Re-visiting seasonal predictions using subseasonal scenarios extracted by fuzzy k-means

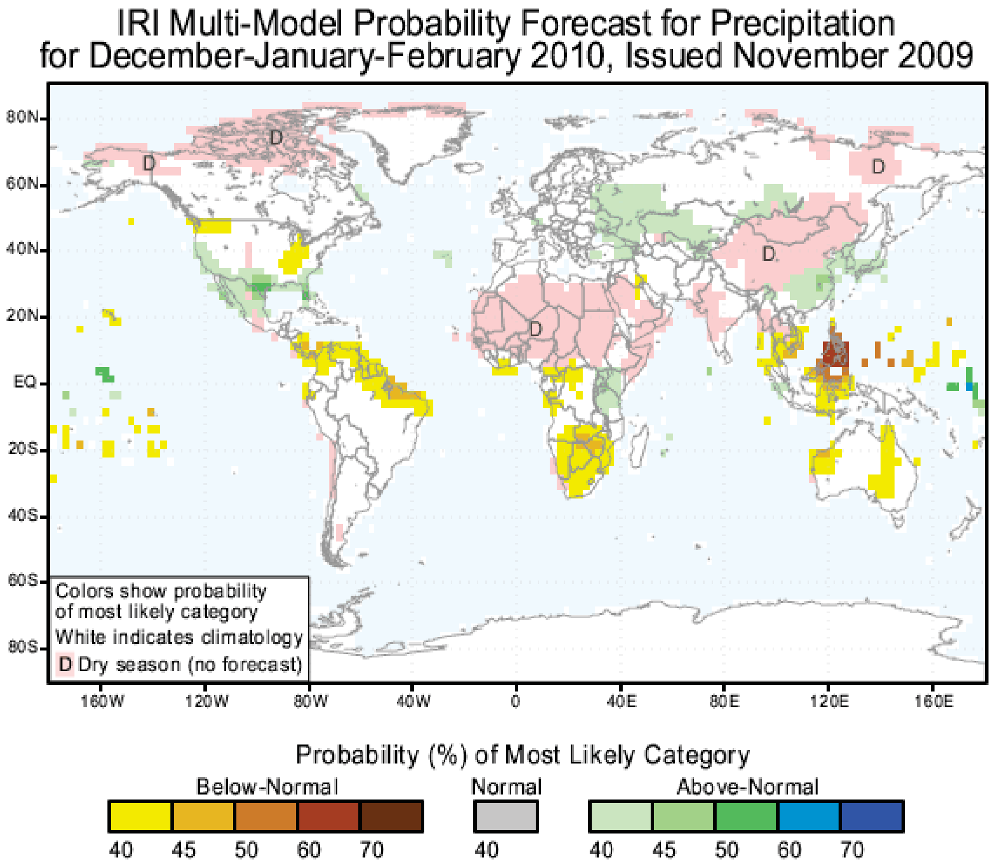
Vincent Moron
Aix-Marseille Université, CEREGE and IRI

collaborative work with **Andrew W. Robertson** (IRI) and **Pierre Camberlin** (U. Bourgogne)

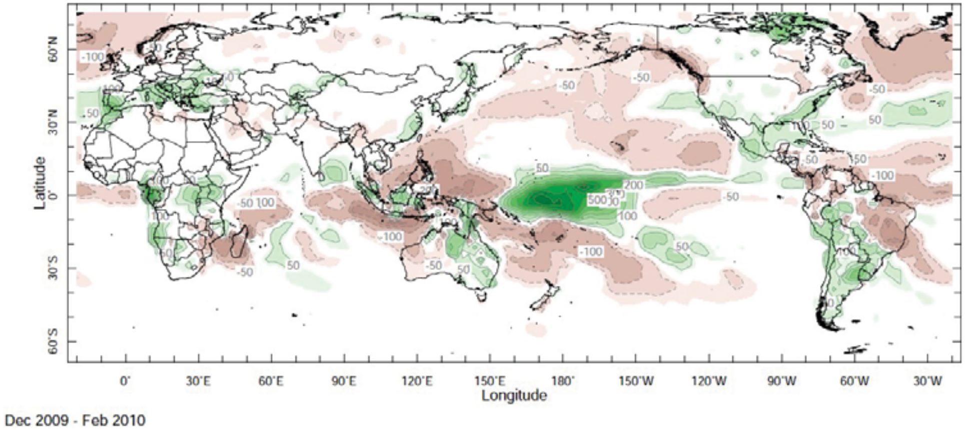
moron@cerege.fr & research page at
https://www.researchgate.net/profile/Vincent_Moron/?ev=hdr_xprf

Current seasonal predictions of tropical rainfall (i.e. Goddard et al., 2001) ?
- at least 3-month amount of rainfall (expressed as interannual anomalies)
- grid-points and over (till regional-scale or domanial scale as All-India rainfall)

Predicted anomalies



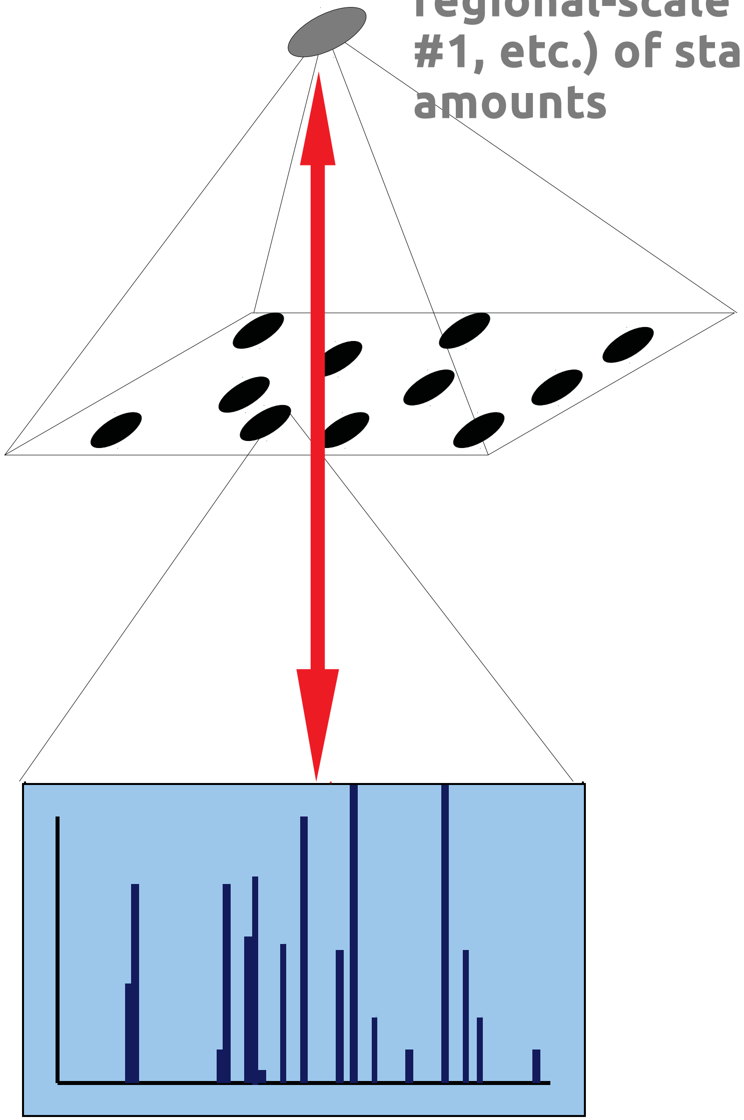
Observed anomalies



(http://iri.columbia.edu)

Current seasonal predictions of tropical rainfall assume, at least implictly, two interrelated underlying hypotheses ;

regional-scale average (spatial mean, EOF #1, etc.) of standardized local-scale seasonal amounts

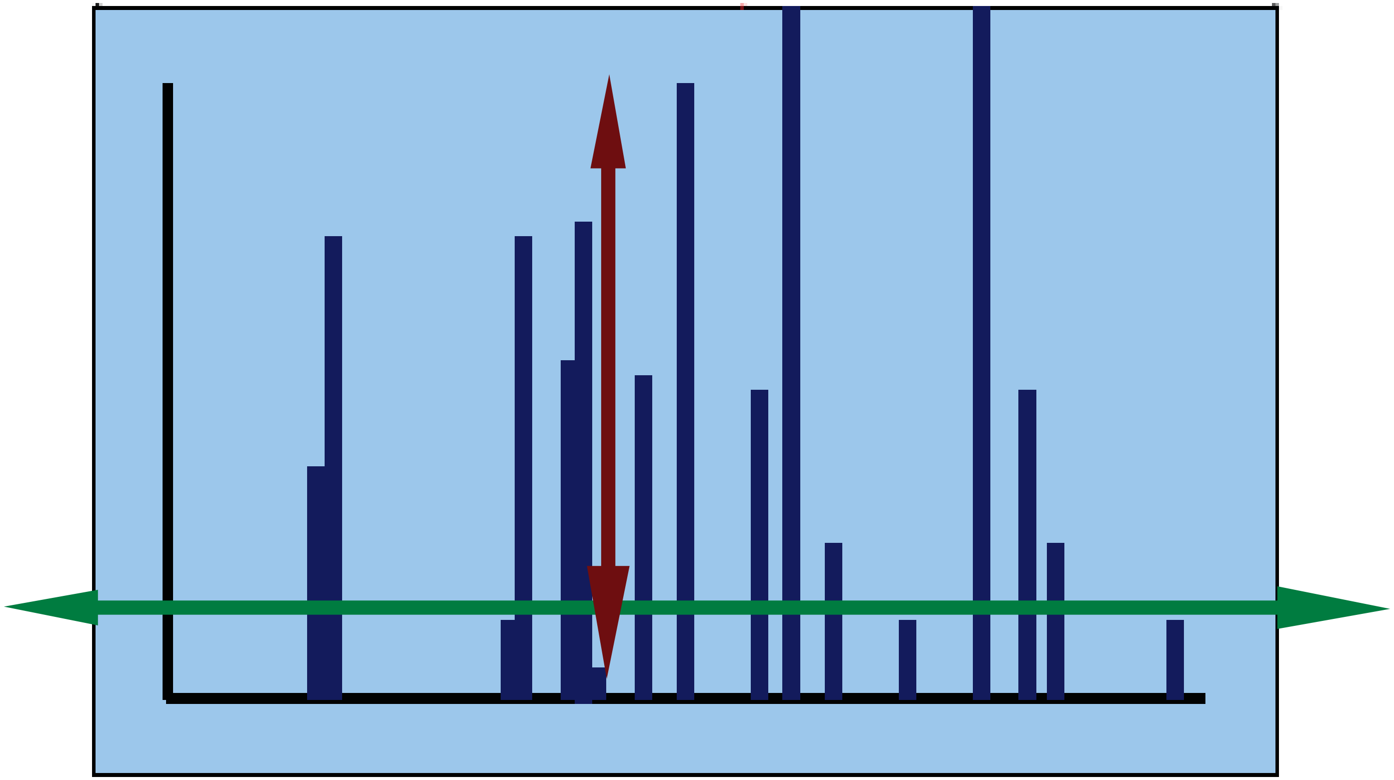


(1) : spatial average increases the signal-to-noise ratio -SNR- (« signal » is the forced component of the atmosphere and « noise » is all unpredictable sources at seasonal time scale including unforced atmospheric dynamics)

(2) : the external (SST or other sources) forcing is slow and imprints a quasi-constant anomaly

The fact is that a wet season is largely dominated by few wet spells : how the forcing is translated toward the intraseasonal characteristics (ISC) ?

