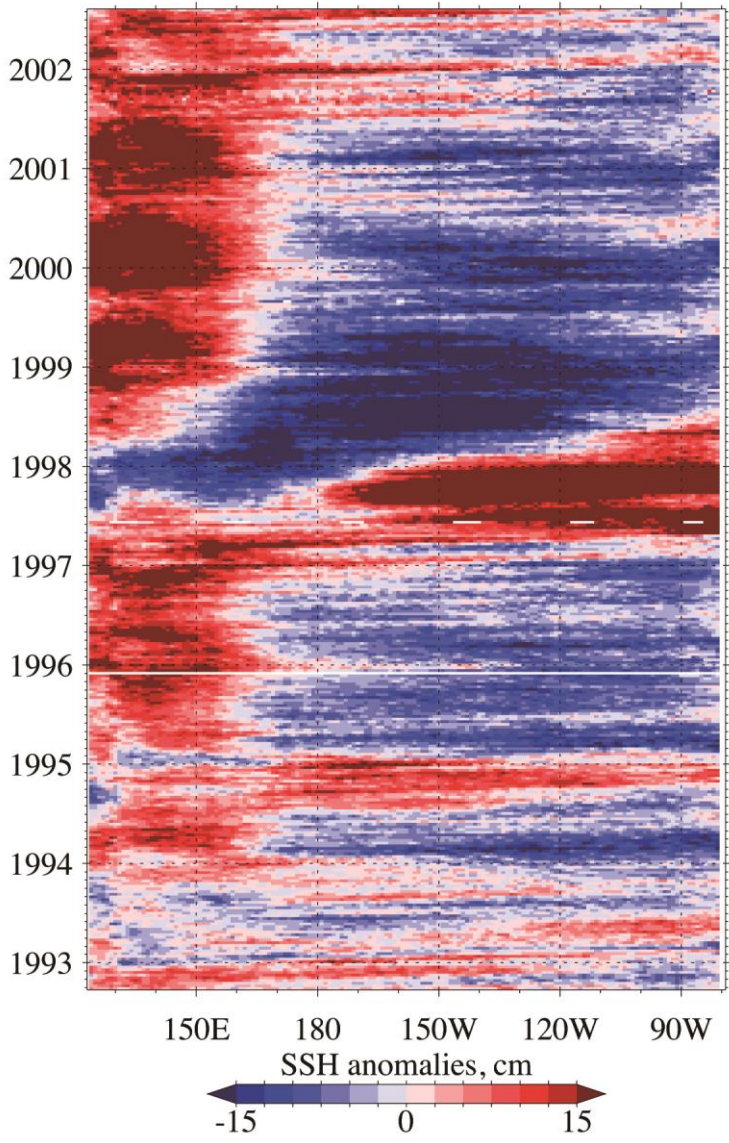


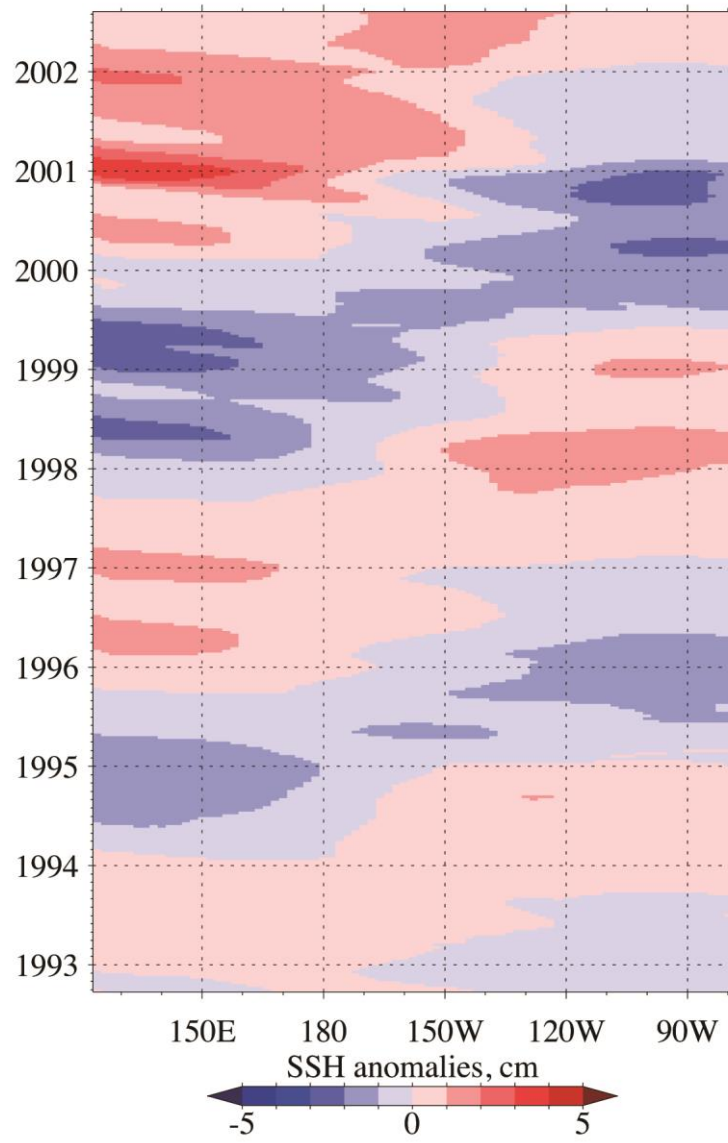
# データ同化と観測誤差共分散行列推定

上野 玄太      モデリング研究系 准教授

観測データ



シミュレーション



## 【データ同化の胆】

データ同化は、観測データをシミュレーションに織り交ぜて現実的な推定値を得る作業であり、時系列解析における状態空間モデルの状態推定にあたる。得られる状態推定値は、あらかじめ設定する初期状態や誤差項の統計的性質に大きく依存する。そのため、「最適」「妥当」な統計的性質を設定したい。

## 状態空間モデル

$$\begin{aligned} \mathbf{x}_t &= \mathbf{f}_t(\mathbf{x}_{t-1}) + \mathbf{G}_t \mathbf{v}_t \\ \mathbf{y}_t &= \mathbf{h}_t(\mathbf{x}_t) + \mathbf{w}_t \end{aligned}$$

