Workshop on Air-sea Gas Flux Climatology:
Progress and Future Prospects

Impact of Climatology Data Geometry on the Results of EOF Analysis

Ge CHEN and Xuan WANG

Department of Marine Technology

College of Information Science and Engineering

Ocean University of China

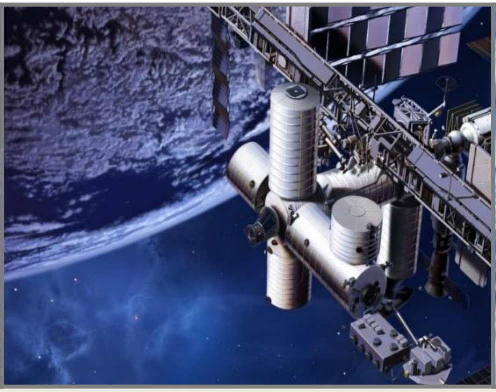
September 25, 2013· Brest, France

Contents

Big Marine Data: Opportunity

- The past half century experienced an explosive expansion in the volume of oceanic and atmospheric data, thanks to the substantial advances in earth observation technologies such as satellite remote sensing and Argo floats.

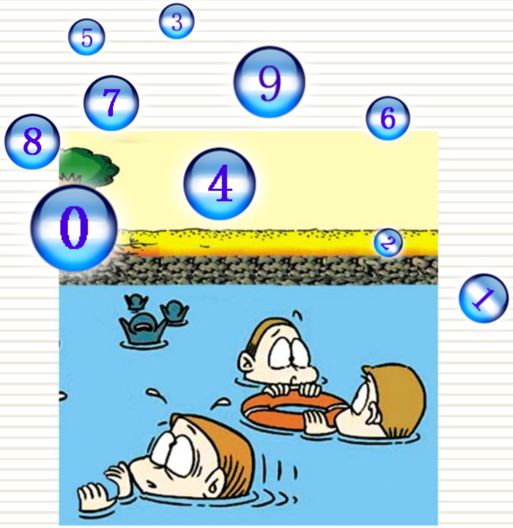
- These new platforms and systems have greatly increased the domains of space, time and spectrum for earth observation in terms of resolution, duration, coverage and continuity, leading to an unprecedented opportunity of having big marine data in huge dimensions.

Big Marine Data: Challenge

- Obviously, we are not fully prepared for handling terrabytes or even petabytes of new data on daily basis.

- How to analyze, interpret and make use of such a massive volume of data has become a serious challenge to the geophysical community.



Big Marine Data: Solution

- A statistical technique called EOF analysis was introduced over fifty years ago (e.g., Lorenz 1956), and is nowadays widely used in ocean and atmosphere sciences for data compression, dimensionality reduction, and pattern extraction.

Lorenz, E. N., 1956: Empirical orthogonal functions and statistical weather prediction. Statistical Forecasting Project Rep. 1, MIT Department of Meteorology, 49 pp.

Big Marine Data: Problem

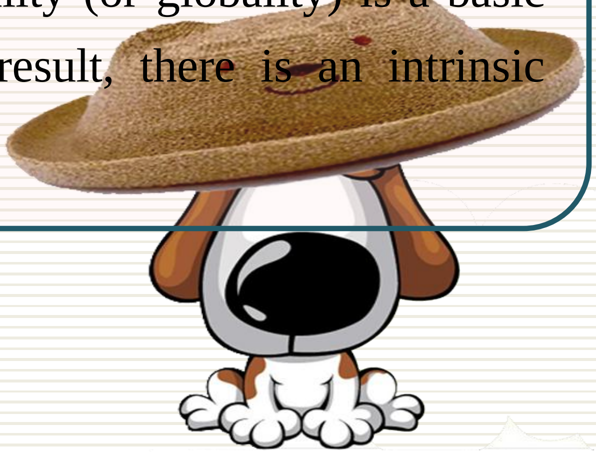
Statistical tools are likely to leave their imprint on the output, so that the results obtained are influenced by the features of the filters. EOF analysis is not free from this “bias”.

- Mode correspondency: Individual EOF modes will not correspond to individual dynamical modes, nor will they correspond to individual kinematic degrees of freedom.

- Spatial orthogonality: The EOFs are constrained to have structures over most of the domain with significant amplitude, when in fact one expects the patterns to be more localized.

