STATISTICAL RELATIONSHIPS OF THE TROPICAL RAINFALL MEASUREMENT MISSION (TRMM) PRECIPITATION AND LARGE-SCALE FLOW

A Thesis

by

KYLE J. BORG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2010

Major Subject: Atmospheric Sciences

STATISTICAL RELATIONSHIPS OF THE TROPICAL RAINFALL MEASUREMENT MISSION (TRMM) PRECIPITATION AND LARGE-SCALE FLOW

A Thesis

by

KYLE J. BORG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Ramalingam Saravanan
Committee Members,	Kenneth Bowman
	Mikyoung Jun
Head of Department,	Kenneth Bowman

May 2010

Major Subject: Atmospheric Sciences

ABSTRACT

Statistical Relationships of the Tropical Rainfall
Measurement Mission (TRMM) Precipitation and Large-Scale
Flow. (May 2010)
Kyle J. Borg, B.A., Austin College
Chair of Advisory Committee: Dr. Ramalingam Saravanan

The relationship between precipitation and large-flow is important to understand and characterize in the climate system. We examine statistical relationships between the Tropical Rainfall Measurement Mission (TRMM) 3B42 gridded precipitation and large-scale flow variables in the Tropics for 2000–2007. These variables include NCEP/NCAR Re-analysis sea surface temperatures (SSTs), vertical temperature profiles, omega, and moist static energy, as well as Atmospheric Infrared Sounder (AIRS) vertical temperatures and QuikSCAT surface divergence. We perform correlation analysis, empirical orthogonal function analysis, and logistic regression analysis on monthly, pentad, daily and near-instantaneous time scales. Logistic regression analysis is able to incorporate the non-linear nature of precipitation in the relationship. Flow variables are interpolated to the 0.25° TRMM 3B42 grid and examined separately for each month to offset the effects of the seasonal cycle.

January correlations of NCEP/NCAR Re-analysis SSTs and TRMM 3B42 precipitation have a coherent area of positive correlations in the Western and Central Tropical Pacific on all time scales. These areas correspond with the South Pacific Convergence Zone (SPCZ) and the Inter Tropical Convergence Zone (ITCZ). 500mb omega is negatively correlated with TRMM 3B42 precipitation across the Tropics on all time scales. QuikSCAT divergence correlations with precipitation have a band of

weak and noisy correlations along the ITCZ on monthly time scales in January. Moist static energy, calculated from NCEP/NCAR Re-analysis has a large area of negative correlations with precipitation in the Central Tropical Pacific on all four time scales.

The first few Empirical Orthogonal Functions (EOFs) of vertical temperature profiles in the Tropical Pacific have similar structure on monthly, pentad, and daily timescales. Logistic regression fit coefficients are large for SST and precipitation in four regions located across the Tropical Pacific. These areas show clear thresholded behavior. Logistic regression results for other variables and precipitation are less clear. The results from SST and precipitation logistic regression analysis indicate the potential usefulness of logistic regression as a non-linear statistic relating precipitation and certain flow variables.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Precipitation, and the associated latent heat release, has a large impact on radiative energy balance and circulation. The majority of global precipitation occurs in the Tropics, see Figure 1. Some favorable conditions in this region include high annual insolation, large amounts of free moisture, convergence, and convective instability. The convergence coincides with the returning branches of the Hadley cell. The location where the North and South branches meets is known as the Inter-Tropical Convergence Zone (ITCZ). This belt shifts North and South of the equator with seasonal insolation changes and has large amounts of deep convection and precipitation associated with it. Some regions along the equator with exceptionally high annual rainfall include South America, Africa, Indonesia and the Pacific Ocean. The seasonal migration of the ITCZ, greater over land than the oceans, leads to dramatic changes in precipitation near subtropical arid regions. On finer temporal scales, the diurnal cycle has immense affect on local precipitation patterns. Diurnal variations over the ocean lead to a precipitation maximum in the late morning, while the maximum over land occurs in the afternoon. A look at smaller spatial scales finds that orographic effects, land-sea contrasts, atmosphere-ocean interactions, and atmospheric waves all have considerable effect on precipitation. The dependencies mentioned above are but a few of many that make precipitation vary greatly due to local conditions.

The complex nature of precipitation leads to a major problem for observational and modeling studies that try to characterize precipitation. Observational studies are

---

The journal model is *Journal of Climate*.

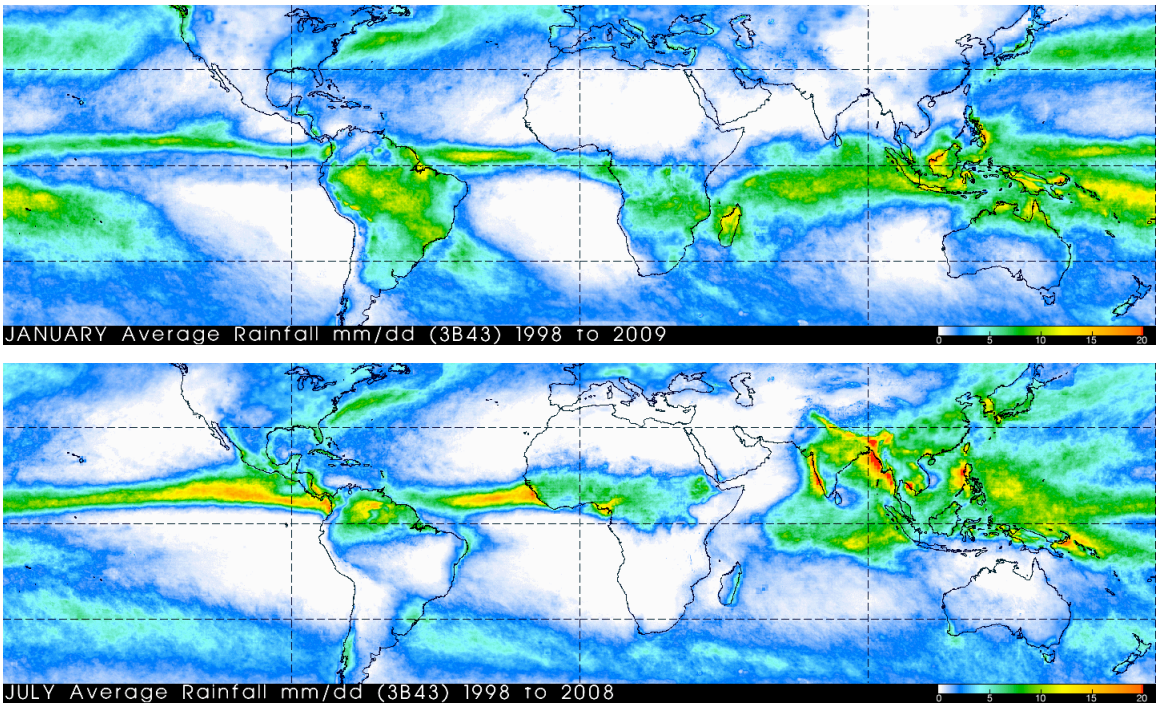

Fig. 1. Top: January climatological TRMM 3B43 precipitation data from 1998–2008. Botom: July climatological TRMM 3B43 precipitation data from 1998–2008. Data is on a 0.25° grid from 40°S to 40°N. Image generated by NASA.

plagued with sampling and accuracy issues. Prior to the satellite era, precipitation was measured primarily via rain-gauge data. Rain gauges are essentially a bucket that collects rain and takes periodic measurements of the amount of water in the bucket. These are, potentially, high resolution temporal estimates of precipitation at a single point. Thus, it was possible to obtain a good understanding of the properties of precipitation for that location, but harder to extrapolate the findings to larger areas. Over the continents, especially in populated areas, large networks of rain-gauges were constructed to help alleviate this problem. Measurements over the oceans were barely existent, relying on a few buoy and ship observations.

The advent of observation systems placed aboard orbiting satellites made it possi-

ble to obtain precipitation estimates across the globe in areas where it had previously been inaccessible. This has had great scientific and societal impacts. Tropical storm detection and warning systems are now in place that have greatly alleviated the societal impacts when these storms make landfall. Total coverage over the oceans has also provided a wealth of knowledge of the general circulation and energy budgets in these areas. While satellite measurement of atmospheric circulation has greatly increased the areal coverage of observations, it suffers from coarse temporal resolution and accuracy issues. Temporal sampling at a location is reliant on the orbital parameters of the satellite and accuracy issues are raised because they observe precipitation indirectly. Originally, infrared emission (IR) from cloud tops was used to infer rain rates. Passive microwave sensors provided an improvement to this method. These instruments aboard satellites measure a volume integrated microwave emission, from which a surface rain rate may be deduced. Passive microwave sensors provide more accurate estimates than IR measurements, but are still less than perfect. They work particularly well over oceans where the surface appears as a uniform cold background which contrasts well with raining areas, but have trouble over land where surface emissivity values are large and can vary dramatically. The newest development, in precipitation measurement, is space-based precipitation radars which can provide direct information on the vertical distribution of precipitation. The first of this kind is aboard the Tropical Rainfall Measuring Mission (Simpson et. al., 1998), referred to as TRMM for the duration of this paper. For a better understanding of the differences between satellite and rain-gauge measurements see Bowman (2004). It provides a comparison of satellite observations from TRMM and a network of ocean buoy rain-gauges.

Precipitation has an inherently non-linear nature; leading to thresholds and

clearly non-Gaussian behavior. The combination of non-Gaussian behavior and poor observational coverage make statistical analysis difficult. Typically precipitation data is averaged to larger spatial and temporal scales, resulting in more Gaussian behavior. Some previous studies examine temperature and precipitation statistics over the United States for seasonal (Madden and Williams, 1978) and monthly (Zhao and

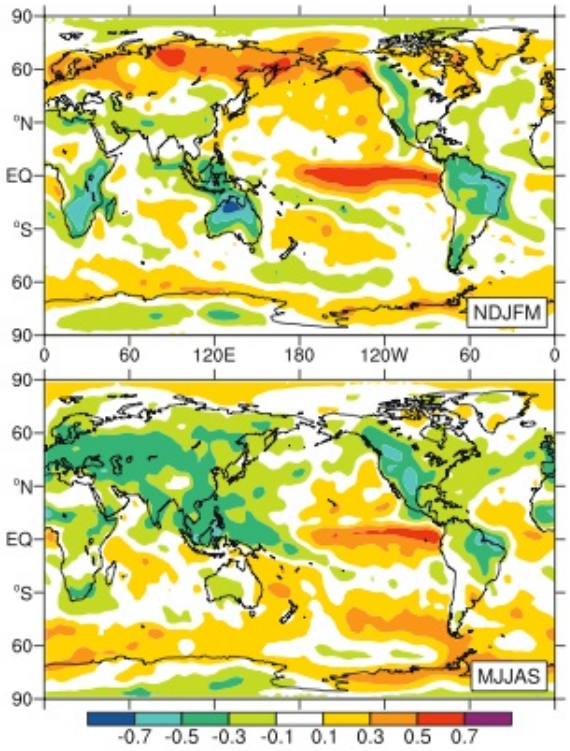

Fig. 2. Correlation between observed monthly mean anomalies of surface air temperature and precipitation amount for groups of months of NDJFM and MJJAS for 1979 – 2002. (Trenberth and Shea, 2005)

Khalil, 1993) mean data. More recent publications use large new data sets, but retain the convention of examining averaged data. Trenberth and Shea (2005) correlate surface temperatures, from the ERA-40 Reanalysis data, and precipitation from the Global Precipitation Climatology Project version 2. Some results are presented in Figure 2. The highest correlations are found in the ENSO region, and there are strong

negative correlations over land during the summer in both hemispheres. There is a coherent region of negative correlation over the maritime continent. Bretherton et al. (2004) use daily and monthly SSM/I water vapor pressure and precipitation, on a 2.5° grid, to calculate correlations over the tropical oceans for the years 1998 – 2001. Most recently, Back and Bretherton (2008a) create a simple thresholded linear model relating precipitation from the Global Precipitation Climatology Project (GPCP) and the European Center for Medium-range Weather Forecasts 40 year Re-analysis (ERA-40) sea surface temperatures (SSTs). It is referred to as a null-hypothesis model in this case, because it constitutes the simplest relationship for rain over the tropics. Namely that rain will occur when the SST is greater than some threshold. The simple model is constructed with a threshold of 298K, determined from a non-linear optimization function in MATLAB. This model predicts a wider and less intense ITCZ than is seen in observations. In order to try and explain this discrepancy Principal Component Analysis (PCA), sometimes referred to as Empirical Orthogonal Function (EOF) analysis, is performed on the vertical profiles for convergence and vertical motion. Linear combinations of the first two eigenvectors from the EOF analysis of vertical motion yield a profile associated with a shallow mode in the Eastern Pacific, and a profile representing a deep mode in the Western Pacific. Figure 3 shows the original eigenvectors and linear combinations. Once these modes are identified the paper expands to a more complex model based upon the modes for convergence.

In many ways the the two mode model that is suggested in Back and Bretherton (2008a) is a derivative from previous studies characterizing convection and precipitation in the Tropics. In light of this fact, we will take a moment to examine some of these studies. A popular line of inquiry is to attempt to diagnose the relationship of SSTs and atmospheric conditions that lead to the onset of deep convection (Fu

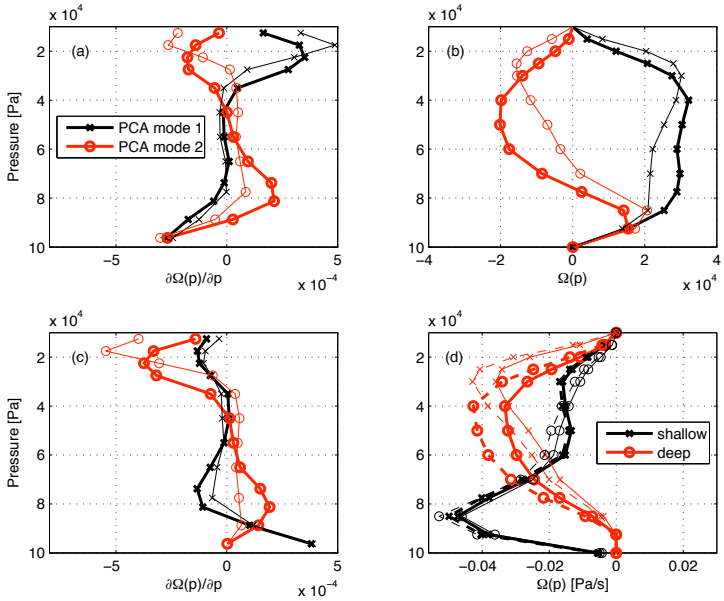

Fig. 3. Empirically-derived horizontal convergence and vertical motion modes from ERA-40 (thick lines) and NCEP (thin lines). Left panels show horizontal convergence profiles, right panels corresponding vertical motion profiles. Top panels (a) and (b) show modes as estimated PCA of horizontal convergence; bottom panels (c) and (d) show the linear combinations of these modes used in subsequent analysis (Back and Bretherton 2008a).

et. al., 1993; Zhang, 1993; Sherwood, 1998; Bielli and Hartmann, 2004). A general result from these studies is that there is large spatial variability shown throughout the Tropics. Zhang (1993), Fu et. al. (1993), and Sherwood (1998) show that in the Western Pacific warm pool high SST anomalies create a benevolent environment for deep convection. Due to this environment onset of deep convection is frequent and dependent on atmospheric disturbances for initiation.

Sherwood (1998) finds that low to mid-tropospheric moisture is the dominant factor controlling convective outbreak in this area. One interesting approach taken in this paper is to examine the relationship between circulation variables and convec-

tion with a non-linear statistical model. It assumes a simple conjunctive relationship. Conjunctive hypothesis tests were performed on the data. This consists of testing for thresholds in each predictor variable that will maximize the likelihood for convection to occur. An interesting result from this analysis is that convection is found 86% of the time above a relatively low CAPE value of 300 J/kg. While this result is interesting and a validation for using non-linear models in regards to CAPE, the results were less successful for other variables such as CIN and RH. This paper goes on to suggest that in order to capture spontaneous convective outbreak finer horizontal scales for the data is vital. Results in the Eastern Tropical Pacific, where lower SSTs are prevalent, are found to be very different. Fu et. al. (1993) find strong convergence is necessary to initiate deep convection in this area. This result is supported by Back and Bretherton (2008a). As described above, it finds a simple thresholded model based on SST predicts a wider and less intense precipitation band in the Eastern Pacific. It is shown that this belt where rainfall is under-predicted corresponds with an area of high surface convergence.

