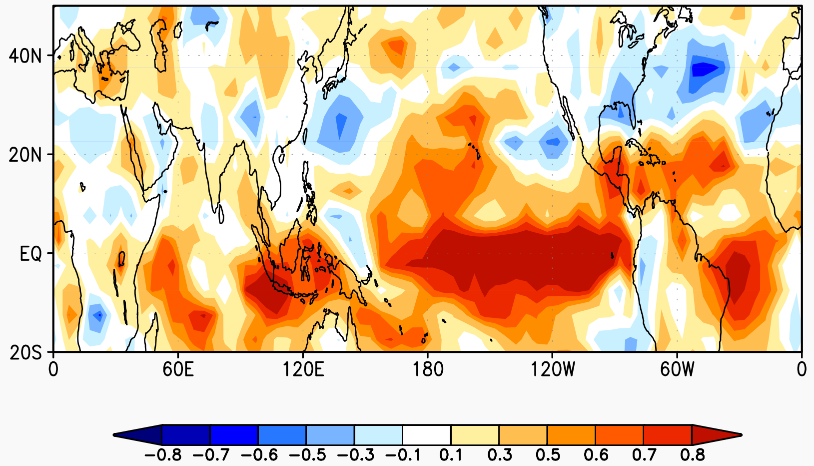
# Initialization Issues of Coupled Ocean-atmosphere Prediction System

In-Sik Kang

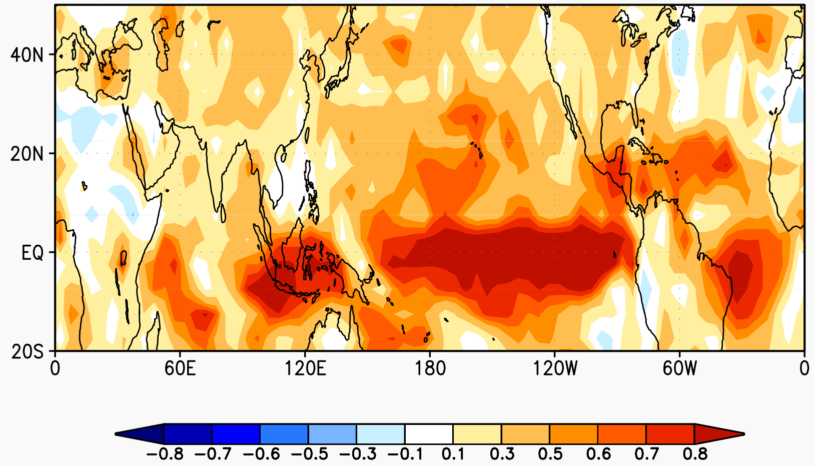
Climate and Environment System Research Center
Seoul National University, Korea

