Tralynn Chen
Statistician
Central Weather Bureau

多變量統計後處理

Empirical Methods of Multivariate Ensemble Statistical Post Processing : Affine Kernel Density and Shuffling Technique

AGENDA

Introduction and Background

Univariate Statistical Post Processing

Multivariate Ensemble Statistical Post Processing

Research Examples

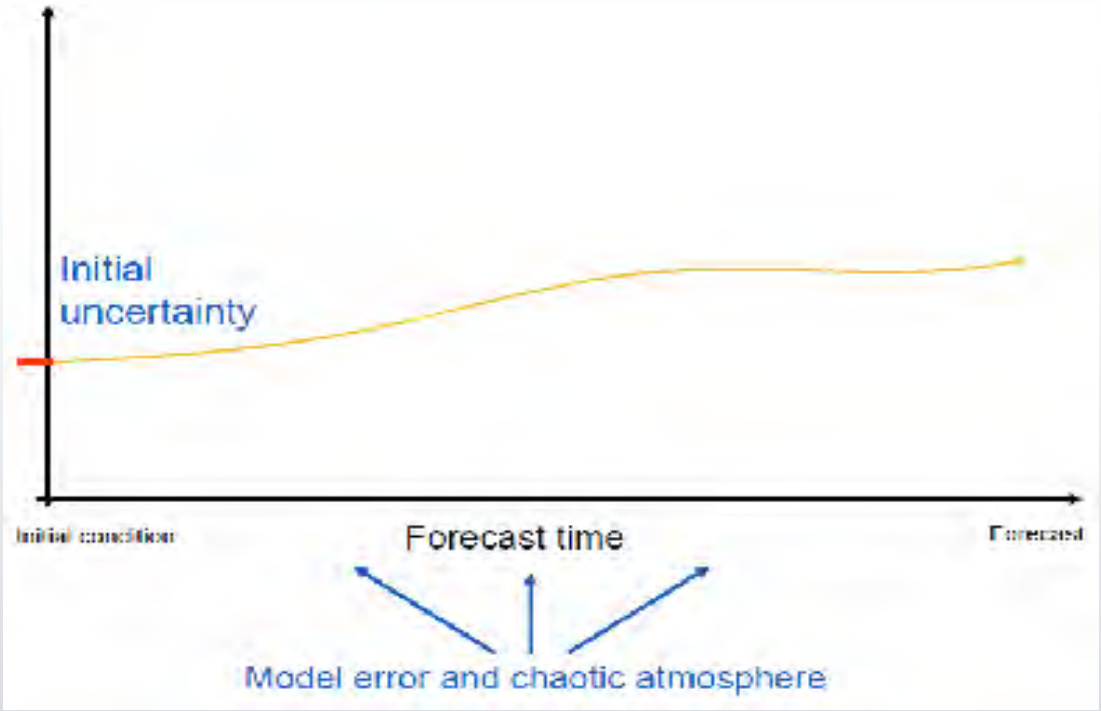
Summary

Introduction and Background

Numeric Weather Prediction System (NWP)

- Model Imperfection.
- Deficit Initial Condition
- Probability forecasting based on multiple members from NWP

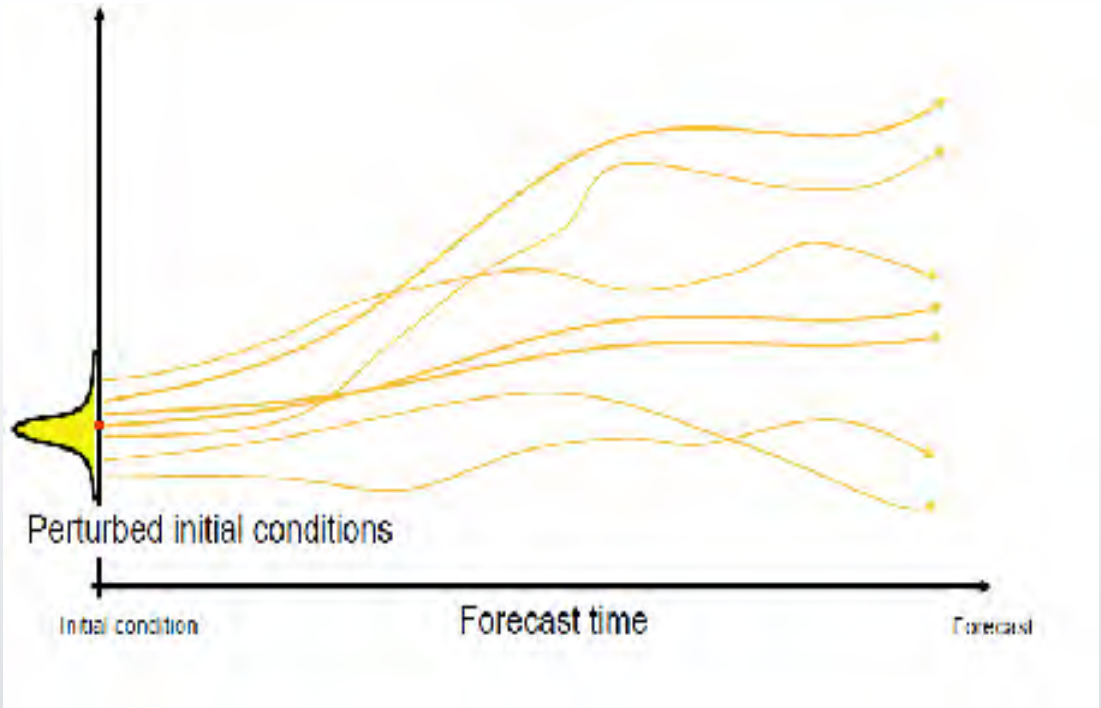
Result In : Forecast error increase as integrated along the forecast leadtime.



Numeric Weather Prediction System (NWP)

- Model Imperfection.
- Deficit Initial Condition
- Probability forecasting based on multiple members from NWP

Result In : Forecast error increase as integrated along the forecast leadtime.



Develop Model Output Statistics (MOS)

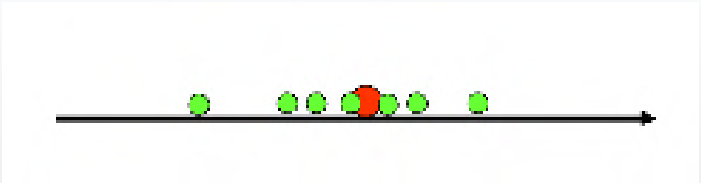
Model Output Statistics

- Bayesian Model Average (BMA)
- (Extended) Logistic Regression
- Ensemble Model Output Statistics
- Affine Kernel Dressing

Methodology : Applied SPP on specific variable, with each forecast leadtime.

