

海洋物理における データ同化システムの発展と 今後の展望

石川洋一 (海洋研究開発機構・地球情報研究センター)

はじめに

- 「データ同化」は地球科学において、一般的なものとなってきている
- さらに、データ同化はこれまで使われてこなかった分野にも広がっており、ますます重要な技術となっている
- 惑星科学分野では観測データの量が問題となる
 - 海洋分野では気象分野に比べ、データが少ない中でデータ同化研究が進められてきた

データ同化とは

- 観測データと数値モデルを組み合わせてデータセットを作成すること

観測データからみると...

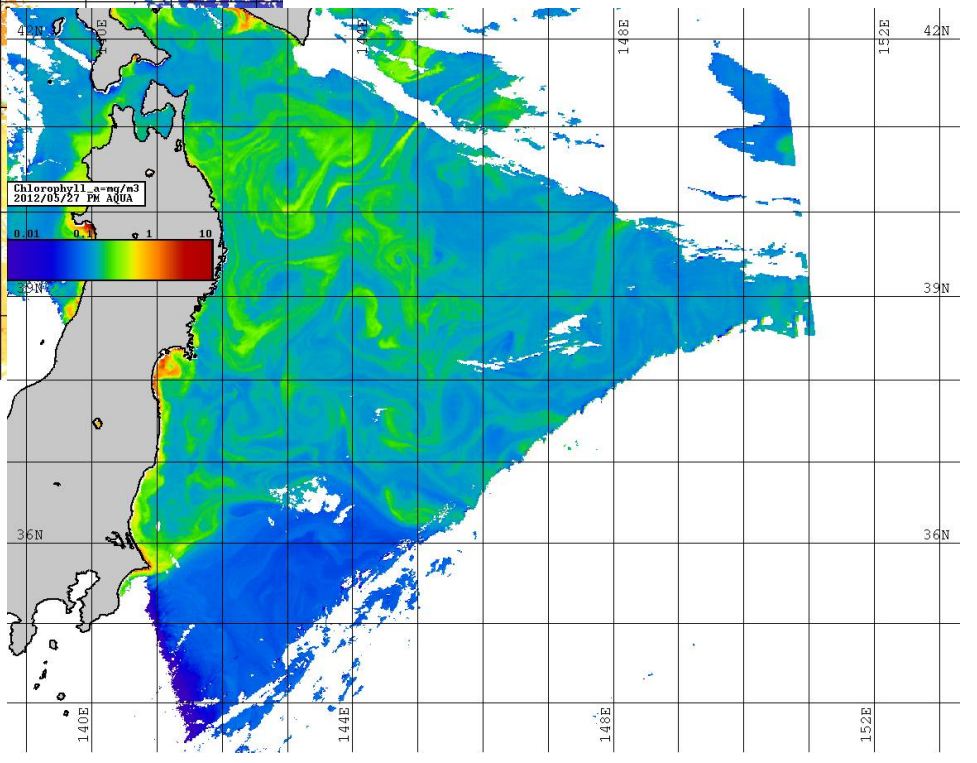
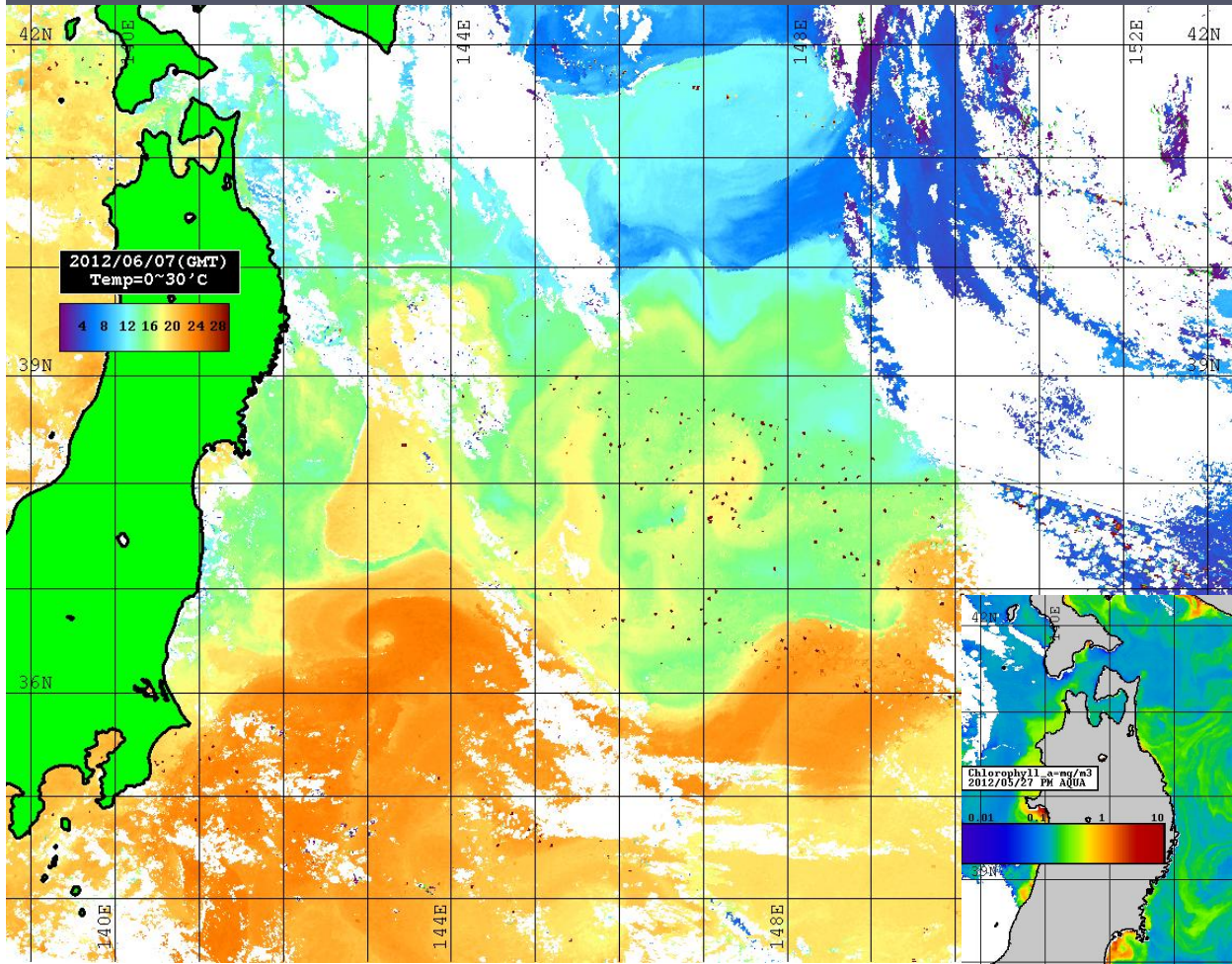
数値モデルを利用した4次元的な内挿／外挿／グリッド化
異なる種類の観測データの統合

数値モデルからみると...

モデルの入力パラメータ(初期条件、境界条件など)の
チューニングの自動化

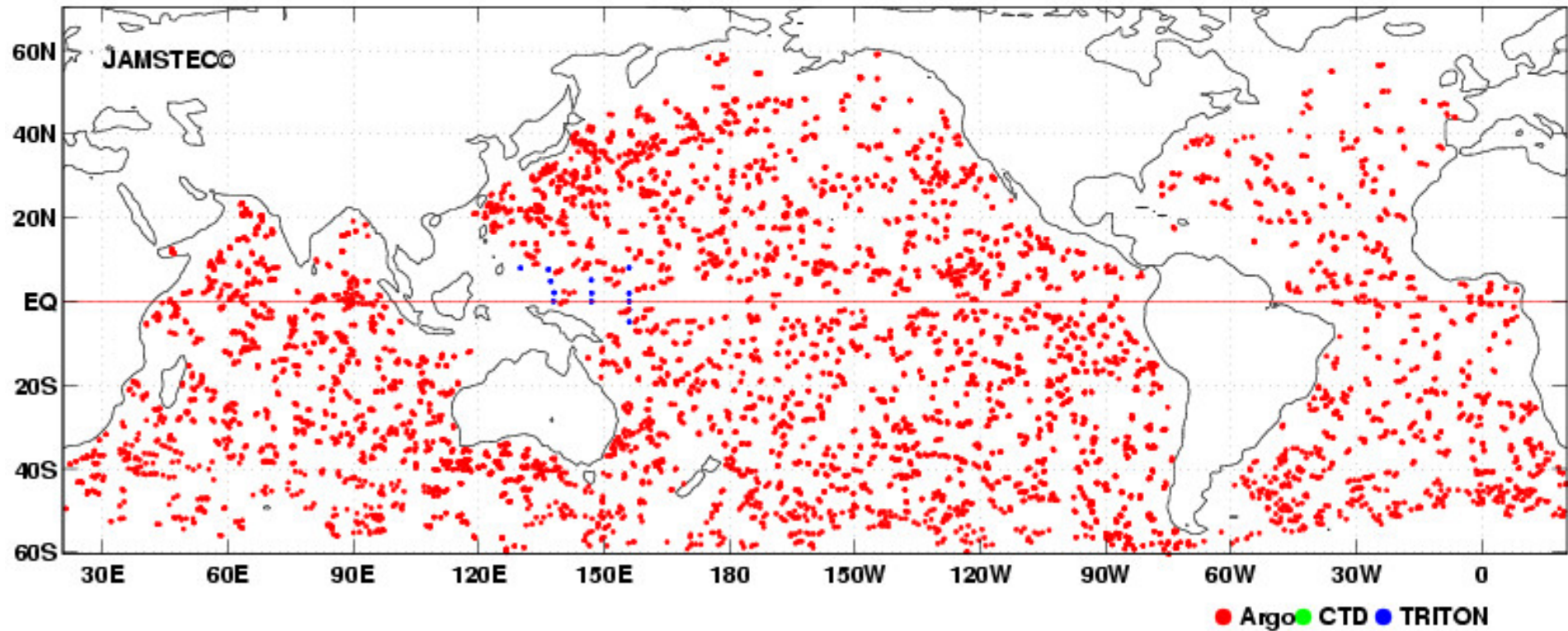
海洋観測データ

- 人工衛星リモートセンシングデータ
 - 海面水温、海面高度、海色など
 - 面的にカバーされ、最も多いが海面の情報のみ
- 現場観測データ
 - 水温、塩分プロファイル
 - 船舶からの観測
 - **ARGO**フロート:自動的に昇降して水温、塩分プロファイルを観測し、衛星通信によってデータを送る

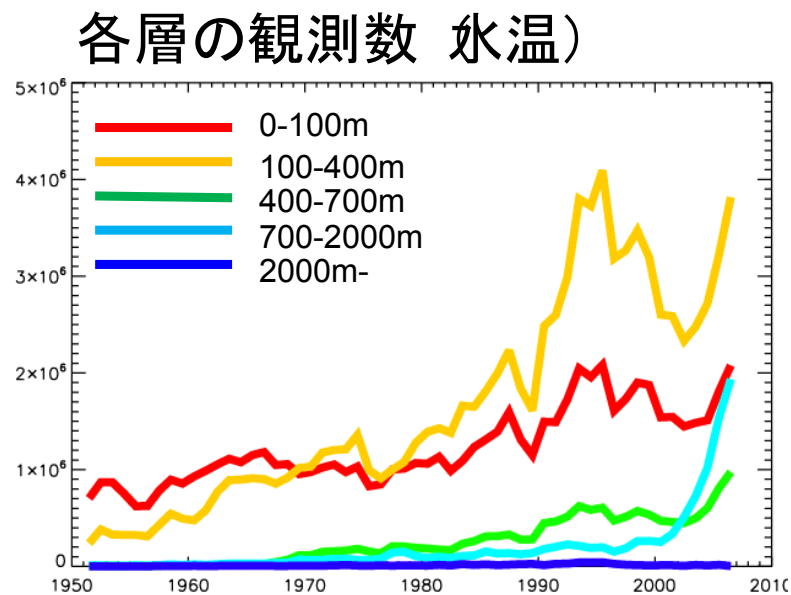


海洋観測データ

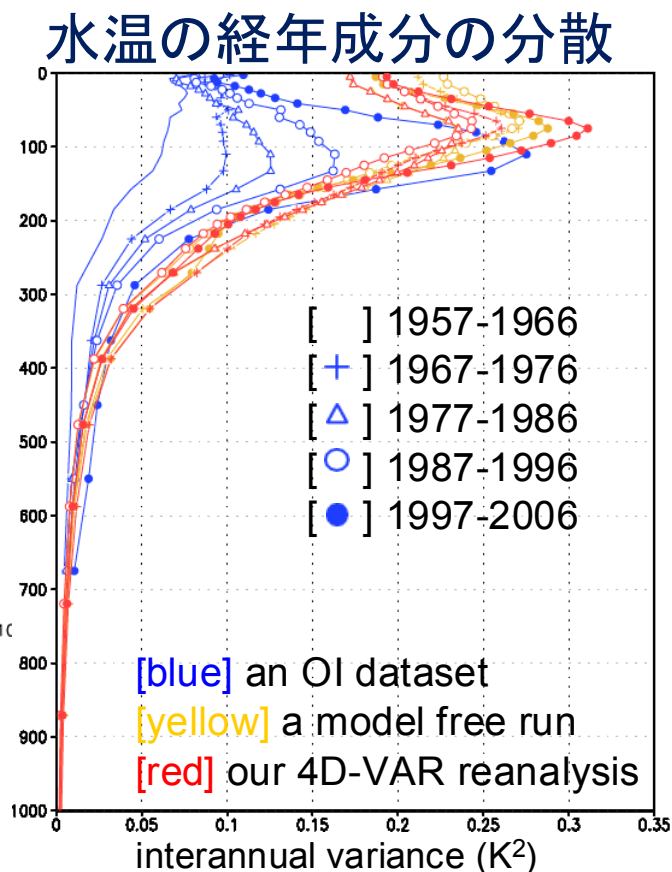
Observed Data Point Map for OI
[Oct. 2009 Pres.=10]



