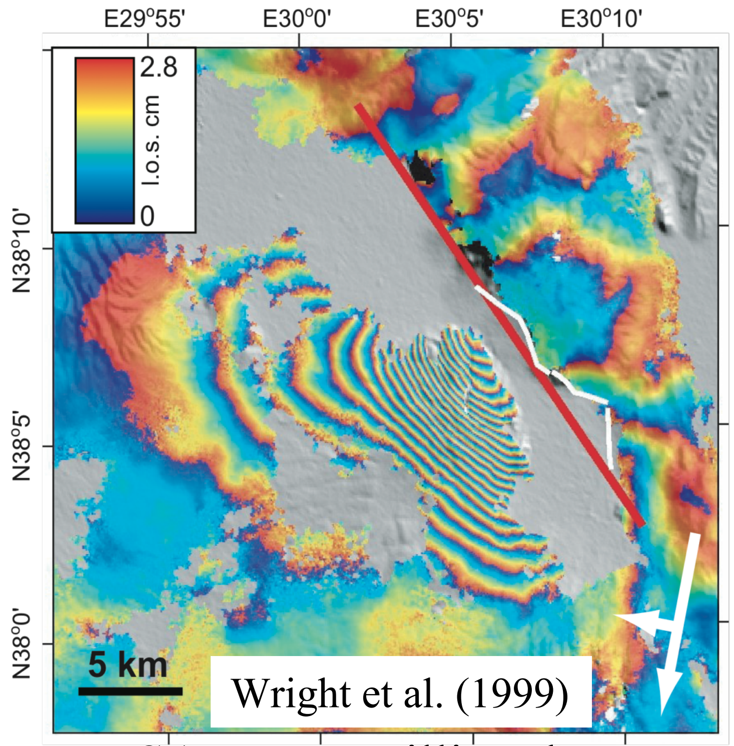
2010.6.15 EOST, Strasbourg

How much information do observed data have?: Importance of covariance components in inverting densely sampled observed data

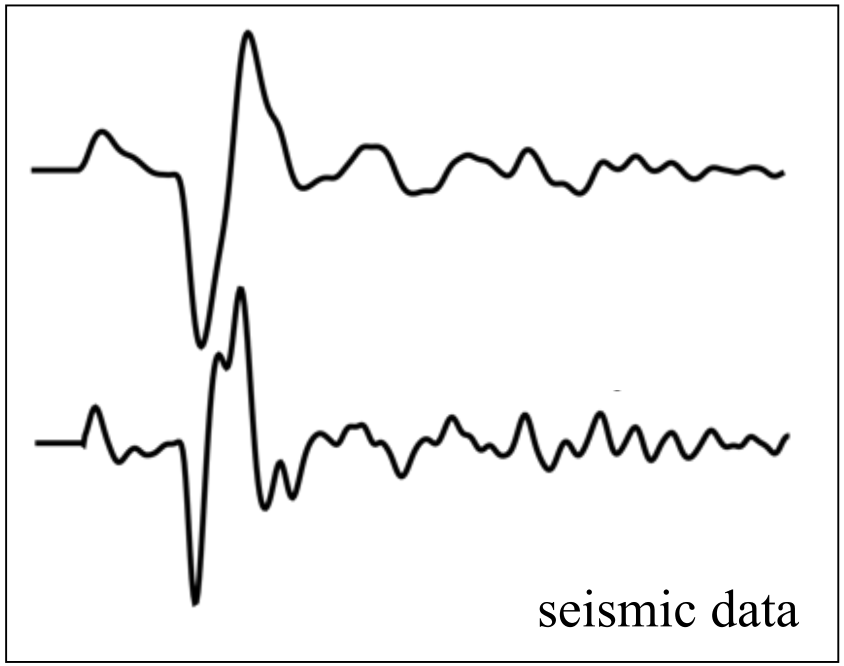
Yukitoshi Fukahata (Kyoto University)
Yuji Yagi (Tsukuba University)

Motivation

We have nominally continuous observed data



InSAR: one-million data



How should we sample and invert such data?

Contents

1. Framework of inversion analysis
   (Introduction of ABIC)

2. Covariance for observation error

3. Covariance for modeling error

Point of Inversion Analysis

We have two sorts of information.

1. Observed data

2. Prior Information (Jackson, 1979; Tarantola, 1987)
   e.g., density of the crust
   velocity of P and S waves
   smoothness in the slip

The relative importance between them is objectively determined by ABIC from observed data

Algorithm for linear inversion analysis (1)

(Based on Yabuki & Matsu'ura, 1992)

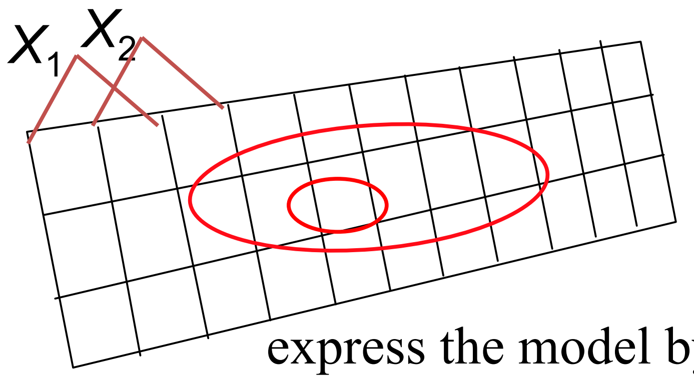
0. Relation between data and model

$$d(\mathbf{x}) = \int G(\mathbf{x};\xi)u(\xi)d\xi$$

1. Parametrization

$$u(x) = \sum_{m=1}^{M} a_m X_m(x)$$

u: model, quantity we want to know (e.g., slip distribution; density)
a_m: model parameter
$X_m(x)$: basis function
d: observed data
G: green's function



express the model by super-position of basis functions

Algorithm for linear inversion analysis (2)