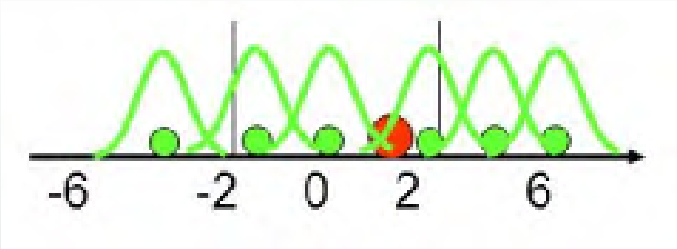
Affine Kernel Dressing

- Forecast from Ensemble members of NWP system



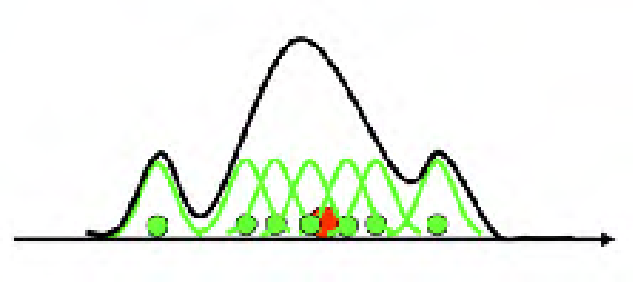
Affine Kernel Dressing

- Forecast from Ensemble members of NWP system
- Applies unit kernel function, K (e.g. Gaussian), to each ensemble member x_i



Affine Kernel Dressing

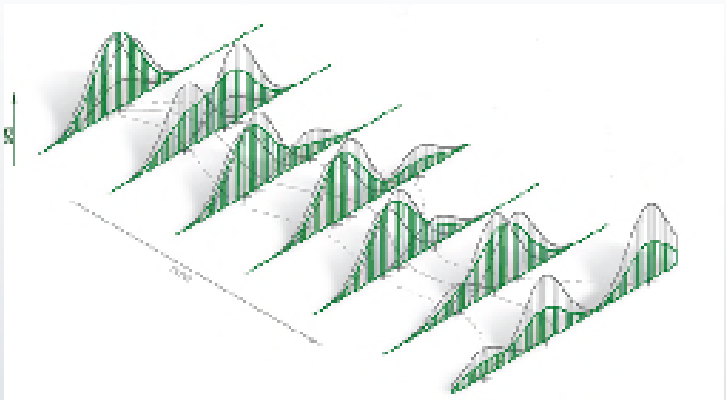
- Forecast from Ensemble members of NWP system
- Applies unit kernel function, K (e.g. Gaussian), to each ensemble member x_i
- Combines and sum up the kernel functions



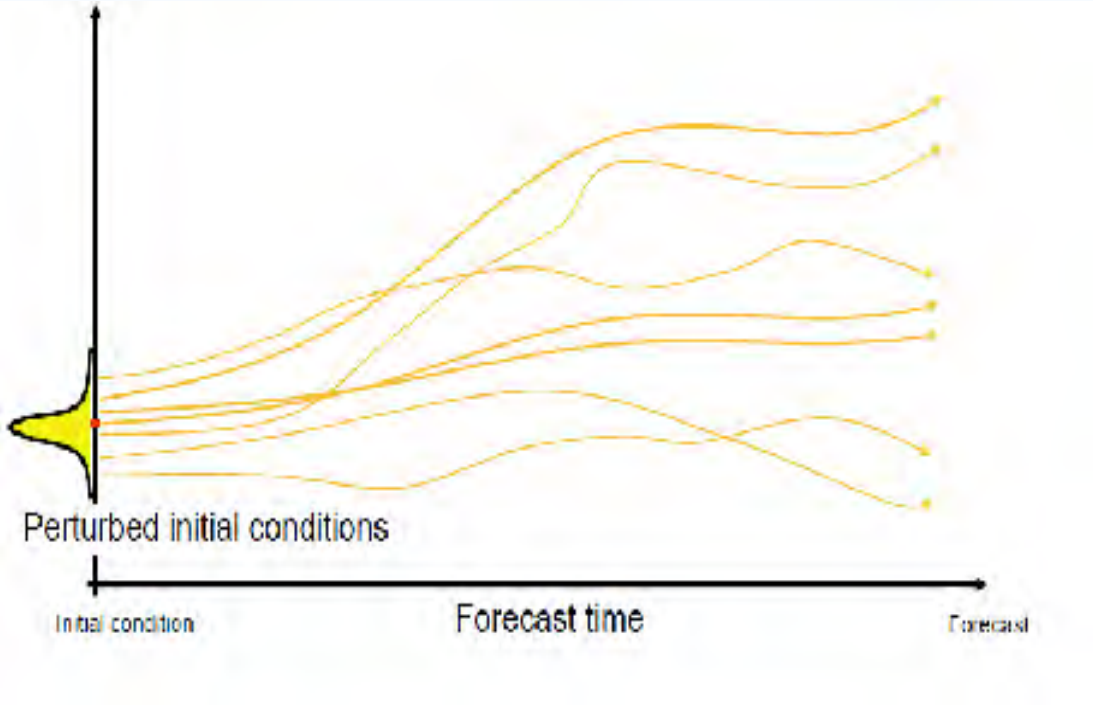
Affine Kernel Dressing (Brocker and Smith, 2008)

- Applies unit kernel function, K(e.g. Gaussian), to each ensemble member x_i and then combines them:

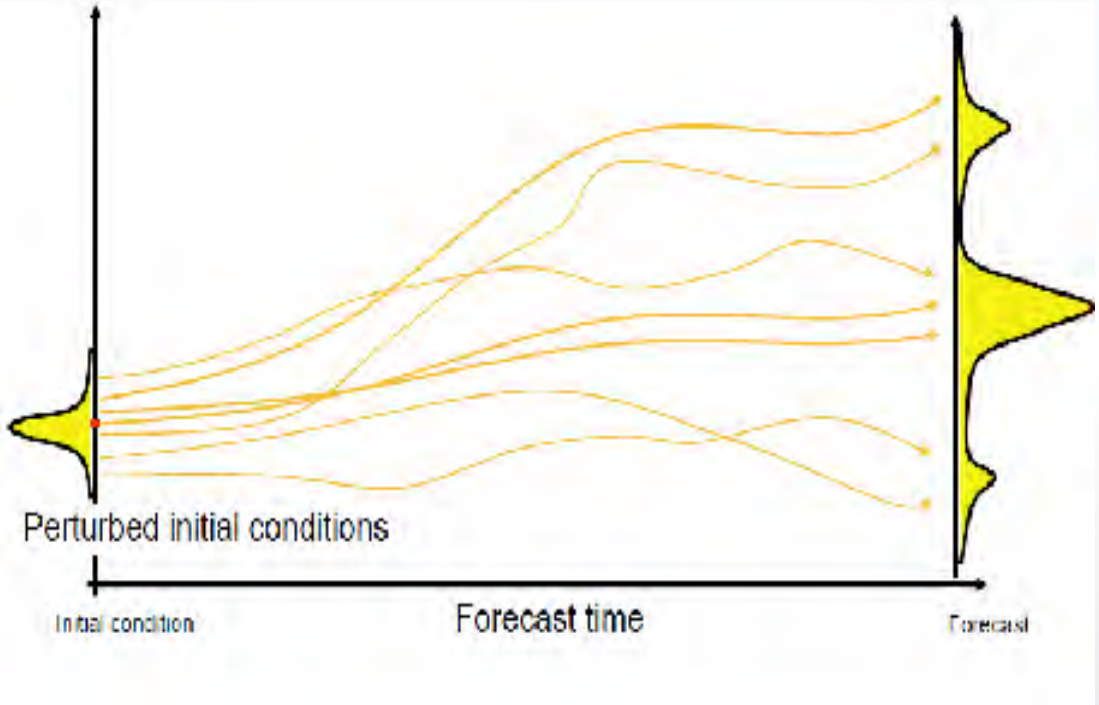
$$p(y; x, \theta) := \frac{1}{N\sigma} \sum_{i=1}^{N} w_i \cdot K\left(\frac{y - ax_i - \omega}{\sigma}\right)$$