# Statistical correction (SNU AGCM)

**Before Bias correction**




**After Bias Correction**



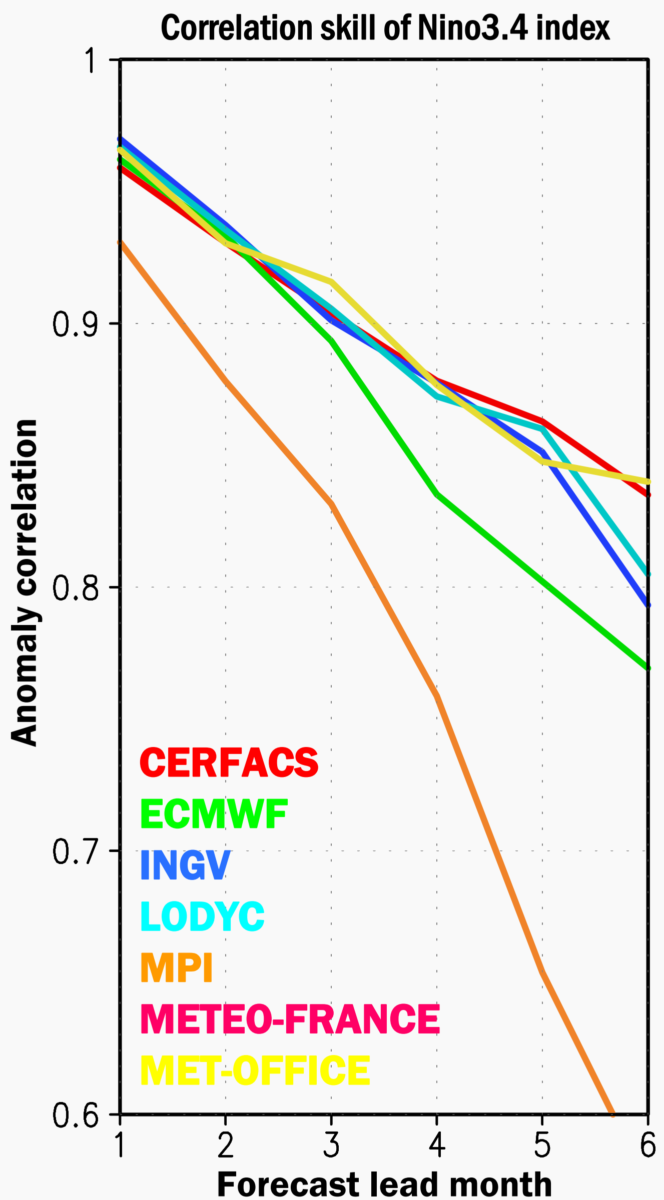

**Predictability over equatorial region is not improved much**

# Is post-processing a fundamental solution?

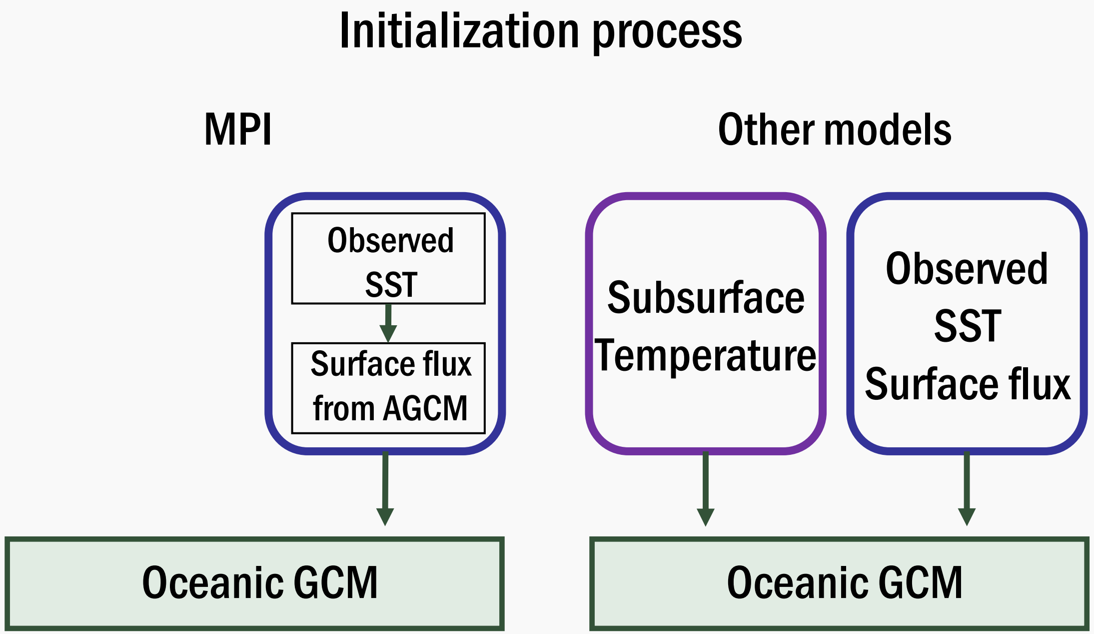
Forecast error comes from ( model imperfectness / poor initial conditions )

( **Model improvement** / **Good initialization process** ) is essential to achieve to predictability limit

# Impact of Initial condition



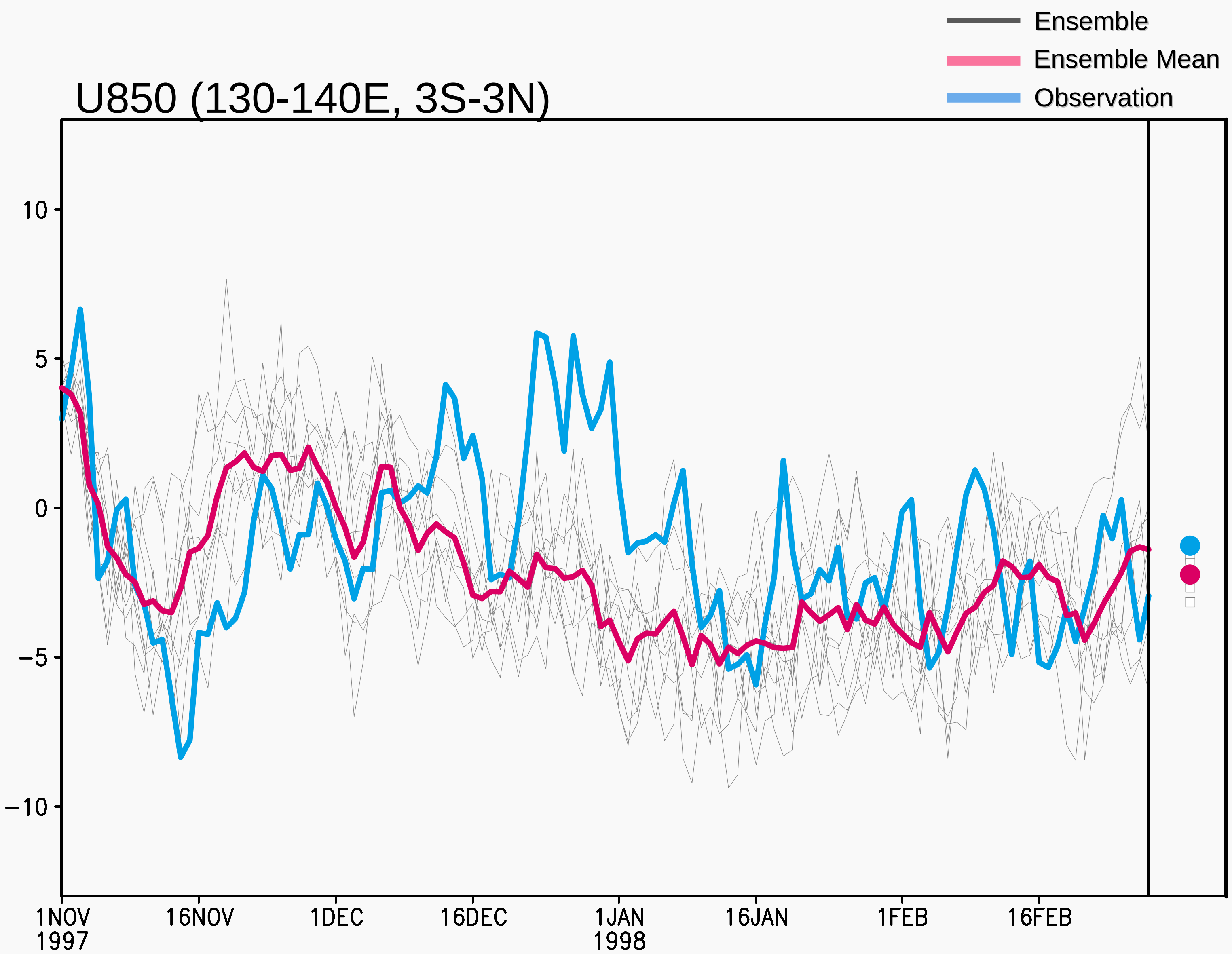

**MPI : Poor initialization process compared to others**




Surface flux is not enough to force the ocean

Temperature observations plays an important role to reflects observational states

# Ensembles Forecasts with Small Initial Perturbations



# What determines quality of Initial conditions?

- **Optimal (accurate & well-balanced) initial conditions**

  → Data Assimilation

  - **Accurate : Errors in initial condition is minimized**
  - **Well-balanced : Physical laws in forecast model is satisfied**

