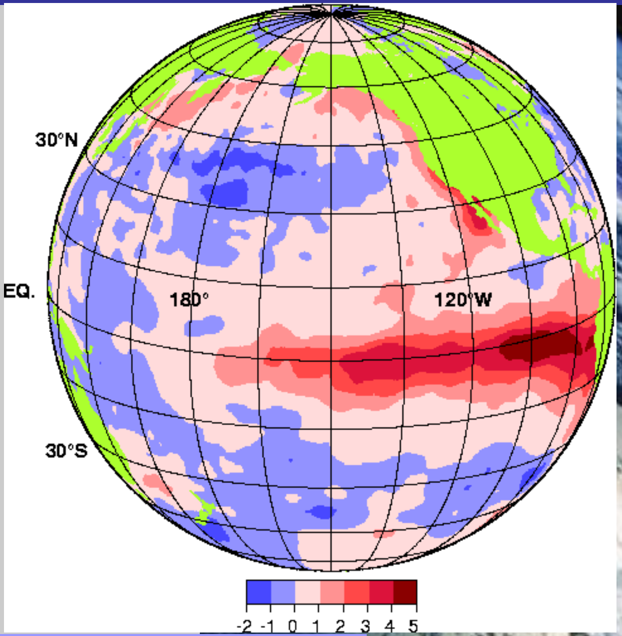
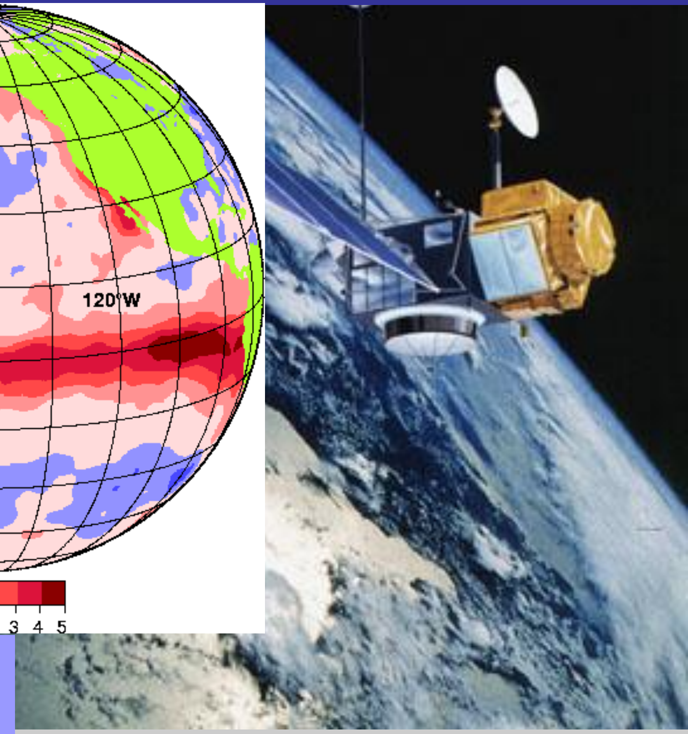
データ同化モデルにおける観測ノイズ分散共分散行列のベイズ推定法

1. 状態空間モデル
2. 観測ノイズ分散共分散行列の推定





上野玄太 (統計数理研究所)

準備：状態空間モデル

- シミュレーションモデルでは思うように現象が再現・予測できない。そのため、シミュレーションモデルの大規模化・精緻化・階層化などを進める[いわゆる「演繹一本槍」]

- シミュレーションモデルと観測データの間にも自由度を持った関係を構築する。

- 以上の２つをそれぞれシステムモデル、観測モデルと呼び、それらを合わせて状態空間モデルという。データ同化の基本となるモデルである。

シミュレーション (1/2)

x の時間発展を与える微分方程式

$$\frac{dx}{dt} = g(x,t)$$

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u = -\frac{1}{\rho}\nabla p + \nu \nabla^2 u + f$$

$$\frac{\partial f}{\partial t} + v \cdot \frac{\partial f}{\partial x} + \frac{F}{m}\frac{\partial f}{\partial v} = \left(\frac{\delta f}{\delta t}\right)_{coll}$$

離散化

$$\frac{x_t - x_{t-1}}{\Delta t} = g\left(x_{t-1}, t\right)$$

移項

$$x_t = x_{t-1} + g\left(x_{t-1}, t\right)\Delta t$$
$$=: f_t\left(x_{t-1}\right)$$

シミュレーション (2/2)

シミュレーションモデル

$$x_t = f_t(x_{t-1}), \quad t = 1, 2, \cdots, T$$

•x :シミュレーション変数をまとめたベクトル量で、**状態**もしくは**状態ベクトル**と呼ぶ。
•x_0: **初期状態**もしくは**初期条件**と呼ぶ。シミュレーションの計算の前にあらかじめ与える

上の設定から当たり前のことではあるが、
初期状態 x_0 を与えると、その後の状態の列
$(x_1, x_2, \ldots, x_T)$ が**一意に決まる**。

時間発展の不確定性、システムノイズ

$$x_t = f_t(x_{t-1}) \quad \rightarrow \quad x_t \approx f_t(x_{t-1}), \quad t = 1, 2, \cdots, T$$

近似等号を表現するために、付加項 v_t を含めて等号で結ぶ：

$$\rightarrow \quad x_t = f_t(x_{t-1}) + v_t, \quad t = 1, 2, \cdots, T$$

付加項 v_t の確率分布を仮定する：

$$v_t \sim p(v_t), \quad t = 1, 2, \cdots, T$$

v_tを**システムノイズ**と呼ぶ。

システムモデル

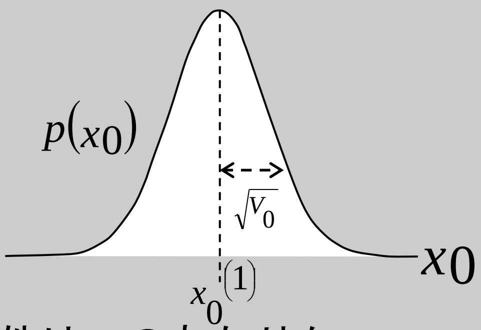
シミュレーションモデル

$$x_0 = x_0^{(1)}$$

初期条件はこうだ！

$$x_t = f_t(x_{t-1}), \quad t = 1,2,\cdots,T$$

時間発展はこうだ！



システムモデル

初期状態の確率分布

$$x_0 \sim p\left(x_0; x_0^{(1)}, V_0\right)$$

初期条件はこのあたりか？

$$x_t = f_t(x_{t-1}) + v_t, \quad t = 1,2,\cdots,T$$

時間発展はこんなものか？

$$v_t \sim p(v_t; Q_t), \quad t = 1,2,\cdots,T$$

システムノイズの確率分布

モデルとデータのつき合わせ

状態

$x_t, \quad t = 1, 2, \cdots, T$

k次元ベクトルの列

つき合わせたい

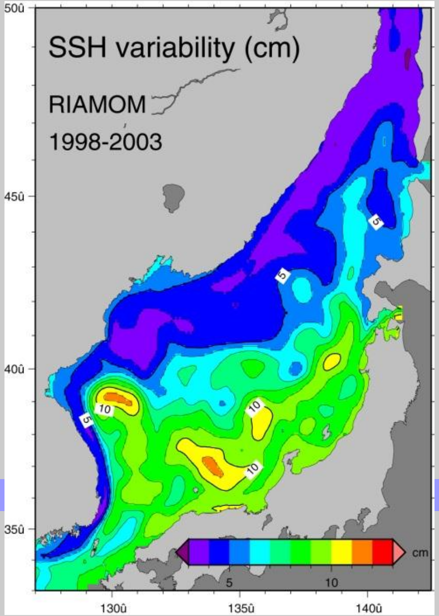
観測データ

$y_t, \quad t = 1, 2, \cdots, T$

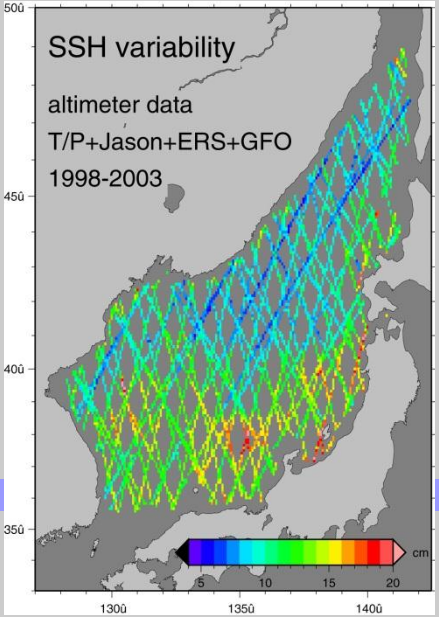
l次元ベクトルの列

状態は多くの変数を要素に持つベクトルであるが、
実際に観測されるのは一部の変数である ($l < k$)。

x_t



y_t



広瀬・万田
(2006)