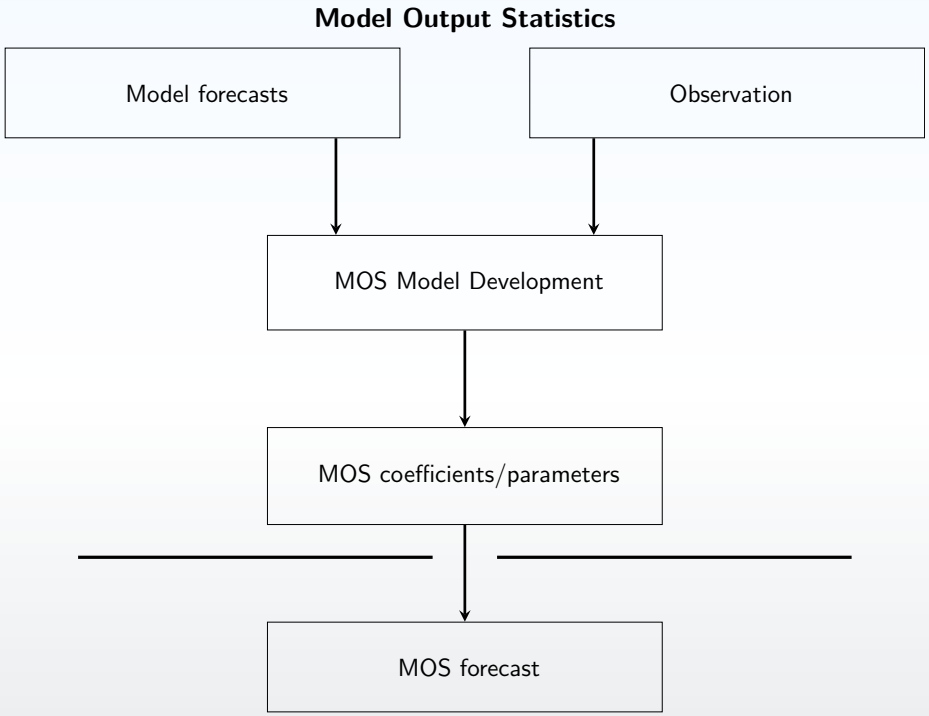
Ensemble Members from NWP



Ensemble Members from SPP



Development of AKD parameter



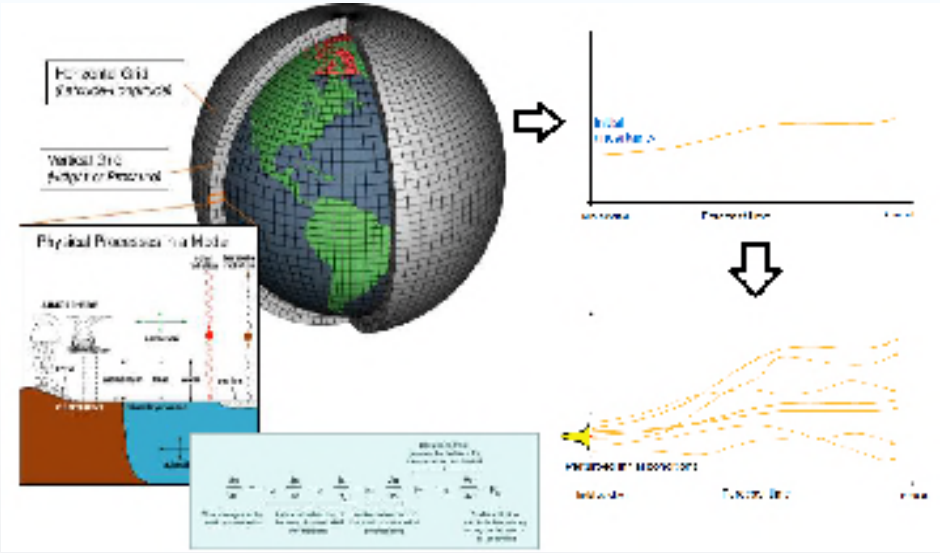
AKD Model Development

- Reforecast/Hindcast Data
- Learning/Training process for AKD parameter

AKD parameters for the further output processing

Embryonic form on Machine Learning

Global Forecast Model

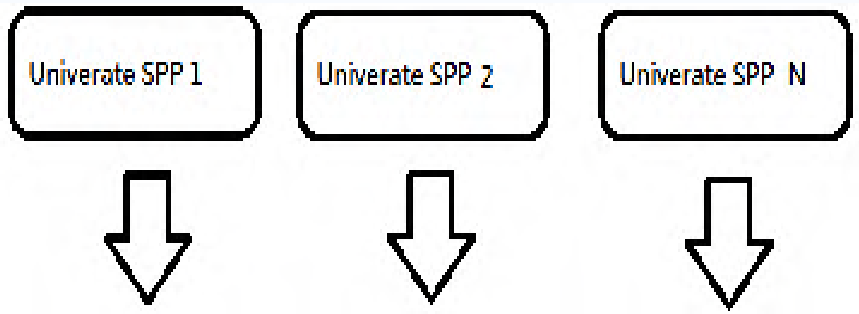


Multivariate Definition

- Several variables (t2m, pressure,...) is included in a Model
- Variables are presented in a grid with several layers
- Forecast leadtime frame a time dimension in a Model