Focus of the Present Work

Locality (or regionality) is a fundamental feature of most geophysical modes in the ocean/atmosphere system. Meanwhile, nonlocality (or globality) is a basic characteristic associated with EOF analysis. As a result, there is an intrinsic mismatch between the “head” and “cap”.



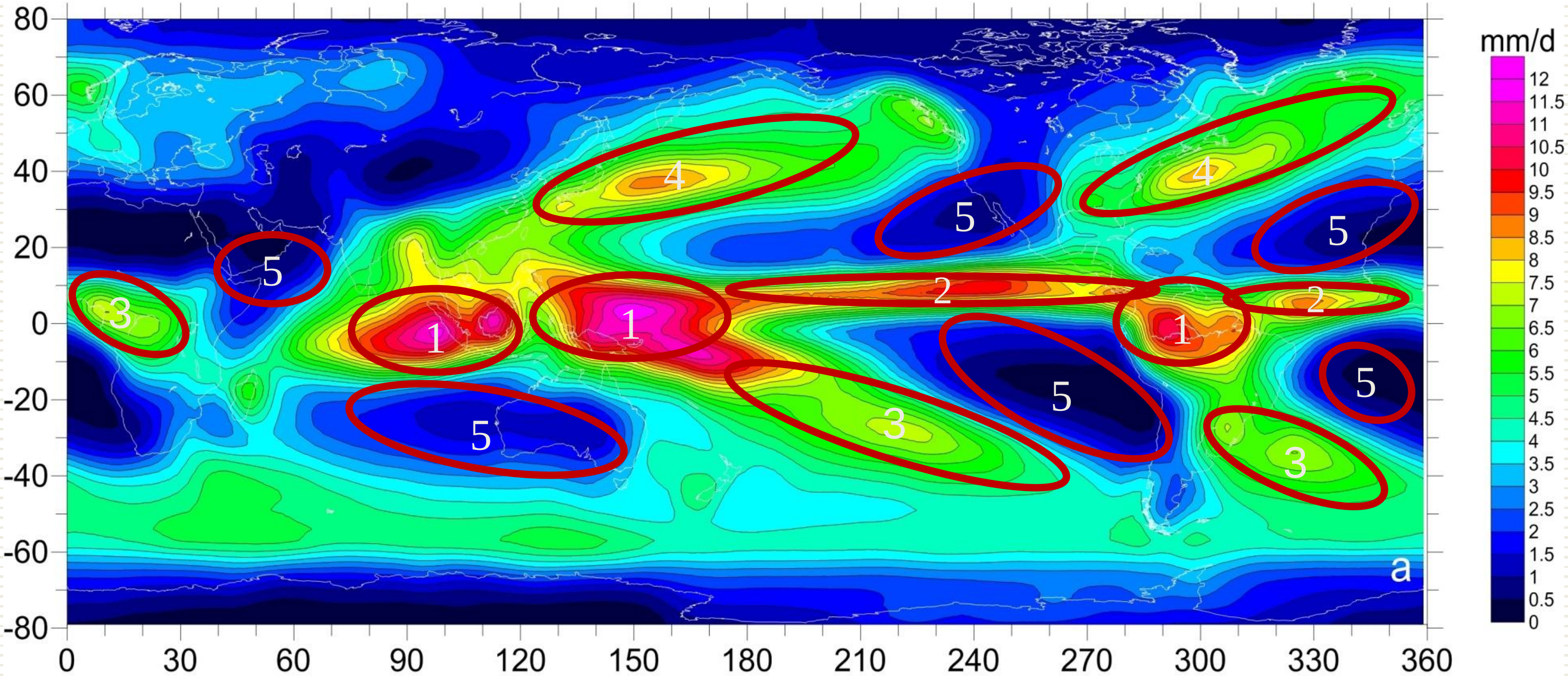
Impact of Domain
Geometry on the Results

Contents

Data Description

Dataset	GPCP Version 2.1 (Global Precipitation Climatology Project)
Type	Merged satellite-gauge product
Duration	01/1979 – 12/2011
Spatial coverage	Global
Temporal resolution	Monthly
Spatial resolution	$2.5^{o} \times 2.5^{o}$

Global Precipitation Climatology



(1) three prominent rain zones in the western tropical Pacific, eastern tropical Indian Ocean, and over the tropical South America;

