The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

SMR/1849-19

Conference and School on Predictability of Natural Disasters for our Planet in Danger. A System View; Theory, Models, Data Analysis

25 June - 6 July, 2007

Seasonal Prediction & Predictability from Signal to Noise

In-Sik Kang

Climate and Environment System Research Center
Seoul National University, Korea

Seasonal Prediction & Predictability

from Signal to Noise

In-Sik Kang

Climate and Environment System Research Center
Seoul National University, Korea

lecture flows

Part I

Overview about Seasonal Prediction & Predictability

- Potential predictability – Signal to Noise ratio
- Real predictability – Tier-2 system
- Coupled model predictability – Tier-1 system

Part II

Predictability in air-sea coupled system

- Air-sea interaction
- Tier-1 vs Tier-2 prediction system
- Predictability of various coupled models

Part III

Access to upper limit predictability

- Error correction
- Multi Model Ensemble Prediction
- Noise dynamics

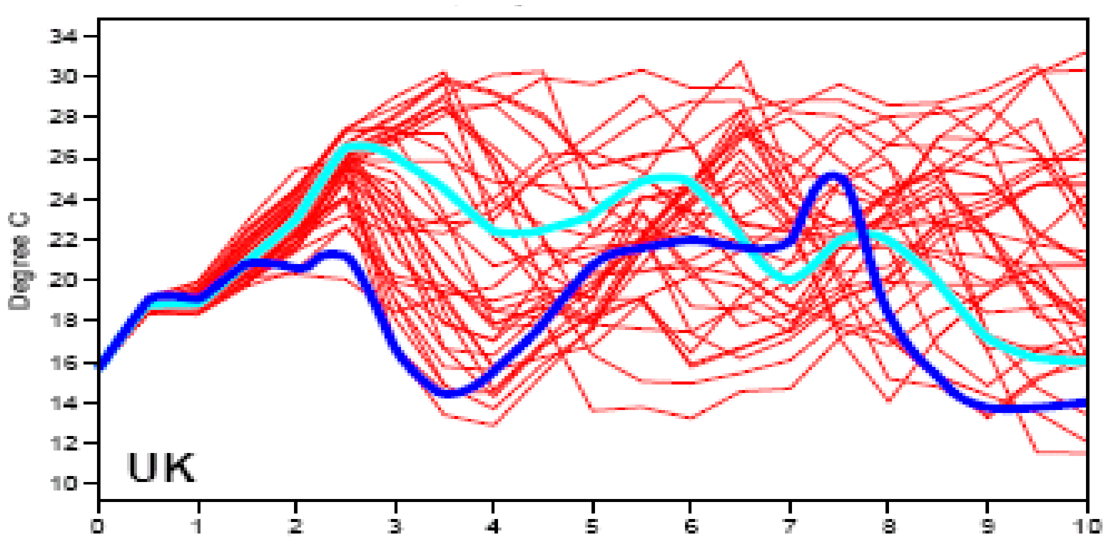
Part I.

Overview of Seasonal Prediction & Predictability

The Nature of the Seasonal Prediction

Uncertainty of seasonal prediction

A Realization = Signal + Noise



Ensembles from
Small Initial Perturbation

Ensemble
Ensemble Mean
Observation

The Nature of the Seasonal Prediction

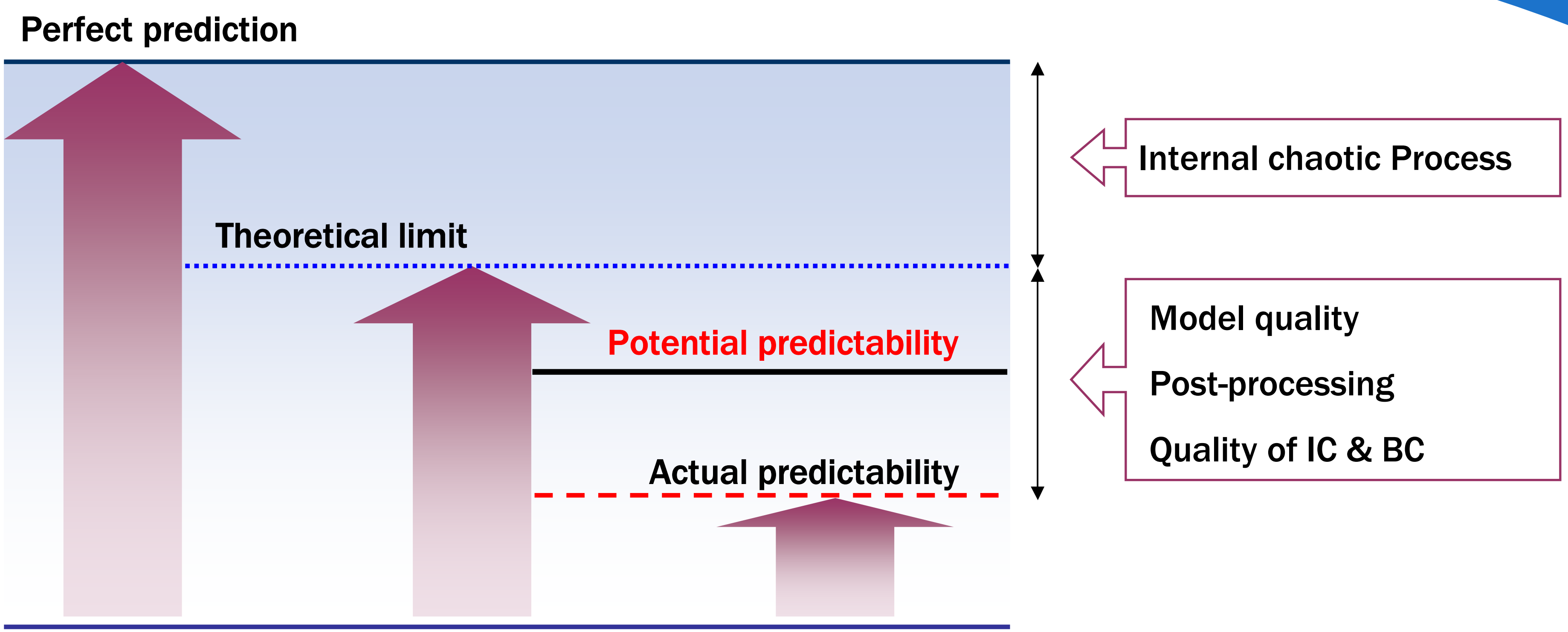
Uncertainty of seasonal prediction

A Realization = True + Error + Noise

- **Uncertainty from Initial Condition**
- **Uncertainty from Model Physics and Dynamics**

Imperfectness of Model → Systematic Error

Predictability of Seasonal Prediction



In climate prediction, potential predictability is regarded as the predictability with full information of future boundary condition (e.g., SST). Thus, predictability is varied with similarity between the response of real atmosphere and prediction method to the same BC.

Establish "potentially" possible prediction skill with state-of-art prediction system

Decomposition of climate variables

Climate state variable (X) consists of predictable and unpredictable part.

- **Predictable part = *signal* (Xs) : forced variability**
- **Unpredictable part = *noise* (Xn) : internal variability**

$$X = Xs + Xn$$

The dynamical forecast (Y) also have its forced and unforced part.

forecast *signal* (Ys) : forced variability of model

forecast *noise* (Yn) : internal variability of model

$$Y = Ys + Yn$$

The internal variability (noise) is stochastic

If the forecast model is not perfect, Xs≠Ys. (there is a systematic error)

Upper limit of prediction

Maximizing correlation in the presence of error in signal and noise

$$Observation : x = x_s + x_n$$

$$Forecast : y = y_s + y_n = \alpha x_s + y_e + y_n$$

Noise y_n and Error y_e are not correlated with others.

α : regression coefficient of signal

$$Cov(x, y) = (\overline{xy}) = \overline{(x_s + x_n)(\alpha x_s + y_e + y_n)}$$

$$= (\alpha \overline{x_s^2} + \alpha \overline{x_s x_n} + \overline{x_s y_e} + \overline{x_n y_e} + \overline{x_s y_n} + \overline{x_n y_n})$$

$$= \alpha V(x_s) + \alpha Cov(x_s, x_n) + Cov(x_s, y_e) + Cov(x_n, y_e) + Cov(x_s, y_n) + Cov(x_n, y_n)$$

$$= \alpha V(x_s)$$

Correlation between observation (x) and forecast (y)

$$\mathrm{Cor}\,(x,y) = \frac{\alpha V(x_s)}{V(x)^{1/2}\left[\alpha^2 V(x_s) + V(y_e) + V(y_n)\right]^{1/2}}$$

The correlation coefficient is maximized by removing *$V(y_e)$* and *$V(y_n)$*

➔ The most accurate forecast will be the *SIGNAL* of perfect model.

Maximum prediction skill : potential predictability

When the forecast produces a perfect signal, the correlation coefficient is

$$\mathrm{Cor(x,y)}= \frac{\alpha V(x_s)}{V(x)^{1/2}\left[\alpha^2 V(x_s)+V(y_e)+V(y_n)\right]^{1/2}} = \frac{V(x_s)}{\left[V(x)V(x_s)\right]^{1/2}}$$

$$= \frac{V(x_s)^{1/2}}{V(x)^{1/2}}$$

$$\sqrt{\frac{V_{Signal}}{V_{Total}}} = \sqrt{\frac{\rho}{1+\rho}}, \quad \rho = \frac{V_{Signal}}{V_{Noise}} \equiv SNR$$

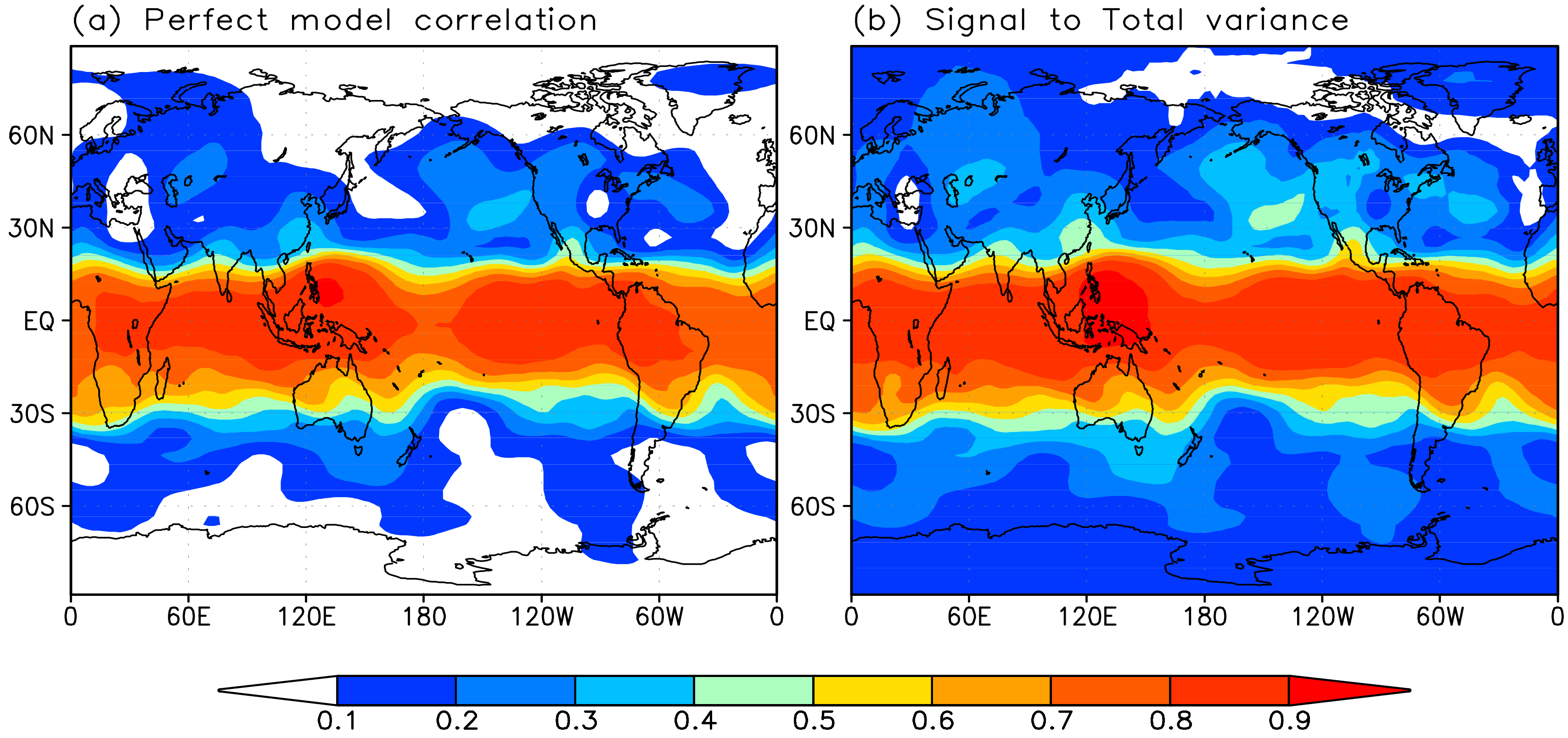
$$\rho : Signal\ to\ Noise\ Ratio$$

$$V_{Total} = V_{signal} + V_{noise}$$

Maximum prediction skill (=potential predictability of particular predictand) is a function of Signal to Noise Ratio

Perfect model correlation & Signal to Total variance ratio

Z500 winter (C20C, 100 seasons, 4 member)



Although the 4 member is not enough to estimate Potential predictability precisely, the patterns of 2 metrics are quite similar

Strategy of Prediction

The strategy of seasonal prediction is to obtain "perfect signal" as close as possible.

(i.e. reducing variance of systematic error and variance of noise)

1. Reduction of Noise

- **Averaging large ensemble members**

(if number of ensemble members is infinite, Noise will be zero in the ensemble mean)

2. Correct signal

- **Improving GCM**
- **Statistical post-process (MOS)**
- **Multi-model ensemble**

Contents

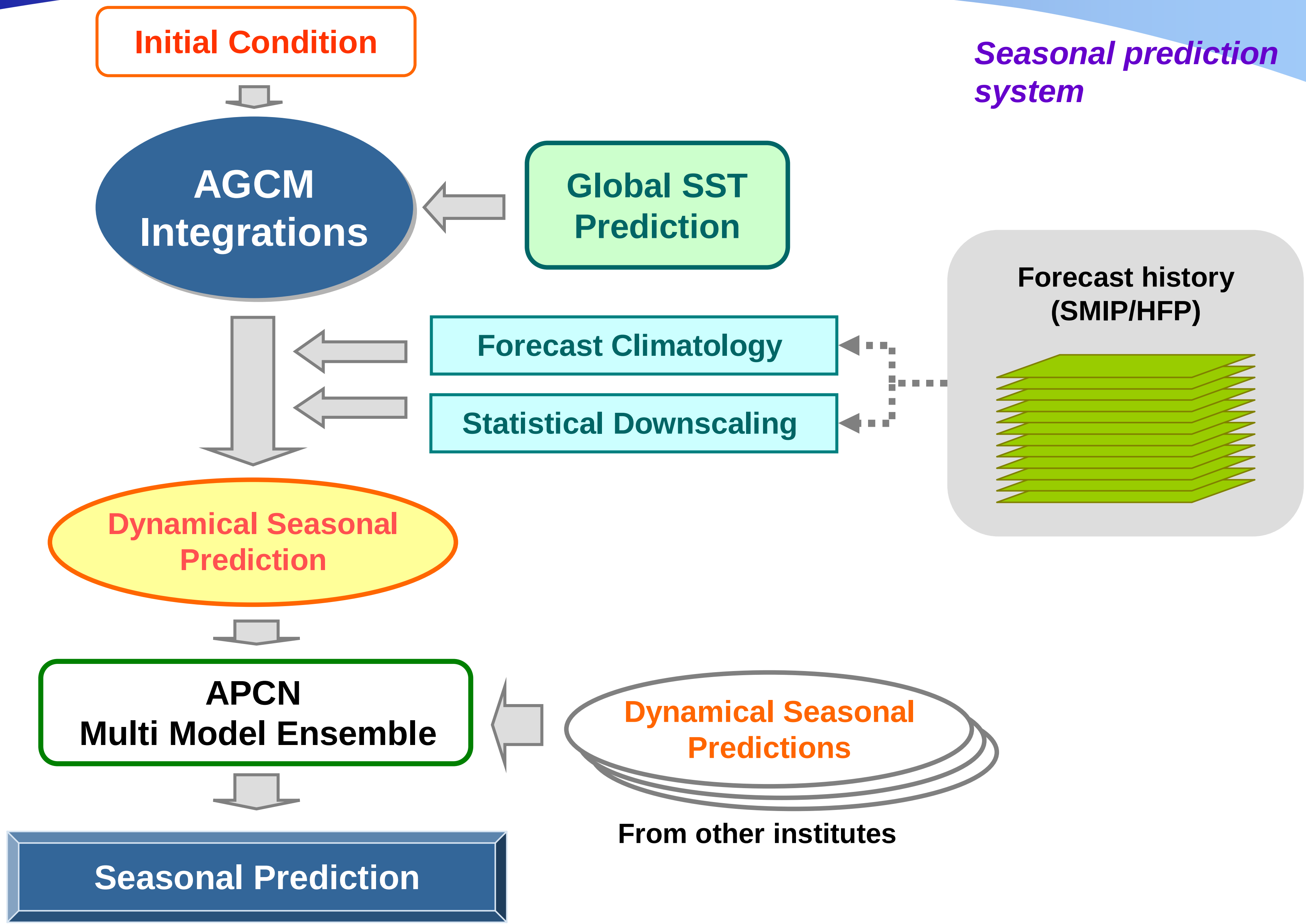
Overviews about seasonal predictability

1. Potential predictability – Signal to Noise ratio
2. Real predictability – Current Tier-2 system
3. Coupled model predictability – Tier-1 system

Scientific issues

Maximize signal	Predict noise
1. Coupled processes 2. Error Correction 3. Multi-model ensemble	Noise dynamics

Multi-model Seasonal Prediction (Two-Tier system)



SMIP project

Seasonal prediction Model Intercomparison Project

- **Organized by**

World Climate Research Programme
Climate Variability and Predictability Programme (CLIVAR)
Working Group on Seasonal to Interannual Prediction (WGSIP)

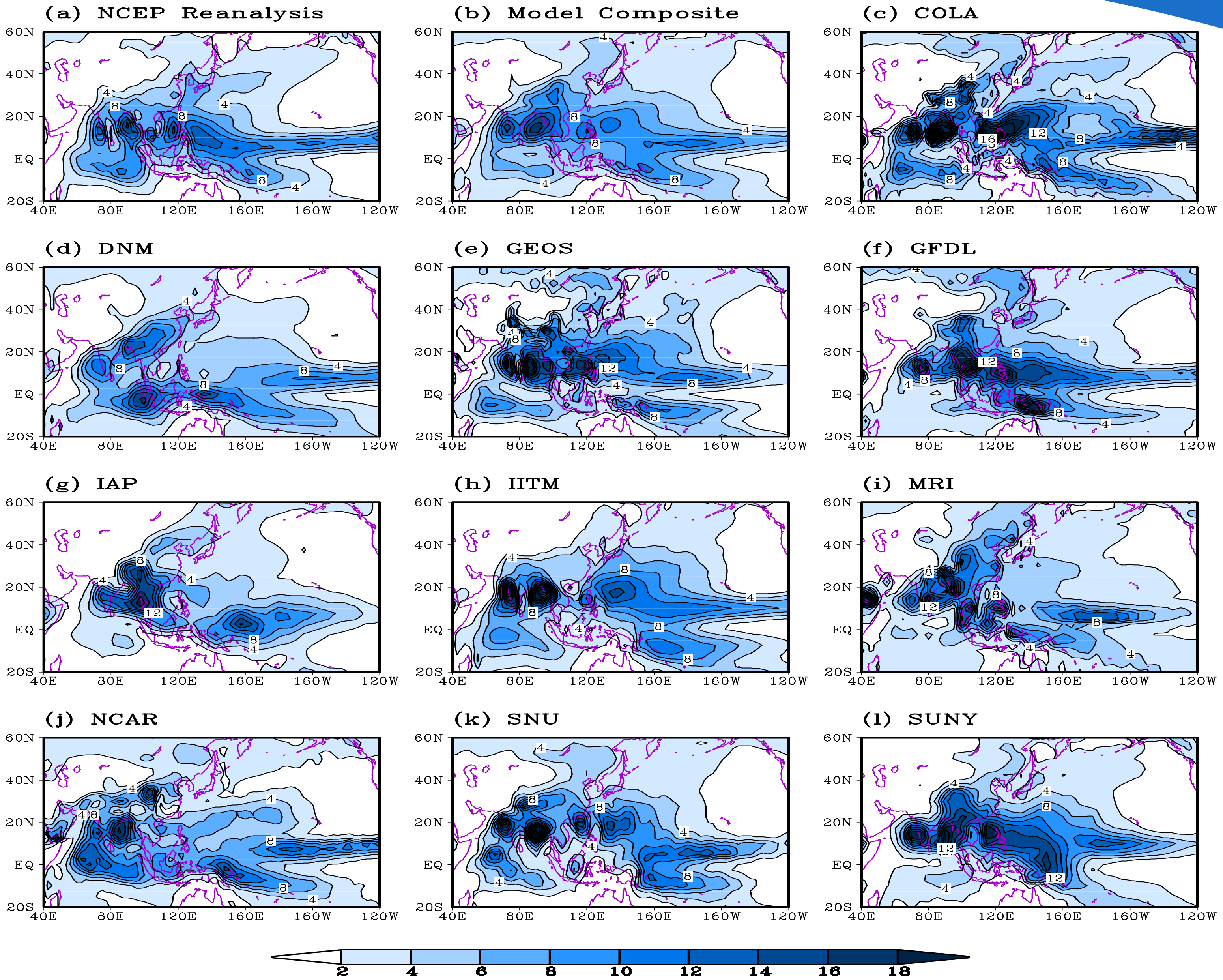
- **Coordinators**

G. Boer(CCCma), M. Davey (UKMO), I.-S. Kang (SNU), and K. R. Sperber (PCMDI)

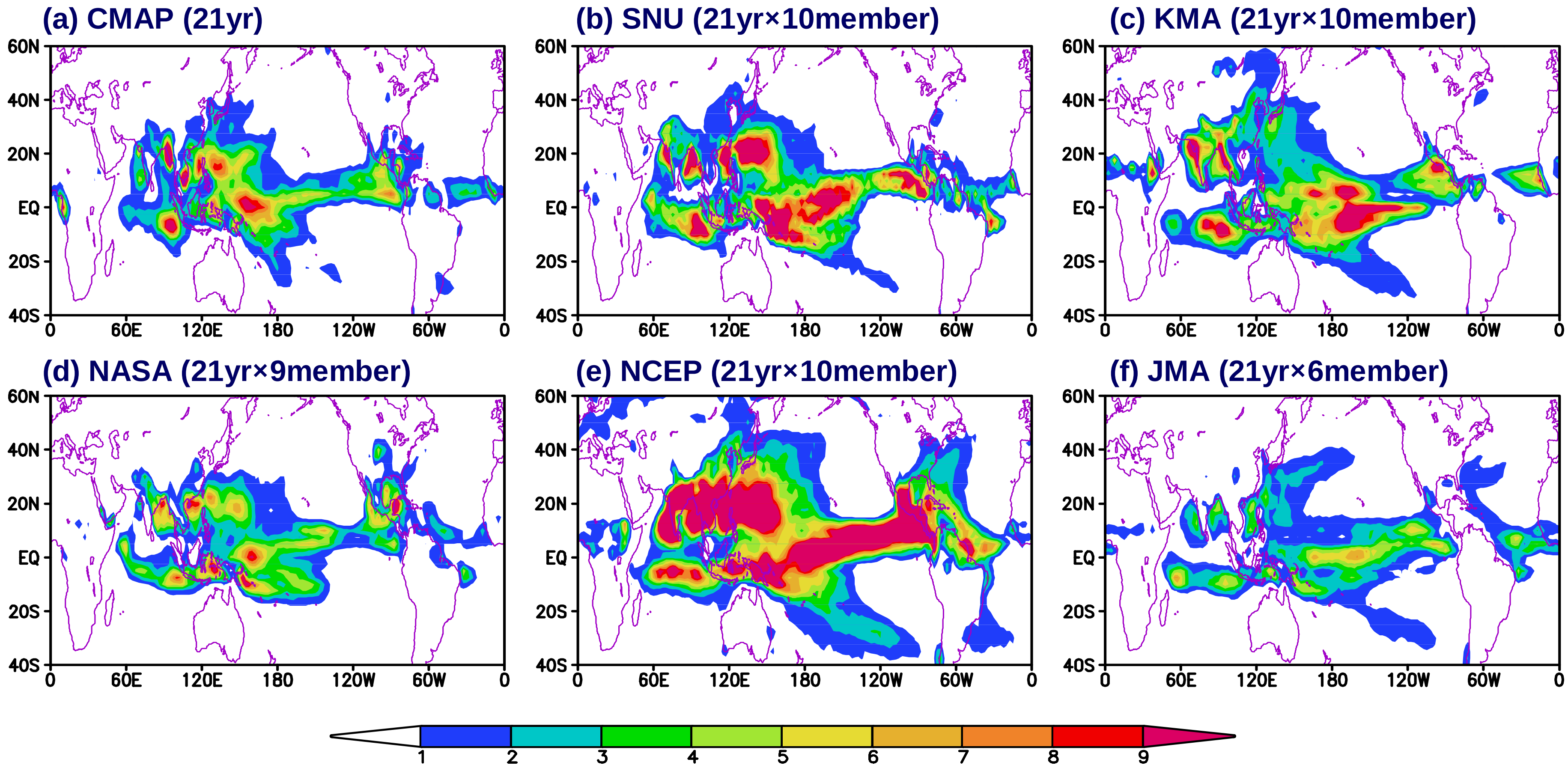
Purpose

Investigate 1 or 2 season potential predictability based on the initial condition and observed boundary condition