Motivate Multivariate Ensemble Statistical Post-Processing

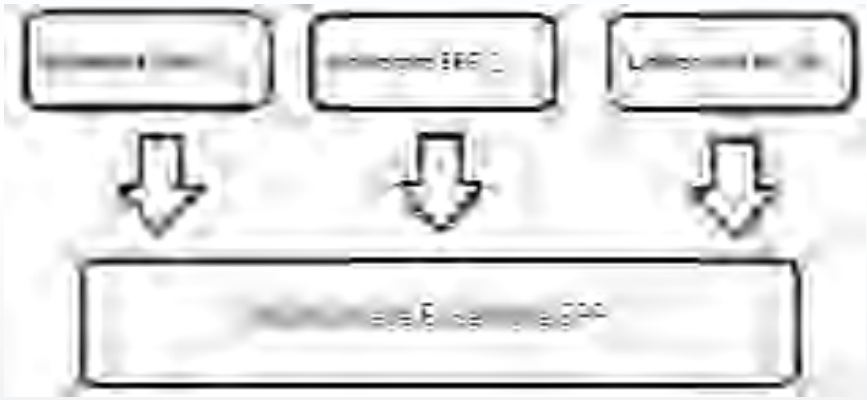
Scope of Multivariate



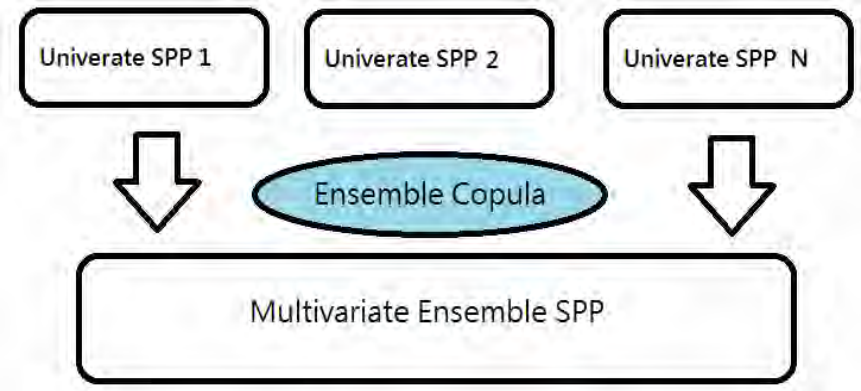
Developing Multivariate Ensemble Statistical Post Processing

- Inter-Variable
- Spatial Variable
- Time Series

From Univariate Statistical Post Processing to Multivariate Ensemble Statistical Post Processing



From Univariate Statistical Post Processing to Multivariate Ensemble Statistical Post Processing



Dependence Modeling – Sklar Theorem (Sklar, 1959)

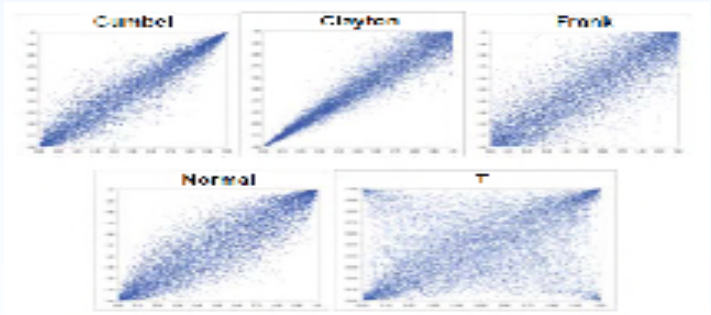
For every joint probability distribution H, there is a copula C such that:

$$H(x, y) = C(F(x), G(y))$$

If F_{XY} is continuous then C is unique

- C Copula function
- F marginal distribution of X
- G marginal distribution of Y

Parametic Coupla



Assume variable $y(i) = Y(v, s, \tau)$ where $i = 1, \ldots, M$ and

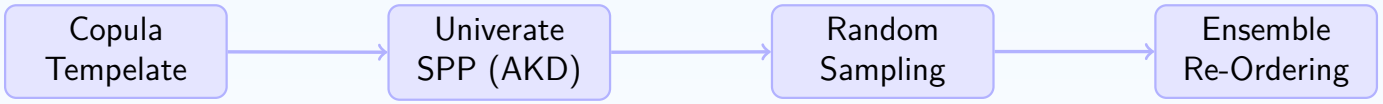
- v : Variable
- s : Location
- τ : Forecast Leadtime

$$F(y_1, \ldots, y_M) = C(F(y_1), \ldots, F(y_M))$$

Suppose 5 variables (t2m,tp,rh,...) with 10 layers and $100X100$ grids on 4-week forecast leadtime

Problem with Parametric Copula : High Dimension Parameter!

Nonparametric Method : Schaake's Shuffle



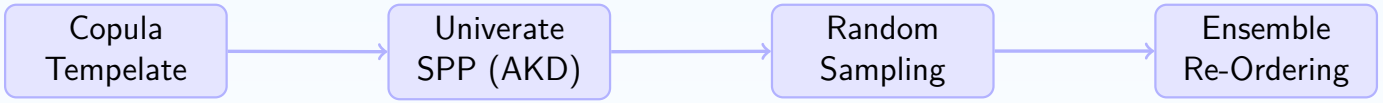
Copula Template

- Let x be current forecast, x_H be (re)forecast data set in the past,

$$\Delta^{t_d} = \sqrt{(\mu(x) - \mu(x_H))^2 + (\sigma(x) - \sigma(x_H))^2}$$

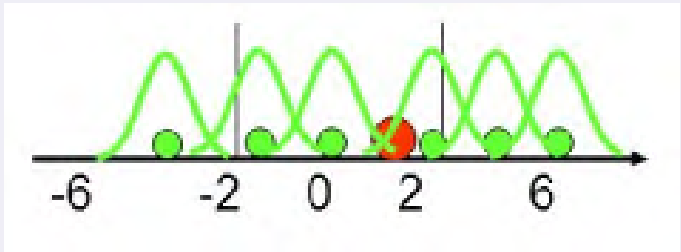
  is the similarity criterion measurement.
- Find the top Δ^t and align it with its corresponding observation data
- Rank the observation data as the ensemble copula

Nonparametric Method : Schaake's Shuffle



Univariate SPP

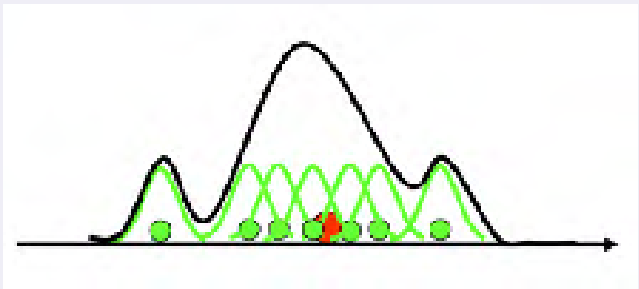
Applied AKD for each location, time and variable and obtain marginal CDF



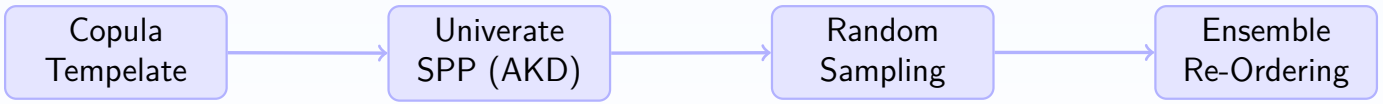
Nonparametric Method : Schaake's Shuffle

Random Sampling

Random sampling from marginal CDF of forecast



Nonparametric Method : Schaake's Shuffle



Ensemble Re-Ordering

Applied copula template on new sampling data for forecast.

