^{N_{\text{fg}}}$ refines $\sigma_\omega$ into $\sigma$ (ambiguous).

## Scaling of errors

- Background error covariance matrix: $\mathbf{B}_\omega = \mathbf{\Gamma}_\omega \mathbf{B} \mathbf{\Gamma}^{\mathrm{T}}_\omega$,
- Observations/representativity/model errors: $\mathbf{R}_\omega$, to be discussed later.



# Up and down the scale ladder (2/4)




## Bayesian choice of a prolongation operator

- Idea: Use prior $\sigma \sim \mathcal{N}(\sigma_b, \mathbf{B})$ to refine the source. Knowing $\sigma_\omega$ in representation $\omega$, then from Bayes' rule, the most likely refined source is given by the mode of

$$q(\sigma|\sigma_\omega) = \frac{q(\sigma)}{q_\omega(\sigma_\omega)}\delta(\sigma_\omega - \mathbf{\Gamma}_\omega \sigma)$$

# Up and down the scale ladder (3/4)

## Bayesian choice of a prolongation operator

- Refinement is now a statistical process ! But the prolongation operator will be defined as the most likely refinement operation.
- Thus the (estimate of the) refined source is

$$\sigma^\star = \sigma_b + \mathbf{B}\mathbf{\Gamma}_\omega^{\mathrm{T}}\left(\mathbf{\Gamma}_\omega\mathbf{B}\mathbf{\Gamma}_\omega^{\mathrm{T}}\right)^{-1}(\sigma_\omega - \mathbf{\Gamma}_\omega\sigma_b)$$

which suggests the (affine) prolongation operator

$$\mathbf{\Gamma}_\omega^\star \equiv (\mathbf{I}_{N_{\mathrm{fg}}} - \mathbf{\Pi}_\omega)\sigma_b + \mathbf{\Lambda}_\omega^\star \,,$$

where the linear part of $\mathbf{\Gamma}_\omega^\star$ is

$$\mathbf{\Lambda}_\omega^\star \equiv \mathbf{B}\mathbf{\Gamma}_\omega^{\mathrm{T}}\left(\mathbf{\Gamma}_\omega\mathbf{B}\mathbf{\Gamma}_\omega^{\mathrm{T}}\right)^{-1}, \quad \text{and} \quad \mathbf{\Pi}_\omega \equiv \mathbf{\Lambda}_\omega^\star\mathbf{\Gamma}_\omega \,.$$

# Up and down the scale ladder (4/4)

## Up and down

- Must consistently satisfy $\mathbf{\Gamma}_\omega \mathbf{\Gamma}^\star_\omega = \mathbf{I}_N$.
- Down and up: $\mathbf{\Gamma}^\star_\omega \mathbf{\Gamma}_\omega = (\mathbf{I}_{N_{\mathrm{fg}}} - \mathbf{\Pi}_\omega)\sigma_b + \mathbf{\Pi}_\omega$

## Properties of $\mathbf{\Pi}_\omega$

- $\mathbf{\Pi}_\omega$ is a projector since $\mathbf{\Pi}^2_\omega = \mathbf{\Pi}_\omega$.
- It is also $\mathbf{B}^{-1}$-symmetric: $\mathbf{\Pi}_\omega \mathbf{B} = \mathbf{B}\mathbf{\Pi}^{\mathrm{T}}_\omega$.

## Observation equation in representation $\omega$

- Then $\mathbf{H}$ becomes $H_\omega = \mathbf{H}\mathbf{\Gamma}^\star_\omega$, and

$$\mu = H_\omega \sigma_\omega + \varepsilon_\omega = \mathbf{H}\mathbf{\Gamma}^\star_\omega \mathbf{\Gamma}_\omega \sigma + \varepsilon_\omega \,,$$

so that

$$\mu = \mathbf{H}\sigma_b + \mathbf{H}\mathbf{\Pi}_\omega(\sigma - \sigma_b) + \varepsilon_\omega \,.$$

# Accounting for aggregation errors

- Consistent observation equations:

$$\mu = \mathbf{H}\sigma + \varepsilon = \mathbf{H}_\omega \sigma_\omega + \varepsilon_\omega \,.$$

- Assuming aggregation is the only source of scale-dependent errors, one has $\mathbf{H}\sigma + \varepsilon = \mathbf{H}\sigma_b + \mathbf{H}\mathbf{\Pi}_\omega(\sigma - \sigma_b) + \varepsilon_\omega$, leading to the identification

$$\varepsilon_\omega = \varepsilon + \mathbf{H}\left(\mathbf{I}_{N_{\mathrm{fg}}} - \mathbf{\Pi}_\omega\right)(\sigma - \sigma_b) = \varepsilon + \varepsilon^c_\omega \,.$$

- Assuming independence of the error and source priors, the computation of the covariance matrix of these errors leads to

$$\mathbf{R}_\omega = \mathbf{R} + \mathbf{H}\left(\mathbf{I}_{N_{\mathrm{fg}}} - \mathbf{\Pi}_\omega\right)\mathbf{B}\mathbf{H}^{\mathrm{T}} \,.$$

- In that case, one checks that the innovation statistics $\mathbf{D} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}$ are scale-independent $(\mathbf{R} + \mathbf{H}_\omega \mathbf{B}_\omega \mathbf{H}_\omega^{\mathrm{T}} \longrightarrow \mathbf{R}_\omega + \mathbf{H}_\omega \mathbf{B}_\omega \mathbf{H}_\omega^{\mathrm{T}} = \underline{\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}}})$.

# Optimal representation mitigates aggregation effect

- $\omega$ maximizes DFS (bocquet et al., 2011): normalized uncertainty reduction $(\mathbf{B}-\mathbf{P}^a)\mathbf{B}^{-1} = \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{H}$

$$\mathrm{DFS}_\omega = \mathrm{Tr}(\mathbf{\Pi}_\omega \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{H}).$$

Optimal information propagation from observation sites to the whole domain

- The aggregation effect can be quantified by:

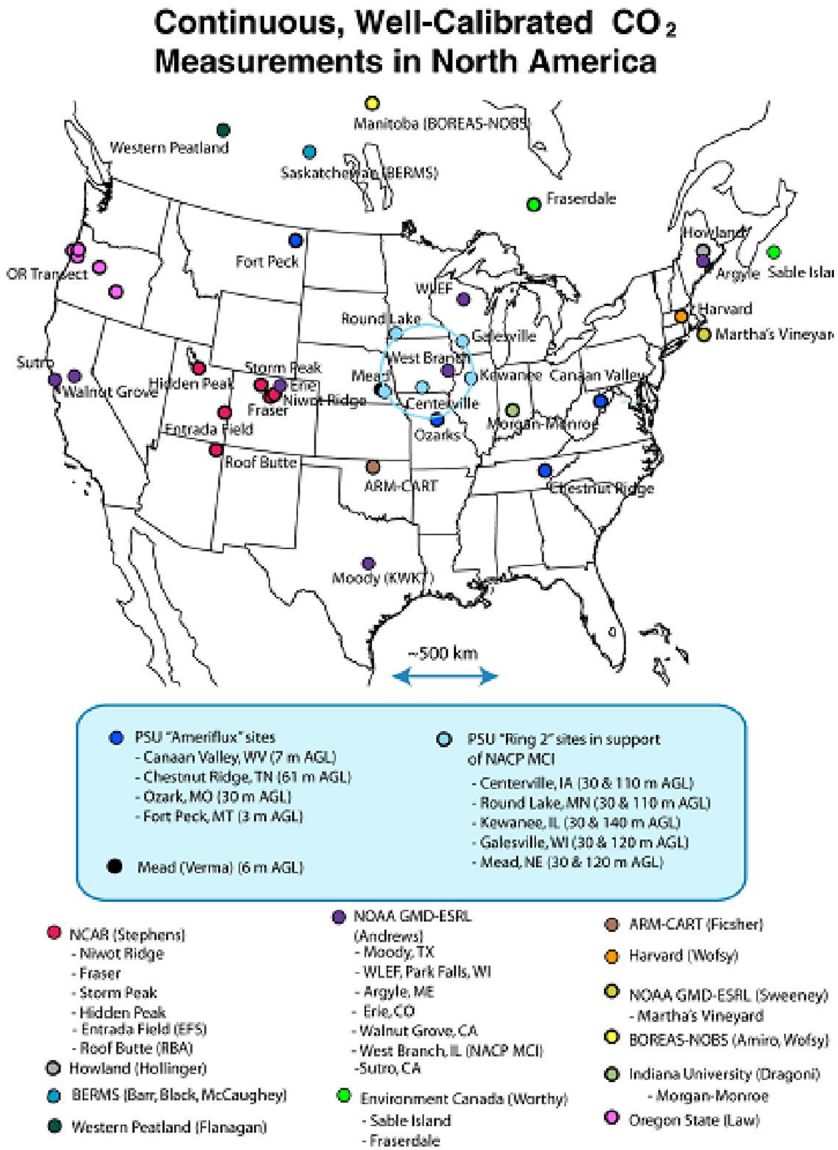
$$\widehat{\mathscr{I}_\omega} = \mathrm{Tr}\left[\mathbf{R}^{-1}(\mathbf{R}_\omega - \mathbf{R})\right] = \mathrm{Tr}\left(\mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right) - \mathrm{Tr}\left(\mathbf{\Pi}_\omega \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right).$$