Another paper by Back and Bretherton (2008b), provides evidence that the climatological distribution of convergence may be regarded as a cause of deep convection, rather than an effect. This is a foundation in the accompanying paper, Back and Bretherton (a) mentioned above, that breaks convergence in the Tropical Pacific into two modes in an attempt to model precipitation. Sobel and Neelin (2006) conducted a modeling study attempting to diagnose contributions of boundary layer and free atmospheric thermodynamic conditions to precipitation. They found that the boundary layer is an important part of providing a narrow, intense ITCZ. This appears to coincide with the results from Back and Bretherton (2008b), reinforcing the validity of the assumptions made in their companion paper, Back and Bretherton

(2008a).

We plan on extending the conceptual foundations of the papers outlined above to search for statistical relationships between precipitation and the atmospheric circulation on finer spatial scales. We will examine how these relationships change for different time scales. We will examine the relationships between surface temperatures and mid-tropospheric vertical winds, from re-analysis data sets, and satellite precipitation. Relationships for satellite surface convergence and instability will also be explored. Empirical orthogonal functions of vertical temperature, from satellite retrievals, will be used as a proxy for convective instability. In light of the results from the non-linear conjunctive model of Sherwood (1998), we will also apply a popular statistical tool, logistic regression, to this problem. Logistic regression is a commonly used non-linear technique that is frequently used in genetics and opinion polling. This technique has not found widespread use in atmospheric research. This regression may be applied to any of the variable stated above to search for thresholded behavior. The next chapter describes the data and methods of analysis that we use. The third chapter explores some simple statistical measures between satellite recorded precipitation and the general circulation. The fourth chapter covers empirical orthogonal function analysis of vertical temperature profiles. The fifth chapter presents the results of logistic regression analysis. The final chapter contains a discussion and conclusions.

CHAPTER II

DATA AND METHOD

A. Introduction

The data we use includes Tropical Rainfall Measuring Mission (TRMM) 3B42 precipitation, several NCEP/NCAR Re-analysis fields, AIRS vertical temperature profiles and QuikSCAT surface wind retrievals. These data sources vary from providing more sparse and precise measurements (AIRS, TRMM and QuikSCAT) to very complete, less precise, fields (TRMM 3B42 and NCEP/NCAR Re-analysis). Correlation and empirical orthogonal function (EOF) analysis of the more complete fields may provide useful, simple, information of large scale patterns, but may also be misleading due to the imprecise nature of the measurements. Similar analysis with more precise data may act as a check on large scale patterns from the larger fields and also provide more precise local relationships. Logistic regression analysis is a more novel way of finding a relationship between atmospheric variables that has not found widespread use as of the present. It is very common in the broader statistics community and provides information in regards to the relationship between a binomial, or multi-nomial) response variable and a continuous indicator variable. This may provide interesting results in regards to precipitation and atmospheric flow variables. Now we present the data and methods mentioned above in more detail.

B. Data

1. Precipitation

The Tropical Rainfall Measuring Mission (Simpson et. al., 1998) was launched in November 1997 to provide detailed precipitation estimates over previously under-

sampled areas in the tropics. TRMM has several instruments aboard including a microwave imager (TMI) and a precipitation radar (PR) that provide high precision estimates of rainfall. The 3B42 data product consists of a 0.25° spatial grid, from 50S to 50N and spanning all longitudes with a temporal resolution of three hours. Only about a quarter of each year's data missing due to coverage and retrieval errors. This product allows for near instantaneous relationships with up to eight measurements at a site per day, sampling across the diurnal cycle. TRMM 3B42 is a combination of microwave precipitation estimates from TRMM, Special Sensor Microwave/Imager (SSM/I) , the Advanced Microwave Scanning Radiometer (AMSR), the Advanced Microwave Sounding Unit (AMSU) and geostationary IR estimates. In a given grid box, 0.25° horizontal coverage and a three hourly temporal window, any microwave measurements are taken as truth. These microwave measurements have an implicit 15 minute temporal averaging, hence the terminology 'near instantaneous' used above. If no microwave estimates are present in the space time window for the grid box then the IR estimates from the geostationary satellite observations are used. All TRMM products are distributed by the Goddard Distributed Active Archive Center (GES DISC DAAC).

2. Re-analysis

The National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) Re-analysis (Kalnay, 1996) is a complete set of atmospheric flow variables covering the entire globe with a 2.5° grid at 17 vertical levels every six hours. This reanalysis covers the time period from 1948 to the present. The re-analysis uses numerous data sources along with a data assimilation scheme and a model to create global estimates of all variables. All variables are classified by

the amount of observational data that is incorporated into their estimation. Variables classified as A are well defined by observations. B classification variables are partially defined by observations but strongly influenced by the model. The C classification signifies variables that are not assimilated and estimates come solely from the model. Examples of A classifications variables are zonal and meridional winds, geopotential height, and absolute vorticity. Vertical velocity and specific humidity are classified B, and precipitation is a C classified variable. The ERA-40 Re-analysis (Uppala et. al. 2005) data set is another commonly used re-analysis data set that possesses the same space-time grid as NCEP/NCAR, but covers the time period from mid-1957 to 2001. The NCEP/NCAR Re-analysis data is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. The primary distribution point of ERA-40 Re-analysis in the United States is by the Computation Information and Systems Laboratory (CISL) at NCAR.

3. Other Satellite Data

NASA's Quick Scatterometer (QuikSCAT) was launched in June 1999. The SeaWinds instrument on QuikSCAT mission was originally designed to be a "quick recovery" mission to fill the gap created by the loss of data from the NASA Scatterometer (NSCAT). SeaWinds has exceeded the original mission lifetime of three years and continues to provide ocean surface wind data today. SeaWinds measures backscatter from the ocean surface and then derives surface wind stress, as well as wind speed and direction. The SeaWinds instrument consists of a 13.4 GHz radar and an antenna. The antenna is a one-meter-in-diameter rotating dish that sweeps a circular pattern. SeaWinds produces a 1,800-kilometer-wide swath each orbit and covers 90% of Earth's oceans each day. It can measure wind speeds from 3 – 20 m/s with an

accuracy of 2 m/s and has a horizontal wind vector resolution of 25 km. SeaWinds data is distributed by NASA's Jet Propulsion Lab's (JPL) Physical Oceanography Distributed Active Archive Center (PO.DAAC).

Launched in May 2002, the Atmospheric Infrared Sounder, AIRS, is one of six instruments on board the Aqua satellite. Aqua occupies a polar, Sun-synchronous orbit with a period of 98.9 minutes and altitude of 803 km. It is part of NASA's "A-train", a series of high-inclination, Sun-synchronous satellites in low Earth orbit designed to make long-term global observations of the climate. AIRS provides twice daily coverage with a swath width of 1650 km. In low latitudes some locations are missed due to non-overlaping swaths. These points are usually scanned within 2 – 3 days. AIRS provides estimates of the vertical structure of temperature, water vapor as well as measuring clouds, abundances of trace components in the atmosphere including ozone, carbon monoxide, carbon dioxide, methane, and sulfur dioxide. AIRS has a high spectral resolution, 2378 spectral channels, supplying very accurate temperature and moisture profiles. When used in conjunction with microwave instruments aboard aqua AIRS can provide measurements even during cloudy conditions. AIRS data is available through the GES DISC DAAC.

C. Methods

1. Correlation Analysis

To carry out our statistical analysis we interpolate NCEP/NCAR data in space and time to the TRMM grid. This provides a simple means for pairwise deletion of missing data points. Pearson correlation coefficients are calculated for three hourly, daily,

pentad (five day) and monthly averaged time series. This range of time scales may reveal any sensitivities of the correlation coefficients due to the time averaging. A significance value is calculated for each correlation coefficient. This is done in the standard way, via a students t-test at the 99% level. Initial comparisons of results from this and the more common 95% level show little difference. In light of this, we choose to use the more stringent 99% level of confidence. These calculations are done, separately, for each point on the spatial grid. The seasonal cycle has a dominating affect on correlations of atmospheric variables. We are not interested in the seasonal cycle, therefore we calculate correlations for one month at a time. This method samples along small parts of the seasonal cycle and minimizes its impact.

2. Empirical Orthogonal Function Analysis

The structure of a vertical temperature profile gives an indication of the convective instability at that location. As a result, temperature profiles can have a strong influence on precipitation. To examine this relationship, we calculate empirical orthogonal functions (EOFs) of the vertical temperature profile over the Tropical Pacific(150W - 80W, 15S - 15N) for the time period 1999 – 2007. Then we correlate the principal components of the eigenvectors with precipitation. This is done from the surface up to 100 hPa. Due to irregular spacing of the pressure grid, we weight the data by the square root of the layer thickness centered on each level. Un-weighted EOFs are also calculated for comparison. No significant differences in structure were found between the the weighted and un-weighted EOFs. The principal component time series are found by projecting the weighted data onto the EOFs, then correlation coefficients with precipitation are calculated. To create the EOFs we use all spatial and temporal data at each level to calculate a single covariance matrix, C, that relates the vertical

levels for all the data in our domain, then use this to solve the eigenvector problem;

$$CE = E\Lambda \tag{2.1}$$

where Λ is a diagonal matrix containing the eigenvalues λ_i of C. E is a matrix with column vectors $\vec{e_i}$ that are eigenvectors of C corresponding to eigenvalues λ_i. The principal component time series $\vec{a_i}$ of an EOF is found by projecting the data, R, onto the eigenvector;

$$\vec{a_i} = R\vec{e_i} \tag{2.2}$$

These principal component time series for each eigenvector are then processed in the same manner as described above to obtain correlations with rainfall.

3. Logistic Regression

A new way to examine the relationships of flow variables and precipitation is the logistic regression. This is a common statistical method often used in genetics and opinion polling for relationships between one or more explanatory variables and a discrete response variable. First the binomial probabilities of an event occurring are calculated, then the log of these odds (logits) are modeled as a linear function with respect to the independent variable. Mathematically this is shown by first considering the logistic response function, $\pi(x)$, which is simply the probability of success for an event, Y, conditioned on the independent variable X_i:

$$\begin{aligned} \pi(x) &= Pr(Y = 1|X_1 = x_1, ..., X_p = x_p) \\ &= \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+...+\beta_p x_p}}{1 + e^{\beta_0+\beta_1 x_1+\beta_2 x_2+...+\beta_p x_p}} \end{aligned} \tag{2.3}$$

where the second line is the expanded form of the conditional probabilities. We then use the logit function:

$$ln\frac{\pi(x)}{1-\pi(x)} \tag{2.4}$$

to obtain transformed conditional probabilities, $\pi'(x)$, that are linear with respect to the independent variable X_i:

$$\pi'(x) \;=\; ln\frac{\pi(x)}{1-\pi(x)} \;=\; \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p \tag{2.5}$$

and the parameters β_i describe the fit line.

The coefficients must be estimated with a computational maximum likelihood estimation. The slope parameters describe the change in the log-probability of success for the outcome variable per unit change of the explanatory variable.

Now we walk through the process of how we calculate logistic regression parameters for our data. First we gather the precipitation and indicator variable, SST for this example, for the specified region we are interested in. The regions we choose for this work are outlined in more detail in chapter V. After we have gathered the data we convert the precipitation into a binomial time series. This is accomplished by assigning precipitation rates greater or equal to 1 mm/day a value of one. Rates less than 1 mm/day are given a value of zero. Next, we do pairwise deletion of missing data points and reform the binomial precipitation and continuous SST data from a three dimensional array into one single vector each. The SST data is not treated any further than the deletion of data points that correspond with missing precipitation data. These two vectors are then fed into the IDL function NR_LOGISTIC, adapted to idl from a matlab script (NR_LOGISTIC) by John Burkett from the University of

Rhode Island. This function returns fit parameters, p-values of the fit parameters, and a fit line based on these parameters. We then present the fit parameters in a table and plot the fit line. We finish by over plotting probabilities calculated from the data that is fed into NR_LOGISTIC. These probabilities are found by separating the SST data into 60 bins, summing the corresponding binomial precipitation data, and dividing by the number of data points in that bin. The number of bins (60) is subjectively decided. This last step is simply to aid visualization for the figure and does not arise from the logistic regression calculation. The figure on page 52 shows several of these plots, please refer back to this description for aid in understanding what they show, or there is another description of the process in the beginning of chapter V.

CHAPTER III

CORRELATION ANALYSIS

A. Introduction

This chapter describes the results of correlation analysis between TRMM 3B42 precipitation and flow variables from re-analysis and satellite data sets. We calculate a correlation coefficient for the precipitation and flow variable time series at each spatial point on the TRMM 3B42 grid for monthly, pentad, daily, and three hourly time averages. The results are presented as maps of the correlation coefficients and scatter plots of the raw data from some example locations. Contoured areas are significant at the 99% level as determined by a t-test. First we start with a verification of the results presented in Trenberth and Shea (2005). Then we proceed to examine correlations between TRMM precipitation and NCEP/NCAR flow variables including surface temperature and 500mb omega. These environmental properties are a common theme when trying to associate precipitation with large scale flow, as we saw in the introduction. We also calculate Moist Static Energy from the NCEP/NCAR re-analysis for correlations with precipitation. This follows in the footsteps of Back and Bretherton (2006) and Neelin and Held (1988), who have previously examined relationships between Moist Static Energy and atmospheric convection. In an attempt to use more direct atmospheric observations than re-analysis, we examine correlations between Quikscat divergence and TRMM precipitation. We also provide results from correlations of Quikscat surface divergence and NCEP/NCAR surface temperatures. All of the analysis is performed on multiple time scales, the finest being near instantaneous, which has not been presented in any of the papers considered earlier.

B. Re-analysis

1. Surface Temp

Initial comparisons of correlations between TRMM 3B42 precipitation and the NCEP/NCAR and ERA-40 surface temperatures from the same time period yield maps with similar structure across the tropics. For the rest of the analysis we use the NCEP/NCAR re-analysis alone due to the longer period of overlap with TRMM precipitation data. Figure 4 shows NCEP/NCAR correlation maps, for 1999 – 2007, calculated similar to Trenberth and Shea (2005). First monthly averages are computed, then the winter months (NDJFM) and summer months (MJJAS) are collected. Next, anomalies from the mean are found for each season (winter and summer) and these anomalies are correlated with precipitation. See Figure 2 for a comparison with that paper's results. The winter map we calculate does not show the large area of positive correlations in the Central and Eastern Pacific found in Trenberth and Shea (2005). The main reason for this discrepancy is lack of a strong El Nino or La Nina in our data period. Figure 6 shows a stitched together time series of SST for the month of January. We see a large change in SST values from 1998 to 1999 due to the large El Nino in 1997/1998. As explained below we purposefully discard the one year with large a El Nino signal. The Trenberth and Shea (2005) calculation has a time period extending from 1979 to 2001, allowing for more ENSO events to be in the data record than our nine years.