Monsoon Predictability: Climatological JJA Precipitation



Total Variance of JJA Precipitation Anomalies



Forced & Free variance

Forced variance

Climate signals

caused by external forcing (e.g. SST)

$$\frac{1}{N-1}\sum_{i=1}^{N}(\overline{X}_i - \overline{\overline{X}})^2 - \alpha$$

$\alpha = \frac{1}{n}\cdot$ **Free variance**

☐ **Ensemble mean variation with respect to time**

Free variance

Intrinsic transients

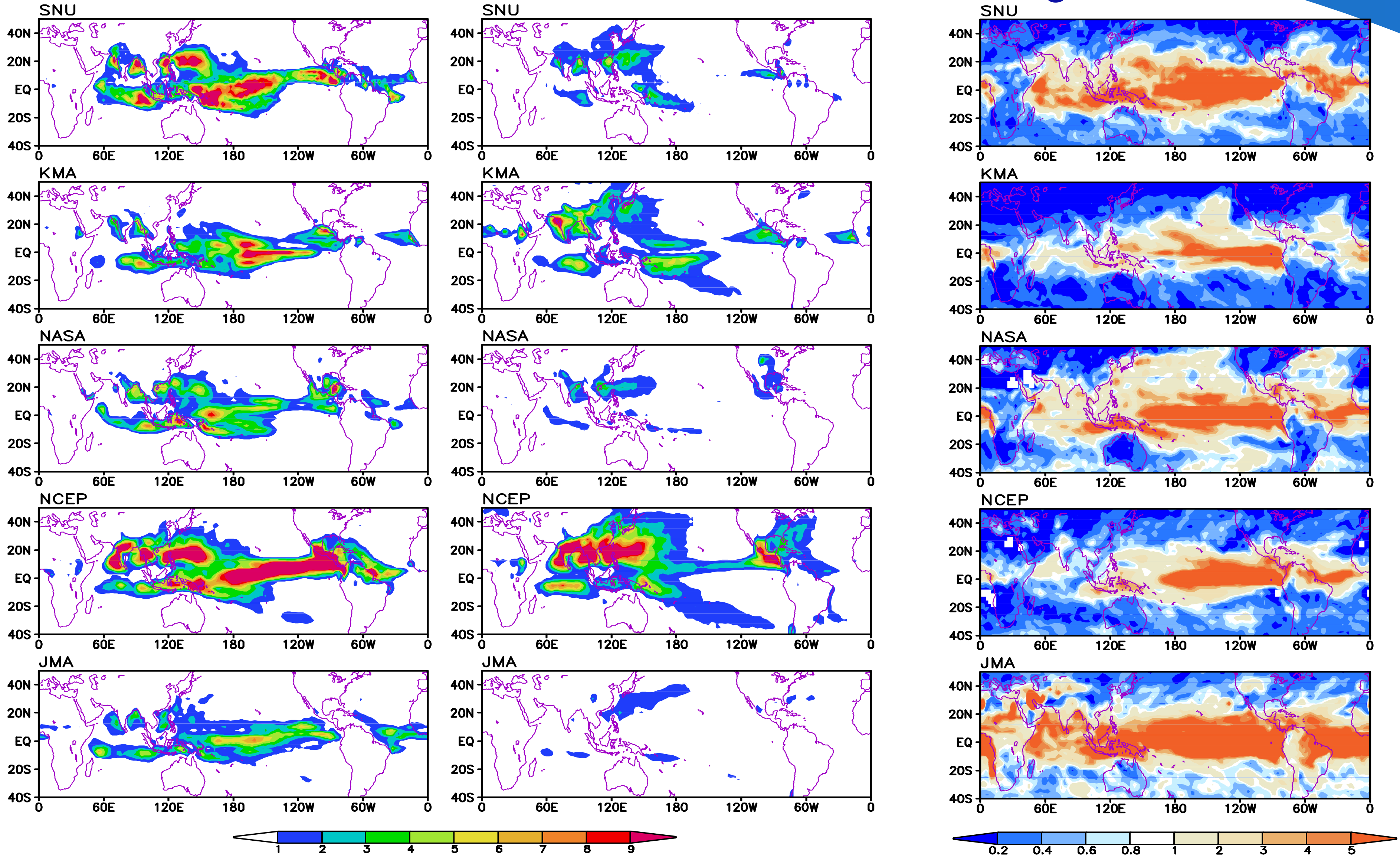
due to natural variability

$$\frac{1}{N(n-1)}\sum_{i=1}^{N}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2$$

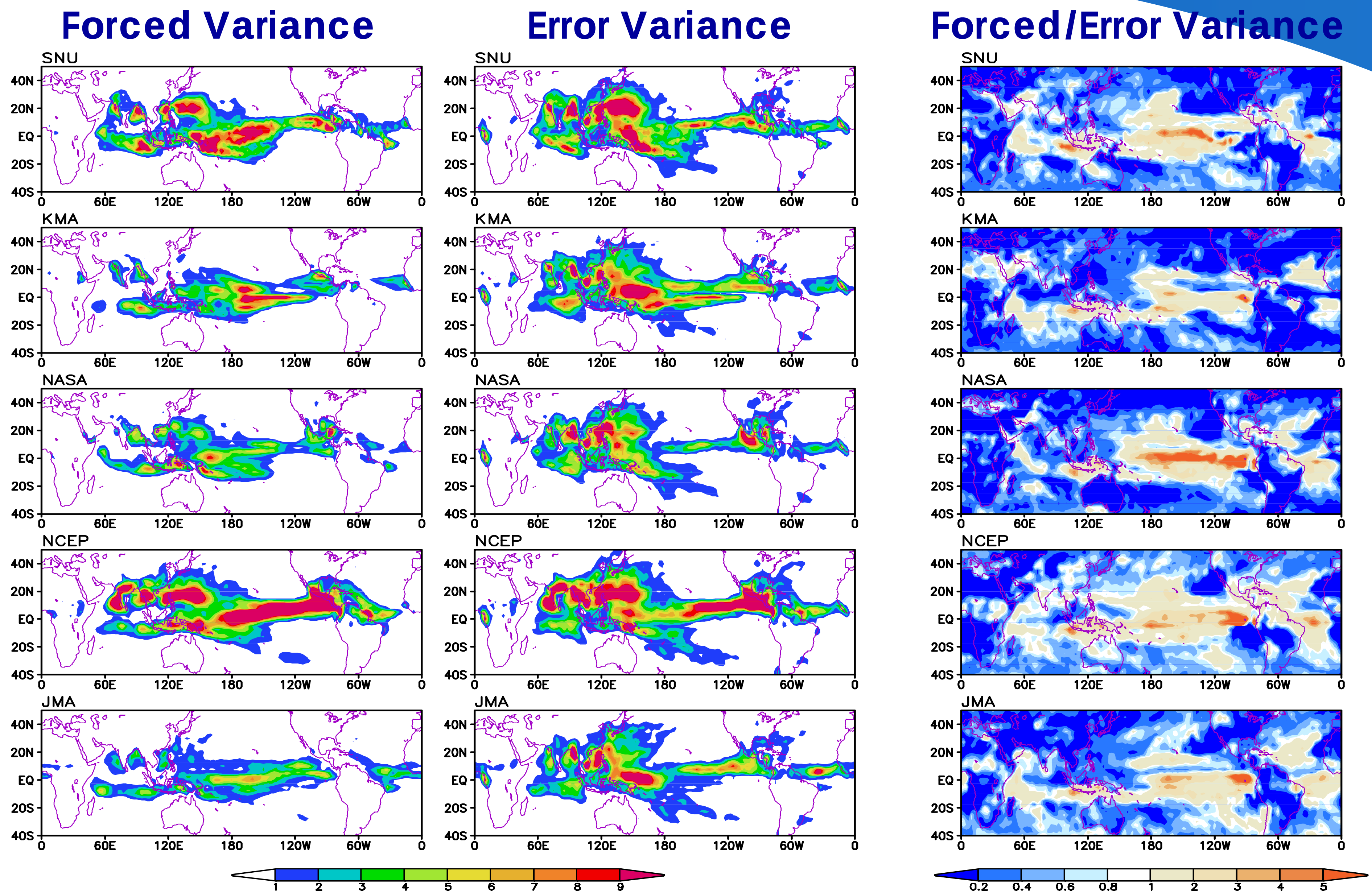
Ensemble spread

Variance analysis of JJA Precipitation Anomalies

Forced Variance | Noise Variance | Signal/Noise

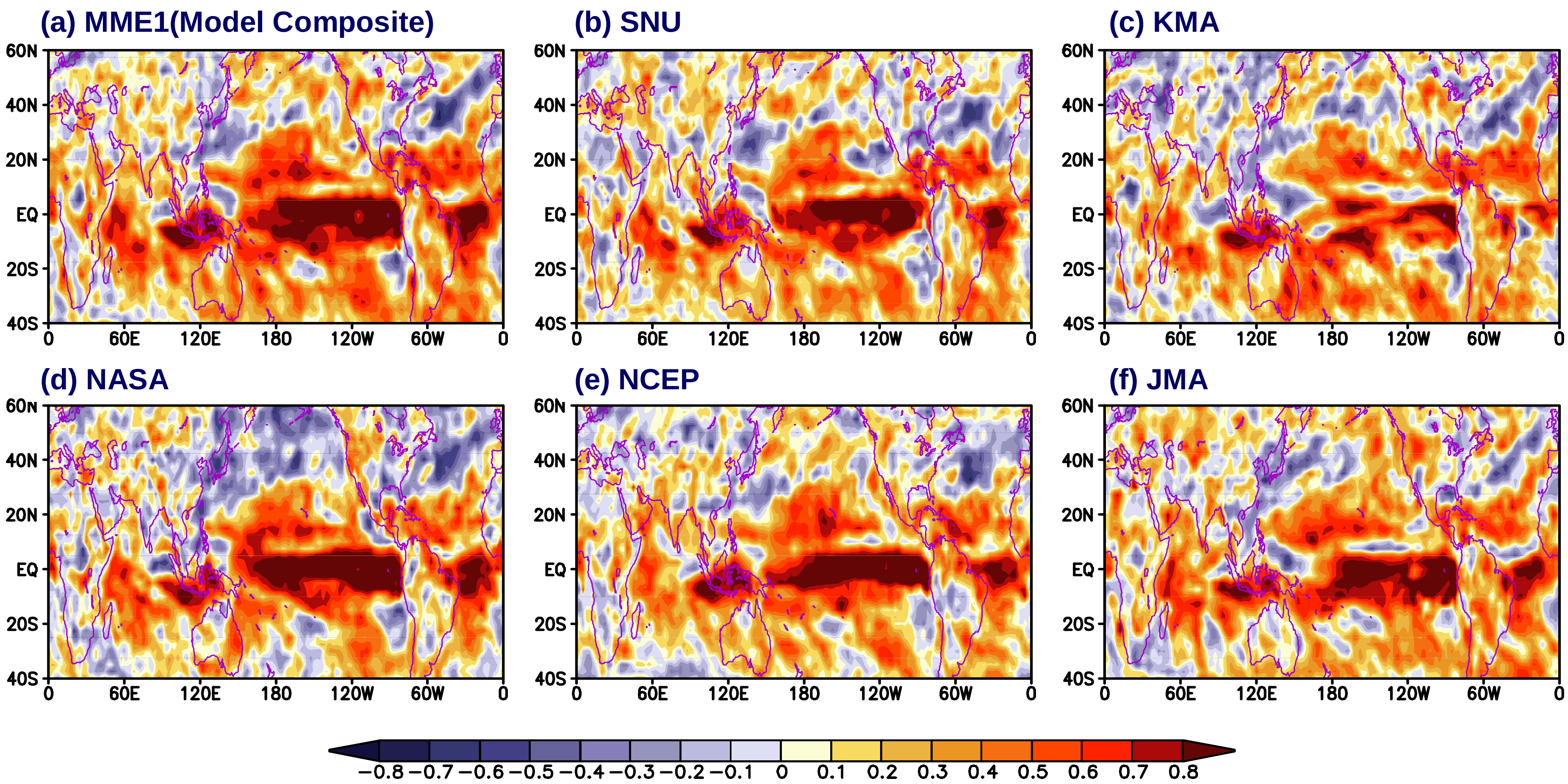


Variance analysis of JJA Precipitation Anomalies



Prediction Skill of JJA Precipitation (21 yr)

Temporal Correlation



How is real predictability?

SMIP/HFP (Historical Forecast Project)

Investigate 1 season real predictability based on the observed initial condition and predicted boundary condition

Atmosphere

↑ ↑ ↑ ↑ ↑ ↑

Boundary Condition

Predicted SST

CES Global SST Prediction System

❖ In the 2-tier climate prediction system, global SST forecasts are required for boundary conditions of AGCM in seasonal climate prediction.

❖ The Ensemble Global SST Prediction System was developed for the seasonal climate prediction

Combined System of Dynamical and Statistical Models

Dynamical Model

- Intermediate El Nino Prediction model

+

Statistical Model

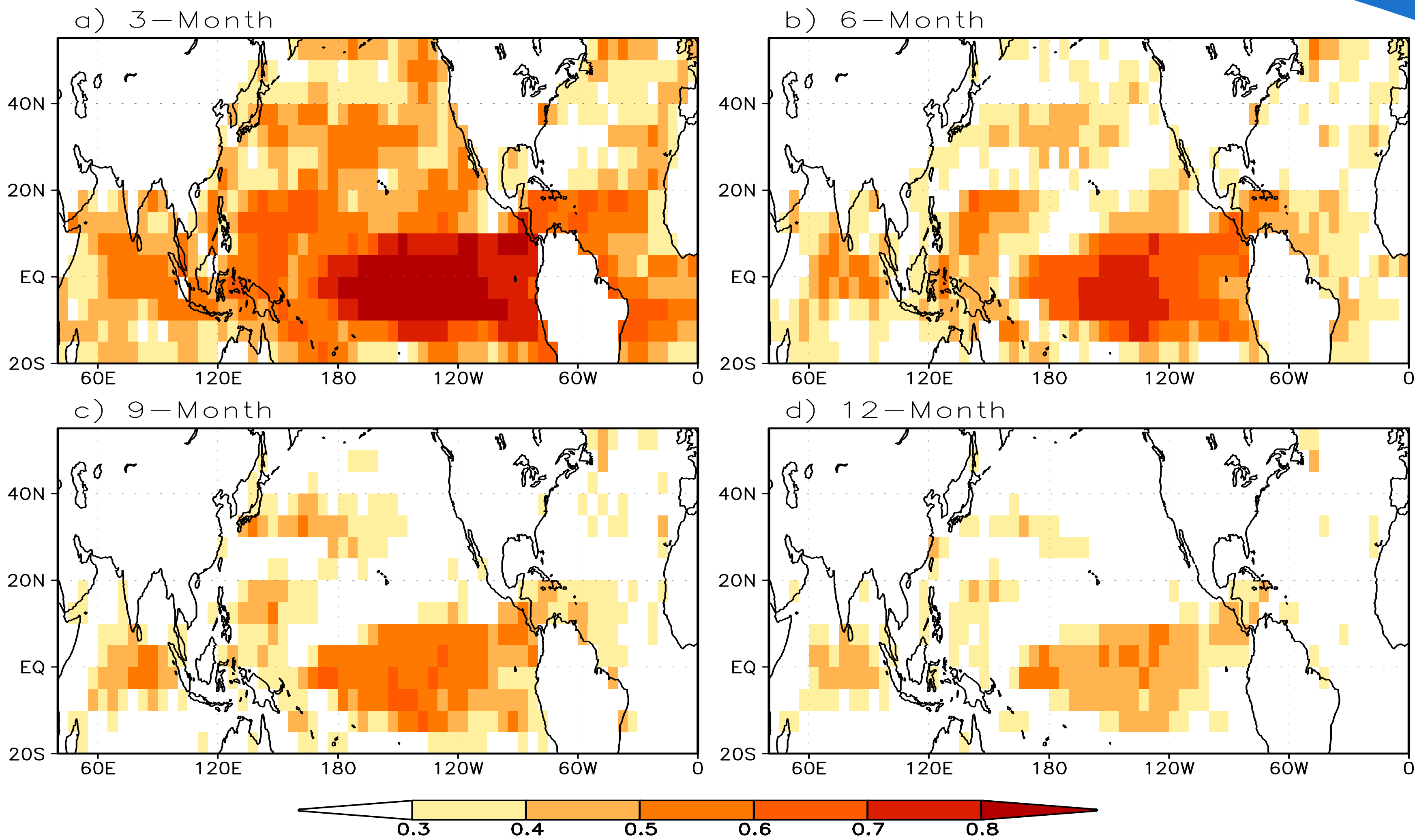
- Lagged Linear Regression Model (LLR)
- Pattern Projection Model (PPM)

Ensemble Procedure

Global SST Prediction

- Boundary Condition of AGCM
- Priori indication of Regional climate Variability

Correlation Skill of SST Ensemble Prediction

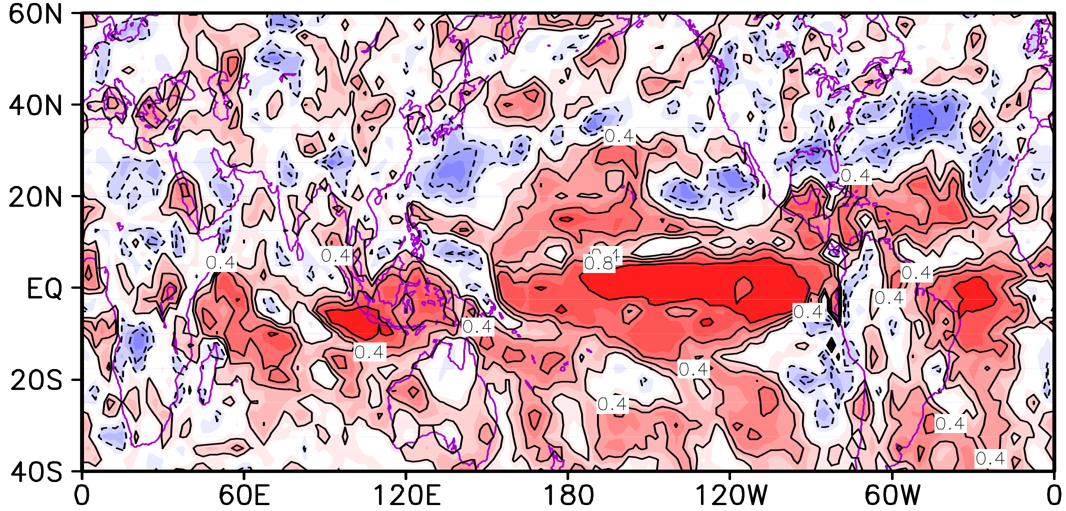


AGCM prediction system

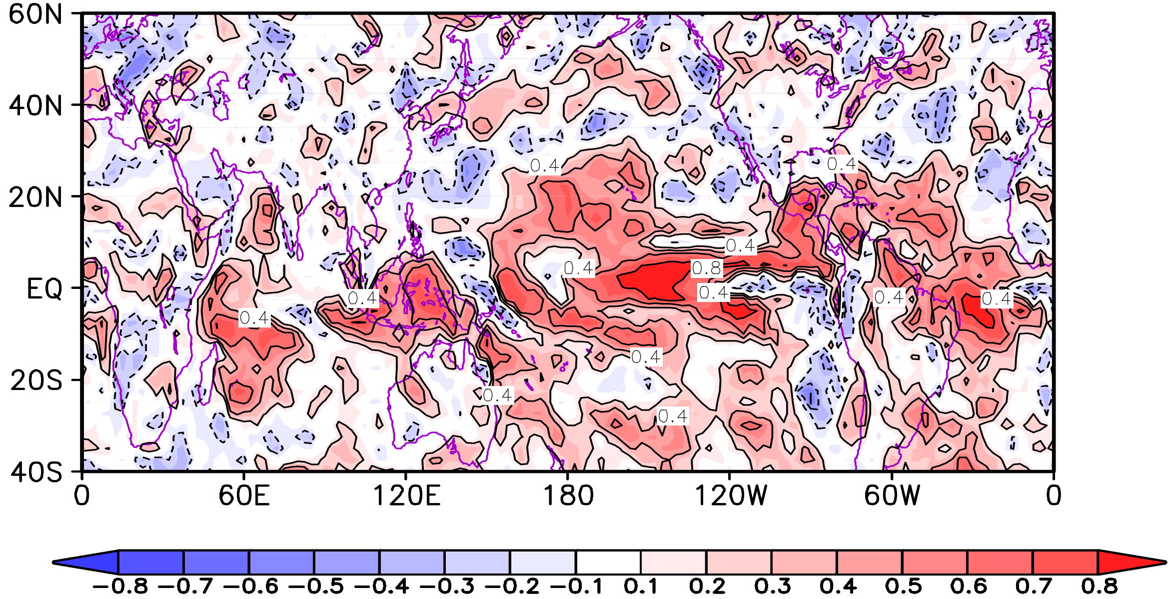
Institute	AGCM	Resolution	Ensemble
FSU	RSUGCM	T63 L27	10
GFDL	AM2	2.0lat x 2.5lon L24	10
SNU	GCPS	T63 L21	6
CAM2/UH	CAM2	T42 L26	10
ECHAM/UH	ECHAM4	T31 L19	10

