Can we predict short term highs and lows in the Indian Monsoon? Identifying and Evaluating Monsoon Intraseasonal Oscillations in Climate Data

Patrick Orenstein

Thesis Advisor: Professor Baylor Fox-Kemper
Readers: Professors Jung-Eun Lee and Timothy Herbert

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Geology-Physics/Mathematics

Department of Earth, Environmental, and Planetary Sciences
Brown University
December 2018

Table of Contents

Chapter 1: Introduction

All of my work is divided into three parts; the first defines an atmospheric phenomenon, the second attempts to predict its behavior, and the third evaluates the ability of a global climate model to simulate it. That phenomenon is the Indian monsoon intraseasonal oscillation (MISO), which causes brief periods of especially intense rainfall during the Asian monsoon on the Indian subcontinent.

The Asian monsoon is a continental-scale seasonal rainfall pattern which accompanies a change in wind direction across the Indian ocean. It affects agriculture and natural disaster risk on the Indian subcontinent, so understanding its behavior on multiple spatial and temporal scales is very societally important. While the annual arrival of the monsoon is very predictable, its variations both year to year (interannual) and over the course of a single season (intraseasonal/subseasonal) are much harder to predict, and have been a topic of significant interest to researchers (Goswami et al. 2016).

At the most basic level, the MISO is defined as a deviation from the seasonal monsoon rainfall trend, which gradually increases over the course of the summer, peaks around late July, then decreases to its offseason intensity. This is generated in part by the annual north-south movement of the Intertropical Convergence Zone (ITCZ), the boundary between Northern and Southern hemisphere atmospheric circulation. As a result, MISO events occur in an extremely complex circulation context, making them difficult to predict more than a few weeks in advance. They are also inherently marine phenomena, involving significant ocean-atmosphere heat exchange, and are extremely dependent on the particular geometry and physical characteristics of the Bay of Bengal (Goswami et al. 2016).

Previous work has looked at the relationship between the interseasonal and intraseasonal variations in the Indian monsoon. Goswami and Mohan (2000) found that while the two behaviors act on different time scales, they are not independent phenomena. They exhibit similar spatial patterns, and the interseasonal variation can be alternatively viewed as a year-to-year change in *intraseasonal* activity. The authors inferred that the chaotic nature of intraseasonal oscillations therefore spelled defeat for researchers trying to predict year-to-year trends. The work in this thesis and research like it represent a possible pathway to approaching this problem anew.

Not only are MISOs important to the intensity of the monsoon overall, but positive oscillation phases have been shown to be correlated with a greater frequency of synoptic weather systems moving inland from the Bay of Bengal. Moreover, those storms associated with a positive MISO phase tend to form at a central point in the northern Bay of Bengal and travel northwest across India, steered by the low pressure border known as the monsoon trough. The place where the storms form is a relatively small region associated with the point of greatest MISO variation in the Bay, meaning that the movement and change of the MISO over time may

effect the origin and path of weather events in that region of the subcontinent (Goswami et al. 2003).

Similarly, the most significant mode of East-West tropical intraseasonal variation, the Madden-Julian oscillation (MJO), has been shown in global climate models to nearly double in simulations with quadrupled atmospheric CO_2 levels. The precipitation anomalies associated with the MJO are projected to increase by 10% with every degree C of surface temperature warming, partly due to increases in surface heat flux, but primarily due to a significant increase in vertical atmospheric circulation (Arnold et al. 2015). Since MISOs exhibit a similar mechanism, they may respond to climate change similarly.

Although researchers are interested in the effect of MISO on weather over land, this work particularly focuses on the Bay of Bengal for a few reasons. Firstly, while they interact with terrestrial weather, MISOs form in the Indian Ocean to the south and move northward, and previous studies have found a zone of peak variation to be centered on the bay. Secondly, studying a purely oceanic region eliminates the issue of local topography from the equation, allowing us to treat the entire section of interest as essentially a uniform field. Finally, this work originated as part of the US Office of Naval Research's (ONR) MISO-BoB project (Monsoon Intraseasonal Oscillation-Bay of Bengal). The research described here uses simple mathematical tools to address some of the questions raised by that project, hopefully shedding some light on how well current models simulate MISOs and laying out some paths to improving them.

Chapter 2: Detecting MISOs in daily mean rainfall data using extended empirical orthogonal functions (EEOFs)

Abstract

The first part of my research uses a statistical method to look for the signal of the Monsoon Intraseasonal Oscillation in precipitation data from the Bay of Bengal. Using an extended empirical orthogonal function (EEOF) developed by Suhas et al. (2012), I identified the signal of a 30-60 day long oscillation in both reanalysis data and results from two global climate models with different parametrizations of small-scale ocean turbulence. While EEOFs do not necessarily have a physical meaning, the results from all three models point to a real mode of north-south variability.

Introduction

Given the societal and scientific importance of the Indian monsoon, it makes sense that a great amount of effort has been put into studying short-term variations in the monsoon (Goswami et al. 2016). Previous studies have defined MISOs using a variety of indices, including atmospheric vorticity at 850 hPa (Goswami et al. 2003) and sea surface temperature (Sengupta et al. 2001). This chapter tries to approach the same problem using a different technique to parse spatiotemporal data looking for the signal of short-term oscillations. Empirical orthogonal function (EOF) analysis decomposes complex data sets into their primary modes of variability, revealing which geographic and temporal patterns are most significant to the overall variability. This allows oscillatory patterns to be revealed in a computationally cheap manner, and without having to make any assumptions about model physics.

Extended empirical orthogonal function (EEOF) analysis take this a step further, repeating this process starting at different points in time (see Methods). EEOF analysis can reveal patterns in sequential phenomena that change over time (Weare & Nasstrom 1982). This work applies an EEOF method based on the work of Suhas et al. (2012) and sets up a methodology which can be applied in future chapters to issues of predictability and bias measurement. Suhas et al. (2012) defined indices for intraseasonal variation based on EEOF analysis, termed MISO1 and MISO2. The work in this chapter essentially recreates those indices and applies them to additional data.

While increased rainfall does not define a MISO, it is a well-established result (and the most impactful on human activity). As a result, precipitation can be used as a proxy for MISO intensity, but the EEOF method may yield a more meaningful index than rainfall data on its own, either from a particular station or as a regional average.

Methods

1. *Data*

Following Suhas et al. (2012), I used the Global Precipitation Climatology Project (GPCP) as my observational data set. The GPCP is a reanalysis based on both satellite and historical observations (Huffman et al. 2009). This was compared to output from the National Center for Atmospheric Research (NCAR) Community Earth System Model, version 1.2 (CESM1.2). Two CESM data sets were used, 30-year model runs both with and without a parametrization for upper ocean Langmuir turbulence (Li et al. 2016). CESM data has been shown to have the smallest bias in simulating the monsoon compared to other global climate models (Anand et al. 2018). Both data sets had the annual mean and first three harmonics removed using the NCAR Command Language (NCL), so that the variation leftover were precipitation anomalies with only sub-seasonal variations. To focus on the Bay of Bengal, the following analysis was performed on a zonal average of precipitation data between 12.5 °S-30.5°N and 60.5-95.5°E.

2. *Extended Empirical Orthogonal Functions*

Empirical orthogonal functions (EOFs) are an application of singular value decomposition (SVD) which treats the decomposed values as representations of the temporal and spatial variability of a data set. Equation 1 shows the archetypal SVD in matrix notation, where M is a spatiotemporal data set, U and V are its left and right singular vectors, and S is the matrix of singular values.

$$\mathsf{M}_0 = \mathsf{USV}' \tag{1}$$

V has the same dimensions as the spatial dimension of the original matrix and can be taken to represent modes of spatial variability in the data, while U has the time dimensions and is viewed as representing the corresponding time series of those modes of variability. S is a diagonal matrix, representing the relative importance of those modes. Terminology is highly variable, but here the spatial modes are referred to as empirical orthogonal functions and the time series are principal components (e.g. EEOF1, EEOF2, and PC1, PC2).

By definition, the singular values are ordered from greatest to smallest. The special characteristic of SVD is that the majority of the singular values can be ignored without significantly affecting the accuracy of the recomposed M matrix. In most cases the first 20 EOFs can account for the majority of the variance in the data (Fox-Kemper 2004). This fact is applied in Chapter 3 to try to define a prediction method based on only the first few EOFs.