To minimize the aggregation effect is equivalent to the maximization of the Fisher criterion (Wu et al., 2011):

$$\mathrm{Tr}(\mathbf{\Pi}_\omega \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}).$$

which is the limiting case of the DFS criterion when **R** is inflated or when **B** vanishes.

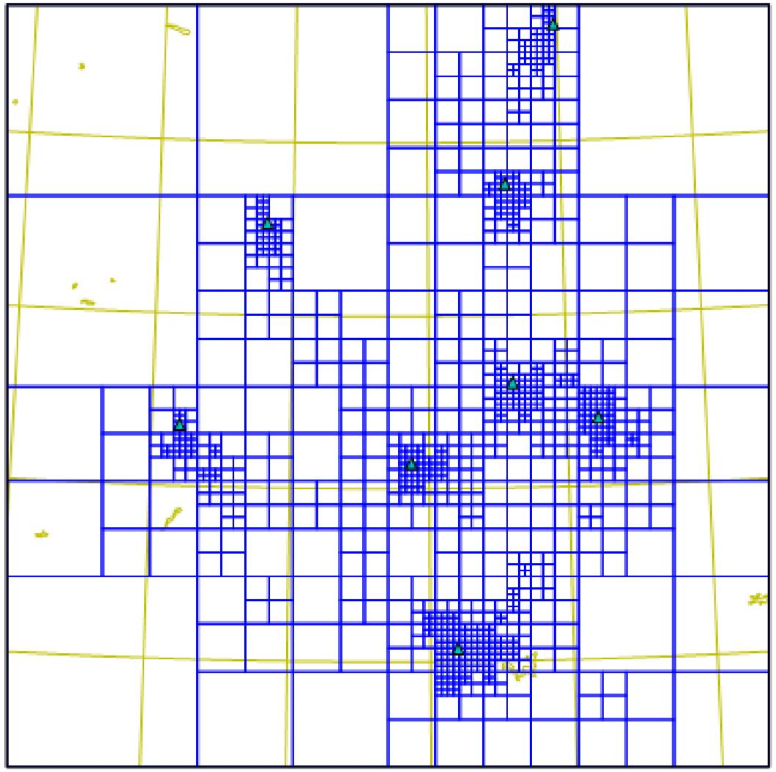
# Inversion system: Experimental setup

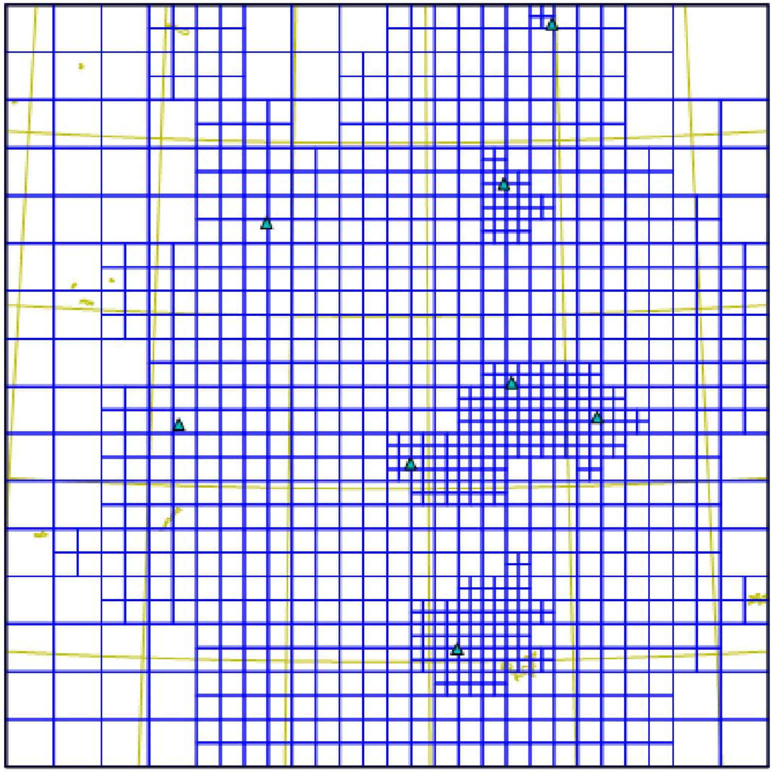



## Setup

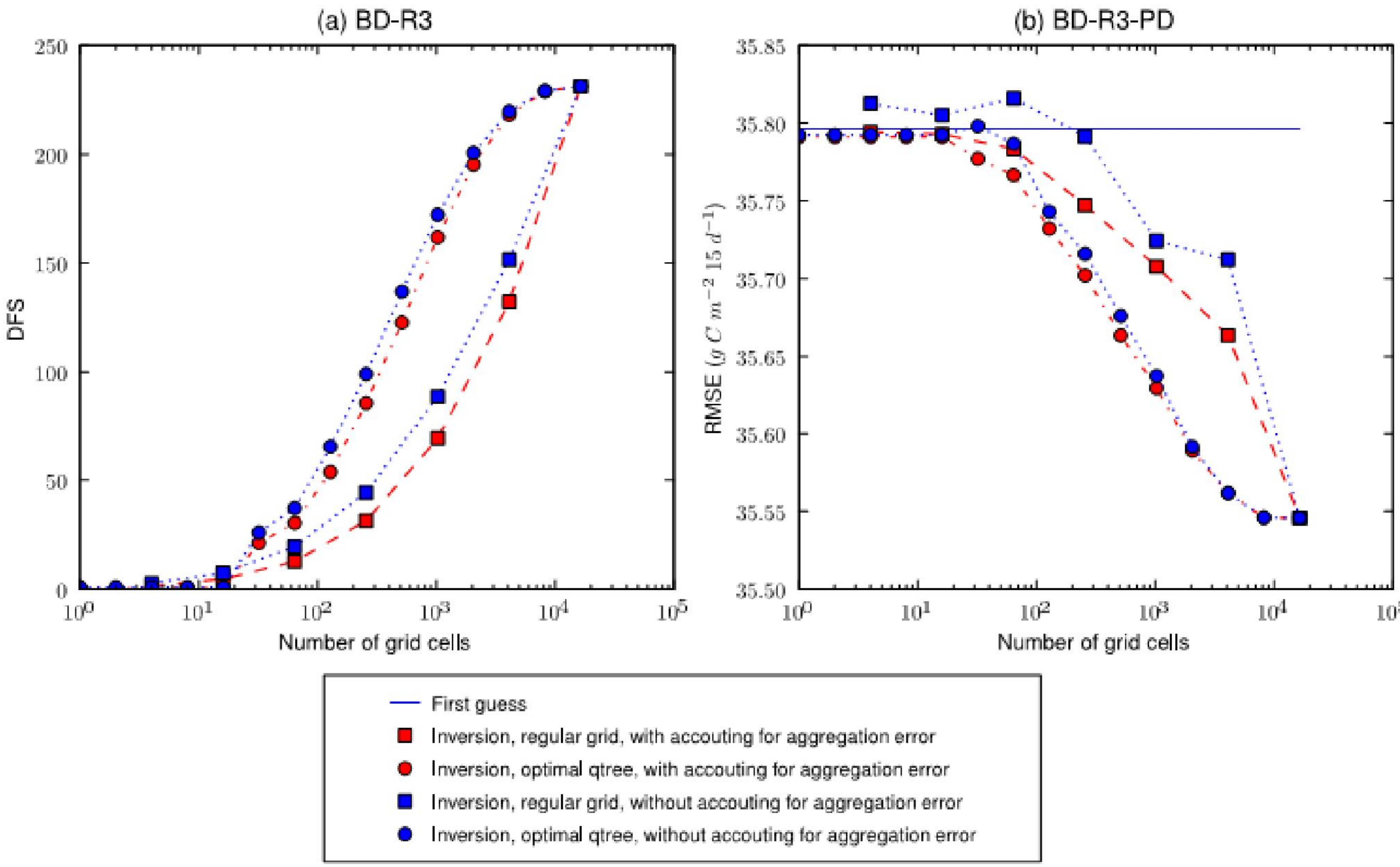
- **Domain**: 980km$\times$980km with 20km$\times$20km grid cell
- **Period**: 01$\sim$15 June 2007 or weekly inversions
- $\mu$: hourly synthetic or real observations from 8 towers
- $\sigma^b$: SiBcrop fluxes
- **H**: Computed from particles generated by Lagrangian model LPDM
- **R**: Diagonal or temporal correlations
- **B**: Diagonal or Balgovind parameterization