To obtain familiarity with the data, Figures 5 and 6 show the raw time series of three hourly rain rate and surface temperature for the month of January. These plots are created by stitching together the time series from January for the years 1998 – 2007. Missing data are assigned a value of zero anomaly for the rain rate

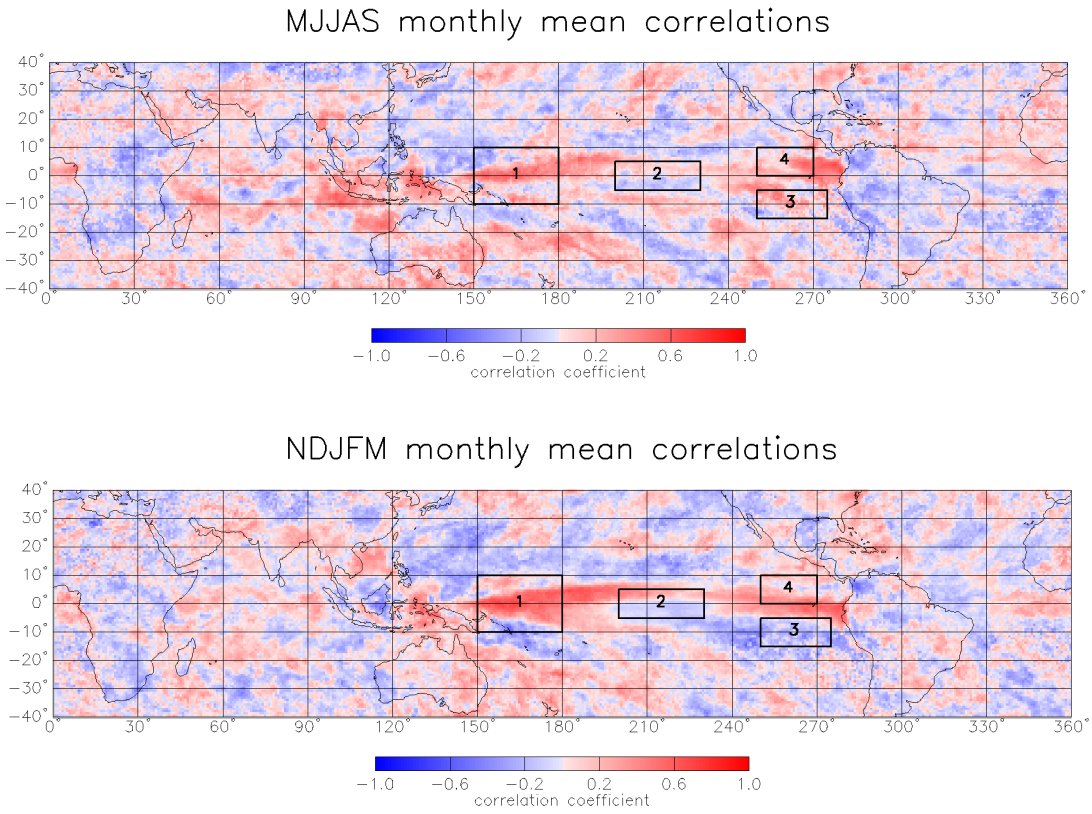

Fig. 4. Correlation between observed monthly mean anomalies of NCEP skin temperature and precipitation amount for groups of months of NDJFM and MJJAS for 1998 – 2007. The four regions used for later logistic regression analysis are outlined and numbered.

plots. The rain rate time series are mostly missing or at low values with extremes scattered throughout. The surface temperature time series demonstrate much less variability than the precipitation rates. There is a noticeable jump in several of the temperature time series between 1998 and 1999, which is probably associated with the strong 1997/1998 El Nino. We are not interested in the effects of a single large El Nino, so the rest of the maps are calculated from 1999 – 2007.

The monthly averaged data for January, seen in Figure 7, has a large region of positive correlations spanning the Western and Central Pacific ocean. The band above the Equator corresponds with the seasonal location of the ITCZ. Positive correlations are prevalent in the central Indian ocean, and over parts of the Maritime Continent. There is a band of negative correlations in the South Central and South Eastern Pacific centered just below the Equator. As the correlations switch to finer timescales spatial structures stay similar to the monthly averaged data. The correlation values decrease on finer timescales. The areas of positive correlations in the extra-tropical oceans increase. Correlations for July show narrow bands of correlations in the Central and South Central Pacific. There are no outstanding spatial structures, unlike in the January data. Finer time scales show smaller correlation coefficients. Patterns are still similar across the different temporal averaging time periods.

Figure 8 shows scatter plots for monthly, pentad, daily and near instantaneous data at a point in the Western Pacific (0N,160E). This location was chosen to show the non-linear nature of the data as well as examine a location with 'good' correlations. There is a major clustering of data points near zero rain rates with relatively few at higher values. January scatter plots have a clearly positive relationship and a strong outlier at each time interval. The daily and near instant time scales are

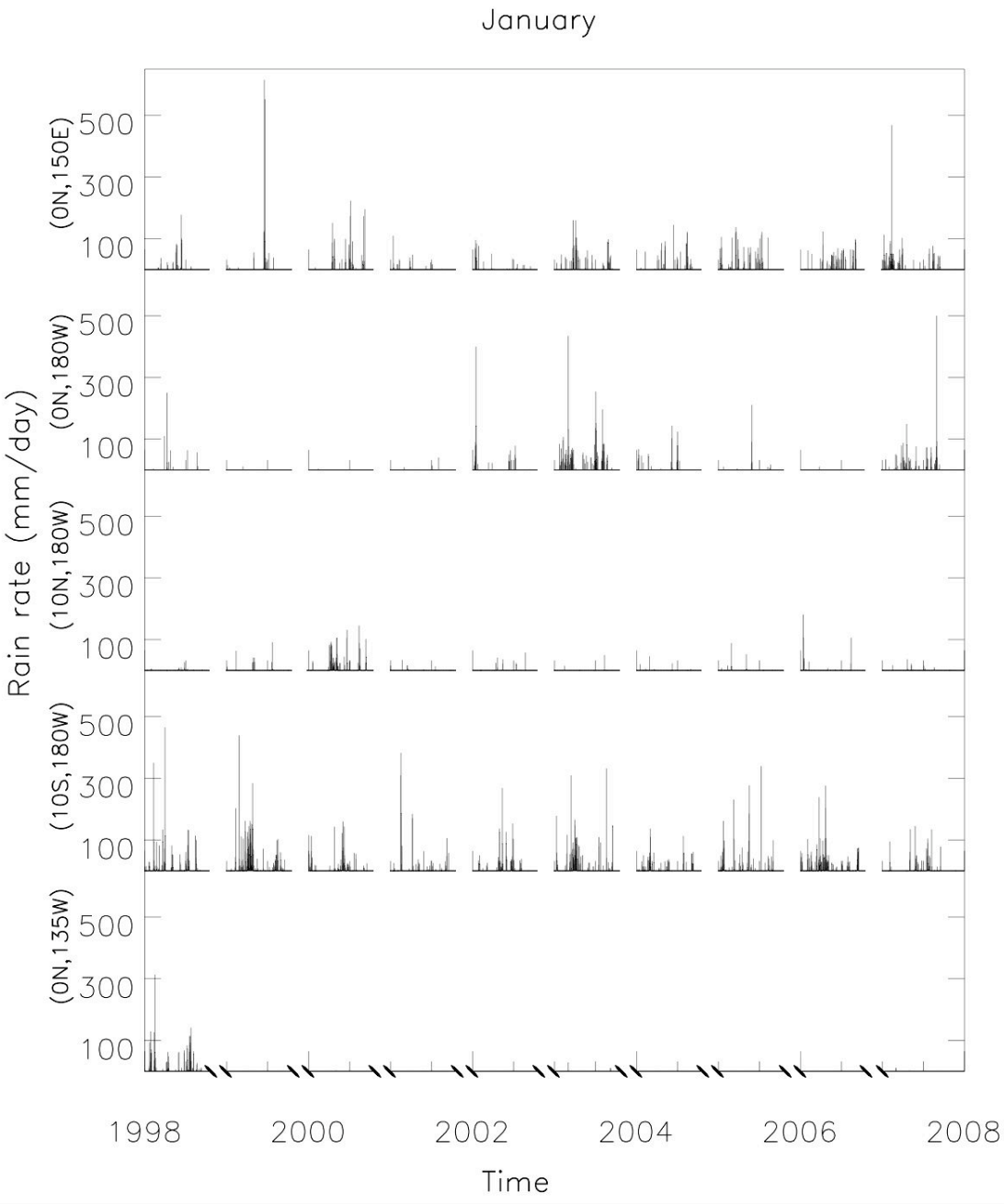

Fig. 5. Three hourly TRMM 3B42 rain rate anomaly time series for the month of January. Missing data are assigned a value of zero.

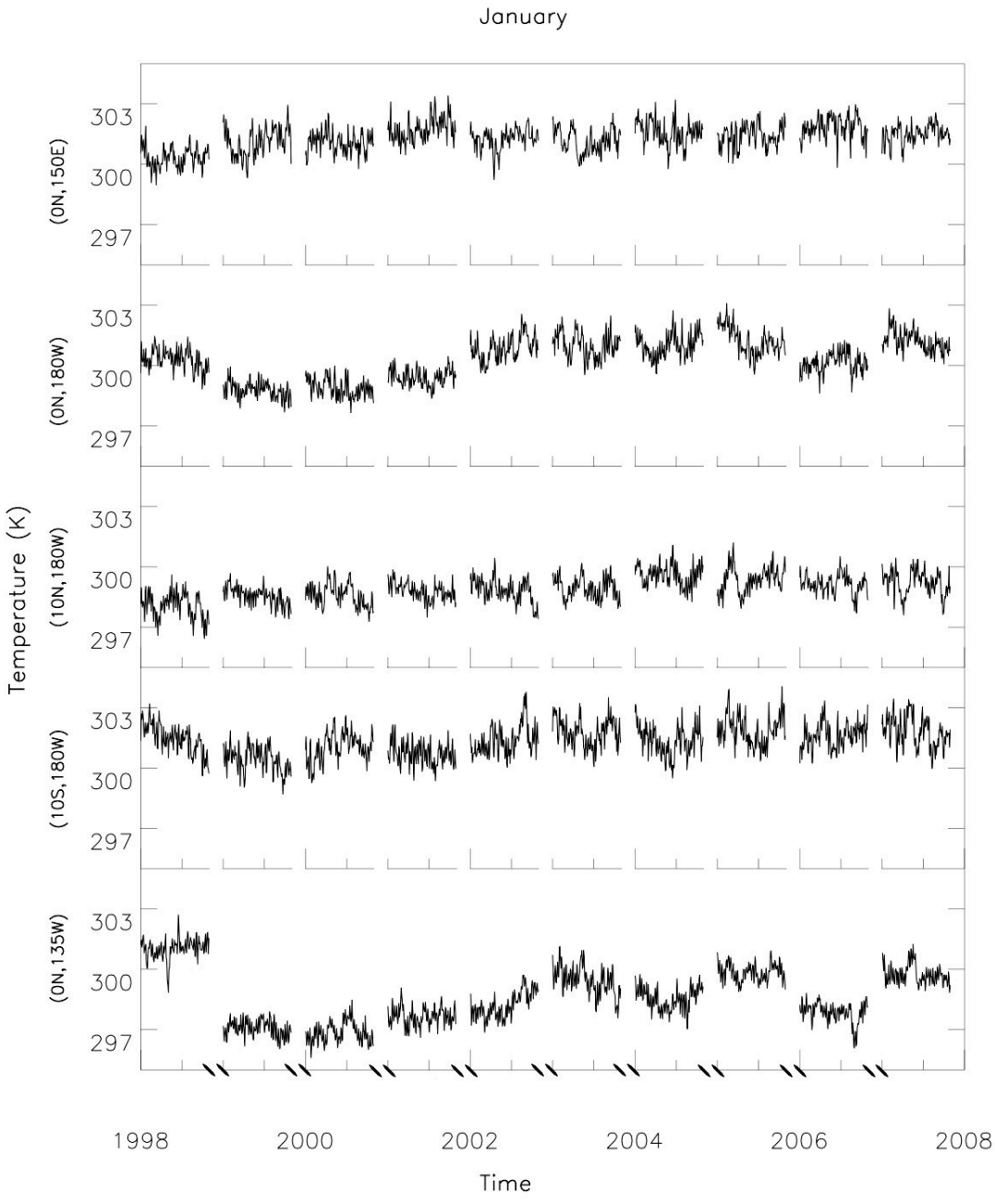

Fig. 6. Three hourly NCEP surface temperature anomaly time series for the month of January.

dominated by low and zero rain rates. A few higher rain rates do appear with some consistency at higher SST values. The monthly July scatter plot shows an increase in precipitation as SSTs grow larger. This pattern holds on the finer time scales. The January maximum rain rates for this location are consistently higher than those in July. From these scatter plots we can infer that positive correlations correspond with hot and wet or cold and dry conditions. Conversely, negative correlations correspond with cold and wet or hot and dry conditions. These negative correlations are most likely due to surface temperatures being driven by precipitation.

2. 500mb Omega

Figure 9 shows correlation maps for NCEP 500mb omega and precipitation. The monthly map for January shows strong negative correlations, negative omega indicates upward motion, in the Central Pacific. Negative correlations dominate in the oceans except for areas where there is descent of the Hadley cell, such as the large area off the West coast of South America. Over land negative correlations are persistent on the eastern coasts of North and South America, Africa and Asia. On finer time scales correlation values decrease in magnitude. Areas with positive correlation coefficients fail to be statistically significant at the 99% level. The negative correlations in these maps indicate upward motion of air at 500mb is associated with precipitation. The positive correlations imply downward motion of air is associated with rainfall. The positive correlations are dominated by extreme values and disappear on finer timescales as the number of observations increase.

The monthly and pentad averaged data for July have similar patterns. Strong negative correlations in the Western and Northeast Pacific. The Central Pacific has

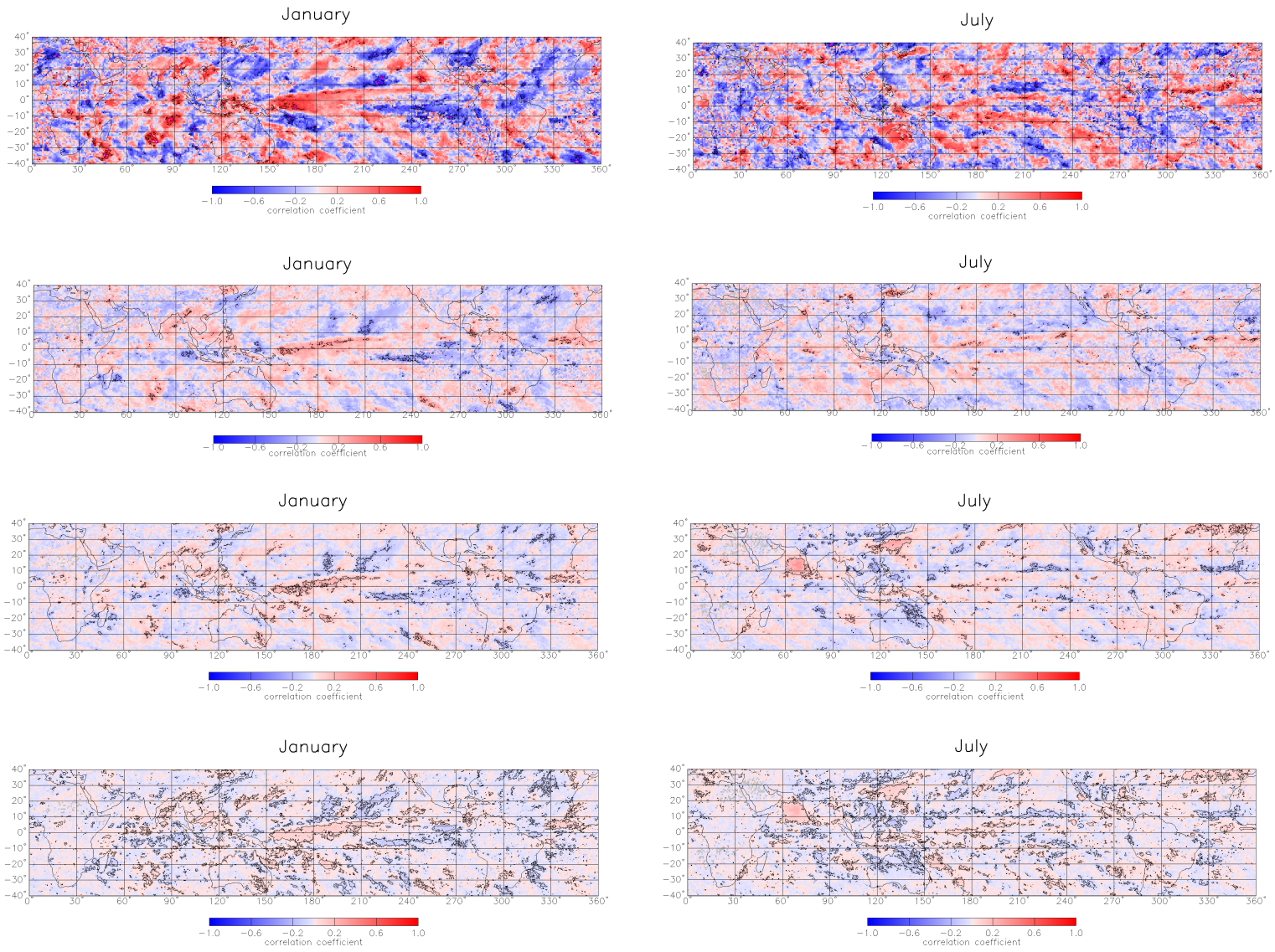

Fig. 7. Top: Monthly 3B42 and NCEP surface temperature correlations. 2nd row: Pentad correlations. 3rd row: Daily correlations. Bottom: near instant correlations.