An *extended* EOF involves expanding the original M matrix, adding on duplicate data which is offset (or 'lagged') by one additional day (see Equation 2 for a schematic). By simultaneously performing EOF analysis on the same data from slightly different starting times, EEOF analysis can shed light on the evolution of the major modes over the span of days or week.

In this case, an EEOF with lags ranging from 1-16 days is used to look at the short-term evolution of the MISO modes.

$$\mathsf{M} = \left(\begin{array}{ccc|ccc|c} (x_1,t_0) & \cdots & (x_{end},t_0) & (x_1,t_{-1}) & \cdots & (x_{end},t_{-1}) & \cdots \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \cdots \\ (x_1,t_{end}) & \cdots & (x_{end},t_{end}) & (x_1,t_{end-1}) & \cdots & (x_{end},t_{end-1}) & \cdots \end{array}\right) \tag{2}$$

EOF analysis has limitations due to its lack of dynamical cause-and-effect, meaning the decomposition does not imply the oscillatory modes have any physical significance. It has been shown that EOFs can generate false appearances of order in data with no structure (Dommenget & Latif 2001) The easiest proof-of-concept for the EOF method comes from understood periodic events like El Niño, which is easily identifiable when it shows up as the dominant mode of variability for an EOF over a large region like the Pacific or the whole earth.

For the purposes of this work, MISO maxima and minima are identified here as peaks and troughs in PC1 (defined as extreme by being outside the 10th and 90th percentiles). Composite maps of the difference in regional rainfall anomaly between active and break phases of the MISO are calculated as the average difference precipitation between the maximum and minimum MISO stages. While the EEOF is formulated based only on rainfall in the Bay of Bengal, composite plots show a wider region for context, and to show how the MISO phases defined by the EEOF manifest in the Indian Ocean as a whole.

Results

The oscillatory events identified by the EEOF method have an average period of 40 days, which is consistent with previous descriptions of the MISO as approximately 30-60 days long (Goswami et al. 2016). There was some variation between the data sets: while my observational data exhibited MISO phases with an average length of 37 days, data from the CESM model without Langmuir turbulence showed an average phase length of 62 days, and CESM with Langmuir turbulence showed an average phase length of 49 days. Figure 1 breaks this data down, showing the frequency of positive and negative phases of different lengths in each data set. In addition to the greater average phase length, the model data also shows a much broader distribution of periods, indicating greater variability in phase shape. Overall, this provides some indication that CESM with Langmuir does a slightly better job of simulating MISOs as they occur in the real world.

While the EEOF analysis does a good job of uncovering modes of variability, the periods of the modes means that the 16-day lag cannot be expected to represent the full cycle of MISO variability. Indeed, the two most significant modes of spatial variability, EEOF1 and 2 (termed MISO1 and MISO2 by Suhas et al.), show a definite but incomplete cycle of northward

movement over the course of a selection of 8 out of the total 16 days of lagged repeated data (Figure 2a,b,c). MISO1 is essentially stationary in all cases, while MISO2 shows a slow but clear movement north over the lags. This evidence is discussed further in Chapter 4.

Visual evidence of the physical meaning of the EEOF modes can be found by plotting principal components 1 and 2 alongside zonal precipitation (Figure 3). The precipitation data show clear northward-moving phenomena associated with peaks and valleys in the principal components. Points used to generate composite phase maps are plotted on PC1 (near-duplicates are caused by the algorithm used to reduce the number of duplicate points).

Finally, the composite maps (Figure 4a,b,c) based on PC1 show clear north-south modes of variation in precipitation anomaly through the Bay of Bengal. The GPCP data displays a higher amplitude of variation, and another center of variability to the West of India, which is not part of the definition of a MISO but which could be correlated with it.

Discussion

That the phenomenon identified by the EEOF analysis has a period similar to that previously described for the MISO does not in of itself testify to the method's success, but together with the spatial structure of the EEOFs and the patterns of the composite maps they tell a story of periodic motion which is highly consistent with the MISO. While the length of the lags used is insufficient to cover a full MISO event, they do show the characteristic northward trend of the precipitation maximum.

The composite maps shown for CESM (Figures 4b and 4c) data have a lower spatial resolution than the GPCP observational data—current work involves upscaling the GPCP data onto the model grid. This difference explains some of the difference in complexity between the maps, but does not explain the difference in magnitude. GPCP data shows a much greater difference between the oscillation nodes in the north and south of the Bay of Bengal. This could be due to lesser degree of spatial variation characteristic of models which have some degree of simplification in their ocean and atmospheric physics. However, it is noteworthy that the CESM scenario without Langmuir turbulence has more structure than the scenario with it, which would tend to imply the opposite. It may be that the model does not have the diffusivity problem that models have on the scale of oceans, but has some other difference in process parametrization which might explain the lack of structure in the Langmuir scenario.

Conclusions

These results verify the EEOF method of Suhas et al. (2012) and show that it can be applied to other data in order to reveal local modes of variability like the MISO. Overall, it appears that Langmuir turbulence significantly effects MISO behavior in climate models, which makes sense given that upper ocean turbulence plays a large role in ocean-atmosphere interactions and the MISO is an inherently marine weather phenomena. Adding a Langmuir parametrization to the CESM model appears to generate MISO events with a more realistic

period and spatial profile, while model data without it seems to create a simplified, exaggerated MISO phenomenon. Future steps include comparing composite phase maps made using EEOF2 instead of EEOF1 to see how precipitation in the region changes between positive and negative phases of MISO2.