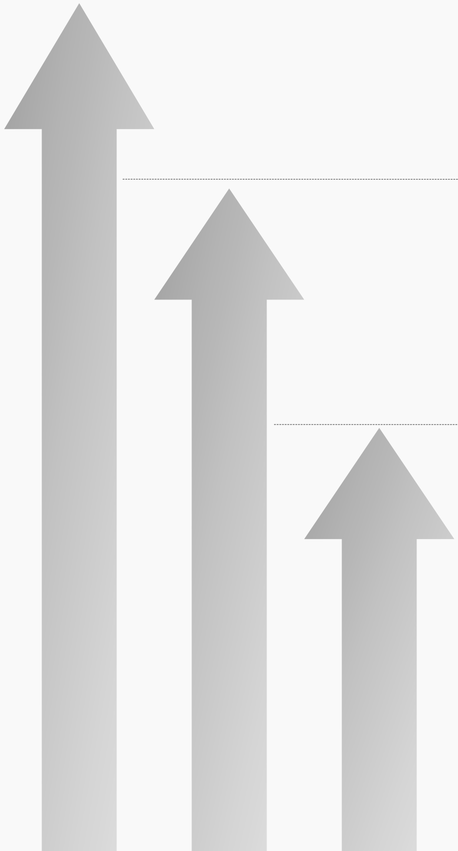
  **(e.g. geostrophic balance, T-S relationship, thermal wind relationship, etc.)**

- **Reducing inherent uncertainties**

  → Ensemble Prediction

  - **Reduction of noise by averaging many ensemble members**
  - **If number of ensemble members is infinite, unpredictable component will be zero in the ensemble mean**

# Issues about initial conditions



**True state + Perfect model**

**True state + Imperfect model**

**→ Is true state a best initial condition of imperfect forecast model?**

**Imperfect IC + Imperfect model**

**→ How to reduce uncertainties ?**

- **Obtain optimal initial condition**
  **-Data assimilation**
- **Reducing inherent uncertainties**
  **-Ensemble prediction**

# Is True state a best initial condition of imperfect forecast model?

**Experiments with Lorenz model**

$$\frac{dx_1}{dt} = -px_1 + px_2$$

$$\frac{dx_2}{dt} = rx_1 - x_1x_3 - x_2$$

$$\frac{dx_3}{dt} = x_1x_2 - bx_3$$

$$\Psi_{Initial\_Condition} = (1-k)\cdot\Psi_{Model} + k\cdot\Psi_{TRUE}$$

**K = nudging coefficient**

**Experimental Design**

**k=1 : TRUE state**

**k= 0.8 : 80% TRUE + 20% MODEL free run**

**k= 0.5 : 50% TRUE + 50% MODEL free run**

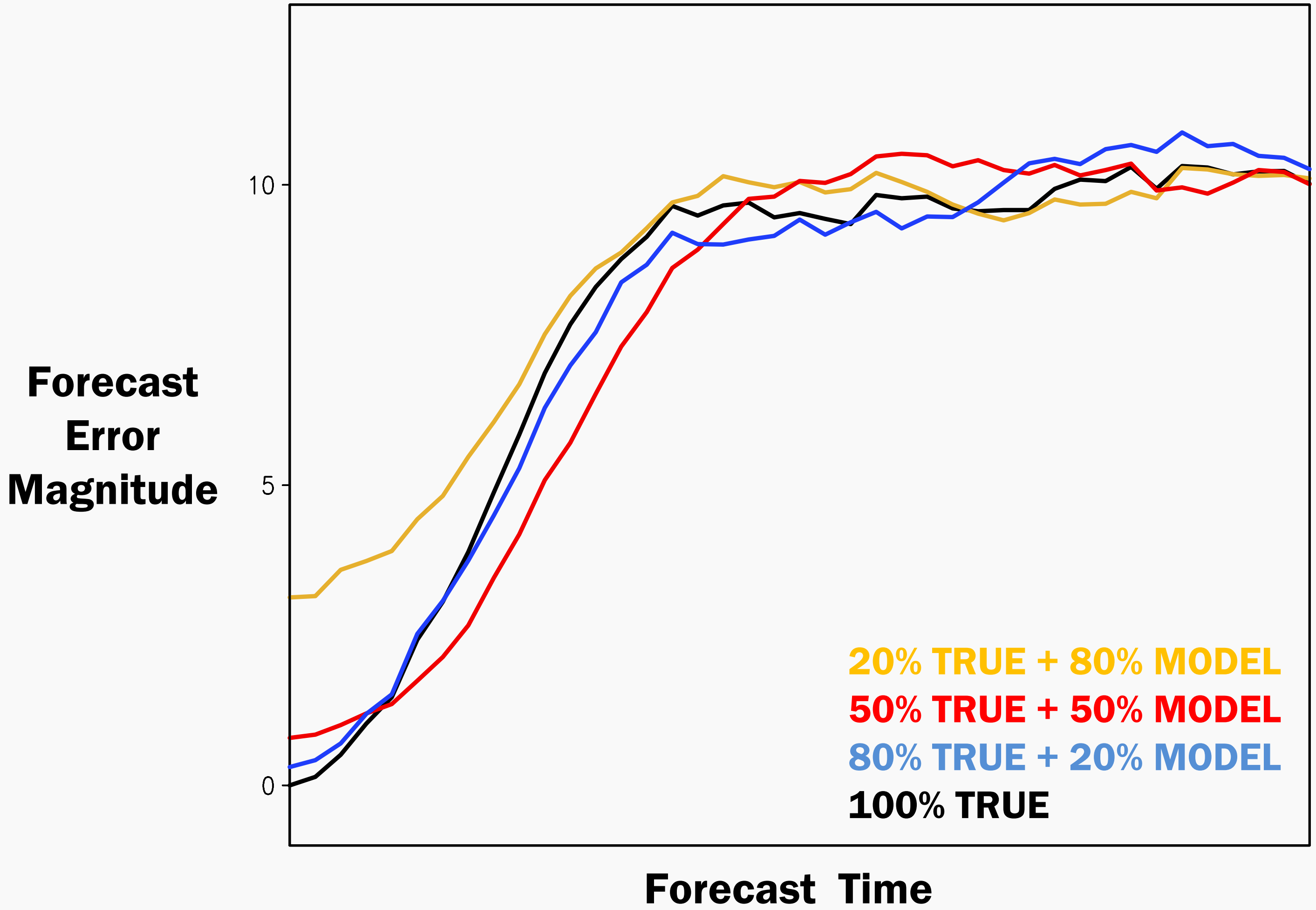
**k= 0.2 : 20% TRUE + 80% MODEL free run**