Temporal Correlation of summer PRCP

SMIP



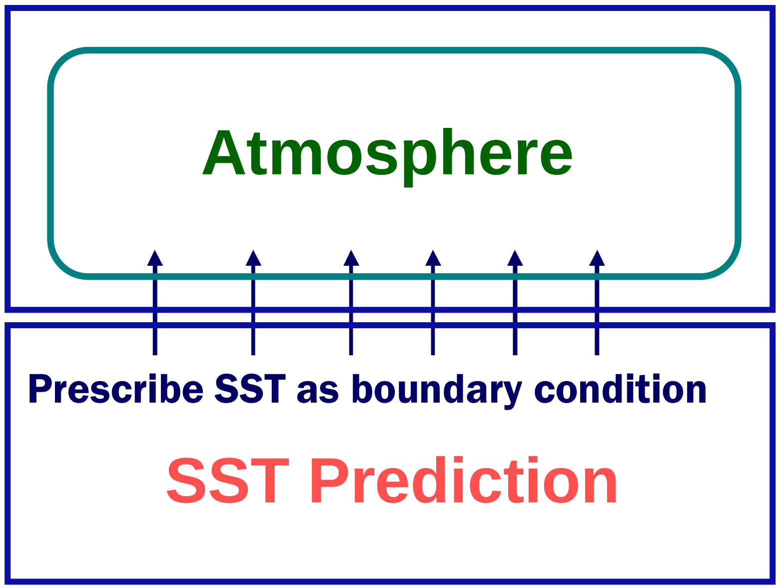
SMIP/HFP



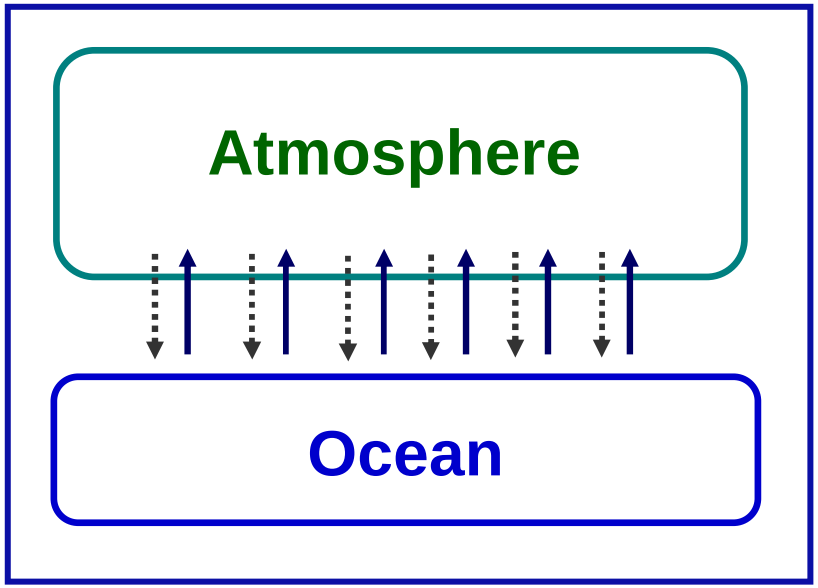
Current activities of seasonal prediction

Climate Prediction System

Two-tier



One-tier



Key SST prediction skill

Coupling of atmosphere and ocean process

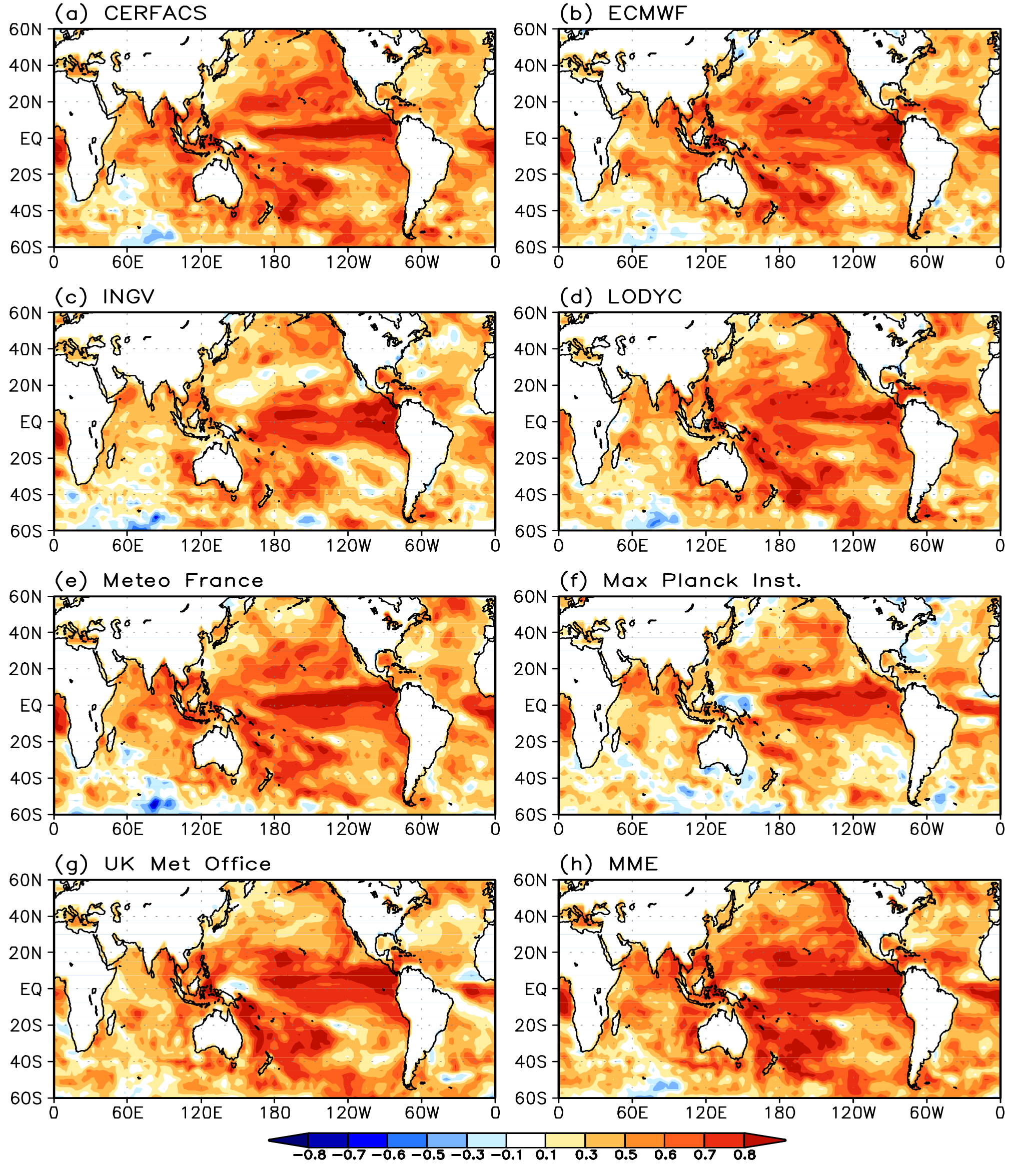
DEMETER/ ECMWF Prediction system

- **One-tier prediction system using CGCM**
- Development of European Multimodel Ensemble system for seasonal-to-interannual prediction
- 9 ensemble members of 7 models; 1980-1999 forecast

Institute	AGCM	Resolution	OGCM	Resolution	Atmosphere initial conditions	Ensemble generation
CERFACE	ARPEGE	T63 31 Levels	OPA 8.2	2.0x2.0 31 Levels	ERA-40	Windstress and SST perturbations
ECMWF	IFS	T95 40 Levels	HOPE-E	1.4x0.3-1.4 29 Levels	ERA-40	Windstress and SST perturbations
INGV	ECHAM-4	T42 19 Levels	OPA 8.1	2.0x0.5-1.5 31 Levels	Coupled AMIP-type experiment	Windstress and SST perturbations
LODYC	IFS	T95 40 Levels	OPA 8.2	2.0x2.0 31 Levels	ERA-40	Windstress and SST perturbations
Meteo-France	ARPEGE	T63 31 Levels	OPA 8.0	182GPx152G P 31 Levels	ERA-40	Windstress and SST perturbations
MPI	ECHAM-5	T42 19 Levels	MPI-OM1	2.5x0.5-2.5 23 Levels	Coupled run relaxed to observed SSTs	Atmospheric conditions from the coupled initialization run (lagged method)
UK Met Office	HadAM3	2.5x3.75 19 Levels	GloSea OGCM based on HadCM3	1.25x0.3- 125 40 Levels	ERA-40	Windstress and SST perturbations

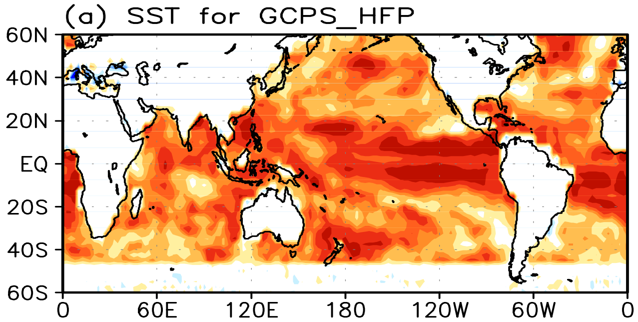
Temporal Correlation of SST

DEMETER

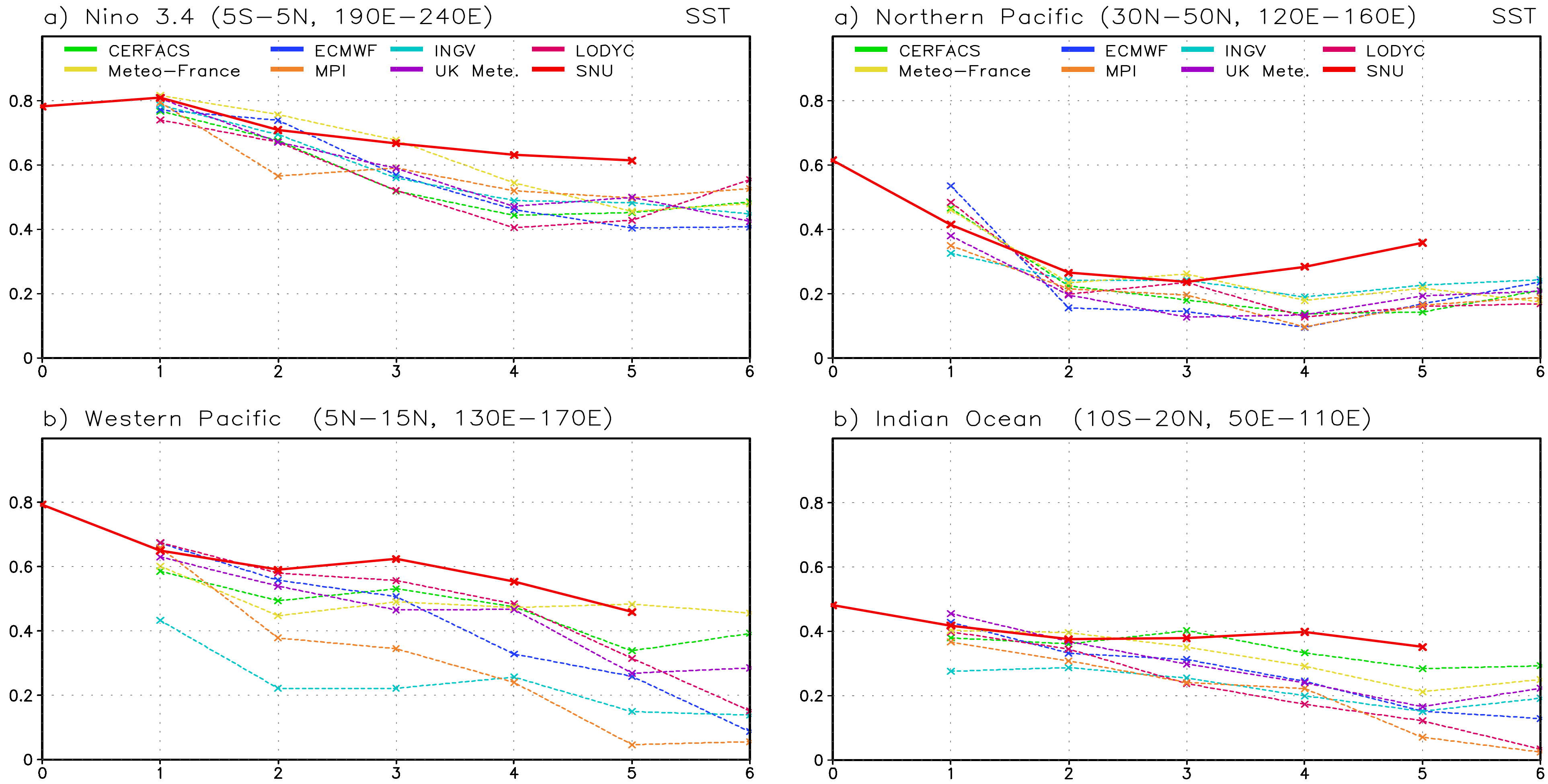


Prescribed SST

(3 month lead forecast)

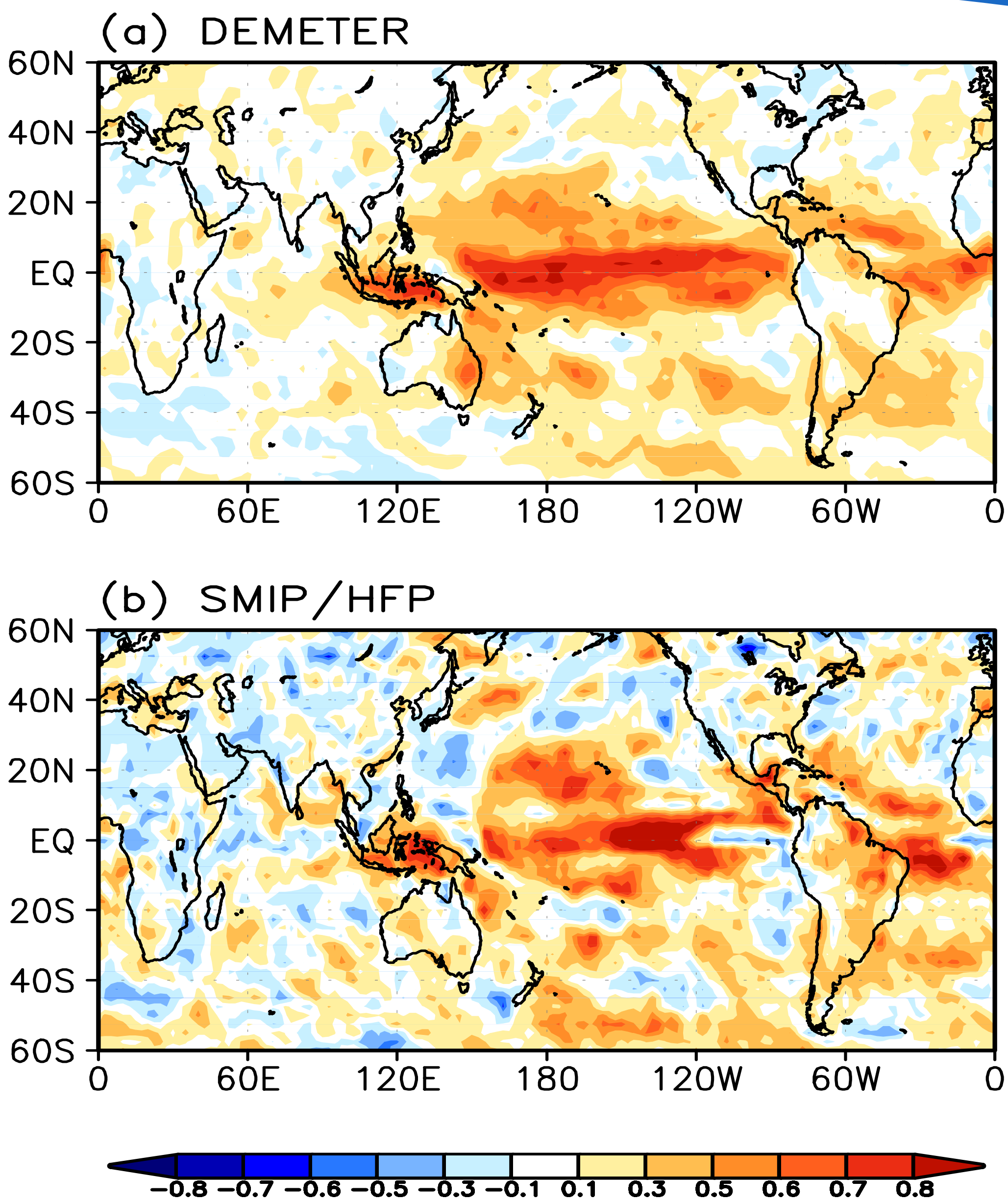


Correlation of area averaged SST



Temporal Correlation of PRCP

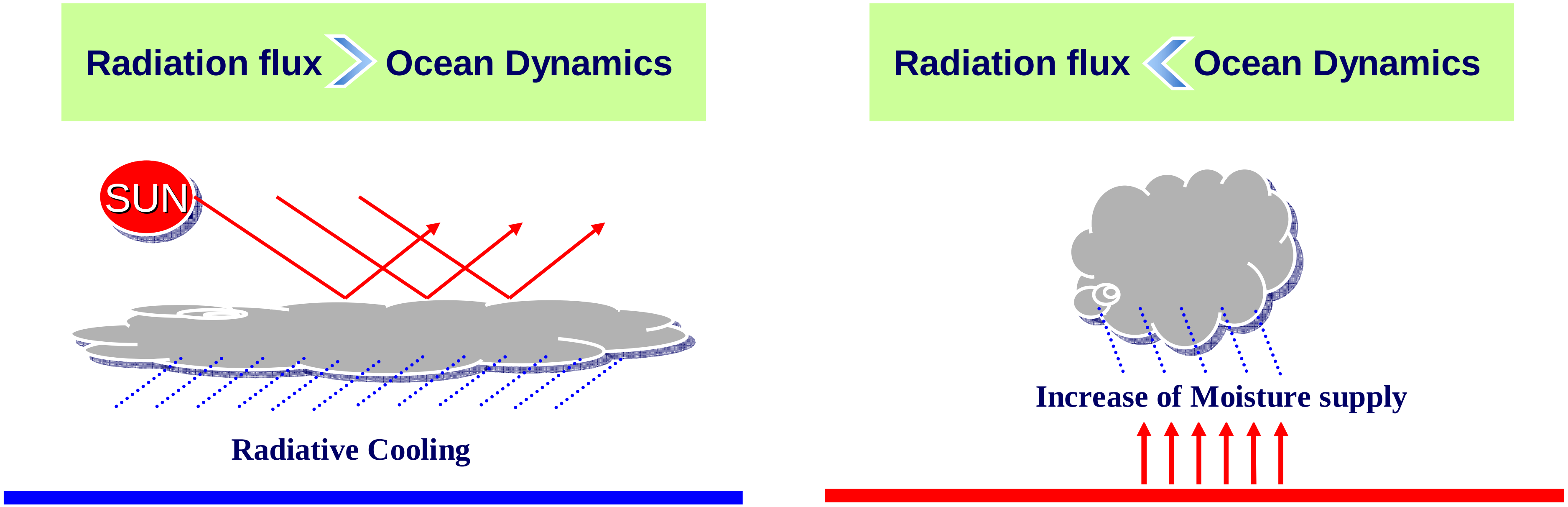
Temporal correlation



Part II.