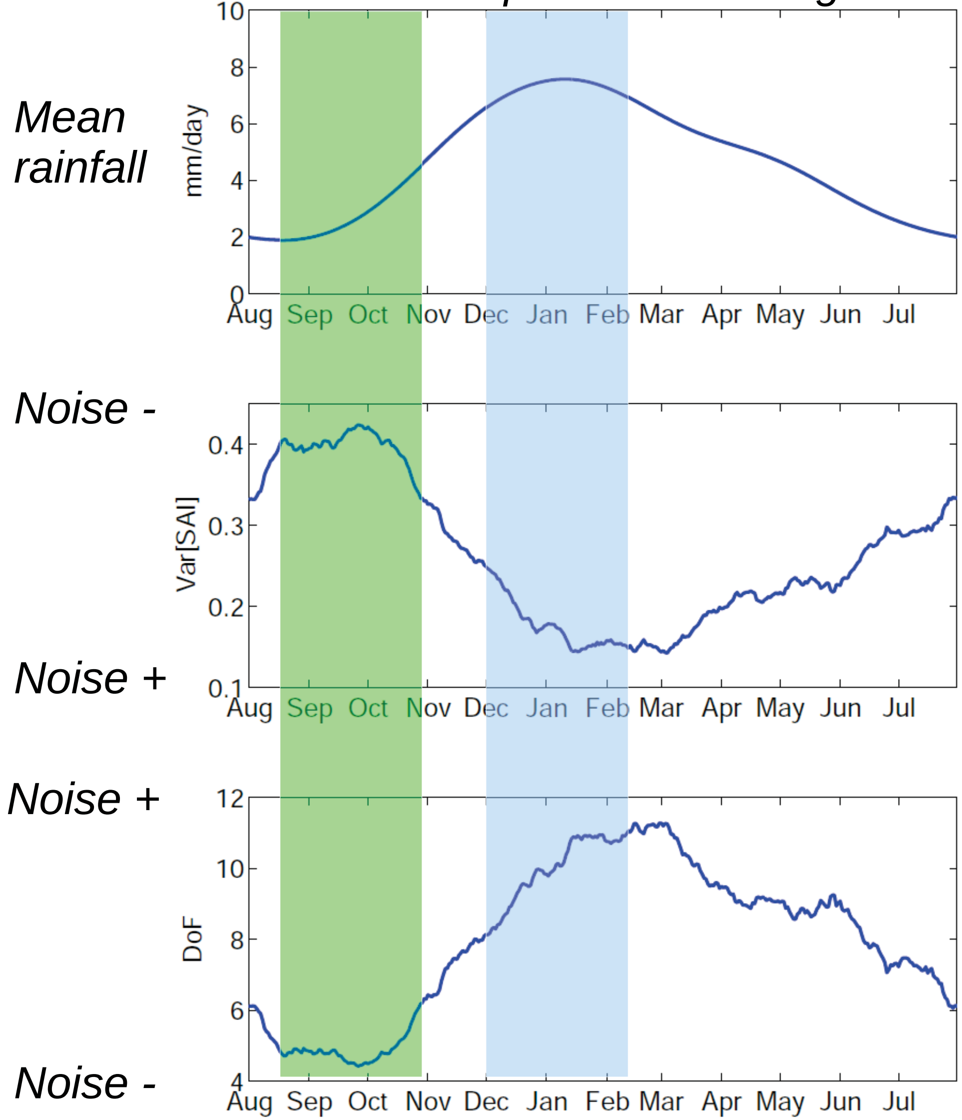
How the forcing could be translated toward local-scale ISC ?



- phase of the wet season : anomalous shift of the season (delayed/early onset/withdrawal and/or seasonal peak)
- internal scale and occurrence : intensity/number/length of wet events

If the strongest impact of boundary forcings, including SST, is on phase, seasonal amounts would not be necessarily the optimal variables for the reduction of the SNR

Statistics are computed on running 31-days



Var[SAI] is the variance of the spatial average of local-scale anomalies. DoF are the spatial degrees of freedom (Fraedrich et al., 1995 ; Moron et al., 2007)

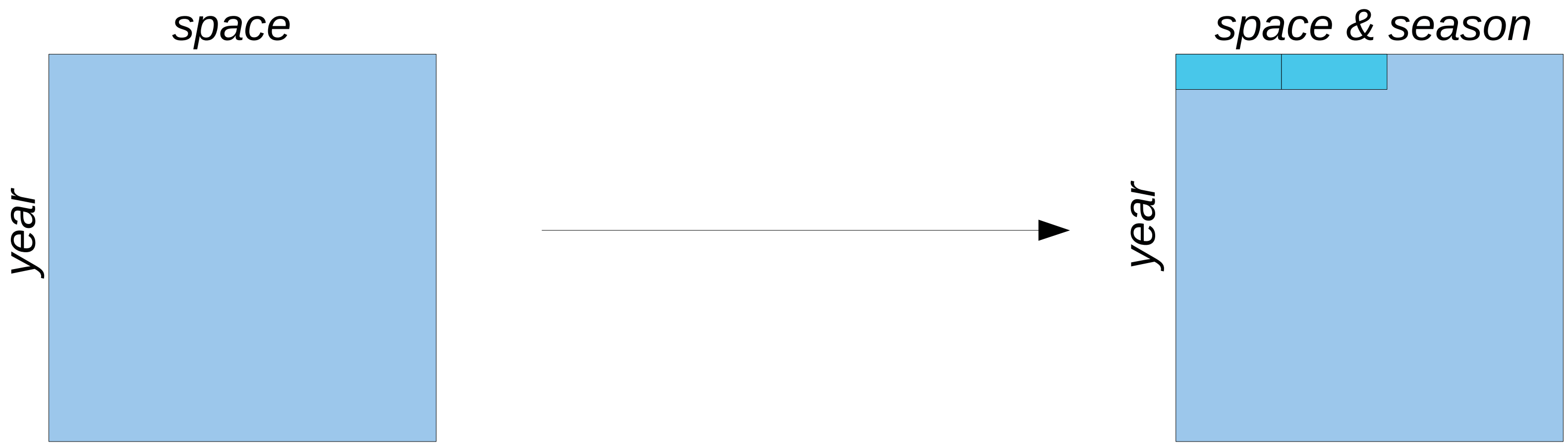
Example of spatial coherence of interannual anomalies of rainfall over monsoonal Indonesia (south of equator. cf. first lecture of last friday)

The climatological peak around January is relatively noisy and any seasonal amounts considering these amounts will have a low SNR

The SNR is better before the core of the rainy season (that is the transition between dry and wet season). This signal could be blurred by large amounts if the whole season is considered

Proposal : why not considering explictly these subseasonal modulations and the regional-scale signals that maximize the spatial covariance across the stations INSIDE a season = extraction of the typical *subseasonal* scenarios ?

classical approach / *this study*



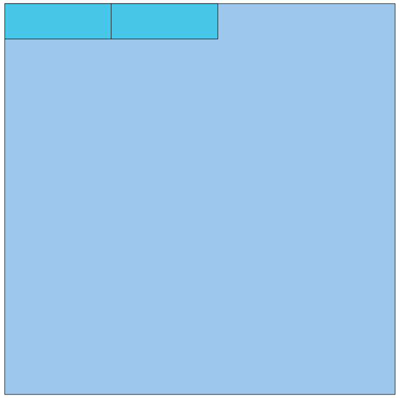
standardized seasonal amounts

Consideration of the daily sequences (low-pass filtered and standardized at local-scale)

Corresponding covariance matrix could be diagonalized to extract the largest scale of interannual variations maximizing the covariance across stations

Extraction of the subseasonal modes of variation that maximize the covariance across stations and across season

« Subseasonal » refers to the fact that intraseasonal variations which is not phase-locked between years is filtered out by the analysis



Daily time series is low-pass filtered (using a recursive filter with a cut-off at 1/31 cycle-per-day) and standardized at local-scale. The stations-seasons are concatenated so that row describe years and column space & intraseasonal variations

This time-lagged matrix is diagonalized with EOF

$$\mathbf{Y} = U \times S \times V'$$

To extract the first modes of variation

$$Z = U \times S$$

*A fuzzy k-means is used to cluster the most typical subseasonal scenarios in the time-lagged unstandardized Principal Components, **Z** according to the cluster's centroid C. The cost function J is*

$$J(c,m) = \sum_{k=1}^{c} \sum_{i=1}^{N} (\mathbf{u}_{ki})^m (Z_i - C_k)^2$$

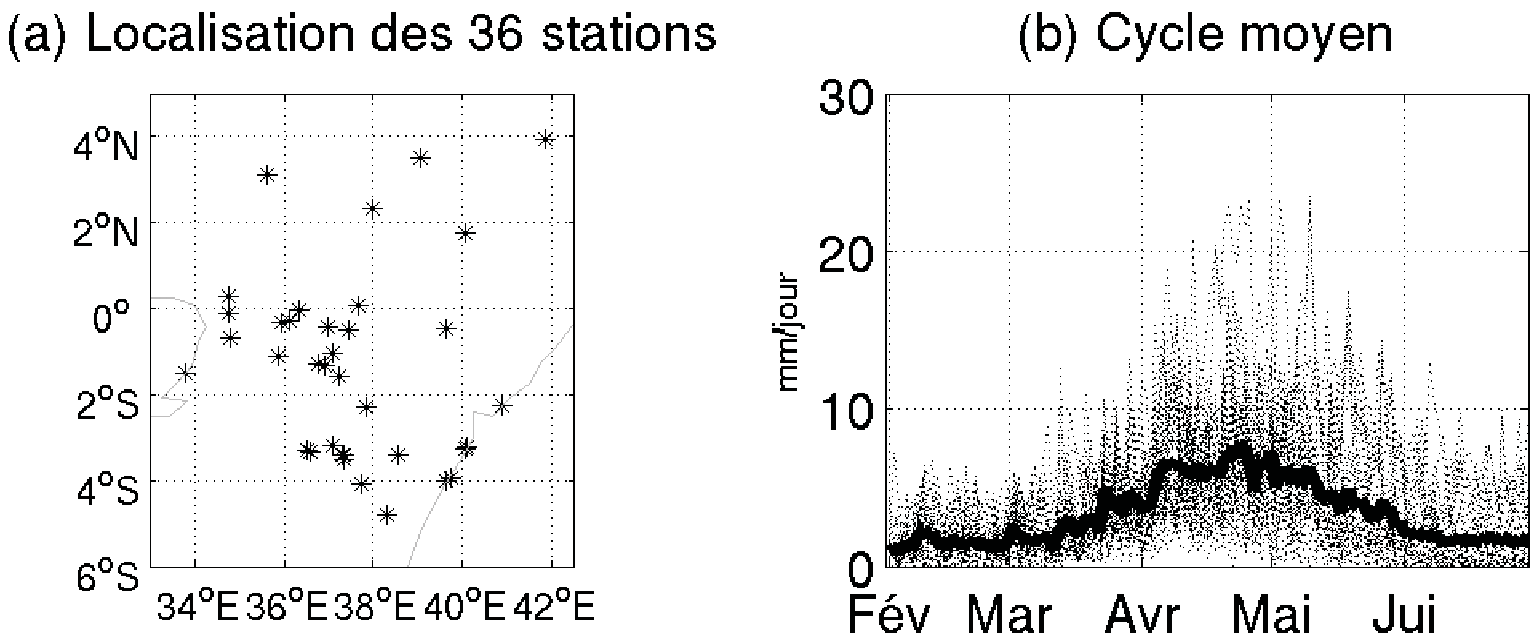
$$\sum_{k=1}^{c} \mathbf{u}_{ki} = 1, \quad \text{where} \quad 0 \le \mathbf{u}_{ki} \le 1.$$

u *are memberships (= 0 or 1 in « hard » k-means) and m is the fuzziness exponent (= 1 is equivalent to « hard » k-means)*

Two examples of extraction in areas where the skill of seasonal amounts is notably low and/or unstable