## 初期状態、システム誤差、観測誤差の仮定

$$\begin{aligned} \mathbf{x}_1 &\sim N(\mathbf{x}_b, \mathbf{B}), \\ \mathbf{v}_t &\sim N(\mathbf{0}, \mathbf{Q}_t), \\ \mathbf{w}_t &\sim N(\mathbf{0}, \mathbf{R}_t) \end{aligned}$$

観測誤差  
共分散行列

## 【問題意識】

なるべくデータの情報を取り込む設定にしたい。ただし、過適合して予測性能を落とすことは避けたい。ここでは観測誤差共分散行列 $\mathbf{R}$ の全要素を「最適化」する。単純ではあるが確実な方法として、グリッドサーチ（総当たり法）やリサンプリングがあるが、以下が問題である。

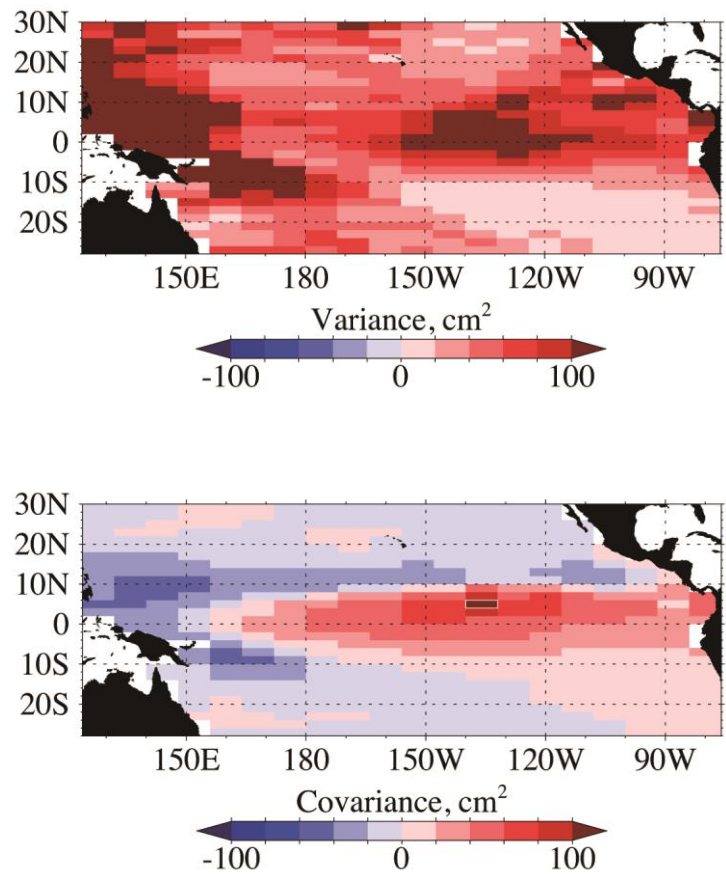
- (1) 計算が大変（行列の要素数は観測地点数の2乗に比例するため）
- (2) 探索範囲にどうしても限界がある（行列の候補を自分で用意する必要があるため）