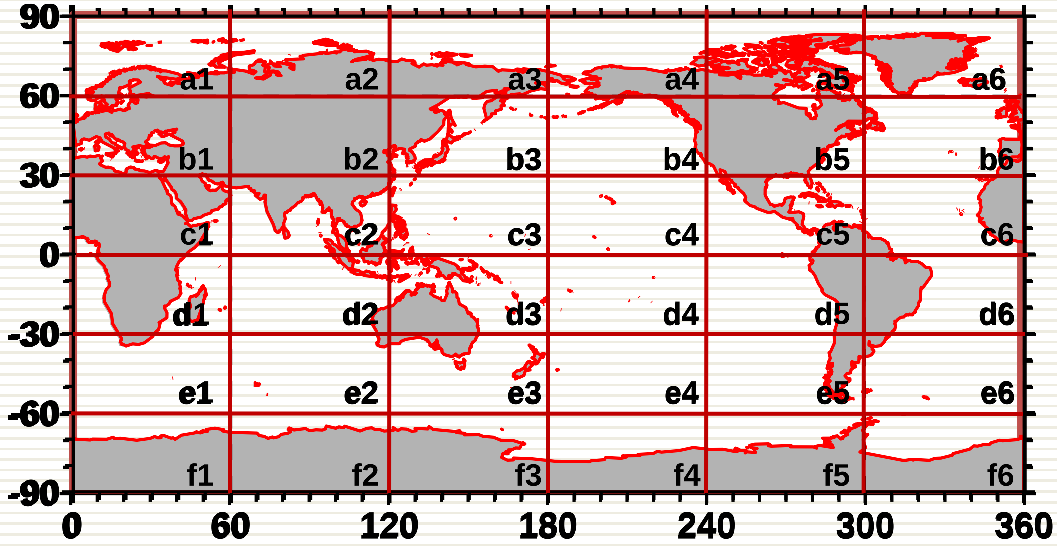
(2) a pronounced rain belt associated with the ITCZ throughout the equatorial region;

(3) a northwest-southeast oriented rain band related to the SPCZ, and its counterparts over the Atlantic and southern Africa;

(4) two southwest-northeast oriented rain bands corresponding to storm tracks in the Northwest Pacific and Northwest Atlantic;

(5) six "marine deserts" in the eastern subtropical areas of the three ocean basins.

Domain Division Scheme



- the entire globe;
- a land/ocean separation;
- six 30º zonal bands;
- six 60º meridional segments;
- thirty-six 30º ×60º tessellated cells.

Fraction of Variance Explained by Modes 1-3

Fraction of variance explained by Modes 1-3 of precipitation EOFs for the globe, land and ocean.

Dividing Scheme	Mode 1	Mode 2	Mode 3
Global	30.98%	8.82%	4.05%
Land	50.70%	6.66%	5.11%
Ocean	24.79%	8.86%	5.12%

- Globally, Mode 1 explains nearly 1/3 of the total variance, in contrast to roughly 1/2 for the land and 1/4 for the ocean. Higher order modes are all below 10%, showing that the primary mode is much more dominant for the land than for the ocean.

Fraction of Variance Explained by Modes 1-3

Fraction of variance explained by Modes 1-3 of precipitation EOFs for six zonal bands.

Dividing Scheme	Mode 1	Mode 2	Mode 3
60^o-90^oN	23.66%	18.46%	6.34%
30^o-60^oN	30.28%	6.94%	4.03%
0^o-30^oN	39.01%	9.75%	5.30%
0^o-30^oS	32.75%	8.14%	6.49%
30^o-60^oS	15.43%	6.43%	4.21%
60^o-90^oS	18.28%	12.76%	9.24%

- As for the six zonal bands, Mode 1 varies from a minimum of 15.43% for 30^o-60^oS to a maximum of 39.01% for 0-30^oN. Higher percentage of variance for Mode 2 appears in polar regions: 18.46% for 60^o-90^oN, and 12.76% for 60^o-90^oS. Higher order modes are found to be lower than 10% for all divided zones.

Fraction of Variance Explained by Modes 1-3

Fraction of variance explained by Modes 1-3 of precipitation EOFs for six meridional segments

Dividing Scheme	Mode 1	Mode 2	Mode 3
0º-60ºE	42.33%	7.76%	6.56%
60º-120ºE	43.66%	8.75%	5.12%
120º-180ºE	24.59%	8.26%	7.16%
180º-240ºE	20.38%	12.42%	8.81%
240º-300ºE	42.56%	8.54%	4.25%
300º-360ºE	34.63%	17.32%	5.35%

- In a meridional perspective, 0º-120ºE and 240º-300ºE are over 40% for Mode 1, while 180º-240ºE has a lowest value of 20.38%, displaying a comparable range with zonal bands.