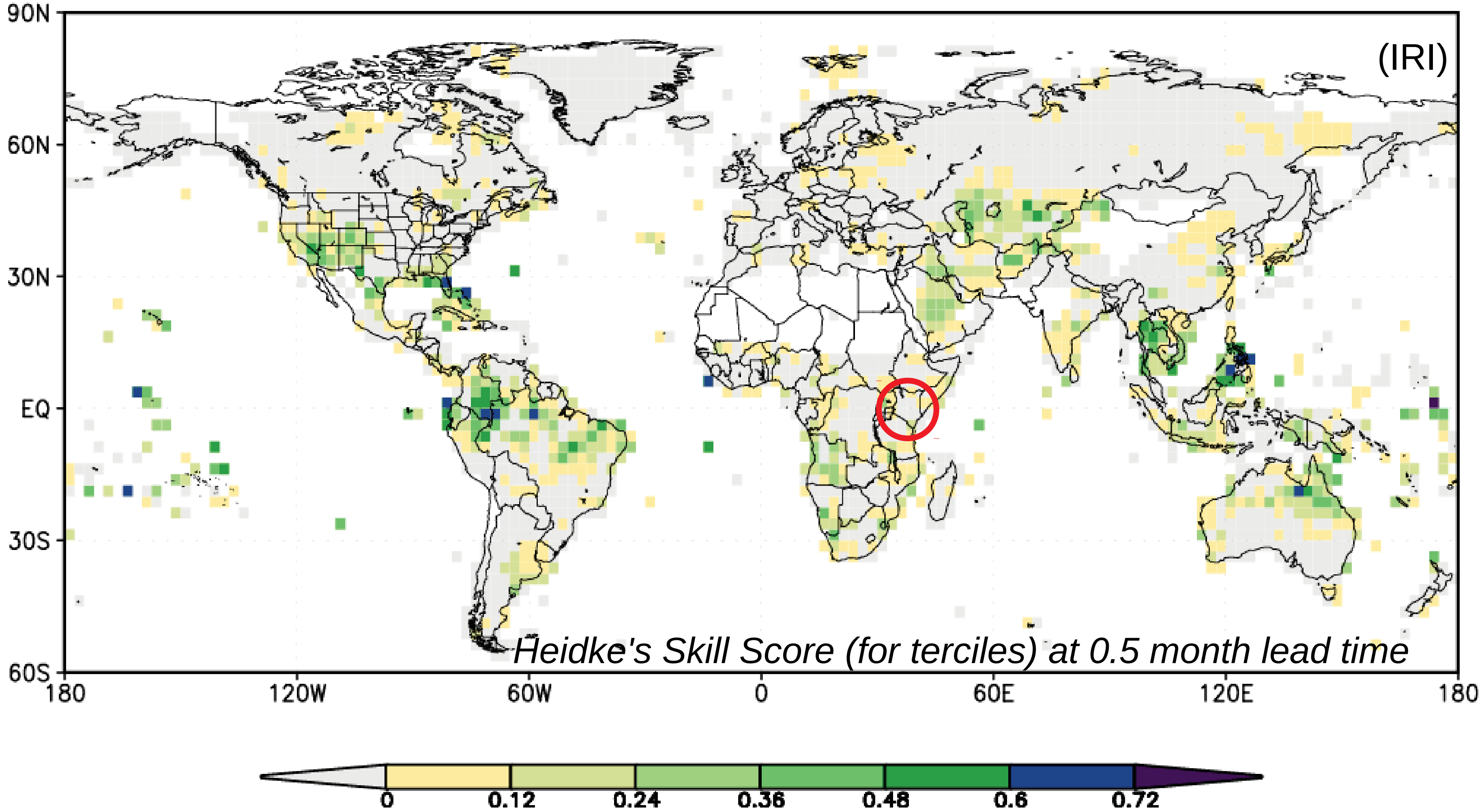
1. « Long Rains » over Equatorial East Africa : The long rain occuring in March-May when ITCZ moves north are difficult to predict from tropical SST (i.e. Camberlin and Philippon, 2002)

2. Monsoonal India in June-September : weak-to-moderate and/or unstable relationship with ENSO (less rainfall in warm ENSO events) and low skill of seasonal predictions (especially during recent decades) (i.e. Kar et al., 2011 ; Sinha et al., 2013)

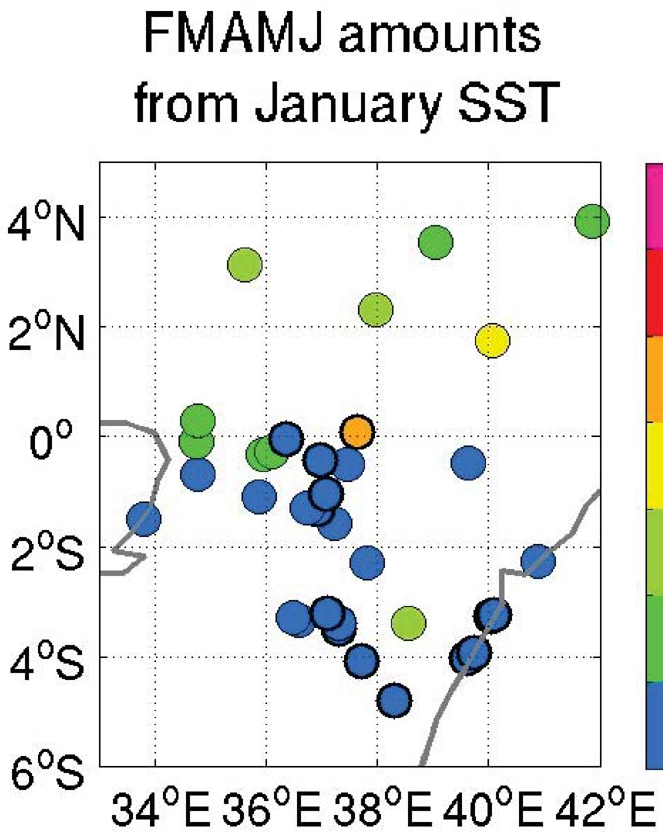


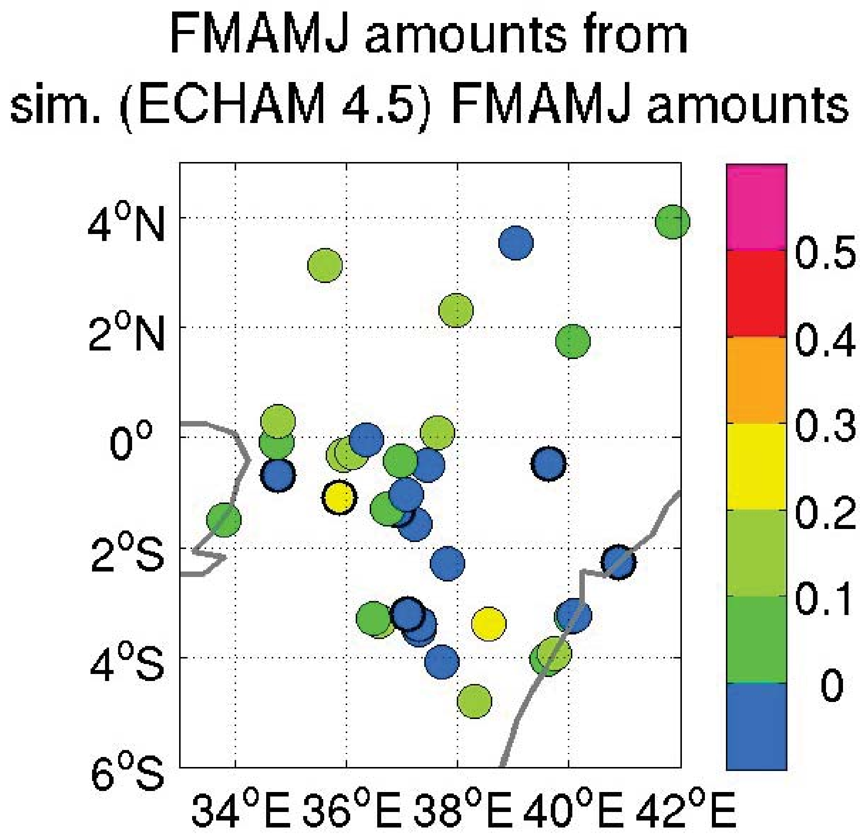
- 36 stations x 41 years (1961-2001) < 5 % missing entries
- large spatial variations with driest sector in NE and valleys and wettest sector along Indian coast, mountains and in Western Kenya
- moderate seasonal cycle with a peak in April, but low rainfall already present in February and also June

Retrospective skill of March-May seasonal amounts from February SST



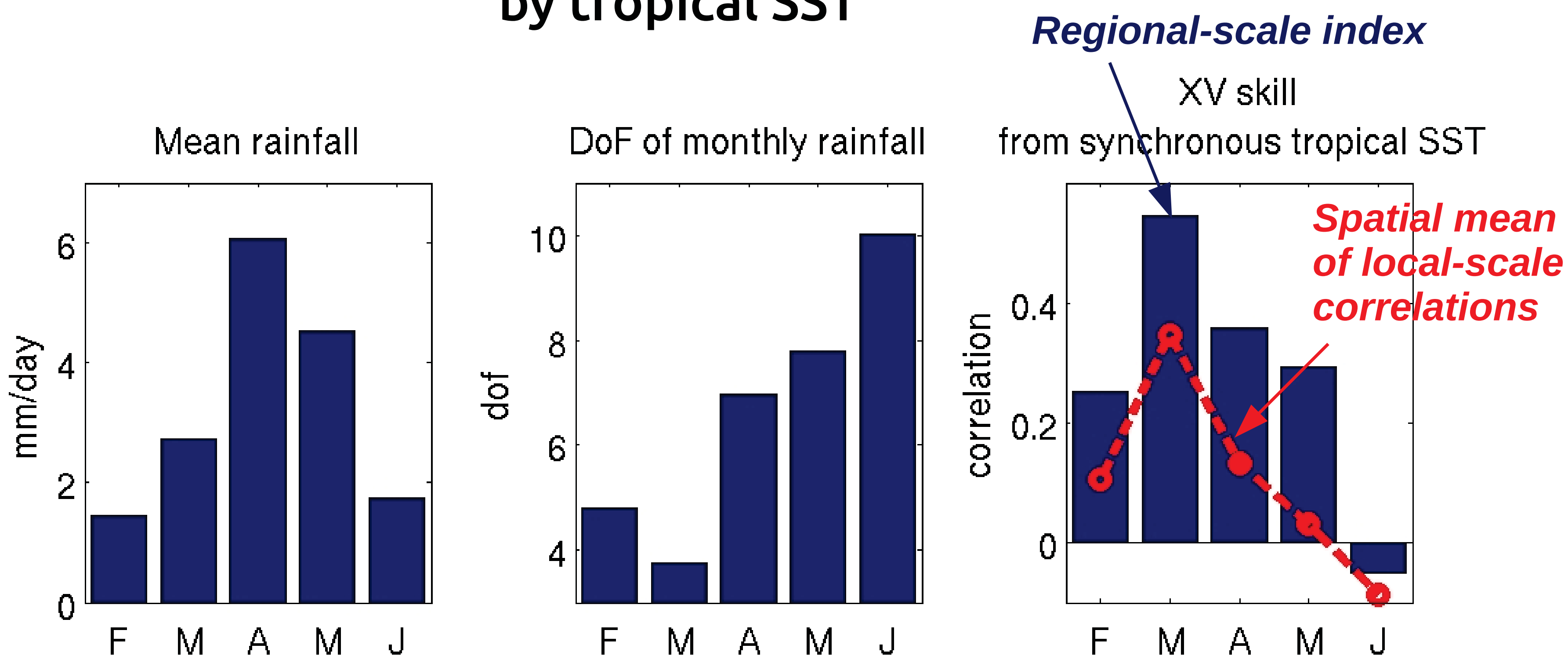
Very poor skill of Feb-June seasonal amounts from either statistical and statistical-dynamical predictions from tropical SST





Correlation (obs vs simulation) of FMAMJ amount from a cross-validated CCA using observed tropical SST in January (on left) and simulated rainfall from ECHAM 4.5 ca_sst – february initialization – (on right)