観測データ数はどれくらいあるか



海面の衛星観測データは数えていない。)



OIデータは観測数にもなって変動性が異なる。モデルシミュレーション、同化結果は変動性が大きく、最近のOIデータに近い。

“同化” 「大辞林 第二版」

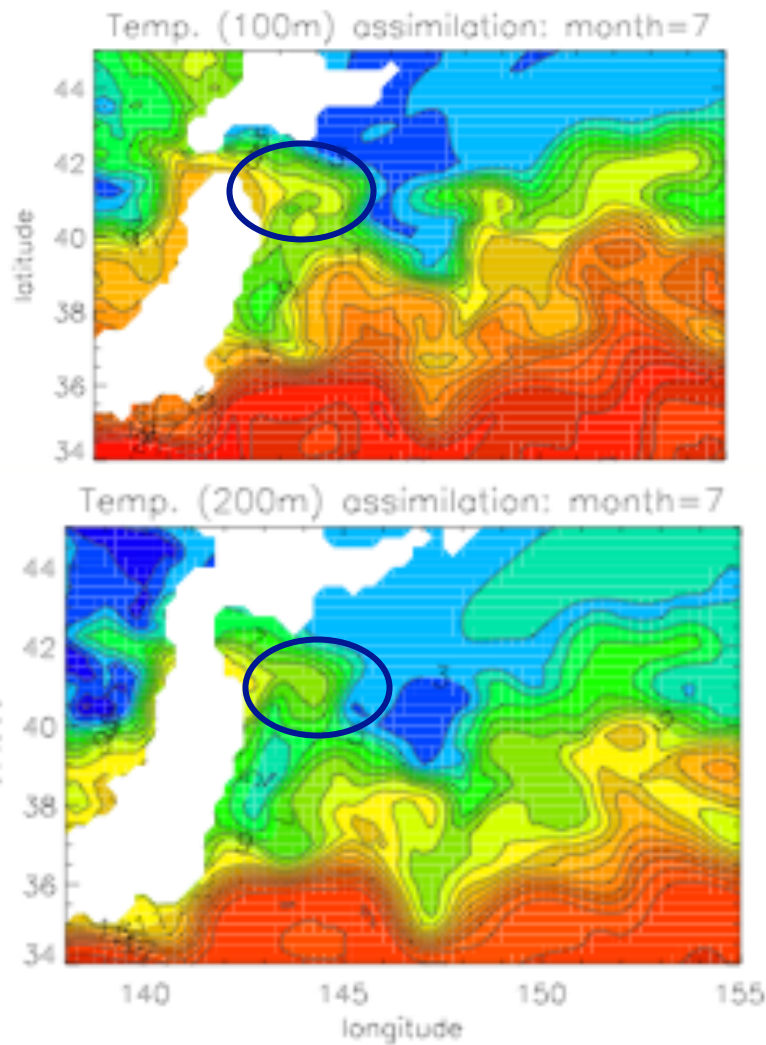
1. 本来異なる性質や考え方が同じものになること。
2. 外から得た知識などを理解して自分のものとする
こと。
3. 生物体が外界から摂取した物質に特定の化学変化
を加え、その生物に固有あるいは必要な物質を作
り出すこと。同化作用。アナボリズム。
4. マグマが周囲の岩石や外来物質を取り込み、一つ
のものに混合すること。同化作用。
5. ある音素が隣接する音素に影響されてそれと同じ、
または似た性質のものに変化すること。エビス
(ebisu) の i が先行する e と等しくなってエベス
(ebesu) となる類。

データ同化とは？

- 観測データと数値モデルを組み合わせて、現実的なデータセットを作るための手法
 - 複数の観測データを統合した時系列データセットを作成する
 - 数値モデルで用いているパラメータなどの最適化を観測データをもとにして行う
 - 数値予報のための初期条件、境界条件を作成する
 - 観測システムの設計のための評価、改善にむけた情報提供

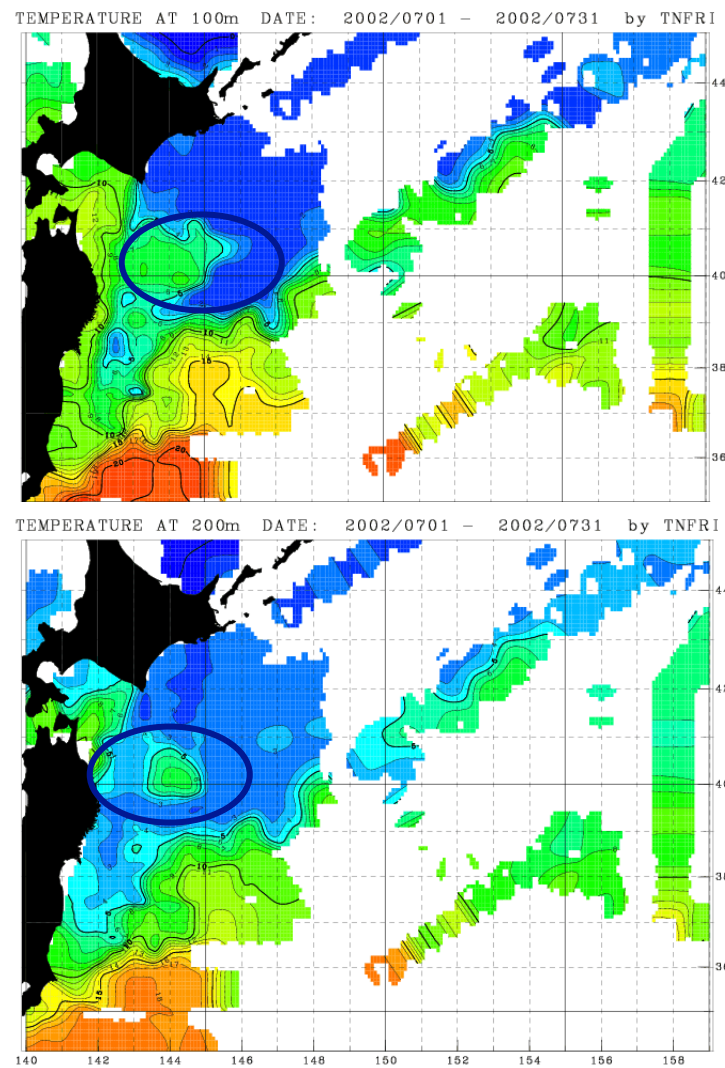
7月 データ同化結果の例 (亜表層水温場)

同化結果 東北海区水温図(東北水研より)



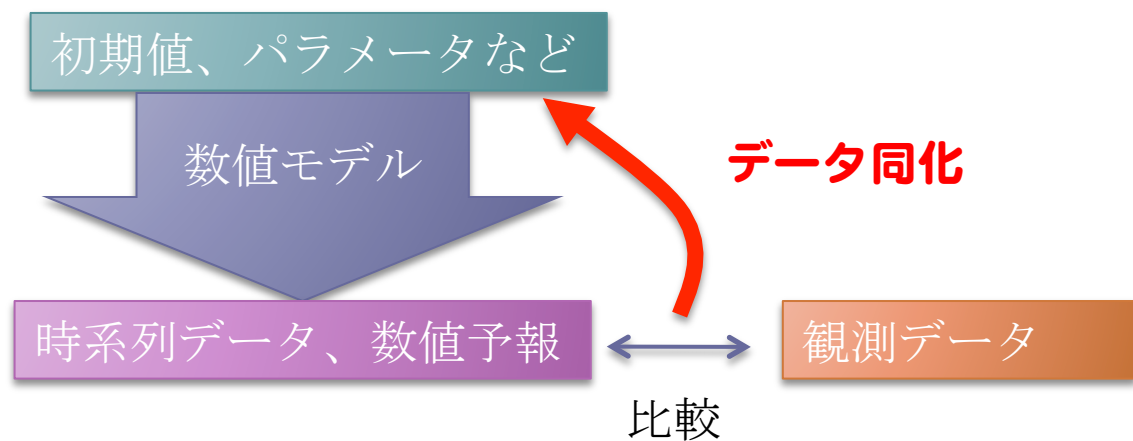
100m

200m



データ同化と関連した数学

- 微分方程式
- 行列、線形代数
- 確率論(確率微分方程式)
- 統計学(誤差論)
- 最適化理論
- 最適制御
- 逆問題

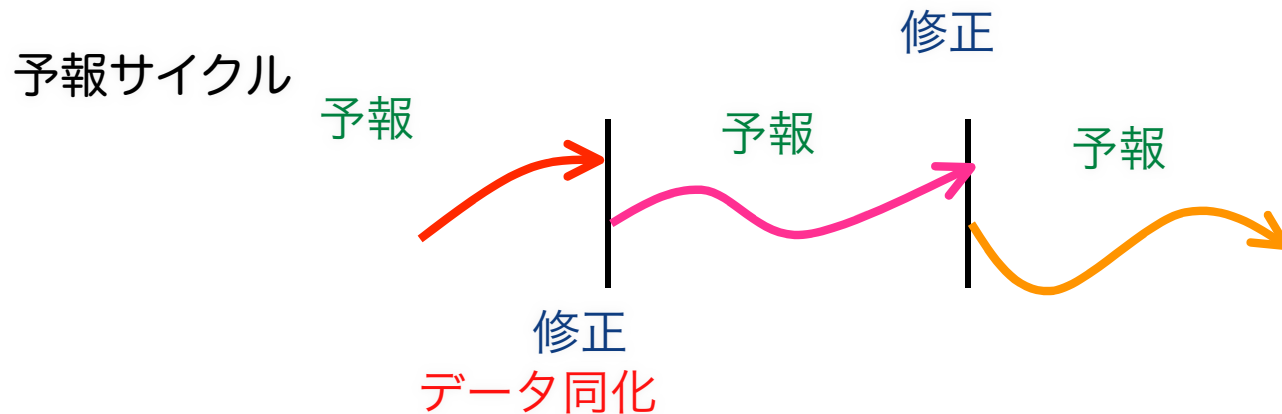


データ同化とは？

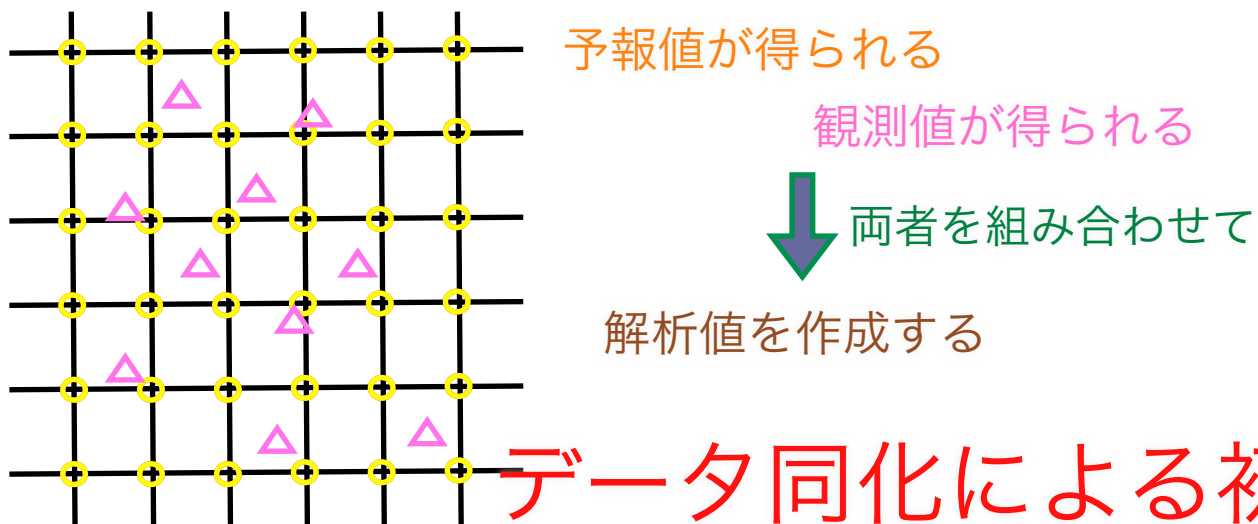
- 観測データと数値モデルを組み合わせて、現実的なデータセットを作るための手法
 - 複数の観測データを統合した時系列データセットを作成する
 - 数値モデルで用いているパラメータなどの最適化を観測データをもとにして行う
 - 数値予報のための初期条件、境界条件を作成する
 - 観測システムの設計のための評価、改善にむけた情報提供