観測モデル、観測ノイズ

考え方：得られた状態は観測と近い

$$y_t \approx h_t(x_t), \quad t = 1,2,\cdots,T$$

近似等号を表現するために、付加項 w_t を含めて等号で結ぶ：

$$\rightarrow \quad y_t = h_t(x_t) + w_t, \quad t = 1,2,\cdots,T$$

付加項 w_t の確率分布を仮定する：

$$w_t \sim p(w_t; R_t), \quad t = 1,2,\cdots,T$$

観測モデル

w_tを**観測ノイズ**と呼ぶ。

観測ノイズ

- 観測ノイズ w_t の定義は、 観測とモデルの差 y_t-$h_t(x_t)$

- 観測機器の特性に起因する、いわゆる「測定誤差」は、観測ノイズに寄与する要素の一部と考えるべき
[観測ノイズ、という言葉から類推しないでください]

- 極端な例として、観測ノイズと「測定誤差」が一致する場合を考えると、モデルが観測機器をとりまく自然現象を完璧に再現できる場合に対応
- 通常はモデルにはそこまでの能力はないため、モデルが再現し切れない部分も差としてカウントされる

状態空間モデル

システムモデル、観測モデルからなるモデルを
状態空間モデルと呼ぶ。

システムモデル

$$x_0 \sim p\left(x_0; x_0^{(1)}, V_0\right)$$

$$x_t = f_t(x_{t-1}) + v_t, \quad t = 1, 2, \cdots, T$$

$$v_t \sim p(v_t; Q_t), \quad t = 1, 2, \cdots, T$$

観測モデル

$$y_t = h_t(x_t) + w_t, \quad t = 1, 2, \cdots, T$$

$$w_t \sim p(w_t; R_t), \quad t = 1, 2, \cdots, T$$

（狭い意味で）データ同化では、初期状態の確率分布、システムノイズの確率分布、観測ノイズの確率分布を与えたときに、観測データに基づいて、状態の確率分布を求める。

データ同化における状態空間モデル

- データ同化 = 状態変数 x_t を推定する問題

$$x_t \approx f_t(x_{t-1}) \qquad y_t \approx h_t(x_t) \quad \Leftrightarrow \quad x_t = f_t(x_{t-1}) + v_t \quad \text{(システムモデル)} \qquad y_t = h_t(x_t) + w_t \quad \text{(観測モデル)}$$

  - x_t : タイムステップ t でのシミュレーションの全変数をまとめたベクトル　　(x_0: 初期条件)
  - y_t : タイムステップ t での観測される全変数
  - v_t : システムノイズ. シミュレーションの不確定部分を表現する確率変数 (不確定部分: 境界条件、パラメータ、 …)
  - w_t : 観測ノイズ

変数の次元　x_t: 10^4~10^6　y_t : 10^2~10^5　$\dim(x_t) >> \dim(y_t)$

データ同化における共分散行列

状態空間モデル

$$x_t = f_t\left(x_{t-1}, v_t\right)$$ システムモデル

$$y_t = h_t\left(x_t\right) + w_t$$ 観測モデル

共分散行列

システムノイズ $v_t \sim N\left(0, Q_t\right)$

観測ノイズ $w_t \sim N\left(0, R_t\right)$

初期値 $x_0 \sim N\left(x_{0|0}, V_{0|0}\right)$

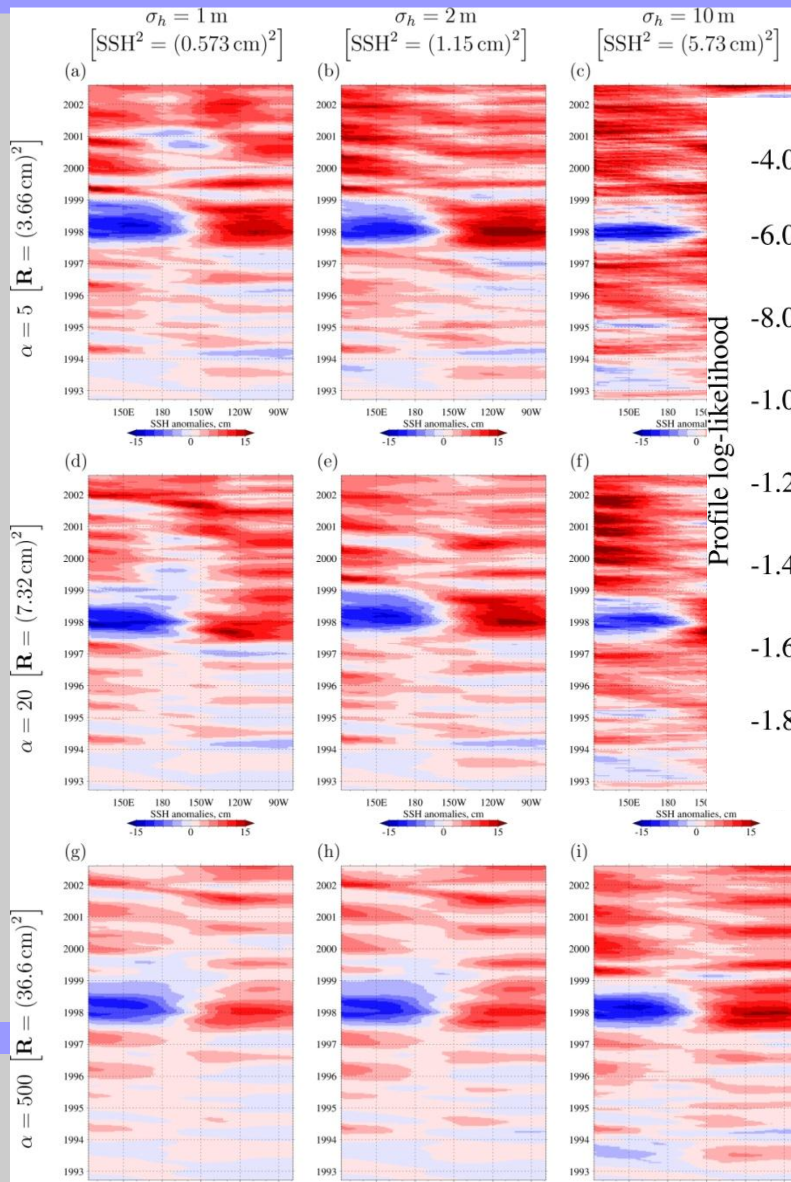
Q_t小 → モデルよりの状態推定値
R_t小 → データよりの状態推定値
$\mathrm{V}_{0|0}$小 → $\mathrm{x}_{0|0}$ よりの初期値

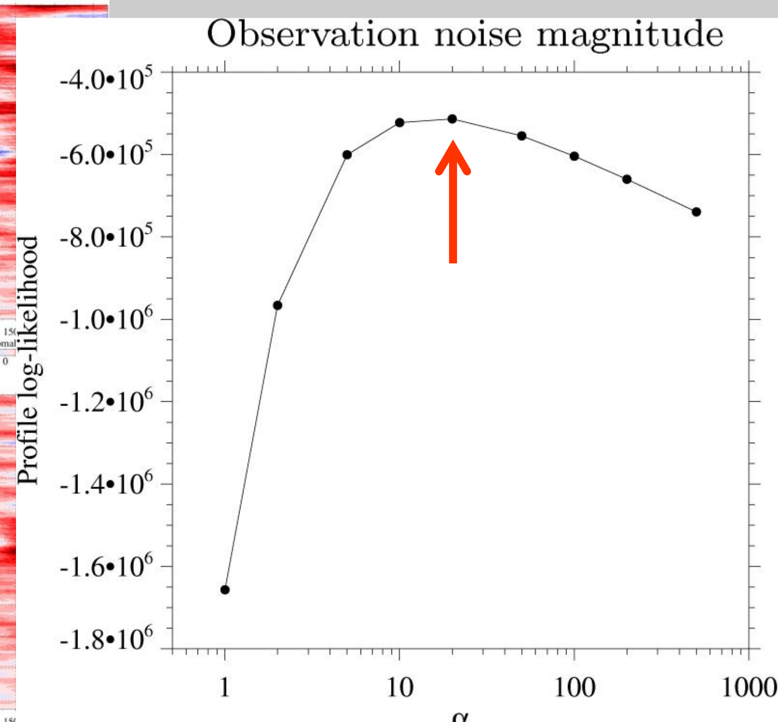
- $\theta = (\mathrm{V}_{0|0}, Q_{1:T}, R_{1:T})$ の最尤推定

データ同化結果

Q大
(σ_h大)

R大
(α大)