2. Observation Equation

$$\mathbf{d} = \mathbf{H}\mathbf{a} + \mathbf{e} \qquad \mathbf{d}\text{: data, } \mathbf{e}\text{: error,} \qquad \mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{E})$$

$$p\left(\mathbf{d} \mid \mathbf{a}; \sigma^2\right) = (2\pi\sigma^2)^{-\frac{n}{2}} |\mathbf{E}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2\sigma^2}\left(\mathbf{d} - \mathbf{H}\mathbf{a}\right)^T \mathbf{E}^{-1}\left(\mathbf{d} - \mathbf{H}\mathbf{a}\right)\right]$$

3. Prior Information (ex. smoothness condition)

$$\int_{XY} \left[\left(\partial u / \partial x\right)^2 + \left(\partial u / \partial y\right)^2\right] dxdy = \mathbf{a}^T \mathbf{G} \mathbf{a} \quad \rightarrow \quad \text{small}$$

$$p(\mathbf{a}; \rho^2) = (2\pi\rho^2)^{-\frac{m}{2}} |\mathbf{G}|^{\frac{1}{2}} \exp\left[-\frac{1}{2\rho^2} \mathbf{a}^T \mathbf{G} \mathbf{a}\right]$$

Algorithm for linear inversion analysis (3)

Observation eq.

$$\mathbf{d} = \mathbf{Ha} + \mathbf{e}$$

$$p(\mathbf{d} \mid \mathbf{a}; \sigma^2)$$

Prior information

$$r = \mathbf{a}^T \mathbf{Ga}$$

$$p(\mathbf{a}; \rho^2)$$

Bayes Thorem

$$p(\mathbf{a}; \sigma^2, \rho^2 \mid \mathbf{d}) = c p(\mathbf{d} \mid \mathbf{a}; \sigma^2) p(\mathbf{a}; \rho^2)$$

$$= c\left(2\pi\sigma^2\right)^{-(n+m)/2} (\alpha^2)^{P/2} |\mathbf{E}|^{-1/2} |\mathbf{G}|^{1/2} \times \exp\left[-\frac{1}{2\sigma^2} s(\mathbf{a}; \alpha^2)\right]$$

$$s(\mathbf{a}) = (\mathbf{d} - \mathbf{Ha})^T \mathbf{E}^{-1} (\mathbf{d} - \mathbf{Ha}) + \alpha^2 \mathbf{a}^T \mathbf{Ga}$$

Maximization of probability $p(\mathbf{a}; \sigma^2, \rho^2 \mid \mathbf{d})$ *gives the optimal solution*

$$p(\mathbf{a};\sigma^2,\rho^2 \mid \mathbf{d}) = cp(\mathbf{d} \mid \mathbf{a};\sigma^2)\,p(\mathbf{a};\rho^2)$$

$$= c\left(2\pi\sigma^2\right)^{-(n+m)/2}(\alpha^2)^{P/2}\left|\mathbf{E}\right|^{-1/2}\left|\mathbf{G}\right|^{1/2} \times \exp\left[-\frac{1}{2\sigma^2}s(\mathbf{a};\alpha^2)\right]$$

$$s(\mathbf{a}) = (\mathbf{d}-\mathbf{Ha})^T\mathbf{E}^{-1}(\mathbf{d}-\mathbf{Ha}) + \alpha^2\mathbf{a}^T\mathbf{Ga}$$

$\alpha^2 = \sigma^2/\rho^2$: ratio of variance between observed data and prior information

This expression includes various solution as a special case

(i) Enough data $\longrightarrow$ $\hat{\mathbf{a}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{d}$: least squares solution

(ii) $\mathbf{G}=\mathbf{I}$ $\rightarrow$ $\hat{\mathbf{a}} = (\mathbf{H}^T\mathbf{H}+\alpha^2\mathbf{I})^{-1}\mathbf{H}^T\mathbf{d}$: damped least squares solution

(iii) $\mathbf{G}=\mathbf{G}$ $\rightarrow$ $\hat{\mathbf{a}} = (\mathbf{H}^T\mathbf{H}+\alpha^2\mathbf{G})^{-1}\mathbf{H}^T\mathbf{d}$: Laplacian condition

Remaining problem: ***How do we determine α^2?***

Introduction of ABIC (Akaike, 1980)

(ABIC: Akaike's Bayesian Information Criterion)

Definition of ABIC

$$\mathrm{ABIC} = -2\log L(\sigma^2, \alpha^2 \,|\, \mathbf{d}) + 2 \times (\text{number of hyperparametes})$$

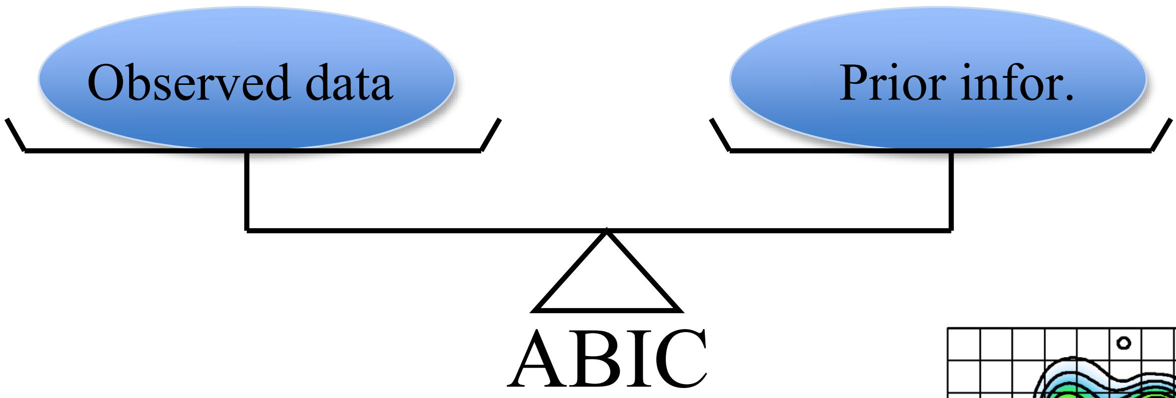
L: marginal likelihood for σ^2 and α^2

$$L(\sigma^2, \alpha^2 \,|\, \mathbf{d}) = \int p(\mathbf{d} \,|\, \mathbf{a}; \sigma^2)\, p(\mathbf{a}; \alpha^2)\, d\mathbf{a}$$