Predictability of air-sea coupled system

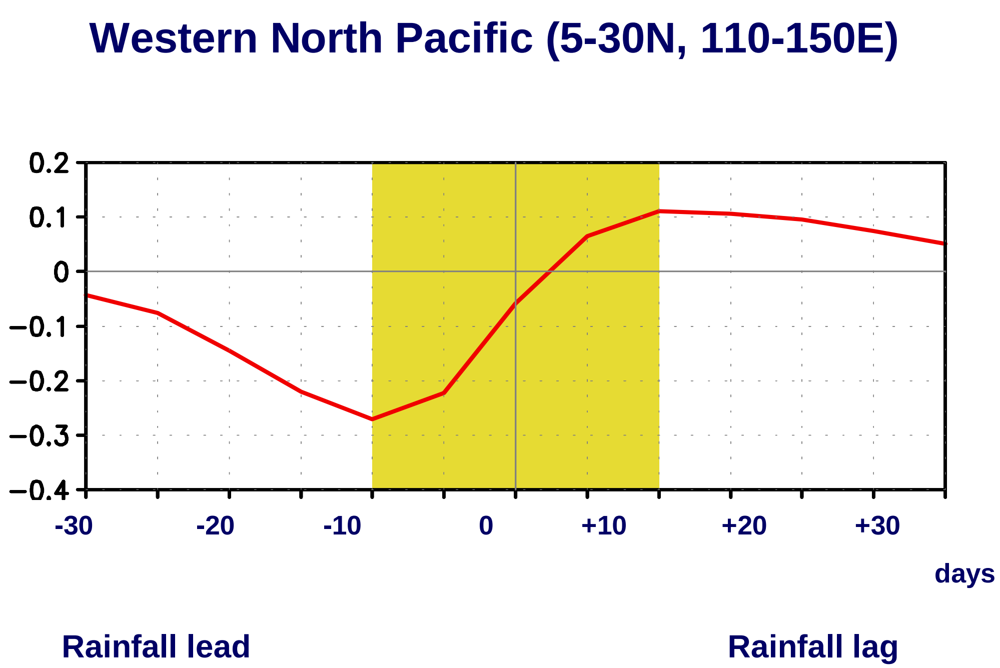
Air-sea interaction in the tropical Pacific

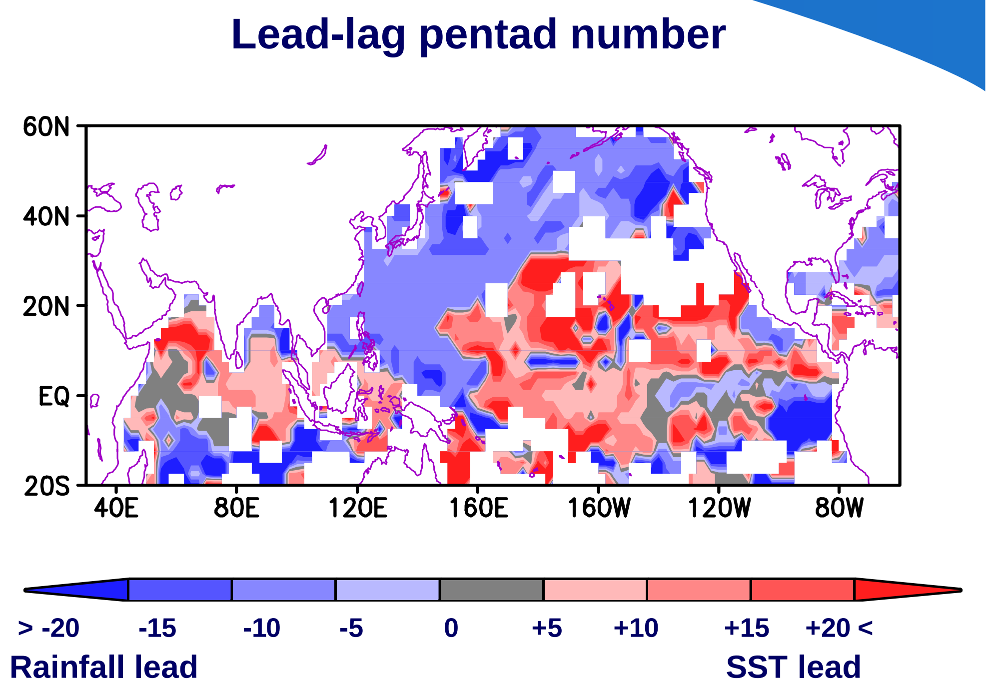


Where radiative flux control the SST…

1. Radiative flux would lead the SST anomalies
2. Temporal correlation between PRCP & SST can be a negative sign

Lead-lag correlation between pentad SST and rainfall data for JJA 82-99





Only more than 95% significance level is shaded

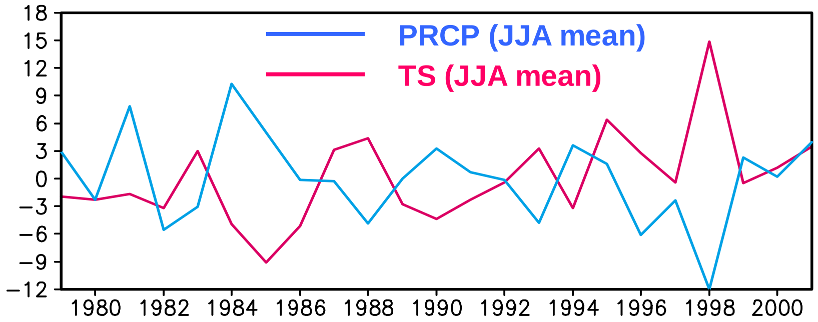
→ **Atmosphere forces the ocean where the correlation coefficients between rainfall and SST show negative.**

Role of the air-sea coupled process on seasonal prediction

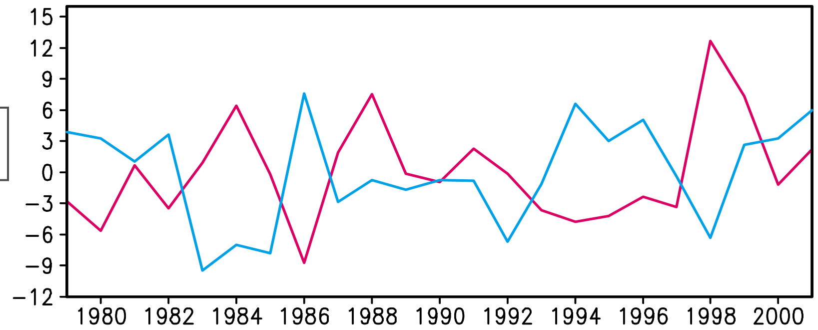
Time Series of Regional mean

(10-20N, 110-130E) PRCP & TS

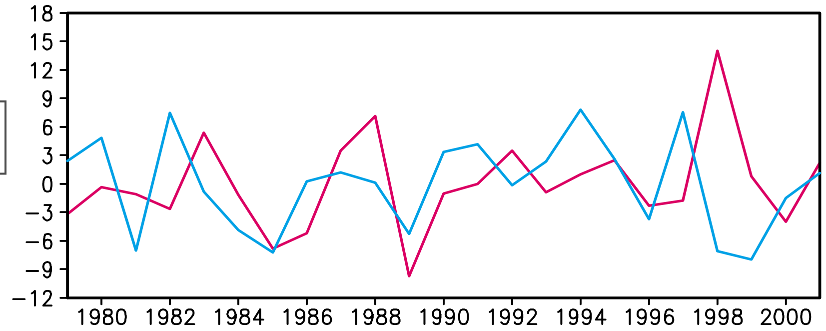
OBS



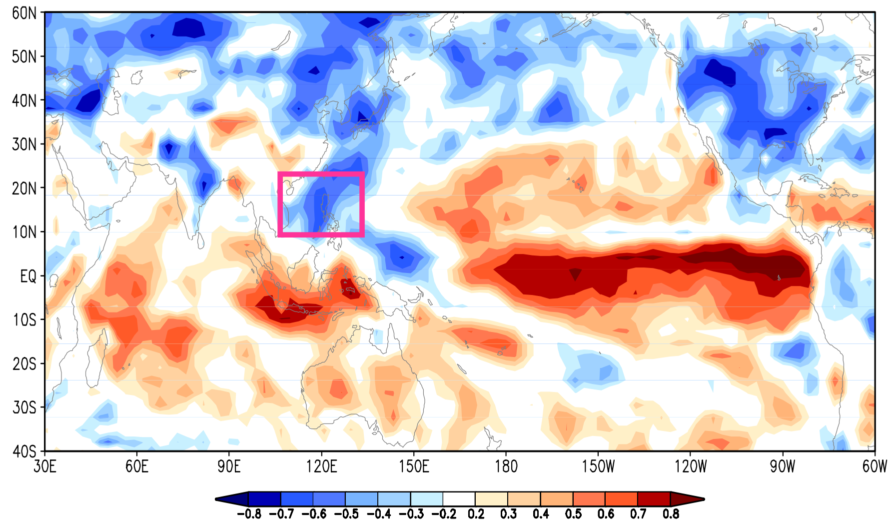
Tier-1



Tier-2



Correlation between OBS JJA precipitation and TS during 1979-2001

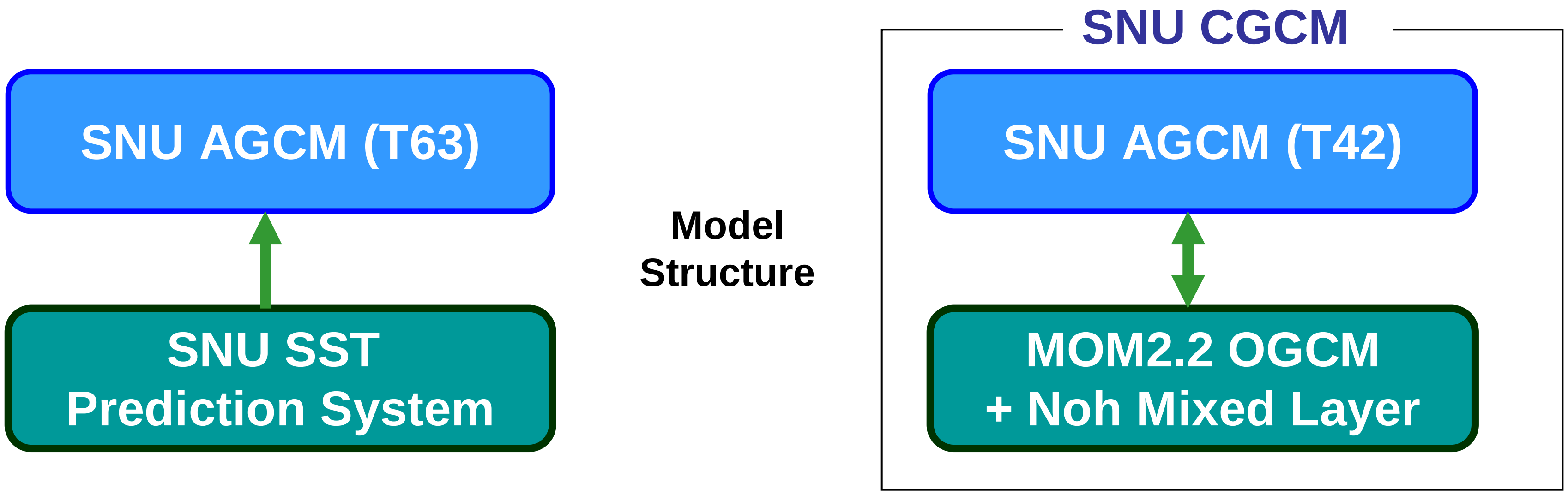


Correlation between PRCP & TS

	OBS	Tier-1	Tier-2
COR	- 0.64	- 0.38	- 0.02

Tier-2 vs Tier-1 Prediction Systems

Tier-2 Prediction System | **Tier-1 Prediction System**



Tier-2		Tier-1
T63	AGCM Resolution	T42
Perfect	SST Climatology	Some systematic Biases
Relatively Superior	SST Anomaly	Relatively Inferior
No	Air-Sea Coupled Process	Yes

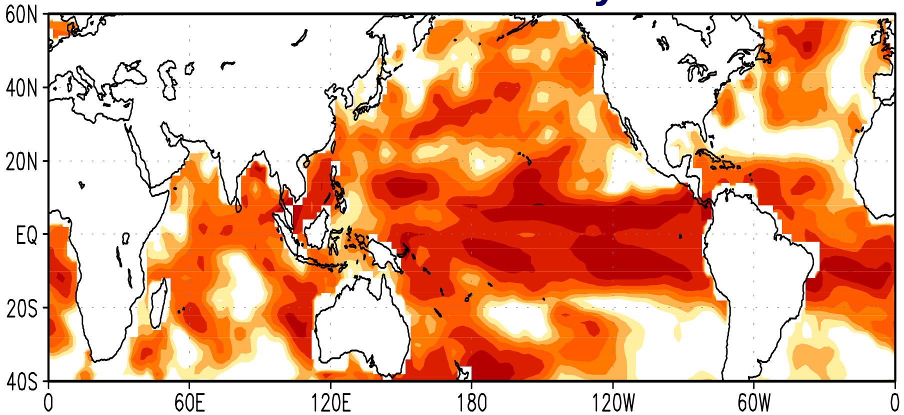
Correlation Skill for SST

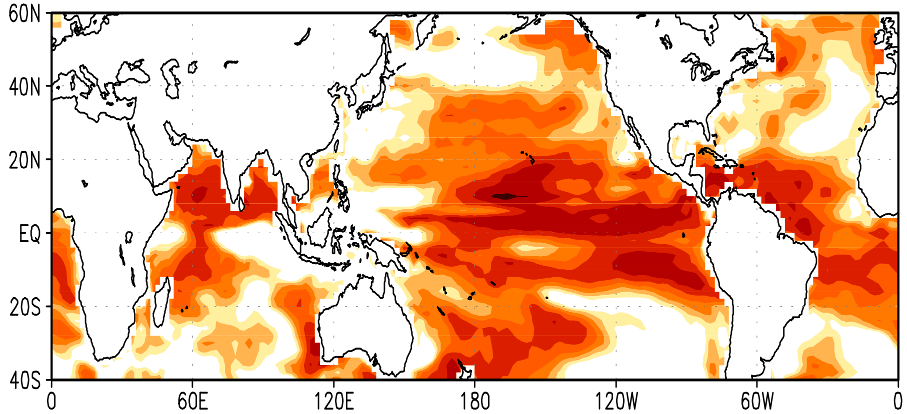
Tier-2 Prediction

(dynamical and statistical ensemble prediction)

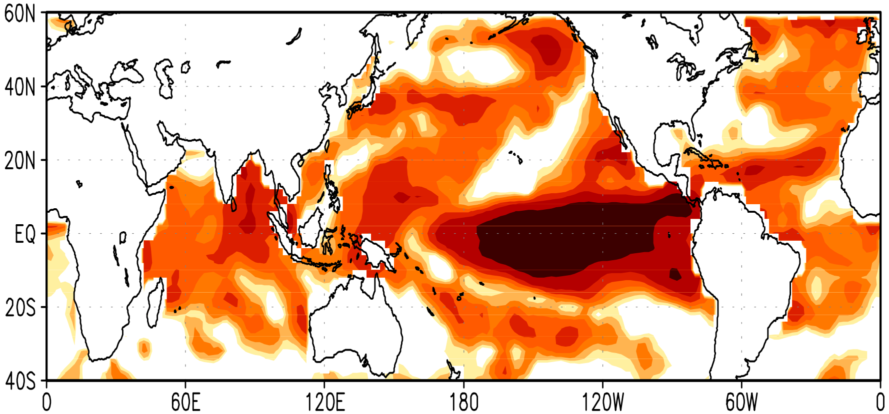
Tier-1 Prediction

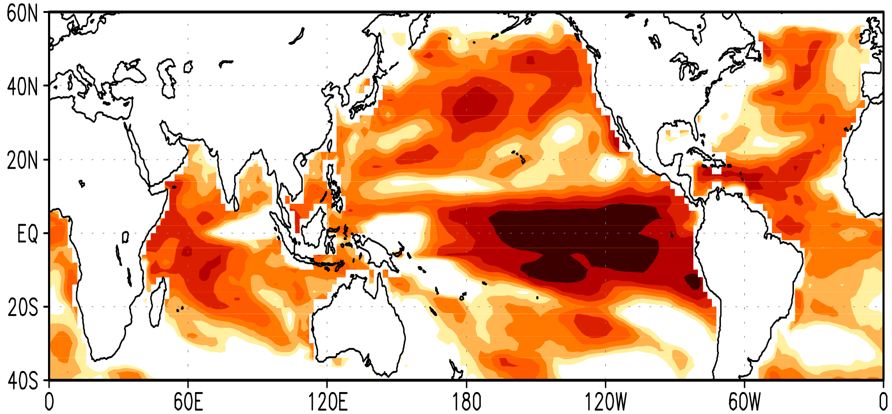
Summer Mean from 1st May

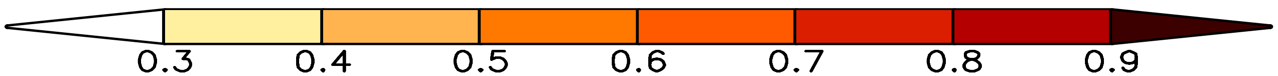




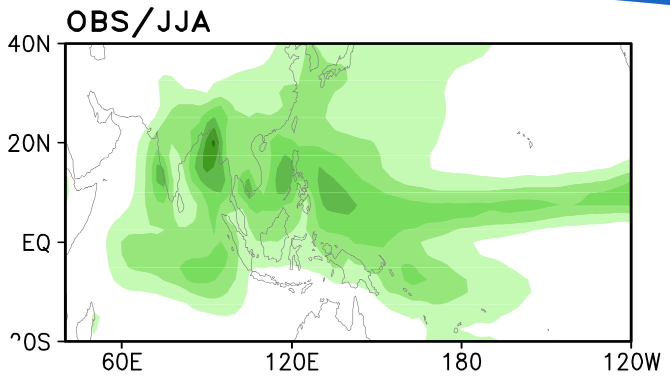
Winter Mean from 1st Nov.







JJA Precipitation Climatology and Mean Bias

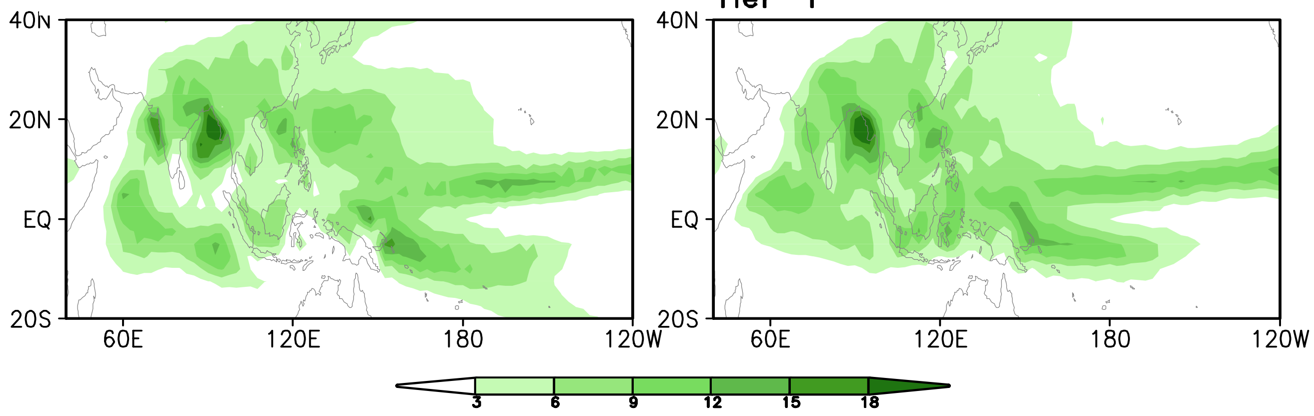


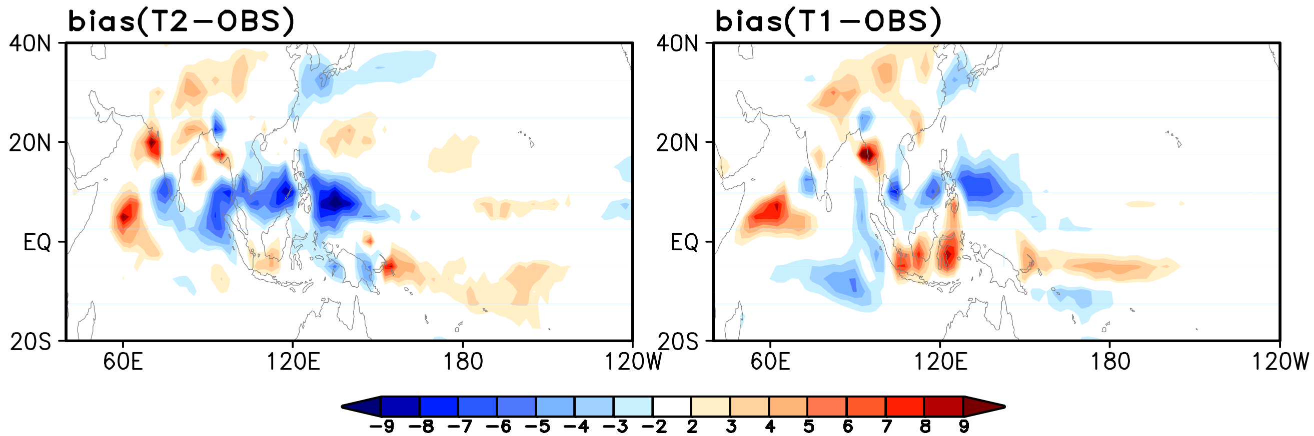
✓ **Perfect SST Climatology**

✓ **Some Mean SST biases**

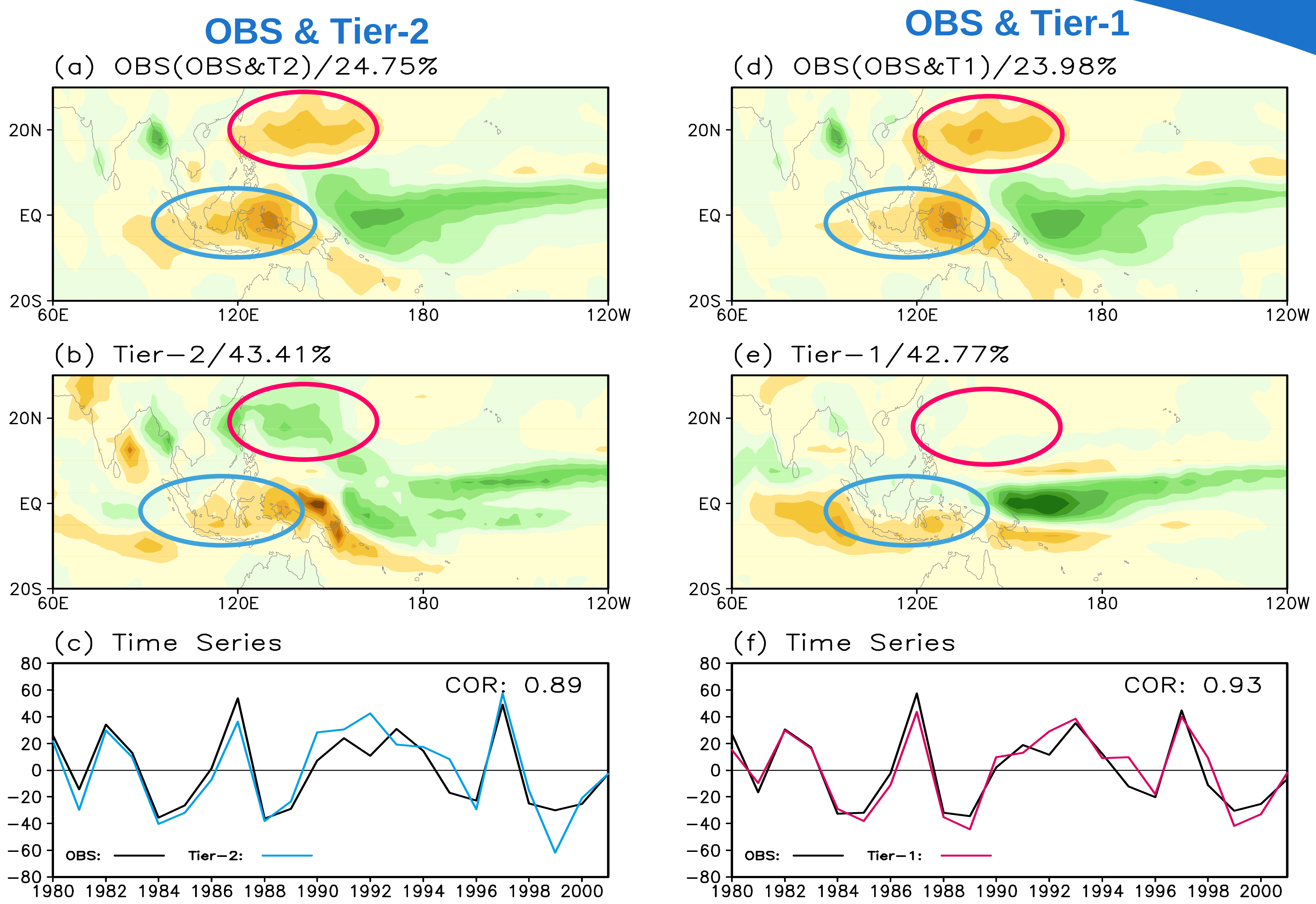
Tier-2

Tier-1

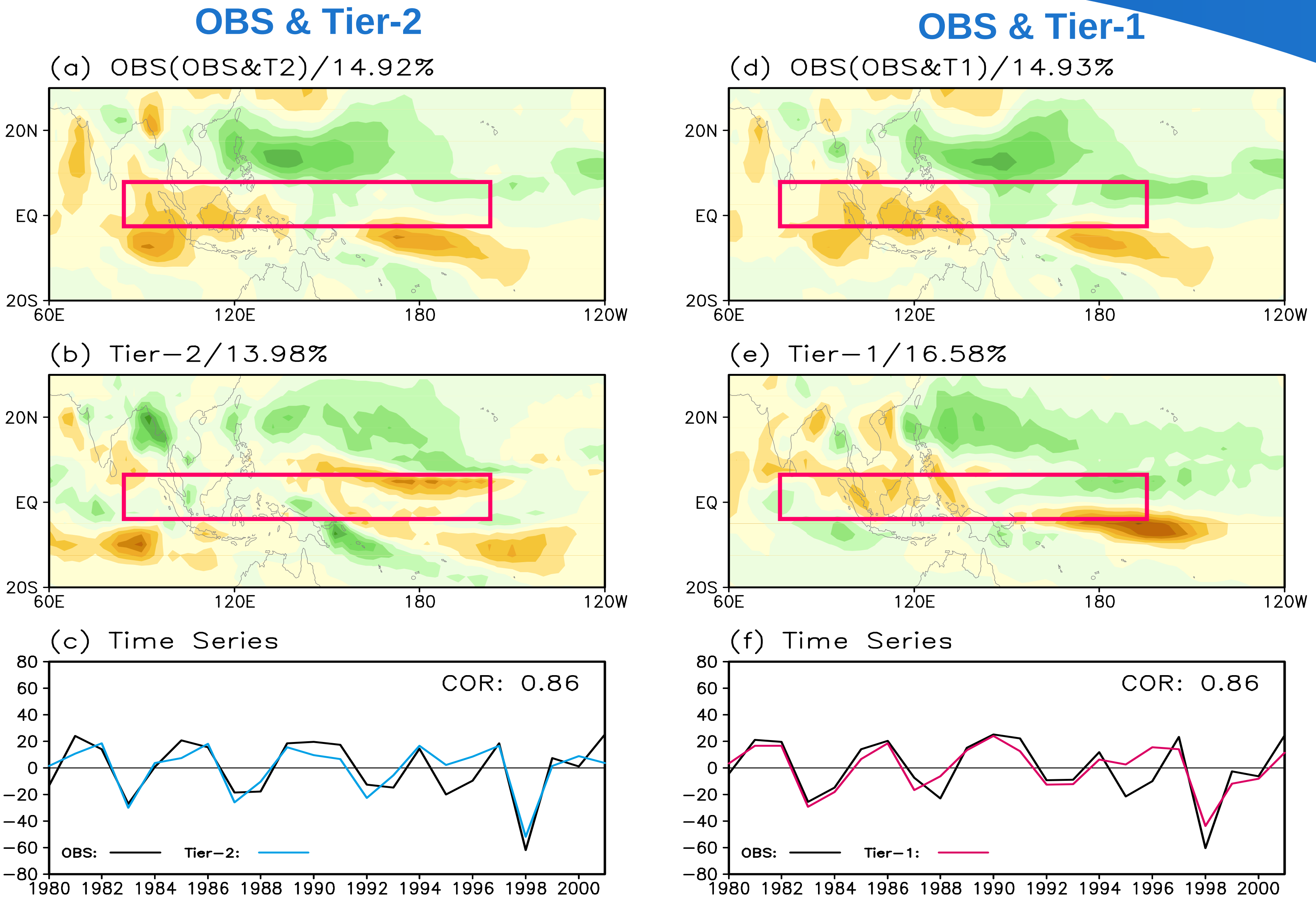




1st SVD Mode for Precipitation (ENSO mode)