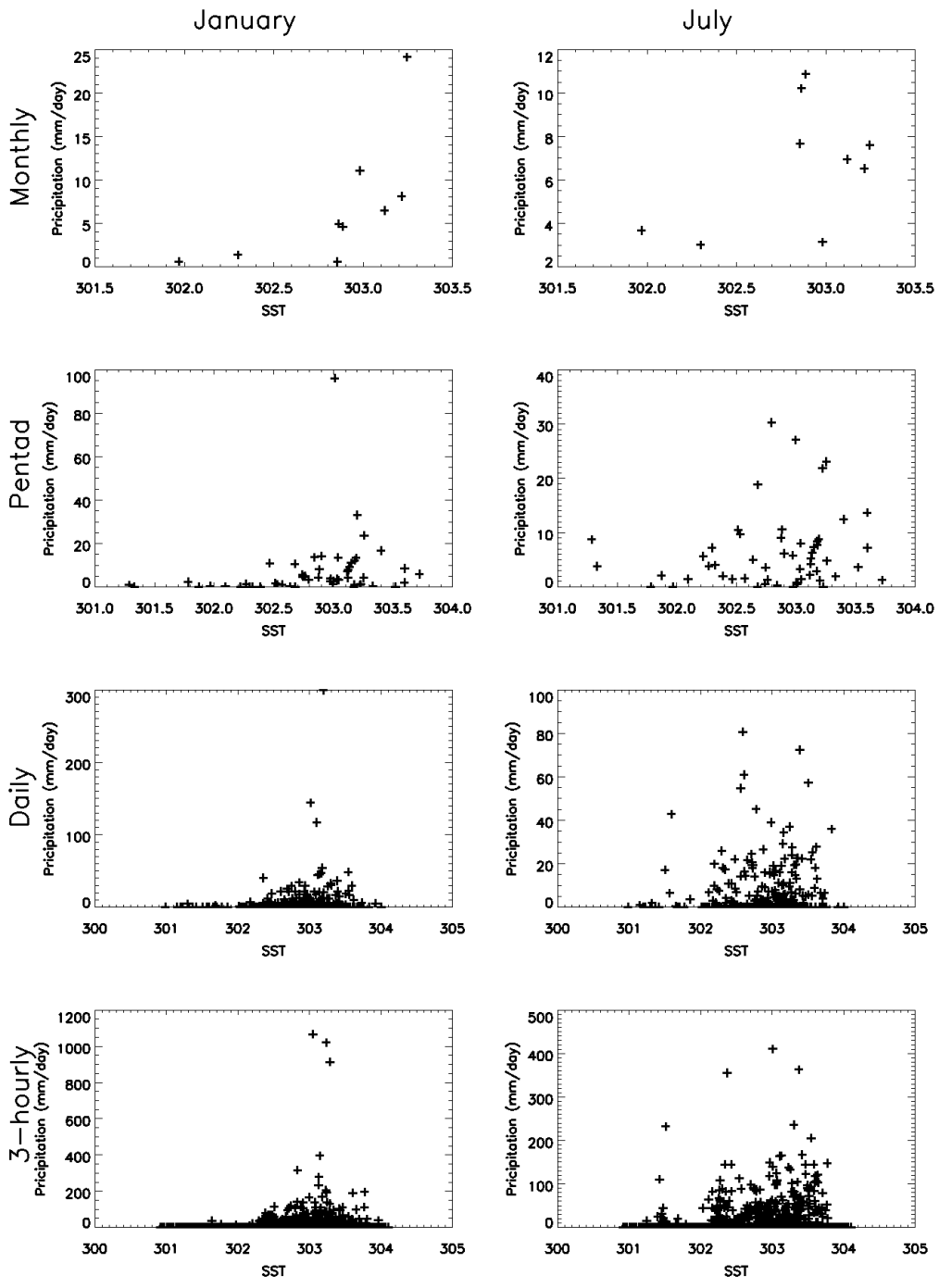

Fig. 8. Monthly, pentad, daily and 3-hourly 3B42 and NCEP sea surface temperature scatter plots for two months at a location (0N,160E) in the Pacific.

a less well defined area of negative correlations as compared to the January data. There are also regions of negative correlations in the extra-tropical Pacific. Positive correlation coefficients are found in areas with little rainfall, refer back to Figure 1 for the distribution of rainfall. The daily and near-instantaneous data do not show strong negative correlations or coherent patterns similar to the monthly and pentad averaged data.

Figure 10 shows scatter plots for monthly, pentad, and daily averaged data at the same Western Pacific location (0N,160E) as the SST scatter plots above. The monthly averaged data for both January and July show a decrease in precipitation as omega becomes less negative. The January data maintains this pattern on the finer time scales. These scatter plots have larger outliers than the July data. The finer time scales for July indicate an increase in precipitation for less negative values of omega. Non-zero rain rates are clustered around an omega value slightly less than zero.

3. Dry and Moist Static Energy

We calculate an integrated moist static energy from the NCEP/NCAR re-analysis. The moist static energy (mse) :

$$mse = \sum_{p'=p_1}^{p_2} (c_p T(p') + gz + L_v q(p'))dp' \tag{3.1}$$

is integrated from $p_1 = 1000mb$ to $p_2 = 300mb$. In this equation c_p is the specific heat of dry air at constant pressure, T is temperature, g is gravitational acceleration, and z is the height. The second term, gz, is the geopotential height and is a variable taken directly from the re-analysis, as is the temperature. L_v is the latent heat of

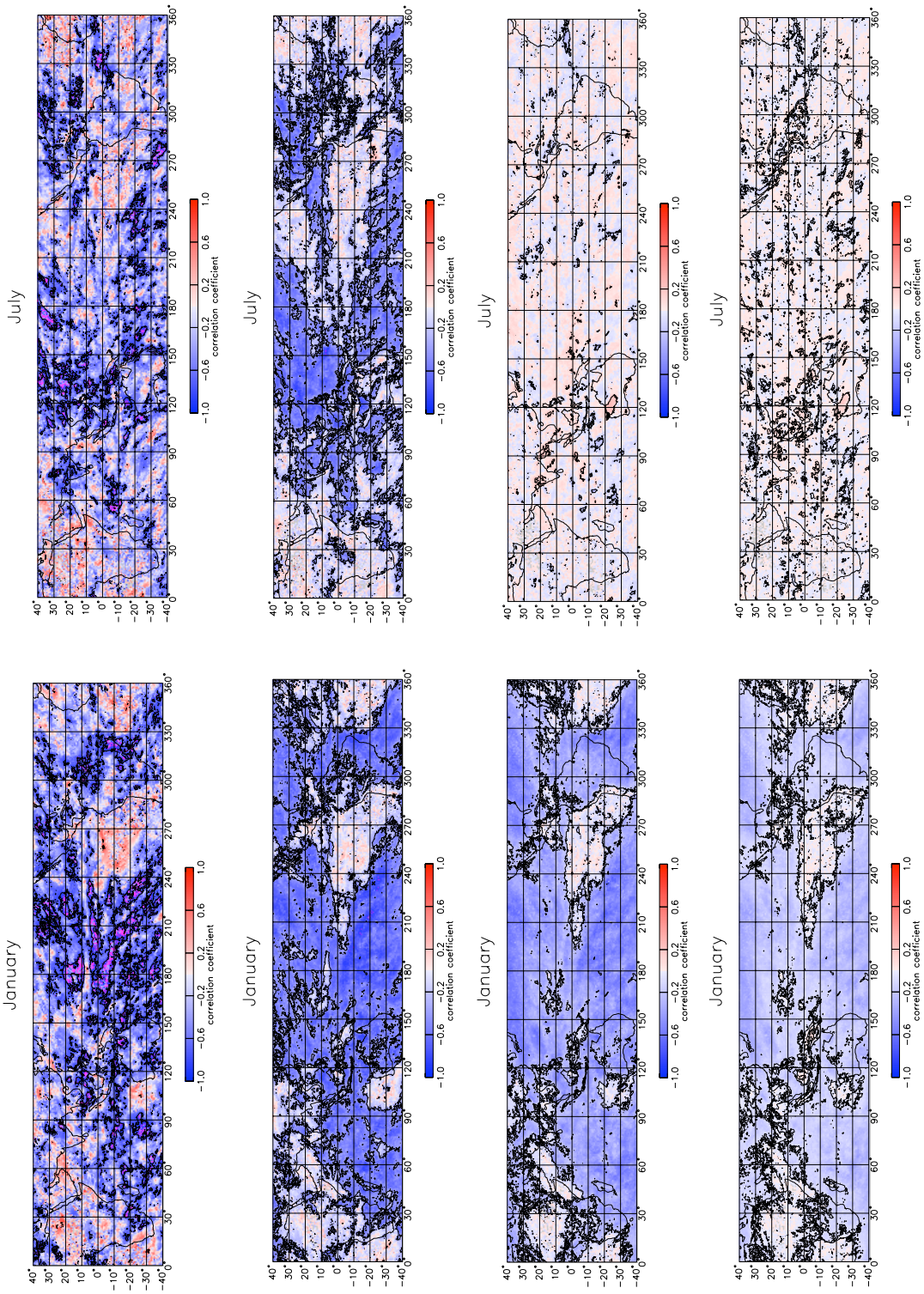

Fig. 9. Top: Monthly 3B42 and NCEP 500mb omega correlations. 2nd row: Pentad correlations. 3rd row: Daily correlations. Bottom: near instant correlations.

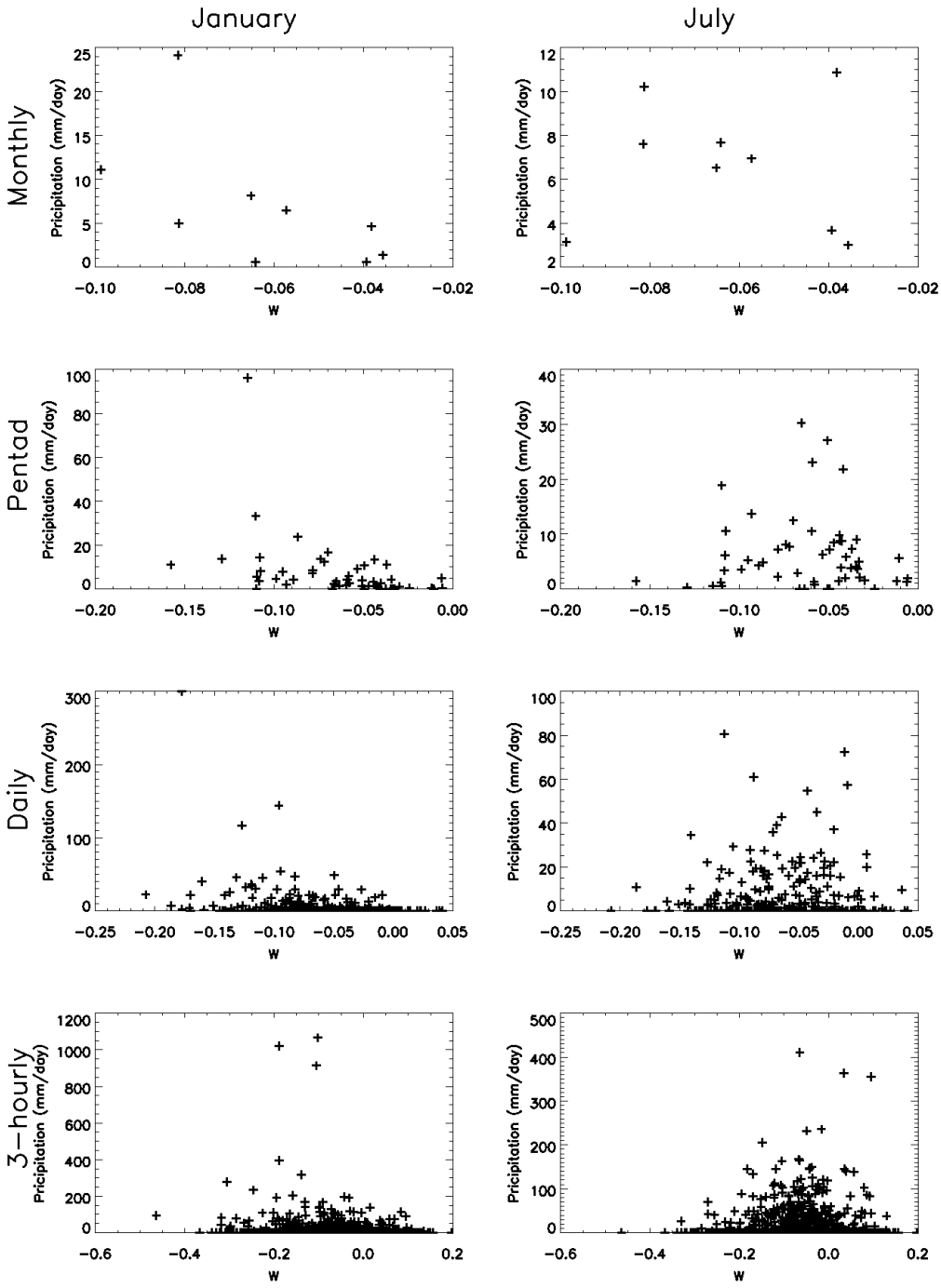

Fig. 10. Monthly, pentad, daily and 3-hourly 3B42 and NCEP 500mb omega scatter plots for two months at a location (0N,160E) in the Pacific.

vaporization and q is the specific humidity.

The most prominent feature in the January correlation maps of precipitation and moist static energy, see Figure 11, is the large region of negative correlations in the Central and Eastern Pacific. Negative correlations imply the precipitation is driving the integrated moist static energy in these areas. There is a strip of positive correlations that starts in the West Pacific and outlines the top of the negative region across the Pacific. There is also a coherent region of positive correlation coefficients in the Indian ocean. There are also positive correlations in the low precipitation region off the West coast of South America. These features hold for all time scales, although the strength of the correlations decrease with time scale.

The July correlation maps show less clear patterns. There is a thin strip of positive correlations starting in the West and carries into the Central Pacific. There is also a region of negative correlations over Australia. The Western Indian and Eastern Pacific have more negative correlations on monthly and pentad time scales. The correlation coefficients decrease on finer timescales and the coherent patterns become harder to perceive.

Scatter plots for the example location (0N,160E) in the Western Pacific are shown in Figure 12. These scatter plots are more ambiguous that those seen above. The precipitation outliers in the January data appear near the mean moist static energy values for each time scale. A very fine look at the near instantaneous scatter indicates a lower density of non-raining values at the largest moist static energy values. The July scatter plots show a decrease of precipitation at the higher moist static energy values for pentad, daily, and near instant time scales.

C. Satellite

1. QuikSCAT divergence

Divergence fields were calculated from the QuikSCAT surface winds data. QuikSCAT measurements have the same spatial resolution as 3B42 but are much less frequent, one ascending and one descending pass per day. We omit results for pentad, daily, and near instantaneous time scales. The analysis from these time scales are dominated by noise. Results for the monthly correlations are shown in Figure 13.