The criterion of ABIC minimum $\rightarrow$ σ^2, α^2

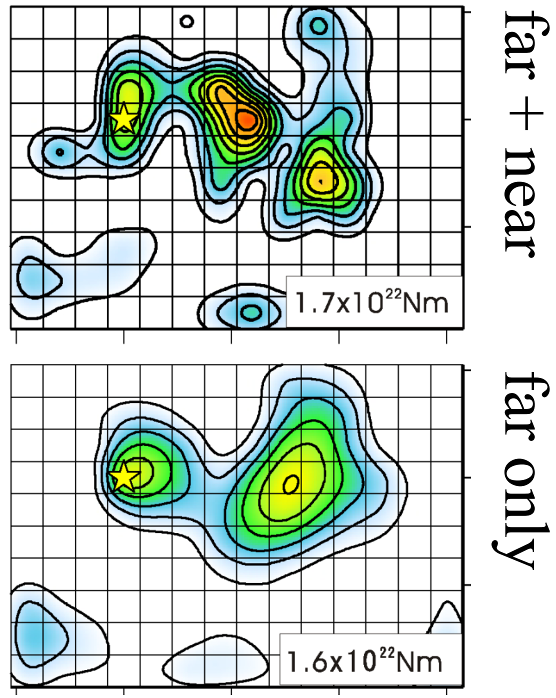
objectively determined from observed data

Characteristic of ABIC



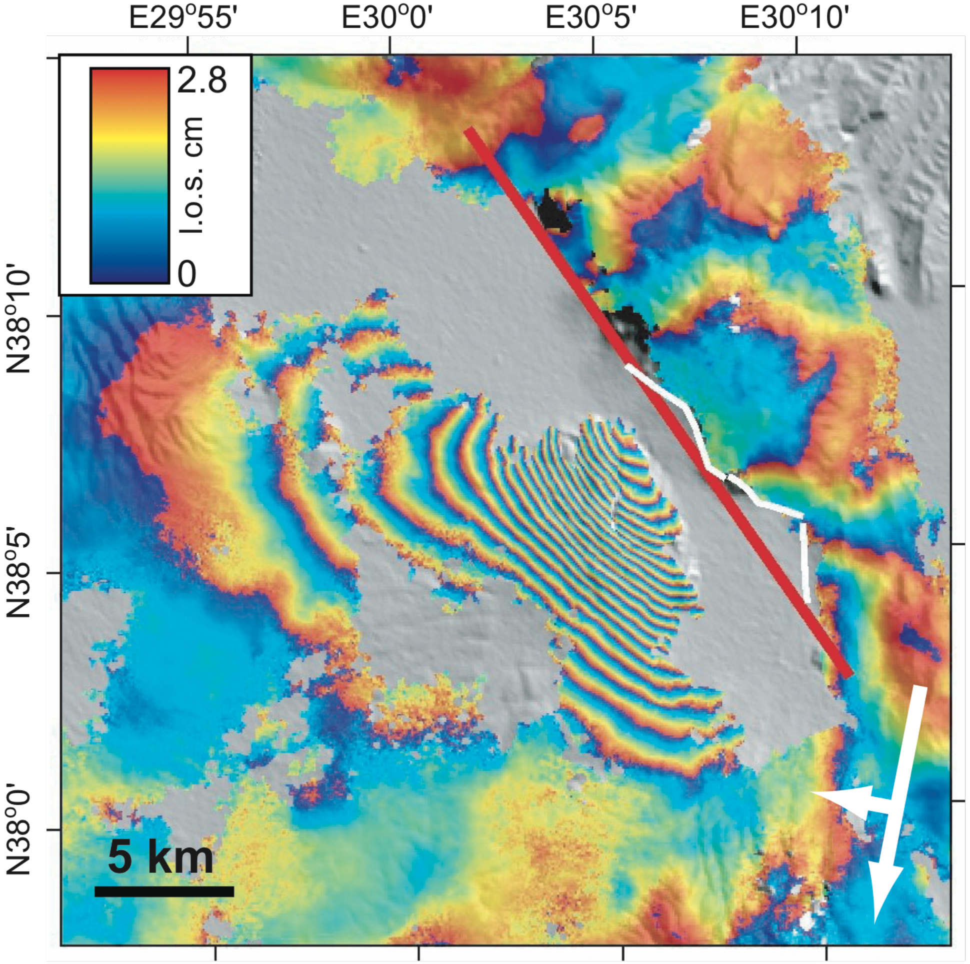
Observed data: accurate, sufficient
→ the model that fits to the data

Observed data: inaccurate, insufficient
→ the model that follows prior



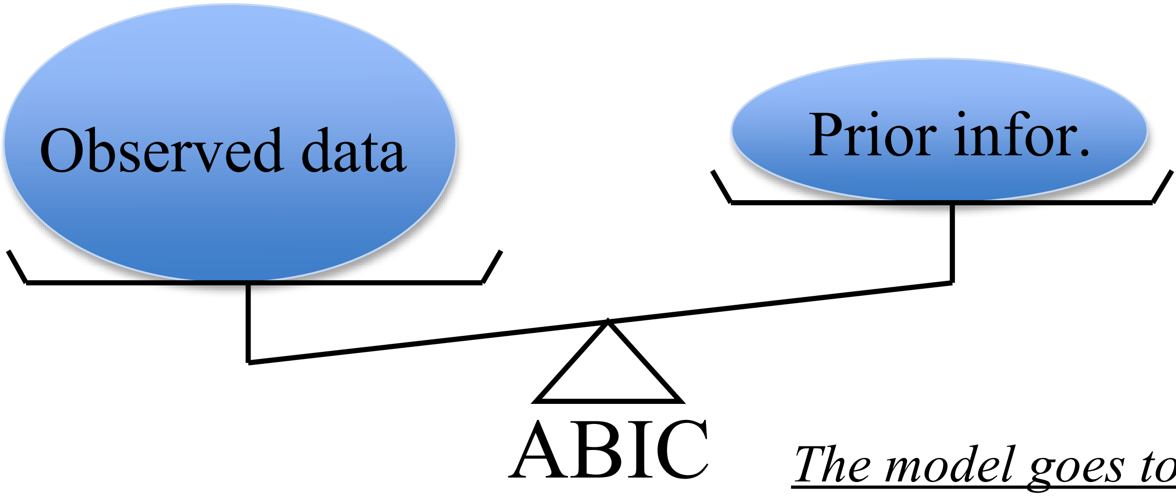
Fukahata, Yagi & Matsu'ura (2003)

How should we sample and invert continuous data?