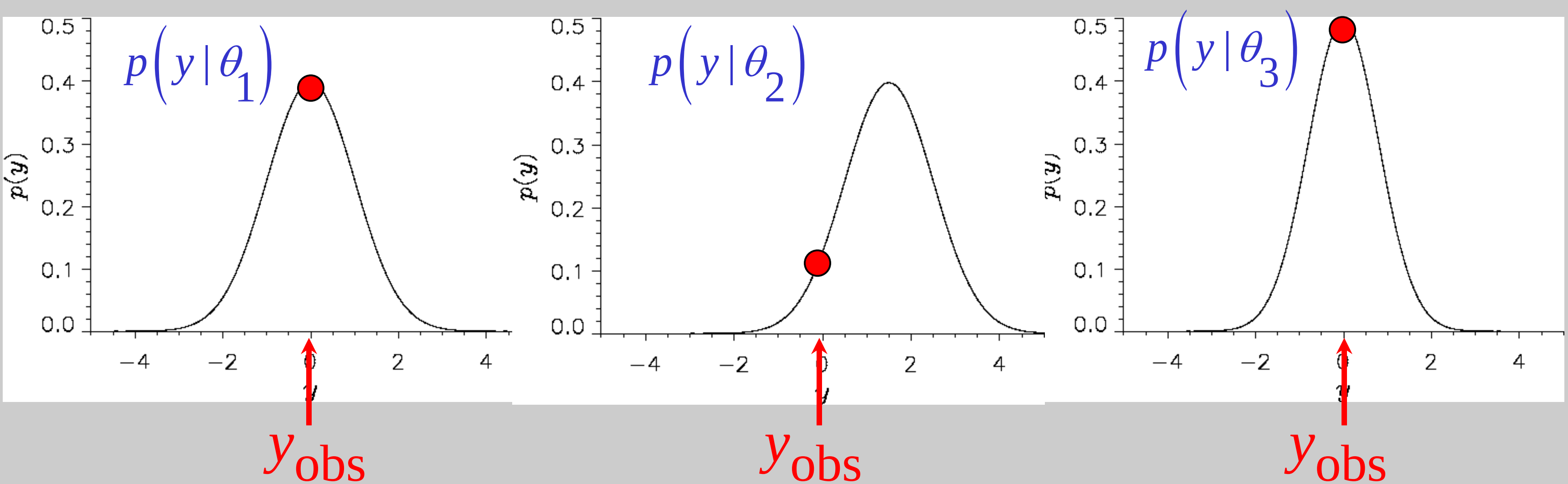




どれがいいのか？

尤度とは

与えられた観測値 y_{obs} が得られるような、
一番もっともらしい確率分布はどれか？



もっともらしさ = 尤度 $L(\tau) = p(y_{\mathrm{obs}}|\theta)$

上の例では θ_3 が一番もっともらしい。

時系列の尤度

$$
\begin{aligned}
L(\theta) &= p\left(y_1, y_2, \cdots, y_T \mid \theta\right) \\
&= p\left(y_1 \mid \theta\right) \cdot p\left(y_2 \mid y_1, \theta\right) \cdot p\left(y_3 \mid y_1, y_2, \theta\right) \cdots \cdot p\left(y_T \mid y_1, y_2, \cdots, y_{T-1}, \theta\right) \\
&= \prod_{t=1}^{T} p\left(y_t \mid y_{1:t-1}, \theta\right)
\end{aligned}
$$

この量を最大とするようなθを求める。
実際上は、対数を最大とするようにすると扱いやすい。

$$
\begin{aligned}
\ell(\theta) &\equiv \log p\left(y_{1:T} \mid \theta\right) \\
&= \sum_{t=1}^{T} \log p\left(y_t \mid y_{1:t-1}, \theta\right)
\end{aligned}
$$

時系列の尤度

$$\ell(\theta) = \sum_{t=1}^{T} \log p\left(y_t \mid y_{1:t-1}, \theta\right)$$

$$= \sum_{t=1}^{T} \log \int_{-\infty}^{\infty} \boxed{p\left(y_t \mid x_t, \theta\right)} \boxed{p\left(x_t \mid y_{1:t-1}, \theta\right)} dx_t$$

尤度　　　　予測分布

観測モデルから決まる

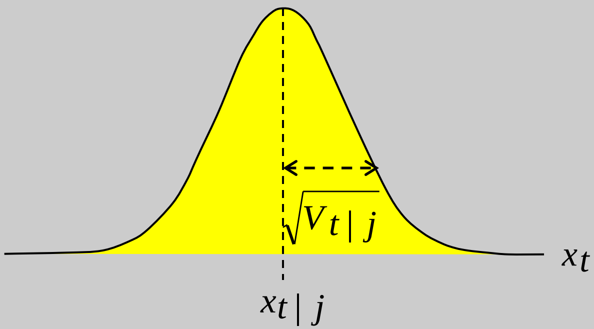
$$y_t = H_t x_t + w_t$$
$$w_t \sim N\left(0, R_t\right) \Rightarrow N\left(H_t x_t, R_t\right)$$

一般にガウス分布ではない
[力学モデルが非線型なので]

でもガウス
だと思うと $N\left(x_t^f, P_t^f\right)$

分布のアンサンブル近似

ガウス分布

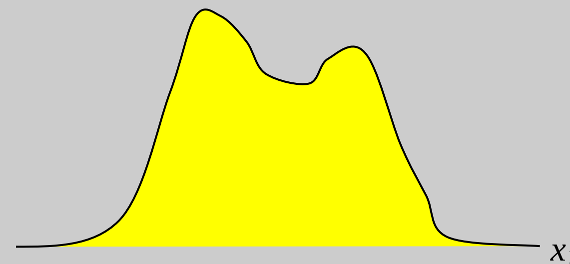


厳密に
表現可能

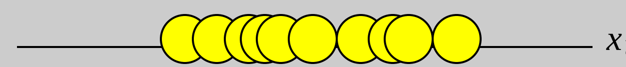
$$N\left(x_{t|j}, V_{t|j}\right)$$

カルマンフィルタ (KF)

非ガウス分布



アンサンブル
近似/粒子近似

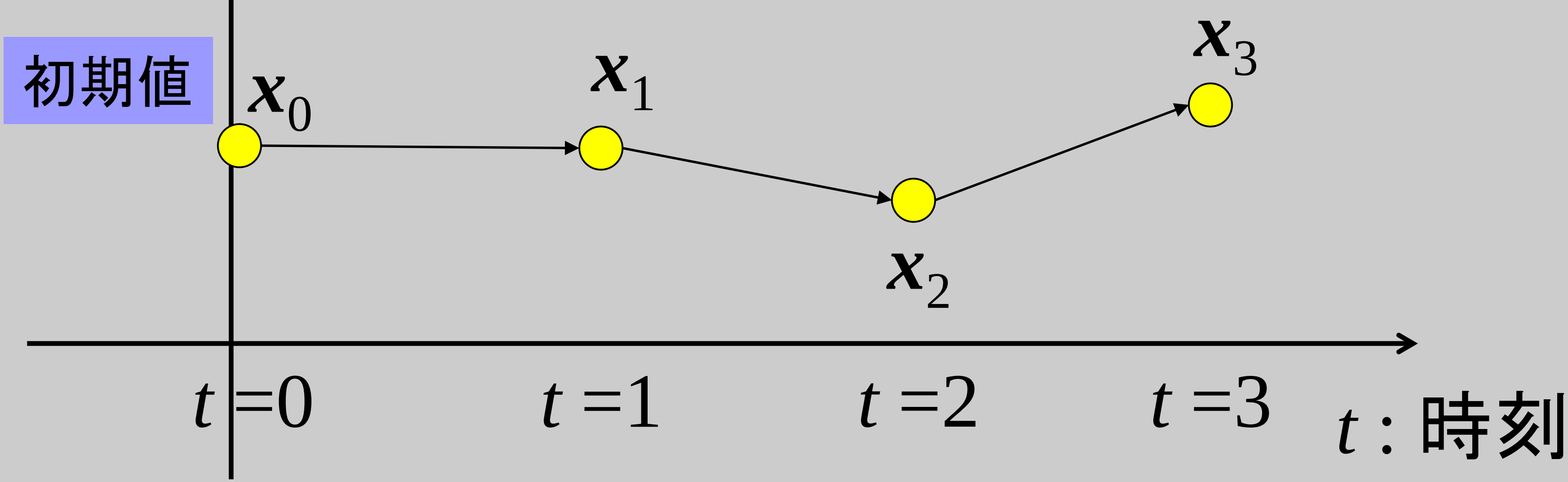


アンサンブルカルマンフィルタ (EnKF),
粒子フィルタ (PF)