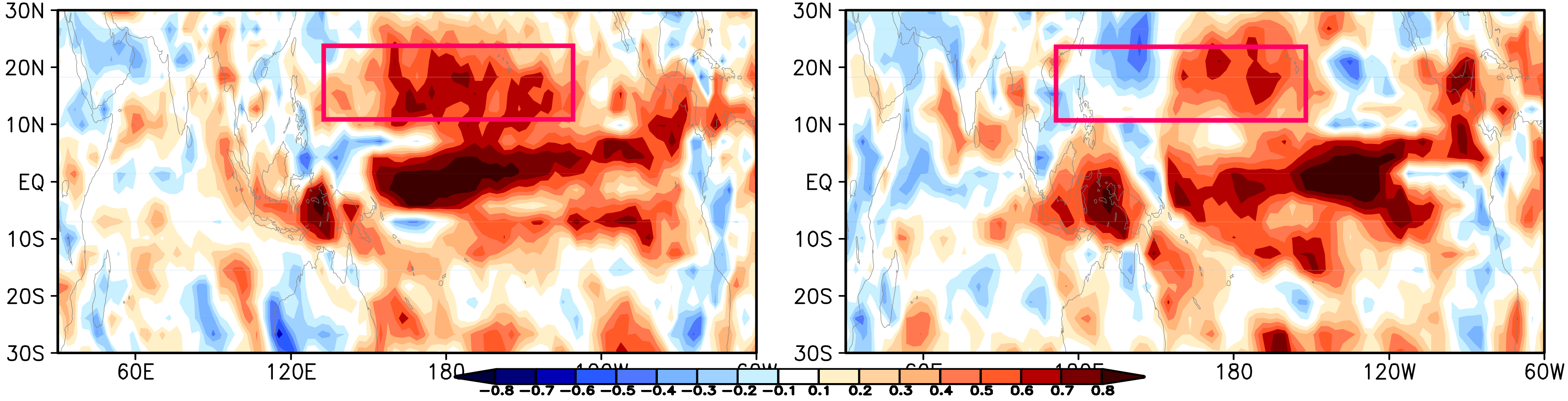
2nd Mode for Precipitation (WNP Monsoon mode)



Correlation skill for JJA Precipitation

Tier-One (CGCM)

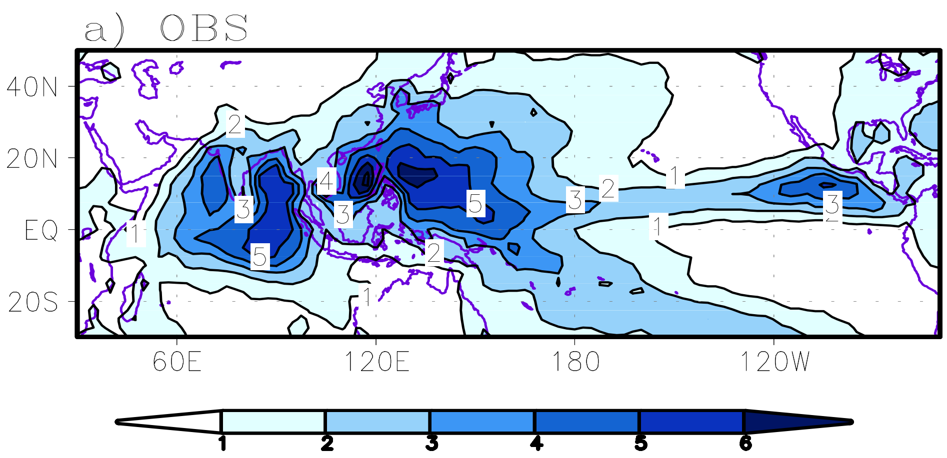
Tier-Two (SMIP/HFP)

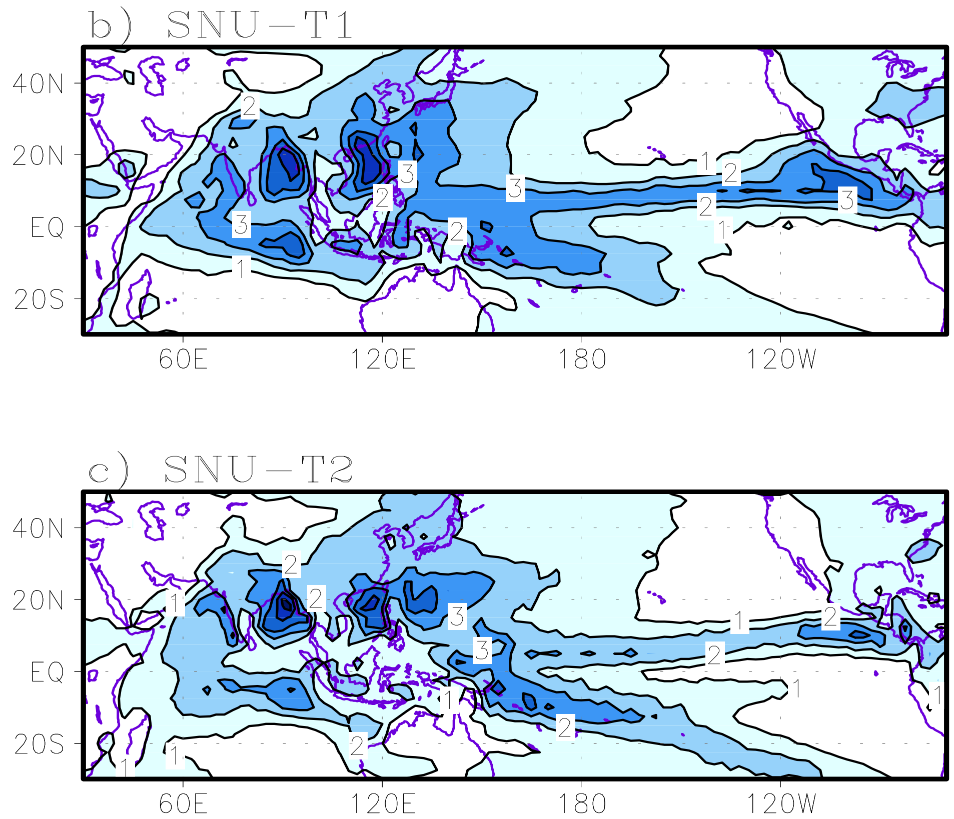


Tier-1 prediction is superior to Tier-2 prediction,
even though the tier-2 SST prediction is better.

Role of air-sea interaction on ISO activity

ISO activity (MJJA)

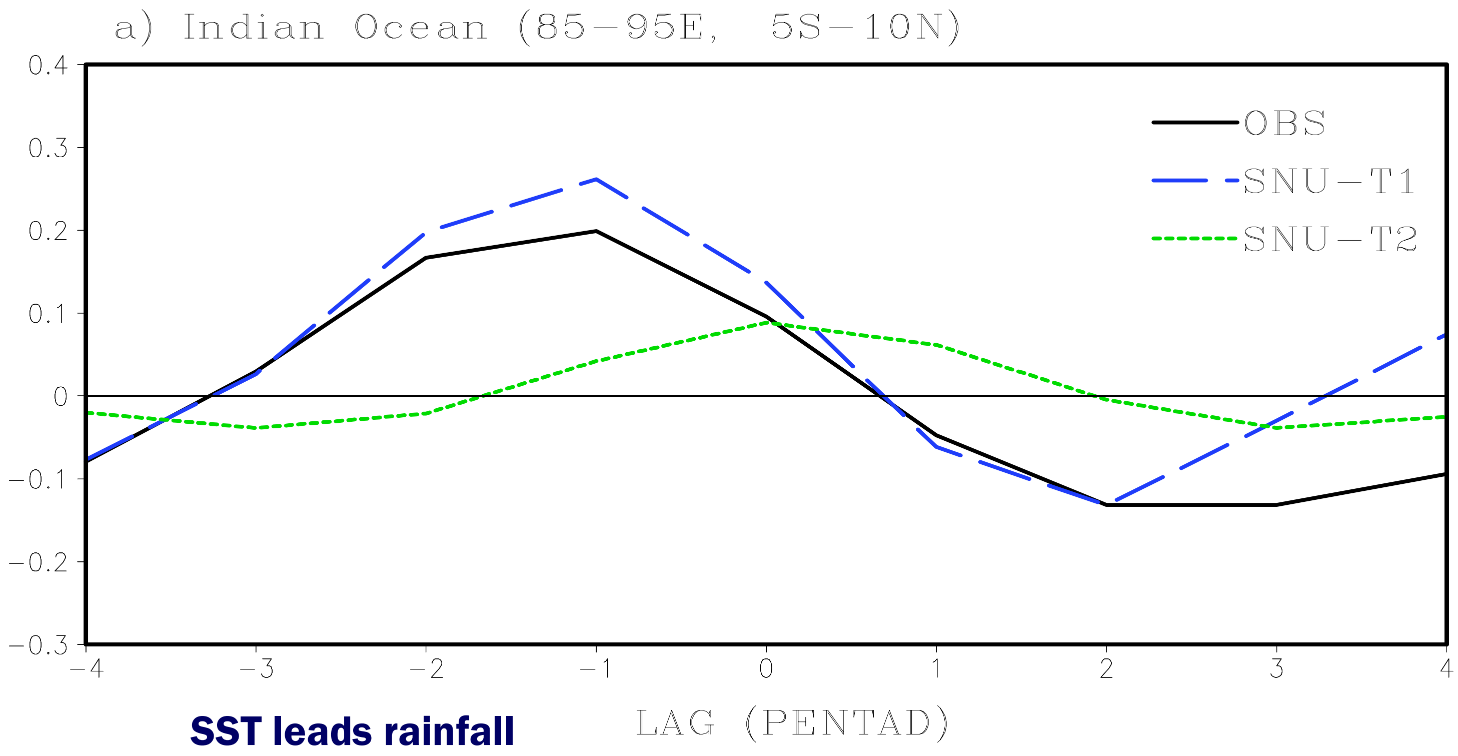




*** ISO activity : STD of 20-90 filtered prcp 1980-2001 MJJA**

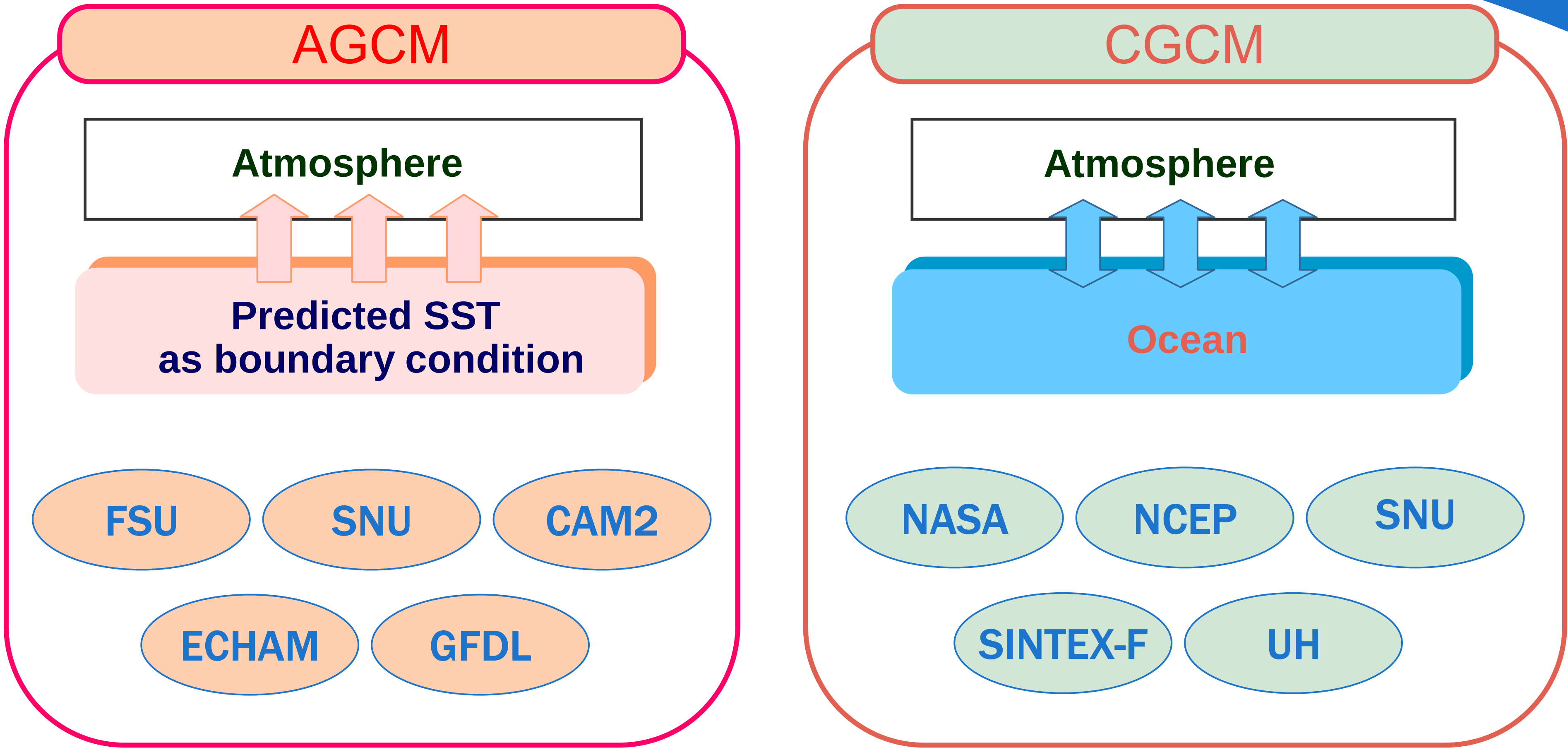
Enhanced ISO activity in Tier-1

Phase relationship with PRCP and SST



- SST leads rainfall by one-two pentads in OBS and T1
- PRCP and SST are almost in phase in T2

Prediction models of various institutes – CliPAS project



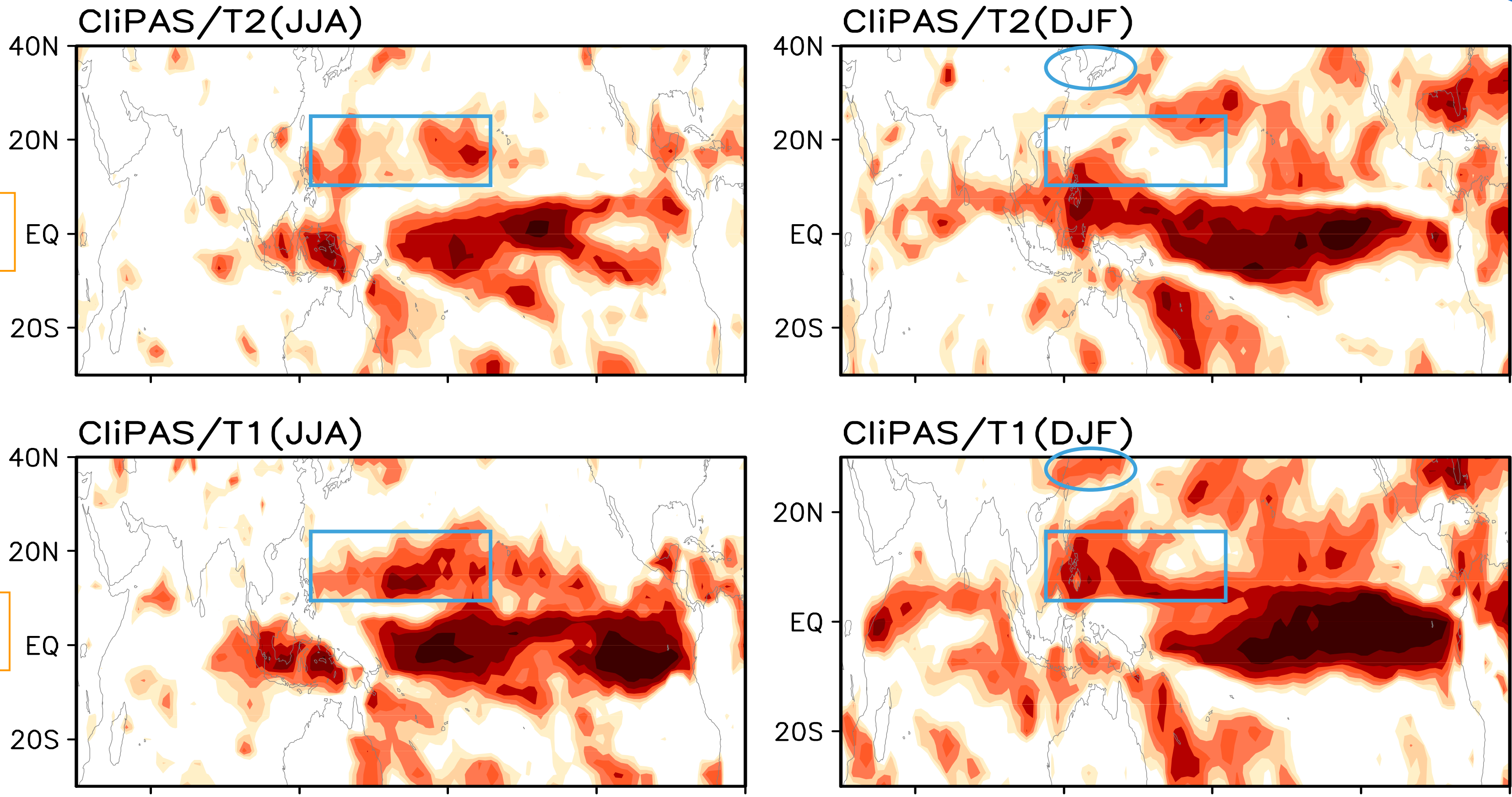
Experiment design

- 1981 – 2004 summer (MJJAS)and winter (NDJFM) seasons for 24 years
- 6-15 Member ensemble for each model
- 4-9 months lead time forecast

Correlation Skill for Precipitation - CliPAS data

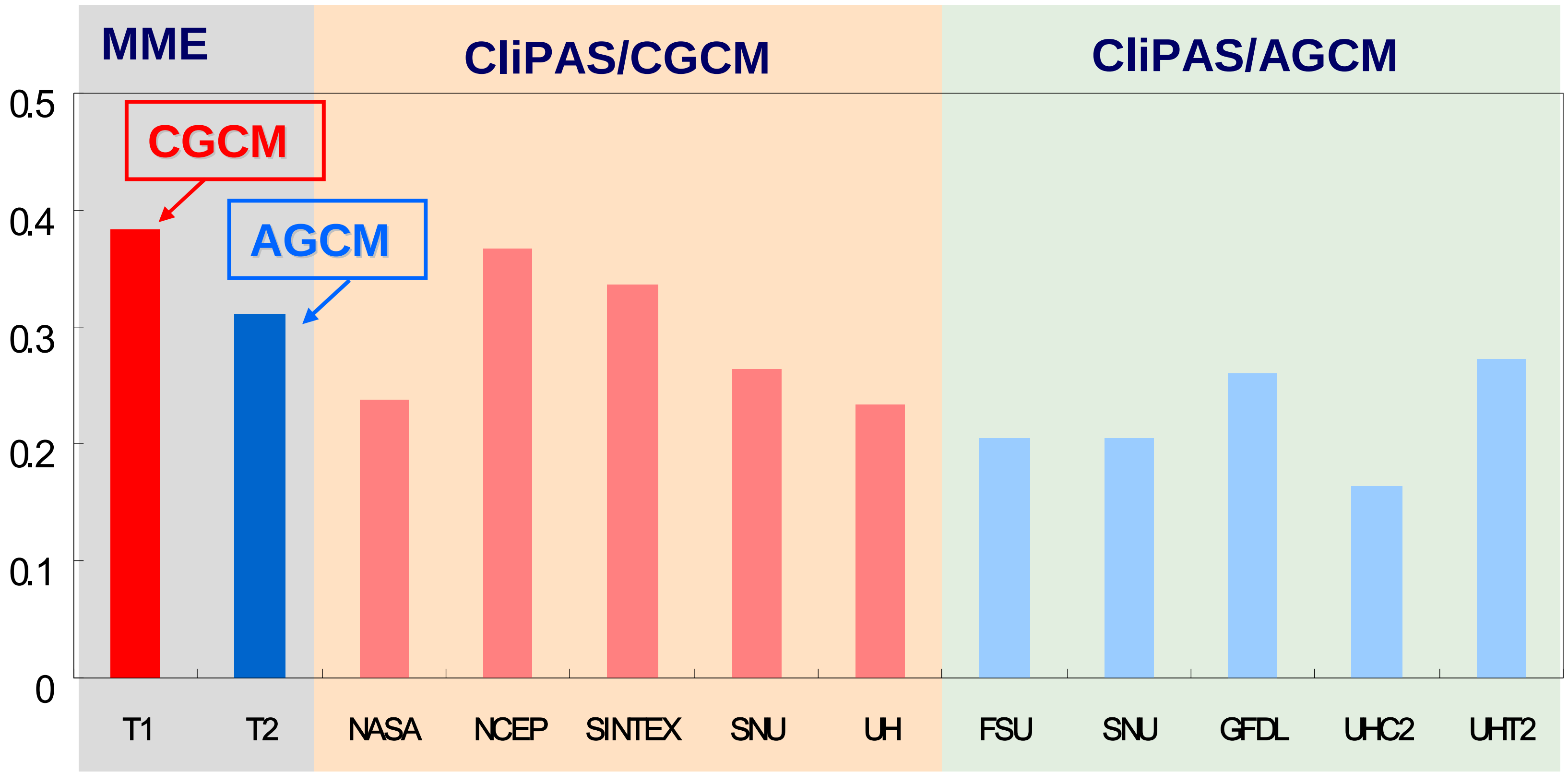
Tier-Two

Tier-One



The state-of-the-art Climate Prediction

Global domain pattern correlation(60S-60N, 0-360)



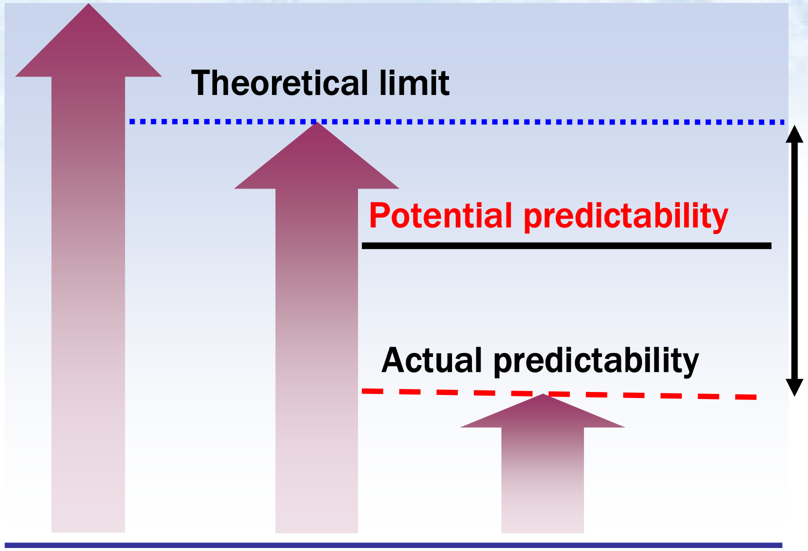
Part III.