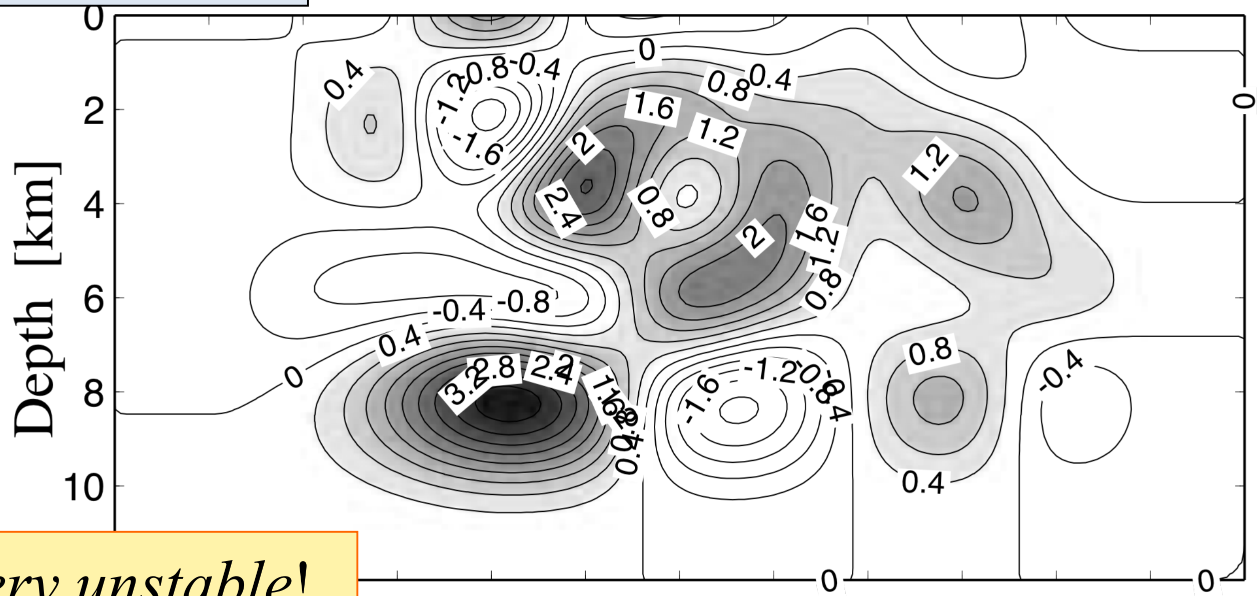
Fukahata & Wright (2008)

For higher sampling rate,
the information from ob. data apparently increases.



The model goes to fit to the data

An inverted result



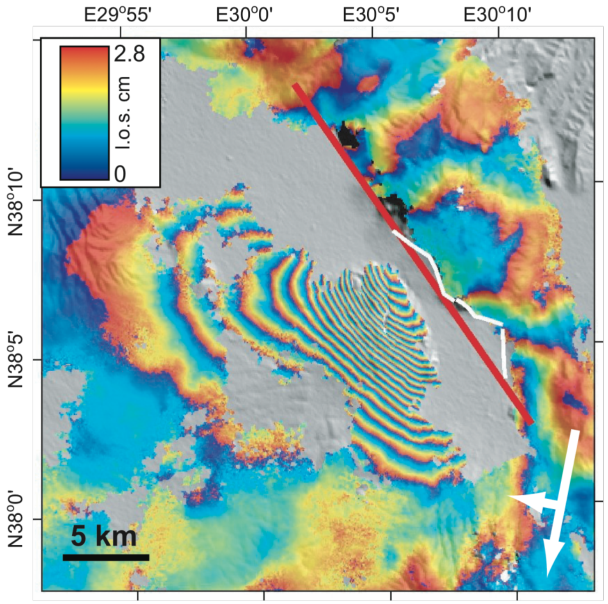
very unstable!

Slip distribution of Dinar earthquake

For InSAR data

Observation Eq. : $\mathbf{d} = \mathbf{Ha} + \mathbf{e}$

Observed data have spatially correlated errors mainly due to atmospheric noise.