数値モデルに観測データを“同化”する。

数値予報とデータ同化



データ同化: 観測データを用いた予報値の修正



データ同化手法の基礎

線形不偏最適推定理論

Best Linear Unbiased Estimation (BLUE)

ある変数 x (未知の真値: x_{true})に対して
異なる2つの推定値 x_1, x_2 および
それぞれの誤差 σ_1, σ_2 が与えられたとき、
 x に対する最適推定値は次式で与えられる。
ただし、 x_1, x_2 の期待値は x_{true} と等しいとする。

$$x_{\text{opt}} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2}$$

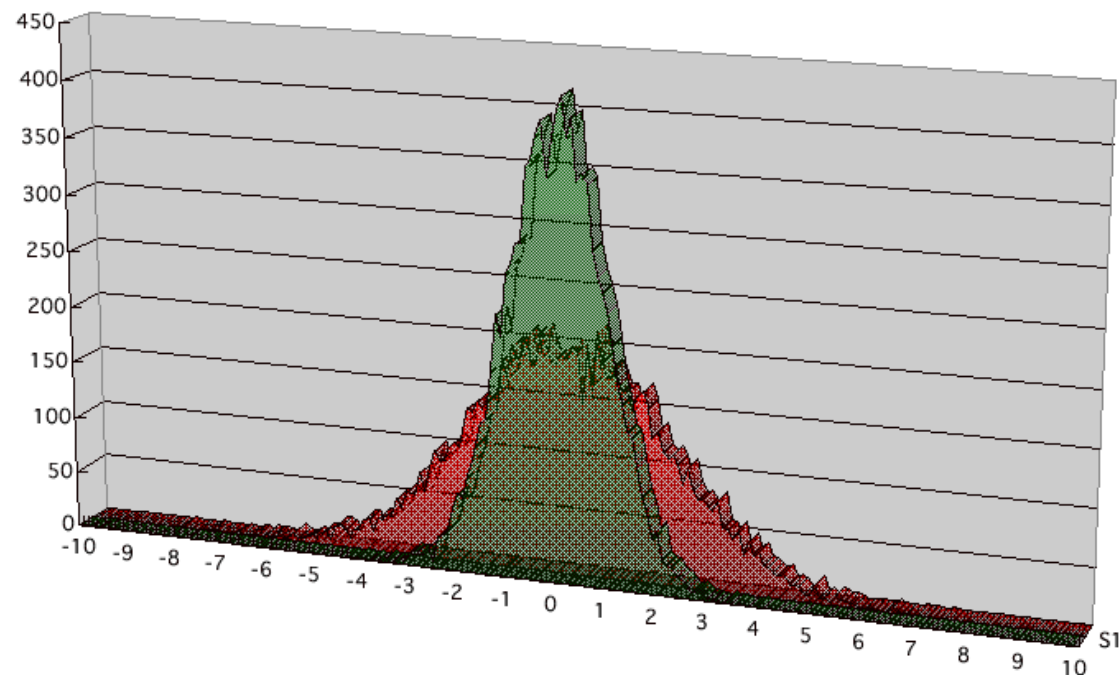
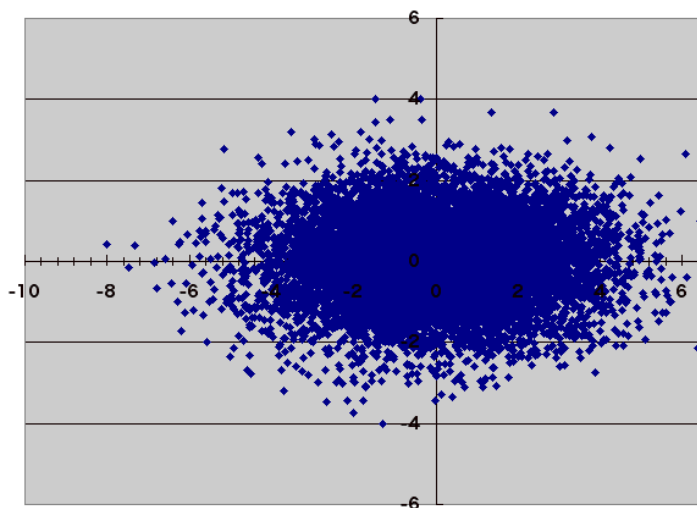
不偏最適推定の簡単な例題

ある変数 x (未知の真の値=0) についての推定値として

x_1 : 期待値 0, 標準偏差 1

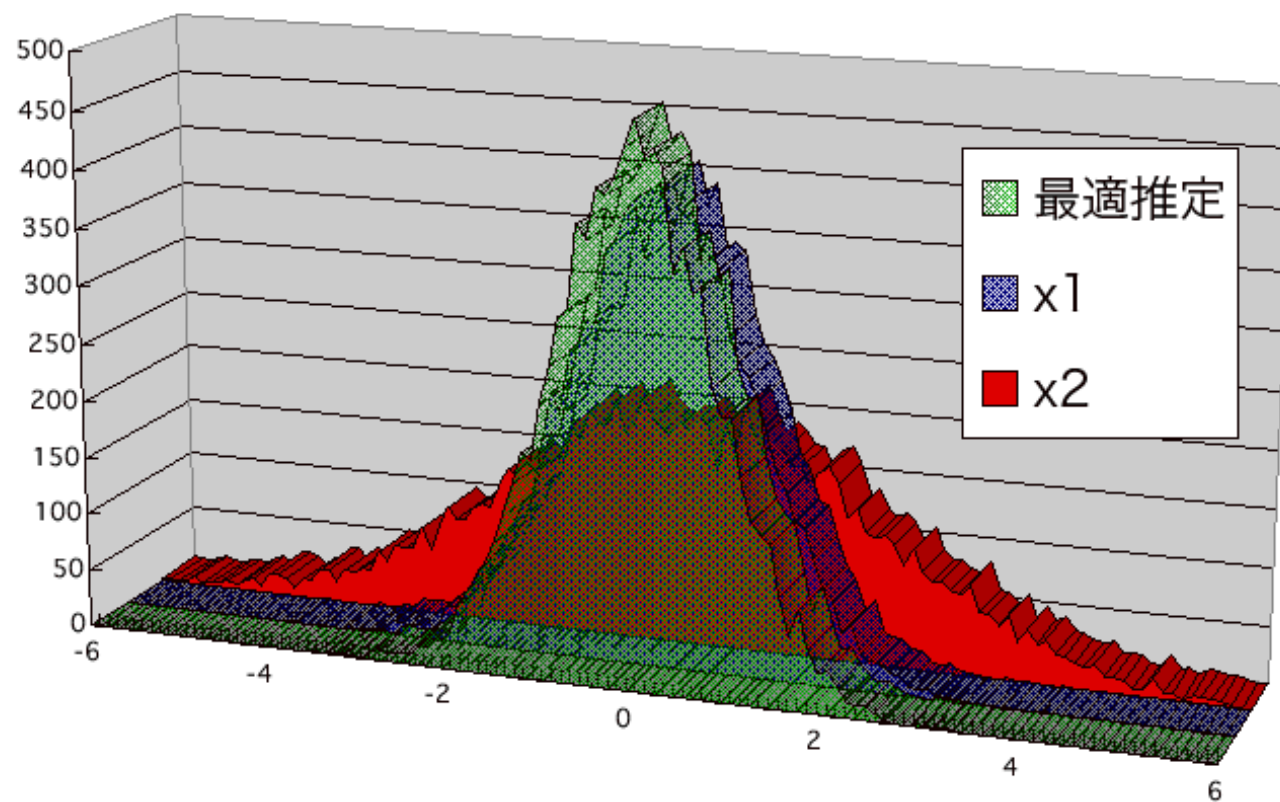
x_2 : 期待値 0, 標準偏差 2

があたえられた時の、最適推定値を求めよ。



不変最適推定理論から最適解は

$$x_{opt} = \frac{4x_1 + x_2}{5} \quad \text{で与えられる。}$$



不偏最適推定理論における注意点

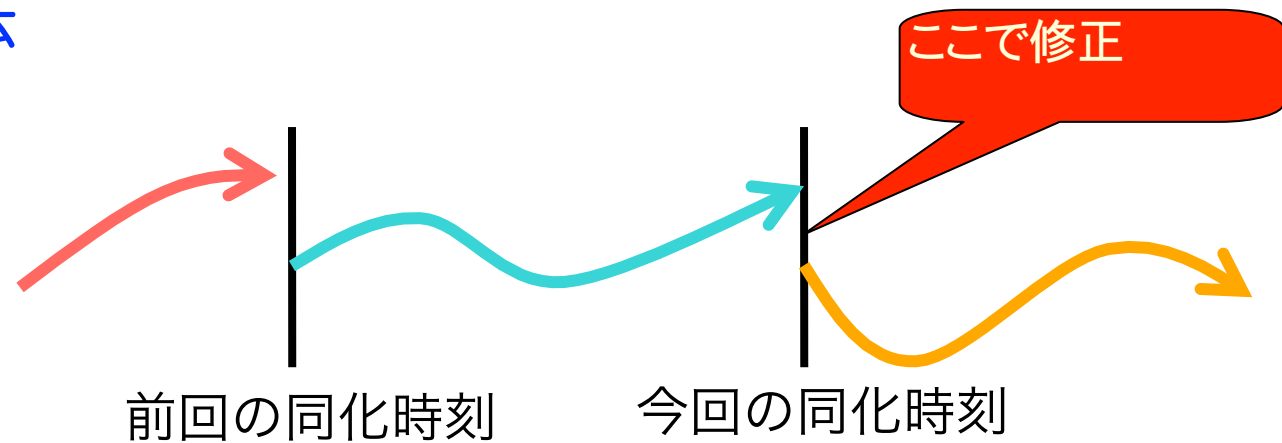
- 求められた最適解は統計的(確率的)に最適であり、個々のケースでは必ずしも真の値に近づくとは限らない。
- 最適解の精度はあらかじめ推定された誤差の精度に依存している。実際に推定誤差の比が重要である。
- 現実には、バイアスがない、誤差に相関がないという仮定が成り立つ保証がない。