Access to upper limit predictability

Contents

- Error correction
- Multi model ensemble prediction
- Noise dynamics

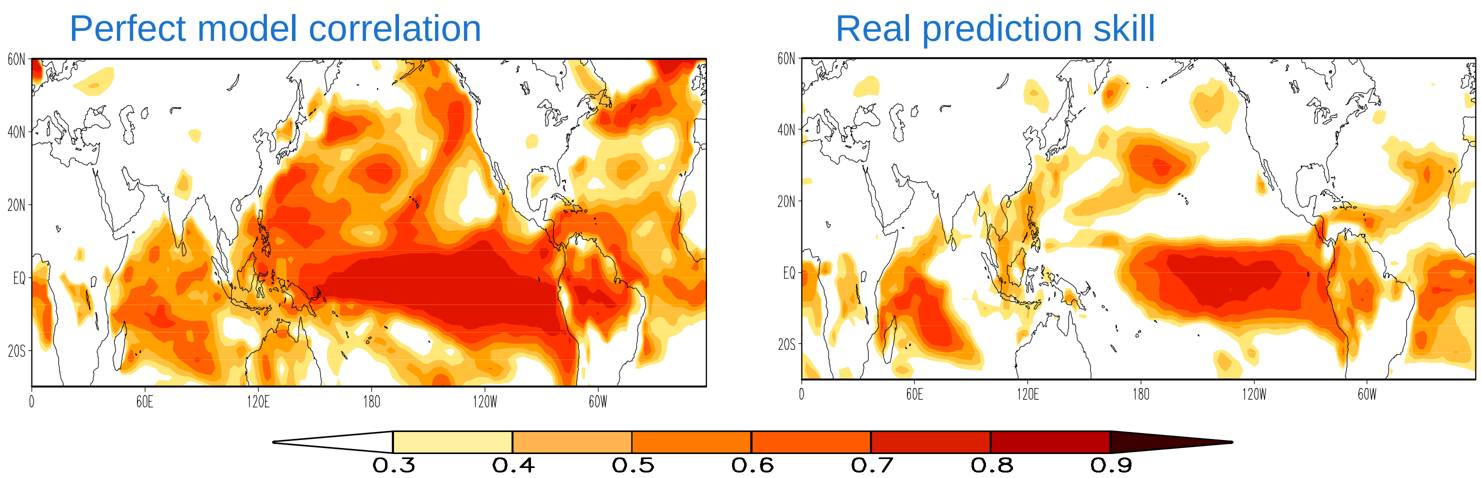
Approach to the theoretical limit



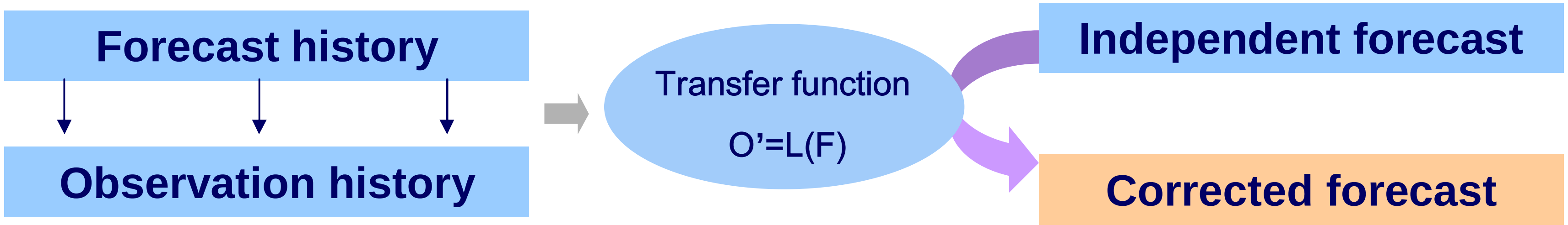
Challenging part

The error comes from…

1. Model physics uncertainties
2. Initial condition uncertainties
3. Internal noise



Correcting signal : Statistical Post process



There are many approaches in post-process, All of them share similar assumption. :
Statistics between forecast and observation is stationary

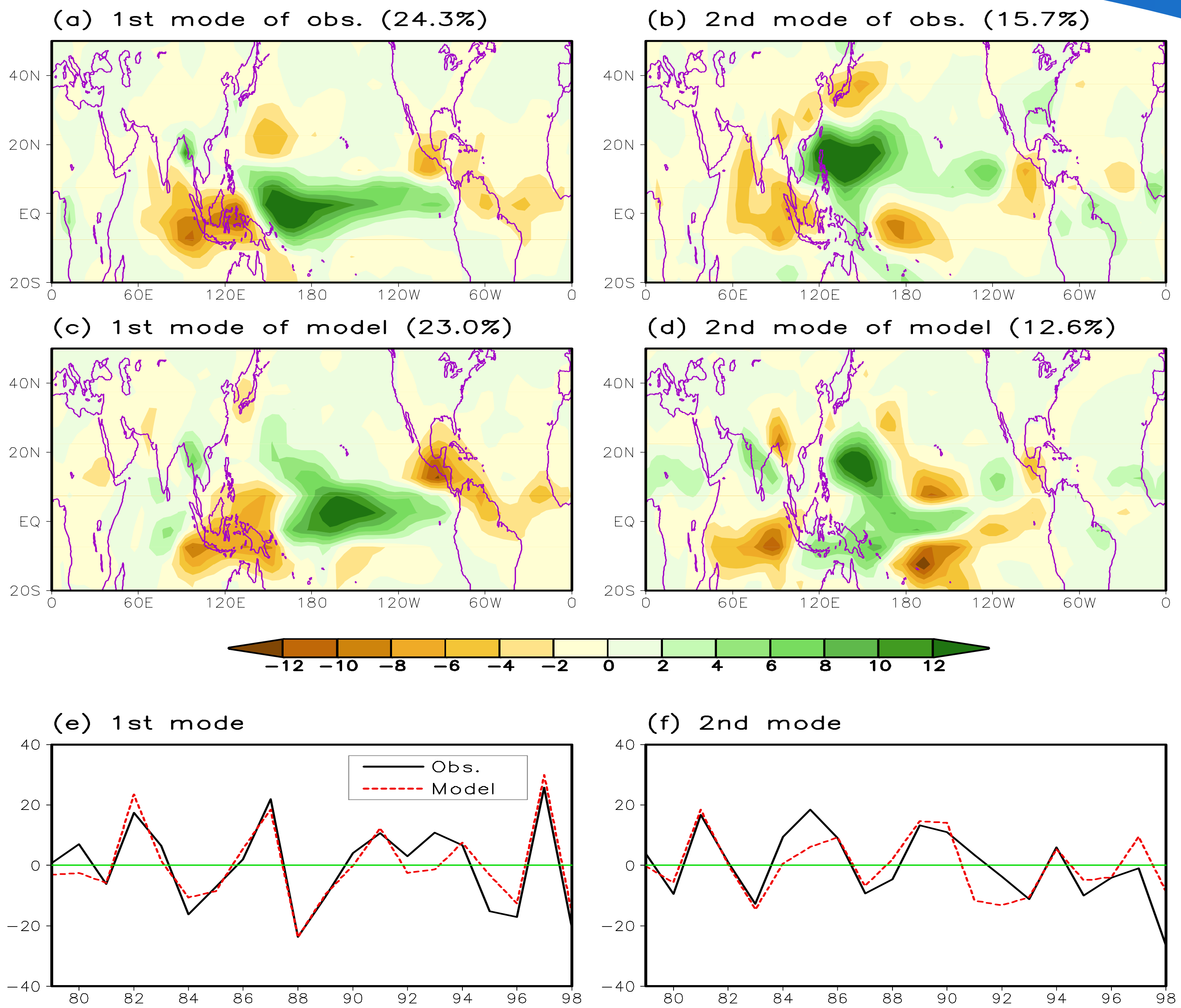
If statistics is not stationary, post-process will not work in independent forecast

Thus, statistical stability is a rule of thumb in the statistical post-process (avoiding overfitting)

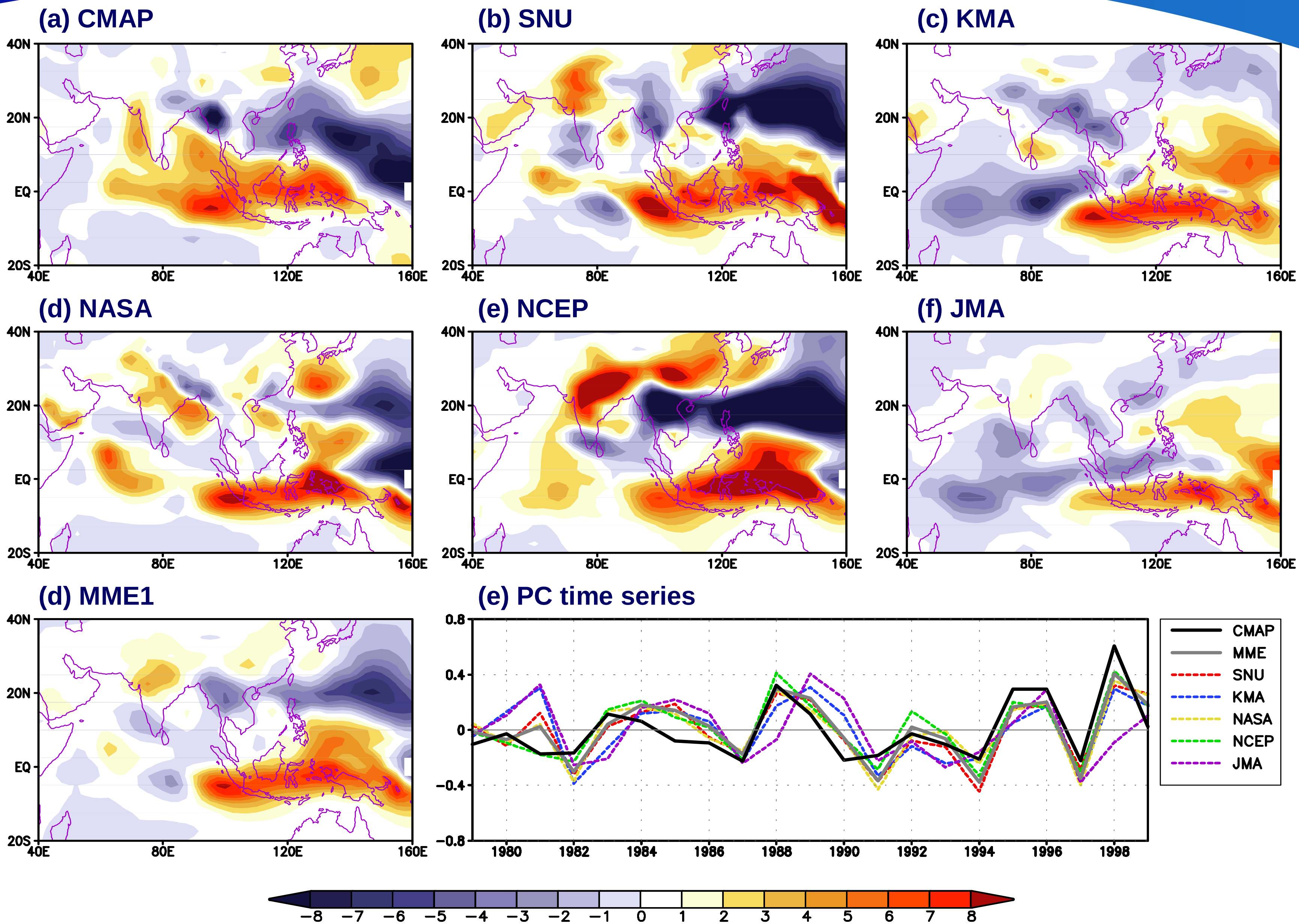
Regarding actual constraints, available large ensemble forecast with well-tuned post process will be an appropriate strategy of seasonal forecast.

➔ Statistically optimized multi model ensemble prediction

EOF of Summer Mean Precipitation



EOF Analysis of Summer Mean Precipitation



Anomaly Bias Correction

Procedure of Anomaly Bias Correction

EOF Analysis

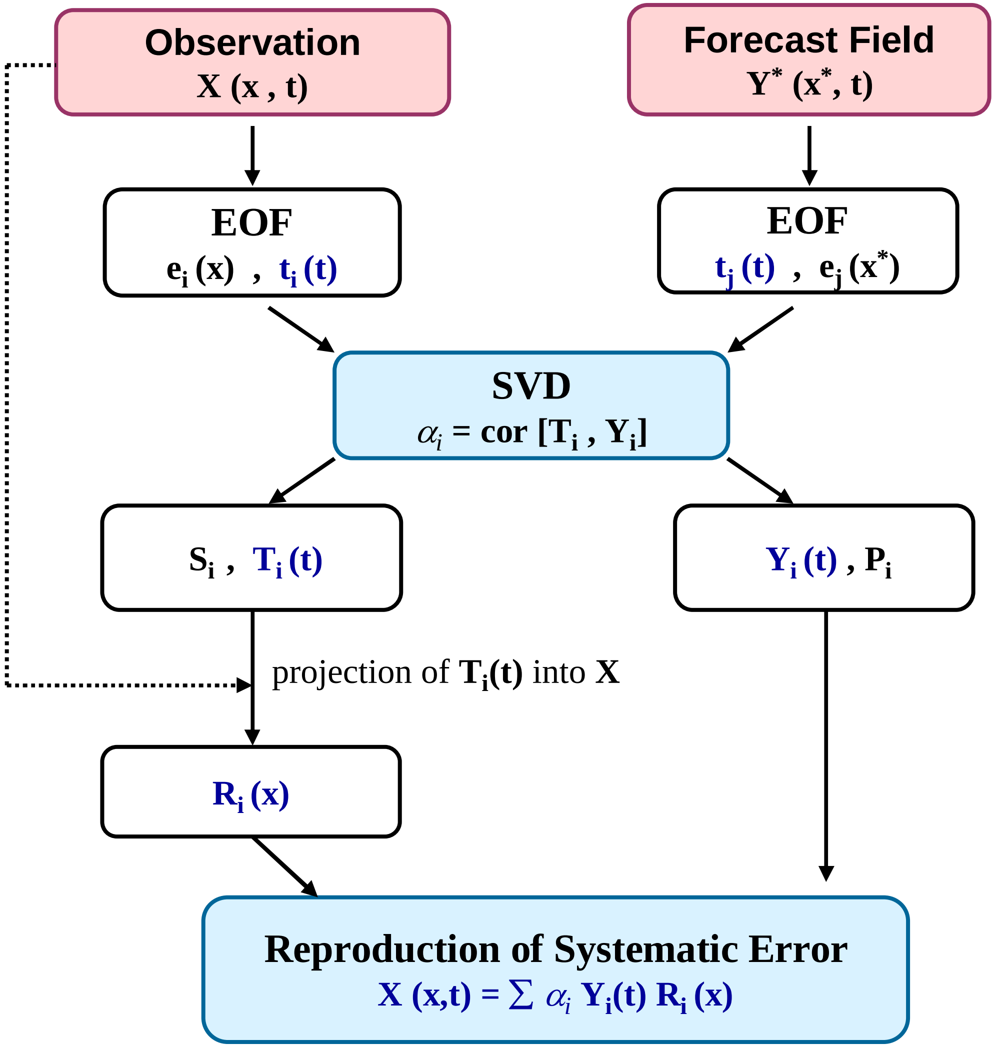
Truncation of small scale noise modes by retaining first 10 EOF modes