$$\mathbf{e} \sim N(\mathbf{0}, \sigma^2 \mathbf{E})$$

$$E_{ij} = \exp\left(-r_{ij}/s\right)$$

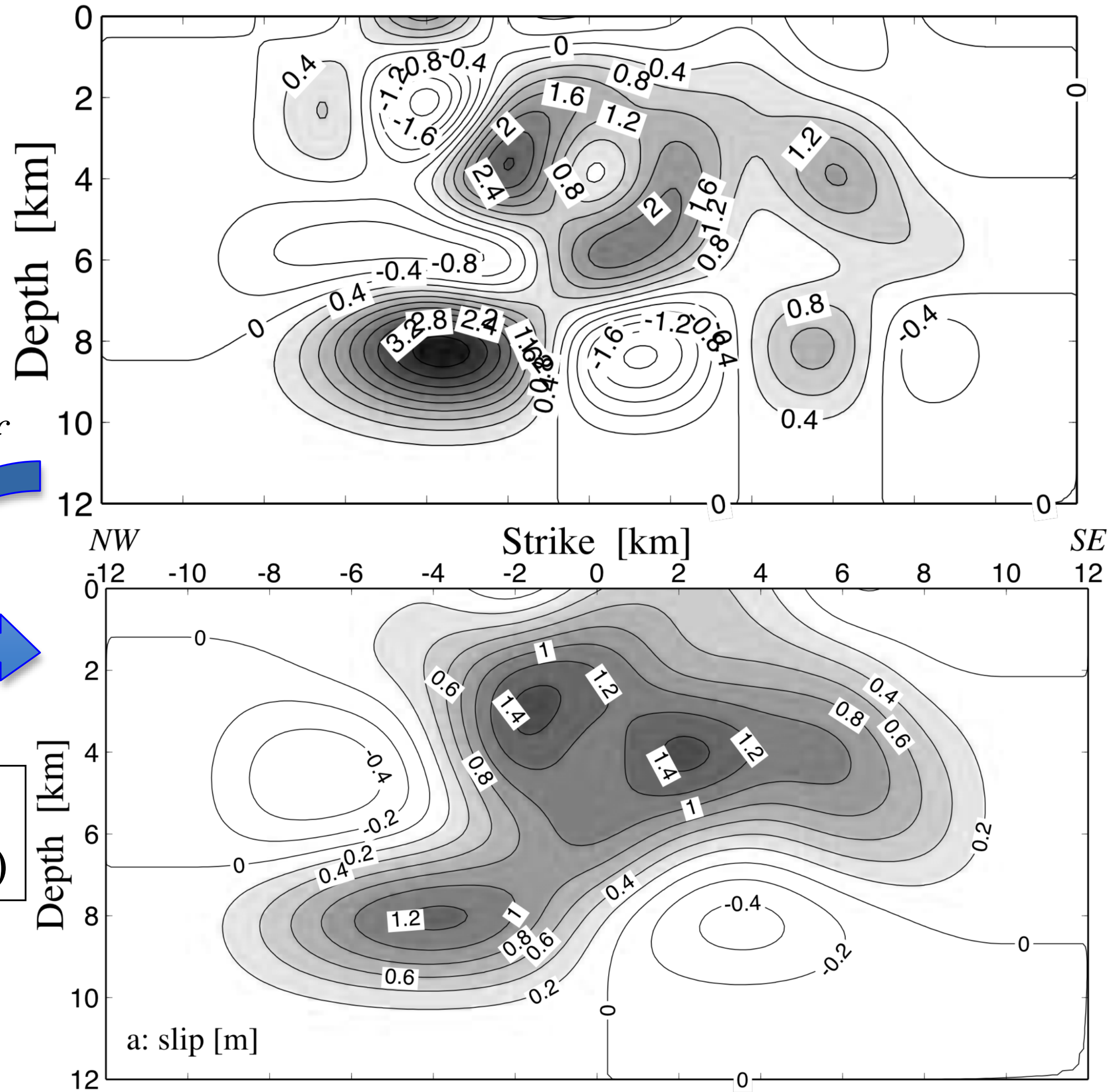
r_{ij} : distance between data i and j

s : typical correlation length (~10km)

$$s(\mathbf{a};\alpha^2) = \underbrace{(\mathbf{d}-\mathbf{Ha})^T \mathbf{E}^{-1} (\mathbf{d}-\mathbf{Ha})}_{\text{square of residual}} + \underbrace{\alpha^2 \mathbf{a}^T \mathbf{G} \mathbf{a}}_{\text{smoothness}}$$

By introducing covariance components,
we can properly control the information from ob. data