# Optimal representations with different settings

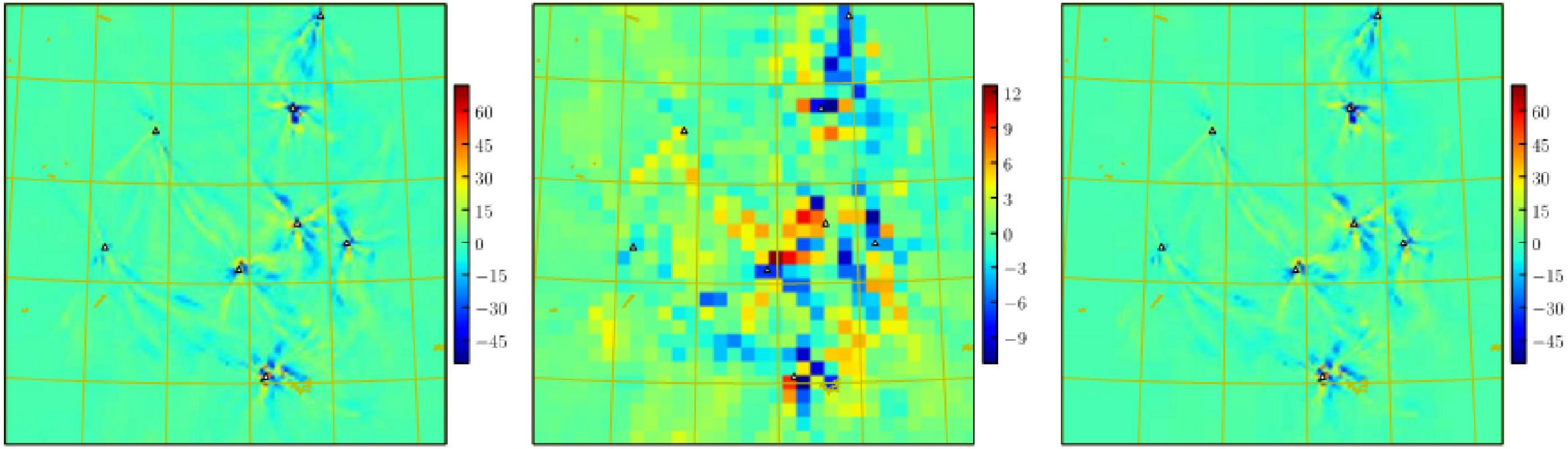

(a) BD-R3-N1024


(b) B50-R3-N1024

# Inversion on regular and optimal representations: diagonal B

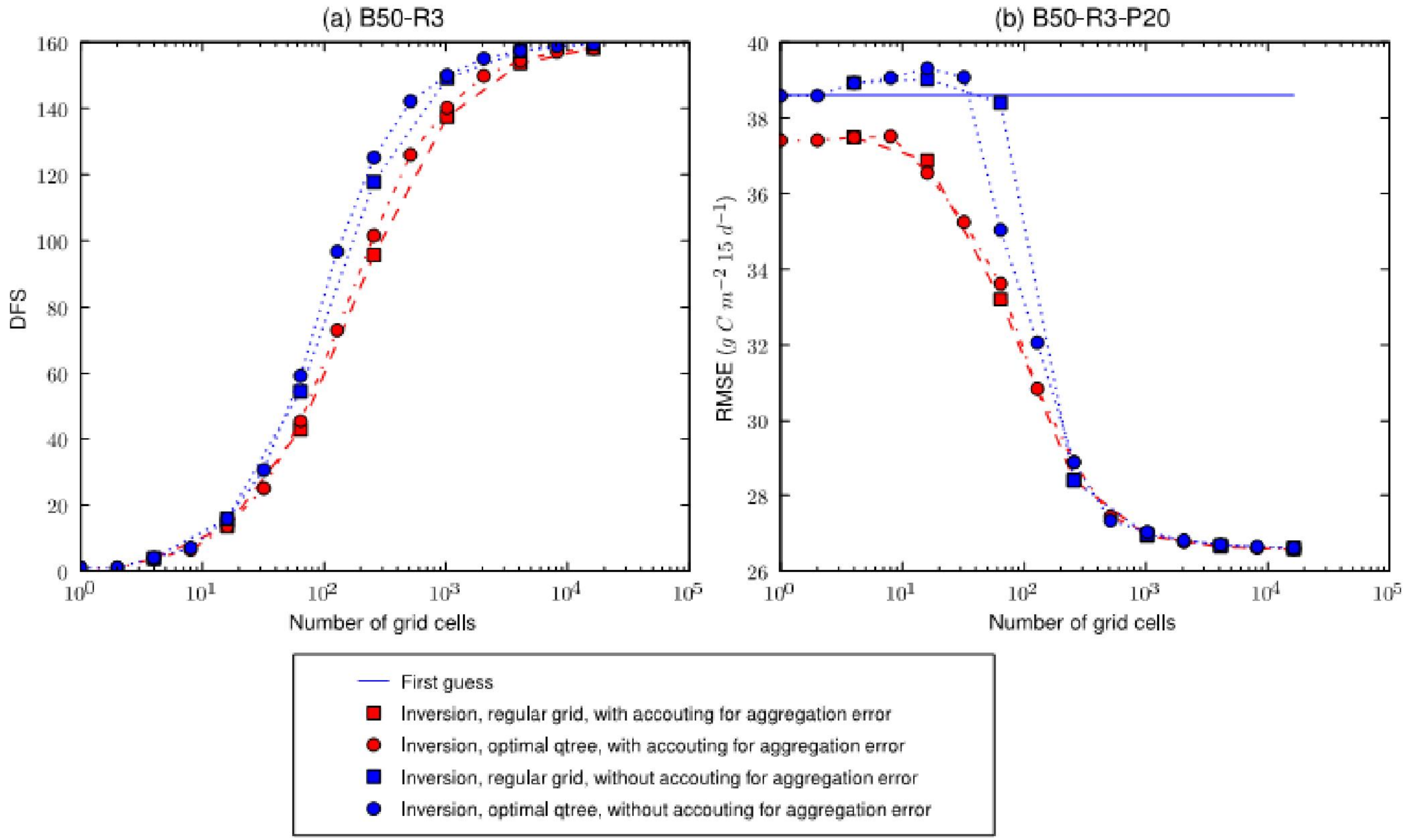


# Performance of optimal grid for diagonal B

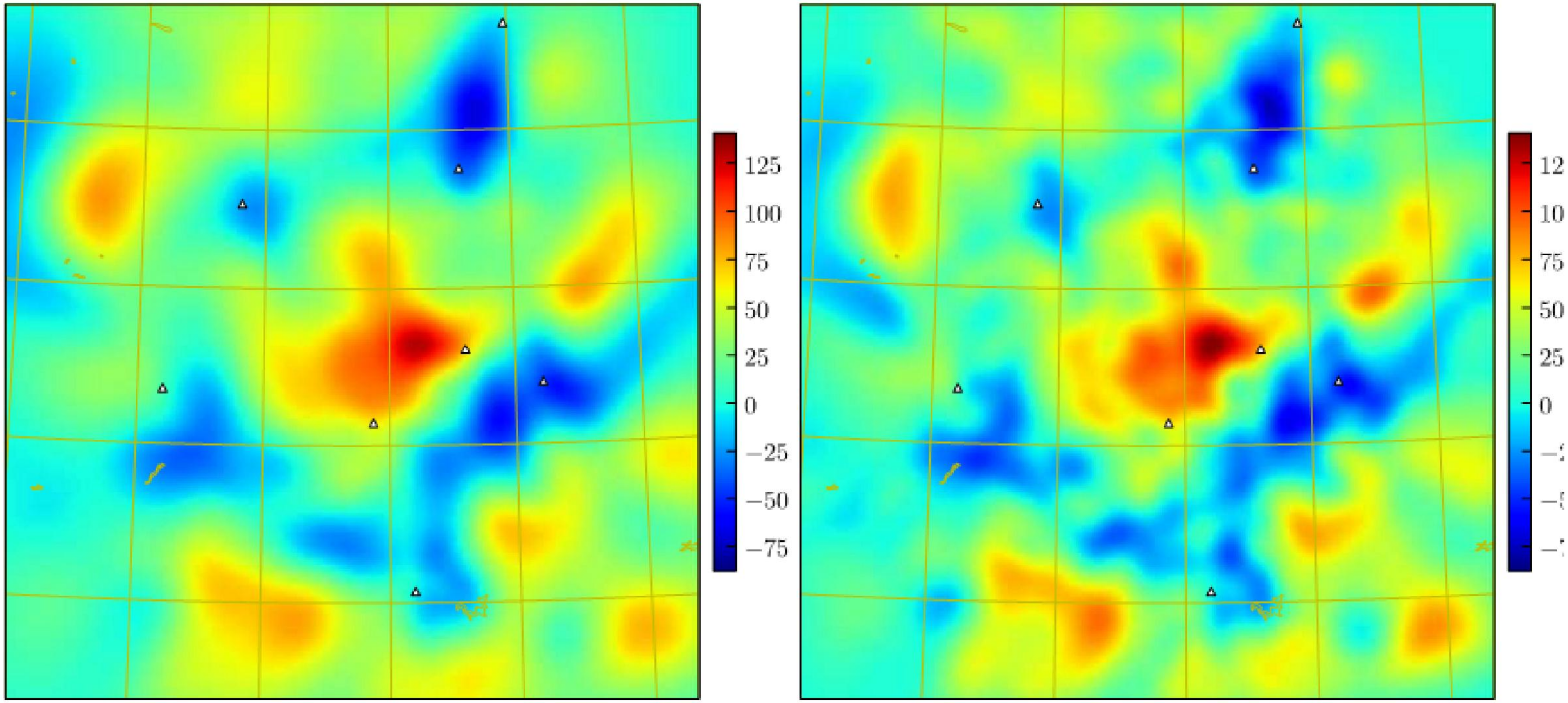



(a) BD-R3-PD, CORR (finest grid) (b) BD-R3-PD, CORR (regular grid) (c) BD-R3-PD, CORR (optimal grid)

# Inversion on regular and optimal representations: correlated B



# Irrealistic correlation length for Balgovind B

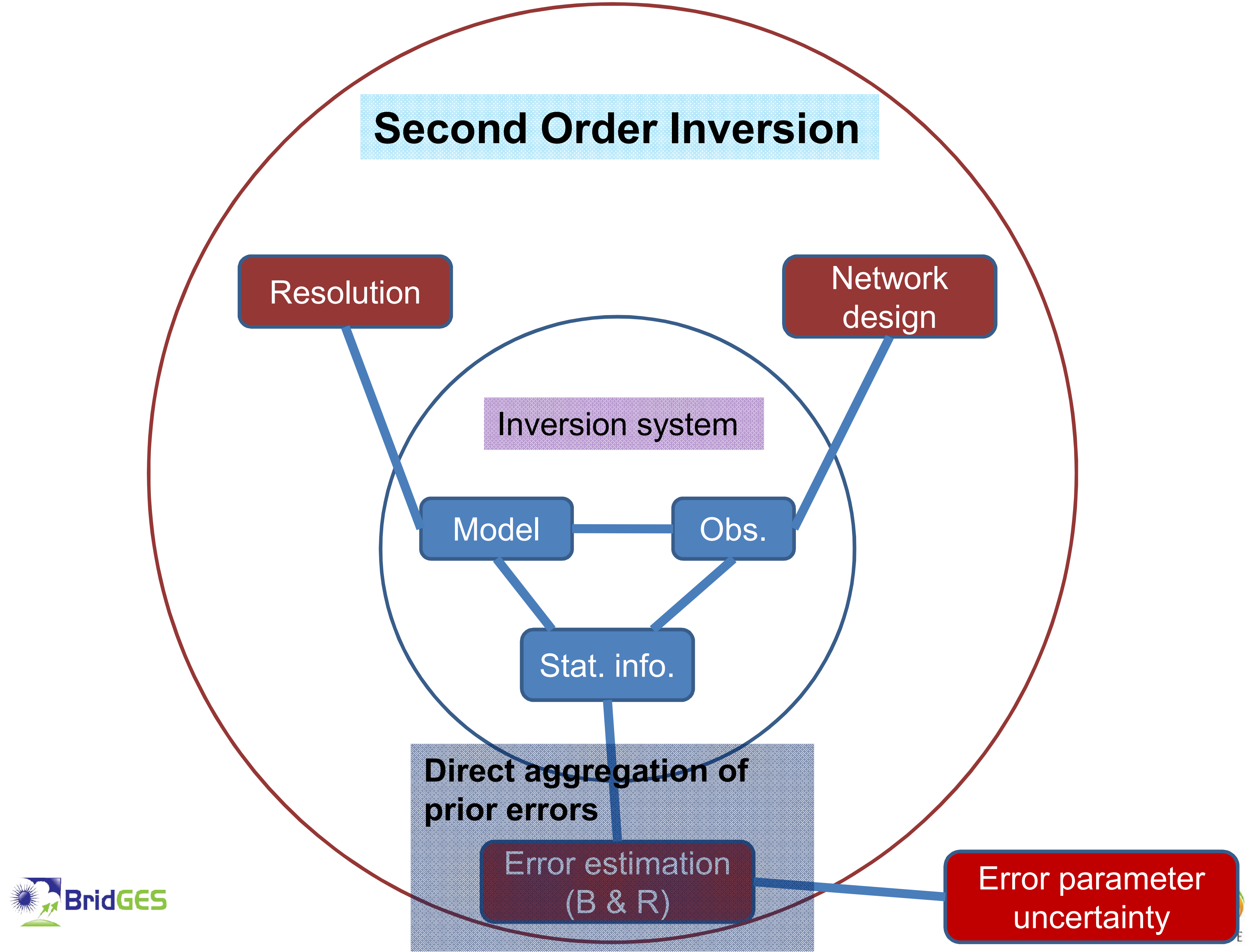



(a) B50-R3-P50, CORR (finest grid)

(b) B20-R3-P50, CORR (finest grid)

## Summary on mutiscale inversion & aggregation error

- A typical second order inversion problem
  - Critereia: e.g. DFS
  - Model configuration: resolution
- An ideal case: model-error-free + finest resolution available
- Explicit aggregation error + information flow map
- Maximizing Fisher criteria = minimizing aggregation error
- Future directions: model error and trade-off between aggregation and estimation error