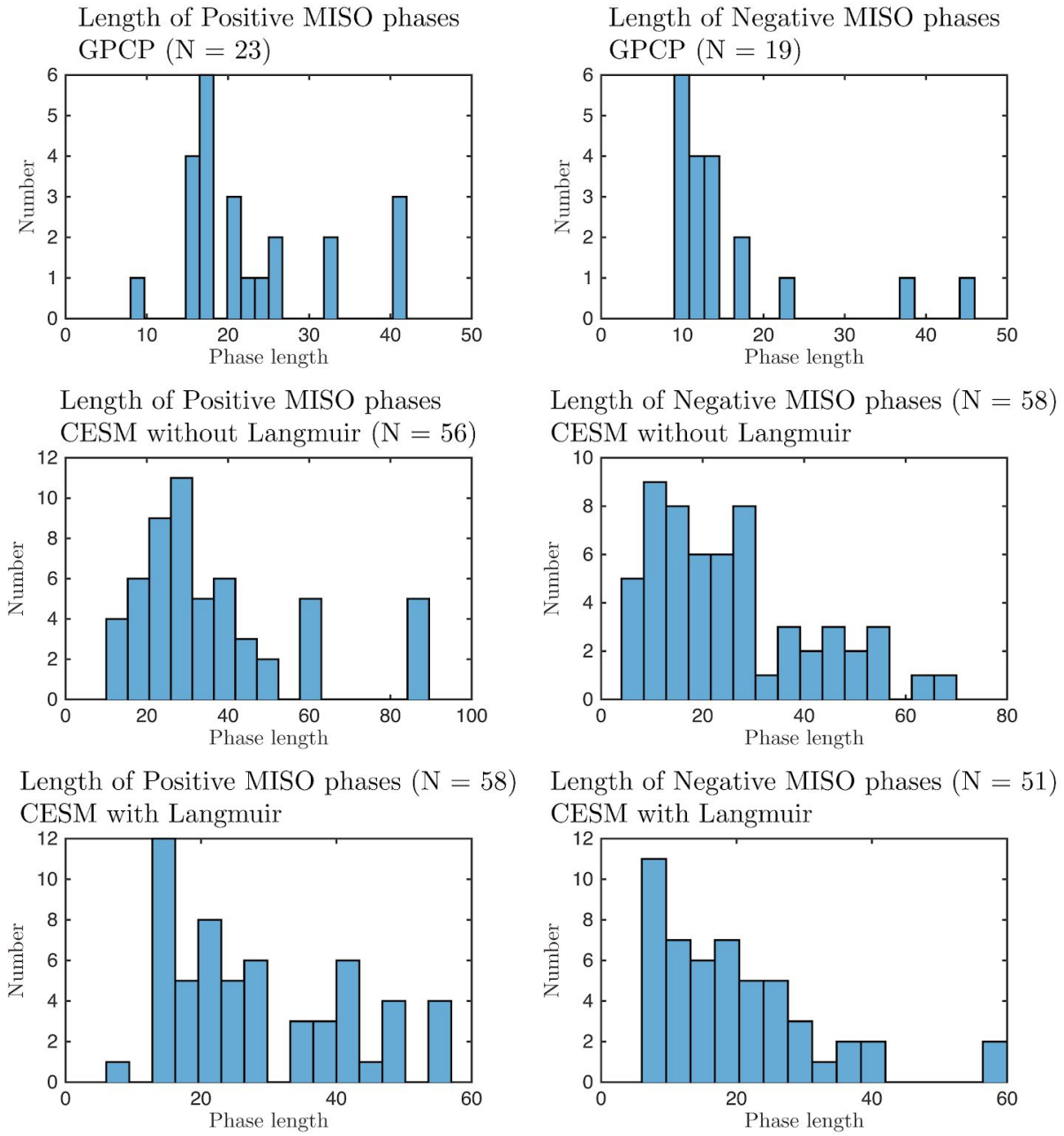

Figure 1: Lengths of positive and negative MISO phases defined by the 10th and 90th percentiles of the Principal Component 1 time series. Not the greater number of phases for the CESM model data than the observational record due to the difference in record length.

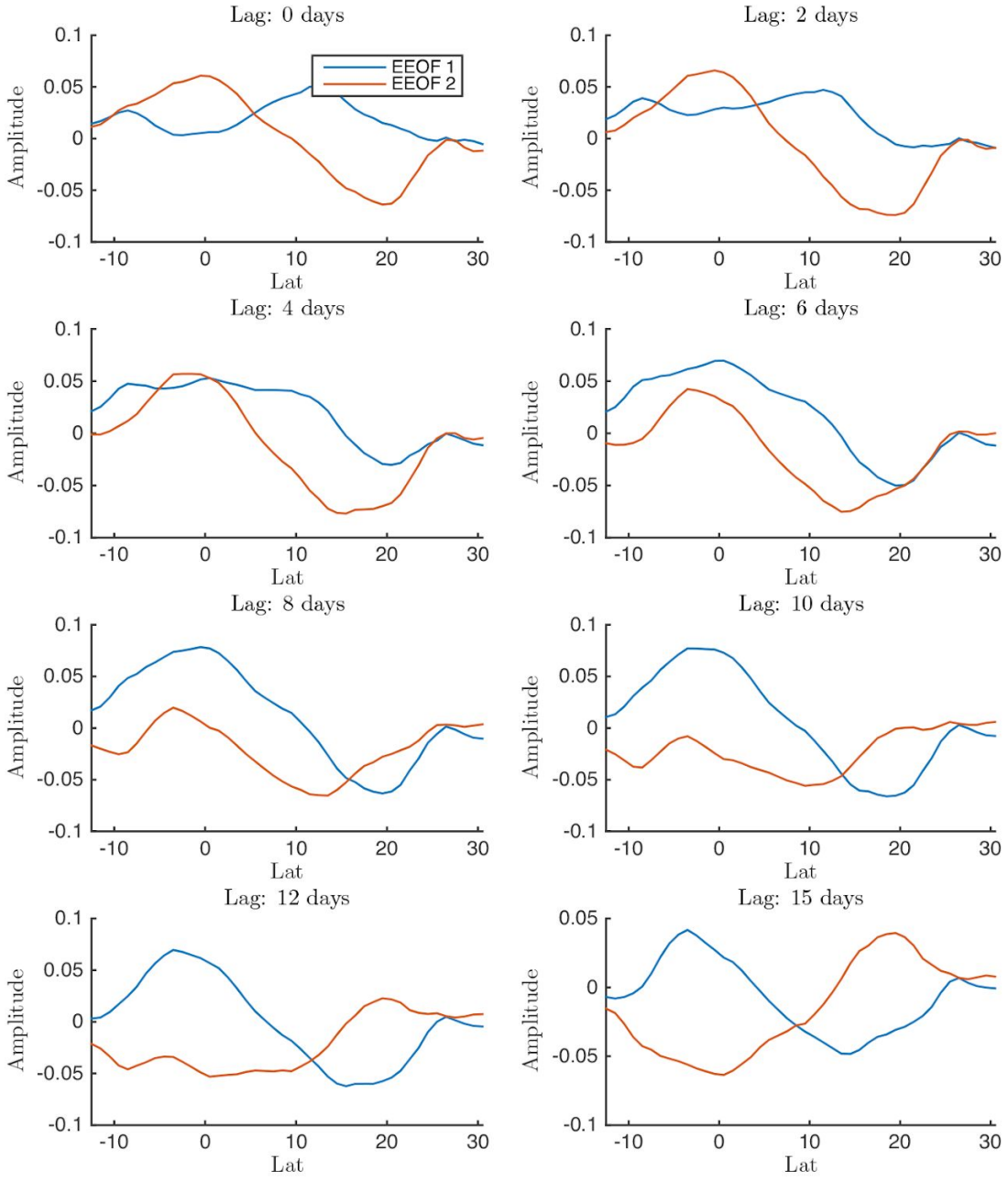

Figure 2a: First two principal components from GPCP observational data, showing 8 selected lags out of a total of 16. PC1 in blue, PC2 in red. Latitude is the horizontal axis.