実際のデータ同化手法はこの不偏最適推定理論を多変量、多次元に拡張したものである。

データ同化手法の分類

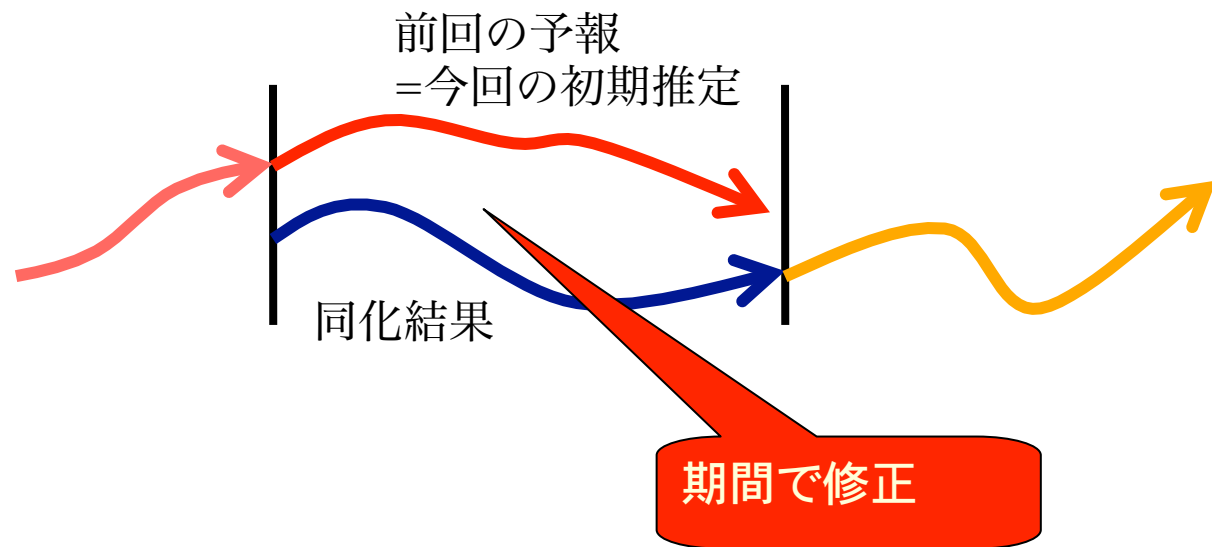
Sequential法

最適内挿法
カルマンフィルタ
3D-VARなど



Iterative法

4D-VAR
RTSスムーザなど

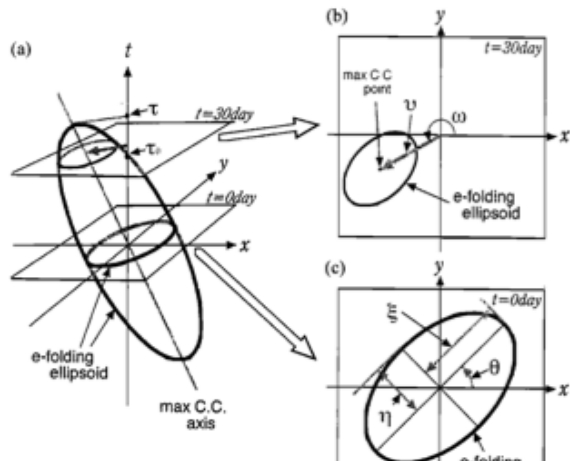


カルマンフィルター

$$x^a = x^f + K(y - Hx^f) \quad K = PH^T(HPH^T + R)^{-1}$$

- 行列Hは観測行列と呼ばれ、モデルの予報値を観測点に射影するオペレータである。
- カルマンフィルターは予報誤差共分散行列の時間発展を予報モデルを用いて求めるものである。
- 予報誤差共分散行列Pはあらかじめ与える必要があり、最適内挿法などの簡単な同化手法では、空間方向にガウス関数を仮定する。

flow dependent covariance



流れや波によって情報が運ばれるので、データ同化による修正は非等方的であるべき

Kuragano & Kamachi 2000

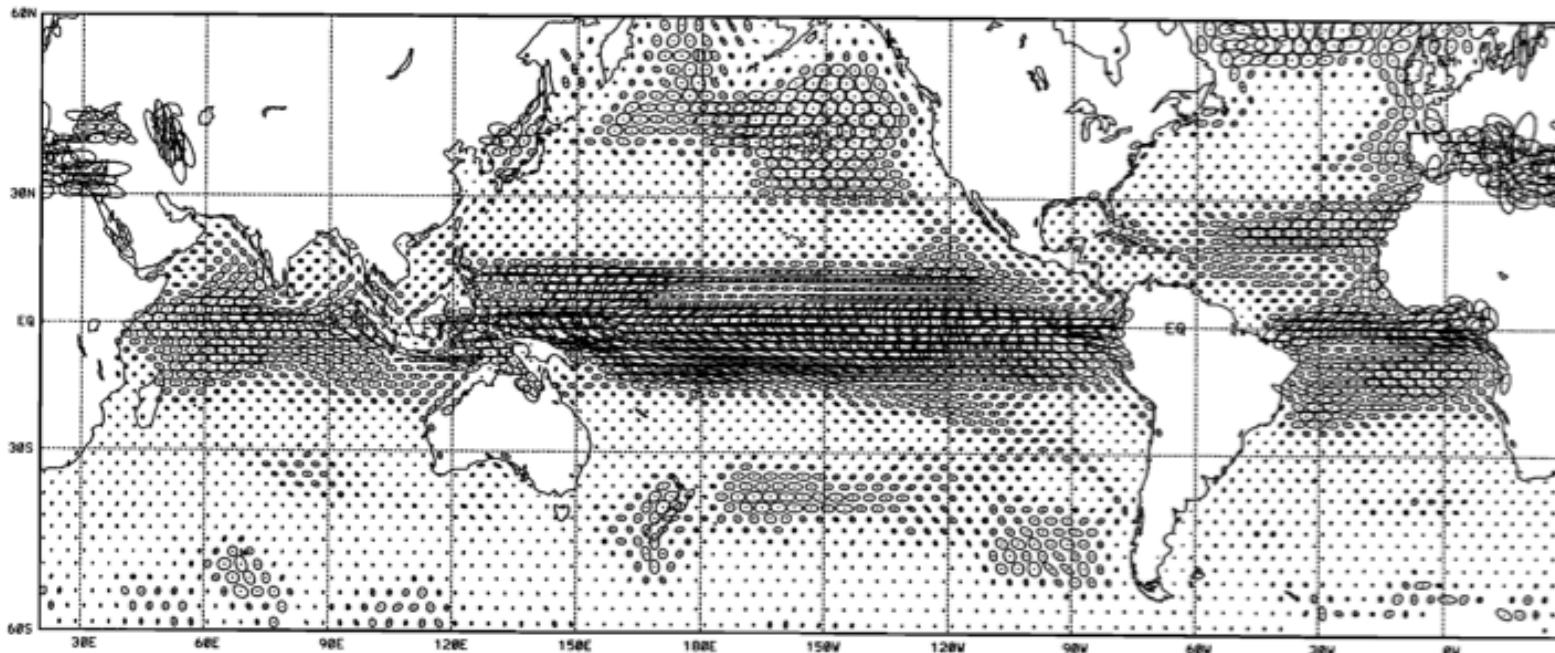


Figure 7. Spatial scales of the true SSHA variability. The elliptic contour lines show a range of the correlation coefficient of 0.95 on a x - y plane with zero time lag. Each ellipse is drawn in a 2° latitude \times 4° longitude staggered grid.

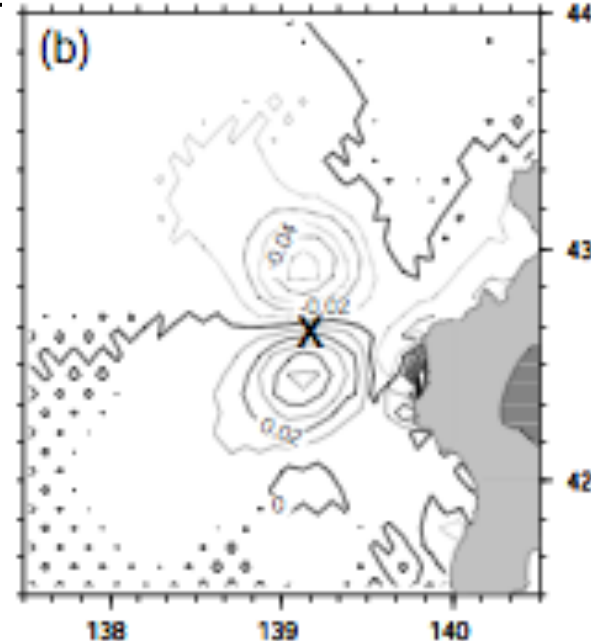
カルマンフィルタ

$$x^a = x^f + K(y - Hx^f) \quad K = PH^T(HPH^T + R)^{-1}$$

- カルマンフィルタは予報誤差共分散行列の時間発展を予報モデルを用いて求めるものである。

- 誤差の空間スケール、変数間の関係がモデルの方程式に対応して求められる

- 流速場などの変動にあわが得られる

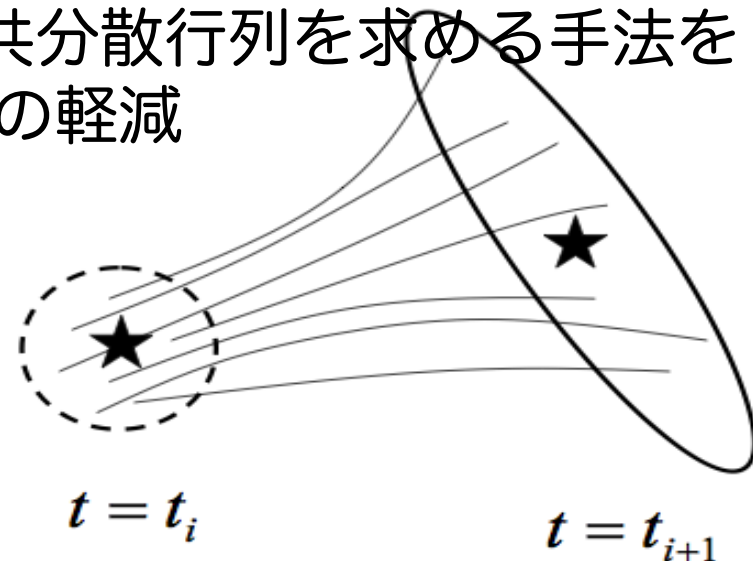


散行列

アンサンブルカルマンフィルター

$$x^a = x^f + K(y - Hx^f) \quad K = PH^T(HPH^T + R)^{-1}$$

- カルマンフィルタは予報誤差共分散行列の時間発展をリカッチ方程式を用いて求める。
- アンサンブルカルマンフィルターでは予報誤差共分散を予報モデルの多数回試行から求める。
- リカッチ方程式における線形化の問題を解決
- 少ないアンサンブル数で誤差共分散行列を求める手法を開発することにより、計算負荷の軽減



3D-VAR(3次元変分法)

$$J = \left(x^a - x^f\right)^T B^{-1} \left(x^a - x^f\right) + \left(Hx^a - y\right)^T R^{-1} \left(Hx^a - y\right)$$

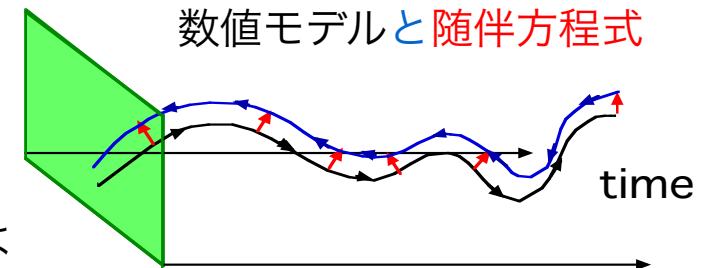
- 誤差共分散行列が同じ場合、最適内挿法と同じ最適解が得られる。
- Hとして非線形の観測オペレータを用いることができるため、観測データの有効利用ができる。
⇒ リモートセンシングデータのリトリバル
- 容易に付加的な制約条件を課すことができる。
例えば、空間的な滑らかさ、力学的バランスなど