そこで、ベイズ法による $\mathbf{R}$ 行列の推定法を開発した。

## 【理想実験による観測誤差共分散行列 $\mathbf{R}$ の推定】

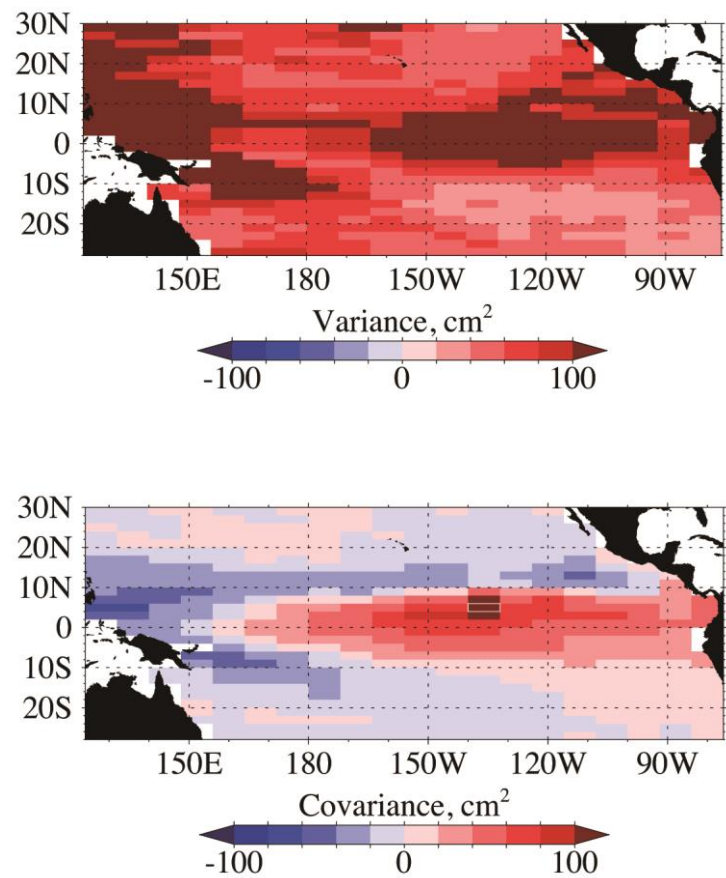
真の観測誤差共分散行列 $\mathbf{R}$ が既知のもとで人工データを生成し、人工データから真の $\mathbf{R}$ 行列をベイズ推定する。 $\mathbf{R}$ 行列の事前分布として、行列 $\mathbf{S}$ をモードに持つ確率分布を仮定し、いくつかの $\mathbf{S}$ の設定値で $\mathbf{R}$ の推定を行った。その結果、 $\mathbf{S}$ として前タイムステップの $\mathbf{R}$ の推定値を正則化したもの（対角成分に小さい数を加える方法、相関関数を用いたテーパーリングによる方法）により、真の $\mathbf{R}$ と近い推定値が得られることがわかる。