Example on Schaake's Shuffle

Assume with a data set at location j and time k with two variables of ensemble size N=5 is describing by the following:

$$Y = \{y^1, y^2\} = \begin{bmatrix} y_1^1 & y_1^2 \\ y_2^1 & y_2^2 \\ y_3^1 & y_3^2 \\ y_4^1 & y_4^2 \\ y_5^1 & y_5^2 \end{bmatrix} = \begin{bmatrix} 11 & 2.2 \\ 14 & 2.3 \\ 13 & 2.5 \\ 12 & 2.1 \\ 15 & 2.4 \end{bmatrix} \Rightarrow R = \begin{bmatrix} 1 & 2 \\ 4 & 3 \\ 3 & 5 \\ 2 & 1 \\ 5 & 4 \end{bmatrix}$$

Example on Schaake's Shuffle

Assume with a data set at location j and time k with two variables of ensemble size N=5 is describing by the following:

$$Y = \{y^1, y^2\} = \begin{bmatrix} y_1^1 & y_1^2 \\ y_2^1 & y_2^2 \\ y_3^1 & y_3^2 \\ y_4^1 & y_4^2 \\ y_5^1 & y_5^2 \end{bmatrix} = \begin{bmatrix} 11 & 2.2 \\ 14 & 2.3 \\ 13 & 2.5 \\ 12 & 2.1 \\ 15 & 2.4 \end{bmatrix} \Rightarrow R = \begin{bmatrix} 1 & 2 \\ 4 & 3 \\ 3 & 5 \\ 2 & 1 \\ 5 & 4 \end{bmatrix}$$

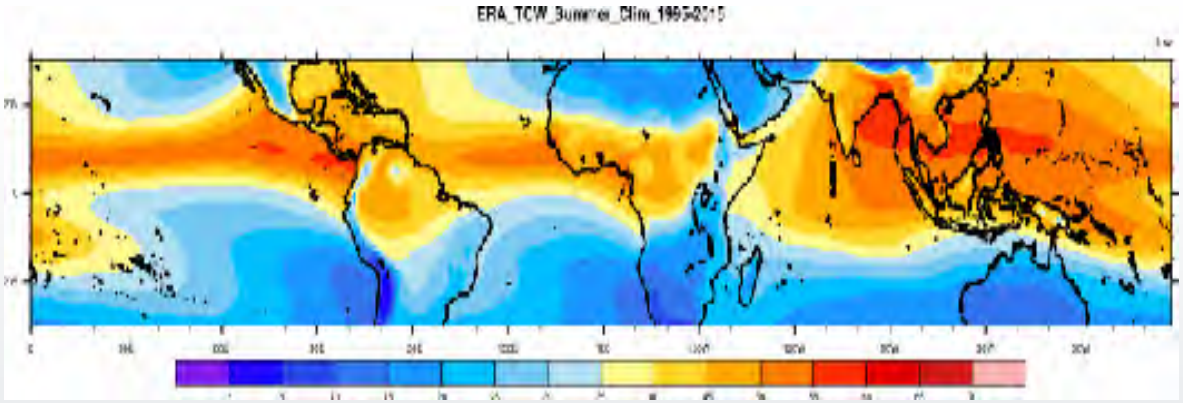
Present forecast data for the two variables as the following:

$$Y' = \{y^1, y^2\} = \begin{bmatrix} 10 & 2.8 \\ 19 & 2.6 \\ 15 & 2.4 \\ 17 & 2.2 \\ 13 & 2.0 \end{bmatrix} \Rightarrow X = order \begin{bmatrix} 1 & 2 \\ 4 & 3 \\ 3 & 5 \\ 2 & 1 \\ 5 & 4 \end{bmatrix} (Y') = \begin{bmatrix} 10 & 2.2 \\ 17 & 2.4 \\ 15 & 2.8 \\ 13 & 2.0 \\ 19 & 2.6 \end{bmatrix}$$

Therefore, rank(X) = rank(Y)

MESPP Example

- ECMWF model
- Total column water (tcw)
- Reforecast data time frame : 1995 - 2015

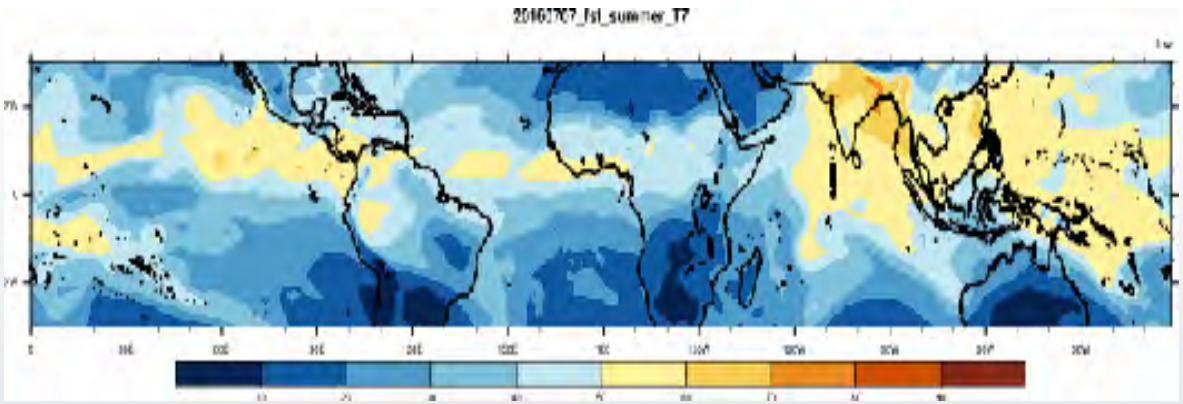


summer average on tcw during 1995-2015 (source: EC-Interim)

AKD

AKD Parameter

Reforecast data set from 1995-2015 to train parameter through AKD scheme of MOS. Then applied kernel unit on the following:

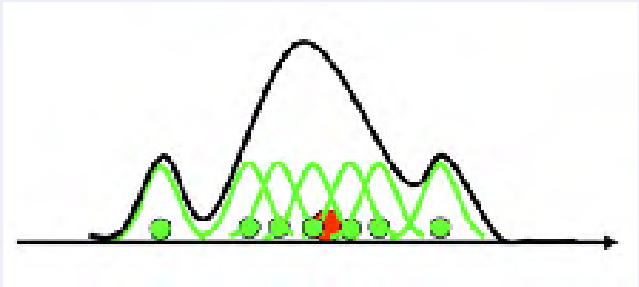


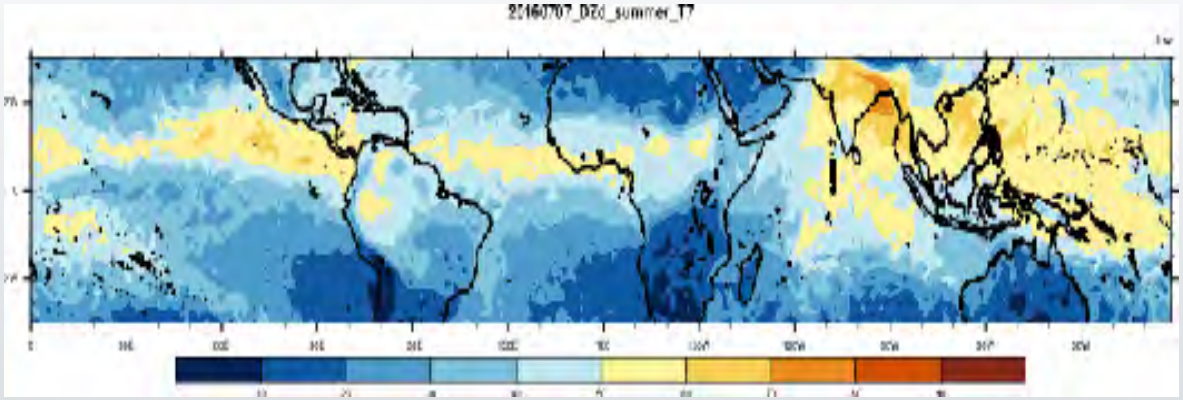
20160707 forecast output with forecast leadtime 7