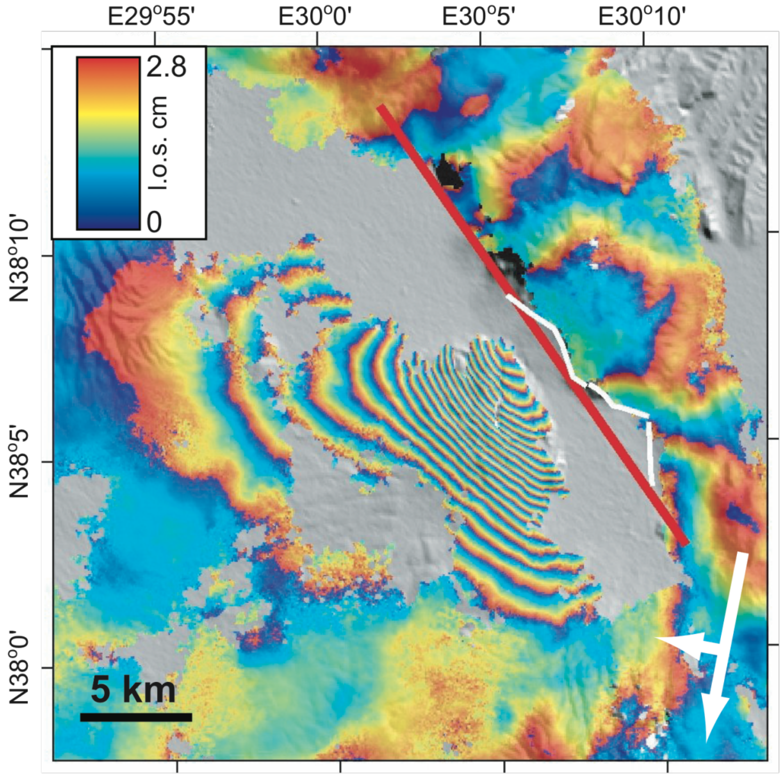
Introduction of covariance

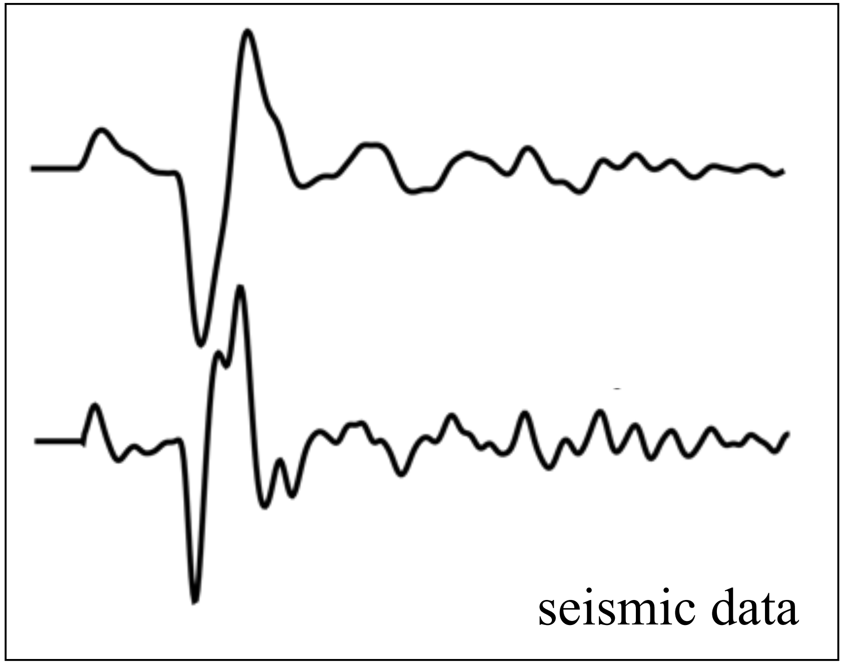


Fukahata & Wright (2008)

cf. Langbein & Johnson (1997); Segall et al. (2000); Lohman & Simons (2005)

When observation error is so small .....

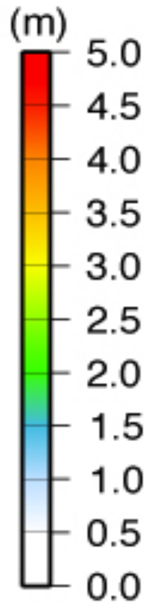




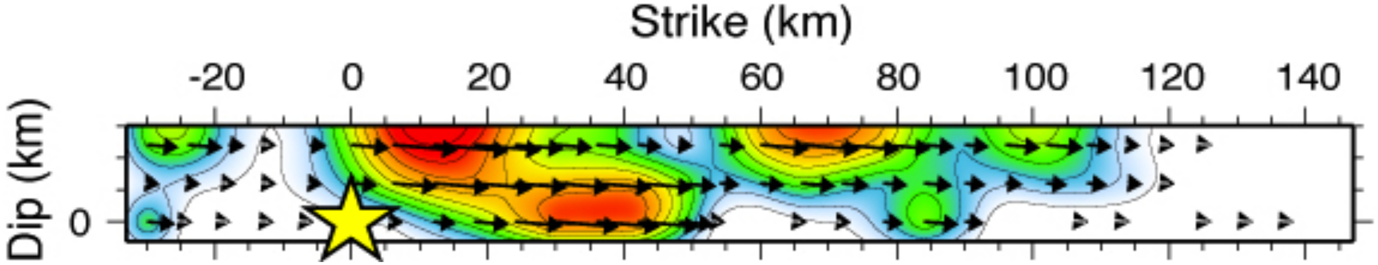
Is the information from observed data proportionally increases according to the sampling rate?

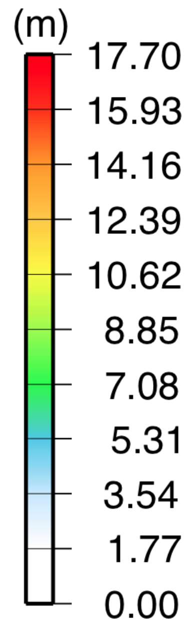
→ *The answer is NO*

Inversion of slip distribution for Manyi, Tibet, earthquake (1997, Mw=7.6)

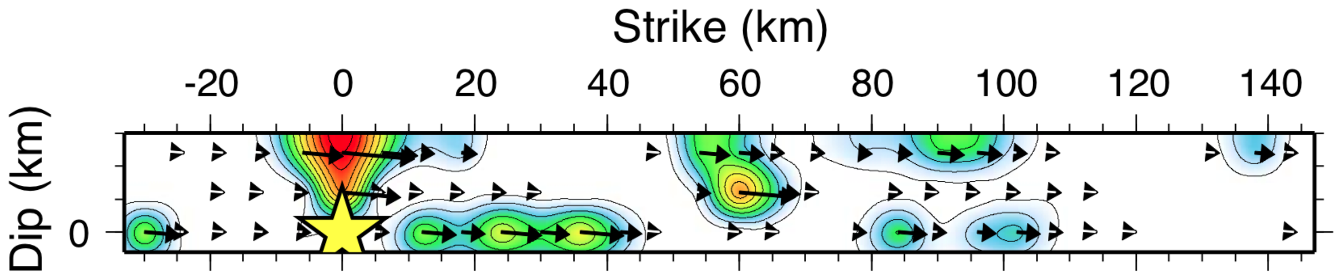


＜Before＞ *reasonable*





＜After＞ *very unstable!*



(On behalf of Funning by Yagi)