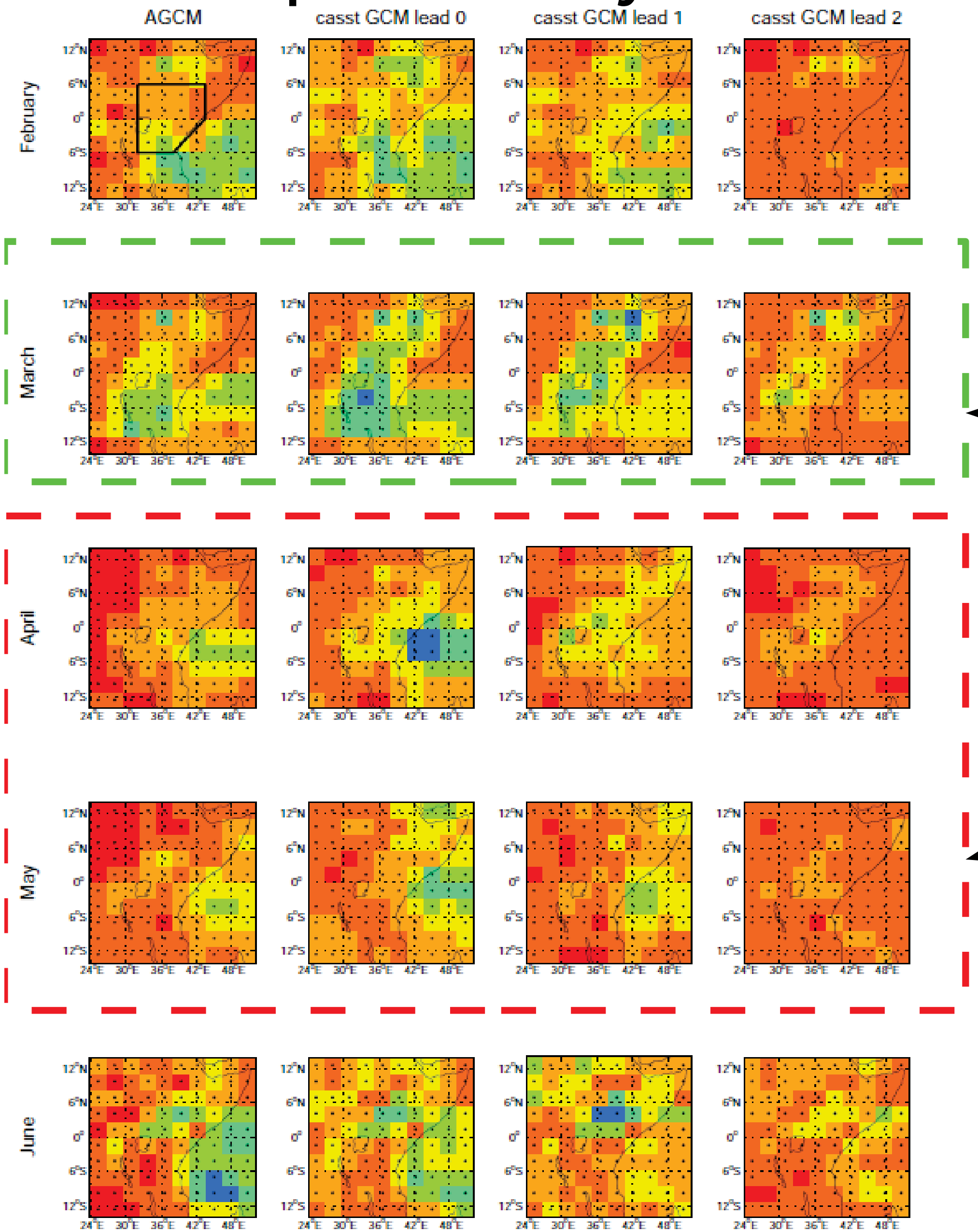
Clear intraseasonal modulation of spatial coherence amongst the stations (max in March) that fits with intensity of variance explained by tropical SST



DoF = spatial Degree of Freedom = 1 for perfect covariant variations and ~ 19 in our case for independent variation (Fraedrich et al., 1995 ; Moron et al., 2007)

(Moron *et al.*, 2013)

Potential predictability from GCM ensembles at monthly time scale …

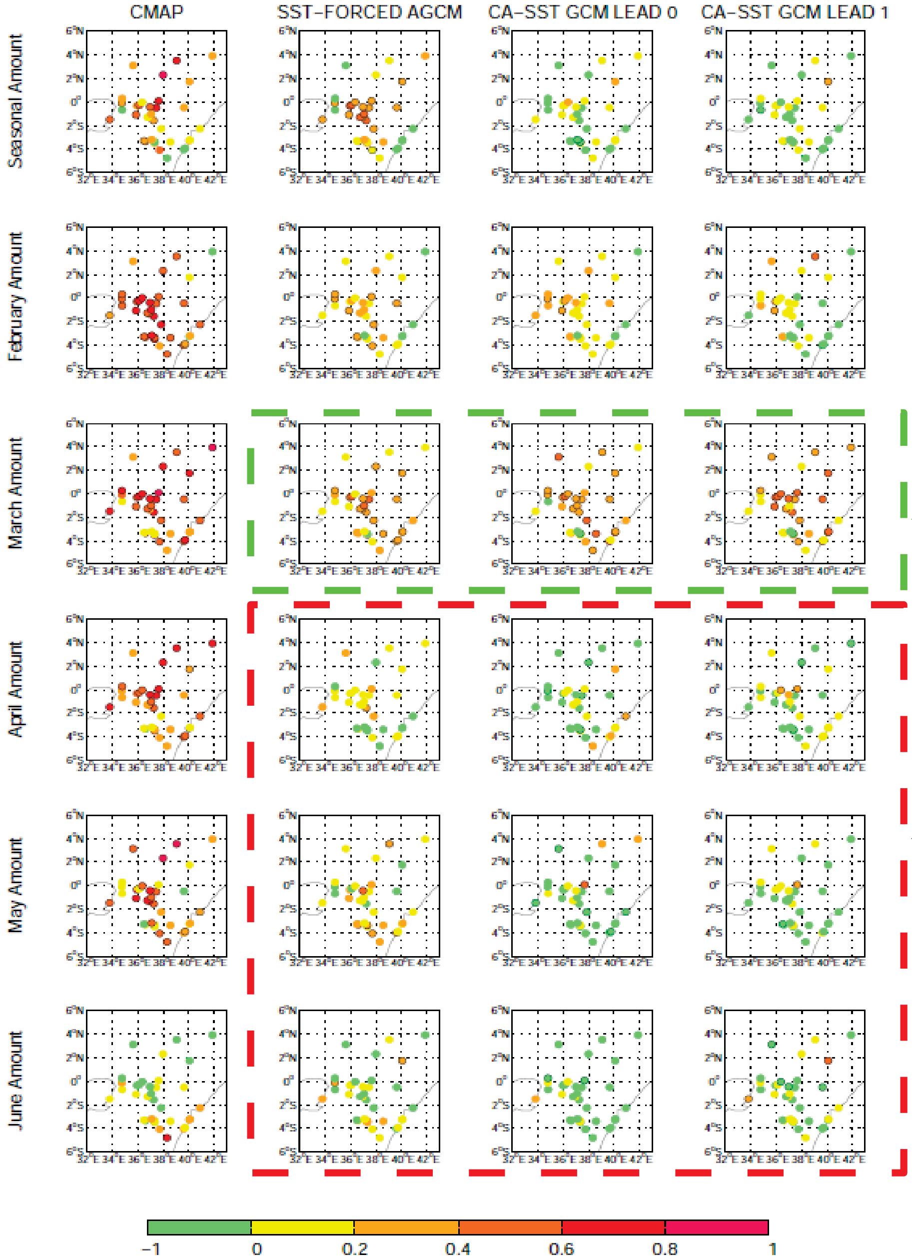


External variance (= common variance amongst an ensemble of 85 runs of SST-forced ECHAM 4.5 and 24 runs of ca_sst ECHAM 4.5 for lead 0, 1 and 2) in monthly rainfall

Maximum PP over Equatorial East Africa in March

Minimum PP over Equatorial East Africa in April-May (= including the seasonal peak of rainfall)

Skill from GCM ensembles at monthly time scale …



MOS skill (given by cross-validated CCA) of the ensemble-mean of 85 runs of SST-forced ECHAM 4.5 and 24 runs of ca_sst ECHAM 4.5) for seasonal and monthly rainfall (CMAP results on first column are displayed as a reference)

Maximum skill over Kenya/N. Tanzania in March. In case of ca_sst, the skill in March is over the one obtained for seasonal amounts themselves

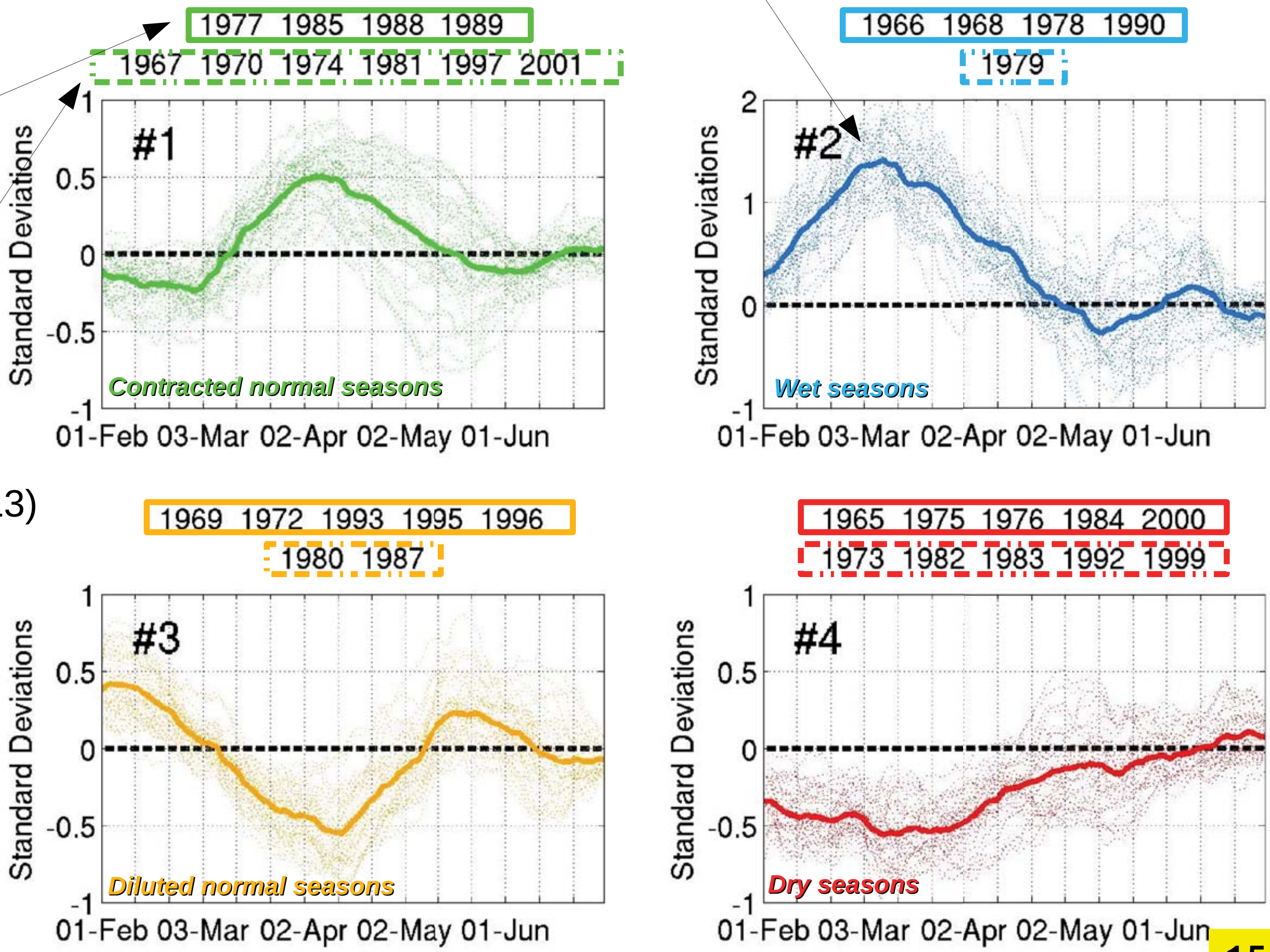
Poor skill in April-June including the seasonal peak of rainfall