**r=28 : true world (perfect forecast model)**

**r=26 : Imperfect forecast model**

# Is True state best initial condition of imperfect forecast model?

**True world ≠ Forecast model (Imperfect forecast model)**




100 cases average

**True state is not always best initial condition of imperfect model**
**→ Physical balance of forecast model should be considered for optimal initial condition**

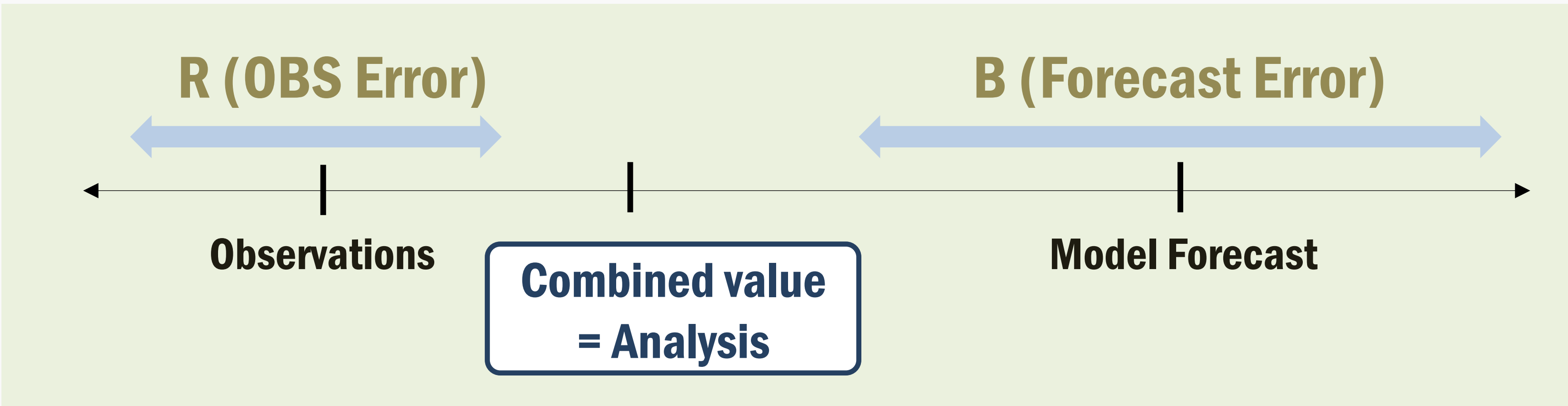
# Data assimilation

" Produce well-balanced initial condition through a
**statistical combination of**
**observations and short-range forecasts** "

$$\Psi_{Analysis} = (1-k)\cdot\Psi_{Model} + k\cdot\Psi_{OBS}$$

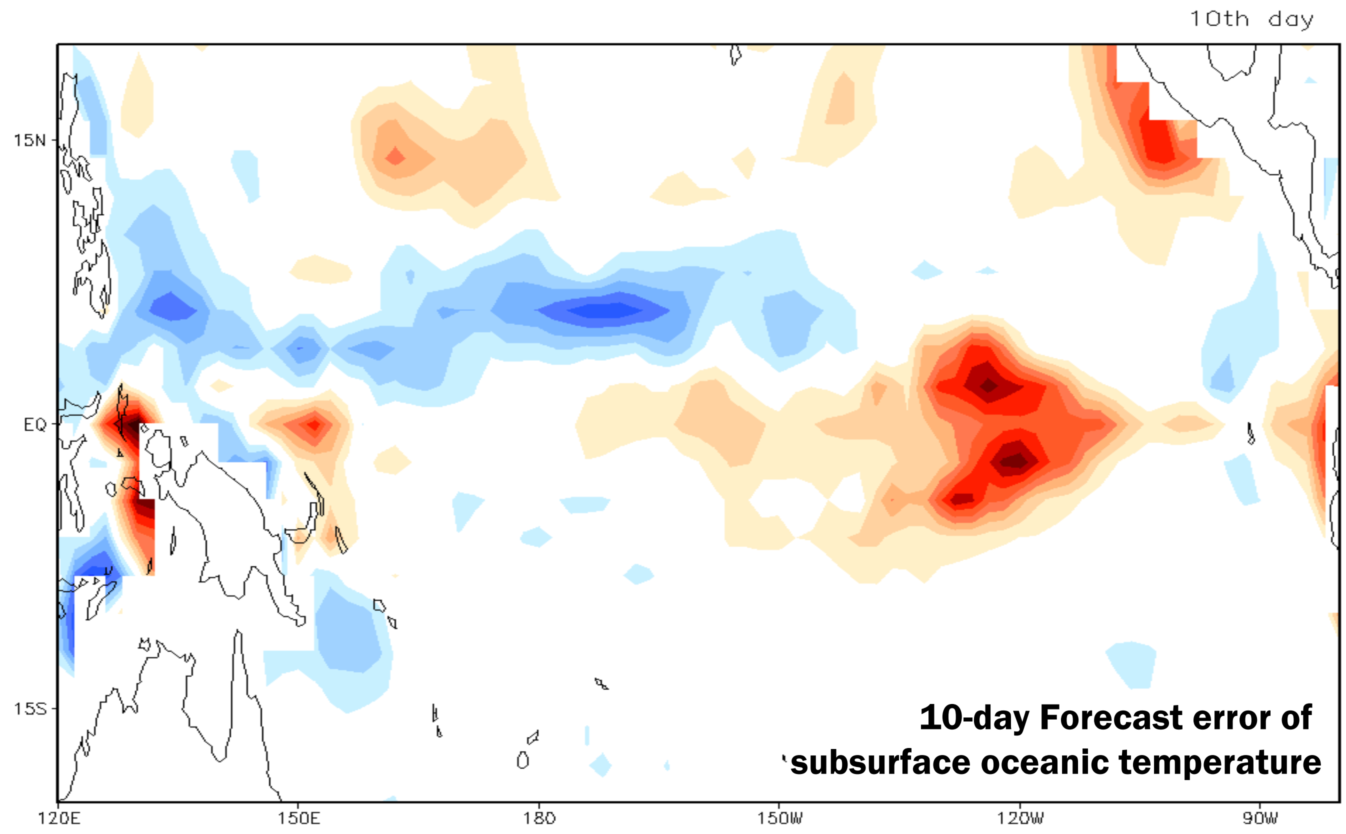
**Weighting coefficient (k)** $= \dfrac{\mathbf{B}}{\mathbf{R+B}}$