Two narrow bands of negative correlations, just above the Equator in the Central Pacific, are the most prominent features in the January monthly averaged correlation map. There is also some clustering of negative correlations in the Indian ocean, but the majority of the map fails to demonstrate significant spatial structure. The bottom band of negative correlations corresponds with a band of negative divergence and the ITCZ. The top band in the Central Pacific appears on the top edge of the band of convergence and the ITCZ. The top band disappears on pentad and daily time scales. The bottom band of negative correlations, corresponding with the ITCZ, is present but faint and more noisy on pentad and daily time scales. The monthly map for July shows less clear spatial structure than the January map. There is blob of negative correlations in the Central Pacific just South of the Equator. There are also areas with coherent negative correlations in the Indian and Equatorial Atlantic oceans. Figure 14 shows scatter plots at a point (0N,160E) in the Western Pacific for the monthly averaged data. The July data has a decrease in precipitation as divergence becomes positive. The relationship for January data is less clear and dominated by an outlier stationed near a divergence value of zero.

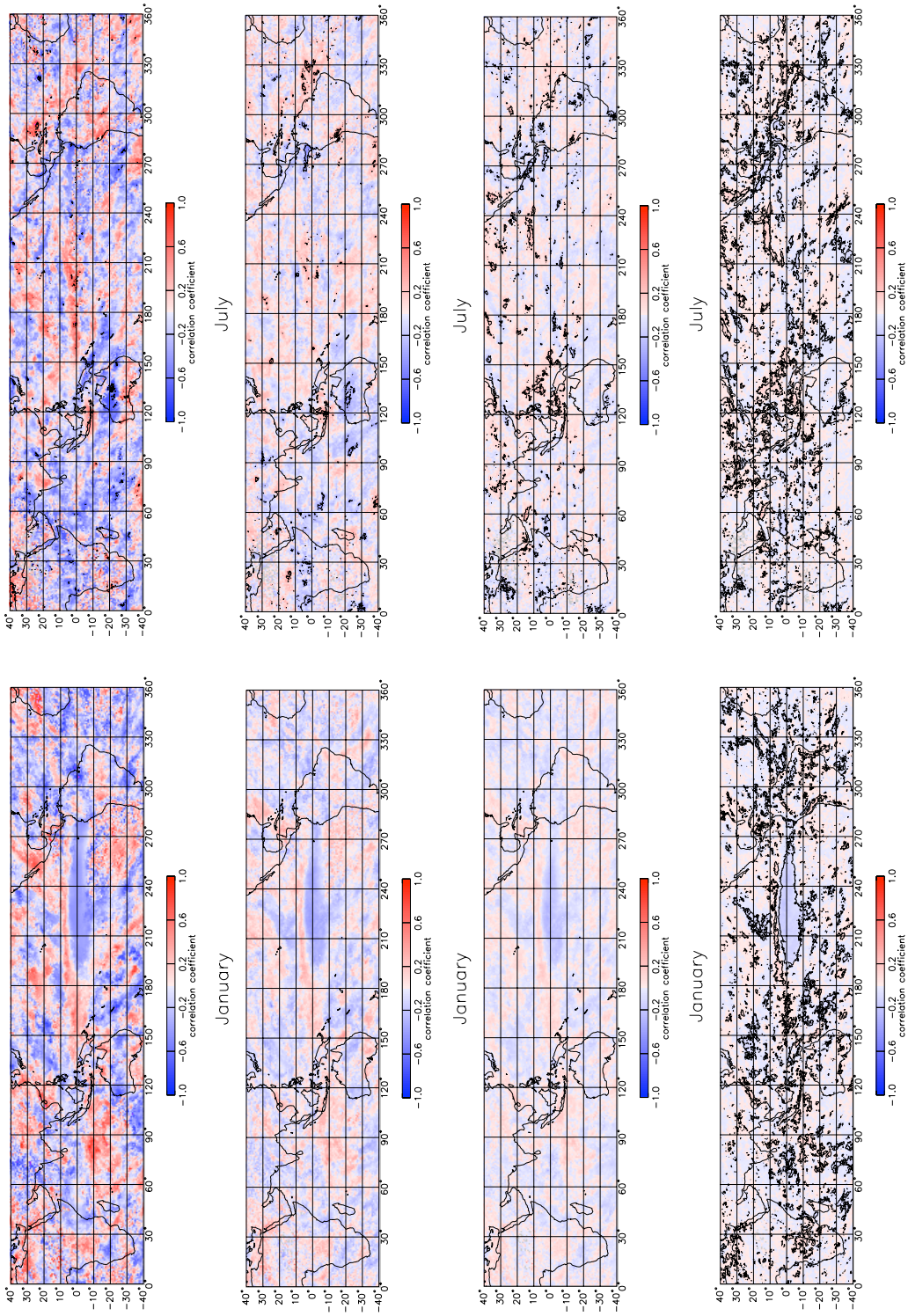

Fig. 11. Top: Monthly 3B42 and NCEP moist static energy correlations. 2nd row: Pentad correlations. 3rd row: Daily correlations. Bottom: near instant correlations.

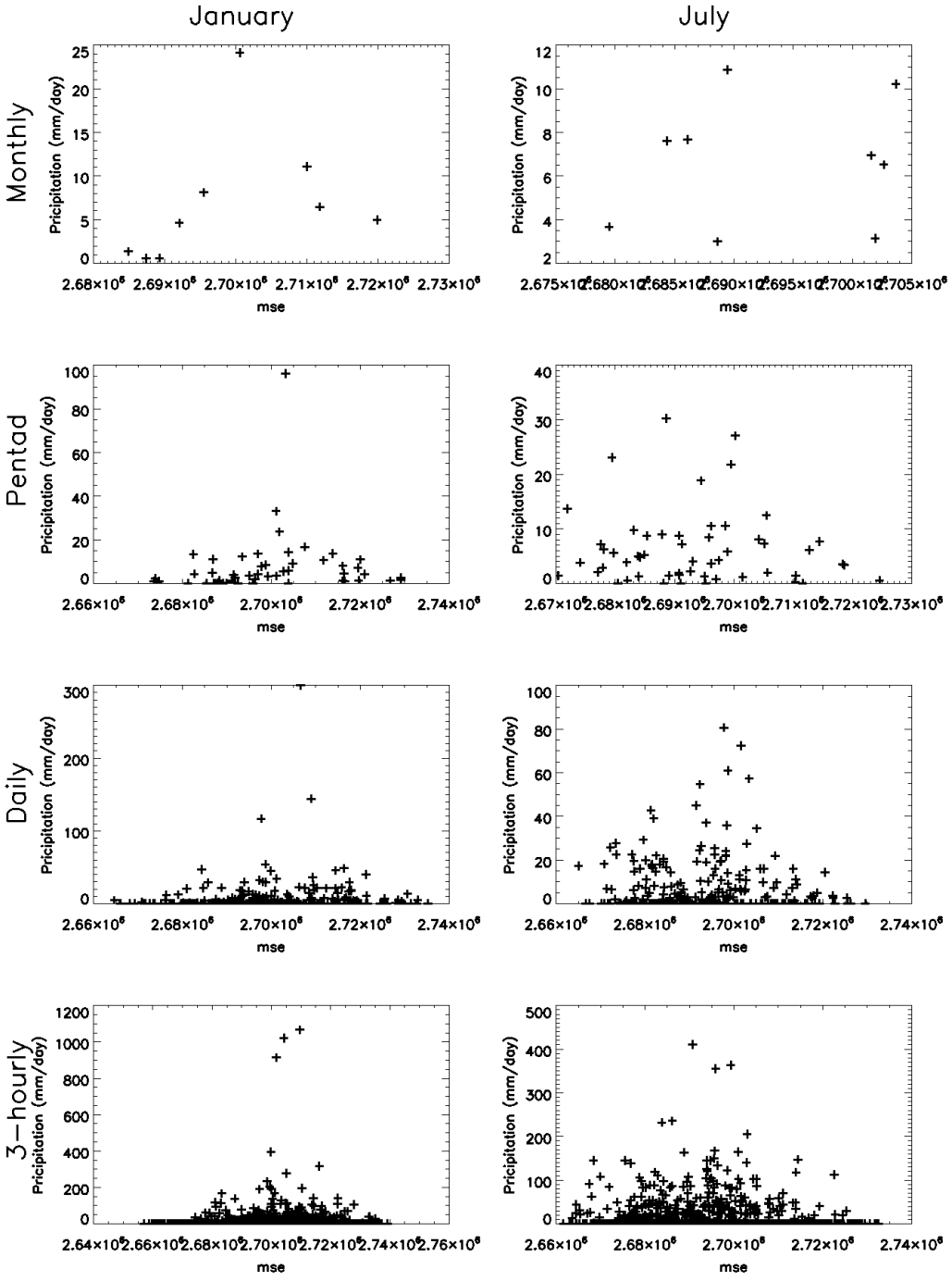

Fig. 12. Monthly, pentad, daily and 3-hourly 3B42 and NCEP moist static energy scatter plots for two months at a location (0N,160E) in the Pacific.

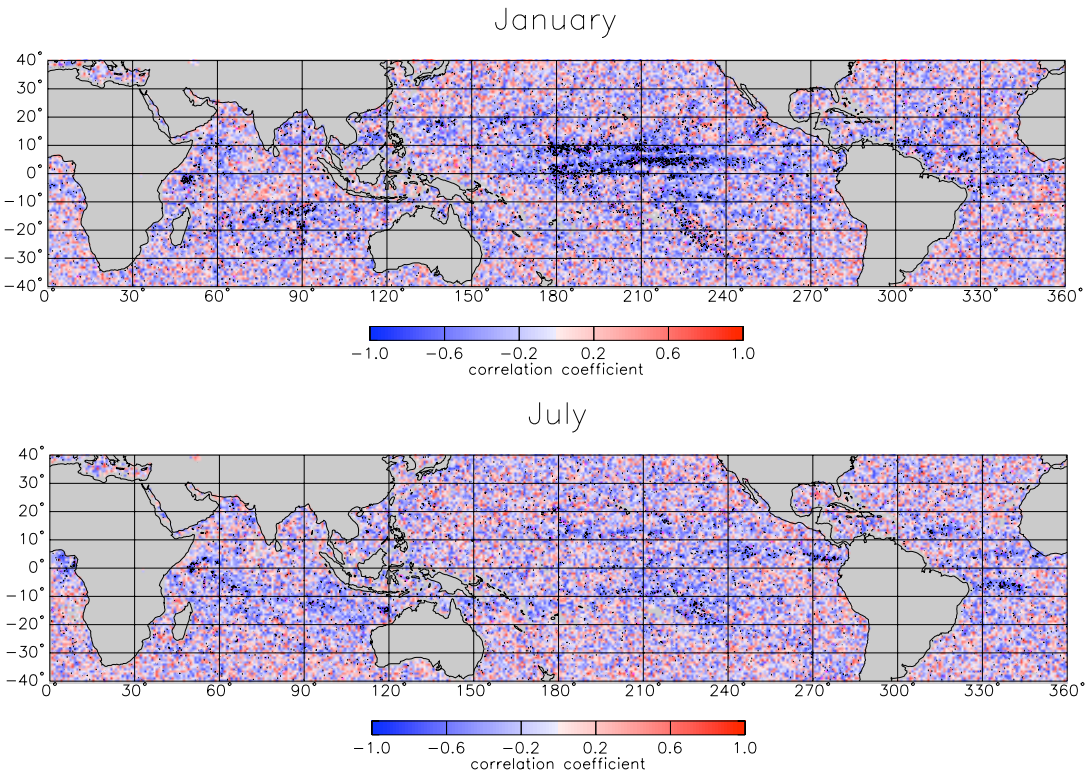

Fig. 13. Monthly 3B42 and Quikscat divergence correlations for January and July from 1999 – 2007.

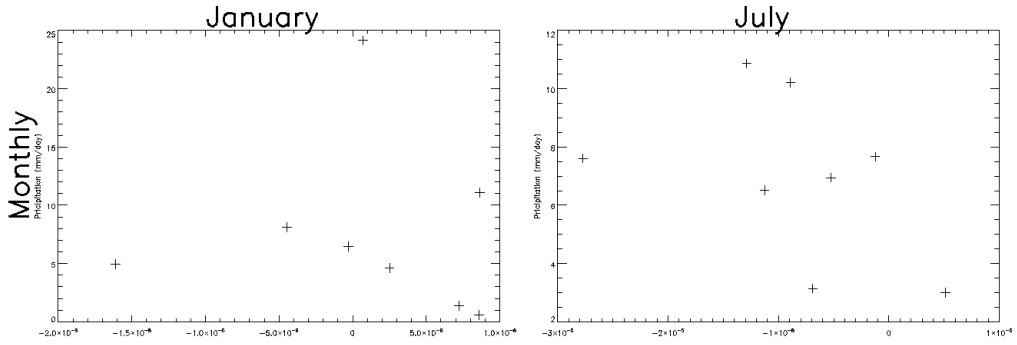

Fig. 14. Monthly 3B42 and QuikSCAT divergence scatter plots for two months at a location (0N,160E) in the Pacific.

We have also calculated correlations between Quikscat surface divergence and NCEP surface temperatures. Correlation maps for the month of January, from 2000 – 2007, are shown in Figure 15. There is a region of negative correlations in the Central and Eastern Pacific along the Equator.There is also an area of negative correlations in the Eastern Indian Ocean. These correlations consist of increase in negative divergence with increase in surface temperature. Positive correlations gird this area to the North and South. These patterns are fairly robust on the different time scales. The area of negative correlations narrows and starts to disappear at daily time scales, while the belts of positive correlations in the Pacific elongate.

D. Summary

We started with a comparison of the results from Trenberth and Shea (2005) with a similar calculation using NCEP/NCAR re-analysis. Results were dissimilar due to a lack of a strong ENSO signal in out time period. We decided to use the NCEP/NCAR re-analysis, rather than the ECMWF re-analysis, for correlation calculations with TRMM 3B42 precipitation because of a longer period of overlap. Correlations between surface temperature and precipitation show coherent structures, of both positive and negative correlations, across the Tropics on each of the four time scales considered. The most prominent of which is a strip of positive correlations in the Western and Central Pacific over the ITCZ and SPCZ. As the time scale decreases the correlation coefficients tend to decrease while the significance increases. The structures have greater spatial extent in January than July. January correlations of 500mb omega and precipitation are largely negative across the tropics on all time scales except in regions of large scale descent of air. Correlations in July are primarily negative on longer time scales, but fade on daily and three hourly scales. Scatter plots indicate

the monthly and pentad correlations are due mainly to extremes. Moist static energy has a larger region of negative correlations in the Central Tropical Pacific. This region is found on all four time scales. Quikscat divergence has a faint and noisy band of negative correlations along the ITCZ in January. Quikscat divergence is negatively correlated with NCEP surface temperatures in the Central and Eastern Pacific along the Equator. The surface temperature and mse correlations show the most coherent structures in the Tropical Pacific. January 500mb omega correlation coefficients on finer time scales are dominantly negative except where they are not statistically significant.