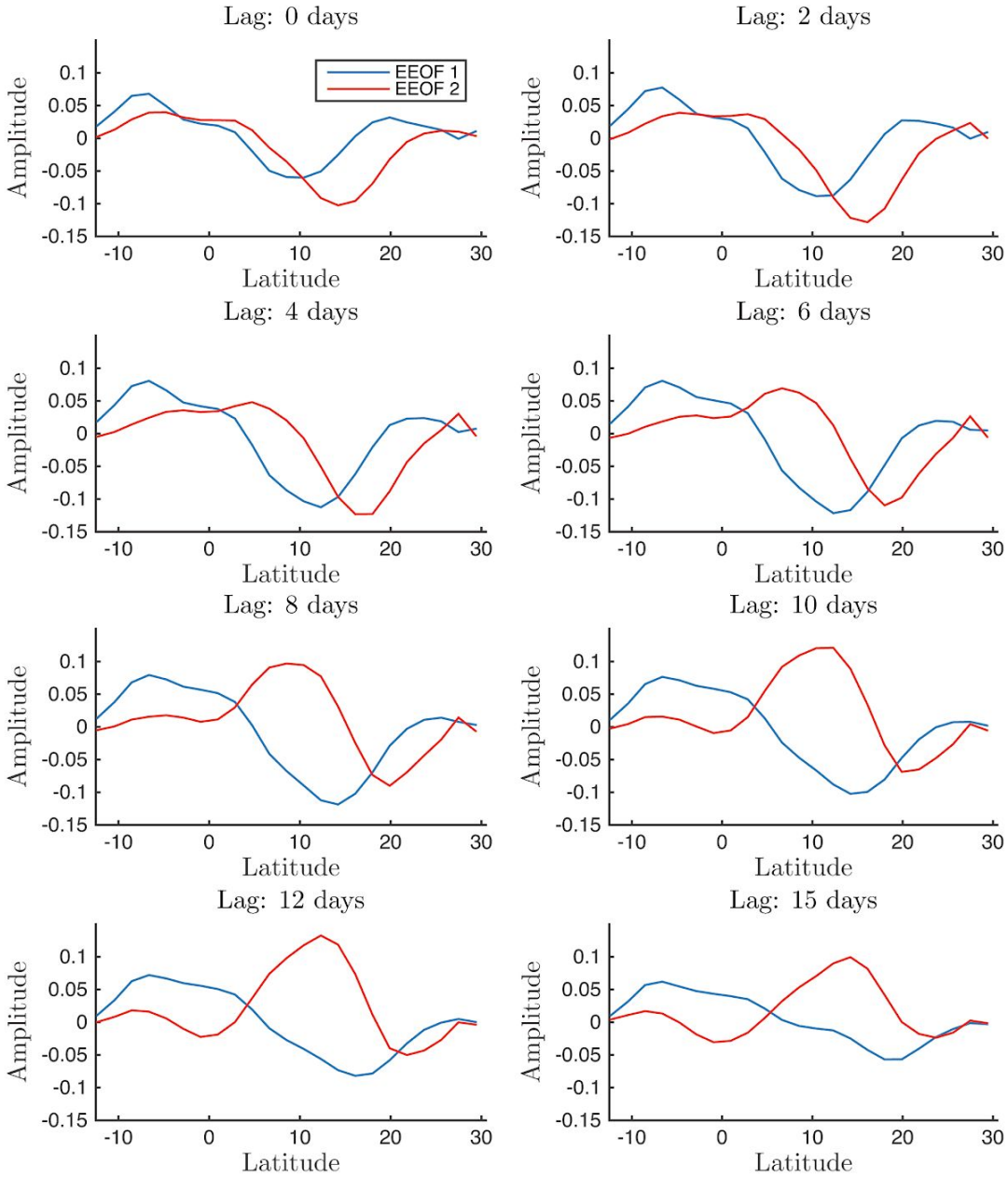

Figure 2b: Same as Figure 2a, for model data from CESM without Langmuir parametrization.

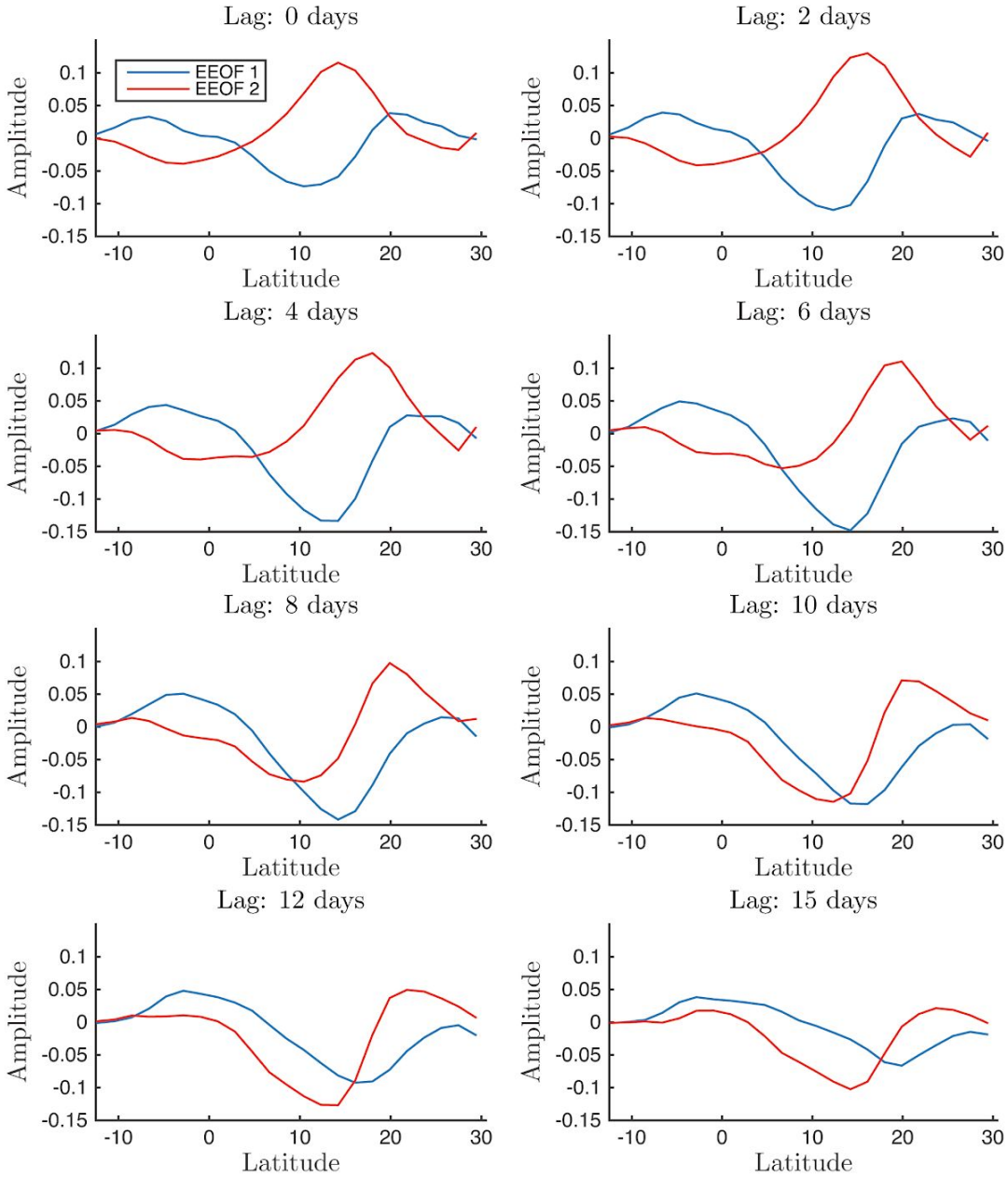

Figure 2c: Same as Figure 2a, for model data from CESM with Langmuir parametrization.

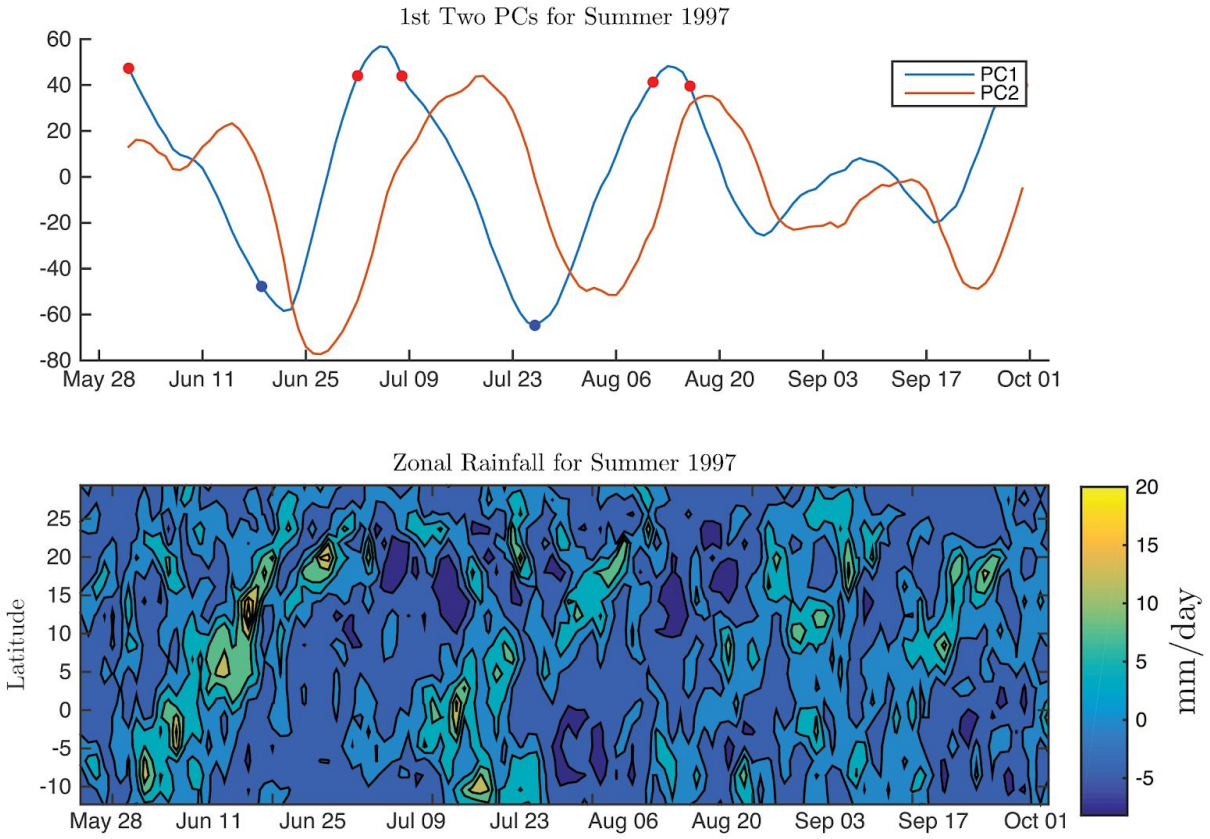

Figure 3: Top: PC1 and PC2 for GPCP precipitation data from the 1997 summer monsoon season; dots show time points used in generated composite phase maps (Figure 4). Bottom: Hovmoller plot showing zonal average rainfall anomaly over the same time span (same latitude range as is used in the EEOF analysis).