Error and observation equation

Observation eq.: $\mathbf{d} = \mathbf{H}\mathbf{a} + \mathbf{e}$

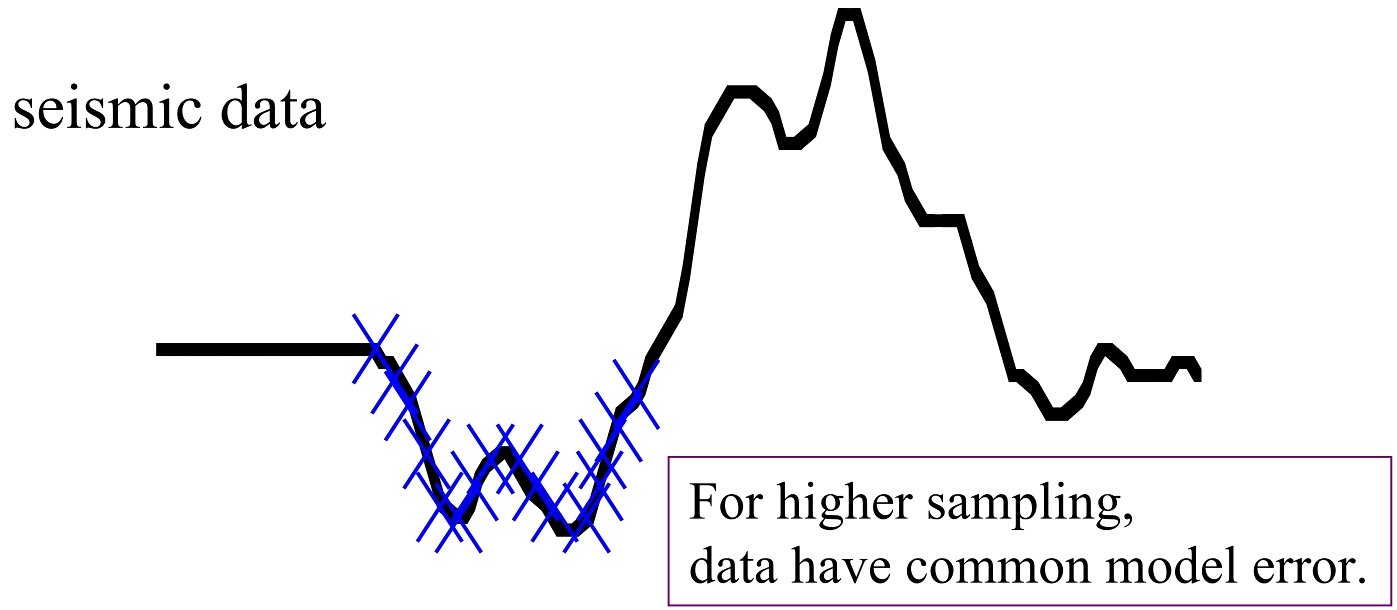
$$\mathbf{e} = \mathbf{d} - \mathbf{H}\mathbf{a}$$
$$= \mathbf{e}^{obs} + \mathbf{e}^{model}$$

The error in obs. eq. is the sum of obs. error and modeling error.

$\mathbf{e}^{obs} = \mathbf{d} - \mathbf{d}^0$: observation error

$\mathbf{e}^{model} = \mathbf{H}\mathbf{a} - \mathbf{H}^0\mathbf{a}$: modeling error

significant correlation



mathematically,

$$\mathbf{E} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \Rightarrow \begin{pmatrix} \sigma^2 & * & * & 0 \\ * & \sigma^2 & * & * \\ * & * & \ddots & * \\ 0 & * & * & \sigma^2 \end{pmatrix}$$

Introduction of covariance components due to modeling error

(Yagi & Fukahata, 2008)

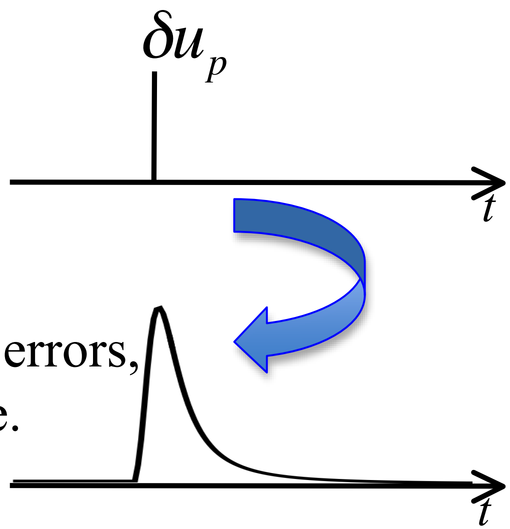
discretization error： $u(\tau) = \sum_{k=1}^{K} a_k T_k(\tau) + \delta u(\tau)$

When parameterizing the problem, discretization error inevitably emerges

Relation between data and model： $d_i(t) = \int_s G_i(t;\tau) u(\tau) d\tau$

Expression of the error in the observation eq.：

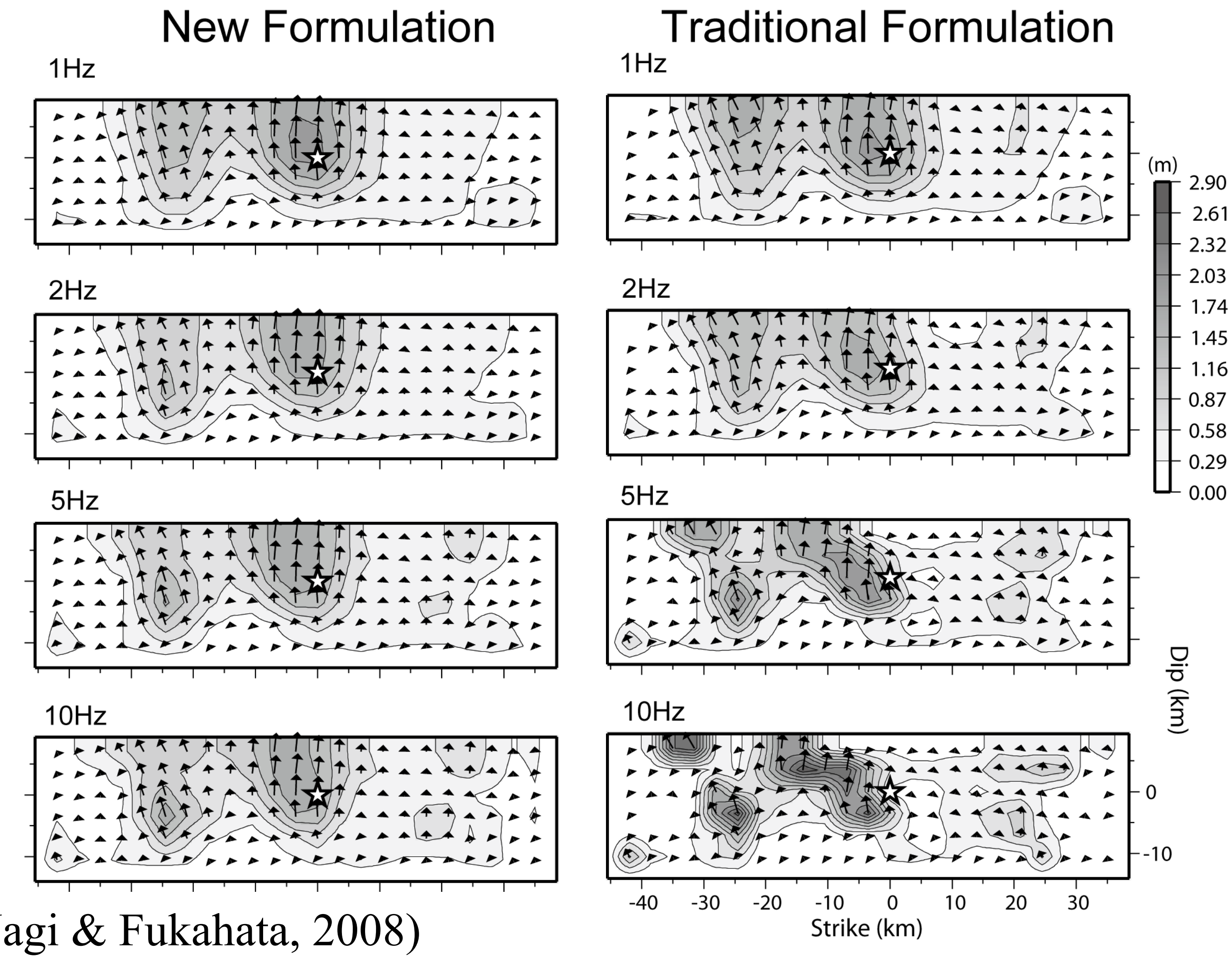
$$e_i^{discre}(t) = \int_s G_i(t;\tau) \delta u(\tau) d\tau$$