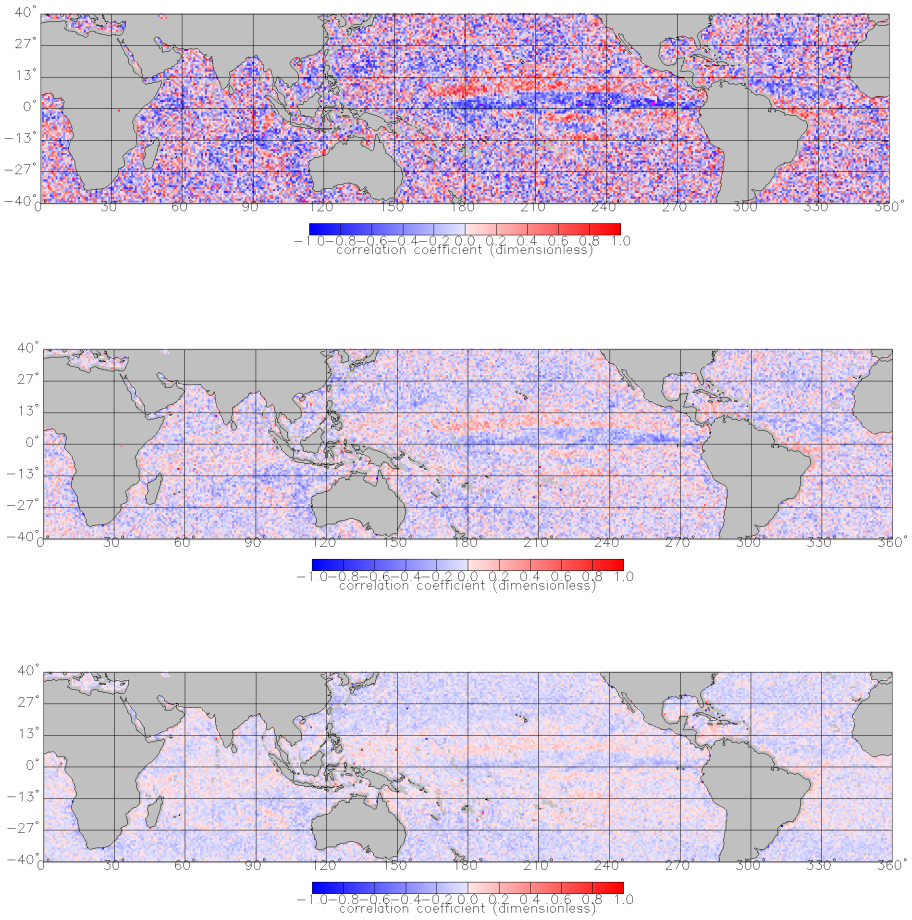

Fig. 15. January NCEP surface temperature and Quikscat divergence correlation maps for monthly (top), pentad (middle) and daily (bottom) averaged data.

CHAPTER IV

EMPIRICAL ORTHOGONAL FUNCTION ANALYSIS

A. Introduction

This chapter describes the results of Empirical Orthogonal Function analysis of vertical temperature and omega profiles in the Tropical Pacific. The structure of vertical temperature profiles is important to convective instability, which is an important factor leading to precipitation. The principal components of the leading EOFs are correlated with TRMM 3B42 precipitation to obtain an idea of how well the corresponding EOF structures are related to precipitation. The principal components for the first few EOFs are correlated with TRMM 3B42 precipitation in the same manner as the flow variables described in chapter III.

B. Omega

To verify the results describing the deep and shallow modes for vertical motion from Back and Bretherton (2008a), we performed an EOF decomposition of the NCEP vertical wind data for the Tropics (20 S – 20 N) during our time period (1999–2007). The first three eigenvectors describe roughly 89% of the variance, similar to the results in the paper. We constructed two new orthogonal base vectors by a linear combination of the first two eigenvectors shown here:

$$\omega_s = -0.44\vec{e_1} - 0.56\vec{e_2} \tag{4.1}$$

$$\omega_d = -0.56\vec{e_1} + 0.44\vec{e_2} \tag{4.2}$$

where ω_s and ω_d are the shallow and deep modes respectively and $\vec{e_i}$ are the first and second eigenvectors. Figure 16 shows both the structure of both the first two

eigenvectors and our linear combinations, which correspond to the shallow and deep modes found in the paper shown in Figure 2. Figure 17 shows variance maps of the principal component time series corresponding with these vectors. The spatial distribution roughly corresponds with the description from the paper, with a stronger deep mode in the Western Pacific. However, we do not see a relatively stronger shallow mode in the Eastern Pacific.

C. Vertical Temperature

Atmospheric temperature profiles are an important parameter for convective instability. In order to take advantage of this relationship we construct Empirical Orthogonal Functions of vertical temperature profiles in the the Tropical Pacific. The EOFs provide information of the dominant modes of variability, and the principal components are a means to relate these modes to precipitation. The vertical temperature EOFs we create are average EOFs for the Tropical Pacific (150E - 80W, 15S - 15N) from the surface to 200 mb. The pressure levels are unevenly spaced, but there is no significant difference between pressure weighted EOFs and unweighted EOFs. The following results are for the unweighted calculations, in the same manner as Back and Bretherton calculated vertical convergence profiles mentioned earlier. First we calculate these EOFs for monthly NCEP re-analysis. Results are shown in the top left of Figure 18. These eigenvectors explain 76% of the variance. The first three eigenvectors explain 52%, 17%, and 8% of the total variance respectively.

In order to be more confident in our results we effect a comparison of NCEP re-analysis and AIRS vertical temperatures in the Tropical Pacific. The top right of

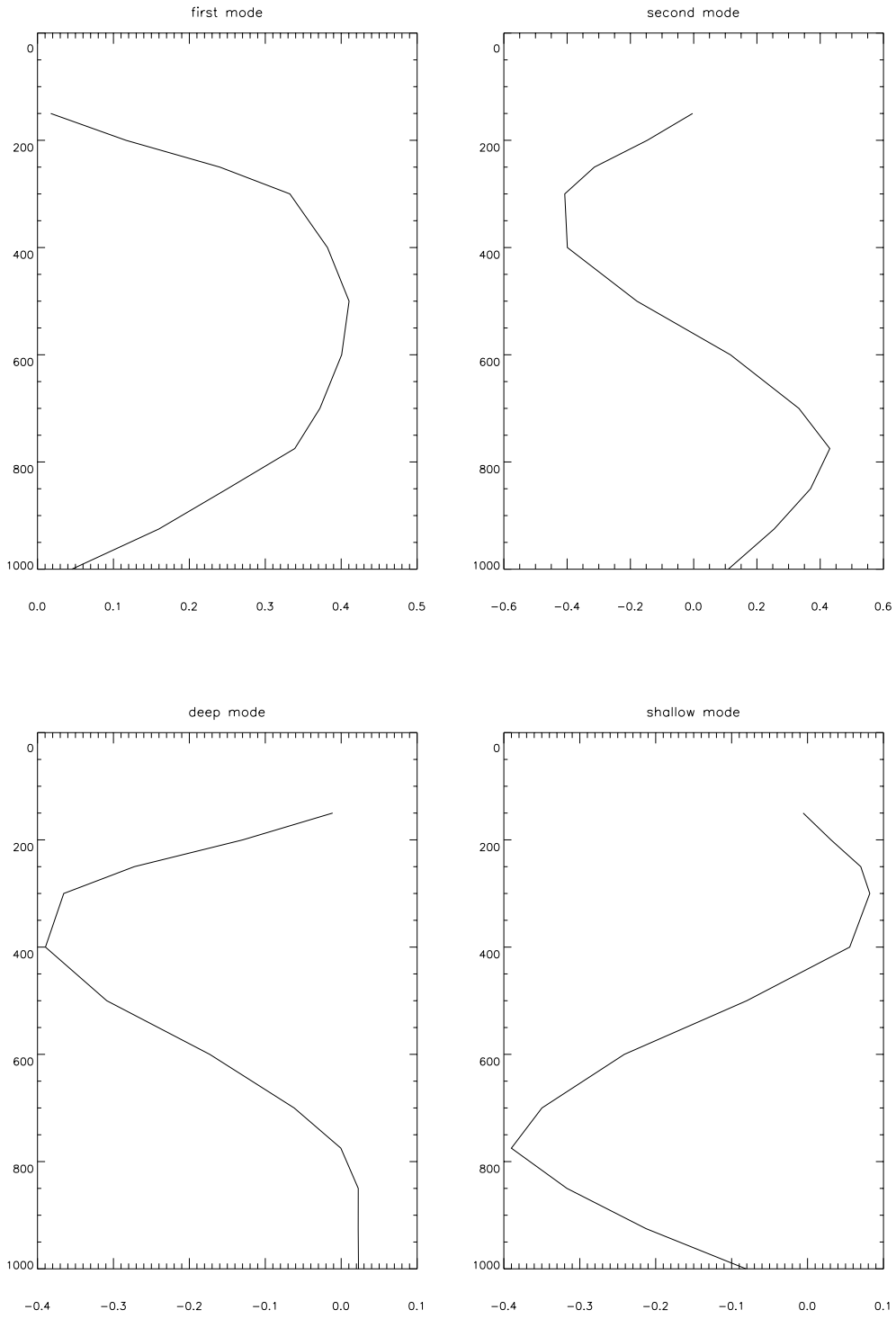

Fig. 16. Top: First two eigenvectors for vertical wind profiles from NCEP Re-analysis. Bottom: Linear combinations of the first two eigenvectors, classified into a deep and shallow mode as described by Back and Bretherton 2008a.

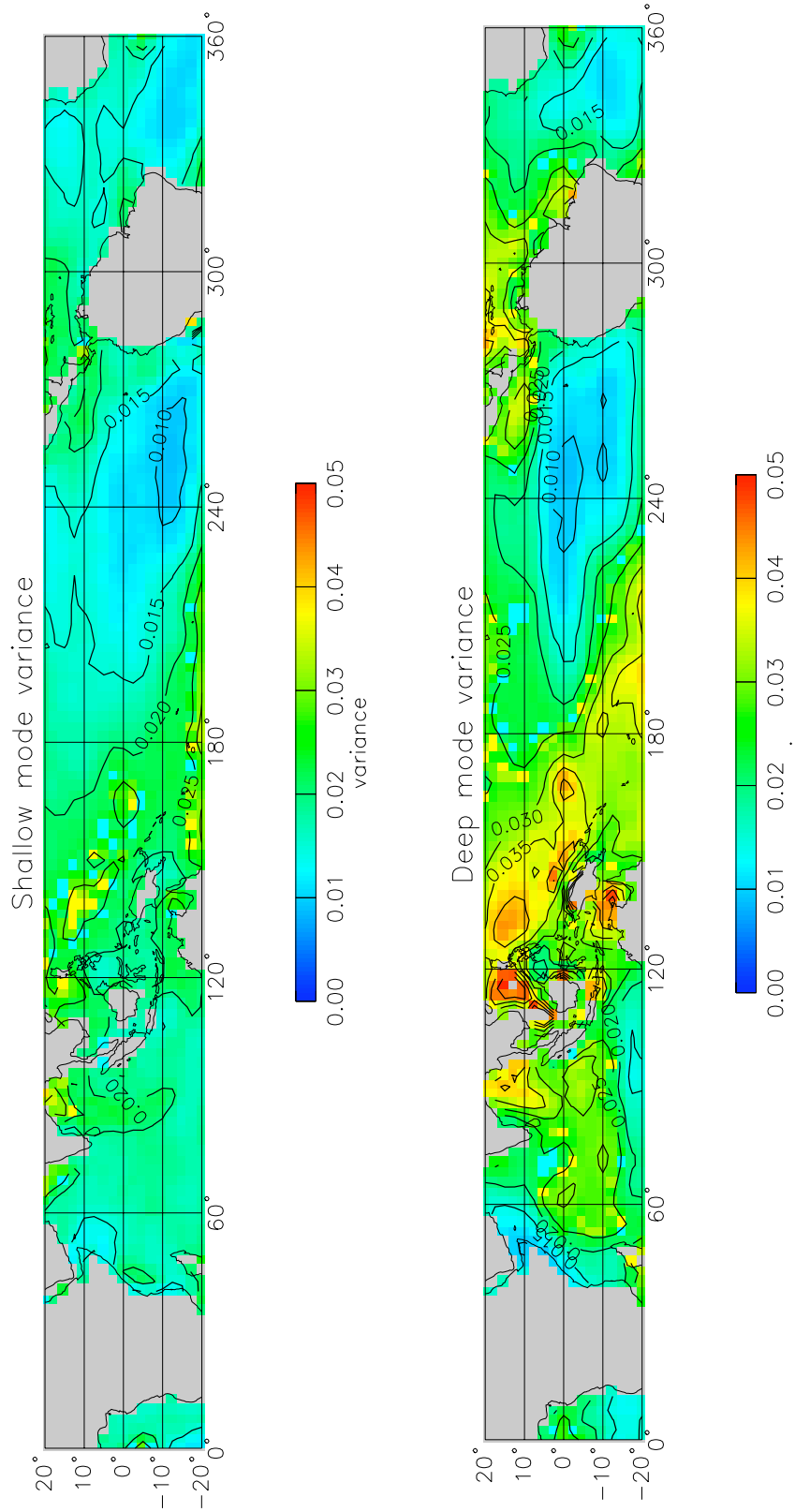

Fig. 17. Variance maps of the principal component time series calculated from the shallow and deep mode profiles shown in figure 8.

Figure 18 shows the first three EOFs from AIRS vertical temperature profiles, formed in the same manner as the re-analysis EOFs. These describe 65%, 20%, and 5% of the total variance. The first two, most dominant, modes are very similar to each other. It is not until the third mode that the patterns start to differ significantly. In light of this we elect to proceed with the AIRS data for further EOF analysis because it provides direct observations with higher vertical resolution than any previous vertical temperature profile measurements. We also constructed vertical temperature EOFs from ERA-40 data (1998 – 2002), which showed similar structure in the first two EOFs as well. To further compare the AIRS and NCEP data sets we created a scatter plot of the two for monthly data at all pressure levels. Figure 19 shows that the two data sets agree quite well in the Tropical Pacific. NCEP seems to have a high bias at the higher altitudes, while AIRS performs less well near the surface. The solid line is the 1:1 line. Figure 18 also shows the EOFs for pentad and daily averaged AIRS data. We see that there is very little change to the structure of the first three EOFs on finer time scales. Table I shows the percentage of variance explained for AIRS EOFS on different time scales. It shows values for EOFs calculated for the entire year, shown in Figure 18, and values for EOFs calculated only for the months of January and July. The January and July EOFs are used for the principal component analysis to follow. The structure of these EOFs are similar to that of the whole year, so we do not reproduce these here in an effort to avoid being repetitious.

1. Principal Components

We calculate correlations between TRMM 3B42 precipitation and the principal components of the first two AIRS EOFs. These correlations are calculated with the same method as those in chapter III. Figure 20 shows correlations of the principal compo-

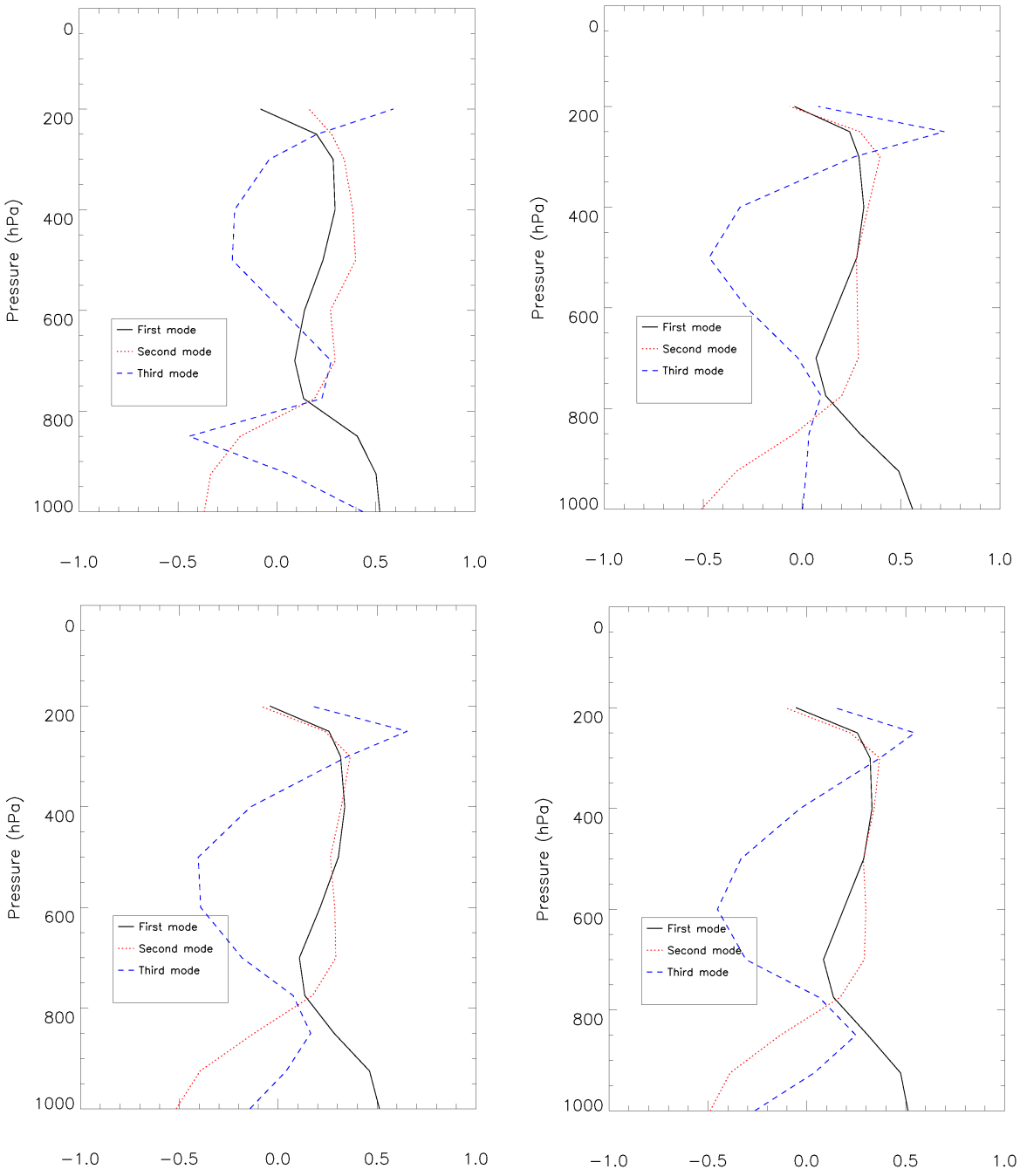

Fig. 18. EOFs of vertical temperature profile in the Tropical Pacific. Top left: Monthly NCEP re-analysis EOFs. Top right: Monthly AIRS EOFs. Bottom left: Pentad AIRS EOFs. Bottom right: Daily AIRS EOFs.