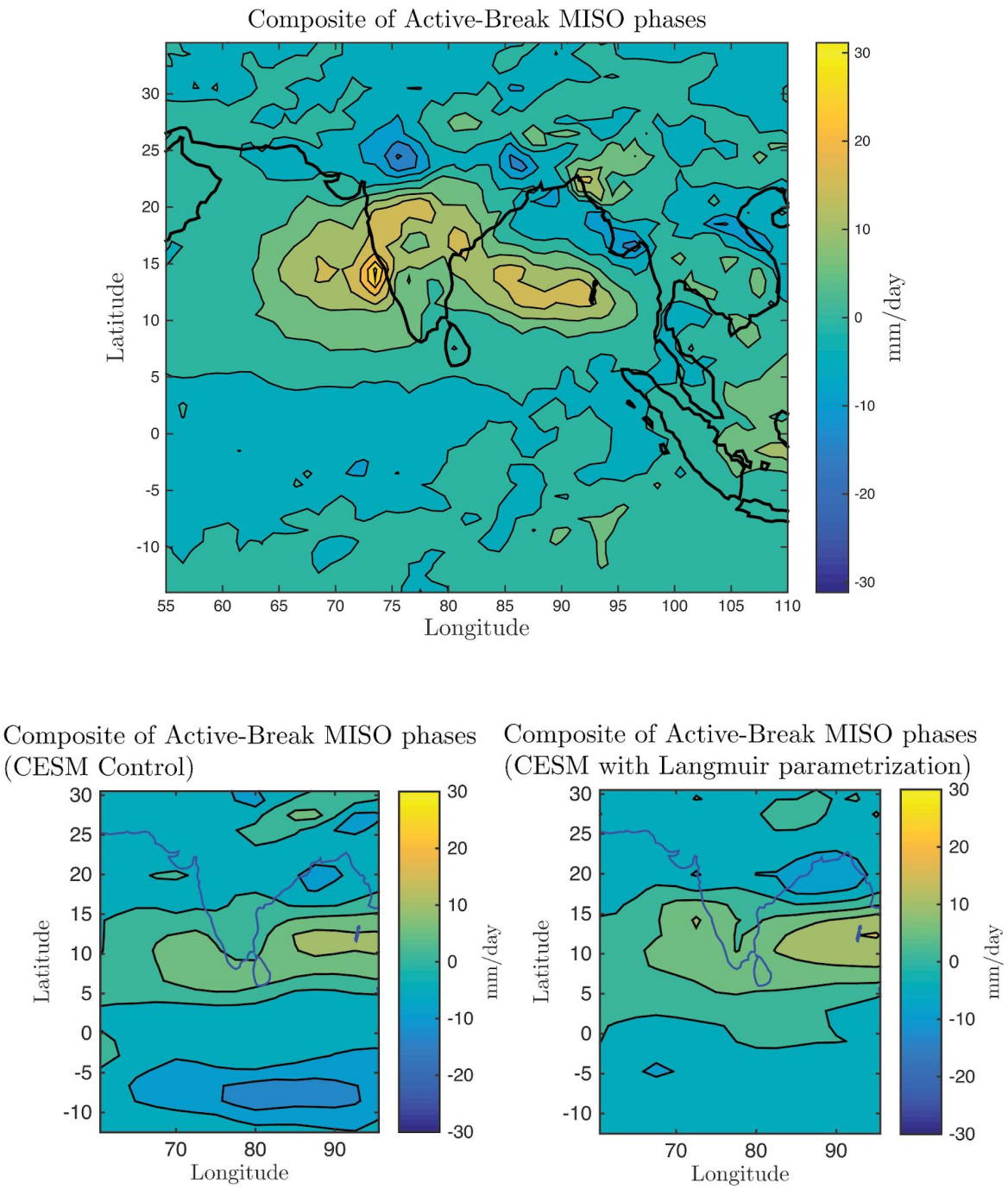

Figure 4a,b,c: Composite plots of the precipitation anomaly difference between positive and negative MISO phases, created by averaging precipitation data from the times of PC1 peaks and troughs. a.) GPCP (observational), b.) CESM without Langmuir parametrization, c.) CESM with Langmuir parametrization.

Chapter 3: Anticipating Intraseasonal Oscillations Using an EEOF-based Prediction System

Abstract

This chapter presents the methodology and preliminary results of using an EEOF-based prediction method to project MISO variability. Since EEOFs deconstruct spatiotemporal data into its most significant modes of variability, it is reasonable to propose a prediction system using only the great modes to recreate a data set from the SVD elements themselves. Here this method is tested using a comparison test involving GPCP data; the data is split into "training" and "prediction" sections, with the first being used to generate EEOFs and the second used as a comparison test as to how well the EEOF reconstruction agrees with actual data.

Introduction

Using an empirical method like an EEOF to devise a prediction method is a very attractive idea. Unlike a dynamic model, it does not depend on taking into account every possible physical process. It would also eliminate human biases like model tuning, since if it cannot take into account ocean or atmospheric physics it can also never include biases like tuning or misunderstandings of processes.

The work which resulted in this thesis explored multiple approaches to developing an EEOF-based prediction scheme, including a version of a previously published method using the first 10 EEOFs for an analog to develop a catalog of MISO events to match to similar events in the hopes of predicting future activity on each scale of variation. Predicting the course of the first 10 principal components (the time series corresponding to each EEOF), then recomposing them with the EEOFs and singular values, obtained a projection for the overall field with predictive skill out to 20-25 days (Xavier & Goswami 2007). The prediction framework in this project uses EEOFs in a different way, but with the same goal (see Methods).

An important distinction to be made is the difference between predictability and model skill. Predictability refers to whether a model gives new, meaningful information different than what can be learned from simply looking at past data. Put another way, it means how long a prediction model output is *different* than climatology or persistence before the best guess is simply repeating previous information (similar to the approach of a farmer's almanac). Predictive skill, on the other hand, refers to how accurate the model's predictions are relative to observations. This means a model might have great predictive skill while simply predicting the same thing as climatology, in which case the phenomenon has low predictability and the model is hardly worth its computation time.

Methods

This part of my research also used GPCP data, this time broken in two to create a quantifiable prediction problem. An SVD is done on the first, "training" part as in Chapter 2, generating a set of EEOFs which represent the modes of variability in only that portion of the data. Lower-order EEOFs are then discarded, generating a truncated right singular vector matrix V_T. A prediction matrix was formulated by applying equation 3 to a synthetic matrix made up of the training period data and a persistence data set the same length as the prediction period.

$$\mathsf{M}_{prediction} = \mathsf{M}_{synthetic}\mathsf{V}_T\mathsf{V'}_T \tag{3}$$

There are multiple options for measuring the success of the prediction method. In this work I looked at the entire data field at once, evaluating the similarity between the entire $M_{prediction}$ and the original data (M_0) matrices with a skill score (Equation 4).

$$SS = 1 - \frac{\sum (\mathsf{M}_{prediction} - \mathsf{M}_0)^2}{\sum (\mathsf{M}_{prediction} - \mathsf{M}_{reference})^2} \tag{4}$$

Results and Discussion

This procedure is still a work in progress, and will be improved as this project continues towards eventual publication. For now, a possible version of an eventual prediction method is shown in Figure 5, which compares the original data for the last 15 days of the training period and the first day of the prediction period to the same time period in the data recomposed from the truncated V matrix using only the first 5 singular values. This time period corresponds to the greatest time lag used to construct the EEOF, which should correlate with the method's eventual time length of predictive skill.