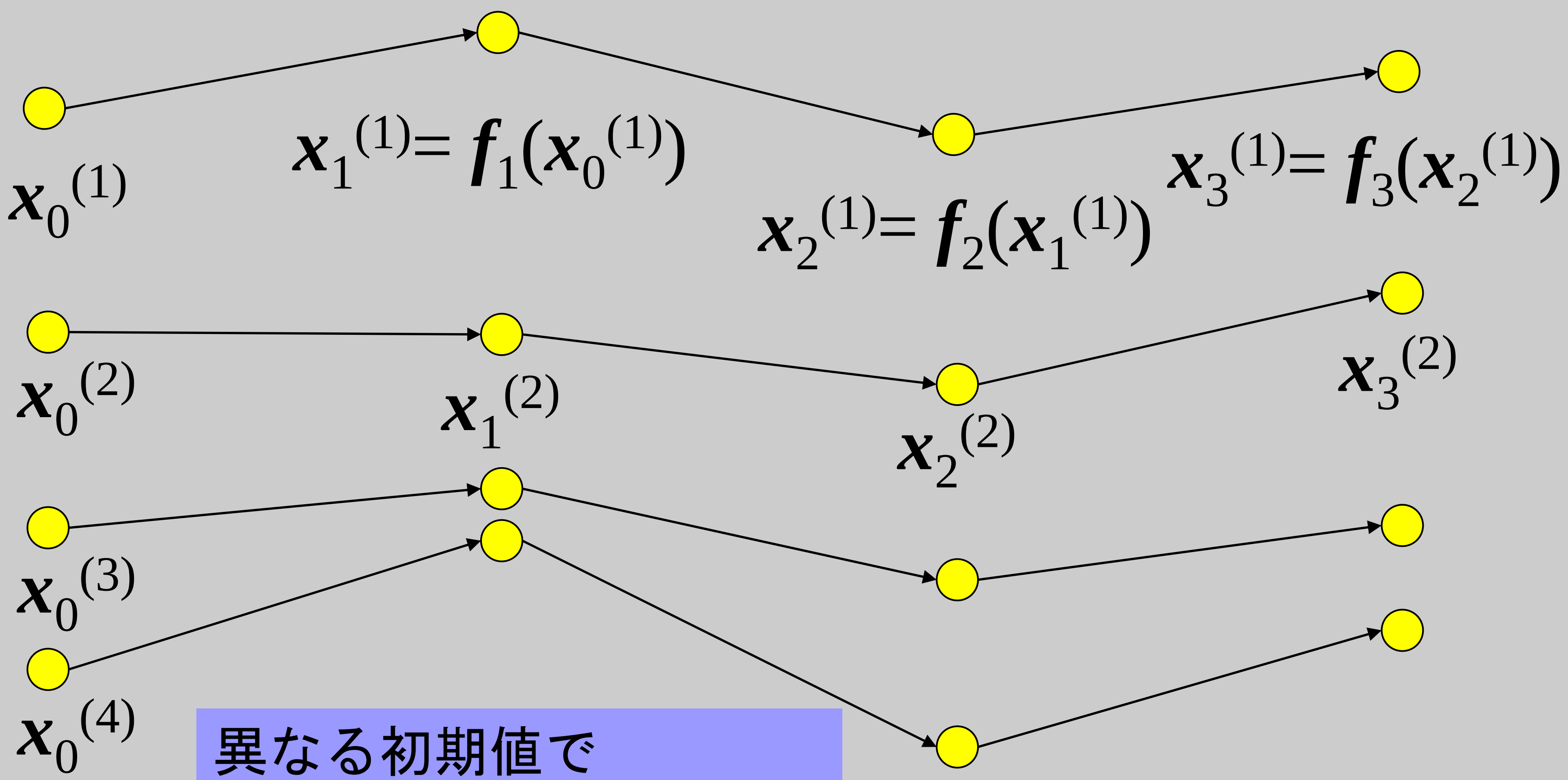
シミュレーションの抽象的表現

$\boldsymbol{x}$：状態　ある時点での全格子点上の全物理量をまとめて表すベクトル

$\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_{t-1})$　シミュレーションによる状態の時間発展



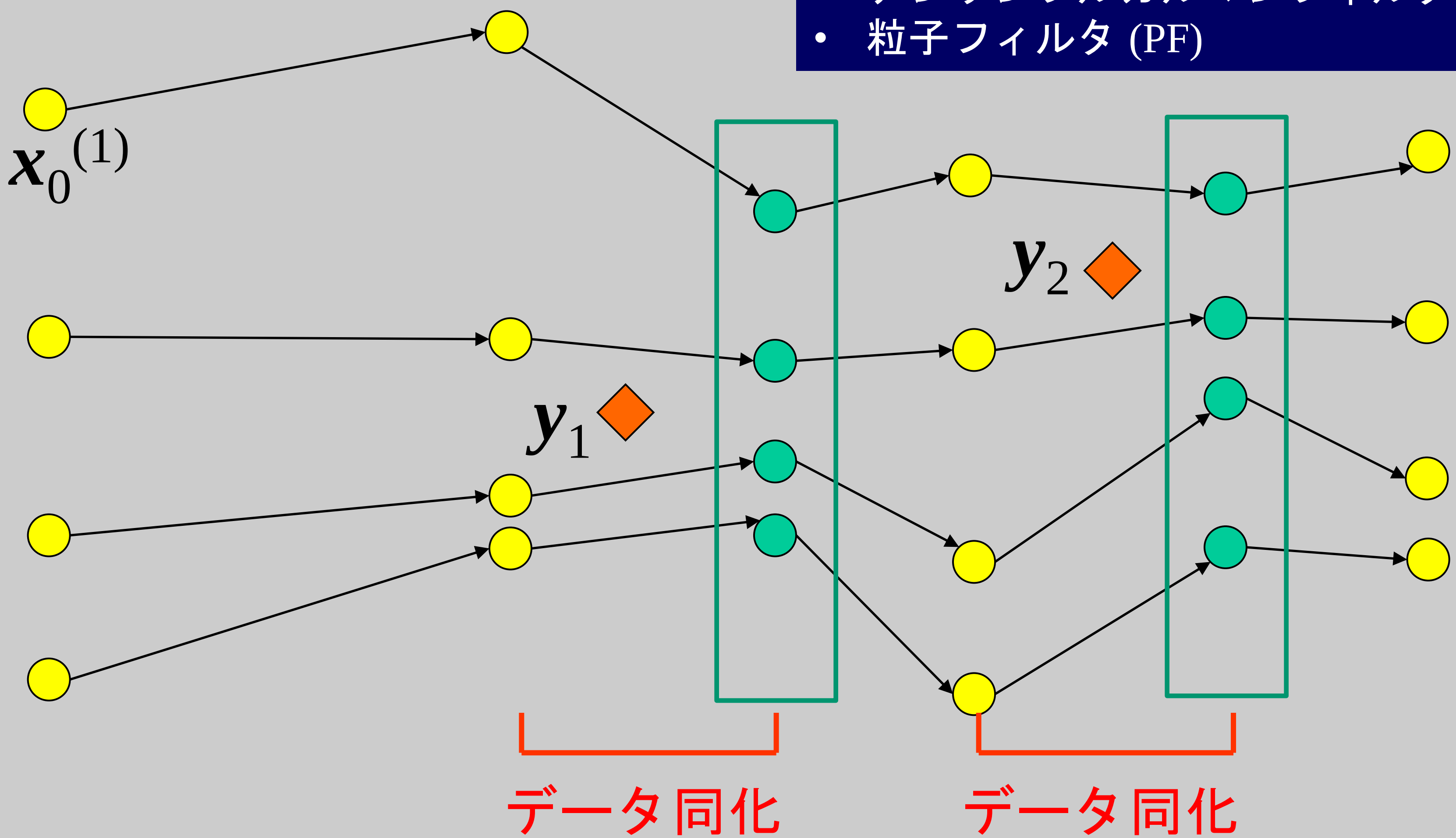
アンサンブル予測



異なる初期値で
シミュレーションを実施

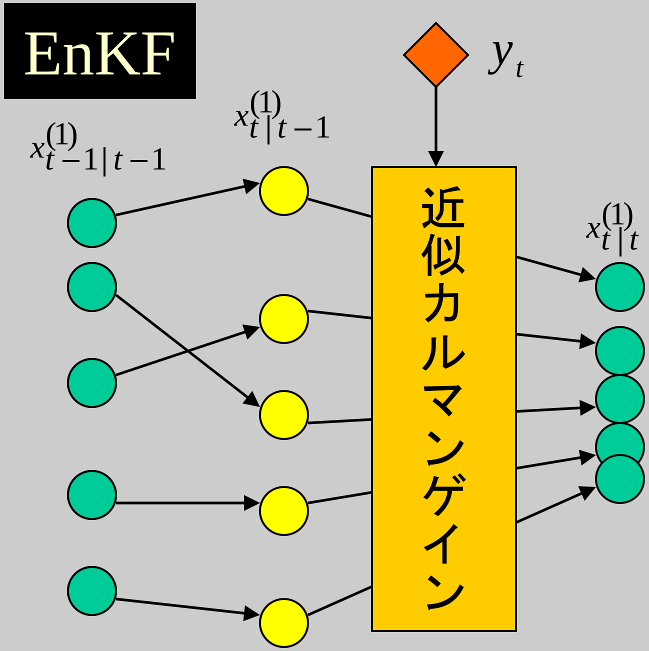
アンサンブル型データ同化

- アンサンブルカルマンフィルタ (EnKF),
- 粒子フィルタ (PF)