Fraction of Variance Explained by Modes 1-3

Fraction of variance explained by Modes 1-3 of precipitation EOFs for thirty-six tessellated cells.

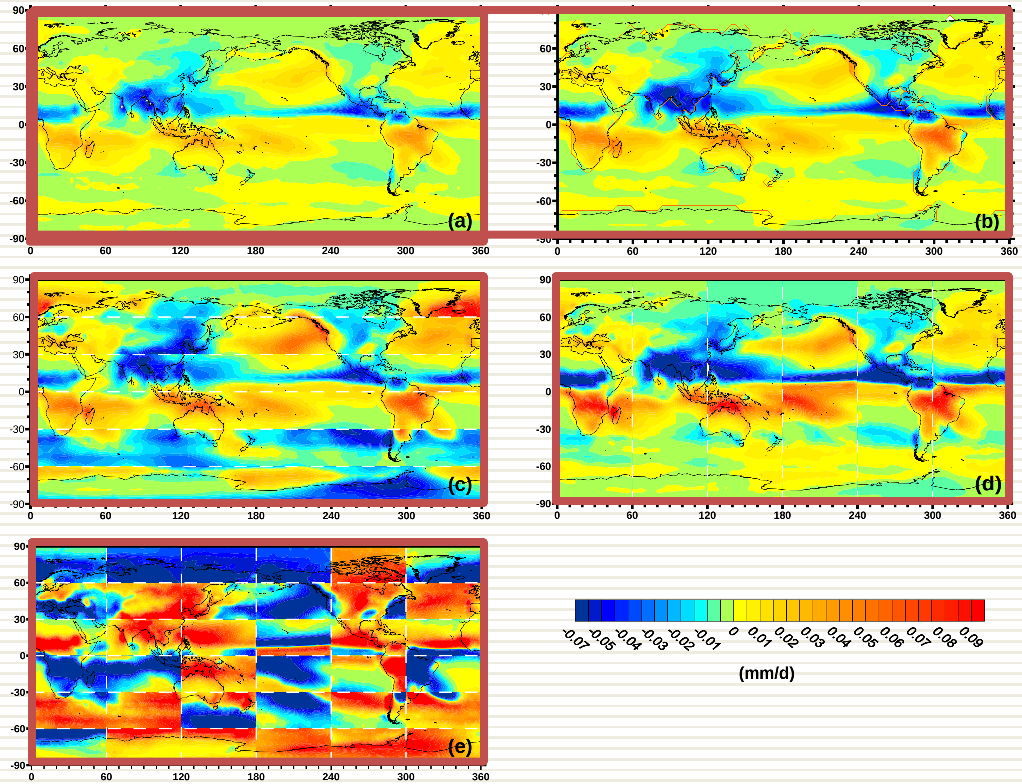
Cell Index	Mode 1	Mode 2	Mode 3	Cell Index	Mode 1	Mode 2	Mode 3	Cell Index	Mode 1	Mode 2	Mode 3
1	39.85%	15.77%	10.35%	13	47.36%	16.31%	6.85%	25	53.33%	11.03%	5.16%
2	23.02%	14.29%	9.93%	14	26.99%	12.93%	9.26%	26	22.17%	9.86%	8.85%
3	60.25%	13.88%	5.04%	15	29.14%	13.20%	10.16%	27	52.28%	9.19%	5.19%
4	56.93%	9.17%	5.69%	16	36.50%	13.01%	7.29%	28	51.21%	15.47%	7.26%
5	23.06%	10.05%	7.00%	17	14.03%	12.52%	9.85%	29	36.36%	11.57%	7.31%
6	45.11%	12.43%	8.03%	18	27.72%	18.69%	11.04%	30	36.06%	17.94%	10.15%
7	36.07%	16.21%	12.65%	19	42.28%	16.61%	10.80%	31	47.10%	13.15%	7.48%
8	59.52%	6.76%	5.66%	20	45.00%	11.53%	5.82%	32	35.69%	9.95%	6.75%
9	56.05%	10.10%	4.41%	21	35.21%	18.51%	6.48%	33	49.51%	13.74%	10.88%
10	30.83%	13.57%	9.59%	22	30.91%	16.65%	9.05%	34	49.89%	21.87%	5.55%
11	25.59%	10.57%	9.21%	23	18.81%	12.63%	9.43%	35	16.14%	10.14%	9.58%
12	31.10%	19.28%	10.23%	24	26.08%	21.07%	12.41%	36	34.99%	24.14%	9.48%

- As far as the thirty-six cells are concerned, however, a very large discrepancy is observed for Mode 1 (ranging from 14.03% for Cell 17 to 60.25% for Cell 3) and Mode 2 (ranging from 6.76% for Cell 8 to 24.14% for Cell 36), indicating that the locality of principal precipitation modes is remarkable. Consequently, the use of a single global mode to represent regional variability is far from adequate.