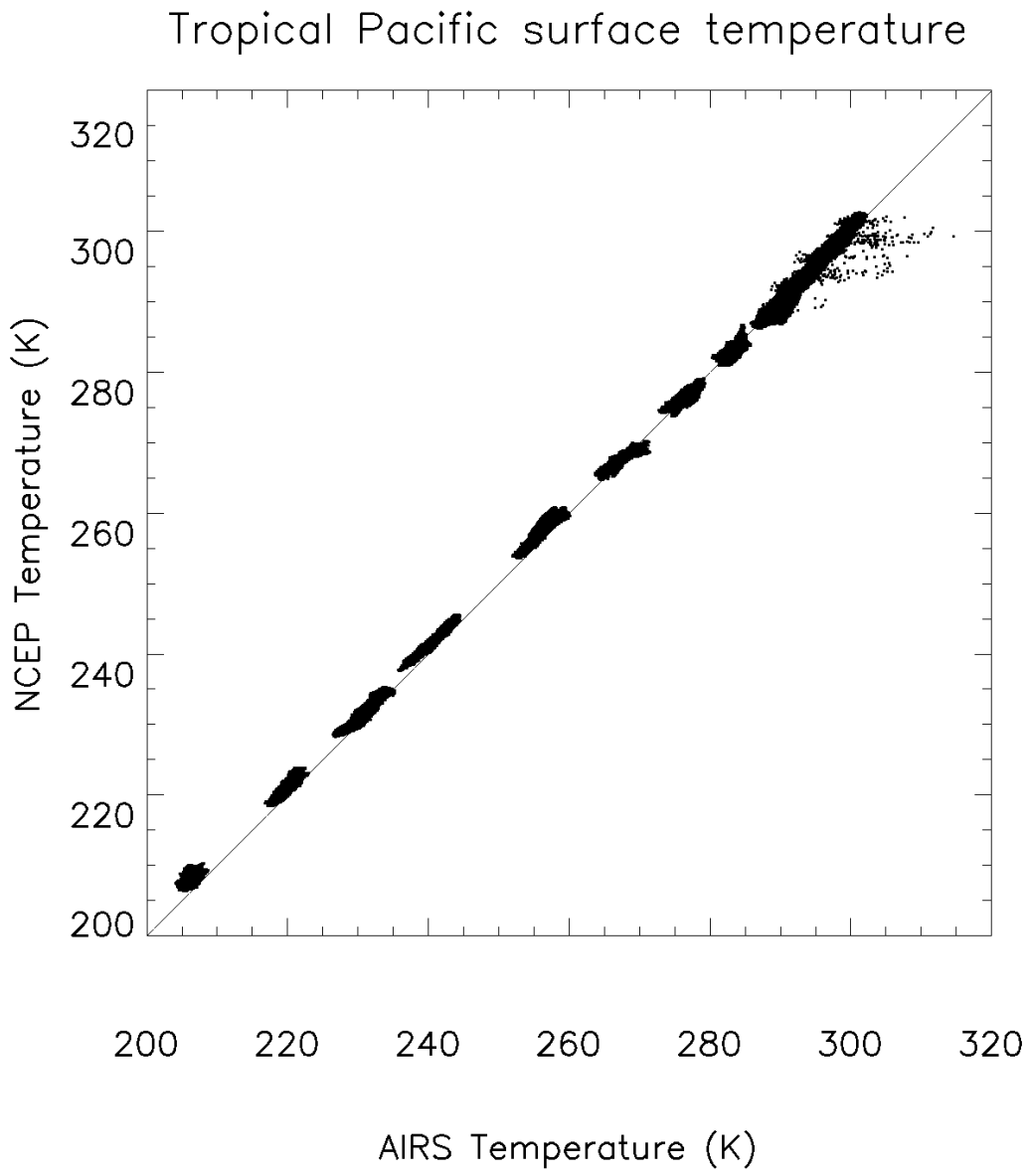

Fig. 19. Scatter plot of monthly NCEP re-analysis and AIRS vertical temperatures, from 1000mb - 200mb, in the Tropical Pacific. Pressure levels are labeled next to the corresponding data and a one-one line is over-plotted.

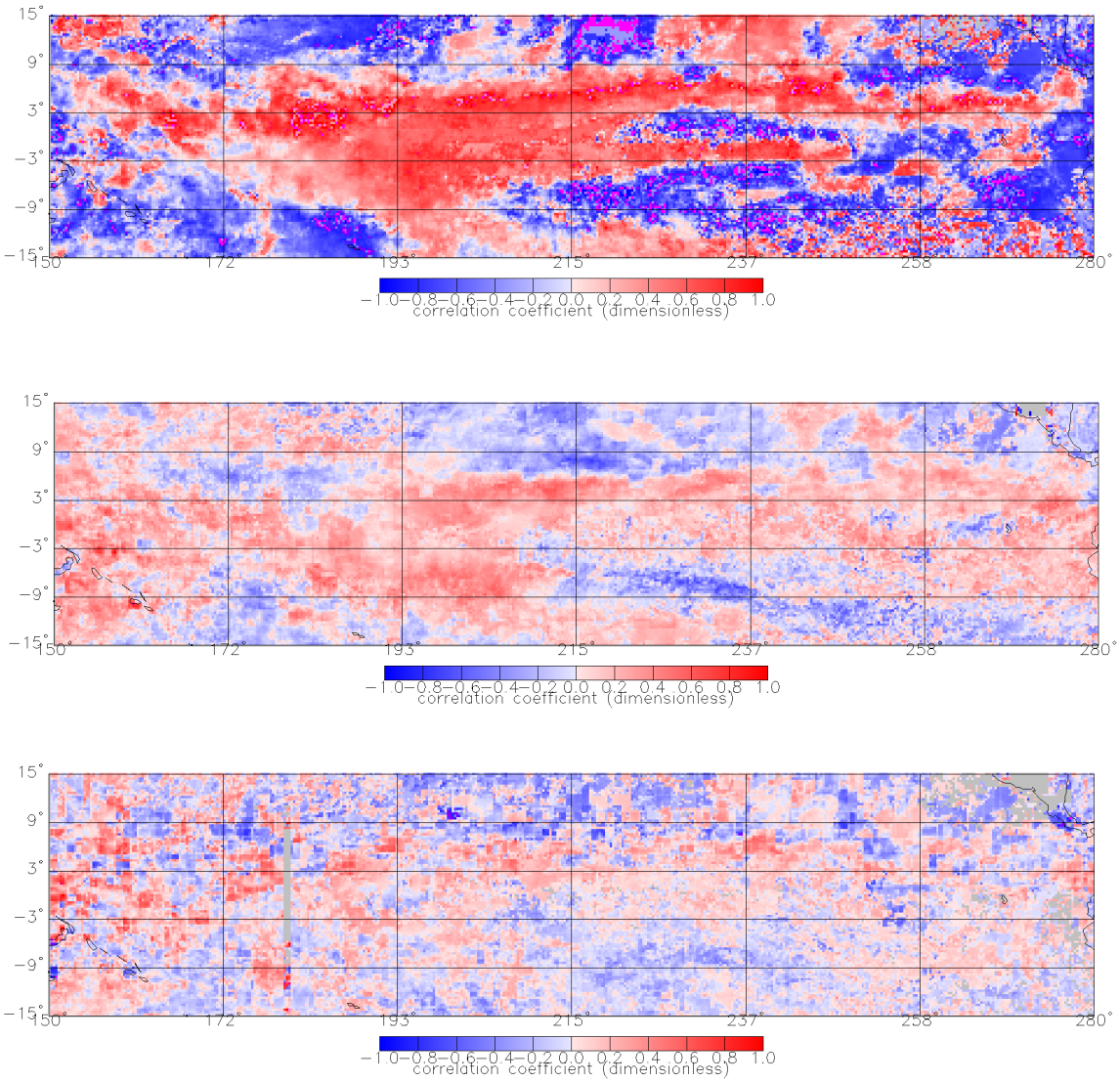

Fig. 20. Correlation maps, for three different time scales, of the Tropical Pacific for January AIRS principal components for the first EOF and TRMM 3B42 precipitation. Top: Monthly averaged data. Middle: Pentad averaged data. Bottom: Daily averaged data.

Table I. Percent variance explained by the first three EOFs of vertical temperature profile in the Tropical Pacific for the entire year, January and July.

EOF #	Monthly	Pentad	Daily
Entire Year			
1	65	57	53
2	20	22	21
3	5	7	8
January			
1	76	61	62
2	14	18	16
3	3	7	7
July			
1	74	60	56
2	13	19	18
3	7	7	9

nents for the first EOF for the month of January. There is banded structure across the region for monthly averaged data. The monthly data for July, shown in Figure 21, has more positive correlations in the western portion of the map. For both times of year the structures dissipate on finer time scales.

The correlations for the principal components of the second AIRS EOF and precipitation are similar to those from the first EOF, see Figure 22. A major difference in the monthly January map is a large swath of negative correlations in the Western Pacific. The bands of positive correlations that stretch across the Pacific are also more narrow.

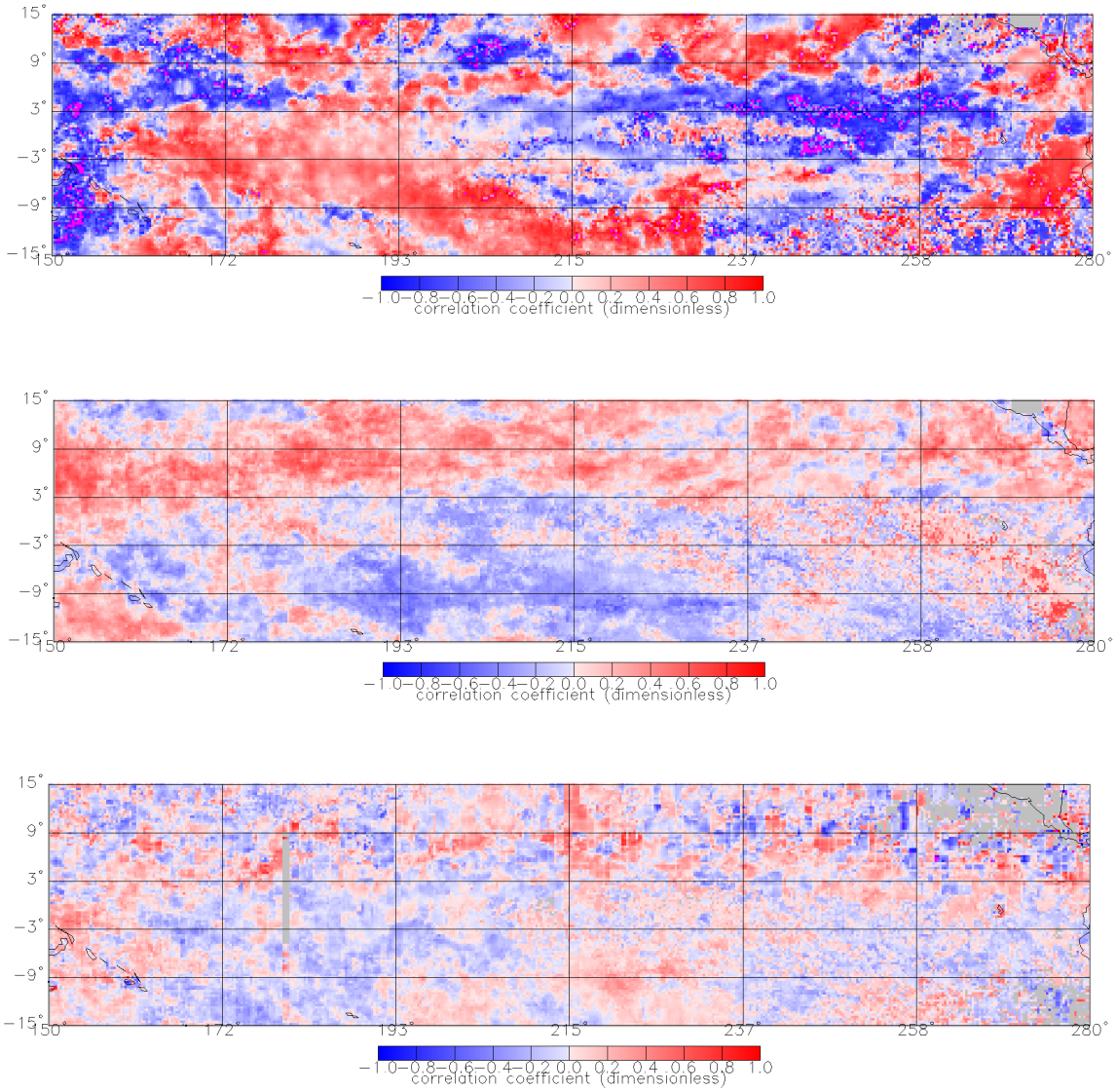

Fig. 21. Correlation maps, for three different time scales, of the Tropical Pacific for July AIRS principal components for the first EOF and TRMM 3B42 precipitation. Top: Monthly averaged data. Middle: Pentad averaged data. Bottom: Daily averaged data.