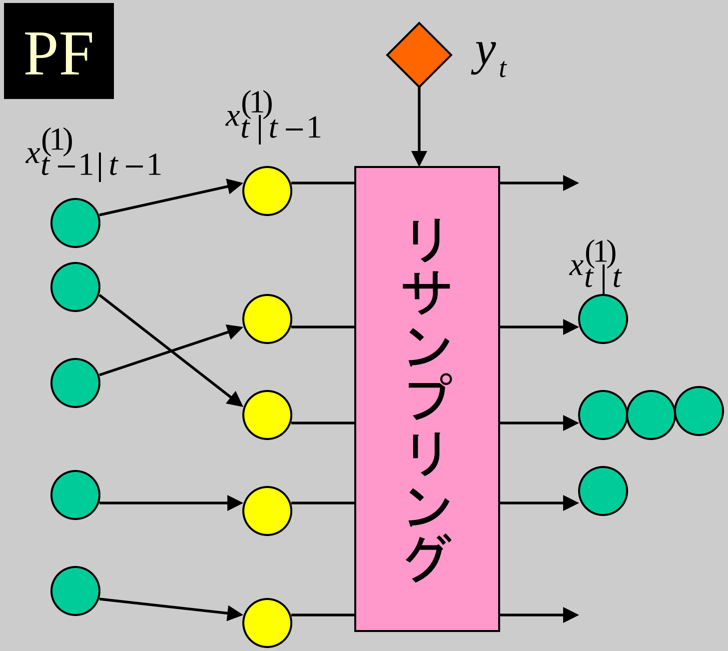
データに応じて、アンサンブルを改変

EnKF と PF

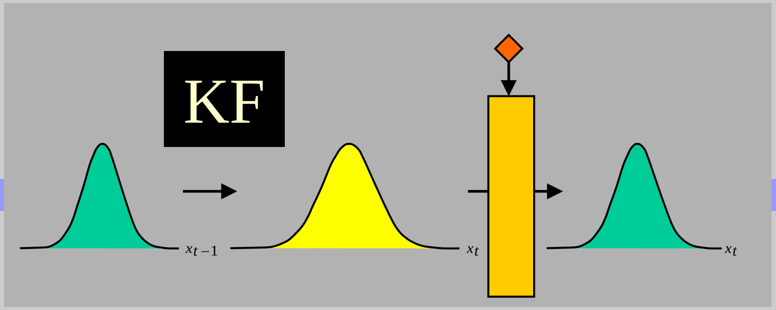


$$K_t = \bar{V}_{t|t-1} H_t \left(H_t \bar{V}_{t|t-1} H_t' + \bar{R}_{t|t-1} \right)^{-1}$$

$$x^{(n)}_{t|t} = x^{(n)}_{t|t-1} + \bar{K}_t \left(y_t + w^{(n)}_t - H_t x^{(n)}_{t|t-1} \right)$$



$$p\left(y_t \mid x^{(n)}_{t|t-1} \right)$$



問題意識

•状態推定の結果は、あらかじめ設定した θ によって、いかようにもできる (正則化パラメータと同様)

•$\theta = (V_{0|0}, Q_{1:T}, R_{1:T})$

$V_{0|0}$: 初期状態の分散共分散行列 (Bと書くことあり)

Q_t: システムノイズの分散共分散行列

R_t: 観測ノイズの分散共分散行列

•なるべくデータの情報を取り込む設定に興味がある

•グリッドサーチ（総当たり）、リサンプリング

✓ 計算が大変

✓ 探索範囲にどうしても限界がある (候補を自分で用意するため)

尤度のアンサンブル近似

尤度は各アンサンブルメンバー $x_{t|t\text{-}1}^{(n)}$ の尤度の平均で近似できる

$$y_t = h_t(x_t) + w_t$$

$$w_t \sim N\left(0, R_t\right)$$

$$
\begin{aligned}
&\ell(\theta) \\
&\equiv \log p\left(y_{1:T} \mid \theta\right) \\
&\approx \sum_{t=1}^{T} \left\{ -\frac{\dim y_t}{2} \log 2\pi - \frac{1}{2} \log \left| R_t \right| \right. \\
&\qquad \left. + \log \sum_{n=1}^{N} \exp \left[-\frac{1}{2} \left(y_t - h_t\left(x_{t\,|\,t-1}^{(n)} \right) \right)' R_t^{-1} \left(y_t - h_t\left(x_{t\,|\,t-1}^{(n)} \right) \right) \right] - \log N \right\}
\end{aligned}
$$

- この量 (対数尤度) を最大化するθ を求める。

観測誤差共分散行列の尤度

今回 $\theta = \left(B, Q_{1:T}, R_{1:T} \right)$ のうち、$R_{1:T}$ のみに着目

$$R_{1:T} \equiv \left\{ R_1, R_2, \cdots, R_T \right\}$$

対数尤度 $$\ell\left(R_1, R_2, \cdots, R_T \right) = \sum_{t=1}^{T} \ell_t \left(R_{1:t} \right)$$

但し

$$\ell_t\left(\boxed{R_{1:t}} \right) = -\frac{\dim y_t}{2} \log 2\pi - \frac{1}{2} \log \left| \boxed{R_t} \right| + \log \sum_{n=1}^{N} \exp\left[-\frac{1}{2} \left(y_t - h_t\left(\boxed{x_{t|t-1}^{(n)}} \right) \right)' \boxed{R_t^{-1}} \left(y_t - h_t\left(\boxed{x_{t|t-1}^{(n)}} \right) \right) \right] - \log N$$

$R_{1:t-1}, R_t$

R_t

$R_{1:t-1}$

推定の方針

$$\ell_t\left(\boxed{R_{1:t}}\right) = -\frac{\dim y_t}{2}\log 2\pi - \frac{1}{2}\log\left|\boxed{R_t}\right| + \log\sum_{n=1}^{N}\exp\left[-\frac{1}{2}\left(y_t - h_t\left(\boxed{x_{t|t-1}^{(n)}}\right)\right)'\boxed{R_t^{-1}}\left(y_t - h_t\left(\boxed{x_{t|t-1}^{(n)}}\right)\right)\right] - \log N$$

$R_{1:t-1}, R_t$

R_t

$R_{1:t-1}$

- $\ell_t(R_{1:t})$ は R_t の推定のためだけに用いる
- 本来は $R_{1:t-1}$ もパラメータだが、これらは過去の対数尤度 $\ell_1(R_1)$, …, $\ell_{t-1}(R_{1:t-1})$ で推定され、現時点では固定とする

- 各時点tで、 R_tが推定される

- 各時点で y_t に応じてR_t を推定するので、ばたばたするかも
- 時点に依存しない一定の R を推定することは できない

- これで最尤推定ができる。のだが、

観測誤差共分散行列のベイズ推定

•動機その１：
最尤推定では、R_t がスカラー倍か対角の場合でしかうまくいかない。本当は非対角（理想は全成分）R_t の推定がしたい。

•動機その２：
R_t が激しく時間変化するというのはいかがなものか

•R_t に事前分布を与えたベイズ推定へ

ベイズ推定

尤度 (今まで)

$$p\left(y_t \mid R_t, y_{1:t-1}\right) □ \sum_{n=1}^{N} \frac{1}{N} \phi\left(y_t; h_t\left(x_{t|t-1}^{(n)}\right), R_t\right)$$

事後分布

$$p\left(R_t \mid y_{1:t}\right) \propto p\left(y_t \mid R_t, y_{1:t-1}\right) p\left(R_t \mid y_{1:t-1}\right)$$

$$□ \sum_{n=1}^{N} \frac{1}{N} \phi\left(y_t; h_t\left(x_{t|t-1}^{(n)}\right), R_t\right) \boxed{p\left(R_t \mid y_{1:t-1}\right)}$$

尤度 **事前分布**

EMアルゴリズム

- 対数事後分布の最適性条件

$$\frac{\partial \log p\left(R_t \mid y_{1:t}, S, \nu\right)}{\partial R_t} = 0 \quad \Rightarrow$$

E ステップ

$$z_n^{(k)} = \frac{\phi\left(y_t; h_t\left(x_{t|t-1}^{(n)}\right), R_t^{(k)}\right)}{\sum_{n=1}^{N} \phi\left(y_t; h_t\left(x_{t|t-1}^{(n)}\right), R_t^{(k)}\right)}$$