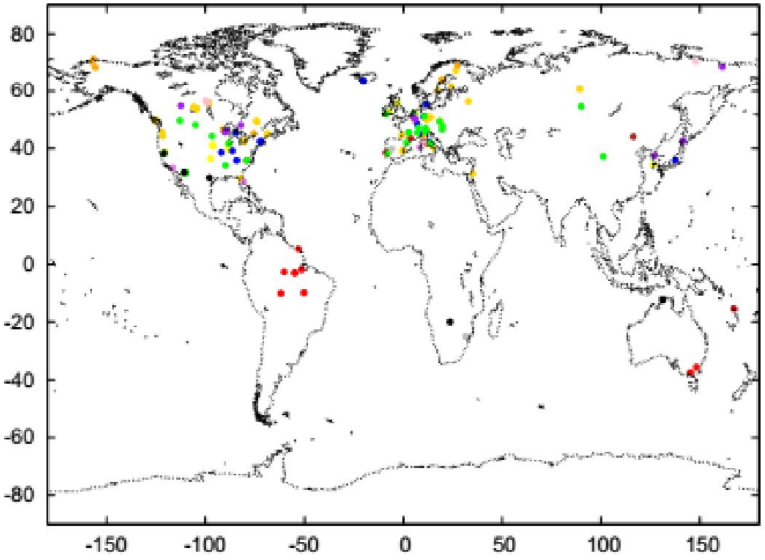
Second Order Inversion
Resolution
Network design
Inversion system
Model
Obs.
Stat. info.
Direct aggregation of prior errors
Error estimation (B & R)
Error parameter uncertainty
BridGES

# Structure of the prior CO2 flux errors

Chevallier et al 2012

Use daily-mean eddy-covariance flux measurements to assign the error statistics of the prior fluxes (Chevallier et al., 2012)

$i$-th day, $1 < i < T_d$, $T_d = 365$
$j$-th year, $1 < j < T_y$, $T_y = 17$
$s$-th site, $1 < s < N$, $N = 156$




Data courtesy from the FLUXNET PIs as part of a La Thuile project. 1991~2007, 156 sites

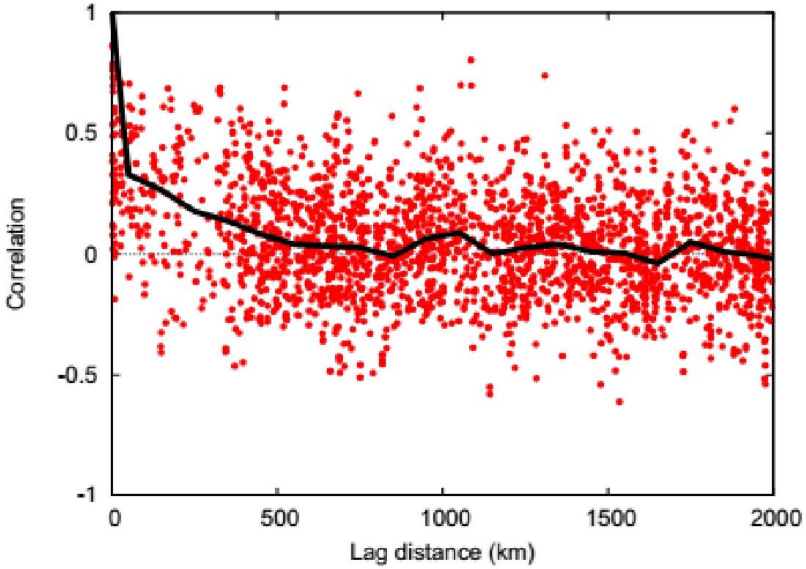
- FLUXNET Observations $[y^s_{i,j}]_{i,j,s}$;
- ORCHIDEE (a process-based terrestrial ecosystem model) simulations $[x^s_{i,j}]_{i,j,s}$.

**Statistics**

- Model-minus-observations
- Observation variability

# Structure of the prior $CO_2$ flux errors

Chevallier et al 2012




Small spatial correlations.

For a given day $i$, for all site pairs, Pearson correlation

$$r_i(s_p, s_q) = \frac{\sum_{j=1}^{T_y} (d_{i,j}^{s_p} - \bar{d}_i^{s_p})(d_{i,j}^{s_q} - \bar{d}_i^{s_q})}{\sqrt{\sum_{j=1}^{T_y} (d_{i,j}^{s_p} - \bar{d}_i^{s_p})^2} \sqrt{\sum_{j=1}^{T_y} (d_{i,j}^{s_q} - \bar{d}_i^{s_q})^2}}$$

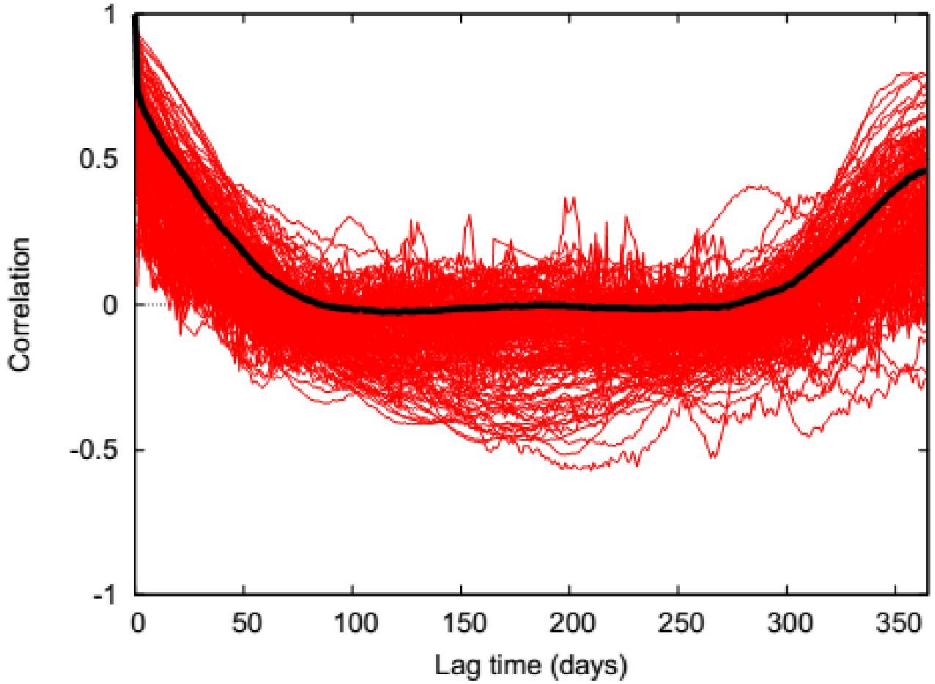
**Model-minus-observations**

$$d_{i,j}^{s} = y_{i,j}^{s} - x_{i,j}^{s}$$

- Short spatial correlation length of few hundred kilometers ($< 0.2$ after 200 km)
- Independent of plant functional types except for deciduous broad-leaved forests

# Structure of the prior CO2 flux errors

Chevallier et al 2012




Large temporal correlations.

For a given site $s$, for date lags in days, Pearson

$$\text{correlation } r_s(t_p, t_q) = \frac{\sum_{j=1}^{T_y} (d^s_{t_p,j} - \bar{d}^s_{t_p})(d^s_{t_q,j} - \bar{d}^s_{t_q})}{\sqrt{\sum_{j=1}^{T_y} (d^s_{t_p,j} - \bar{d}^s_{t_p})^2} \sqrt{\sum_{j=1}^{T_y} (d^s_{t_q,j} - \bar{d}^s_{t_q})^2}}$$