4 subseasonal scenarios extracted by time-lagged EOF + fuzzy k-means

Local-scale rainfall filtered anomalies (dotted line) and spatial average (bold line)

Years where membership > 75 %

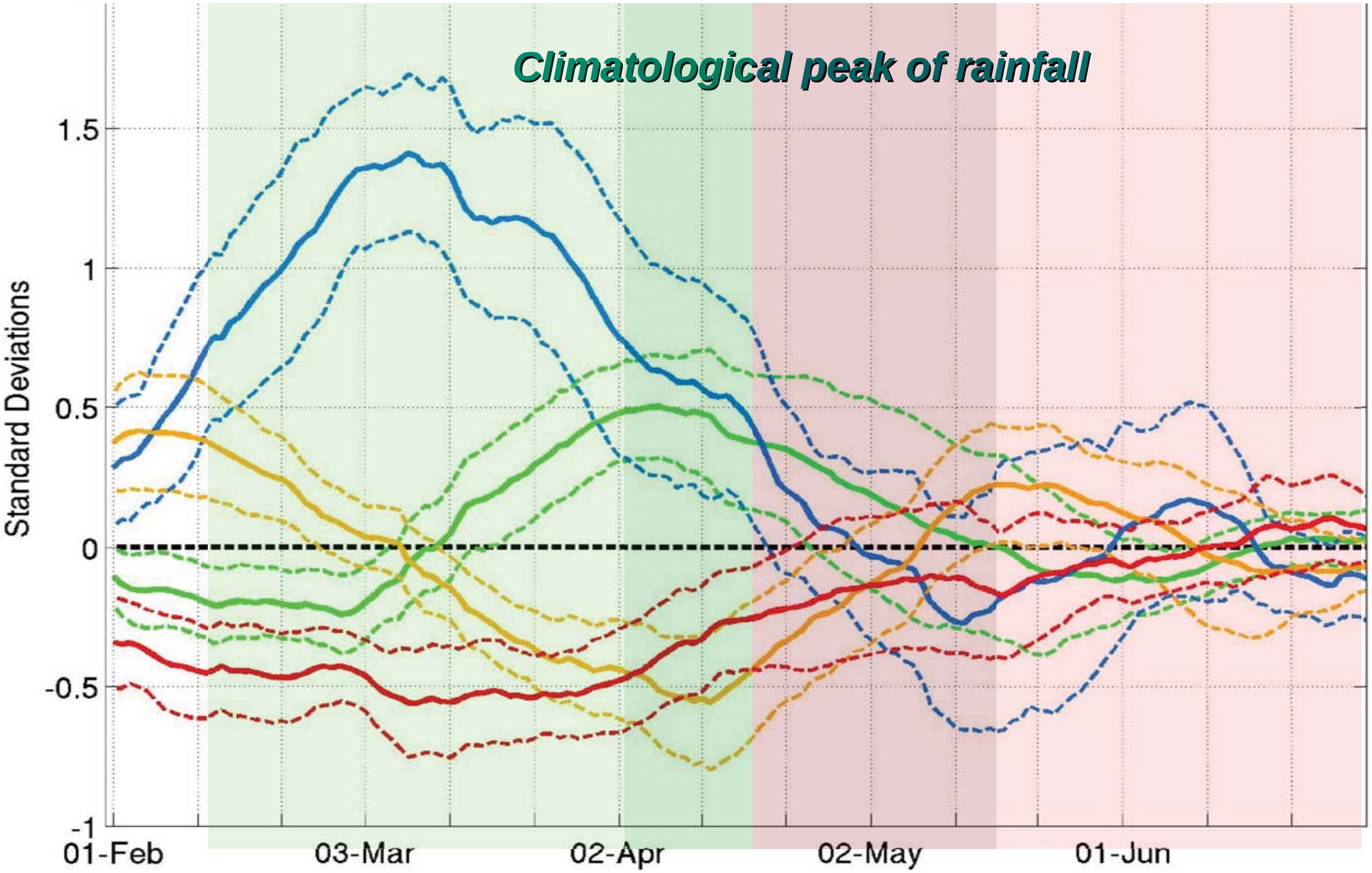
and where it is between 50 % and 75 %



(Moron *et al.*, 2013)

The subseasonal scenarios mostly emphasize the interannual variability around the onset of the rainy season …

Spatial average of the filtered rainfall anomalies for the 4 scenarios with +/- one standard deviation (dashed line)

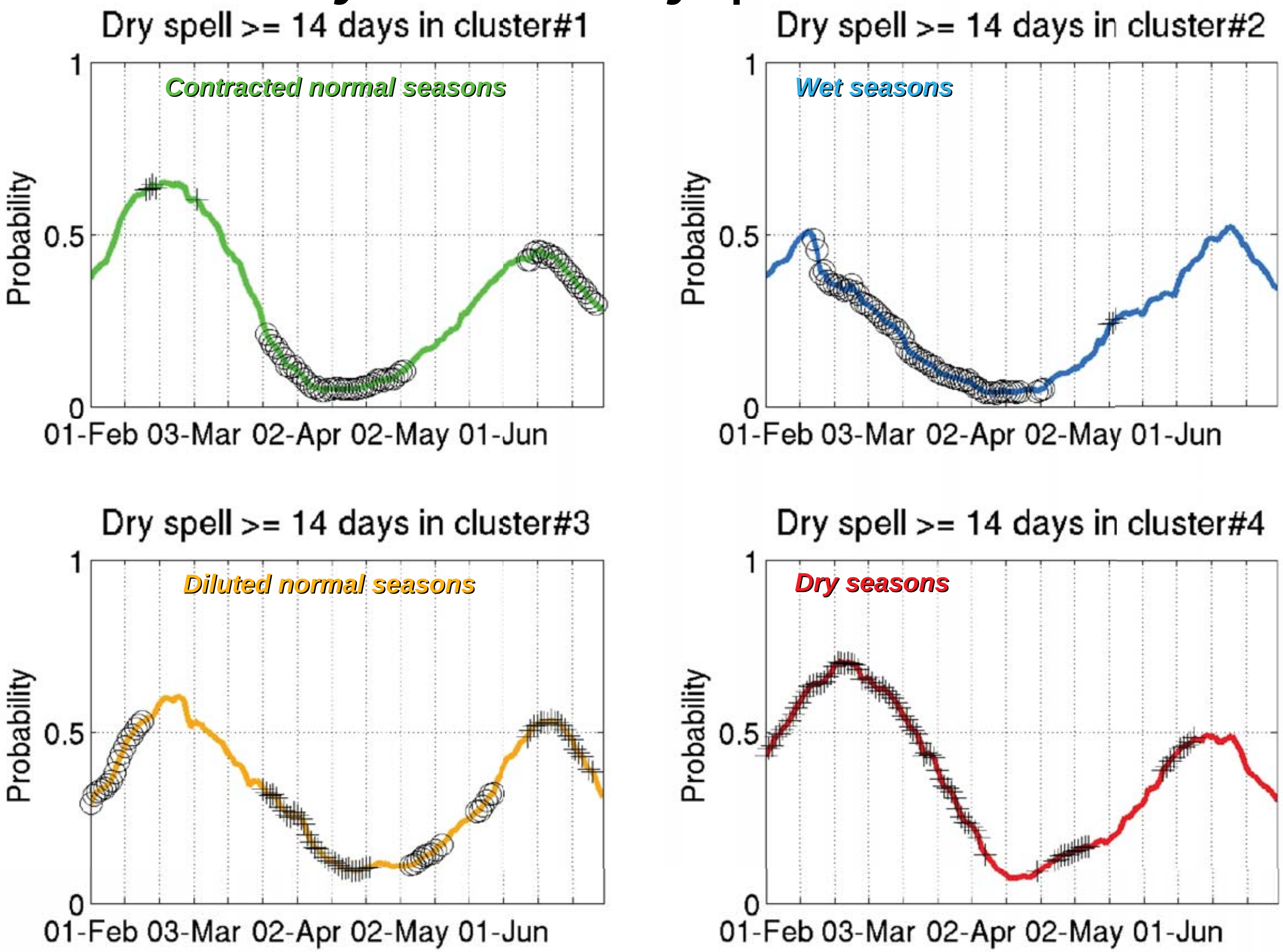


Mean regional-scale signal <> 0 that may be PP

Mean regional-scale signal ~ 0 that may not be PP

the subseasonal scenarios could be easily toward internal characteristics of the rainy season as dry spells ...

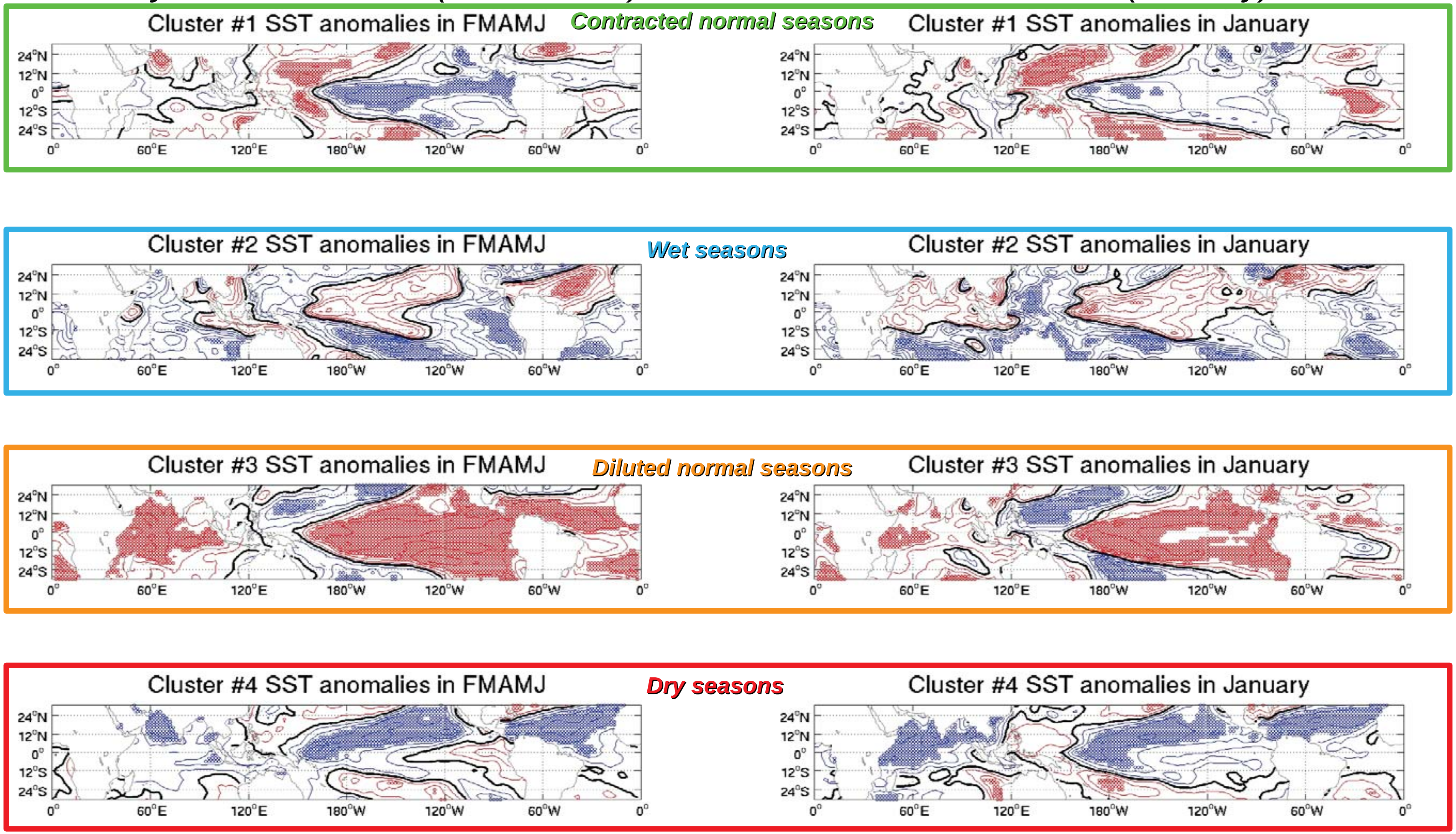
Significant positive (+) or negative (o) probability to observe a dry spell lasting at least 14 days at local-scale



(Moron *et al.*, 2013)

The subseasonal scenarios could also reveal interesting SST forcings ...