Contents

1. Background
2. Data and Scheme
3. Domain Impact on Primary Mode
4. Domain Impact on Secondary Mode
5. Conclusions

Spatial Patterns of GPCP Mode 1



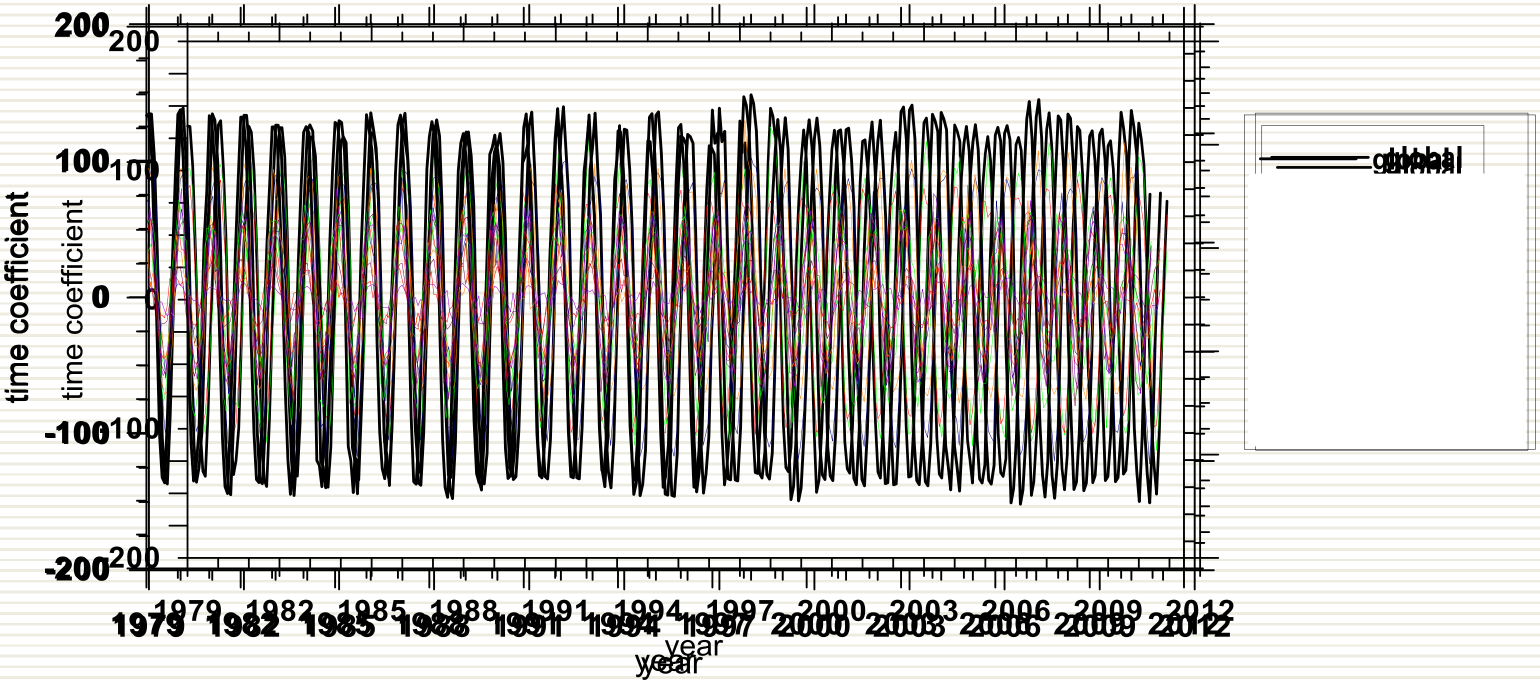
- In contrast, a systematic inconsistency between neighbouring cells is observed in Fig. (e), suggesting that precipitation modal structure in terms of both sign and amplitude is highly regional. Original local features will be highly smoothed in amplitude and 'artificially' adjusted in phase when an EOF analysis is applied on a global scale.