**Model-minus-observations**

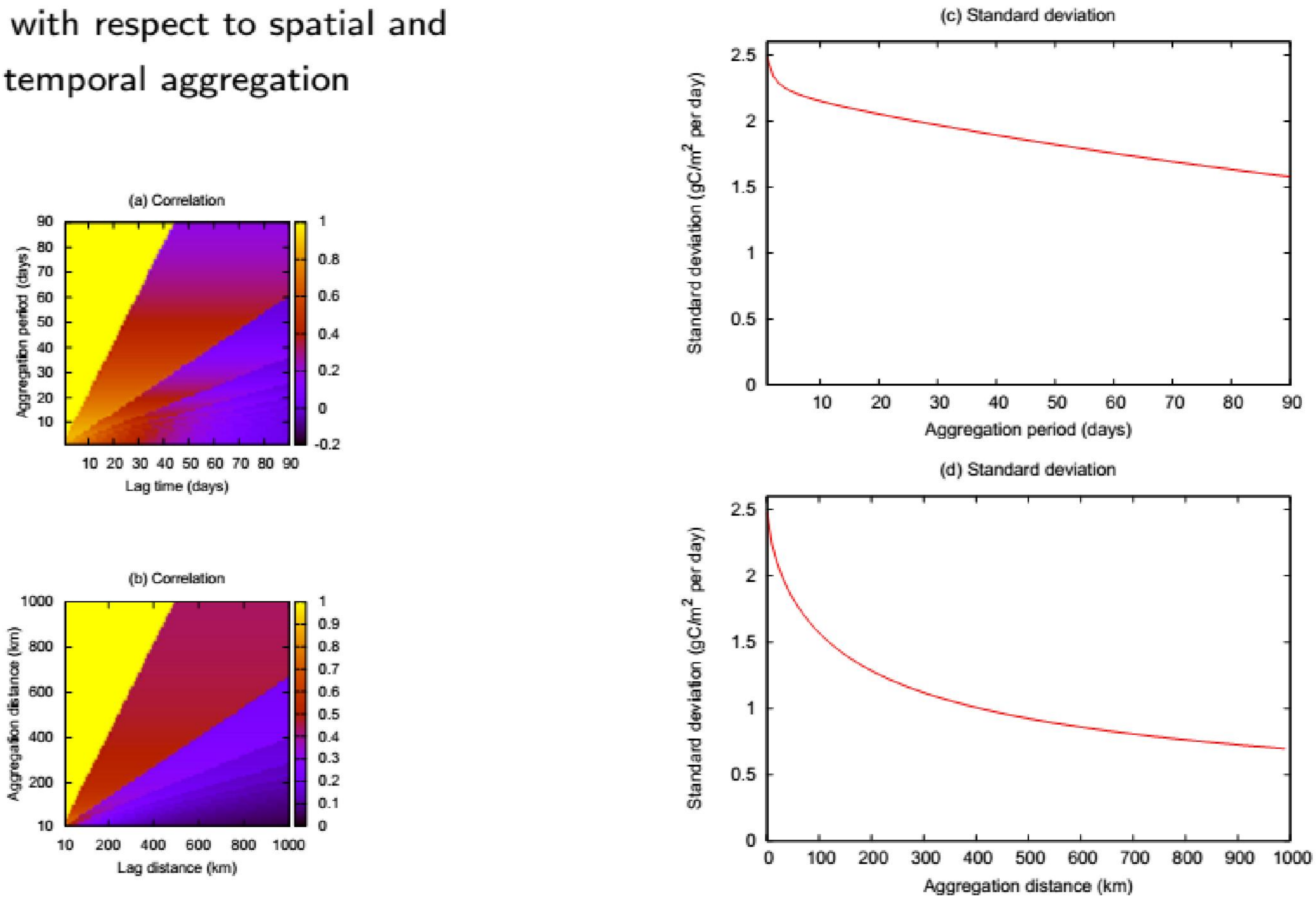
$$d^s_{i,j} = y^s_{i,j} - x^s_{i,j}$$

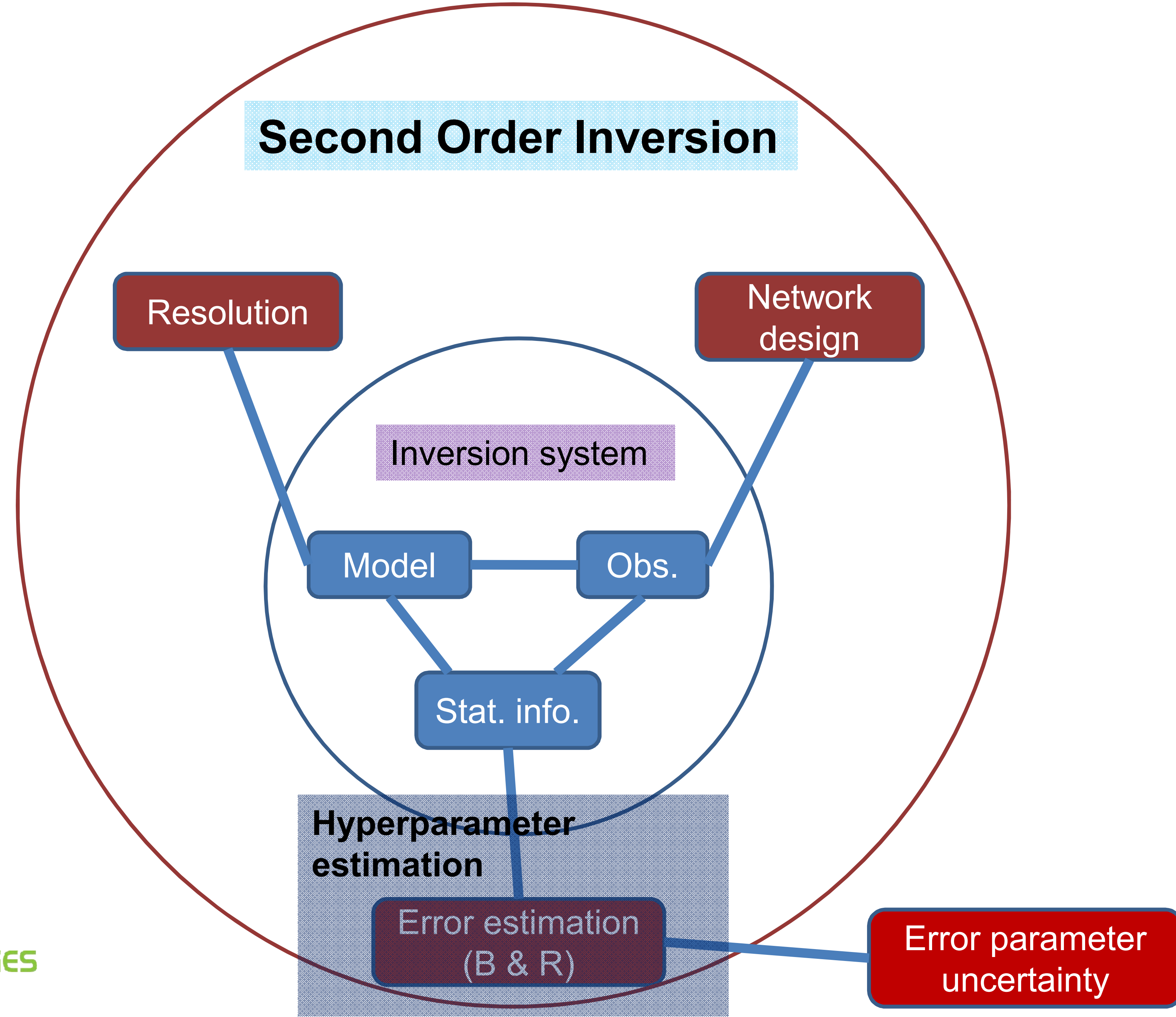
- Large temporal correlation length (positive for lags $< 85$ days and for lags $> 274$ days)
- Reflects systematic errors over weeks

# Direct aggregation of prior errors

Variations of the statistics of the prior errors with respect to spatial and temporal aggregation

Chevallier et al 2012




Second Order Inversion
Resolution
Network design
Inversion system
Model
Obs.
Stat. info.
Hyperparameter estimation
Error estimation (B & R)
Error parameter uncertainty

BridGES

# Uncertainty quantification: hyper-parameter estimation (1/2)

$$C(h) = \kappa^2(1+\frac{h}{L})\exp(-\frac{h}{L})$$

- Hyper-parameters vector $\boldsymbol{\theta} = [\kappa^o, \kappa^b, L]^{\mathrm{T}})$
- Innovation vector $\mathbf{d} = \mu - \mathbf{H}\sigma^b$
- Innovation error covariance matrix $\mathbf{D}_{\boldsymbol{\theta}} = \mathbf{R}_{\boldsymbol{\theta}} + \mathbf{H}\mathbf{B}_{\boldsymbol{\theta}}\mathbf{H}^{\mathrm{T}}$

▶ Likelihood

$$p(\mu|\boldsymbol{\theta}) = \frac{\exp\left(-\frac{1}{2}(\mu - \mathbf{H}\sigma^b)^{\mathrm{T}}\mathbf{D}_{\boldsymbol{\theta}}^{-1}(\mu - \mathbf{H}\sigma^b)\right)}{(2\pi)^{\frac{d}{2}}|\mathbf{D}_{\boldsymbol{\theta}}|^{\frac{1}{2}}}$$