The decrease in coherent structure caused by truncating the singular values is clearly apparent in the training period section of the recomposed data (Figure 5, bottom panel). The first prediction day is close to zero, but Figure 6 is included here to show that it is not, but simply has very small magnitude of variability. This procedure with be subject to improvement in the future in order to generate more realistic values and hopefully extend the method's predictive horizon.

Discussion

While a purely empirical prediction offers many advantages in terms of simplicity, lack of model bias, and computational cost, it has limitations when applied to highly variable phenomena like the MISO. Other work using variations on empirical methods may provide so guidance to improve on these methods in the future. For example, Chen et al. (2018) used nonlinear Laplacian spectral analysis to identify the two greatest patterns of spatial variation of

the MISO in a similar way to the EEOF analysis here. The prediction scheme they developed using those two modes was also constrained using model physics, therefore augmenting the empirical approach with a dynamical understanding of the problem. This method achieved predictive skill out to 20-50 days in advance, and offers the promise of an empirical prediction which can also be used to examine the base mechanics of the MISO, which an EEOF by itself cannot.

Next steps in this work will also involve incorporating multiple methods of evaluating predictability and model skill. For example, Xavier and Goswami (2007) measured prediction skill based on the correlation of a spatial average of the predicted quantity with observations (they used outgoing longwave radiation, a proxy for cloud cover, instead of precipitation). Skill could alternatively be measured by comparing the observed and predicted time series at a particularly meaningful location, such as the center of apparent variation in the northern Bay of Bengal. Overall, methods that allow for the prediction of short-term variability with a minimum of training data can fill a variety of roles from real-time data assimilation to dynamic model diagnosis.

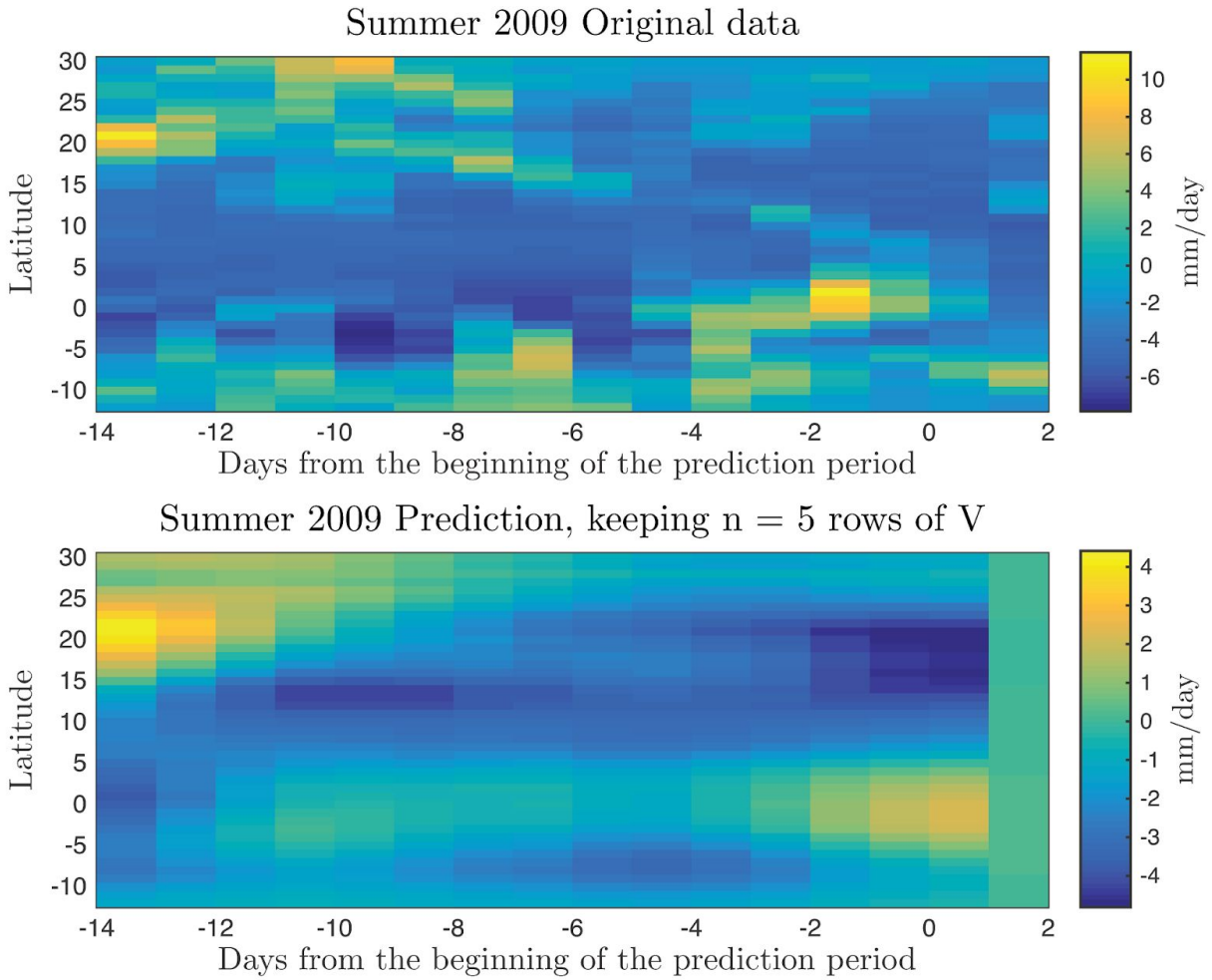

Figure 5: Comparison of original GPCP precipitation data for the two weeks leading up to the prediction period, and the first day of the prediction period, compared to the prediction data set recomposed using on the first 5 singular values (see Equation 3). Note the decrease in coherent structure in the recomposed data set even during the training period due to the absence of any variation beyond the first 5 modes. The first prediction day shows a low amount of variation, but is not entirely without structure (see Figure 6).

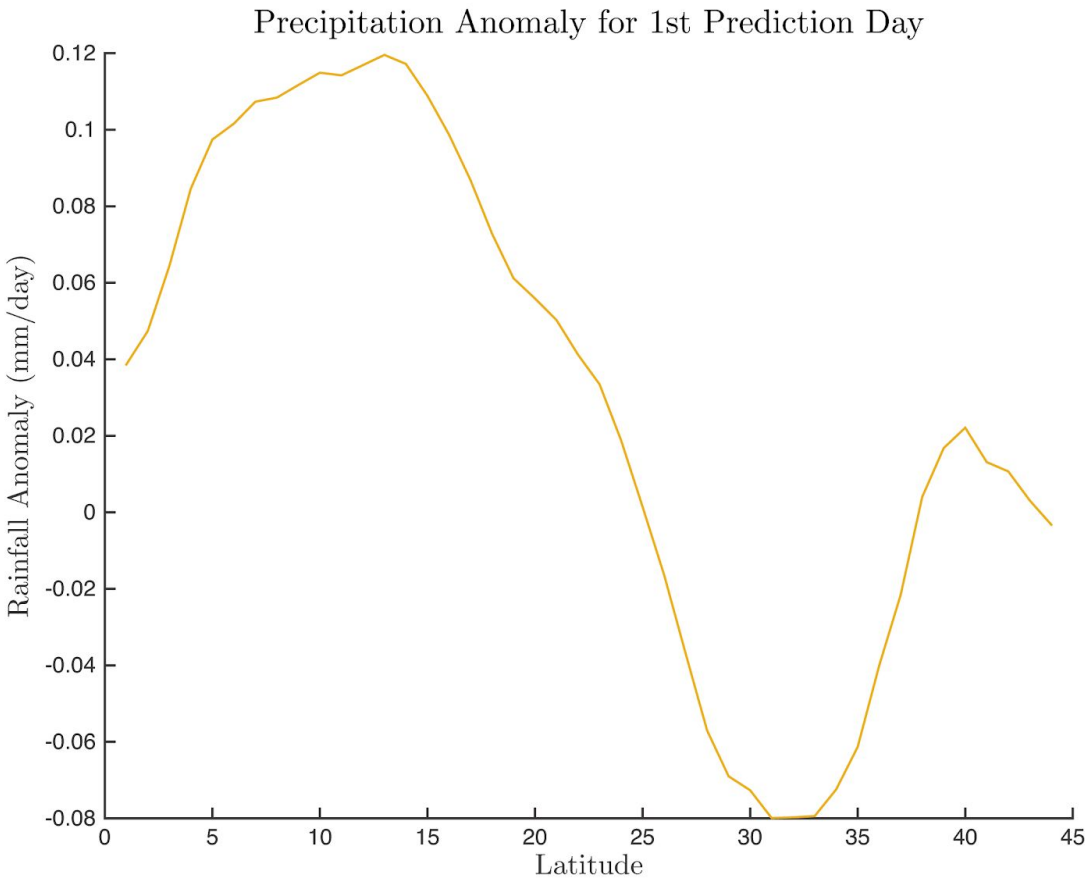

Figure 6: Zonal average precipitation on the first prediction day, recomposed using on the first 5 modes of variability. Note the extremely small magnitude of the rainfall anomaly, which shows up in Figure 5 as close to 0.