- Moving to the combined zonal bands, a large degree of discontinuity can be found outside the tropical regions, especially in the polar regions of the two hemispheres. Poleward from ±60, where annual rainfall is small, a considerable phase discrepancy is seen with respect to the global pattern.

- As for the meridionally divided map, the general global pattern can still be kept, but the contrast becomes even sharper, and apparent discontinuity occurs near the dateline.

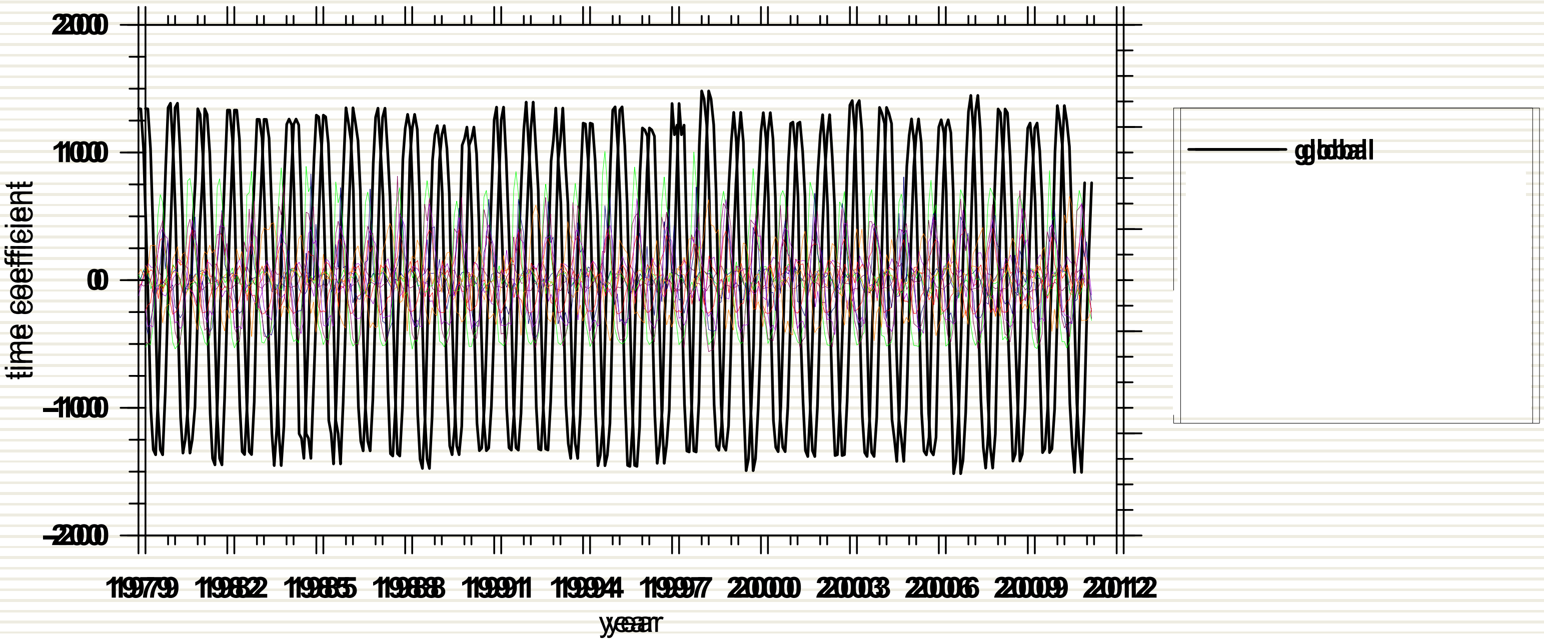
... scheme is useful in preserving the modal structures without causing much boundary effect.

Time Series of GPCP Mode 1



- For the six meridional segments, however, their amplitudes are more or less comparable (except for a few spikes within 180°-240°E), being roughly half of the global amplitude. An annual cycle with a phase consistency remains valid.
- For the six zonal bands, however, the time series are apparently latitude dependent. The amplitude decreases from the tropical areas to the polar regions which is largely proportional to the magnitude of rain climatology. The yearly cycle can be fairly kept until midlatitudes, beyond which the phase pattern appears to be much less regular for both hemispheres.
- The amplitude of associated time series is found to be the largest for the global one, followed by the ocean and the land, being generally proportional to the area of the analyzed domain. The three oscillations all show a dominant annual cycle which are highly in phase with each other.