AKD : Random Sampling

Random Sampling

Random sampling according to akd-derived pdf

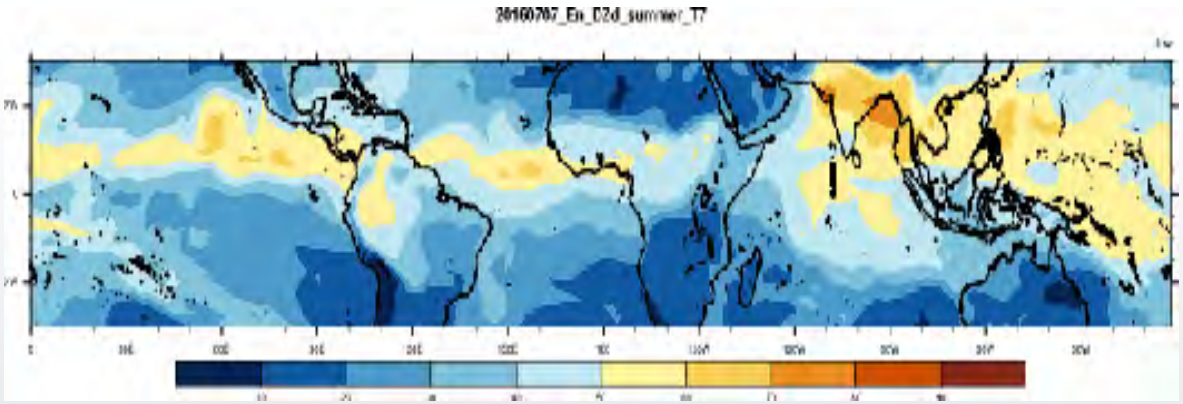




Ensemble Copula

Reordering AKD sample

Apply Ensemble Copula on to AKD sampling for forecast SPP



AKD and Shuffle on 20160707 forecast output with forecast leadtime 7

Skill Score Define

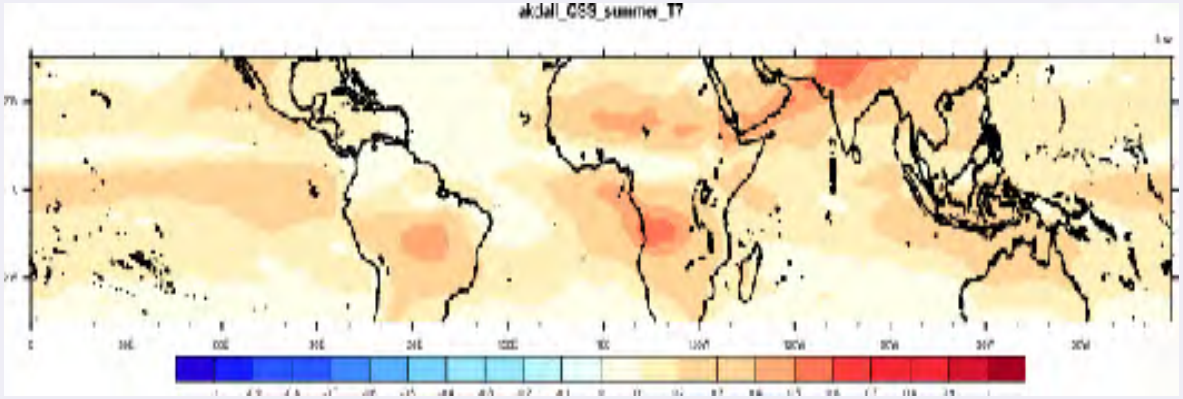
Skill Score

Brier Skill Scores (Murphy and Winkler 1992)

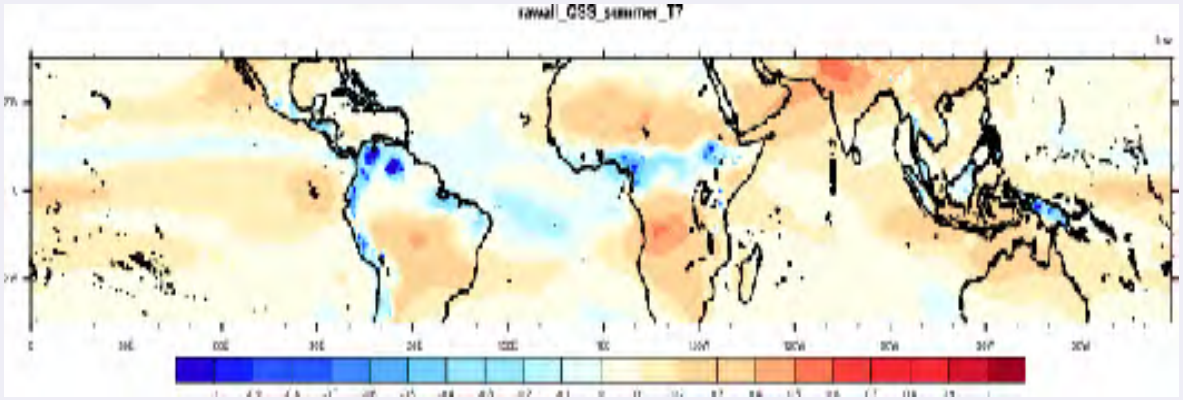
$$SS = \rho_{fx}^2 - \left[\rho_{fx} - \frac{\sigma_f}{\sigma_x}\right]^2 - \left[\frac{\mu_f - \mu_x}{\sigma_x}\right]^2$$