Chapter 4: Using the EEOF prediction method to evaluate climate model accuracy

Abstract

The final component of my research involved using the EEOF prediction system to compare the relative success of multiple climate models in simulating MISOs most similar to those seen in observations. Specifically, I compared the success of the CESM model with and without a parametrization for Langmuir turbulence developed by Li et al. (2016) as used in Chapter 2. Langmuir turbulence increases ocean boundary layer mixing, deepening the mixed layer and possibly impacting ocean-atmosphere interactions. Using the prediction scheme to evaluate the model's skill in simulating MISOs is potentially a more rigorous way of comparing data sets than the comparisons in Chapter 2. Testing the prediction method specifically tests how adding the Langmuir parametrization affects the predictability of modeled MISOs.

Introduction

Having applied the EEOF method to both observational and model data, it makes sense to try to use that process to evaluate the relative skill of the models in simulating MISOs. Using EEOFs to compare the accuracy of CESM with and without Langmuir turbulence could provide another measure of how much of a difference upper ocean mixing has on MISO events, as well as more generally how well a global climate model like CESM simulates small-scale phenomena like the MISO.

The skill score used in Chapter 3 is used there to compare parts of a data set to each other, and may not be as applicable to the problem of evaluating the difference between two models. Additionally, the problem of model skill is not so much one of predictive accuracy as in measuring the model's ability to simulate the MISO phenomenon as a whole realistically. The problem is using the EEOF definition of MISOs to measure how well the models recreate them. This chapter looks into two simple approaches to that and lays the foundation for more work.

Since the EEOF methodology is also relatively computationally cheap, using it as a model evaluation took could provide a quick test of model skill before embarking on the time- and resource-intensive process of model sensitivity analysis. Using EEOFs would provide little insight into *why* a particular model did a better or worse job of simulating the MISO, but the diagnostic value would be helpful in broadly identifying which models (out of a CMIP ensemble, for instance) to focus on. As such, this chapter serves to generally propose the idea of using EEOFs as model comparison tools, especially when it comes to measuring their skill in relation to small-scale phenomena.

Methods

There are many possible ways to go about comparing the predictability of multiple data sets using EEOFs, but for now I only use two simple metrics. One is comparing the first two singular values, which is a metric of how well the first few EEOFs explain the data's variance. A

greater singular value implies that a given EEOF can represent a greater proportion of the observed behavior, theoretically meaning it can be better predicted based on that EEOF alone. Here the singular values are normalized by their sum, meaning all the SVs add to 100% (representing all of the variance).

Second, comparing cross-correlations between the first EEOFs of each model with the observational data can generate a picture of how well the models' EEOFs line up with those of model. More than explaining how much of the variance the EEOF explains, this quantifies how well the model EEOF lines up with that of the observational data. This gives a more robust version of the visual comparison done in Chapter 2, using both the EEOF graphs in Figure 2 and the composite maps in Figure 4. Greater agreement between the data sets would imply that the EEOFs vary on a similar spatial pattern.

Results

All three of the data sets have very broad singular value distributions, meaning the first few EEOFs cannot account for all of the variance on their own. The GPCP's first two EEOFs explain 12.86% and 12.06% of the total variance respectively. For CESM without a Langmuir turbulence parametrization the first two EEOFs explain 9.46% and 7.78% of the variance, while with Langmuir turbulence the first two explain 9.66% and 8.06% (Figure 7). This can be interpreted as showing that MISOs are slightly more predictable in the model with Langmuir turbulence, but the difference is not large enough to be definitive.

The cross correlations of EEOF1 between the models and the observational record are also largely inconclusive. CESM without Langmuir turbulence has a very low covariance coefficient of 0.0006 when compared with observations, with the corresponding value for CESM with Langmuir being 0.0007. Again, the version with Langmuir appears to be slightly more accurate, but the covariance is so negligible as to make the result inconclusive.

Discussion

These two methods for comparing two EEOFs appear to have been largely inconclusive, a negative result which points to the need for different methods of evaluation. It is possible (though unlikely from a dynamics points of view) that Langmuir turbulence has no effect on the behavior of MISOs in the CESM model, but that is nearly impossible to determine using only the two methods shown here.

The extremely low correlation between observational data and either model is hard to interpret, for it is clear that the three data sets are describing similar spatial phenomena. Additionally, the low magnitude of the first two singular values leaves open questions about the robustness of the EEOF analysis in its current form. This is a larger question that will require a more in-depth examination.

Conclusions

Comparing versions of the model with and without Langmuir turbulence has significant value from the perspective of climate mechanisms and model physics. Since the precipitation patterns associated with MISOs form in the Indian Ocean and move North through the Bay of Bengal, this model comparison provides an opportunity to test how important ocean turbulence is to such synoptic scale phenomena.

Clearly more work is needed to find a rigorous way of comparing model results with observational data. One next step is to apply the prediction method from Chapter 3 to both sets of model data. The skill score from that process would indicate how predictable the MISO is *in that model*, which could then be compared with the predictability of the MISO in real life. An important question is whether the MISO has a more or less defined structure in models than in reality—simplifying ocean and atmospheric physics into parametrizations could either lessen the signal of small basin-scale oscillations or exaggerate it.