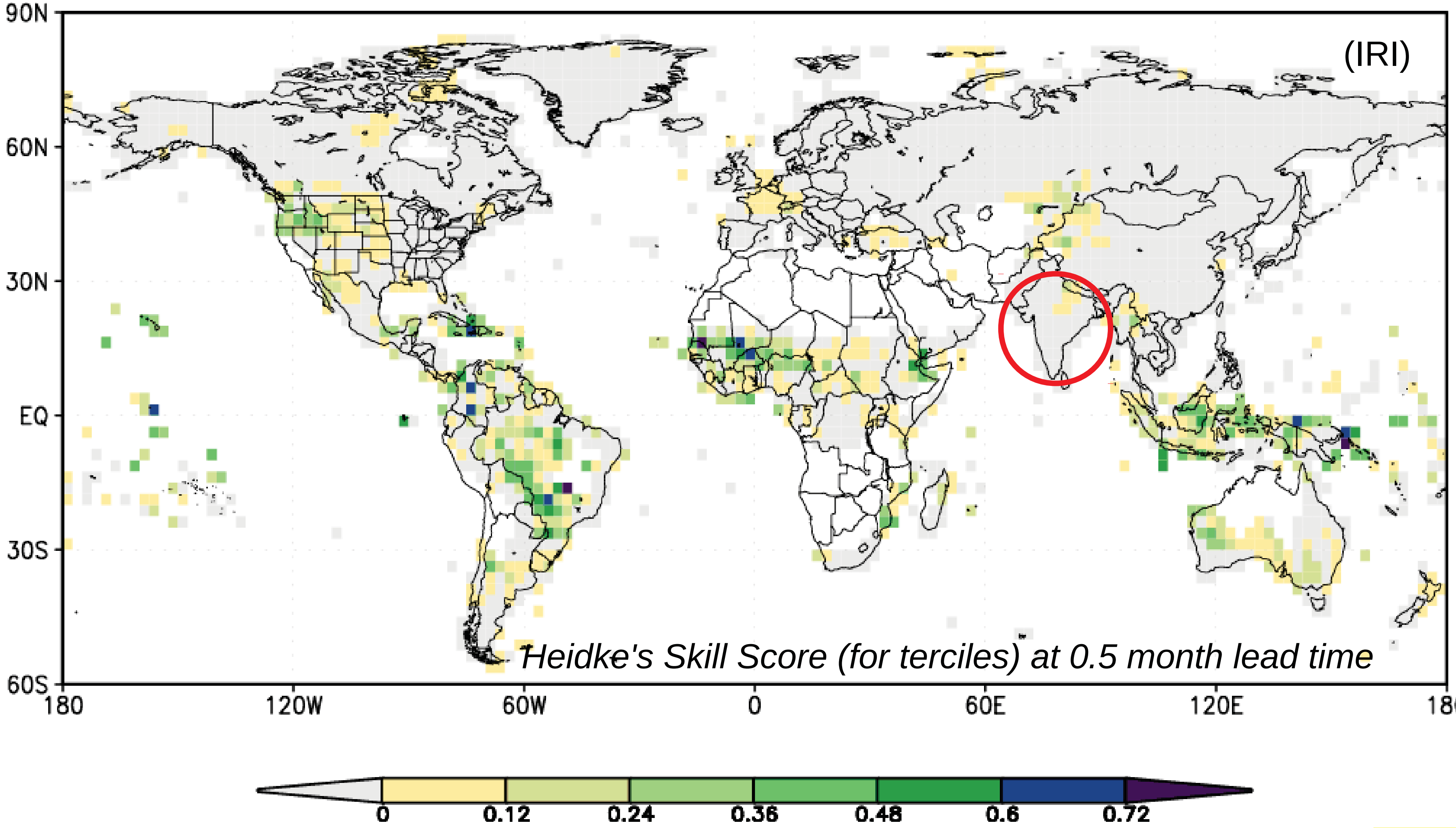
Synchronous SSTA (Feb to June) — *Prior SSTA (January)*



Blue = negative SSTA ('o' : significant at the two-sided 90 % level) & red contours = positive SSTA ('o' : significant at the two-sided 90 % level)

(Moron *et al.*, 2013)

Retrospective skill of June-August seasonal amounts from May SST



Predicted seasonal rainfall anomalies (JJAS) averaged over Monsoonal India from NCMWRF initialized in mid-May (18 runs and SST forecasts from CFS)

Figure 3. Spatial average of JJAS standardized anomaly rainfall over India (SAI) for observed gridded rainfall data (solid line) and model rainfall (interpolated over observed grid points) data (dash line).

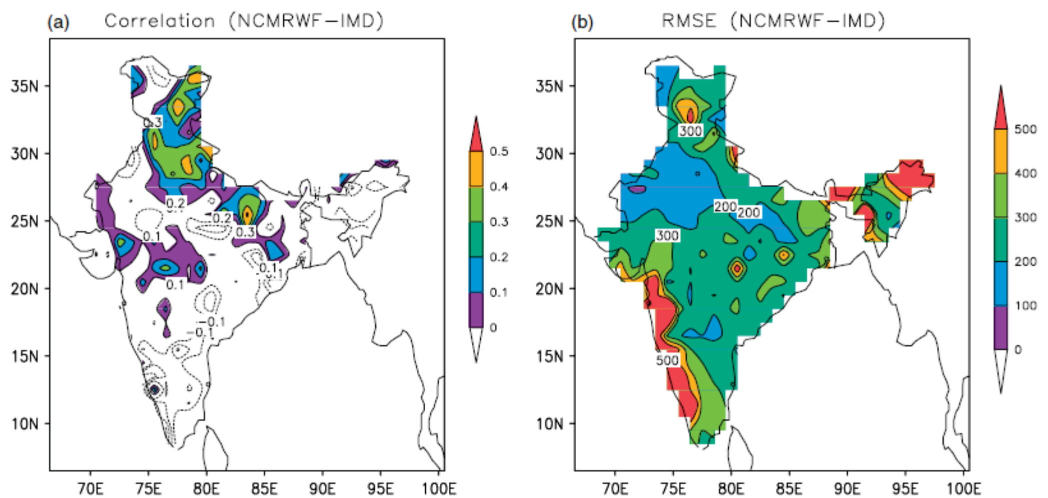
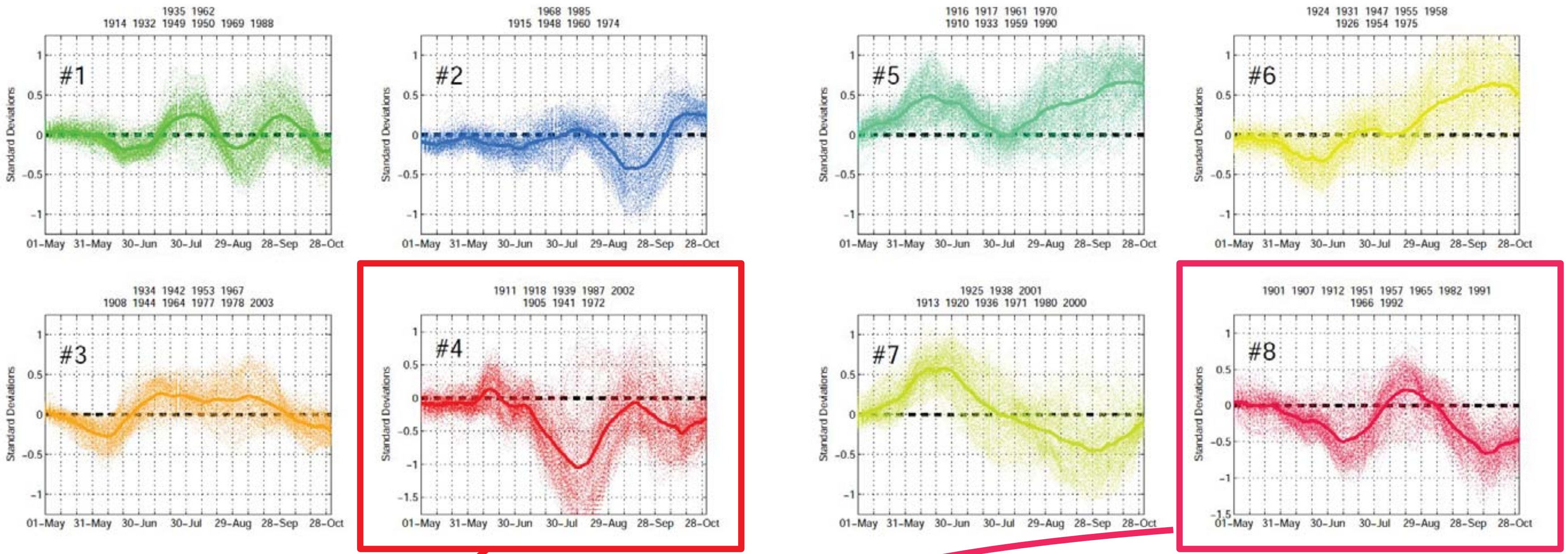
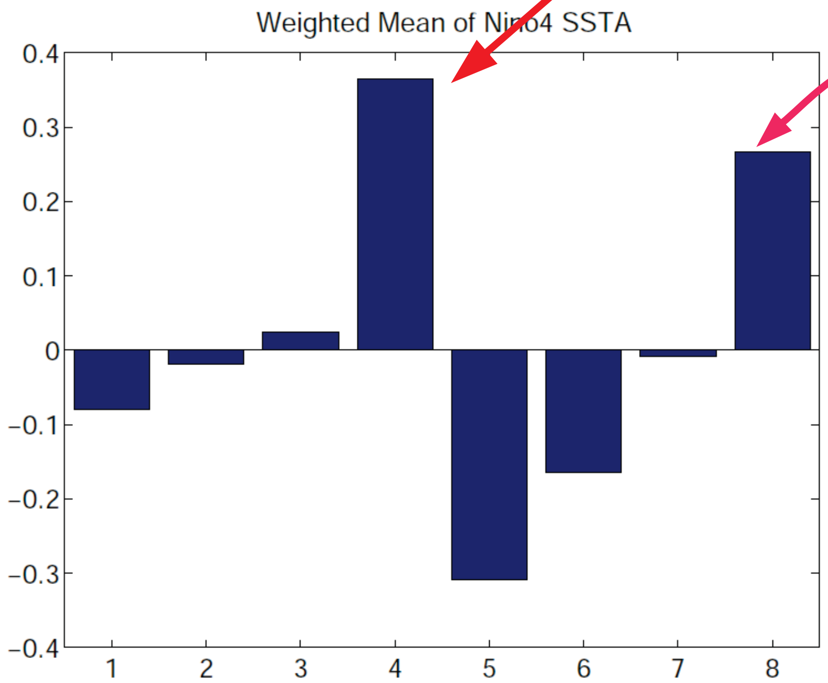


Figure 4. Statistical analysis between observed and model JJAS rainfall totals (model rainfall is interpolated in observed grid) for the period 1981−2008 shown (a) for correlation coefficient (positive in solid contour and shaded and negative are in dash contour) and (b) for root mean square error (mm).