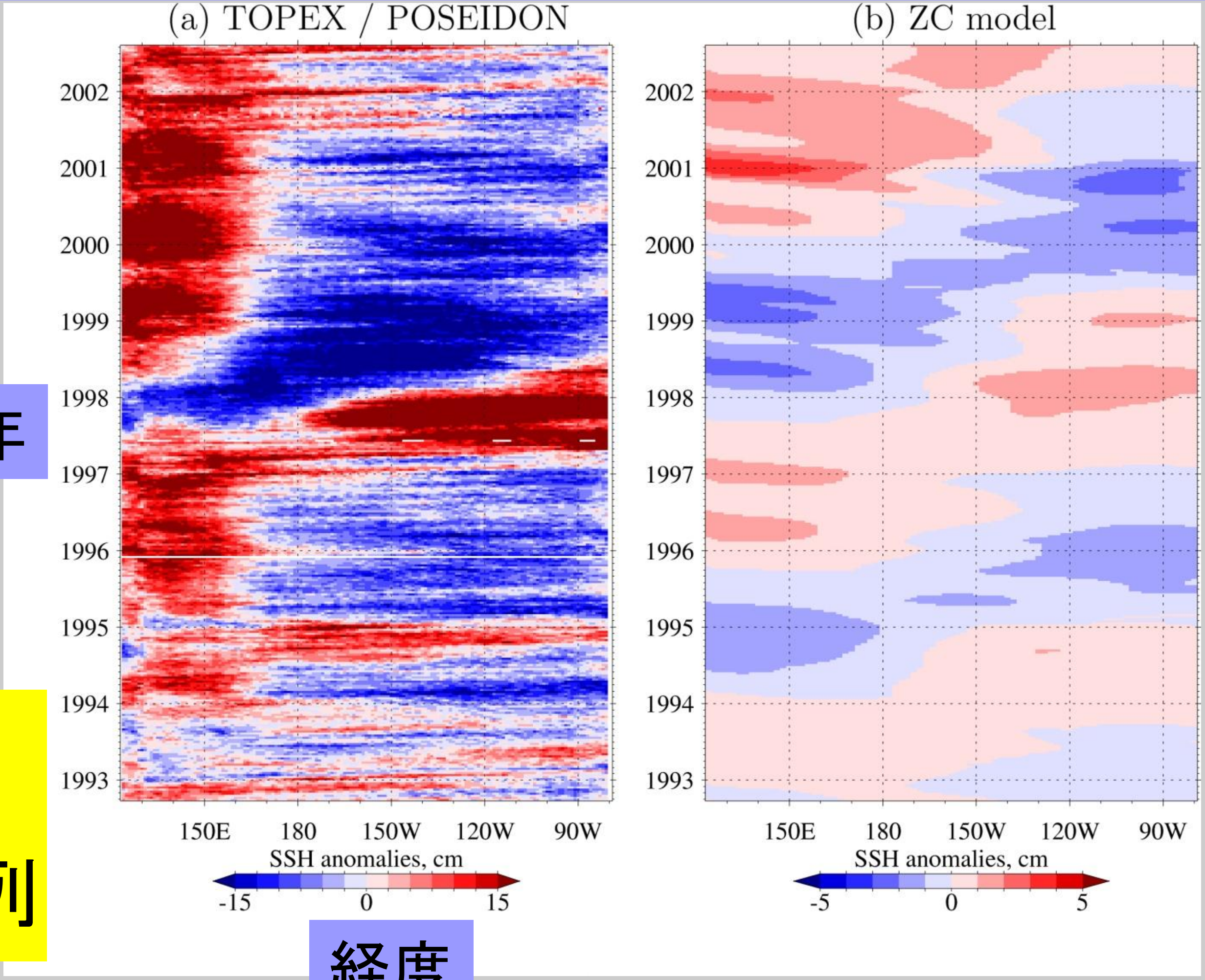
M ステップ

$$R_t^{(k+1)} = \boxed{(1-\lambda)} \sum_{n=1}^{N} z_n^{(k)} \left(y_t - h_t\left(x_{t|t-1}^{(n)}\right)\right)\left(y_t - h_t\left(x_{t|t-1}^{(n)}\right)\right)' + \boxed{\lambda} S$$

where $\boxed{\lambda = \frac{\nu - m - 1}{\nu - m}}$

標本共分散と事前共分散の線型和

データとシミュレーション



年

経度

$$\dim y_t = 503$$

$R_t : 503 \times 503$行列

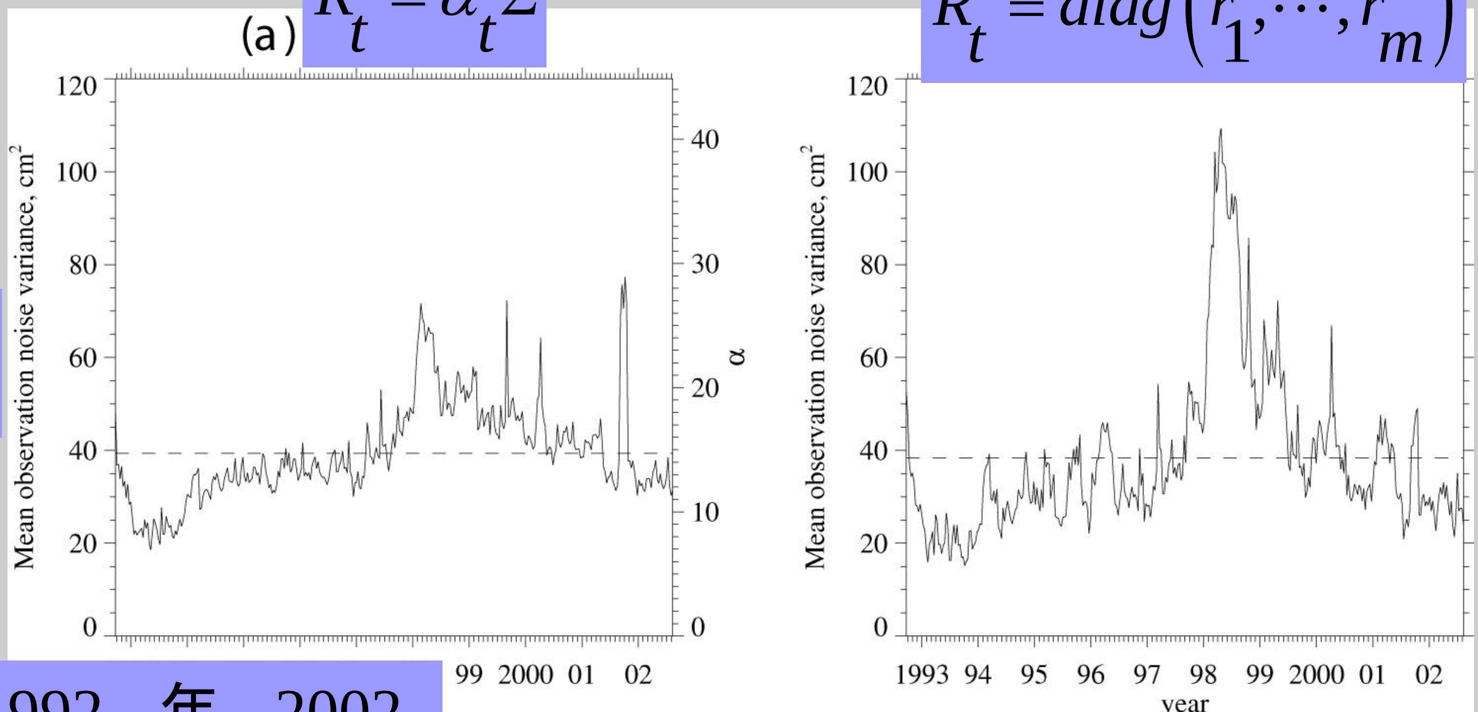
色は海面高度の変動を示している

最尤推定値 [いかがなものか]

$R_t = \alpha_t \Sigma$

$R_t = diag\left(r_1, \cdots, r_m\right)$

分散



1992- 年 -2002

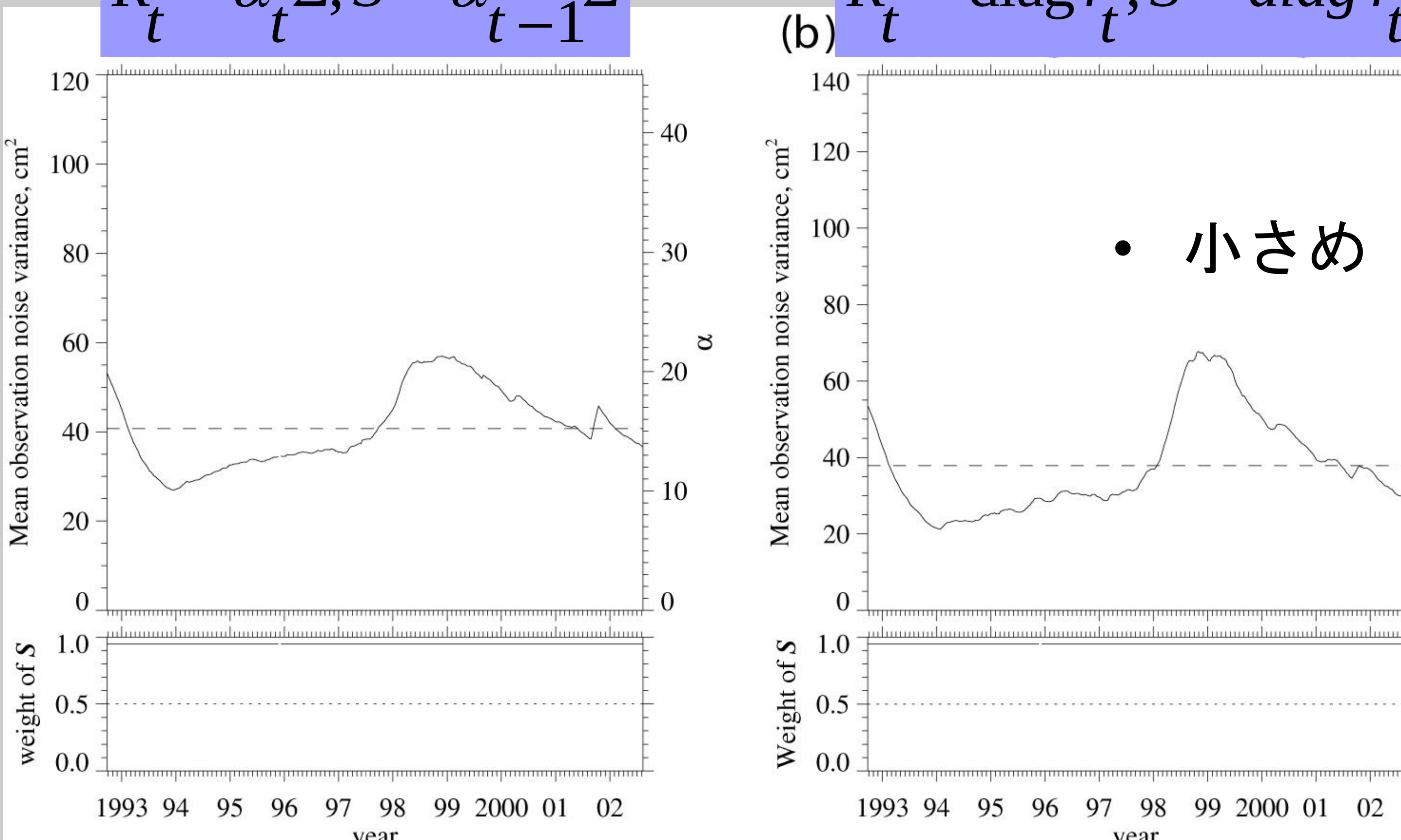
- いずれも、1998年あたりで大
- 平均値はいずれも40cm^2
- α_t Σは振幅小、diag r_t は振幅大