4次元変分法(adjoint法)

$$J = (c^a - c^f)^T B^{-1} (c^a - c^f) + (Hx - y)^T R^{-1} (Hx - y)$$

$$L = J + \lambda^T \cdot Ax \quad (\text{制御変数 } c \text{ は初期・境界条件、パラメータなど})$$

数値モデルの前方積分とadjoint方程式($\lambda^n = A^* \lambda^{n+1}$)
の後方積分を繰り返して、制御変数の最適解を求める。



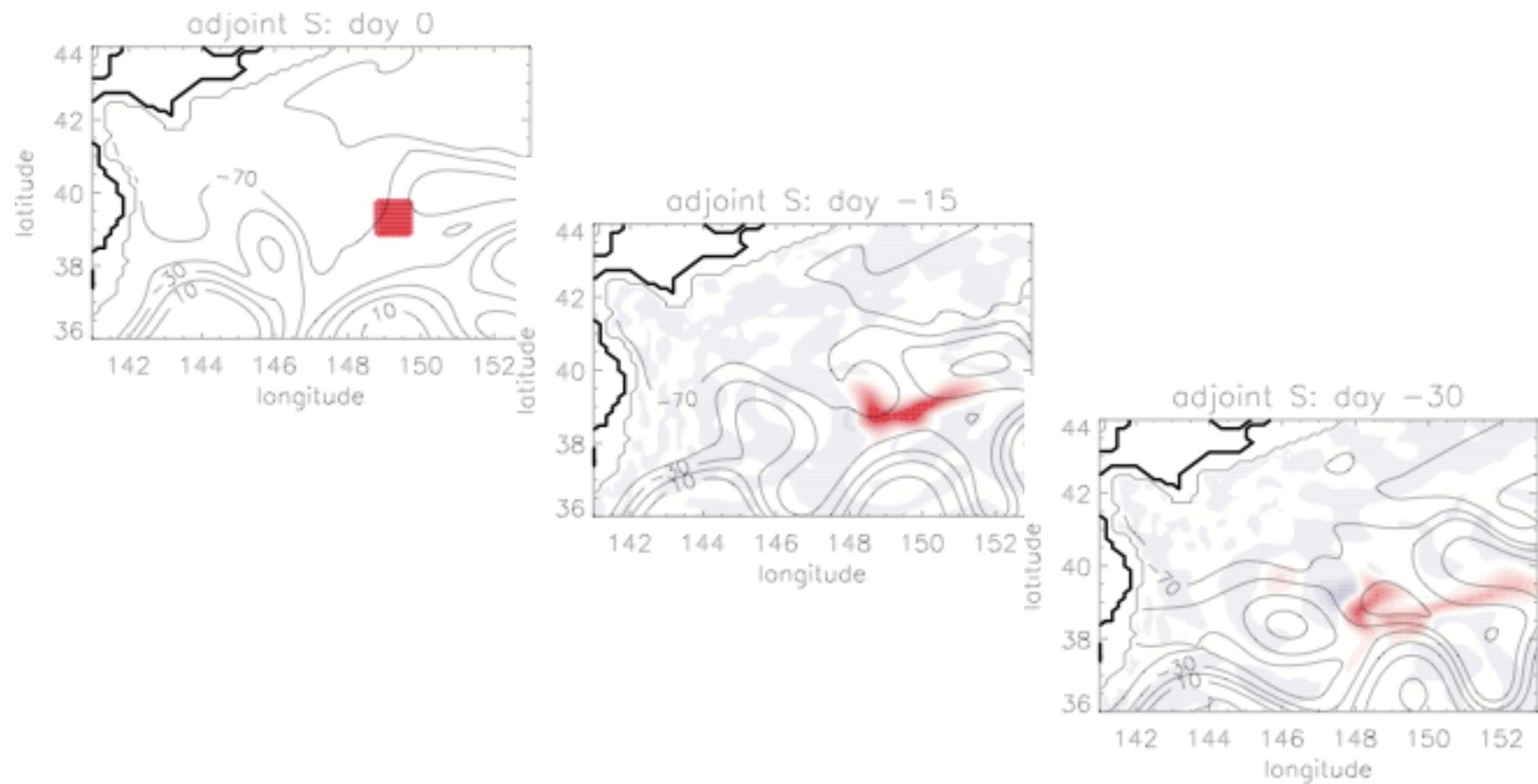
線形システムにおいてadjoint法の最適解は

- 同化期間の最後でカルマンフィルターと一致し
- 期間にわたってRTSスモーカーと一致する

非線形システムでは線形化の方法に依存

アンサンブルカルマンフィルターなどと一致する保証はない

adjointモデルによる後方積分



データ同化システムの設計

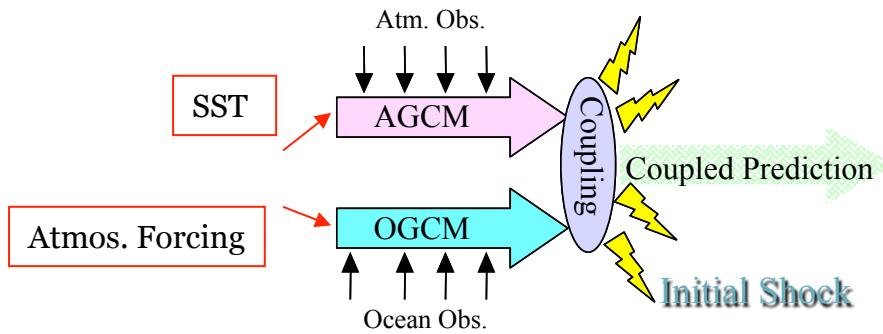
- 観測データ、数値モデル、同化手法のバランスが重要
- 観測データが豊富にあれば、簡単な同化手法でも良い推定値が得られる
- 数値モデルが複雑になると、表現される現象が多様になるので、制御するのが難しい
 - どのような変数／パラメータを修正して、どのような現象を再現するか

結合データ同化システムの概要

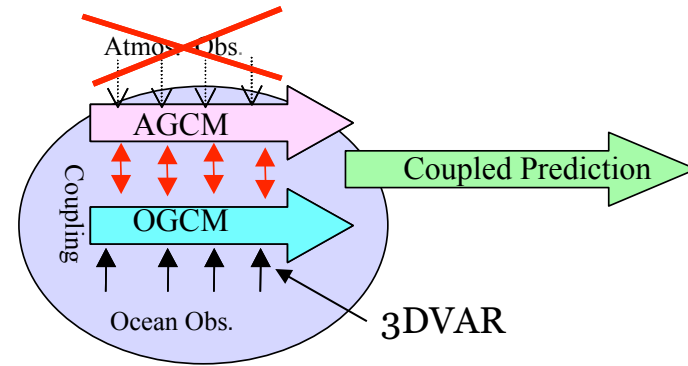
- Assimilation **with a fully coupled GCM**
 - Assimilation method is **4D-VAR**
 - Long assimilation window (9 month)
 - Correction of model climatology by parameter estimation
 - Correction of seasonal to interannual trajectory by initialization
 - Atmospheric data are also assimilated
 - Weather mode is treated as noise
- ➔ 季節から数年の時間スケールをもつ気候変動現象をターゲットにする。Weatherモードの変動についてはモデルの挙動にゆだねるが現象によっては平均場が改善されることで、implicitに物理場が修正されることが期待される。

結合同化システムの3つのタイプ: (Conventional independent/forced A & O, Quasi CDAS, & CDAS)

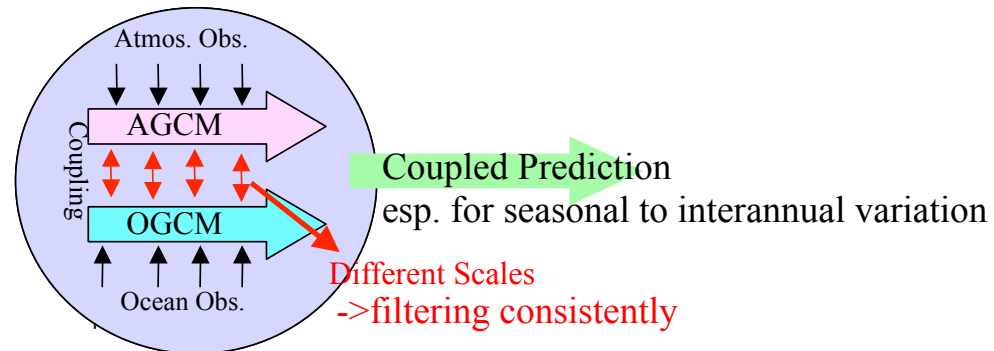
(a) Independent Systems (Conventional DAS)



(b) Quasi CDAS (MOVE-C)



(c) Fully coupled DAS (CDAS: K7)



1. Straight-forward concept
2. No initial shock
3. Higher forecast potential

結合同化システムの概要

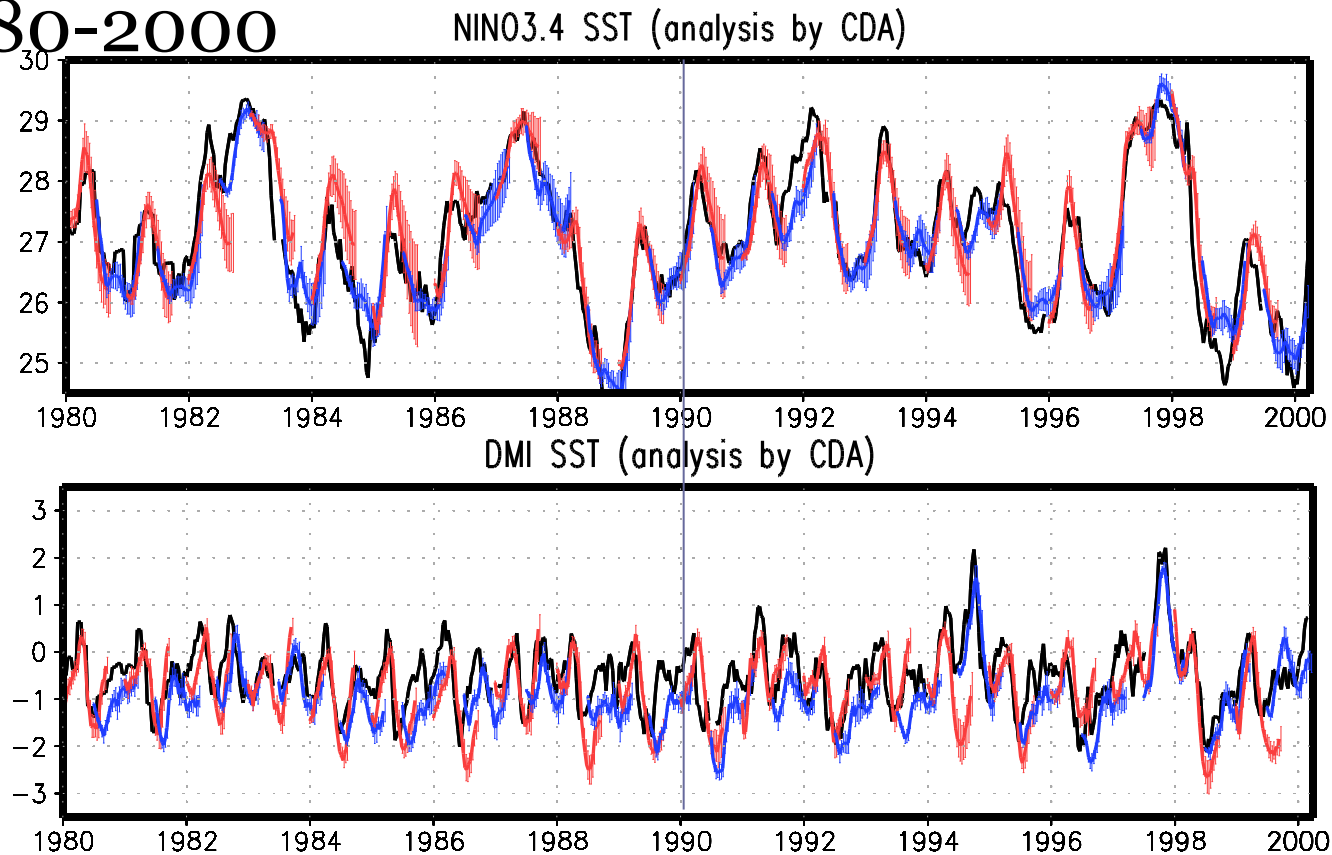
- Coupled Model (CFES):
 - T42L24 **AFES** for AGCM
 - 1x1deg L45 **MOM3** for OGCM
 - IARC SeaIce model
 - MATSIRO Model for Land
- Observational Data
 - Atmosphere:
 - NCEP's **BUFR data** U,V,T,Q (10daily)
 - SSM/I sea wind scalar x ERA40 wind direction (10daily)
 - Ocean:
 - T/P altimeter data(10daily)
 - Reynolds SST (10daily)
 - **WOD data** T,S (monthly)
 - Ocean Data Assimilation Product T,S(monthly)
- Adjoint Code
 - **Adjoint** OGCM and **adjoint** AGCM are coupled [Line by line transformation by TAMC,TAF]
 - **Temporal averaging** of forward field for the adjoint integration is applied to smooth the basic field
 - Adjoint AGCM contains **damping terms** to suppress the strong adjoint sensitivity from weather fluctuations.

$$-\frac{\partial \lambda}{\partial t} = \left(\frac{\partial M}{\partial \mathbf{x}} \right)_{\mathbf{x}=\bar{\mathbf{x}}}^T \lambda - \Gamma \lambda + \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H} \bar{\mathbf{x}} - \mathbf{y})$$

λ : adjoint variables, $\bar{\mathbf{x}}$: temporal average, $-\Gamma \lambda$: damping.