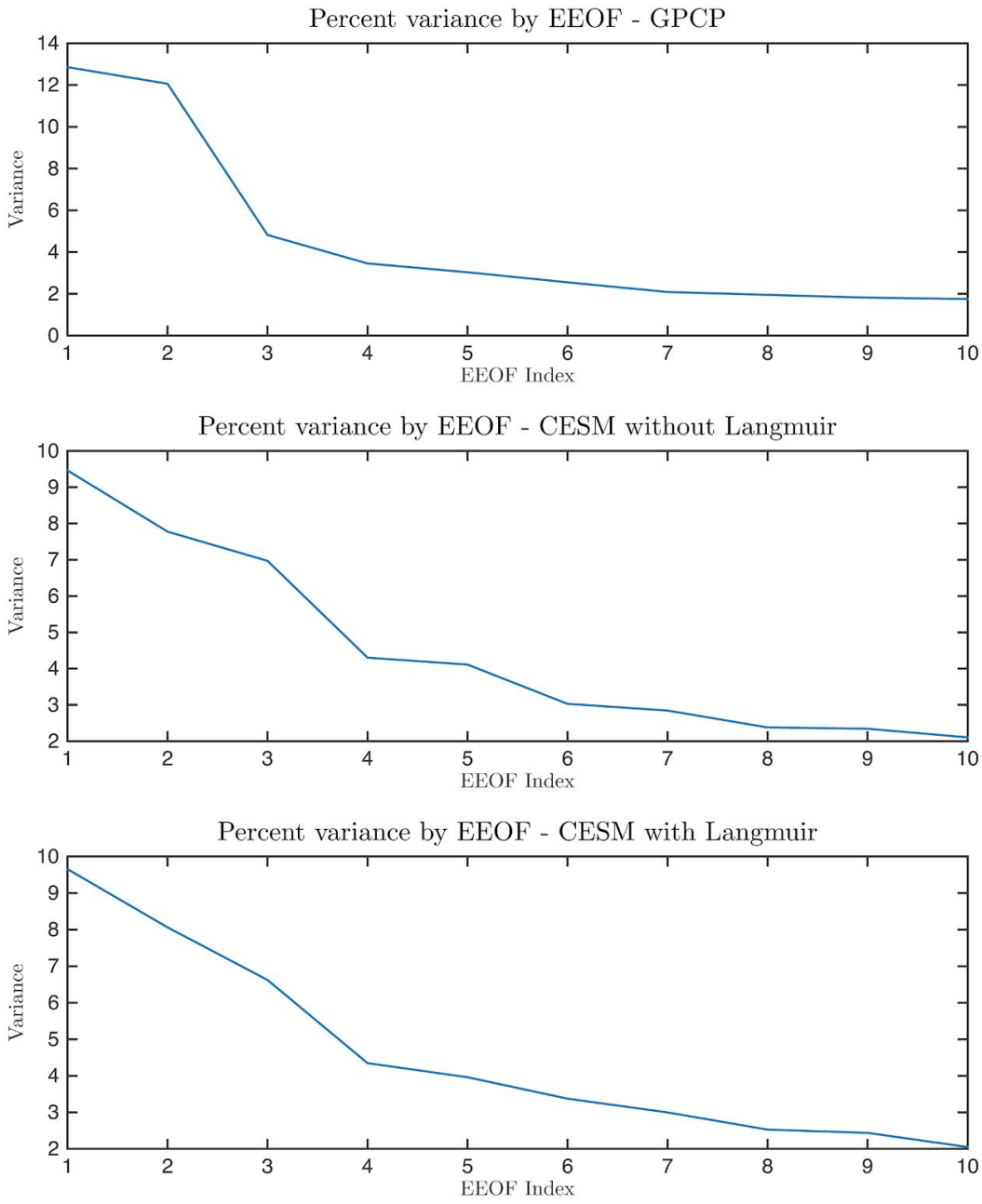

Figure 7: First 20 Singular Values for EEOF decompositions of the three data sets.

Chapter 5: Conclusion

Unlike large climate oscillations like the Atlantic Multidecadal Oscillation or the Indian Ocean Dipole, the MISO's large spatial but small temporal scales make it especially difficult to find in noisy data. Methods like EEOF analysis can help identify such phenomena and define a rigorous way to extract them from observations or climate model outputs. Frameworks like EEOFs also help to limit sampling bias by preventing themselves from seeing patterns in data based on what dynamical processes they are looking for. On the other hand, EEOFs hold the potential to indicate modes of oscillation where there may be none, meaning they have to be applied judiciously and to a limited degree.

In the context of the MISO-BoB project, this work is intended to also provide some small guidance for field observations of MISOs. Their erratic nature, spatial distribution, and irregular periodicity makes them difficult to observe in the field. A research vessel in the Bay of Bengal might have a hard time identifying MISO-like behavior in a given month-long expedition. Similar to using an EEOF with a 16-day lag to evaluate a 40-day oscillation, measurements at one location in the ocean for a limited period of time can show certain aspects of the oscillation, but not a holistic picture. In theory, such an EEOF could be used in reverse as a data assimilation tool, seeing where measured data agrees with what would be predicted for such a time span if a MISO had a given period and spatial distribution. Additionally, this work illustrates how MISOs can be identified from large data sets, opening up the possibility of using satellite data to look at precipitation activity in the month preceding a research cruise and projecting whether a positive phase is imminent.

Based on these results, EEOFs appear to do a reasonably good job of identifying MISOs in different data sets, providing an impartial, dynamics-independent lens with which to view short-term oscillations in noisy data sets. Using them to create a prediction method using truncated singular values, however, will take more work, and settling on a robust measure of model realism is a continuing work in progress. It is clear, though, that empirical methods offer a different approach to evaluating climate data which can be applied to a variety of questions.

Acknowledgments

This research was supported by the US Office of Naval Research through the Monsoon Intra-seasonal Oscillations-Bay of Bengal (MISO-BoB) Directed Research Initiative (DRI), Grant: ONR N00014-17-1-2393.

Additionally, this work owes a great deal to Qing Li, who helped with everything from data processing to code debugging, and many conversations about science and life in between. Critical editing was done by Jacinta Clay and Katherine Orenstein. Discussions with Brodie Pearson, Aakash Sane, Jenna Pearson, Anson Cheung, Abigail Bodner, and Leah Johnson greatly informed the direction of the work at different points and gave me insight into many avenues of study I would never have considered. It owes everything else to Baylor Fox-Kemper.

References

Anand, A., Mishra, S.K., Sahany, S., Bhowmick, M., Rawat, J.S., Dash, S.K., 2018. Indian Summer Monsoon Simulations: Usefulness of Increasing Horizontal Resolution, Manual Tuning, and Semi-Automatic Tuning in Reducing Present-Day Model Biases. Scientific Reports 8, 3522. https://doi.org/10.1038/s41598-018-21865-1

Arnold, N.P., Branson, M., Kuang, Z., Randall, D.A., Tziperman, E., 2015. MJO Intensification with Warming in the Superparameterized CESM. J. Climate 28, 2706–2724. https://doi.org/10.1175/JCLI-D-14-00494.1

Chen, N., Majda, A.J., Sabeerali, C.T., Ajayamohan, R.S., 2018. Predicting Monsoon Intraseasonal Precipitation using a Low-Order Nonlinear Stochastic Model. J. Climate 31, 4403–4427. https://doi.org/10.1175/JCLI-D-17-0411.1

Dommenget, D., Latif, M., 2002. A Cautionary Note on the Interpretation of EOFs. J. Climate 15, 216–225. https://doi.org/10.1175/1520-0442(2002)015<0216:ACNOTI>2.0.CO;2

Fox-Kemper, B., 2004. Wind-driven barotropic gyre II: Effects of eddies and low interior viscosity. J. Mar. Res. 62, 195–232. https://doi.org/info:doi/10.1357/002224004774201690

Goswami, B.N., Ajayamohan, R.S., Xavier, P.K., Sengupta, D., 2003. Clustering of synoptic activity by Indian summer monsoon intraseasonal oscillations. Geophys. Res. Lett. 30, 1431. https://doi.org/10.1029/2002GL016734

Goswami, B.N., Mohan, R.S.A., 2001. Intraseasonal Oscillations and Interannual Variability of the Indian Summer Monsoon. J. Climate 14, 1180–1198. https://doi.org/10.1175/1520-0442(2001)014<1180:IOAIVO>2.0.CO;2

Goswami, B.N., Rao, S.A., Sengupta, D., Chakravorty, S., 2016. Monsoons to Mixing in the Bay of Bengal: Multiscale Air-Sea Interactions and Monsoon Predictability. Oceanography 29, 18–27.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G., 2009. Improving the global precipitation record: GPCP Version 2.1. Geophysical Research Letters 36. https://doi.org/10.1029/2009GL040000

Li, Q., Webb, A., Fox-Kemper, B., Craig, A., Danabasoglu, G., Large, W.G., Vertenstein, M., 2016. Langmuir mixing effects on global climate: WAVEWATCH III in CESM. Ocean Modelling, Waves and coastal, regional and global processes 103, 145–160. https://doi.org/10.1016/j.ocemod.2015.07.020

Sengupta, D., Goswami, B.N., Senan, R., 2001. Coherent intraseasonal oscillations of ocean and atmosphere during the Asian Summer Monsoon. Geophys. Res. Lett. 28, 4127–4130. https://doi.org/10.1029/2001GL013587

Suhas, E., Neena, J.M., Goswami, B.N., 2013. An Indian monsoon intraseasonal oscillations (MISO) index for real time monitoring and forecast verification. Clim Dyn 40, 2605–2616. https://doi.org/10.1007/s00382-012-1462-5

Weare, B.C., Nasstrom, J.S., 1982. Examples of Extended Empirical Orthogonal Function Analyses. Mon. Wea. Rev. 110, 481–485. https://doi.org/10.1175/1520-0493(1982)110<0481:EOEEOF>2.0.CO;2

Xavier, P.K., Goswami, B.N., 2007. An Analog Method for Real-Time Forecasting of Summer Monsoon Subseasonal Variability. Mon. Wea. Rev. 135, 4149–4160. https://doi.org/10.1175/2007MWR1854.1

Addendum: Revised Figure 1