Forecast lead time = 7 / Summer

AKD

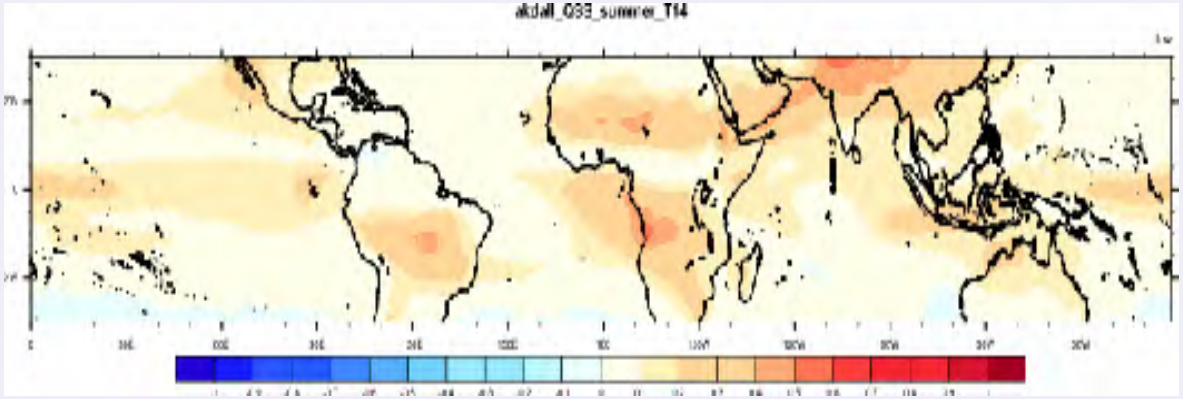


FST

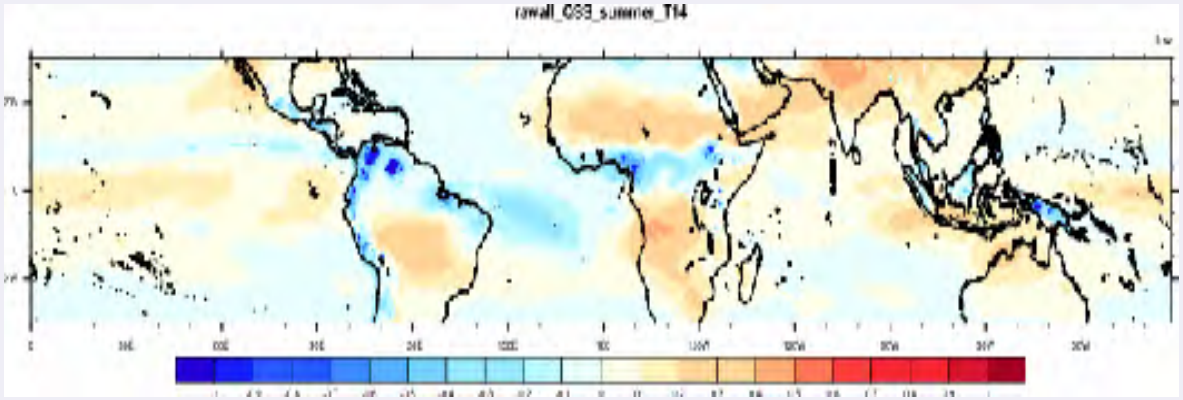


Forecast lead time = 14 / Summer

AKD

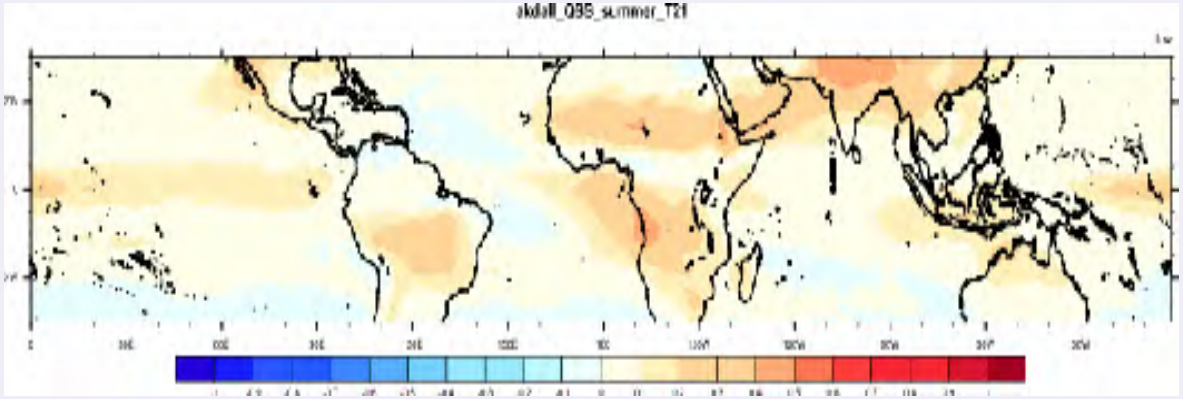


FST

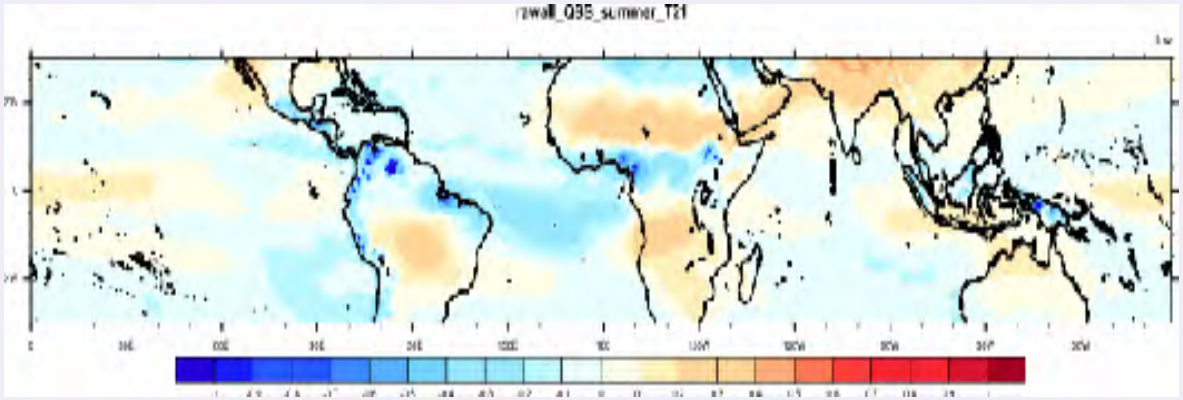


Forecast lead time = 21 / Summer

AKD

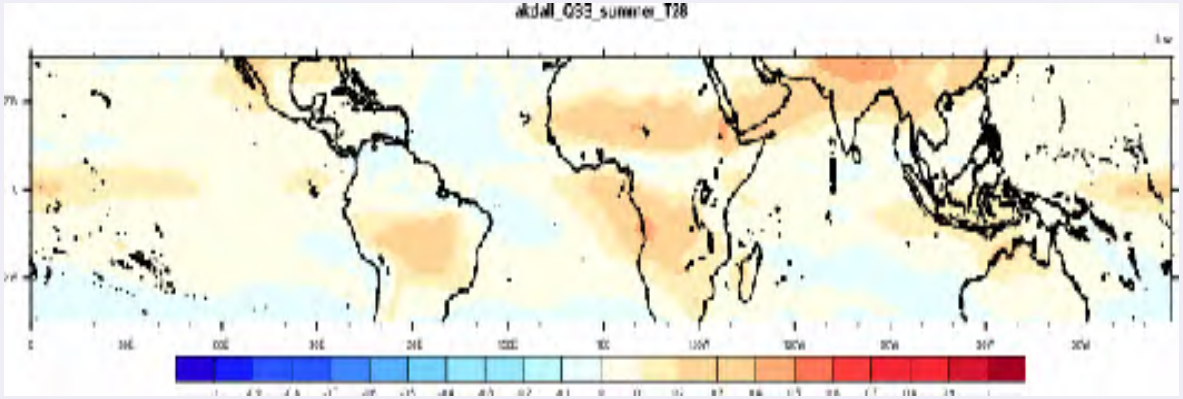


FST

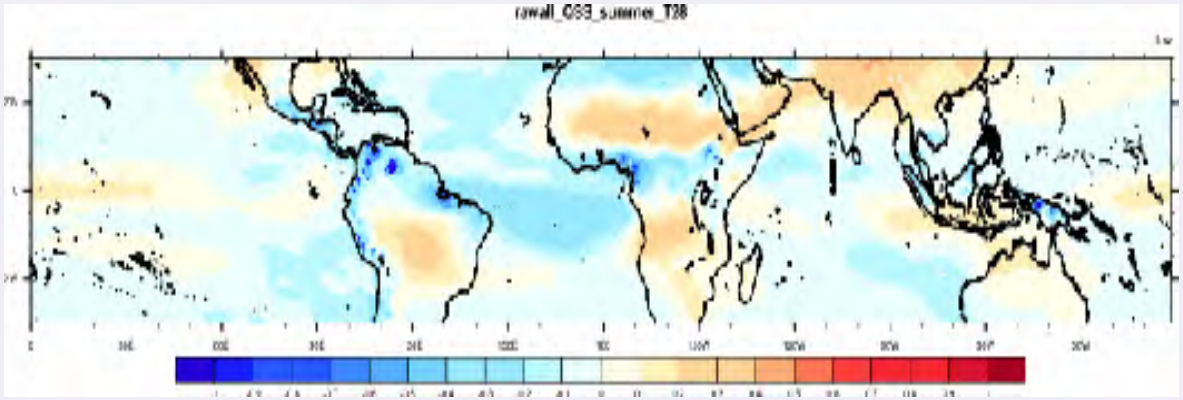


Forecast lead time = 28 / Summer

AKD



FST

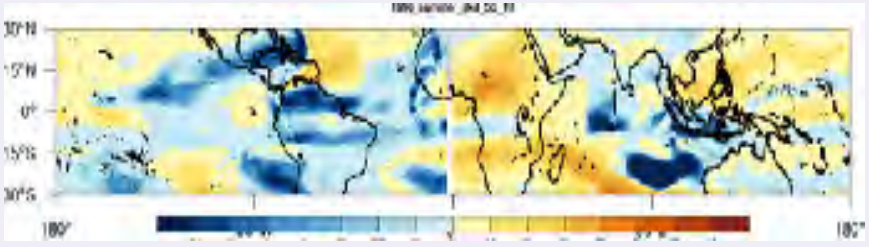


- Potential bulky data storage
- Corresponding reliance of observation data
- Consumption on Computational time

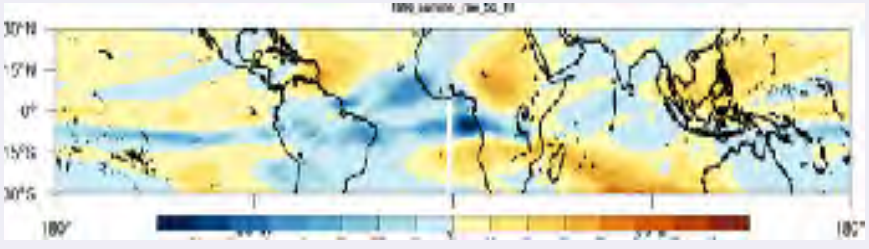
Skill Score of AKD with Cross-Validation on the year 1996 (1995-2016)