ベイズ推定値 (事前重み 0.95)

$R_t = \alpha_t \Sigma, S = \hat{\alpha}_{t-1} \Sigma$

$R_t = \mathrm{diag}\, r_t, S = diag\, \hat{r}_{t-1}$

分散



- 小さめ

1992- 年 -2002

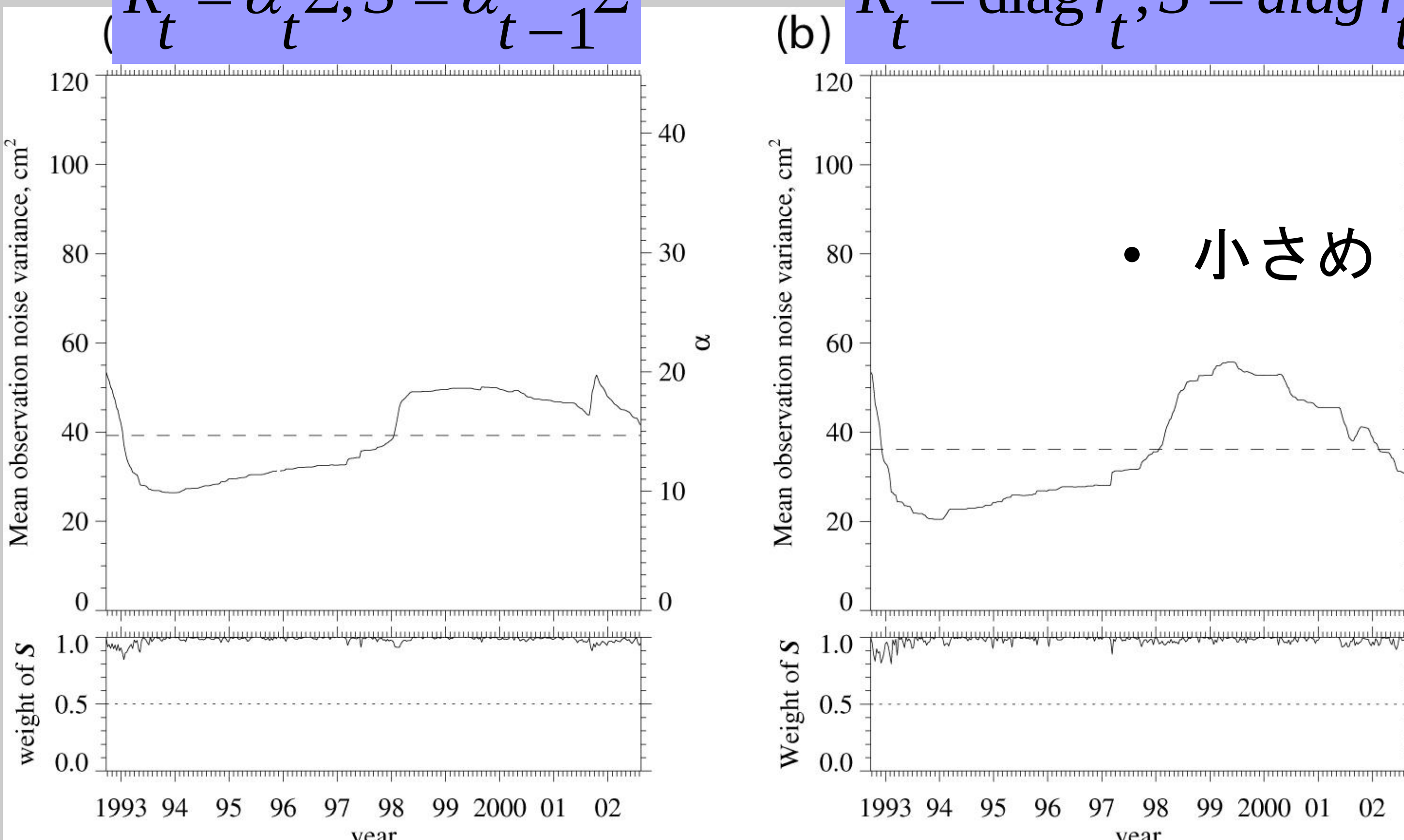
- ばたばた感は解消
- 事前情報に重み0.95

ベイズ推定値 (事前重み推定)