真の $\mathbf{R}$



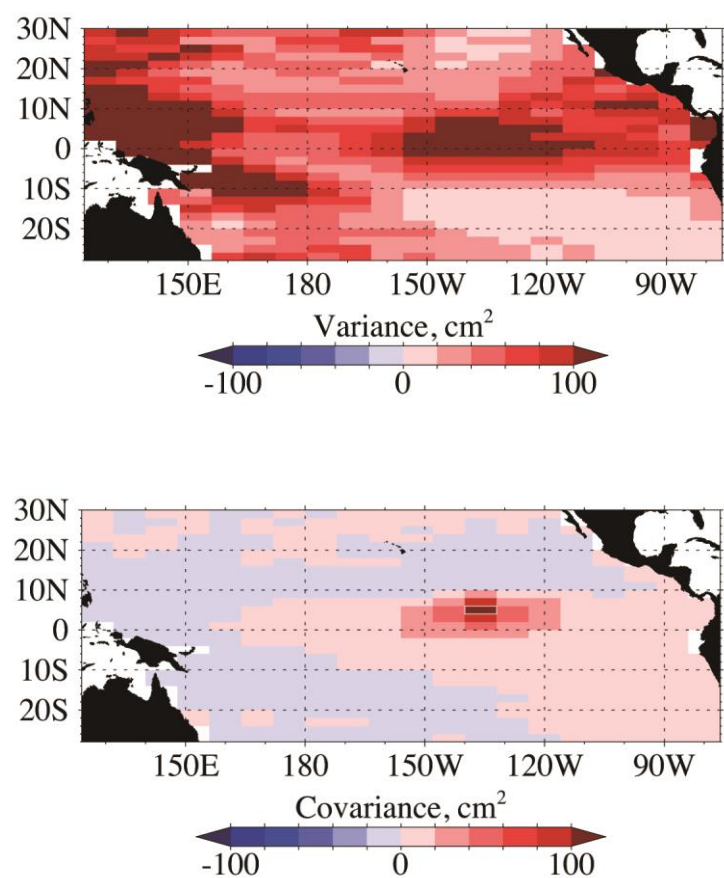
推定した $\mathbf{R}$

$$(\mathbf{S} = \hat{\mathbf{R}}_{t-1} + \sigma^2 \mathbf{I})$$



推定した $\mathbf{R}$

$$(\mathbf{S} = \rho \circ \hat{\mathbf{R}}_{t-1})$$

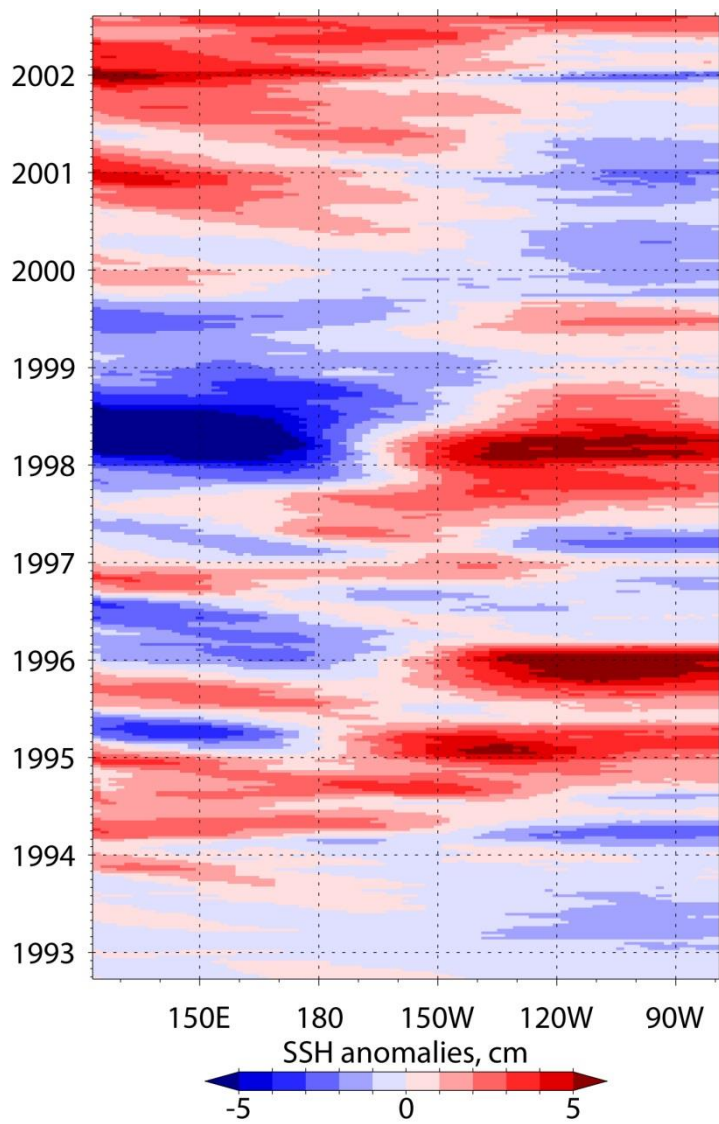


$$\frac{1}{T} \sum_{t=1}^T D(\hat{\mathbf{R}}_t \| \mathbf{R}_t^{\text{true}})$$

$\mathbf{S} = \rho \circ \hat{\mathbf{R}}_{t-1}, c = (76^\circ, 28^\circ)$	227.30
$\mathbf{S} = \rho \circ \hat{\mathbf{R}}_{t-1}, c = (38^\circ, 14^\circ)$	429.62
$\mathbf{S} = \hat{\mathbf{R}}_{t-1} + \sigma^2 \mathbf{I} \times 10^{-1}$	503.95
$\mathbf{S} = \hat{\mathbf{R}}_{t-1} + \sigma^2 \mathbf{I}$	703.78
$\mathbf{S} = \hat{\mathbf{R}}_{t-1} + \sigma^2 \mathbf{I} \times 10^{-2}$	597273.16
$\mathbf{S} = \rho \circ \hat{\mathbf{R}}_{t-1}, c = (152^\circ, 56^\circ)$	4490246.34

## 【 $\mathbf{R}$ の推定値と真の $\mathbf{R}$ との距離】

## データ同化（フィルタ推定値）



## 【まとめ】

ベイズ法による観測誤差共分散行列 $\mathbf{R}$ 行列の推定法を開発した。真の $\mathbf{R}$ 行列を既知とする人工データを用いた理想実験では妥当な推定値が得られることがわかる。

一方で、実データに対しては、シミュレーション（システムモデル）による再現能力の限界を意識せざるを得ない。よりよい状態推定値を得るためには、システムモデルの改良を進める必要がある。

## データ同化（フィルタ推定値） [左図とはスケールが異なる]