Source of observation : Initial forecast time at 00

AKD



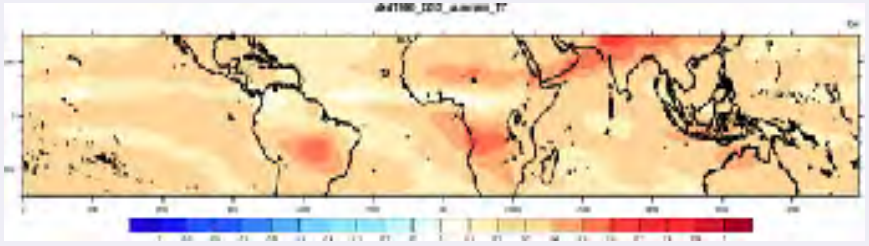
FST



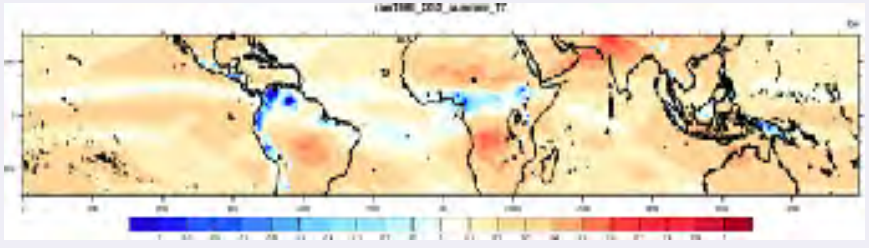
Skill Score of AKD with Cross-Validation on the year 1996 (1995-2016)

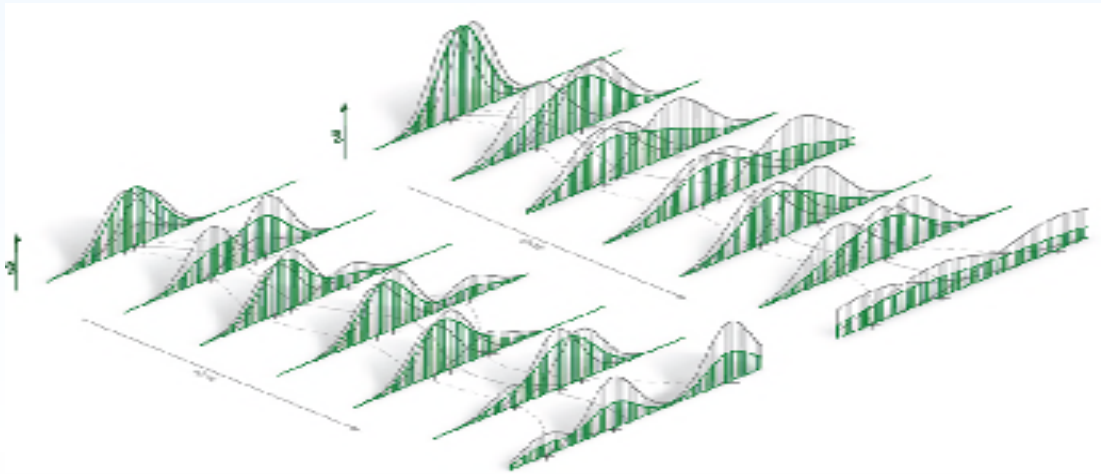
Source of observation : EC-Interim

AKD



FST





- Different scheme on univariate statistical post processing
- Methodologies on Similarity Criterion
- Choices on Copula

Summary

- Model output statistics could be one of many schemes of machine learning
- Stock up history/hindcast/reforecast data is necessary
- AKD could increase skill score with validate corresponding history observation data
- Copula is the pathway from univariate SPP to multivariate SPP

QUESTION

Thank You