(Sinha *et al.*, 2013)

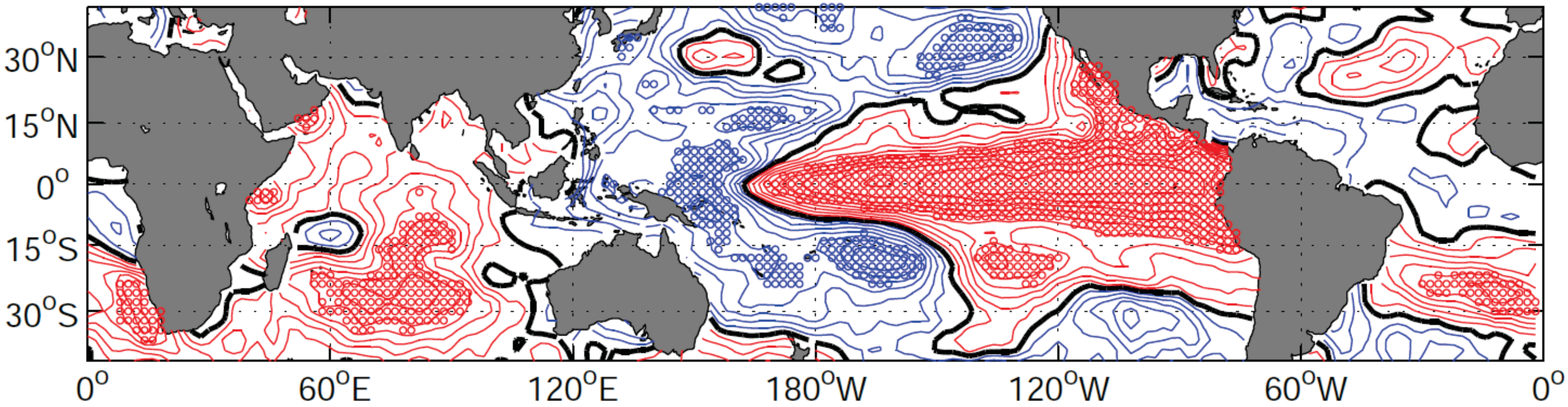




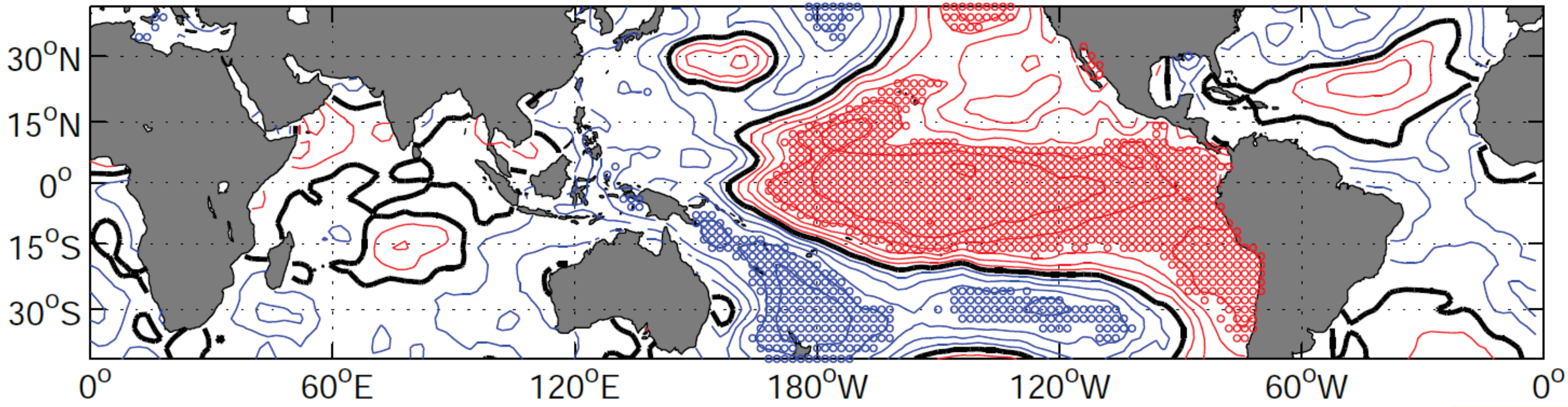
Rainfall anomalies for the 8 scenarios extracted from MJJASO rainfall across monsoonal India (1901-2004, 1° gridded rainfall from IMD)

2 out of 8 subseasonal scenarios (#4 and #8) are related with warm ENSO events. This scenarios leads to the driest seasons (#4 is drier than #8)

Cluster #4 SST anomalies in MJJASO



Cluster #8 SST anomalies in MJJASO

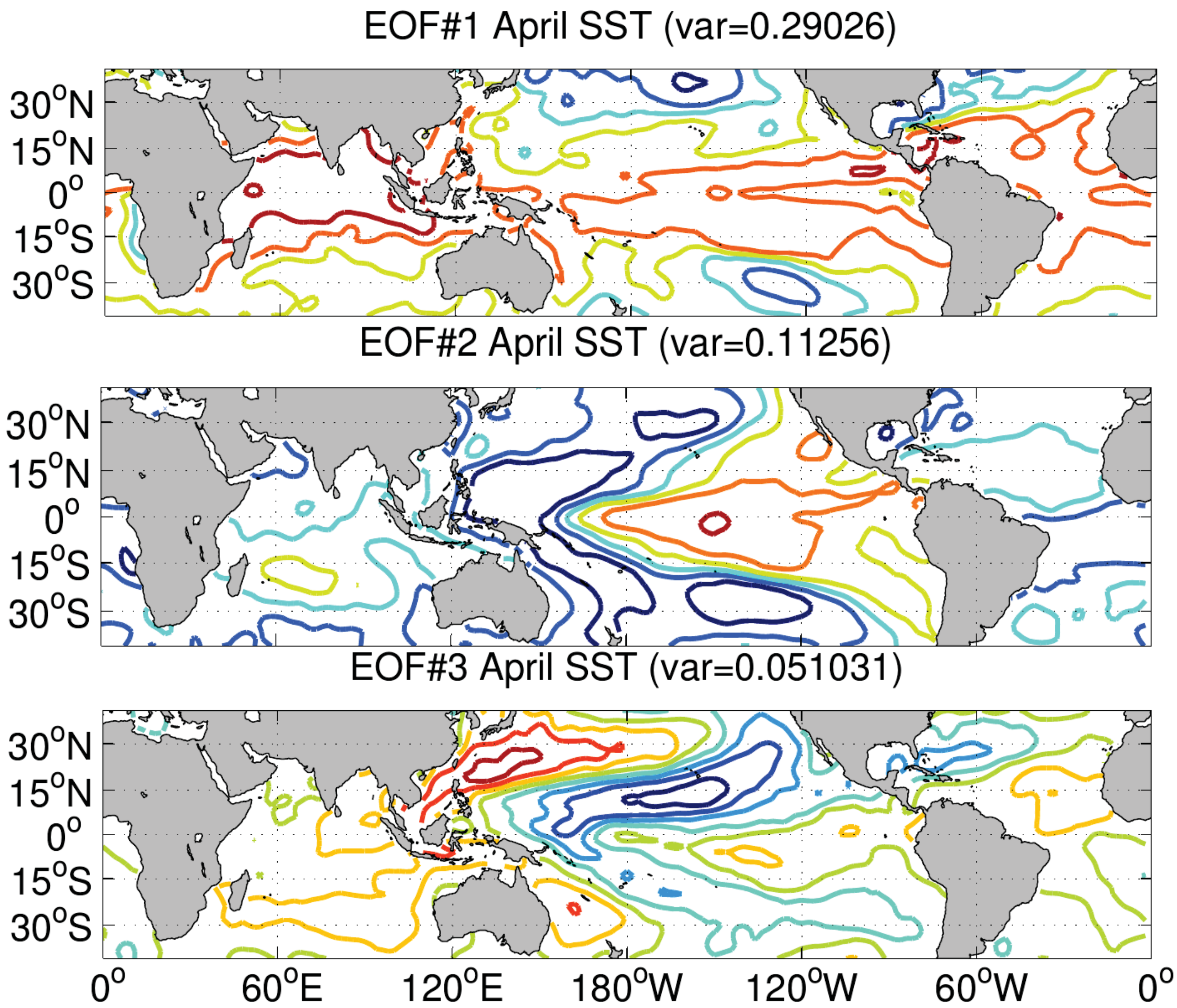


2-tiered predictions of the subseasonal scenario using SST as predictors

First step : multinomial logistic regression is used to predict the membership from April SST (pre-filtered using EOF and considering only the three leading modes)

Second step : stochastic sampling of subseasonal scenarios using a k *nearest analog* approach between the predicted membership (= target) and historical memberships (= library) ; k=500

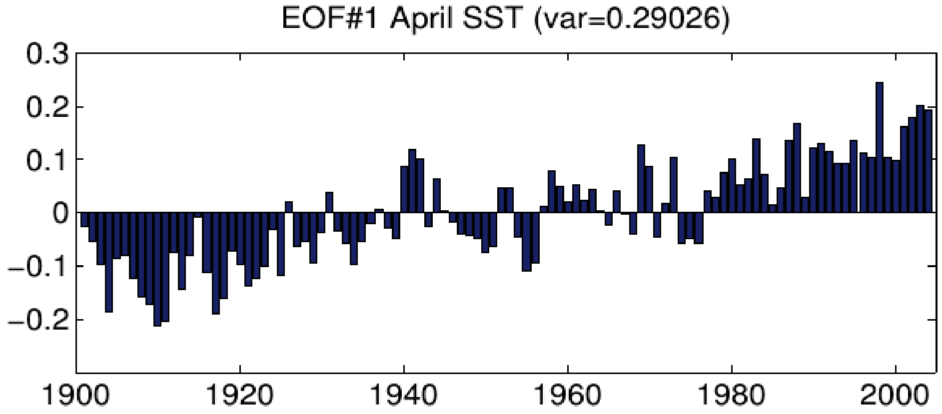
The two steps are fully cross-validated with 10 years withheld at each turn (all steps, i.e. climatology, EOF of SST, time-lagged EOF and fuzzy k-means of rainfall to cluster subseasonal scenarios and k*nn* sampling are cross-validated)