- Maximum likelihood estimation (MLE): minimizing negative log-likelihood

$$\mathscr{L}(\boldsymbol{\theta}) = \frac{1}{2}\ln|\mathbf{D}_{\boldsymbol{\theta}}| + \frac{1}{2}(\mu - \mathbf{H}\sigma^b)^{\mathrm{T}}\mathbf{D}_{\boldsymbol{\theta}}^{-1}(\mu - \mathbf{H}\sigma^b)$$

- Desroziers Scheme: given $L$, solve MLE iteratively for $[\kappa^o, \kappa^b]^{\mathrm{T}}$

# Uncertainty quantification: hyper-parameter estimation (2/2)

- $\chi^2$: Gaussian assumptions lead to $\chi^2$ probability density with number of degrees of freedom equal to the number of observations $d$:

$$\chi^2(\sigma^a) = (\mu - \mathbf{H}\sigma^a)^{\mathrm{T}}\mathbf{R}_{\boldsymbol{\theta}}^{-1}(\mu - \mathbf{H}\sigma^a) + (\sigma^a - \sigma^b)^{\mathrm{T}}\mathbf{B}_{\boldsymbol{\theta}}^{-1}(\sigma^a - \sigma^b)$$

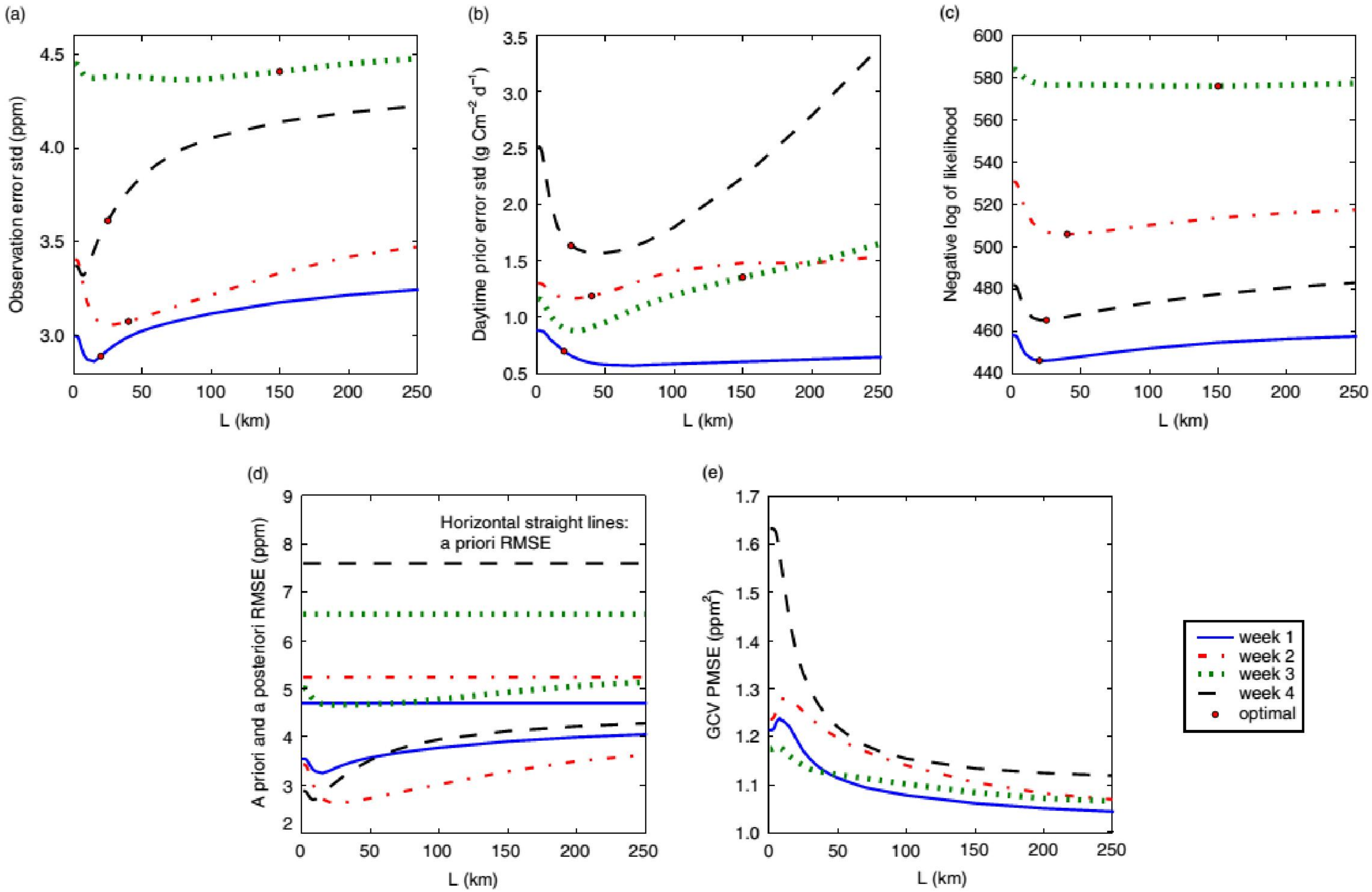
- Information propagation:

$$\mathrm{DFS} = \mathrm{Tr}\left(\mathbf{B}_{\boldsymbol{\theta}}\mathbf{H}^{\mathrm{T}}\mathbf{D}_{\boldsymbol{\theta}}^{-1}\mathbf{H}\right).$$

- General Cross Validation (GCV) minimizes the predictive mean-square error (PMSE) formulated in $\mathbf{R}^{-1}$-norm:

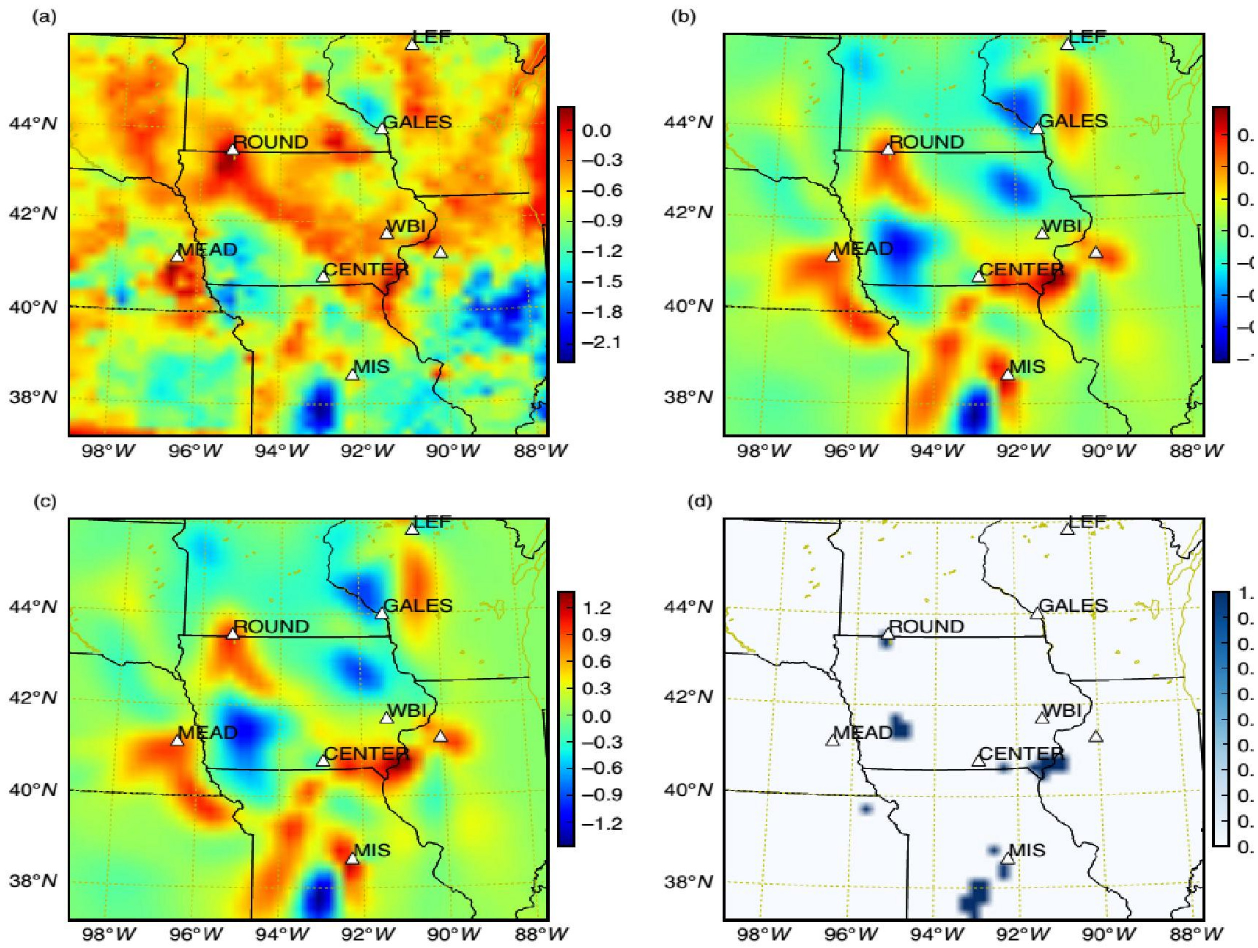
$$P(\boldsymbol{\theta}) = \frac{1}{d}\|\mathbf{H}\left(\sigma^t - \sigma^a\right)\|_{\mathbf{R}^{-1}}^2$$