**B :Forecast error**
**R : OBS error**



# Characteristics of forecast error

**1. Non-local (spatially correlated)**
**2. Time variant**



# Forecast error is assumed as...

## 1. Nudging

- Local & Time invariant

$$\frac{\partial T}{\partial t} = -\bar{v} \cdot \nabla T + \frac{Q}{\rho C_P H} + \boxed{\frac{T_{obs} - T}{\tau_T}}$$

Nudging term in temperature equation

$\tau_T$ is constant

## 2. 3DVAR, Optimal Interpolation (OI)

- Non-local & Time invariant

$$\Psi_{Analysis} = (1-k) \cdot \Psi_{Model} + \boxed{k} \cdot \Psi_{OBS}$$

K = constant, but having a spatial structure

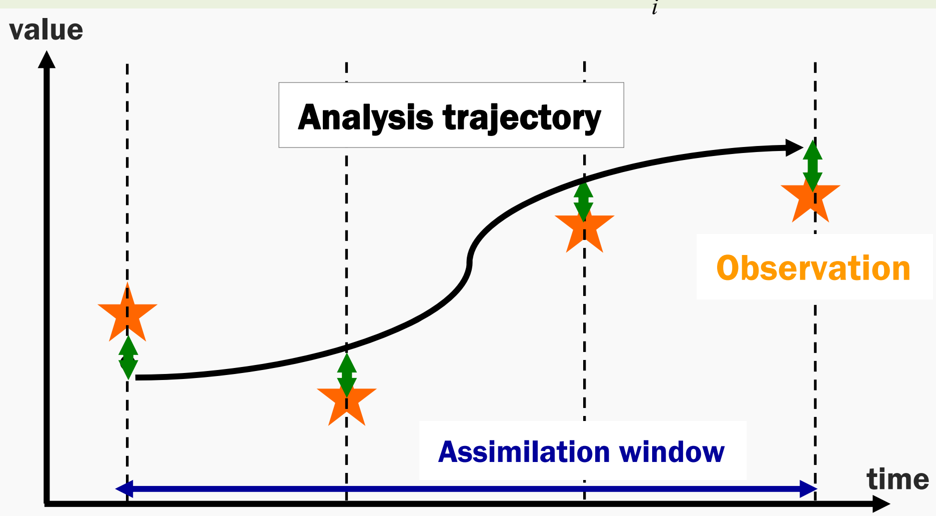
## 3. 4DVAR, Ensemble Kalman Filter

- Non-local & Time variant

# 4DVAR

**Find the initial condition its forecast best fits the observations within the assimilation window by minimizing cost function**

Cost function $= [x(t_0) - x^b(t_0)]^T B_0^{-1} [x(t_0) - x^b(t_0)] + \sum_i [y^o{}_i - H(x_i)]^T R_i^{-1} [y^o{}_i - H(x_i)]$




## Assumptions

1. **Causality :**
   The forecast model can be expressed as the product of intermediate forecast steps.

$$x_i = M_i M_{i-1} \ldots M_1 x$$

2. **Tangent linear hypothesis :**
   The cost function can be made quadratic by assuming that the M operator can be linearized.

$$y_i - H_i M_{0\to i}(x) \approx y_i - H_i M_{0\to i}(x_b) - H_i M_{0\to i}(x - x_b)$$

## Procedures

1. Integrate forecast model to get forecast values $x_{T+\alpha}$
2. Calculate difference between obs and model $y_{T+\alpha} - x_{T+\alpha}$
3. Integrate backward in time using adjoint model

$$\frac{\partial J(x)}{\partial x} = \sum_{i=1}^{\alpha} L_{T+\alpha\to T}{}^T R^{-1} (y_{T+\alpha} - x_{T+\alpha})$$

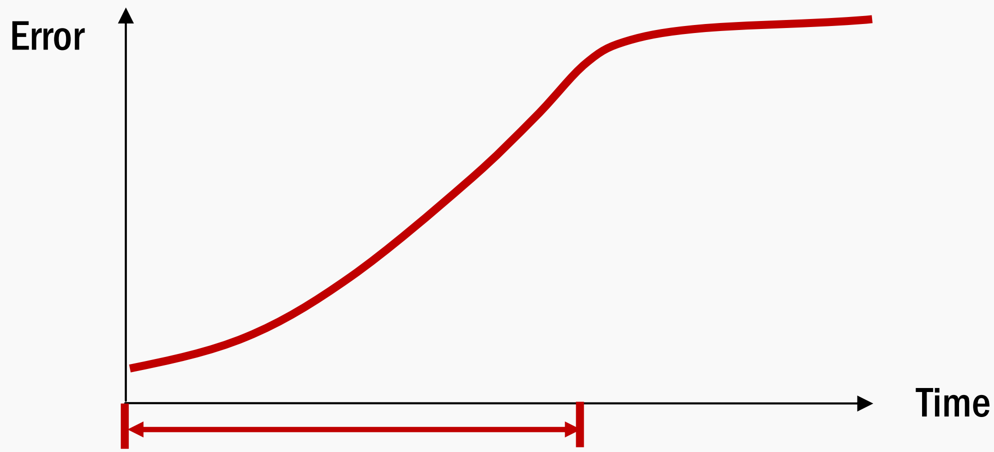
4. Gradient of cost function is used to determine the direction to search the minimum cost function