Following the law of propagation of errors,
covariance components emerge.

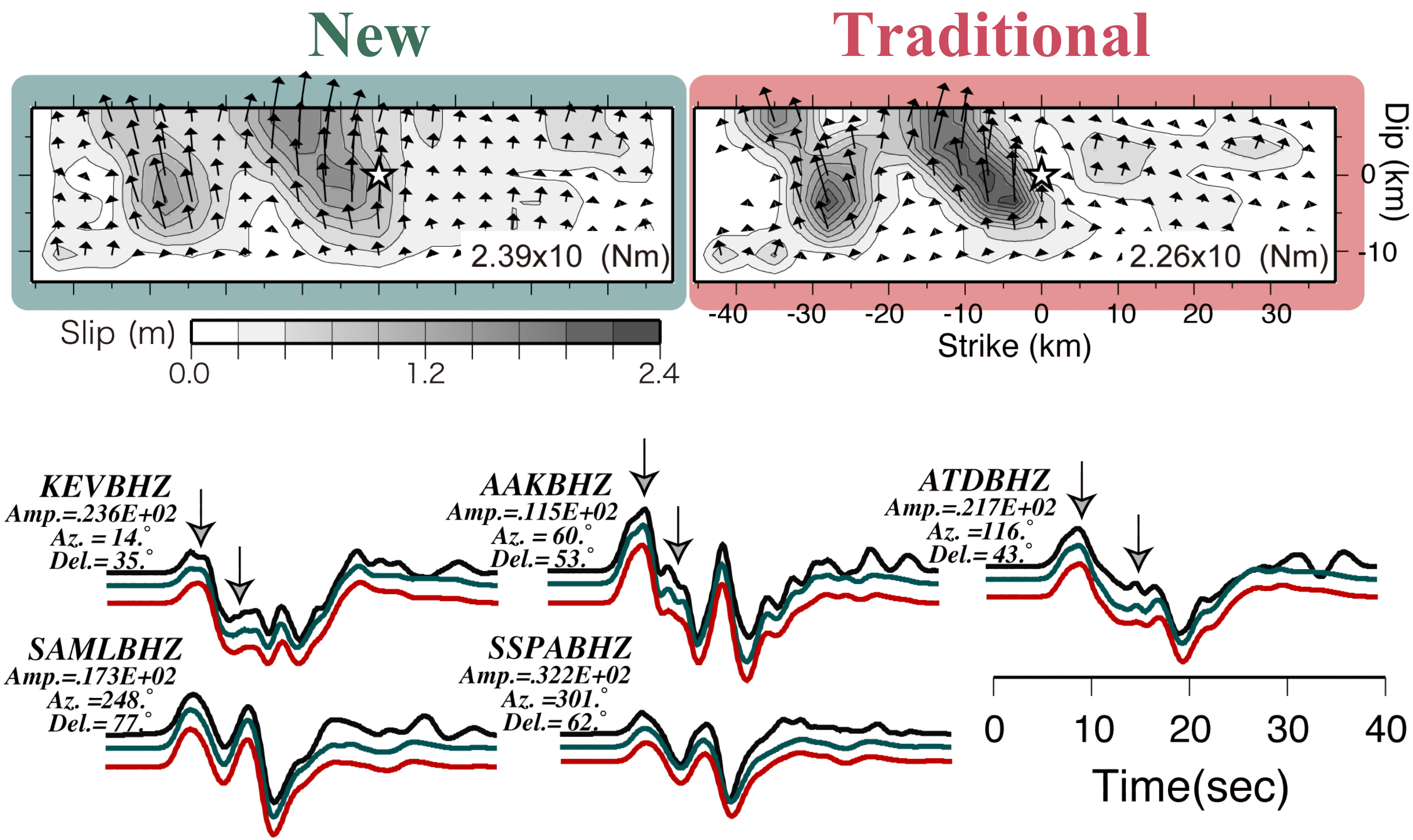
Introducing the covariance components

solved the problem that the solution depends on the sampling rate



(Yagi & Fukahata, 2008)

Reproduction of high frequency components



Note that the residual mean square is *less* in the traditional

Summary

Due to the development of computers, we can now invert continuous observed data with a very high sampling rate.

In inverting densely sampled observed data, the effects of covariance components can be essential.

The error in the observation equation is the sum of observation error and modeling error. So, even if we can neglect observation error, we cannot escape from modeling errors.

By introducing covariance components, we have solved a sampling rate problem for InSAR and seismic data analysis.

More appropriate evaluation of errors (e.g., Green's function; spatially different error) are still needed.