# Results

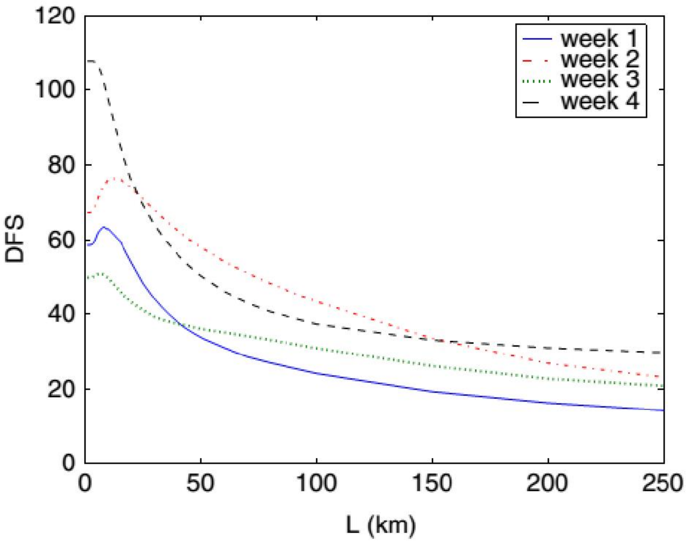


(a)
Observation error std (ppm)
L (km)
(b)
Daytime prior error std (g Cm$^{-2}$ d$^{-1}$)
L (km)
(c)
Negative log of likelihood
L (km)
(d)
Horizontal straight lines:
a priori RMSE
A priori and a posteriori RMSE (ppm)
L (km)
(e)
GCV PMSE (ppm$^2$)
L (km)
week 1
week 2
week 3
week 4
optimal

# Results



(a)
(b)
(c)
(d)
LEF
GALES
ROUND
WBI
MEAD
CENTER
MIS
44°N
42°N
40°N
38°N
98°W
96°W
94°W
92°W
90°W
88°W
0.0
−0.3
−0.6
−0.9
−1.2
−1.5
−1.8
−2.1
1.2
0.9
0.6
0.3
0.0
−0.3
−0.6
−0.9
−1.2




## Uncertainties for hyperparameter estimations

| | $\sigma_o$ | Daytime $\sigma_b$ | $\sigma_b L$ |
|---|---|---|---|
| Week 1 | $2.89 \pm 0.149$ | $3.21 \pm 1.13$ | $20 \pm 6.77$ |
| Week 2 | $3.08 \pm 0.181$ | $5.45 \pm 1.99$ | $40 \pm 13.6$ |
| Week 3 | – | – | – |
| Week 4 | $3.62 \pm 0.241$ | $7.51 \pm 3.48$ | $25 \pm 7.90$ |

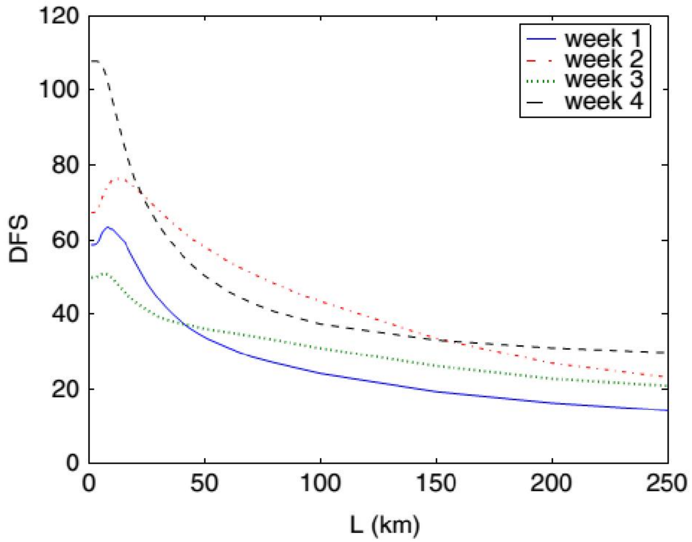
$$\mathcal{H}_{ij}(\boldsymbol{\theta}) = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}. \qquad \mathcal{H}(\boldsymbol{\theta}^*)^{-1}.$$

## Summary

Evaluates diiverse criteria: MLE, $\chi^2$, GCV using real data

Short correlation length: summer time mostly 15 -80 km, occasionally 100 km; confirms the results obtained by direct aggregation of background errors.

When atmospheric transport error is significant, difficult to identify a meaningful optimal L




## Uncertainties for hyperparameter estimations

| | $\sigma_o$ | Daytime $\sigma_b$ | $\sigma_b L$ |
|---|---|---|---|
| Week 1 | $2.89 \pm 0.149$ | $3.21 \pm 1.13$ | $20 \pm 6.77$ |
| Week 2 | $3.08 \pm 0.181$ | $5.45 \pm 1.99$ | $40 \pm 13.6$ |
| Week 3 | – | – | – |
| Week 4 | $3.62 \pm 0.241$ | $7.51 \pm 3.48$ | $25 \pm 7.90$ |

$$\mathcal{H}_{ij}(\boldsymbol{\theta}) = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}. \qquad \mathcal{H}(\boldsymbol{\theta}^*)^{-1}.$$

- **Why DFS is not a proper criterion for hyperparameter estimation?**
- **Why DFS decreases with respect to correlation length?**

# Few words

**Inversion as a system**: three components: model, obs, stat information,
Optimal control theory, success of variational methods (4DVar)
Other system concepts, e.g. observability?

**Inversion as an information machine**: information fusion and flow
relative entropy, entropy dynamics, maximum entropy principle

**Second order inversion**: optimal configuration of inversion system
& **Uncertainty quantification:**
Model resolution & aggregation error, error parameter estimation,

# References


M.Bocquet, L.Wu and F.Chevallier, Bayesian design of control space for optimal assimilation of observations. I: Consistent multiscale formalism, Q.J.R.M.S., 137: 1340{1356, 2011

F.Chevallier, T.Wang, P.Ciais, F.Maignan, M.Bocquet, et al., What eddy-covariance flux measurements tell us about prior errors in CO2 flux inversion schemes, Global Biogeochem. Cy., 26, GB1021, 2012

F. Chevallier  and C. W. O'Dell, Error statistics of Bayesian CO2 flux inversion schemes as seen from GOSAT, GRL, 2013, 40, 1252–1256

L.Wu, M.Bocquet, T.Lauvaux, F.Chevallier, P.Rayner, K.Davis, Optimal representation of sources for mesoscale carbon dioxide inversion with synthetic data, JGR, 116, D21304, 2011

L.Wu, M.Bocquet, T.Lauvaux, F.Chevallier, P.Rayner, K.Davis, Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions, Tellus B, 2013, 65, 20894