$$X_{new} = X_{old} - \frac{\partial J(x_{old})}{\partial x}$$

5. 1-4 procedure is repeated until find stable solution

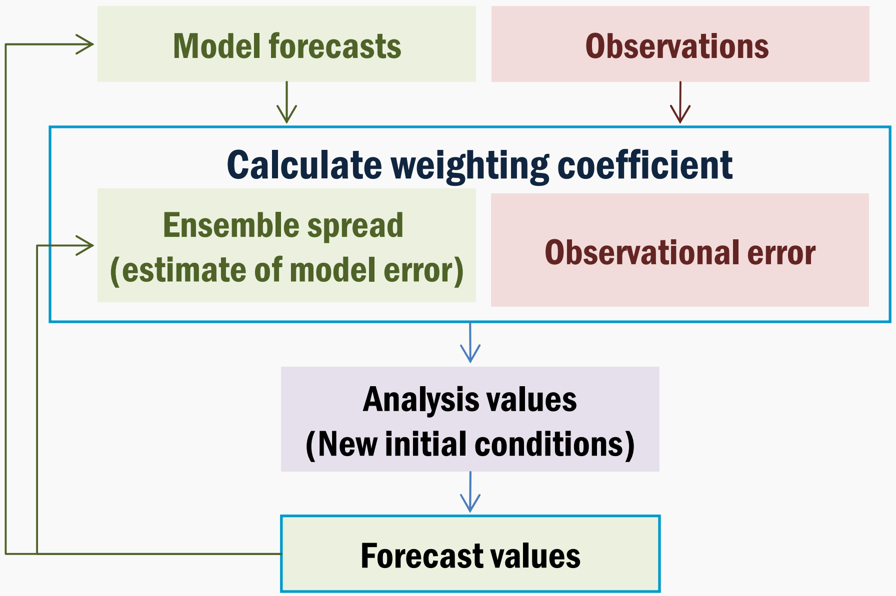
# Is 4DVAR applicable for seasonal prediction? : Linear assumption in CGCM

1. Is linearized CGCM dynamically meaningful?
   - Ocean physics is highly nonlinear
2. Is linear assumption valid in seasonal time scale?
   - Assimilation window is about several months for seasonal prediction
3. Problem in 4DVAR procedure
   - Forward integration with nonlinear model, Backward integration with linearized model } Inconsistency problem




Linear assumption should be valid during several month

# Ensemble Kalman Filter (EnKF)




## Procedures

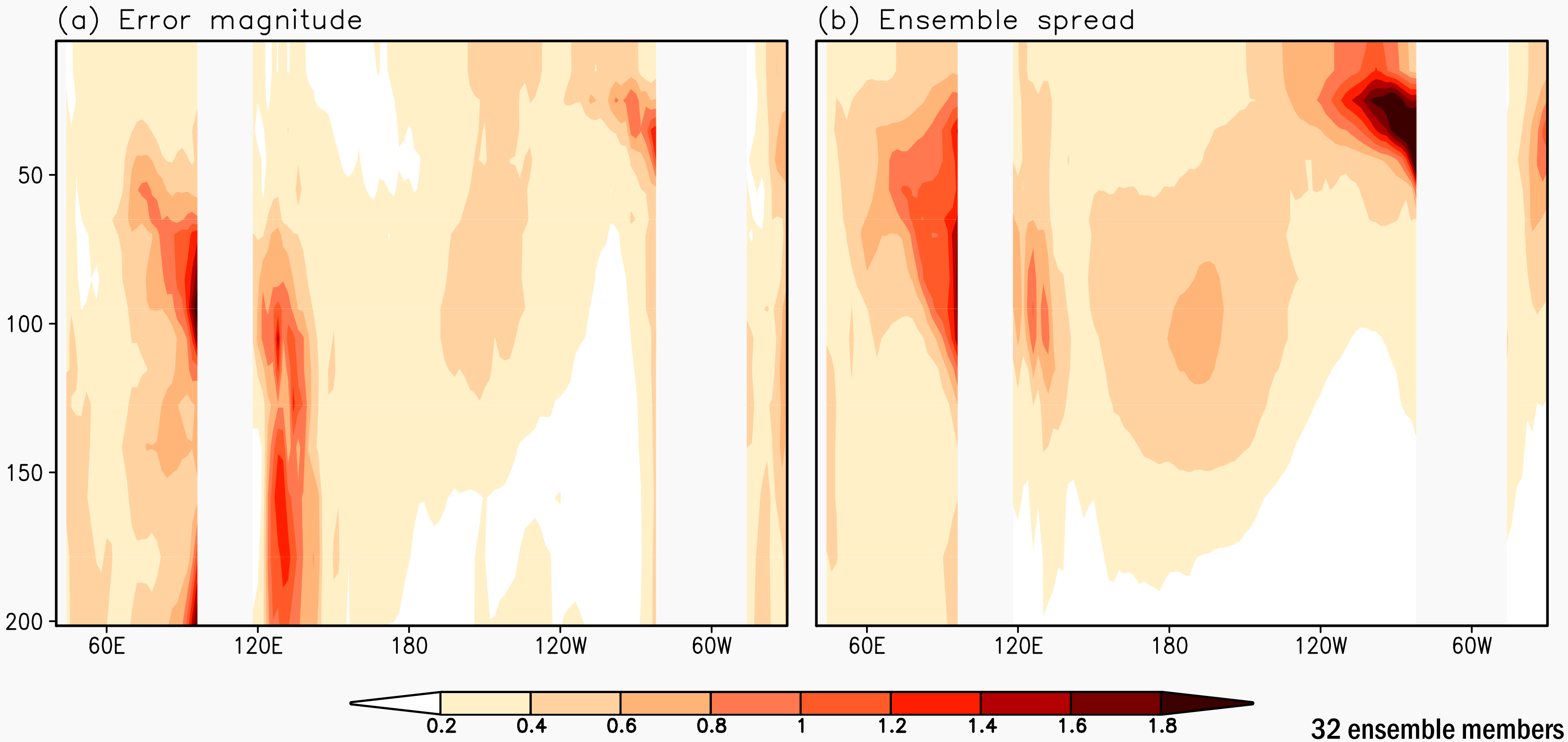
1. Integrate one cycle from random perturbation plus control initial condition
2. Calculate the forecast error covariance based on ensemble spread
3. Update the each ensemble states with analysis equation
4. Repeat 1-3 processes