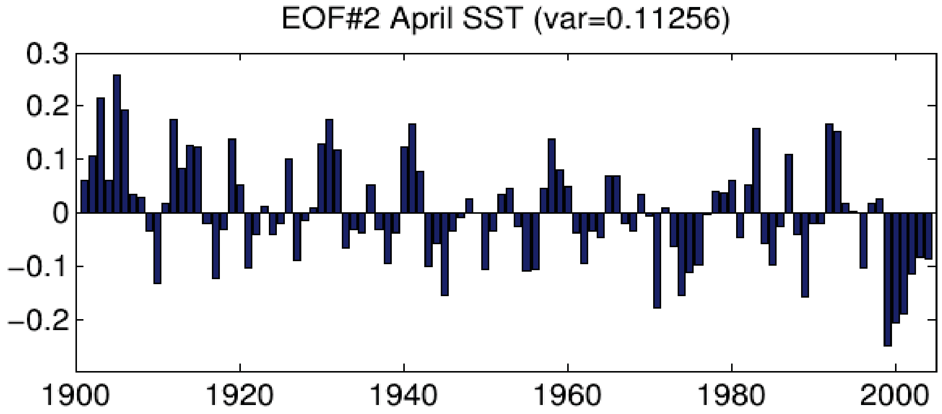
Positive trend

ENSO

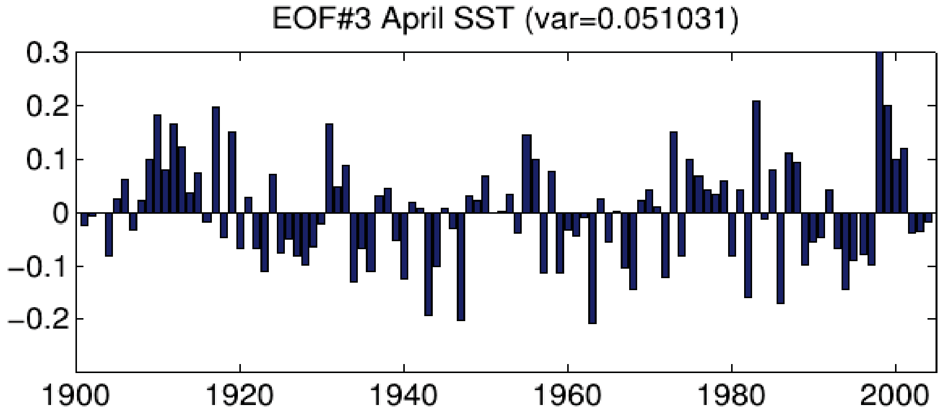
Subtropical Pacific



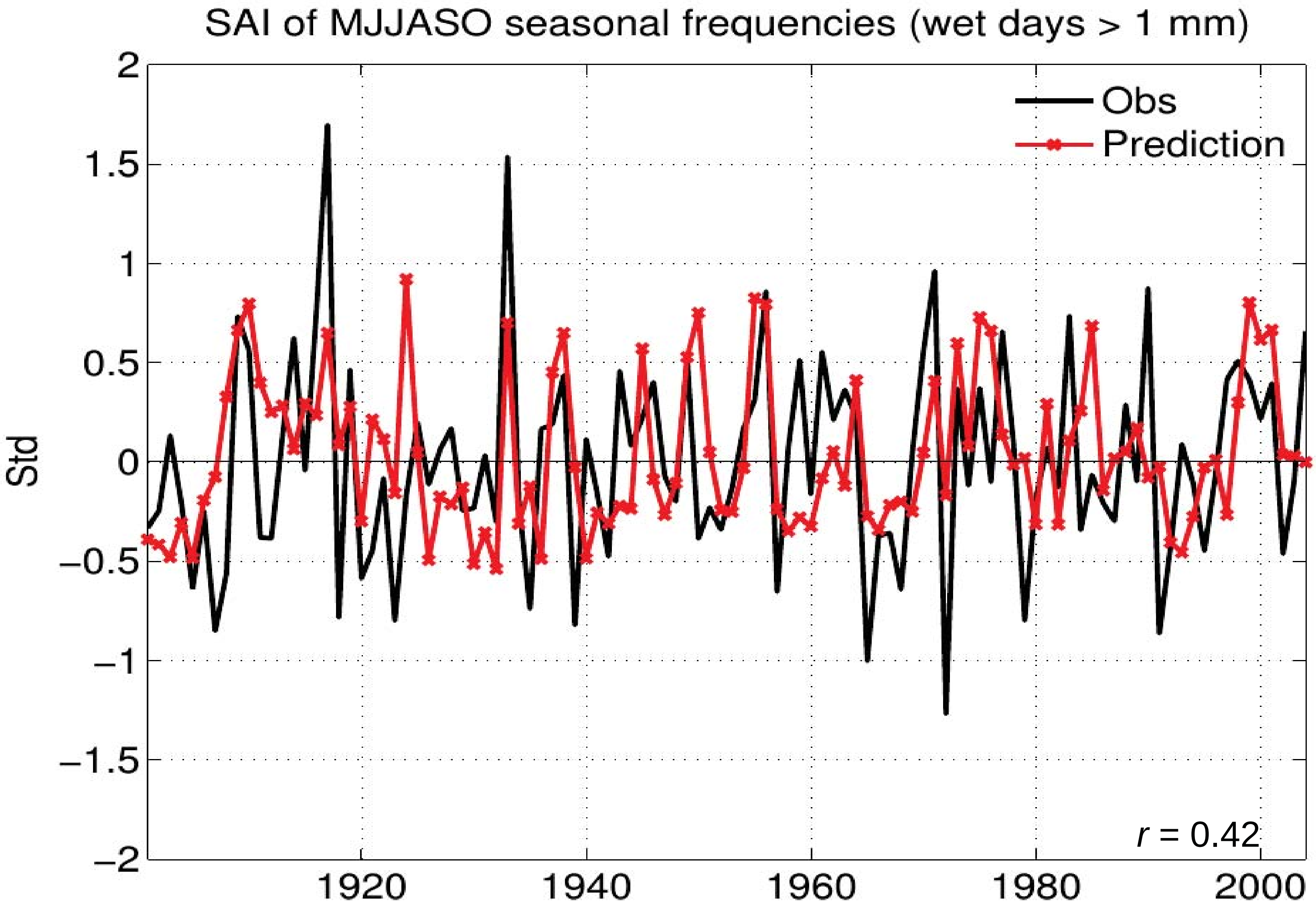
Positive trend



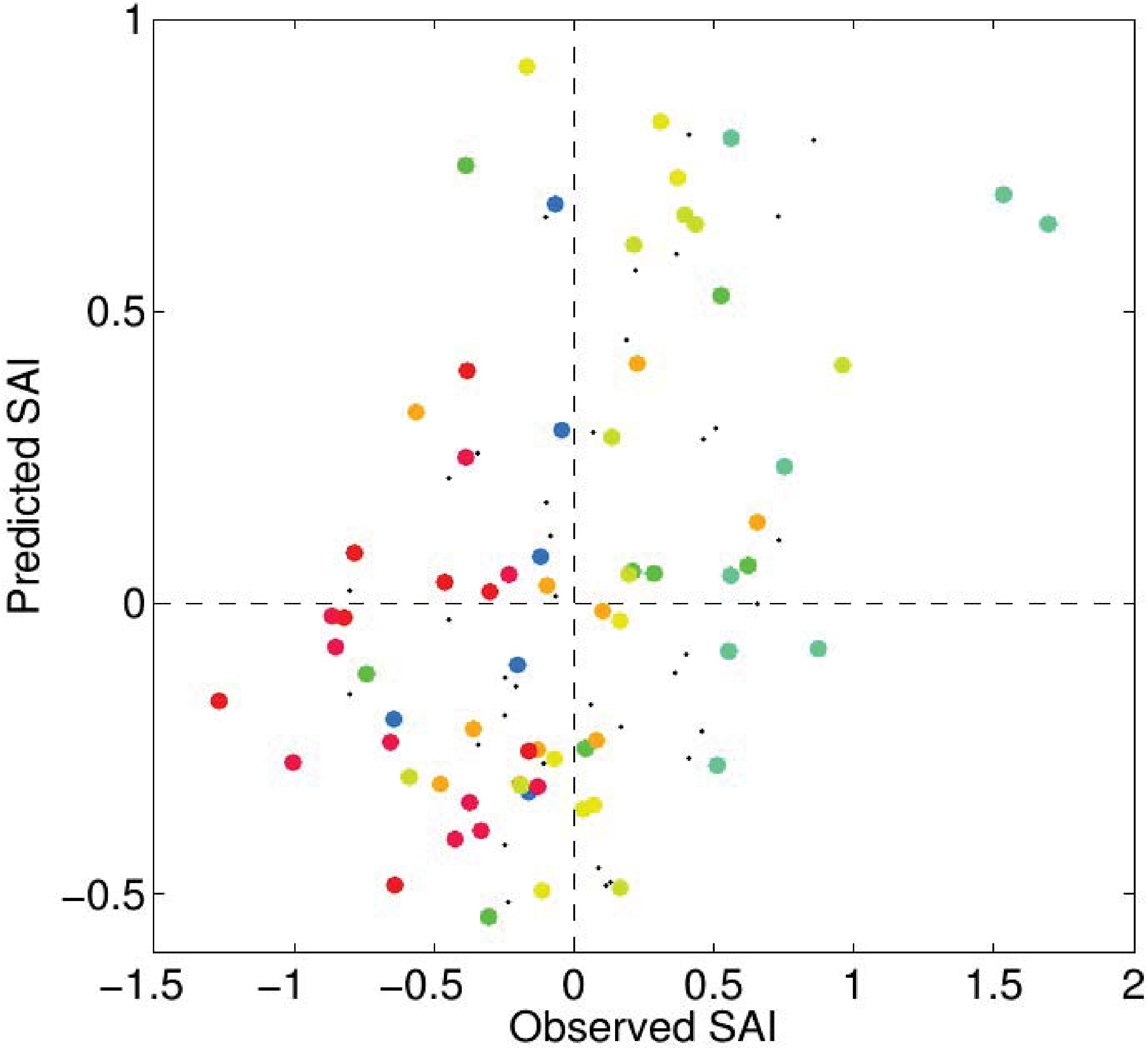
ENSO



Subtropical Pacific



SAI : Standardized Anomaly Index = spatial average of local-scale standardized anomalies (of frequency of wet days from May 1 to October 31)



Unclassified years are those when highest membership < 0.5

		Hits
·	uncl	*20/37*
●	#1	*6/8*
●	#2	*3/6*
●	#3	*6/10*
●	#4	*4/8*
●	#5	*5/8*
●	#6	*4/8*
●	#7	*8/9*
●	#8	*8/10*

Rainfall anomalies related to CP ENSO are better predicted than those associated with EP ENSO

Concluding remarks

- seasonal amounts is not necessarily the optimal variable to be predicted (i.e. when the seasonal peak is noisy while for example the phase of the season is more or less controled by large-scale forcings)

- time-lagged EOF + fuzzy k-means help to extract the most covariant anomalies at interannual and subseasonal time scales

- the subseasonal scenarios could be used to diagnose the most typical time behavior of rainfall anomalies at regional-scale

Drawbacks & next steps

- parameters of the fuzzy k-means (fuzzy exponent + number of clusters) : « internal » vs « external » indices ?

- comparing the skill with classical approaches considering seasonal anomalies as predictands instead of subseasonal scenarios and understanding how subseasonal scenario potentially provide useful information for seasonal predictions

References

- Camberlin P., Philippon N. (2002) The east African March-May rainy season: associated atmospheric dynamics and predictability over the 1968-1997 period. *J. Climate* **15**, 1002-1019.
- Camberlin P., & al. (2009) Components of rainy seasons' variability in equatorial East Africa: onset, cessation, rainfall frequency and intensity. *Theor. Appl. Climatol.* **98**, 237-249.
- Fraedrich K., Ziehmann C., Sielmann F. (1995) Estimates of spatial degrees of freedom. *J. Climate* **8**, 361-369.
- Kar S.C., & al. (2011) Skill of monthly rainfall forecasts over India using multi-model ensemble schemes. *Int. J. Climatol.* **32**, 1271-1286.
- Katz R.W., Glantz M.H. (1986) Anatomy of a rainfall index. *Mon. Wea. Rev.* **114**, 764-771.
- McBratney A.B., Moore A.W. (1985) Applications of fuzzy sets to climatic classification. *Agric. For. Meteo.* **35**, 165-185.
- Moron V., & al. (2007) Spatial coherence of tropical rainfall at the regional-scale. *J. Climate* **20**, 5244-5263.
- Moron V., Camberlin P., Robertson A.W. (2013) Extracting subseasonal scenarios: an alternative method to analyze seasonal predictability of regional-scale tropical rainfall. *J. Climate* **26**, 2580-2600.
- Schwämmle V., Jensen O.N. (2010) A simple and fast method to determine the parameters for fuzzy c-means cluster analysis. *Bioinformatics* **26**, 2841-2848.
- Sinha P., & al. (2013) Seasonal prediction of the Indian summer monsoon rainfall using canonical correlation analysis of the NCMRWF global model products. *Int. J. Climatol.* **33**, 1601-1614.
- Wang W., Yang Y. (2008) On fuzzy cluster validity indices. *Fuzzy Syst.* **158**, 2094-2117.
- Webster P.J., & al. (1998) Monsoons: Processes, Predictability and the prospects for prediction. *J. of Geophys. Res.* **103**, 14451–14510.