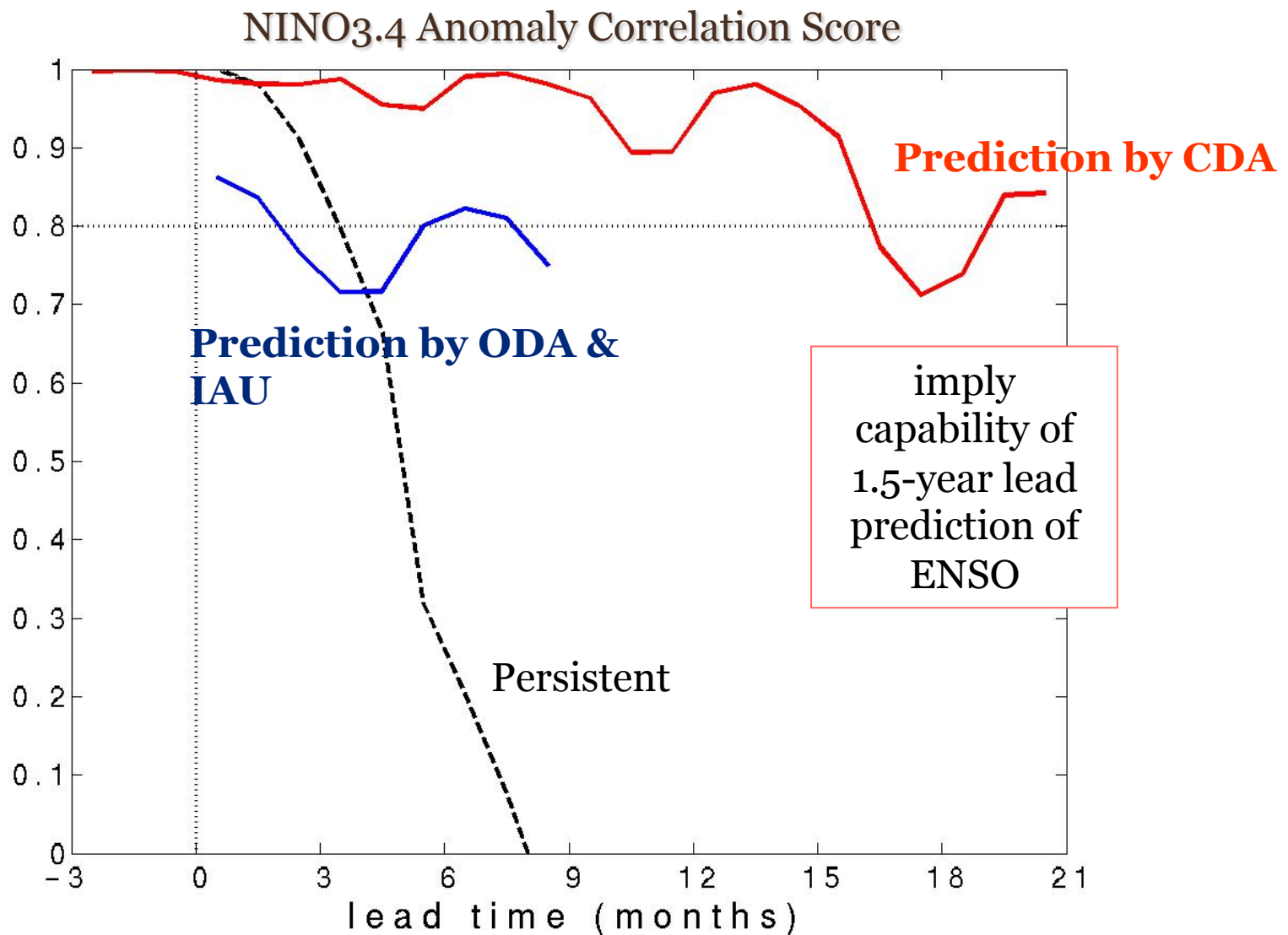
気候変動現象indexの再現性

- CDA Reanalysis experiment for the period of 1980-2000



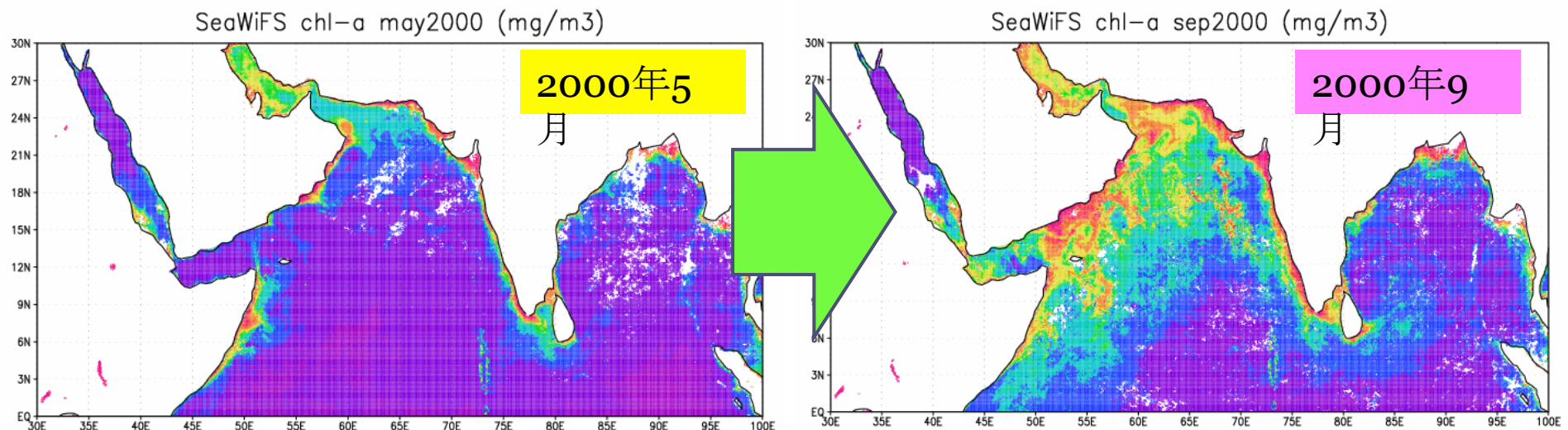
4 D-VAR同化手法を用いたENSO予報研究

(初期値 : Case of 3-month assimilation window;
1997/98)



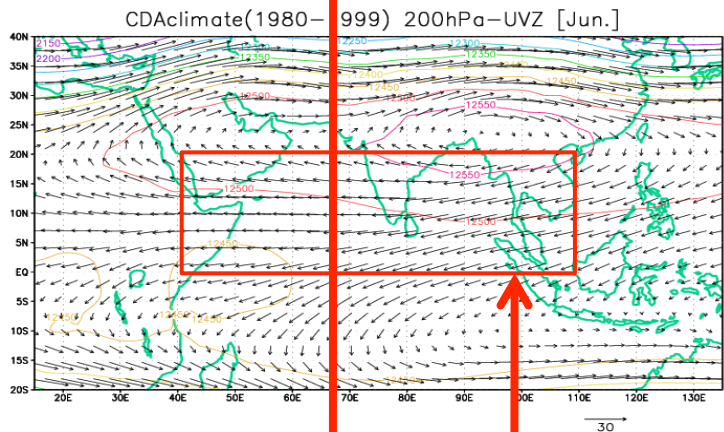
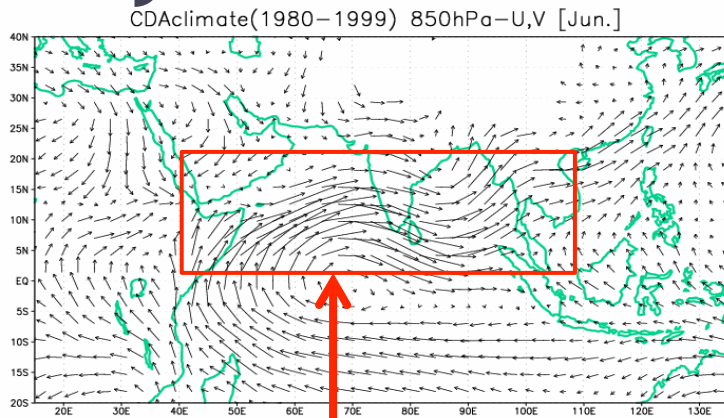
結合同化システムによるインドモンスーン予測

- インド熱帯気象研究所 (IITM) との協働により、気候災害の軽減と水産資源管理に向けたインド洋域での海洋・大気間の熱・水リンクのルート解明及びインドモンスーン予測研究を開始した。今後も継続的に協働研究を進めていく。



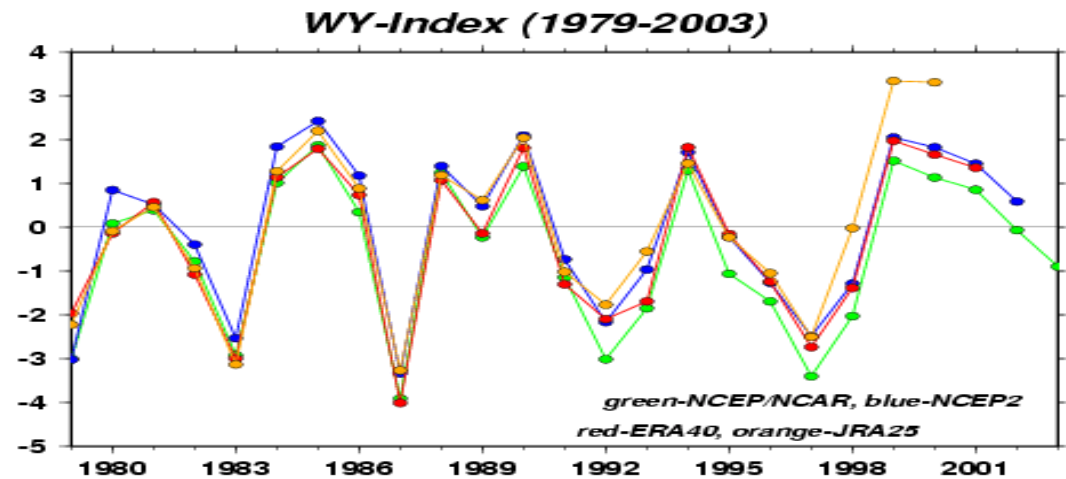
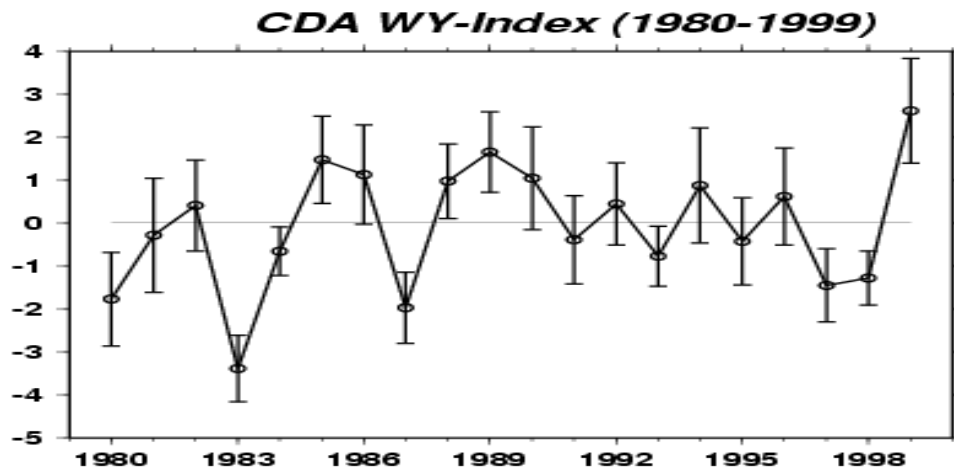
SeaWiFS観測によるchl-a

Dynamical monsoon index



$$DMI = U_{850}^* - U_{200}^*$$

結合同化システムによる
インドモンスーン循環の再
現



Correlation coefficients (vs. CDA)