## Assumptions

1. Ensemble spread is assumed as model forecast error
2. Model & observational error distribution is gaussian

# Forecast error & ensemble spread

## Equatorial Oceanic Temperature
## Error Magnitude & Ensemble Spread at 10 day forecast




32 ensemble members are used.

**PATTERN CORRELATION COEFFICIENT = 0.47**

Forecast error magnitude is represented by ensemble spread to a certain extent

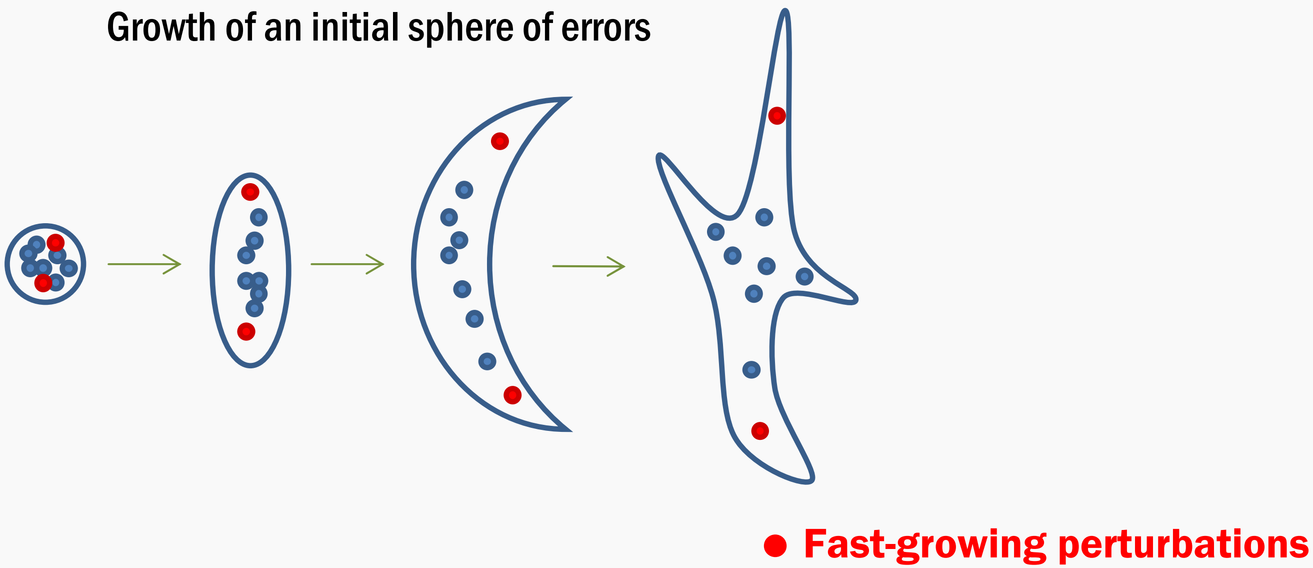
**<u>Is ensemble spread always appropriate tool to measure model forecast error?</u>**

# Optimal Perturbation Method for Ensemble Prediction

**Conditions for optimal Ensemble prediction**

1. Knowledge about initial error distribution
   → **Information about initial error distribution is never given**
2. Infinite ensemble members
   → **Only finite ensemble member is possible**

**Generate few ensemble members to efficiently describe forecast uncertainty → Fast-growing perturbations**



# Optimal Perturbation Method for Ensemble Prediction

**To extract fast growing initial perturbation**

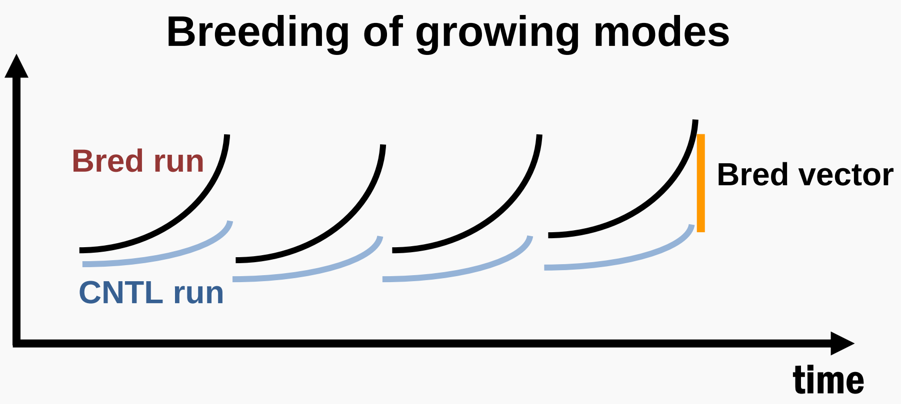
- **Breeding Method**
  - Repeat breeding and rescaling in the (nonlinear) model integration
  - Bred vector as a fast growing mode
  - NCEP for medium-range prediction

- **Singular Value Decomposition (SVD) Method**
  - Linear Stability of linearized model
  - Singular vector as a fast Growing mode
  - ECMWF for medium-range prediction

# Breeding method

## Procedure of Breeding Cycle

1. Integrate one cycle from random perturbation plus control initial condition
2. Calculate error by control run
3. Rescaling the error
4. Integrate one cycle from rescaled error perturbation plus control initial condition
5. Repeat 2-4 processes to initial forecast time