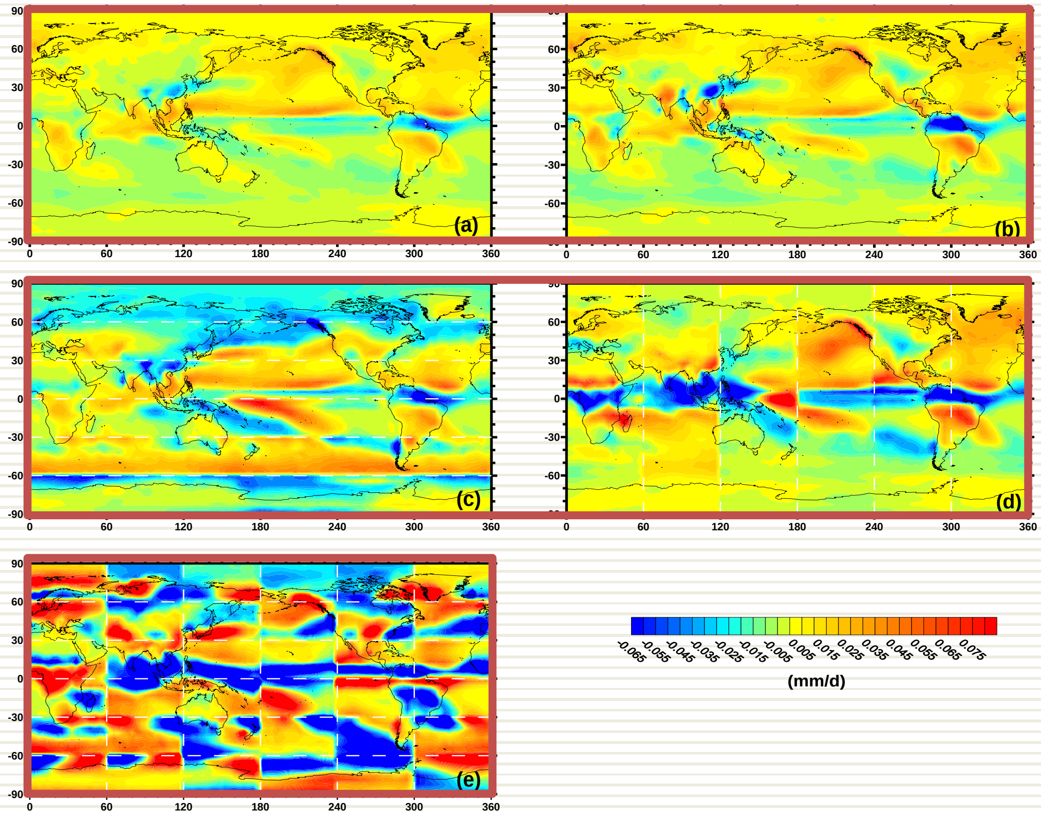
Time Series of GPCP Mode 1



- The above features are further confirmed by similar results for the thirty-six tessellated cells, as shown here for the polar region between 60º-90ºN, and the tropical region between 0º-30ºN.