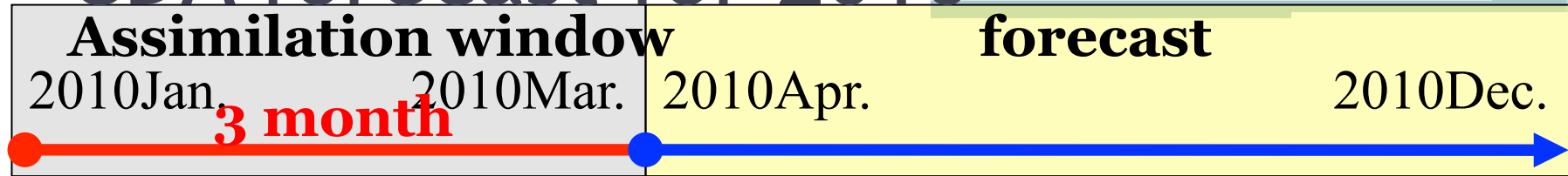
ERA40 → 0.720

JRA25 → 0.765

NCEP1 → 0.674

NCEP2 → 0.676

CDA forecast for 2010

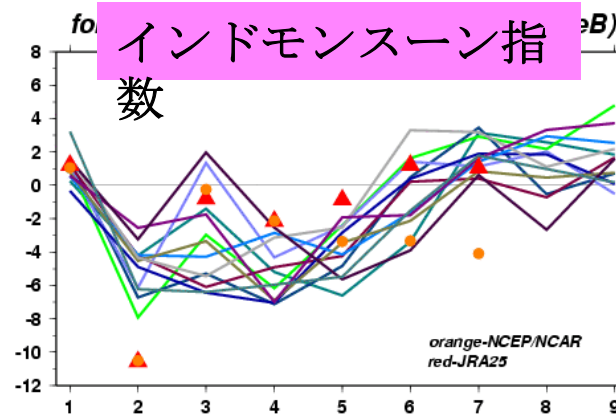
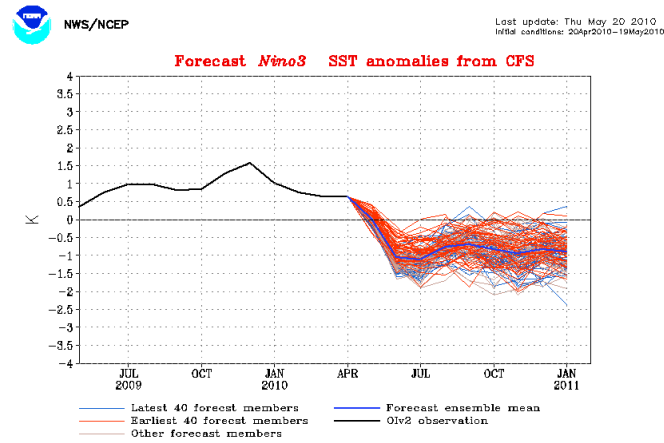
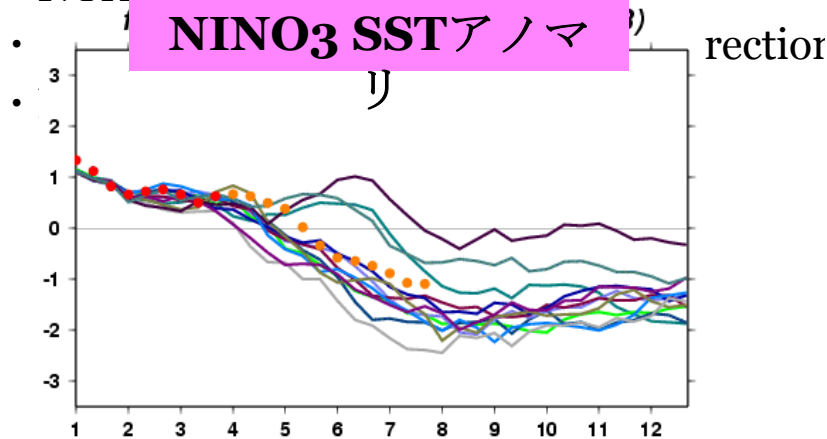


初期条件

- 50年海洋長期再解析から2010年に似たケース(Jan.1995)を使用

同化データ (10-day mean)

- NOAA OISST



結合4DVARによる4月初旬の初期値化予測は、6・7・8月のモンスーンの観測結果を7割程度予測できた。

Parameter Estimation using Green's Function method

Toyoda et al. (in revision, IMBER/IMBIZO special issue)

✓ We divide the model domain (75°S-80°N) into 40 provinces based on Longhurst (1995).

✓ Coastal provinces in Longhurst (1995) are excluded regarding the coarse resolution of the ocean reanalysis.

✓ Japanese Sea, Okhotsk Sea, Costa Rica Dome region, and region off the coast of Peru are added taking into consideration of the specific characteristics of the physical environment.

✓ In each province, the optimal set of the NEMURO parameters are estimated. Note that the region within 2 degree from the coast line (shaded) is eliminated from the optimization.



Optimization

- ✓ We select **32 independent control variables** from all 70 NEMURO parameters.
 - those which differ largely between sites (Fujii et al., 2007; Yoshie et al., 2007)
 - those which have large sensitivity (Yoshie et al., 2007)
 - coefficients of photosynthetic rate, respiration, and mortality for each phytoplankton and of grazing and mortality for each zooplankton are estimated dependently.
- Cost function to be minimized is defined as follows:
 - interannual trend during the experiment (J_{tr})
 - differences between the analyzed and the observations of nutrients (J_{nt})
 - same as J_{nt} but of surface chlorophyll concentration (J_{cl})
 - difference between the analyzed control variables and their first guesses (J_{bg})

$$J = J_{bg} + J_{tr} + J_{nt} + J_{cl}$$

$$J_{tr} = \frac{1}{2} \sum_{\text{variable}} \sum_{\text{grid}} s \cdot \frac{(\text{mean}_{1991-1996} - \text{mean}_{2001-2006})^2}{\text{mean}_{1991-2006}^2},$$

$$J_{nt} = \frac{1}{2} \sum_{\text{NO}_3, \text{Si(OH)}_4} \sum_{\text{month}} \sum_{\text{grid}} s \cdot \frac{(\text{mean}_{1991-2006} - y_{\text{WOA}})^2}{\sigma_{\text{WOA}}^2},$$

$$J_{cl} = \frac{1}{2} \sum_{\text{month}} \sum_{\text{grid}} s \cdot \frac{(\text{chl}^{mn} - y_{\text{SW}}^{mn})^2}{\text{var}_{\text{SW}}^{mn} + \text{var}_{\text{SW}}^{\text{hori}} + \text{var}_{\text{mdl}}^{\text{hori}} + \text{var}_{\text{mdl}}^{1-2}}$$

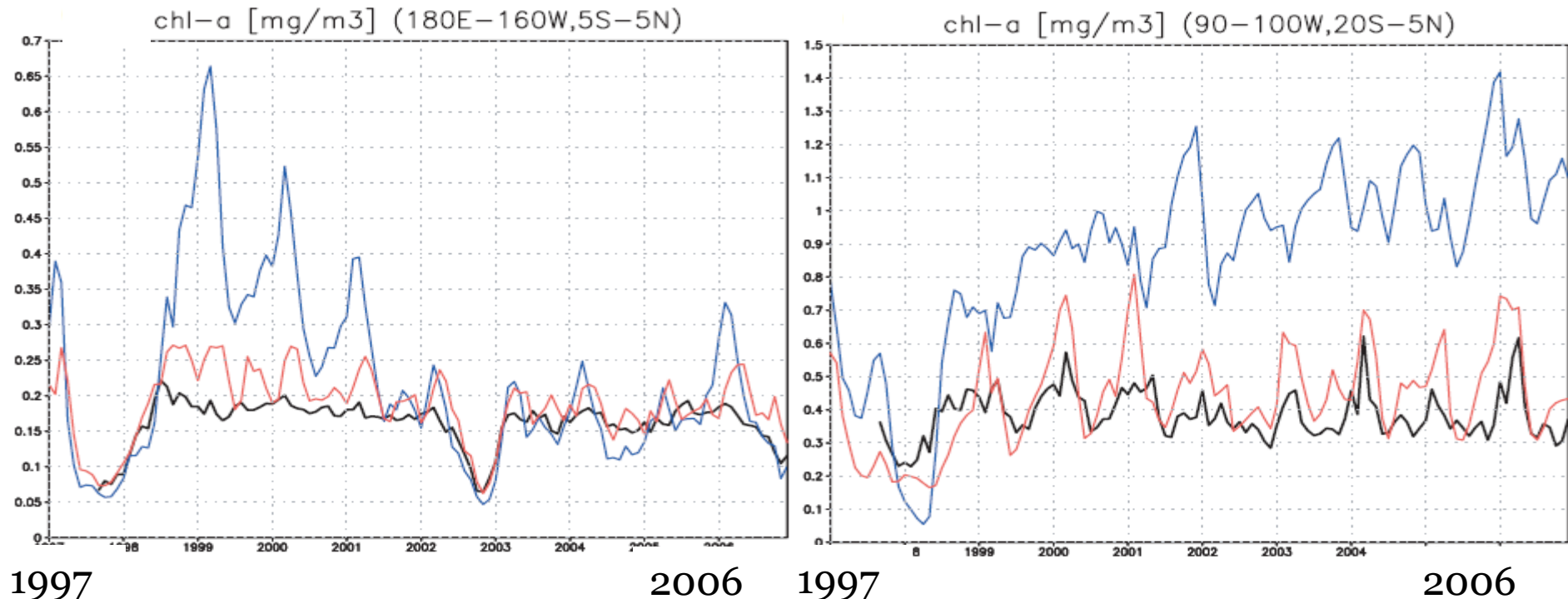
$$+ \frac{1}{2} \sum_{\text{time}} \sum_{\text{grid}} s \cdot \frac{(\text{chl}^{ia} - y_{\text{SW}}^{ia})^2}{\text{var}_{\text{SW}}^{ia} + \text{var}_{\text{SW}}^{\text{hori}} + \text{var}_{\text{mdl}}^{\text{hori}} + \text{var}_{\text{mdl}}^{1-2}},$$

$$J_{bg} = \frac{1}{2} \sum_a s \cdot \frac{a^2}{\Delta a^2} \equiv \frac{1}{2} \mathbf{a}^T \mathbf{B}^{-1} \mathbf{a},$$

Examples of improved time series of surface chlorophyll concentration:
(left) central ant (right) eastern tropical Pacific ($180^{\circ}\text{E}-160^{\circ}\text{W}$, $5^{\circ}\text{S}-5^{\circ}\text{N}$; $90-100^{\circ}\text{W}$, $20^{\circ}\text{S}-5^{\circ}\text{N}$)

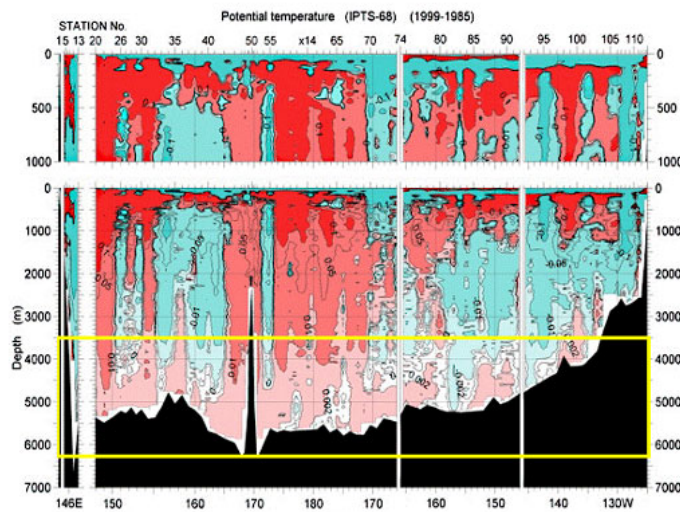
blue: control experiment, red: optimized experiment, black: SeaWIFS data

- In these regions (ones of the most improved regions), the mean and interannual variability of the surface chlorophyll concentration are quantitatively reproduced in the optimized experiment.
- By analyzing the NEMURO variables and the physical fields in the 4DVAR reanalysis, the interannual variability in phytoplankton in these regions is generated by the variability in nutrients, which is attributed mainly by the zonal advection change, not by the vertical one.