## Potential Benefit

- Ensemble prediction using Bred vector
- The bred vector represents fast growing mode of the model and nature

# *Is Bred Vectors applicable for seasonal prediction?*

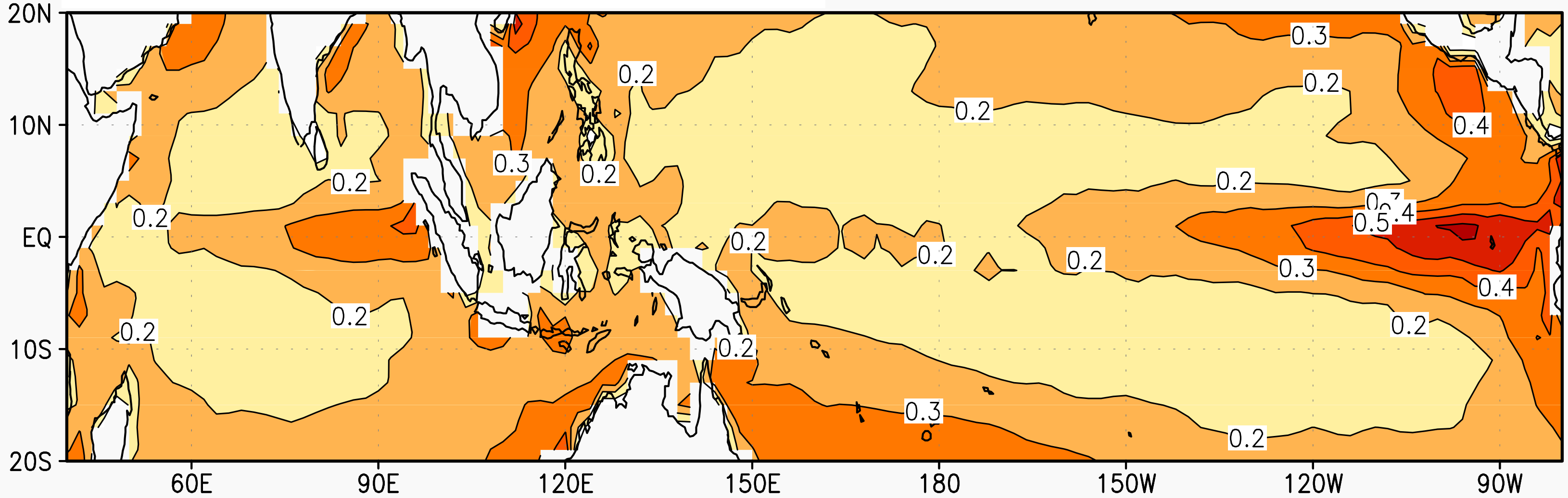
❑ **Rescaling time interval : 1 month**

**- To catch up longer time-scale variability**

❑ **Selection of Norm : RMS of Monthly mean SST over Indo-Pacific Region**

**- To reduce effects of fast atmospheric fluctuation because SST is slowly varying.**

**STD of Bred Vector SST in SNU CGCM**



## Singular Vector method

$$Let\ \mathrm{X} = \Psi(\mathrm{t}),\ \mathrm{Y} = \Psi(\mathrm{t} + \tau)$$

$$\mathrm{Y} = \mathrm{L} \cdot X$$

**By solving singular value of L operator**

$$\mathrm{L} = \mathrm{U}SV^{T}$$

$$\mathrm{UY} = SV\mathrm{X}$$

$$u_i \cdot \mathrm{Y} = \mathrm{s}_i v_i X$$

**If $S_i$ > 1, then $V_i$ is an initial perturbation of growing mode**
**$U_i$ is a final perturbation of growing mode**

**For maximum $S_i$ : fast growing perturbation**

# Singular Vector method

**SVD Method : linear stability of linearized model**

**- Usually linearized model is not available, especially for fully coupled system**

**Empirical linear operator is formulated using time-lag relationship**

**Tangent linear operator**

$$Y = L \cdot X$$

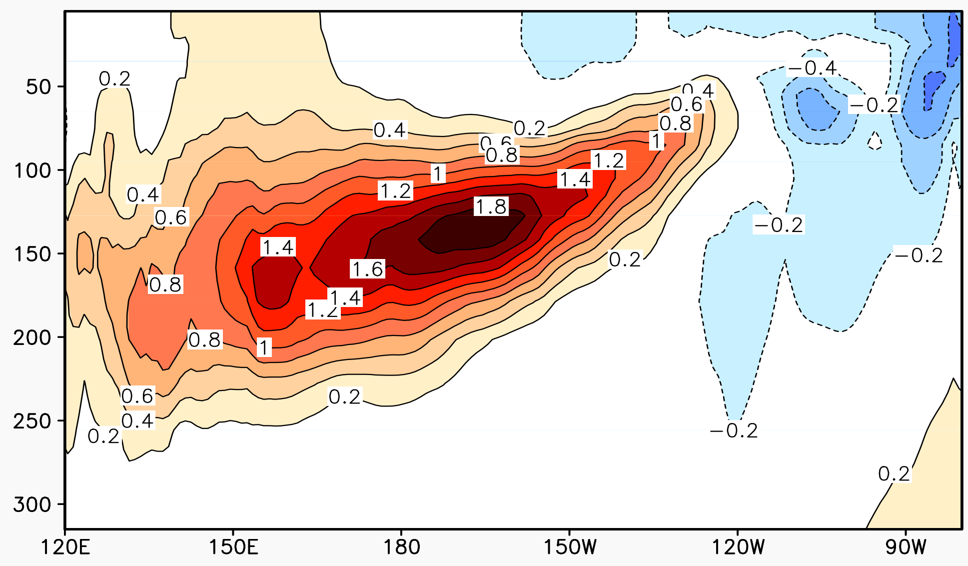
$$L = YX^{T}(XX^{T})^{-1}$$

Variance of X

Covariance between X and Y

**X: 5 EOF modes of thermocline depth**
**Y: 5 EOF modes of SST after 6-month**

**Equatorial Temperature perturbations at May 1st**



# Thank you