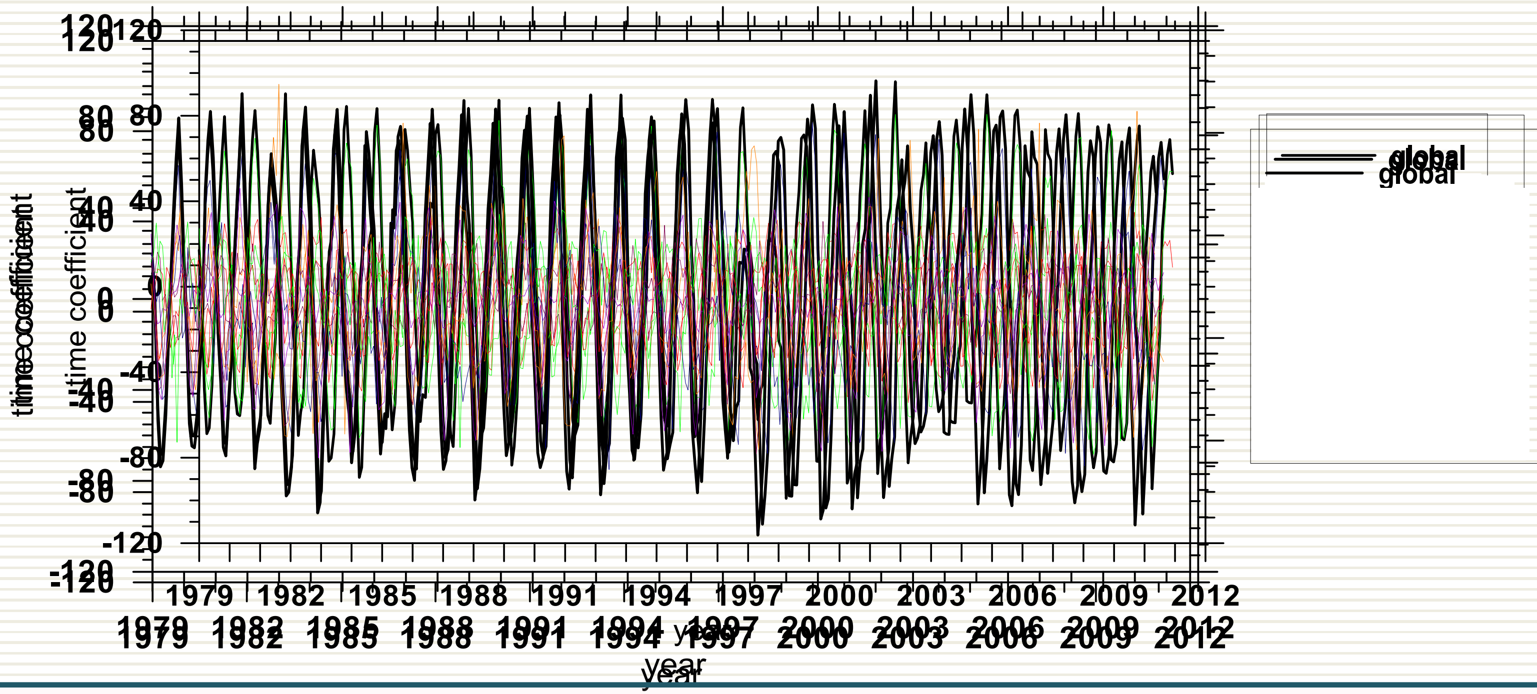
Contents

Spatial Patterns of GPCP Mode 2



- Similar to Mode 1, the land/ocean division does not affect the basic structure of the pattern, but increases the overall contrast by some 20%. One can still find a few slight shifts of the dipole centers such as those between China and India, and around central America.
- In the zonal and meridional combinations, the spatial contrast of Mode 2 continues to enhance. Although the general pattern of Fig. (c) [illegible] more visible.
- Contrary to Mode 1, however, zonal discontinuity appears to be more severe than meridional for Mode 2, implying that the impact of geographic division on EOF analysis is actually data sensitive.
- The montage of thirty-six cells exhibits a highly variable nature similar to Mode 1, confirming the intrinsic locality of precipitation modality.

Time Series of GPCP Mode 2



- The second principal components also bear an annual cycle in terms of periodicity, repeating
- The analysis of Mode 2 principal components indicates that the geographic dependency of EOF is in fact a data distribution dependency, i.e., it has a clear tendency to follow the region where geophysical maxima reside.

Contents

Land/Ocean Division

(1) In a land/ocean separated scenario, the overall effect on EOF result appears to be positive: retaining the basic spatial structure and increasing the dipole contrast with little loss of continuity along coastlines.

(2) Since most of the mode-active zones for precipitation are located over the ocean, the global EOFs are mathematically “tuned” by oceanic variabilities. A land/ocean separation strategy naturally releases such a constraint, and should therefore be recommended.

Zonal/Meridional Division

(1) When a zonal or meridional division is applied, common consequences on EOF result include a considerable increase in spatial contrast around dipoles and a certain discontinuity near boundaries. The time series are more sensitive and dispersive with respect to latitude than to longitude.

(2) Inappropriate or disproportional zonal/meridional divisions may introduce severe distortions to the derived regional EOF modes, and should therefore be avoided.

Concluding Remark

- The nature of geographic dependency of EOF analysis is ultmately attributable to geophysical dependency, implying that the geographic pattern of geophysical climatology and variability should be taken into account when the domain of an EOF analysis is defined.

- An ideal effectiveness can only be achieved by an optimal combination of an appropriately designed filter and a geographically (eventually geophysically) “matched” dataset, i.e., by putting the right “cap” on the right “head”.

Thank You!