$R_t = \alpha_t \Sigma, S = \hat{\alpha}_{t-1} \Sigma$

$R_t = \mathrm{diag}\, r_t, S = diag\, \hat{r}_{t-1}$

分散



- 小さめ

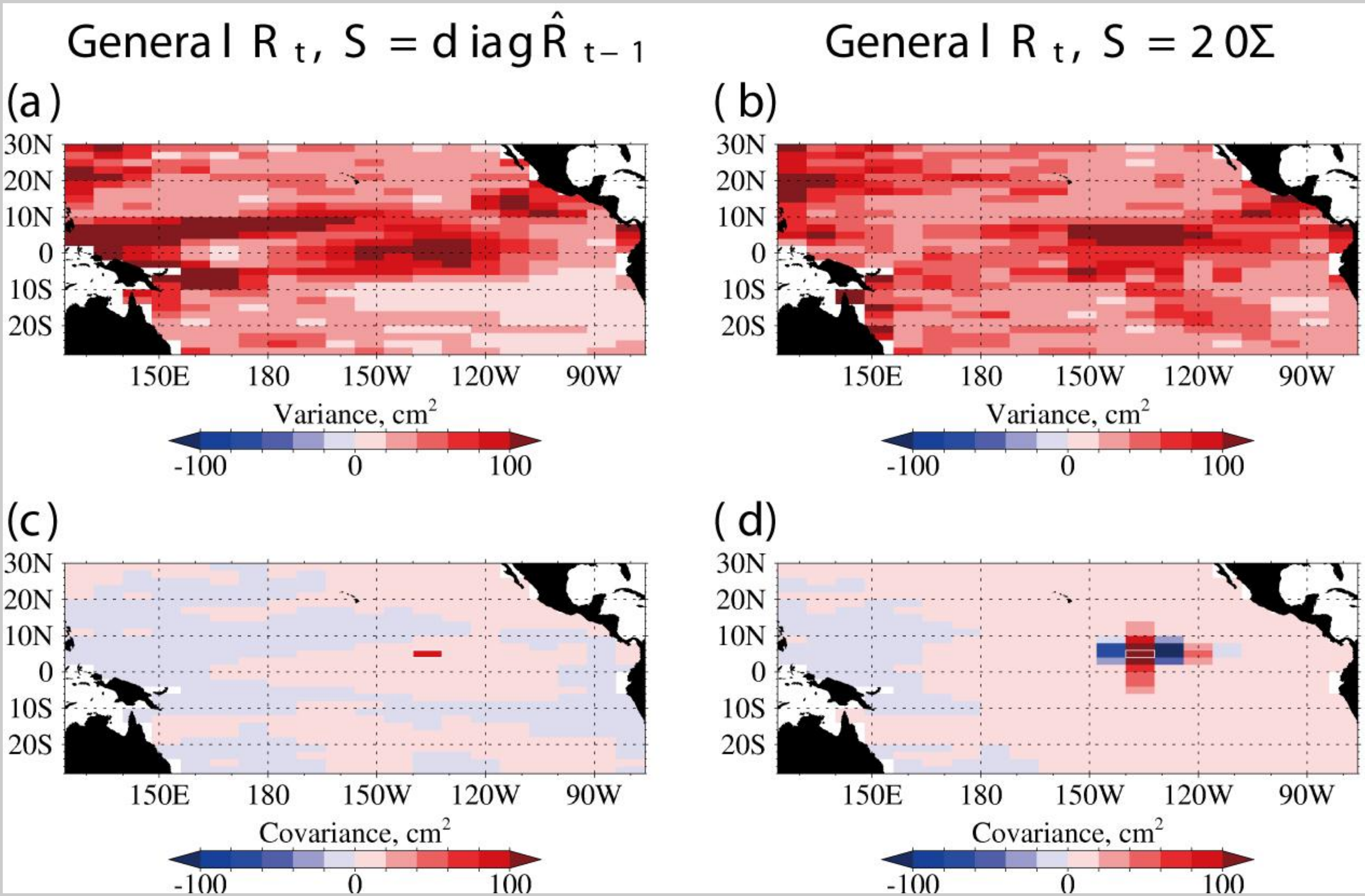
1992- 年 -2002

- ばたばた感は解消
- 事前情報に重み

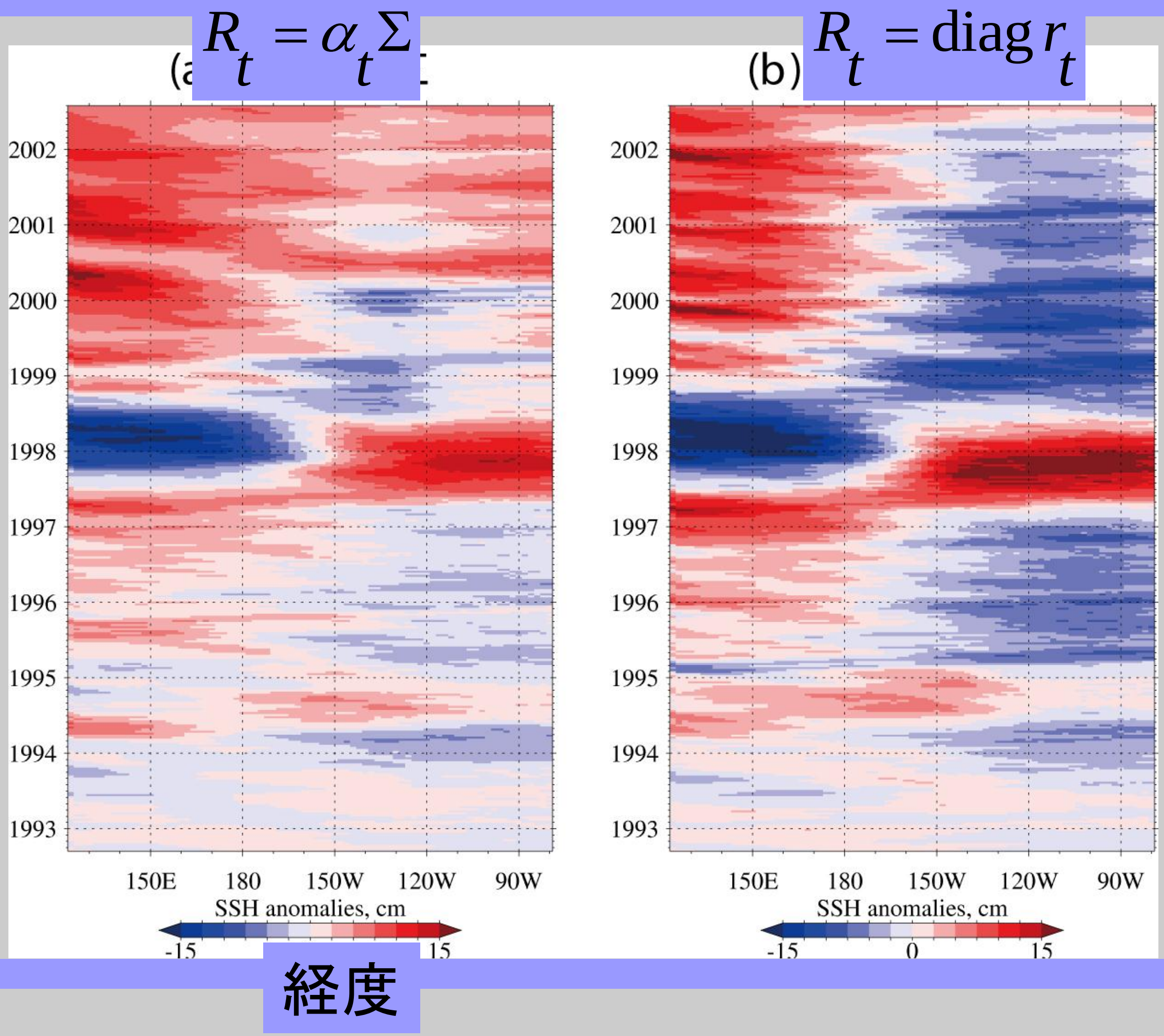
ベイズ推定値 (全要素推定)

分散

共分散



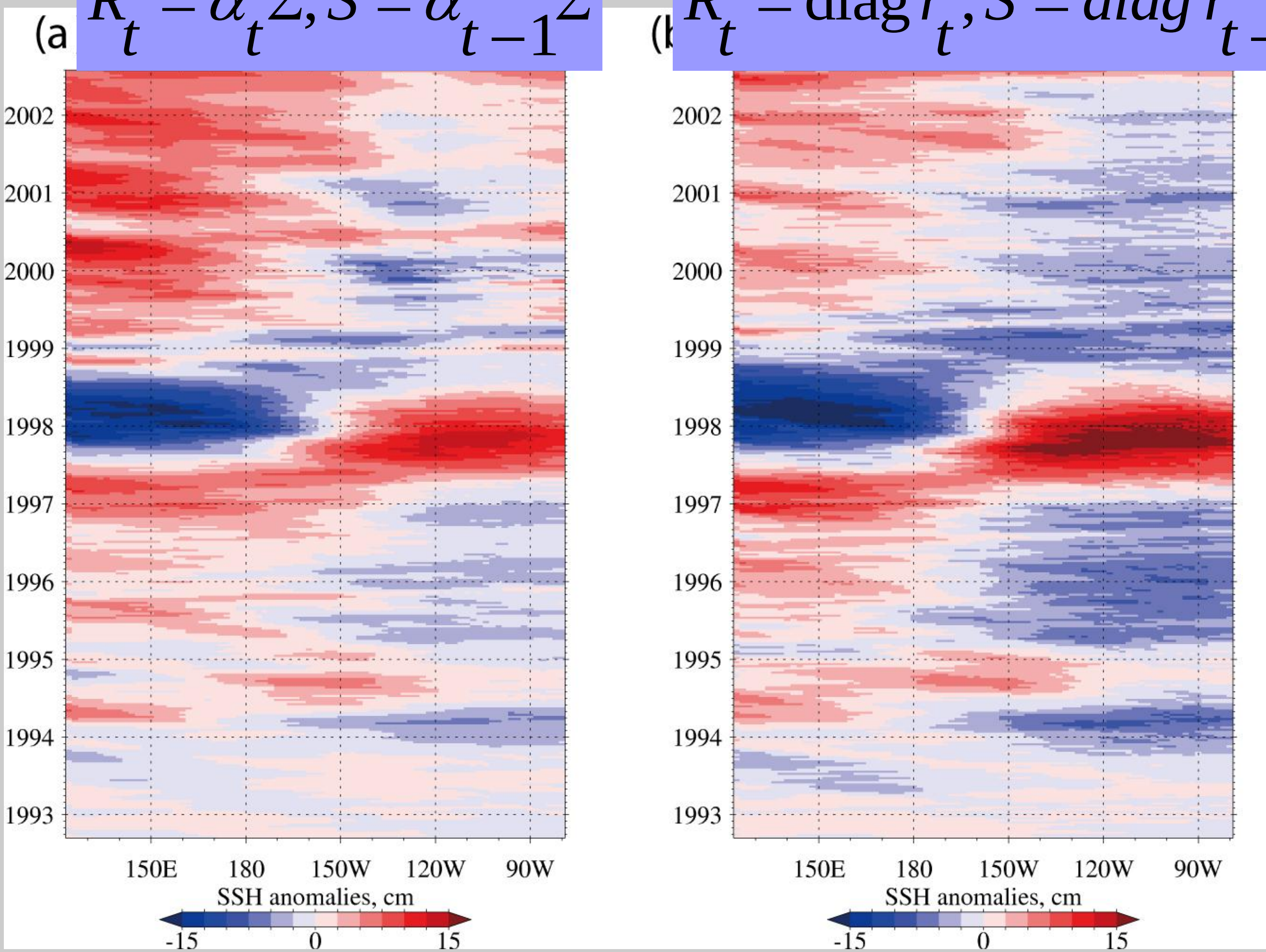
平滑化推定値 (Rt: 最尤推定)



平滑化推定値 (Rt: ベイズ推定)

$R_t = \alpha_t \Sigma, S = \hat{\alpha}_{t-1} \Sigma$

$R_t = \mathrm{diag}\, r_t, S = diag\, \hat{r}_{t-1}$



- ばたばた感が解消した分、データへの追従も抑制気味

まとめ

- 観測ノイズの共分散行列 Rt のベイズ推定

- 全成分を推定することが可能
- 時間変化を滑らかにすることが可能

- EMアルゴリズムを構成 [最尤推定の場合と同様]
- 事前分布のパラメータは直前推定値、経験値、周辺尤度などから
- エルニーニョモデルを用いた数値実験
- 必要な反復回数は5,6回程度