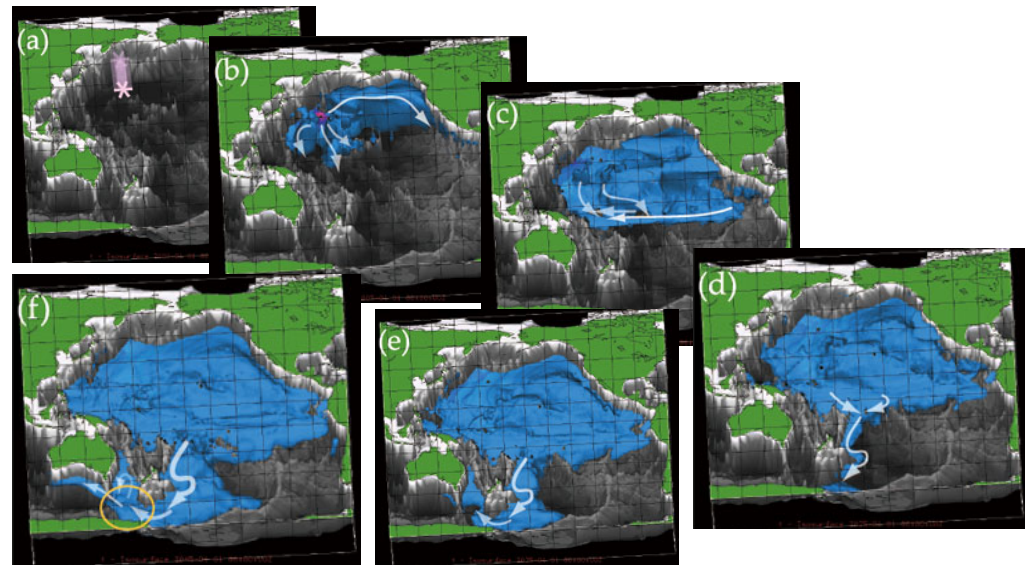


海洋50年再解析を用いた北太平洋最深層部の昇温メカニズム解明

- 4D-VAR海洋長期再解析データセット(1957-2006年)を用いると共に、アジョイントモデルによる感度実験を行うことにより、北太平洋最深層部の昇温メカニズムの解明研究を行いScienceに掲載された。
- 北太平洋（北緯47度）の底層でJAMSTECが発見した水温上昇は、時間を遡って解析するとおよそ40年前の南極アデリーコースト沖周辺での海面フラックス変動の影響であることがわかった。

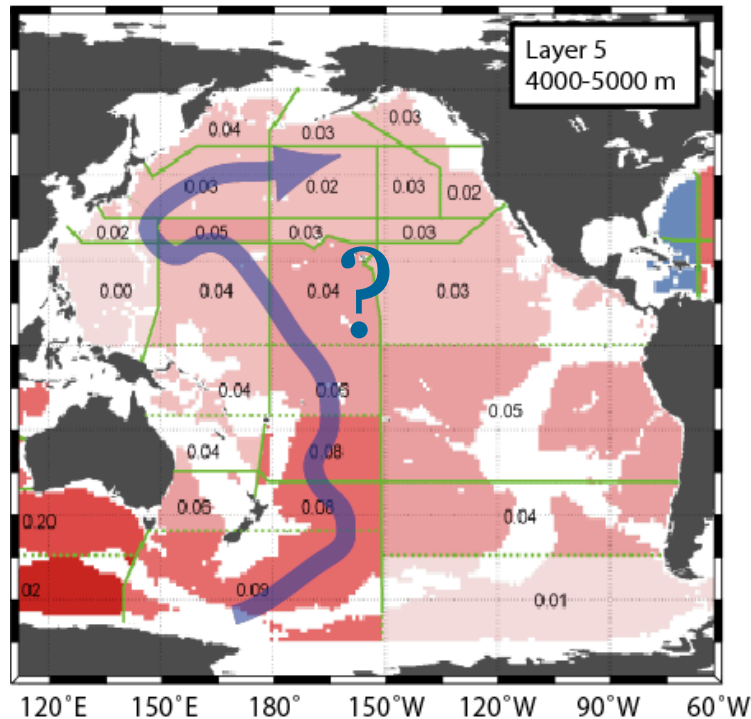


47Nに沿った船舶観測による1999年と1985年の水温差(Fukasawa et al.,2004)



海洋50年再解析データセットを基にした4D-VAR全球全層海洋同化システムによる感度解析結果

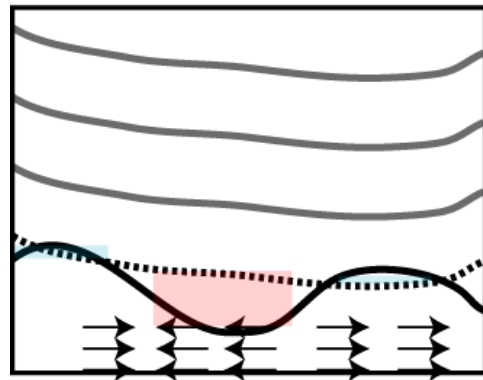
How about the Physical Mechanism?



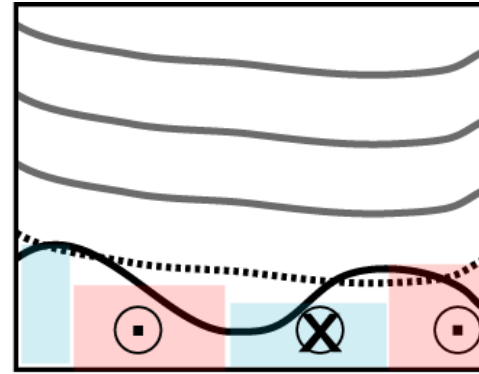
Difference of the heat storage between WOCE-WOCE revisit observational periods.

- ✓ The physical mechanisms governing bottom-water warming are poorly understood since in-situ observations are spatially and temporally sporadic.
- ✓ The changes in heat storage between WOCE-WOCE revisit imply northward running of the warming signal, but...

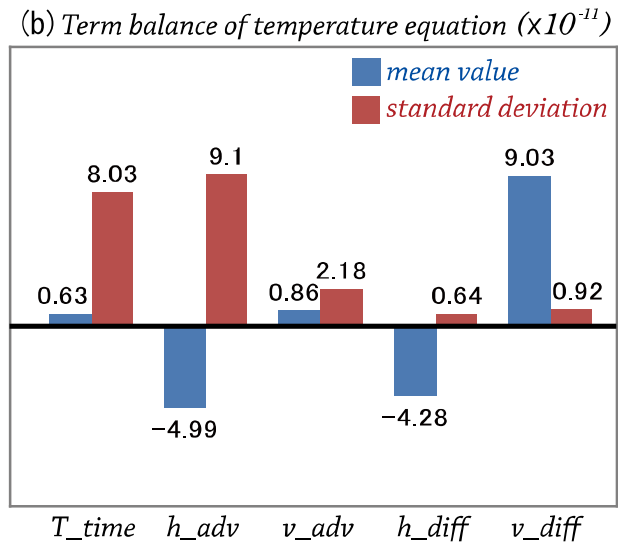
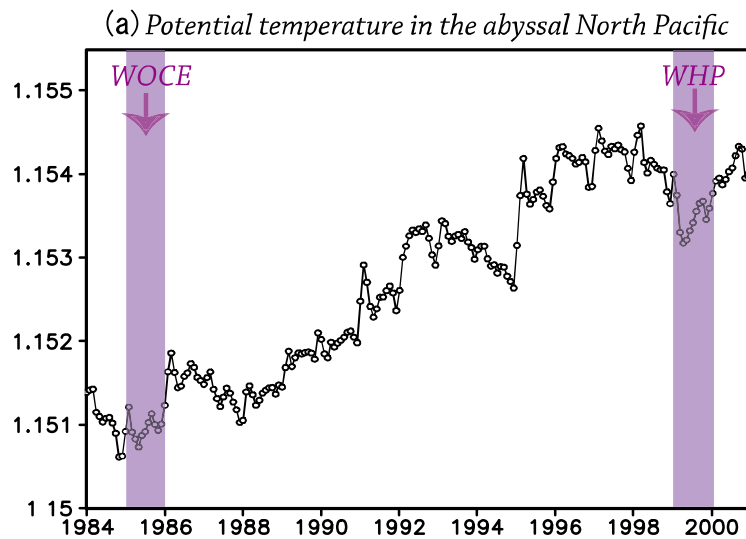
Validation with reanalysis dataset; Term balance of the temperature equation



wave motion
vertical advection
heat distribution

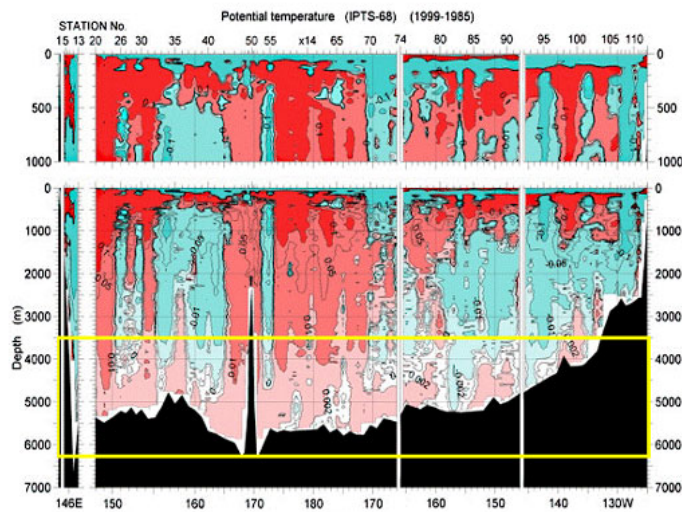


adjusted current
horizontal advection
heat balance

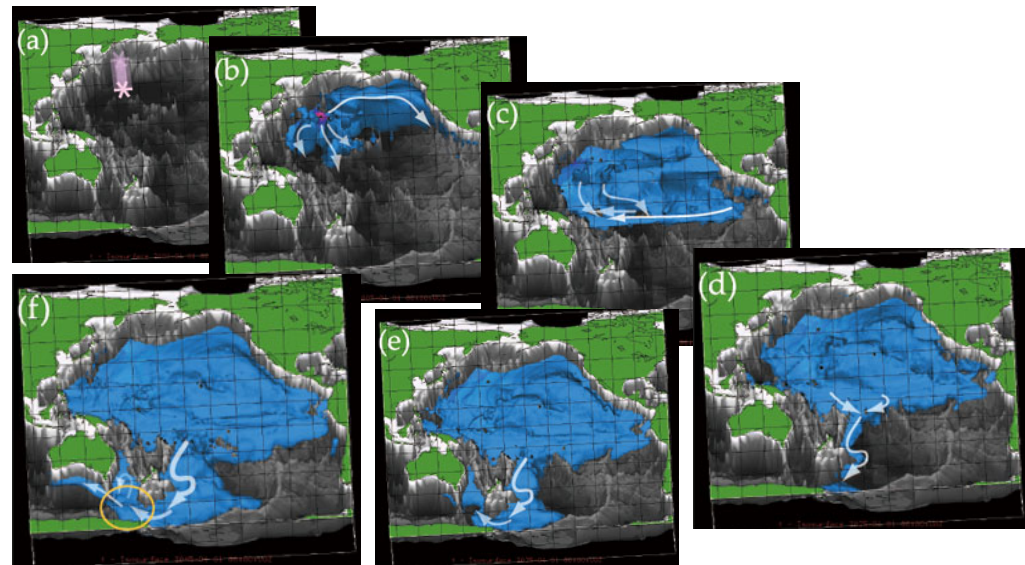


海洋50年再解析を用いた北太平洋最深層部の昇温メカニズム解明

- 4D-VAR海洋長期再解析データセット(1957-2006年)を用いると共に、アジョイントモデルによる感度実験を行うことにより、北太平洋最深層部の昇温メカニズムの解明研究を行いScienceに掲載された。
- 北太平洋（北緯47度）の底層でJAMSTECが発見した水温上昇は、時間を遡って解析するとおよそ40年前の南極アデリーコースト沖周辺での海面フラックス変動の影響であることがわかった。



47Nに沿った船舶観測による1999年と1985年の水温差(Fukasawa et al.,2004)



海洋50年再解析データセットを基にした4D-VAR全球全層海洋同化システムによる感度解析結果

より詳しく勉強したい人へ

- データ同化についての教科書があります。
- 「データ同化」観測・実験とモデルを融合するイノベーション (京都大学出版会)
- データ同化入門 一次世代のシミュレーション技術―(朝倉書店)