SVD Analysis

Couple pattern of observation and model

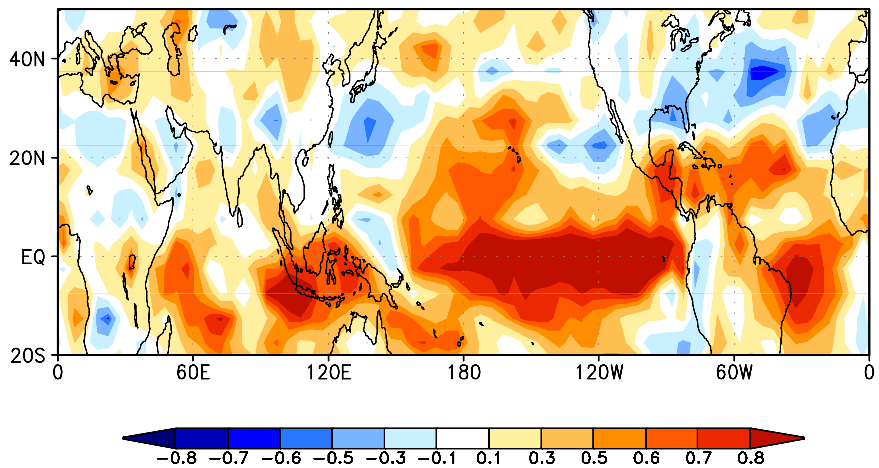
Transfer Function

Replace the model SVD mode to the corresponding observation mode

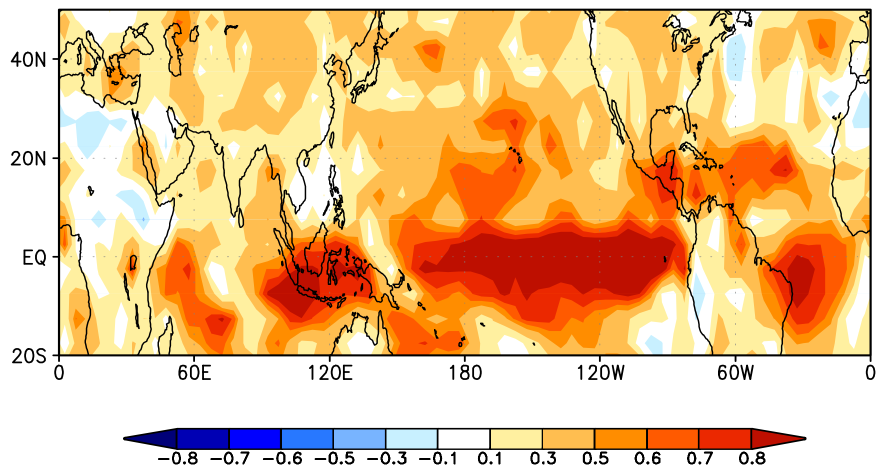


Correlation and Forecast Skill Score

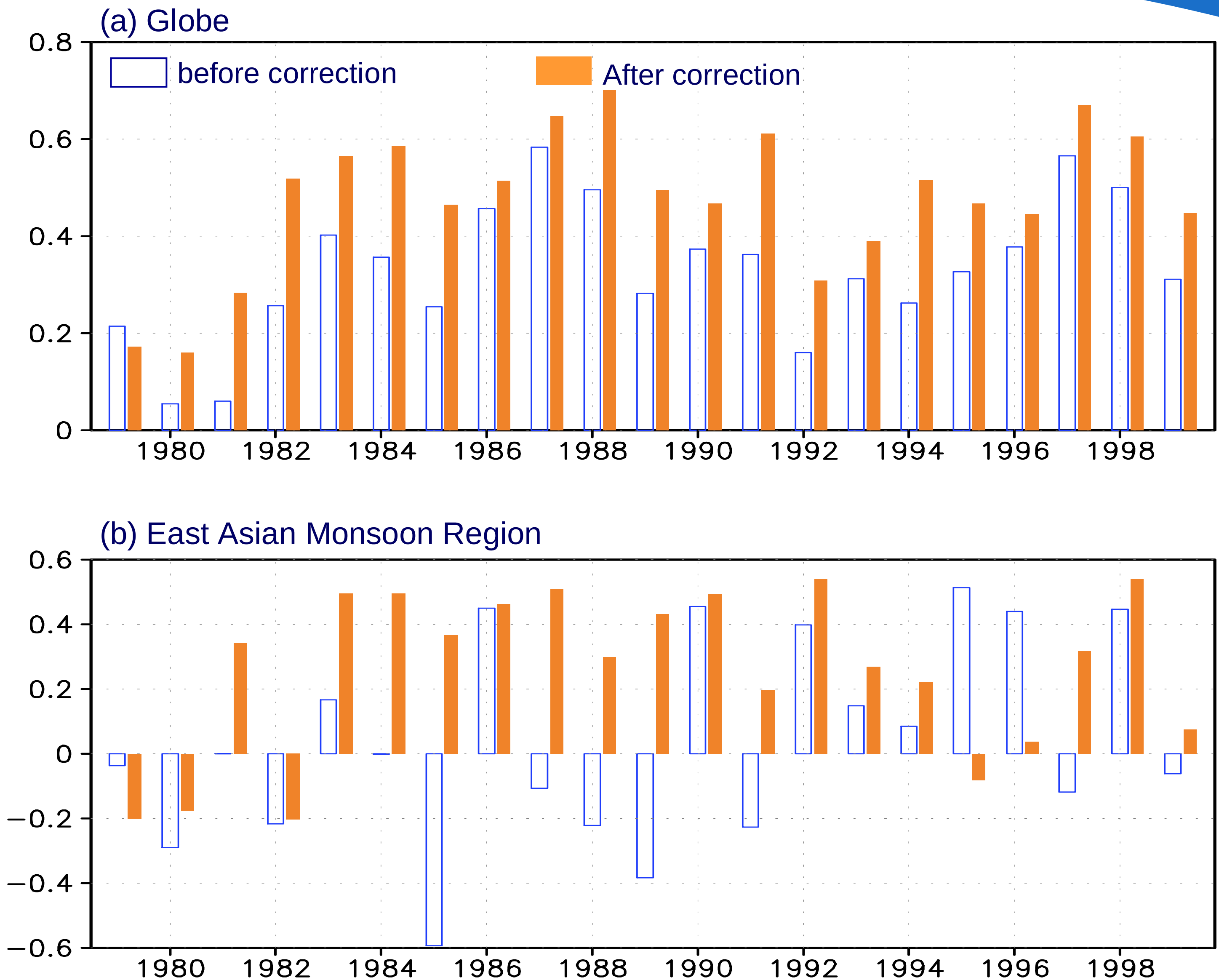
Before Bias Correction



After Bias Correction



Pattern Correlation : Interannual Predictability



SNU correction model

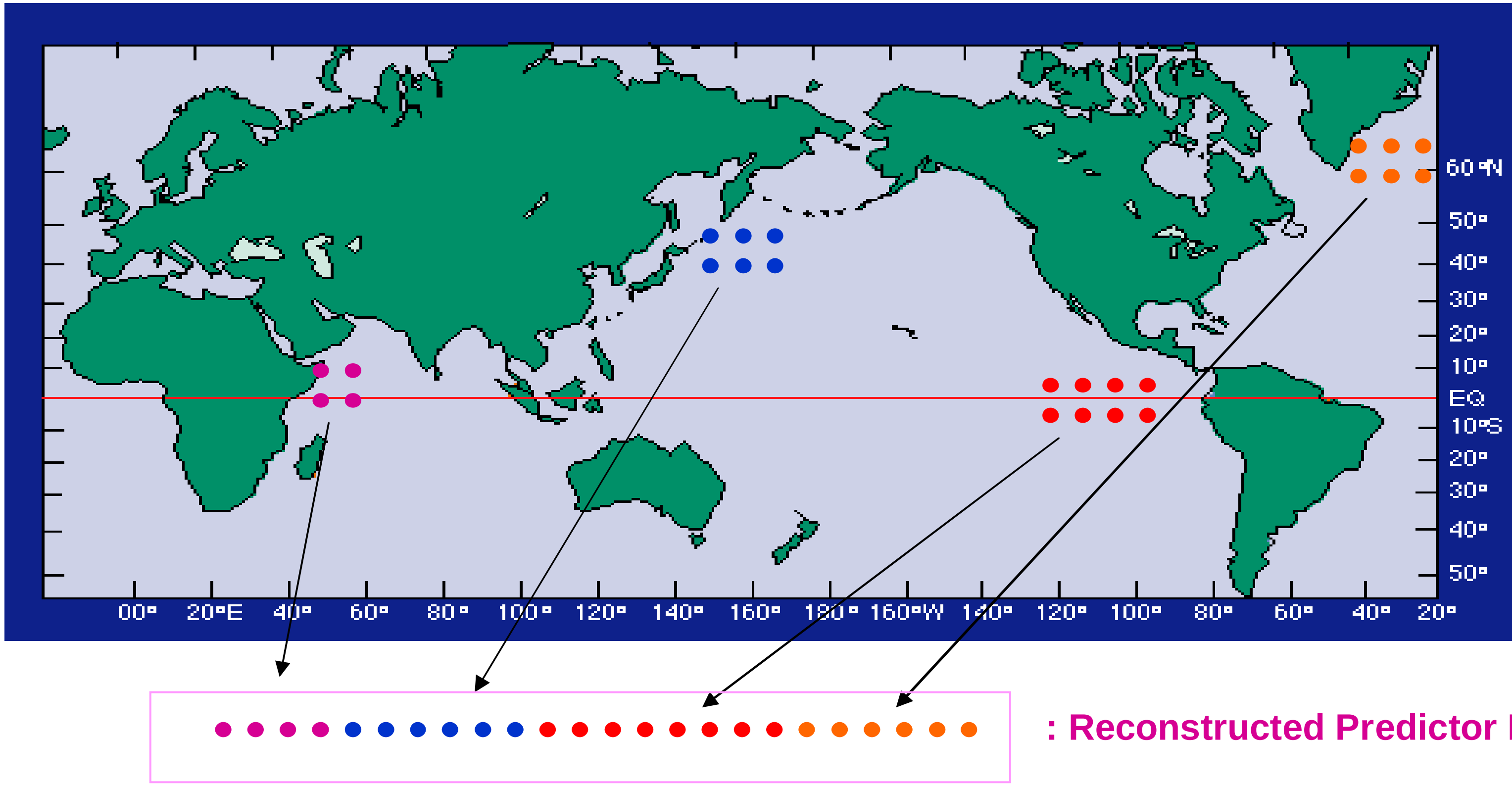
- ## First Step : Prior Predictor Selection

  - Select qualified predictor grid based on correlation for training period
  - Gather split predictors and regard as a predictor pattern

- ## Second Step : Pattern Projection

  - Construct covariance pattern between observation and reconsturcted model pattern
  - Obtain prediction by projecting model pattern on the covariance pattern

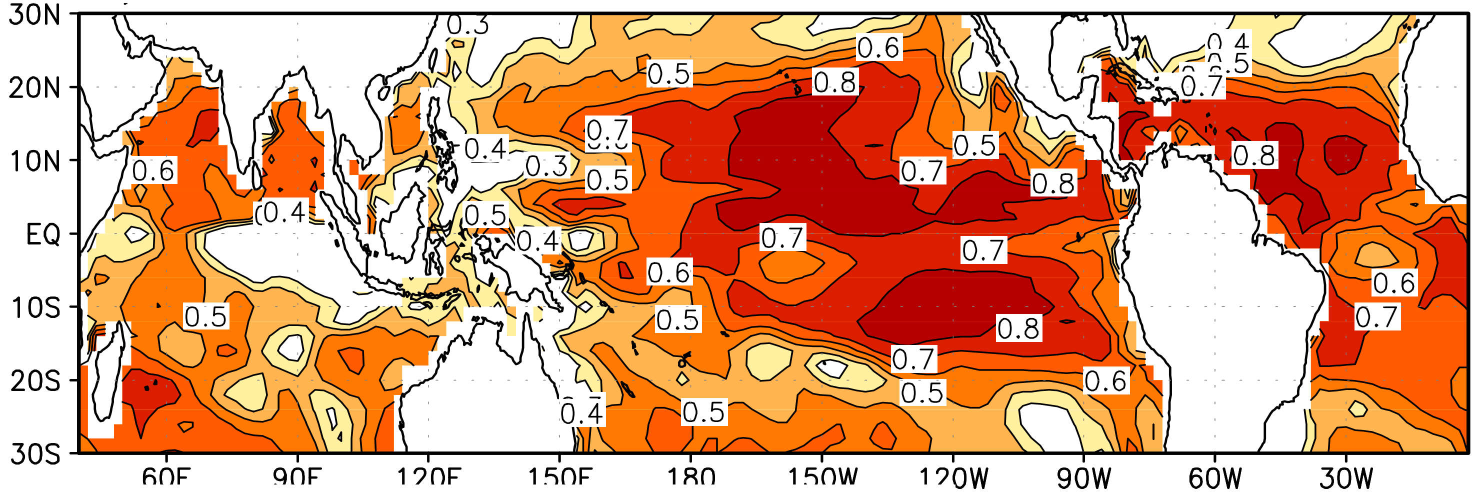
$$X_P(t) = \sigma_Y \sum_{i,j} \frac{COV(i,j) \bullet X(i,j,t)}{\sigma_X^2(i,j)}$$



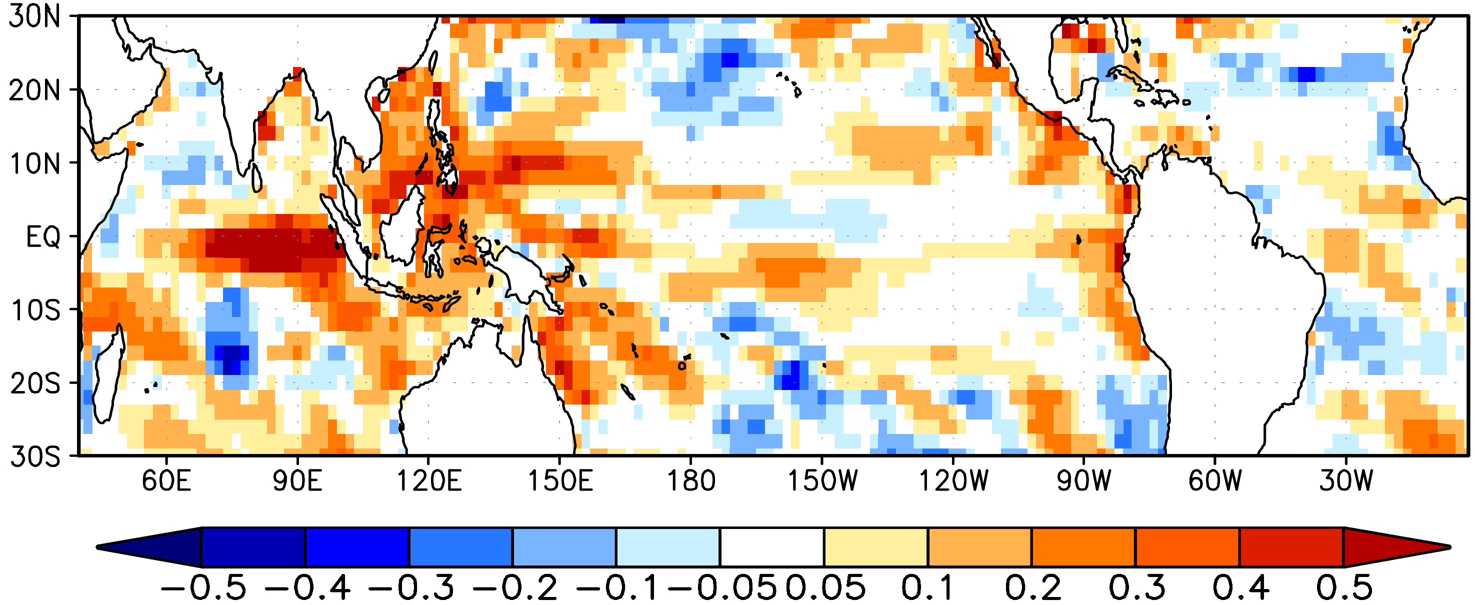
: Reconstructed Predictor Domain

Statistical Downscaling/Correlation Skill for SST

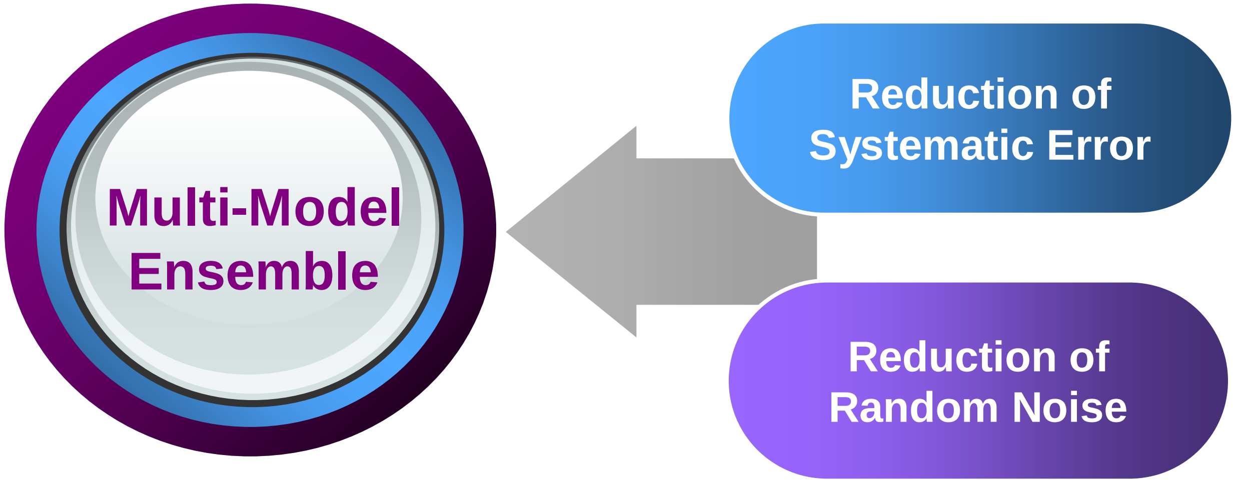
Before Correction



Improvement by Statistical Correction



Multi-model Ensemble Prediction



Cancellation of errors

: Multi-model

More samples

: Ensembles

Benefits of Multi Model Ensemble

$$x_i = y + e_i + \varepsilon_i$$

Forecast = True + Error + Noise

E_M : Error variance of Multi-Model Ensemble $$E_M = \frac{1}{M}\left[\overline{V}(e) + \frac{2}{M}\sum_{ij} COV(e_i, e_j) + \overline{V}(\varepsilon)\right]$$

E_S : Error variance of Single Model $$E_s = \overline{V}(e) + \overline{V}(\varepsilon)$$

[Example] *Two Models (M=2)*

Error variance of Multi-Model Ensemble $$E_M = \frac{1}{2}\left[\overline{V}(e) + COV(e_1, e_2) + \overline{V}(\varepsilon)\right]$$

Error variance of Single Model $$E_s = \overline{V}(e) + \overline{V}(\varepsilon)$$

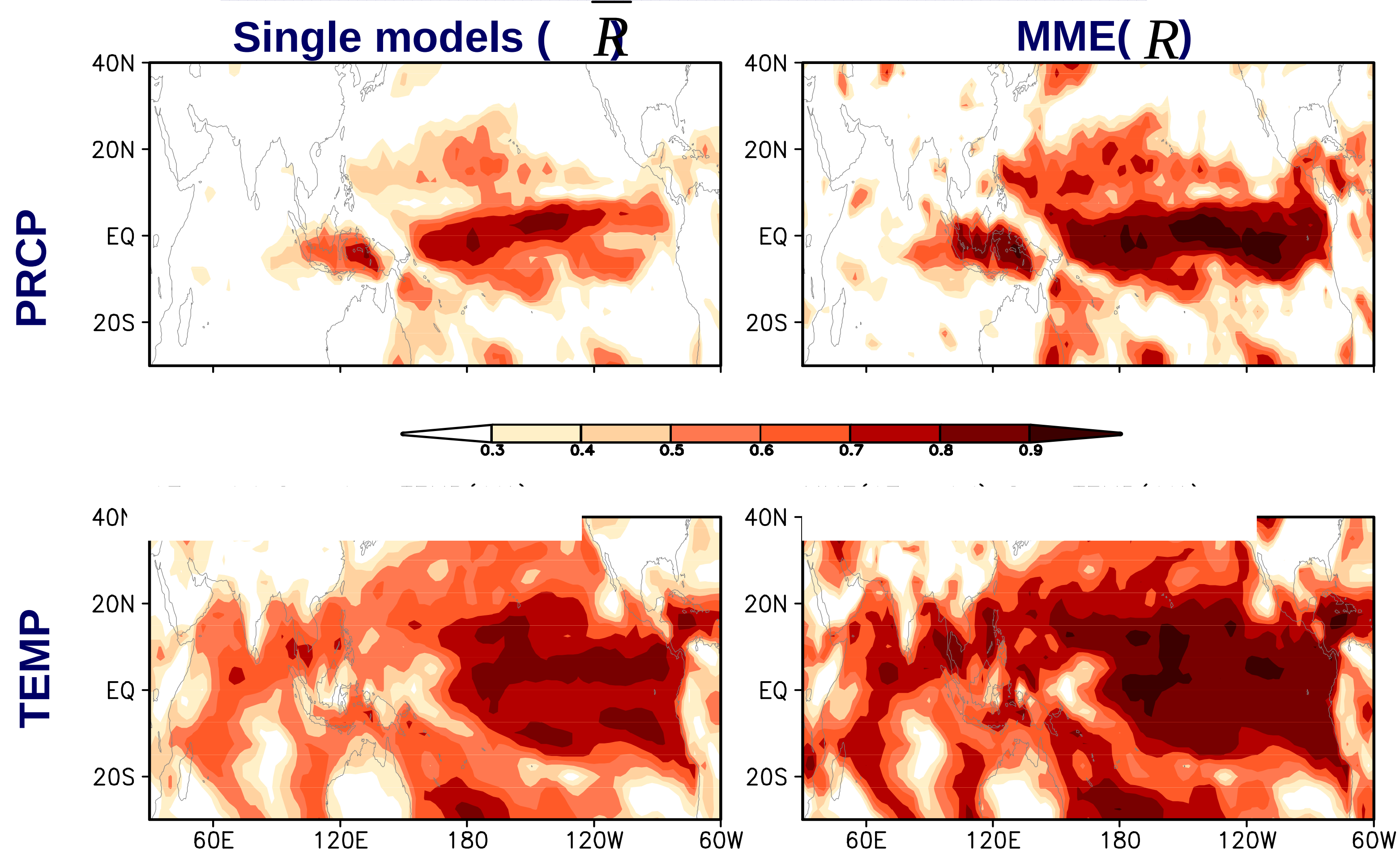
Mean Square Error $$E = E_M - E_s = \underbrace{-\frac{1}{2}\overline{V}(e) + \frac{1}{2}COV(e_1, e_2)}_{\text{Reduction of Systematic Error}} \underbrace{-\frac{1}{2}\overline{V}(\varepsilon)}_{\text{Reduction of Random Noise}}$$

- **For Improvement ➔ E < 0**
- **Where Models are independent each other ➔** $COV(e_i, e_j) = 0$

Correlation skill (JJA)

Correlation skill of MME $R = \frac{\mathrm{cov}(\bar{x}, y)}{\sigma(\bar{x}) \cdot \sigma(y)}$ (x_i: Model, y: Observation)

Mean correlation skill of Single Models $\bar{R} = \frac{1}{M}\sum_{i=1}^{M}\frac{\mathrm{cov}(x_i, y)}{\sigma(x_i)\cdot\sigma(y)}$



17 Models : CliPAS 10 models and DEMETER 7 models

Characteristics of each MME method

MME1

$$P = \frac{1}{M}\sum_i F_i$$

- simple composite
- equal weighting

MME2

$$P = \sum_i a_i F_i$$

- superensemble
- Weighted Ensemble

MME3

$$P = \frac{1}{M}\sum_i \hat{F}_i$$

- simple composite after correction

Issues on Multi Model Ensemble prediction

- **Is a multi model better than a single good model?**

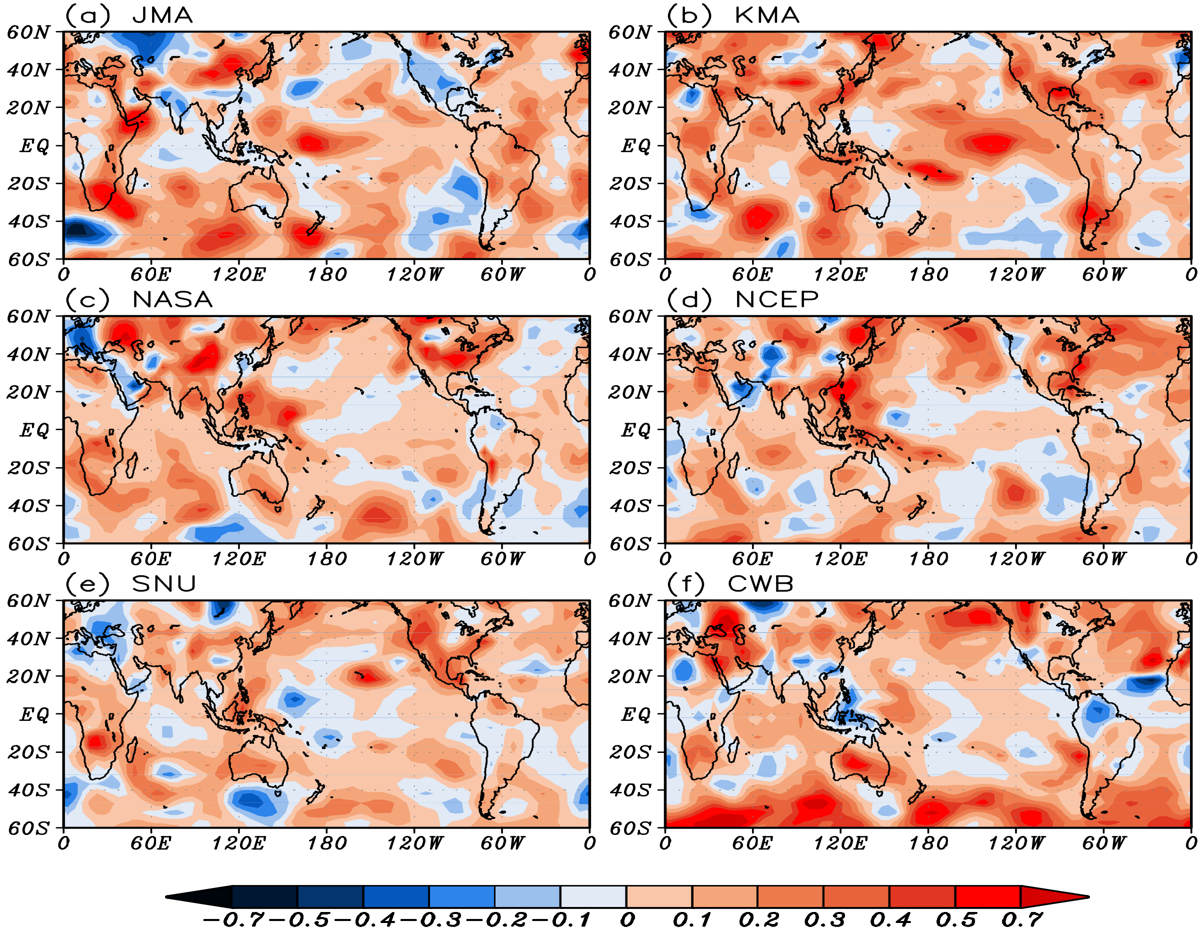
  (Graham et al. 2000; Peng et al. 2002; Doblas-Reyes et al. 2000)

- **Is a sophisticated technique better than a simple composite?**

  (Krishnamurti et al. 2000; Kharin and Zwiers 2002; Pavan and Doblas-Reyes 2000)

Multimodel ensemble forecast in GCM

Temporal correlation of JJA PRCP. : MME1 - single model



For the most of region,

the multi model ensemble (simple composite) is better than single model

Sampling error

Expected value of Normalized error variance of multi model ensemble

- Every model has same variance with observation.
- R : Correlation bet. Obs and model
- r : correlations among the models.

◆ **Predictand :** ϕ

◆ **Forecasts :** ϕ_1, ϕ_2

◆ **Assumption**

$$V(\phi) = V(\phi_1) = V(\phi_2) = \Phi$$

$$Cor(\phi, \phi_1) = Cor(\phi, \phi_2) = R$$

$$Cor(\phi_1, \phi_2) = r$$

$$\phi^* = \sum_{i=1}^{M} a_i \phi_i$$

$$MSE = \overline{(\phi^* - \phi)^2} = \overline{\phi^2} - 2\sum(a_i \overline{\phi_i \phi}) + 2\sum_i \sum_j a_i a_j \overline{\phi_i \phi_j} + \sum_i a_i^2 \overline{\phi_i \phi_i}$$

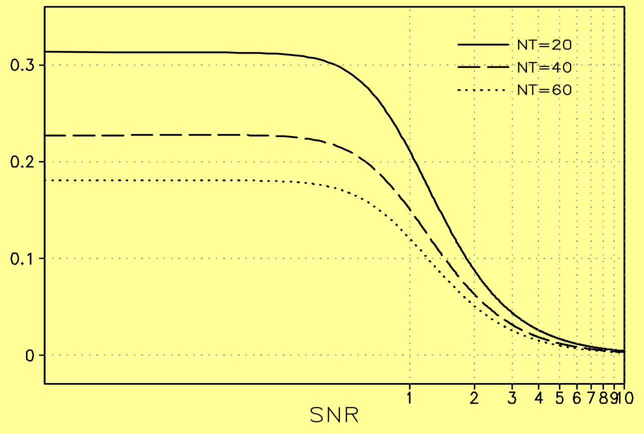
$$E^* = MSE / \Phi = 1 - 2\sum_i a_i R_i + 2\sum_i \sum_j a_i a_j r_{ij}(1 - \delta_{ij}) + \sum_i a_i^2$$

In the superensemble, the training period and forecast period should be different. Then there must be a difference in the statistics of 2 period.

R, R' is a statistical coefficient (correlation) of each period.

The difference of 2 coefficient depends on the sample size and signal to noise ratio

Due to the finite samples, Correlation in the training period (R, r) and forecast period (R′, r′) will be different



Impact of Sampling error in MME2

(a) Simple composite

$$a = \frac{1}{M} \qquad E^* = 1 - 2R' + \frac{(M-1)r' + 1}{M}$$

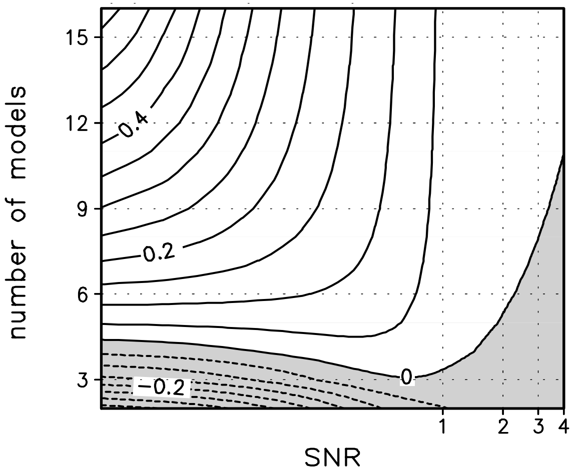
(b) Superensemble

$$a = \frac{R}{1+(M-1)r} \qquad E^* = 1 - \frac{MR'^2}{1+(M-1)r} + \frac{M(R-R')^2}{1+(M-1)r} + \frac{M(M-1)R^2}{(1+(M-1)r)^2}(r'-r)$$

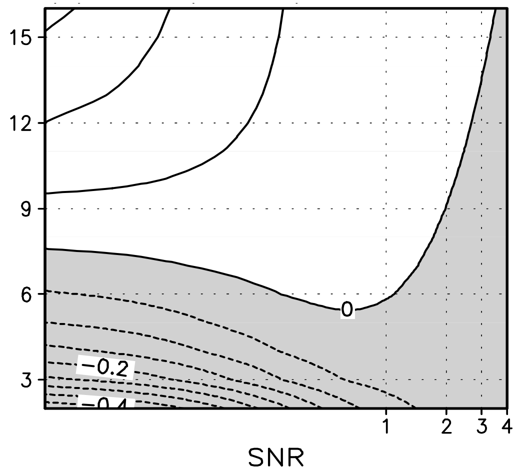
When the statistics of 2 period is different, Error in Superensemble increases

Error variances

(b)-(a) : NT = 20

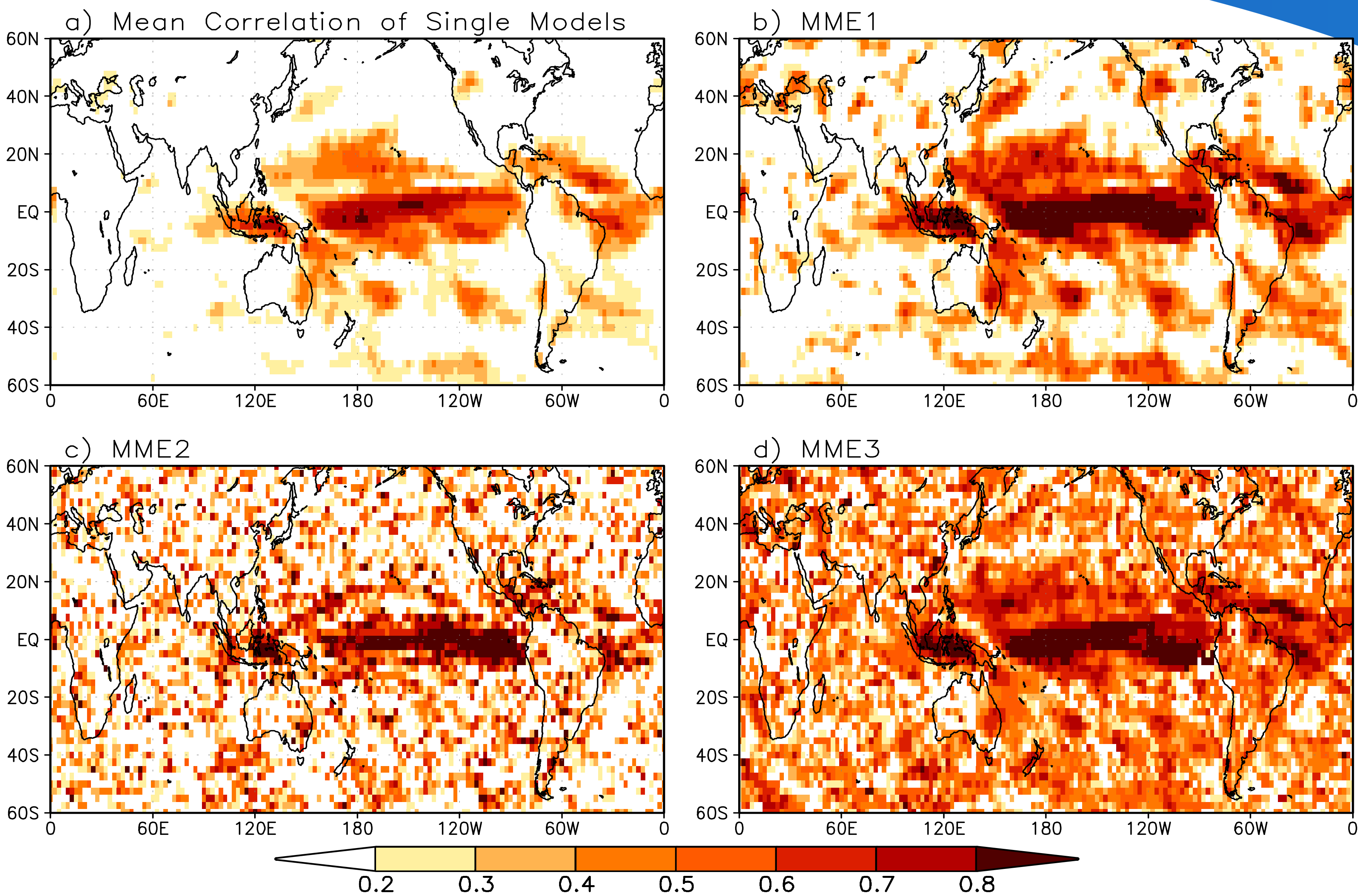


(b)-(a) : NT = 60



Since normal multiple regression has no procedure to prevent overfitting, superensemble is easy to fail in extratropics where the signal to noise ratio is small.

Correlation Skill of MME



- Noise dynamics

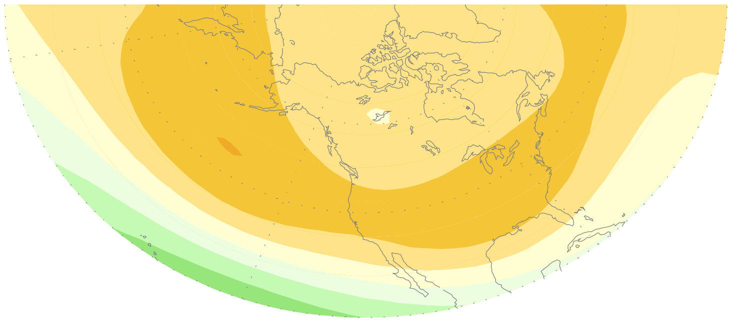
 - Predict an atmospheric noise interact with low-frequency

 - Understand the relationship between noise & low-frequency

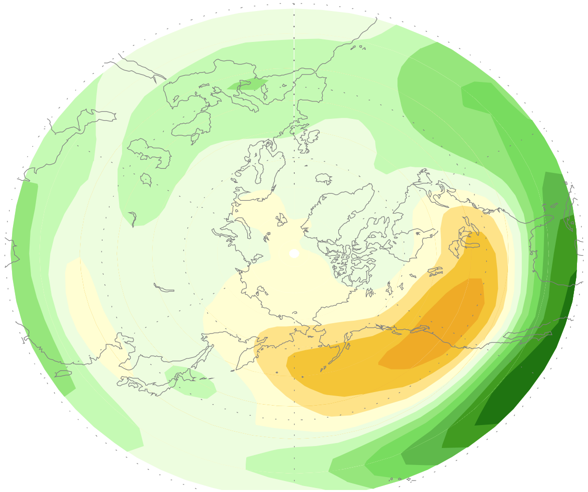
Relationship between seasonal mean flow & eddy activities

SVD 1st mode between 200mb streamfunction and eddy activity

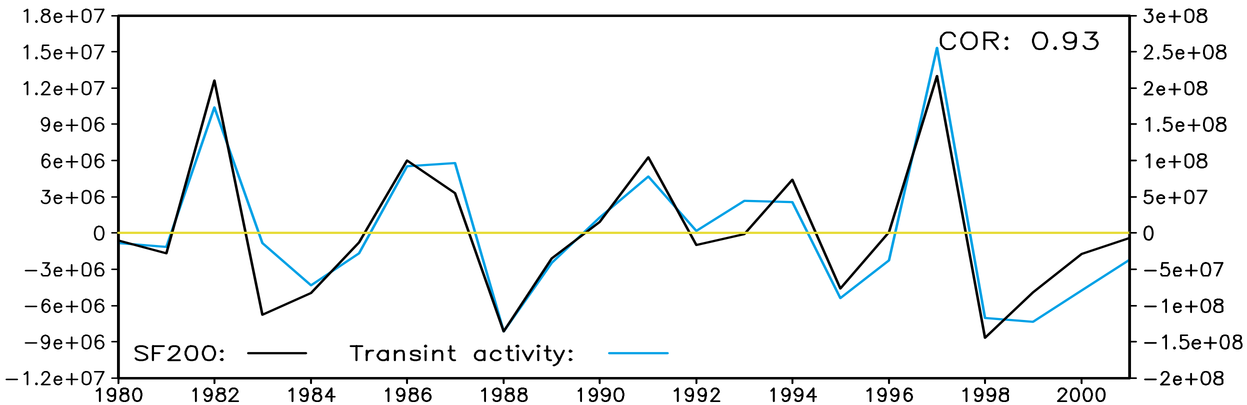
200 mb streamfunction (85.48 %)



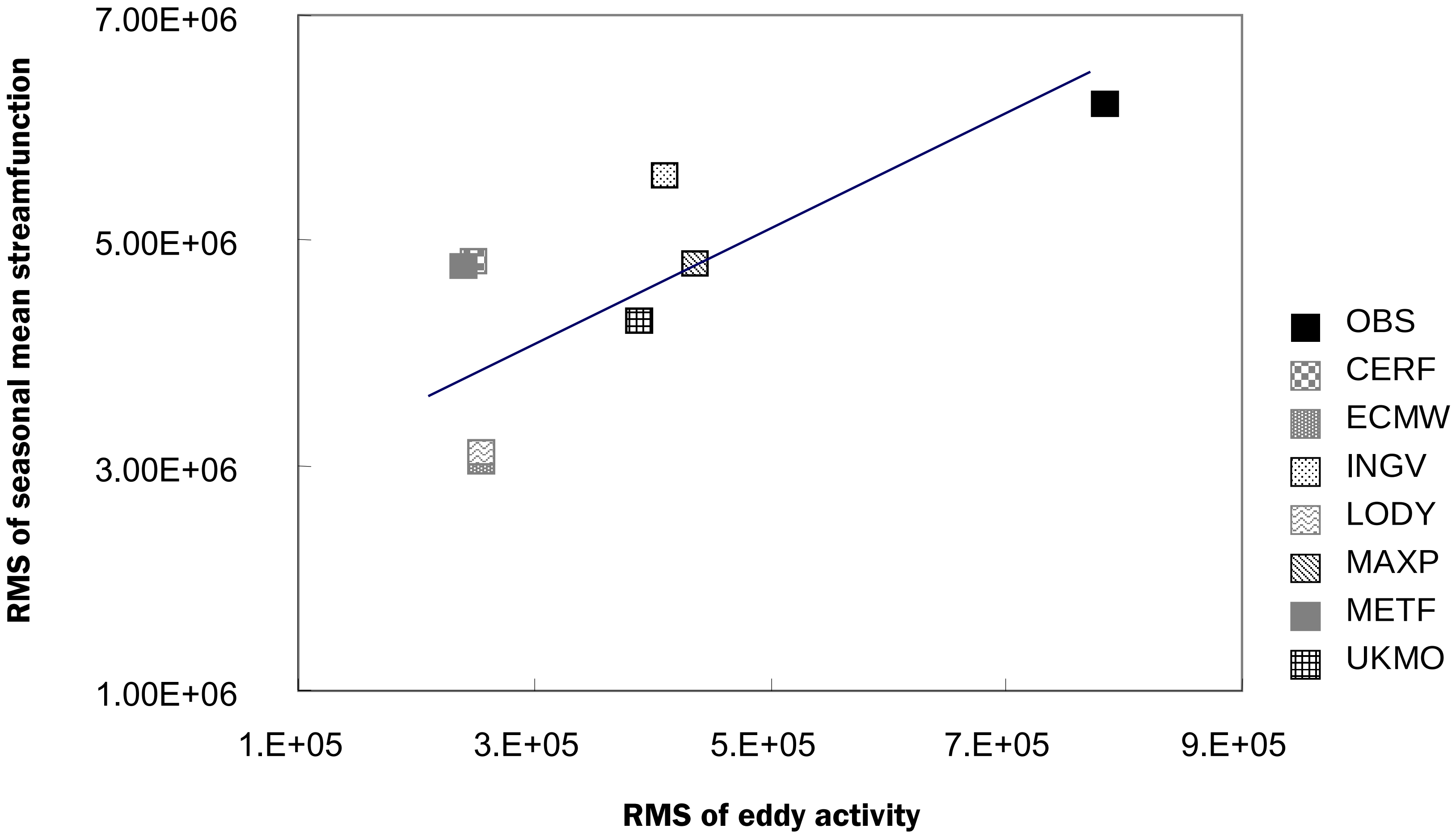
Eddy Activity (69.12 %)



Time Series



Relationship between eddy activity and seasonal mean variability



Eddy activity is proportional to Seasonal mean variability

Dynamical eddy forcing correction

$$L_o(\delta X) = \boxed{E_o - E_m} - (L_o X_m - L_m X_m)$$

Seasonal mean Forecast Error

Error of model eddy forcing from the observed

O : Observation
m : GCM output data

$$L_j X_j = E_j + R_j \ , \ j = 1, \ldots, J$$

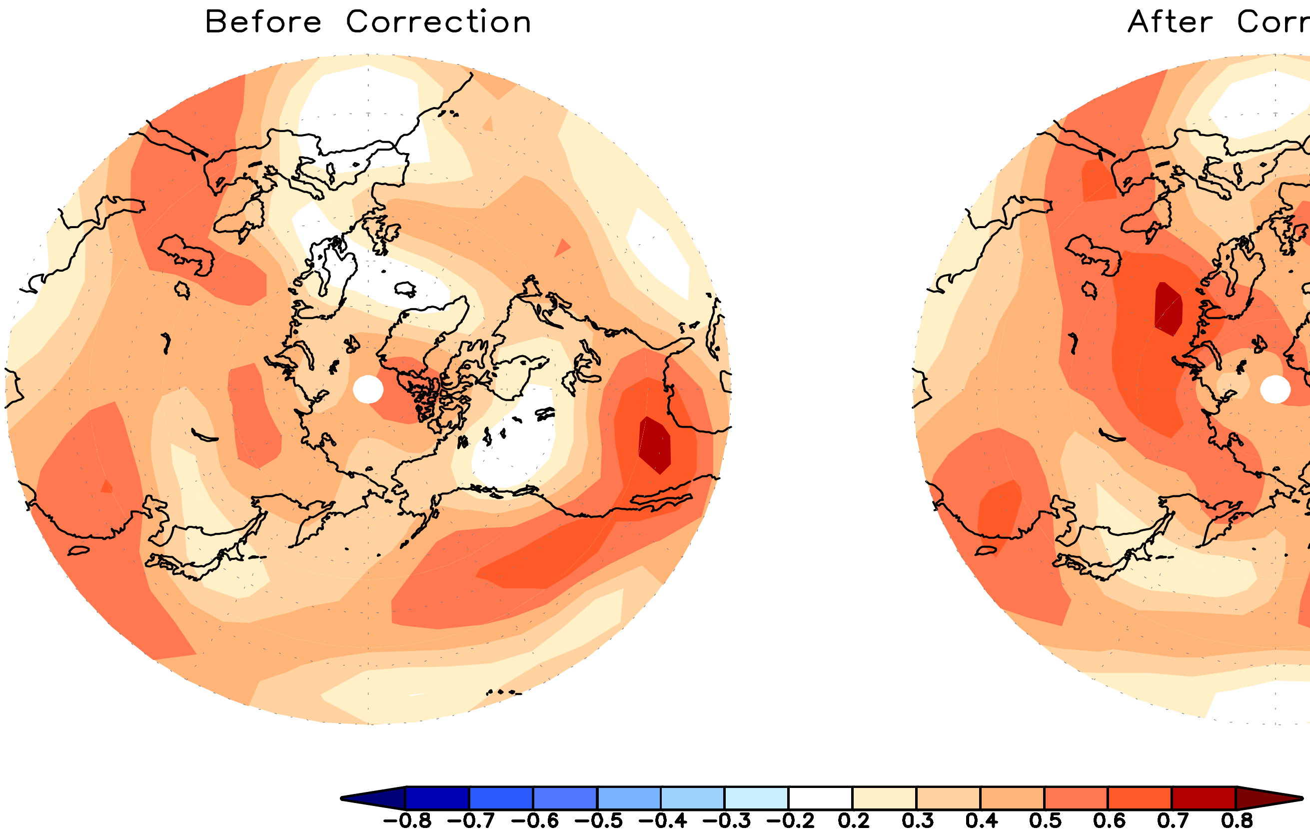
E_j **: The eddy forcing**

R_j **: The residual term**

L_j **: Linearized barotropic model**

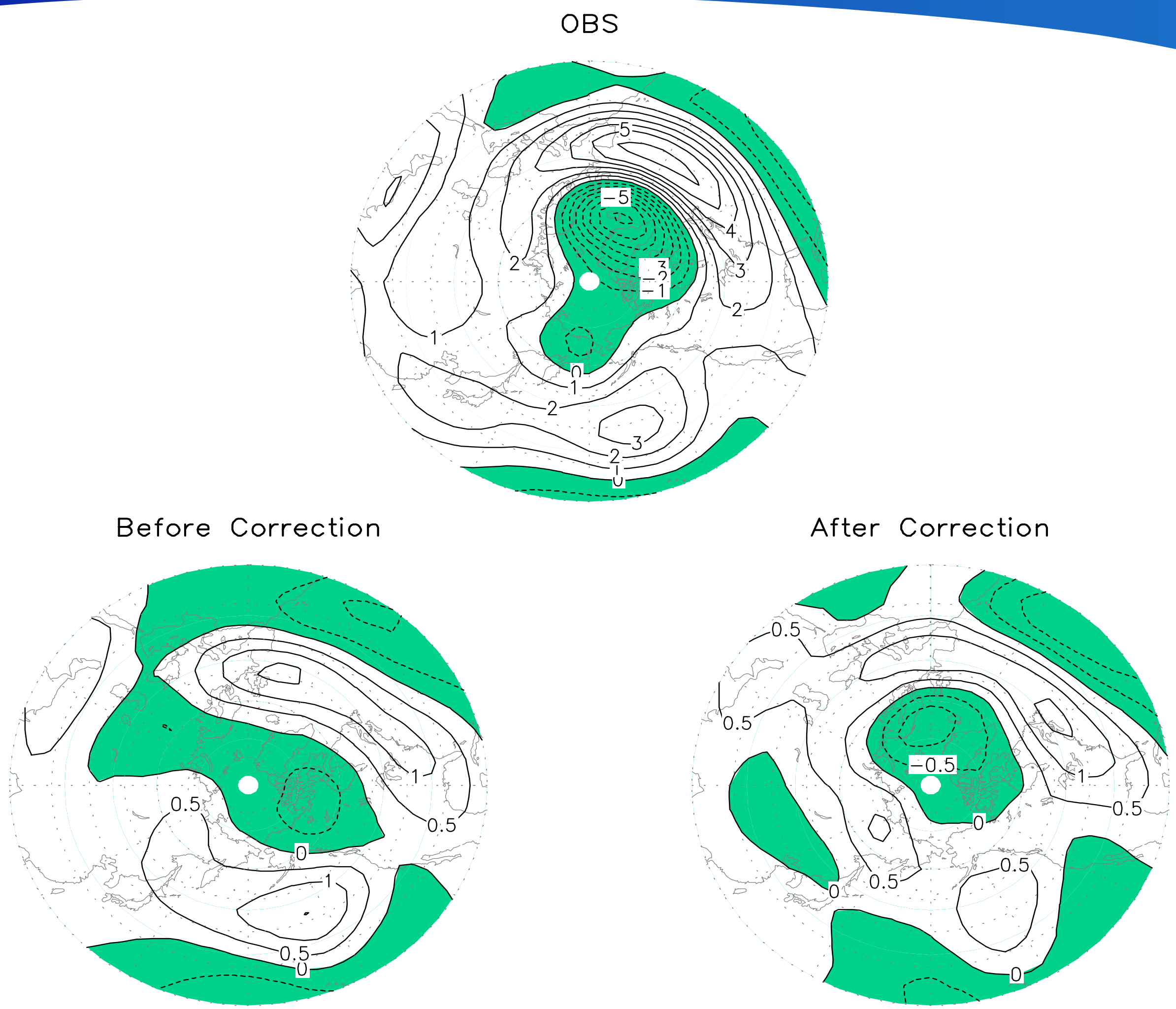
Dynamical Correction - Eddy forcing correction

Correlation Skill of 300mb streamfunction



Before Correction	After Correction
0.32	0.40

Dynamical Correction - Streamfunction regressed by observed NAO index



Before Correction	After Correction
0.54	0.74

Other Important Issues

1. Initialization

2. Model improvement

   - Physical parameterization

   - High resolution modeling

3. Subseasonal (MJO) prediction