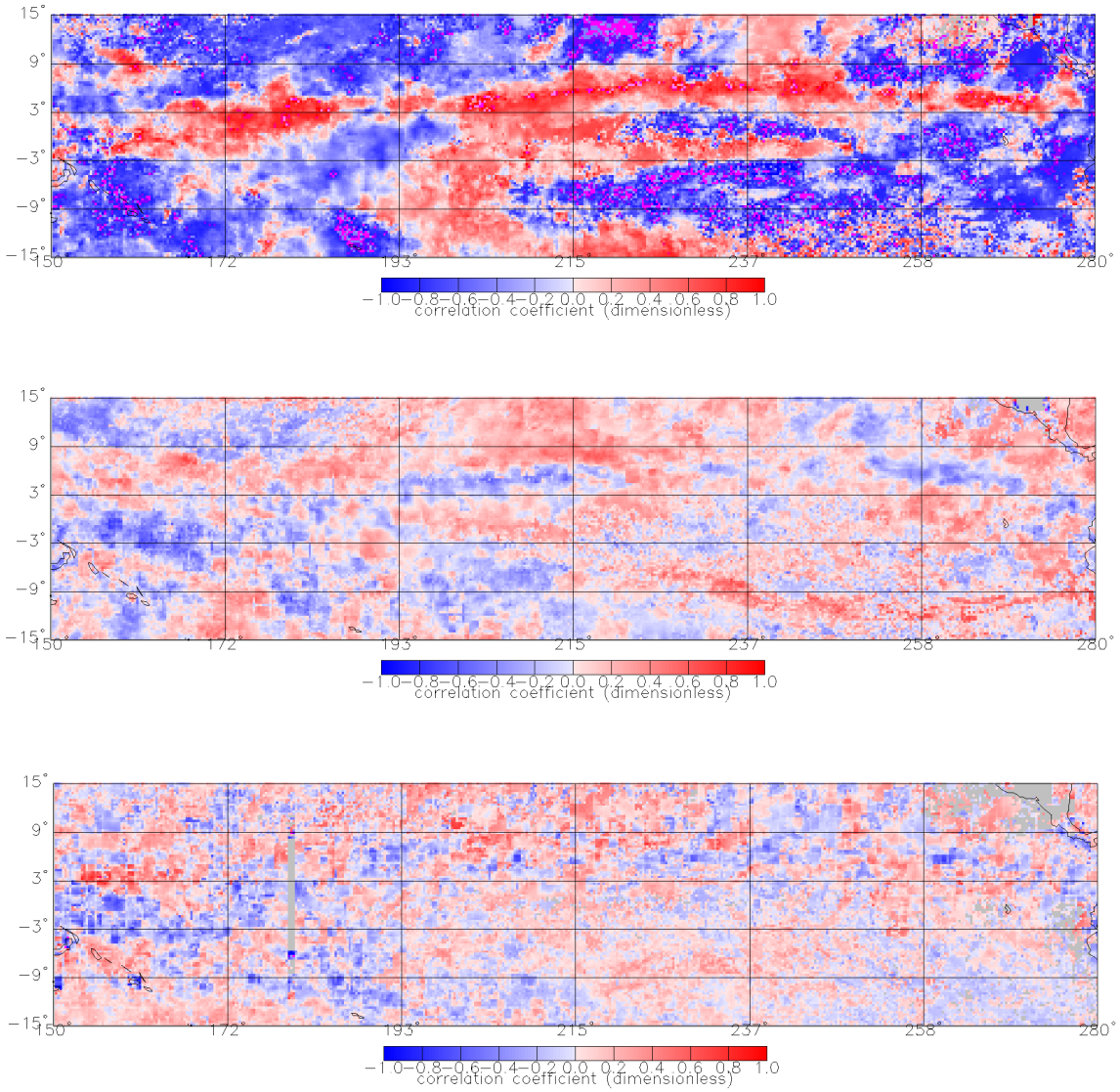

Fig. 22. Correlation maps, for three different time scales, of the Tropical Pacific for January AIRS principal components for the second EOF and TRMM 3B42 precipitation. Top: Monthly averaged data. Middle: Pentad averaged data. Bottom: Daily averaged data.

The monthly July map, for precipitation correlations with principal components from the second EOF, found in Figure 23, shows larger areas of negative correlations in the Western Pacific than are found in the corresponding map for the principal components for the first EOF. In both the January and July cases for the principal components of the second EOF most patterns break down on time scales finer than monthly.

D. Summary

The first few EOFs, of vertical temperature profiles in the Tropical Pacific, from NCEP re-analysis and AIRS observations are very similar in structure. The AIRS EOFs maintain similar shape on monthly, pentad and daily time scales. The amount of variance explained by the first AIRS EOF decreases on finer time scales, while that of the second and third increases slightly. Correlations of the principal components for the first EOF and precipitation in the month of January show a band of positive correlations across the Tropical Pacific. The structure degrades on pentad time scales and dissolves on a daily time average. Monthly averaged principal components and precipitation correlations in July show positive correlations in the Western Pacific and negative correlations in the Eastern Pacific. This structure immediately disappears on pentad and daily averaged data. The maps of correlations between the principal components of the second EOF and precipitation show similar structure to those of the first EOF. Finer time scales lead to more noisy maps of correlation coefficients.

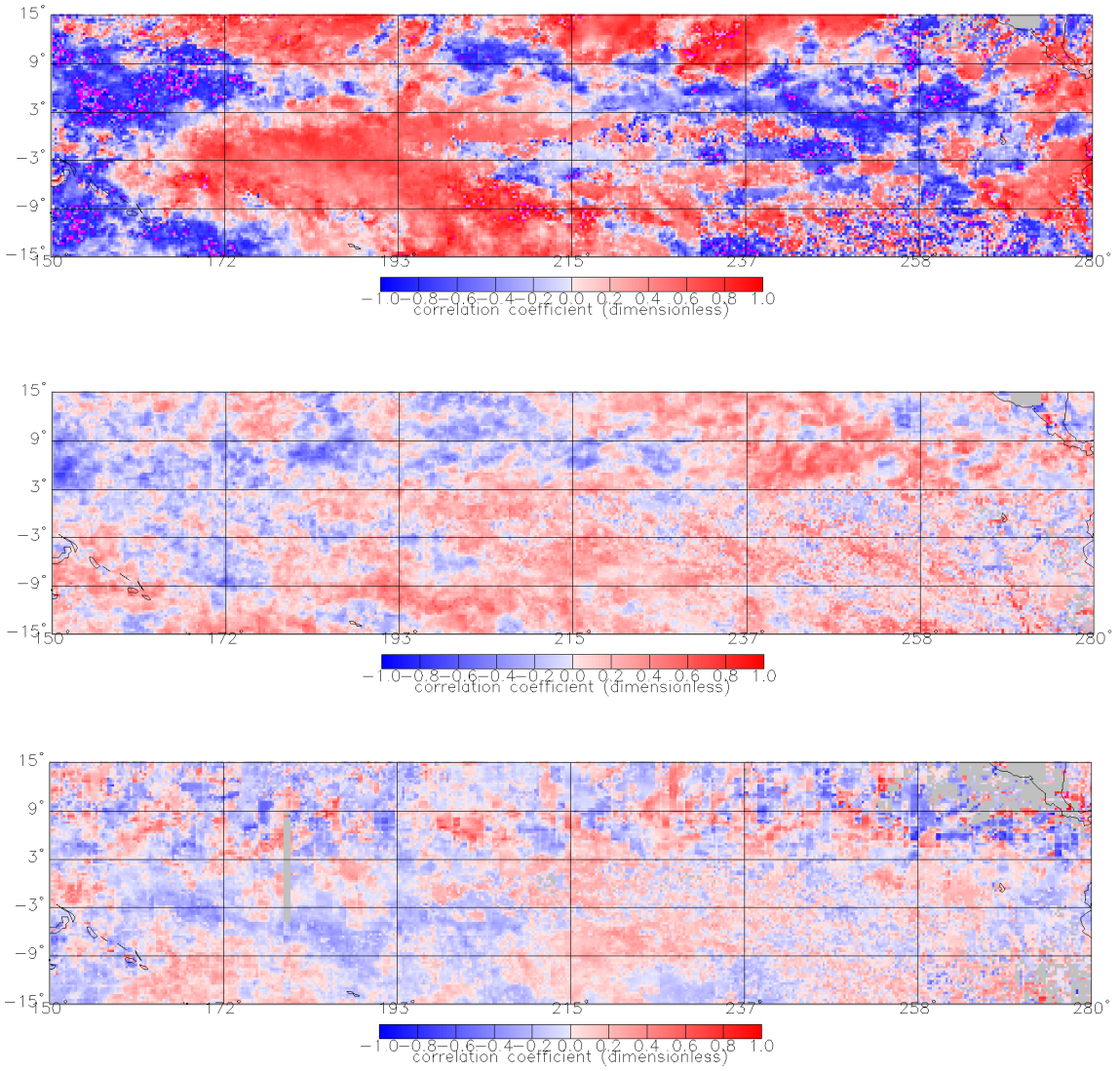

Fig. 23. Correlation maps, for three different time scales, of the Tropical Pacific for July AIRS principal components for the second EOF and TRMM 3B42 precipitation. Top: Monthly averaged data. Middle: Pentad averaged data. Bottom: Daily averaged data.

CHAPTER V

LOGISTIC REGRESSION

A. Introduction

Logistic regression is a common statistical technique to relate a continuous explanatory variable with a binomial response variable. While this is a common technique for statisticians, it has not received attention in the atmospheric science community. This chapter describes the results of logistic regression relationships between TRMM 3B42 precipitation and some of the variables discussed in previous chapters, including NCEP surface temperature and principal components of AIRS vertical temperature EOFs. We specify four regions for logistic regression analysis, shown in Figure 4. Region 1 is in the Western Pacific (150E - 180E, 10S - 10N), region 2 is in the Central Pacific(160W - 130W, 5S - 5N), and regions 3 and 4 are in the South (110W - 84W, 15S - 5S) and North Eastern Pacific (110W - 88W, 0N - 10N), respectively. To perform logistic regression analysis it is necessary to process the precipitation data into a binomial time series. In the binomial time series a value of one identifies a precipitating location and a value of zero indicates a non-precipitating location. Various rain rate thresholds are tested to classify the precipitation data for this step. We select a threshold of 1.0 mm/day. After converting the precipitation into binomial data we compute the logistic regression between this and our continuous variable. In the following plots the probability of precipitation, as a fraction between zero and one, is shown on the ordinate and the continuous regression variable is on the abscissa. Plus symbols are probabilities of precipitation, to provide context, and the solid line is the fit from the logistic regression.

B. SST – Precipitation Analysis

Fits from logistic regression to monthly precipitation and SST data in four regions in the tropical Pacific, shown in Figure 24 shows that the cutoff of 298K in the null hypothesis model of Back and Bretherton (2008a) is a reasonable estimate. Figure 24 shows the logistic regression fits as well as the raw conditional probabilities, for monthly averaged data for the whole year. For these data the threshold for a rain event is 1.0 mm/day. Fit parameters are significant for all regions and shown in Table II. The slope parameter, β_1, describes the change in the log-probability of precipitation per unit change of the explanatory variable, surface temperature in this case. Region 4, in the Eastern Pacific, has a decrease in probability for a rain event for SSTs greater than 300K, causing a divergence from the fit at higher SSTs. Each of these regions shows thresholded relationship between SST and precipitation, as we might expect. The shape of these fits should be considered as a frame of reference for later results.

Table II. Fit parameters for logistic regression analysis of monthly NCEP surface temperature and TRMM 3B42 precipitation for the entire year and the months of January and July. Bold parameters are statistically significant at the 99% level.

	Entire Year		January		July	
Region	β_0	β_1	β_0	β_1	β_0	β_1
1	**-417**	**1.39**	**-523**	**1.75**	**44.3**	**-0.14**
2	**-278**	**0.93**	**-108**	**0.36**	**-158**	**0.53**
3	**-480**	**1.61**	-1183	3.95	-92.3	0.28
4	**-369**	**1.24**	**-33.8**	**0.11**	**-37.3**	**0.13**

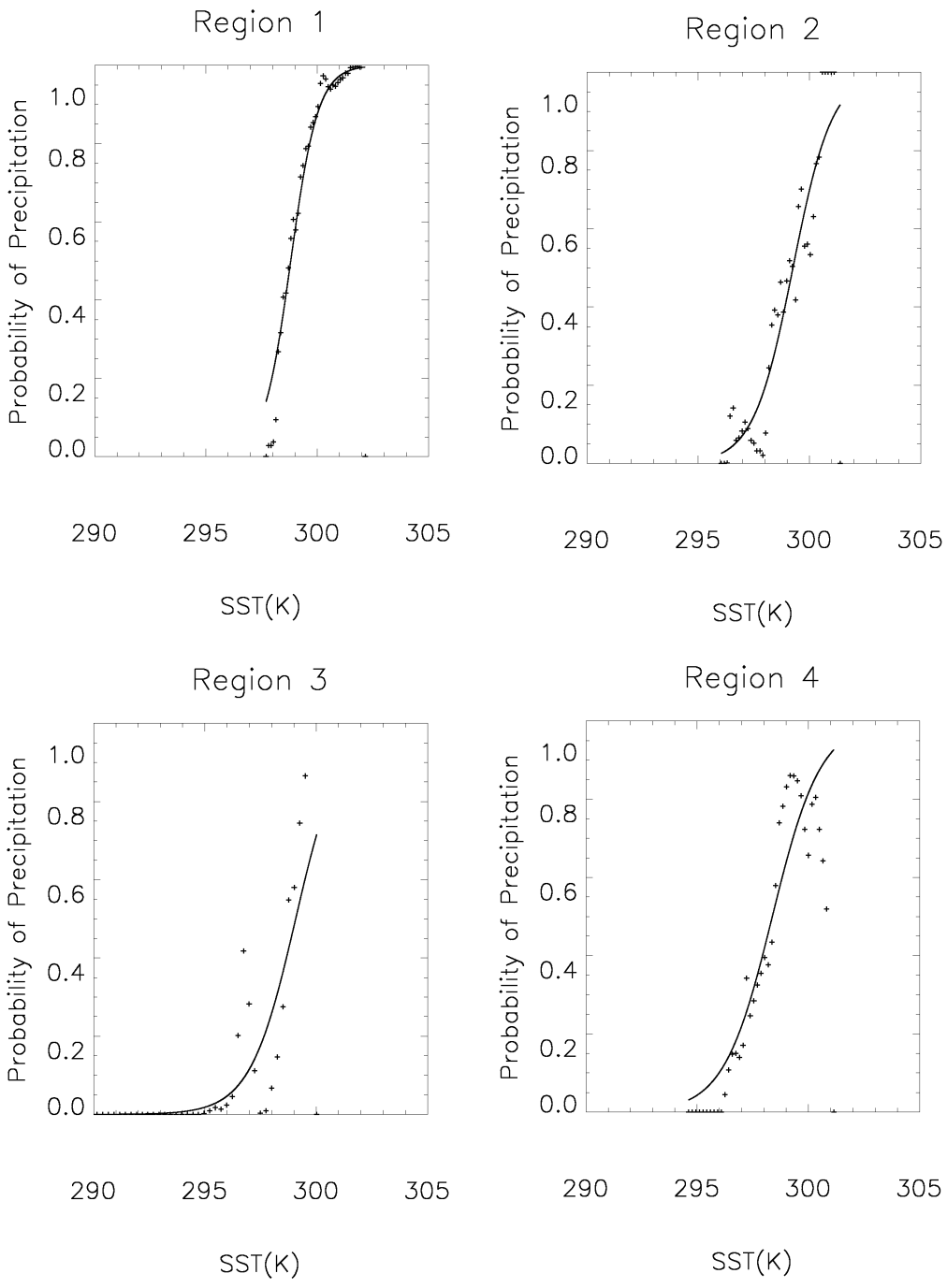

Fig. 24. Probabilities, plus symbols, and the fit from a logistic regression, solid line, for monthly mean SST and precipitation in various regions of the Tropical Pacific for the entire year. The threshold for a rain event is 1.0 mm/day for these calculations.

We perform logistic regression analysis for the months on January and July in the same manner as the correlation analysis in chapters III and IV. The fit parameters for the monthly averaged data are presented in Table II. Results for finer time scales than monthly are not presented because the Newton-Rhapson iterative procedure used to calculate the fit parameters was unable to converge. The fit parameters for the monthly data in January show a strong relationship, $\beta_1 = 1.75$ in region 1. Region 3 does not have significant fit parameters for either month, due to a lack on rain events. Region 2, in the Central Pacific, has larger β_1 parameters than region 4. The fit parameters in these two regions do not change drastically between January and July.

C. Principal Component Analysis

Logistic regressions were fitted for monthly averaged principal component time series of the vertical temperature profiles shown in Figure 10. Figures 25 and 26 show the fits corresponding with the first few eigenvectors in a region. The fits for eigenvectors one and two are shown in Figure 25, and the fits for eigenvectors three and four are found in Figure 26. The fit parameters are printed in Table III and all are significant at the 99% level. In region 1, corresponding to the Western Pacific as seen in Figure 13, high rainfall probabilities preclude a clear thresholded relationship such as found with SST. Region 2 includes more low conditional probability points leading to larger slope parameters from the fit. EOF's 1 and 3 appear to show a thresholded relationship, which can be seen by larger values for the slope parameter. Region 3, a drier region, has very low probabilities for a rain event. All fit parameters are significant, but no clear thresholded relationship is evident. In region 4, EOF 2 has the least threshold like behavior.

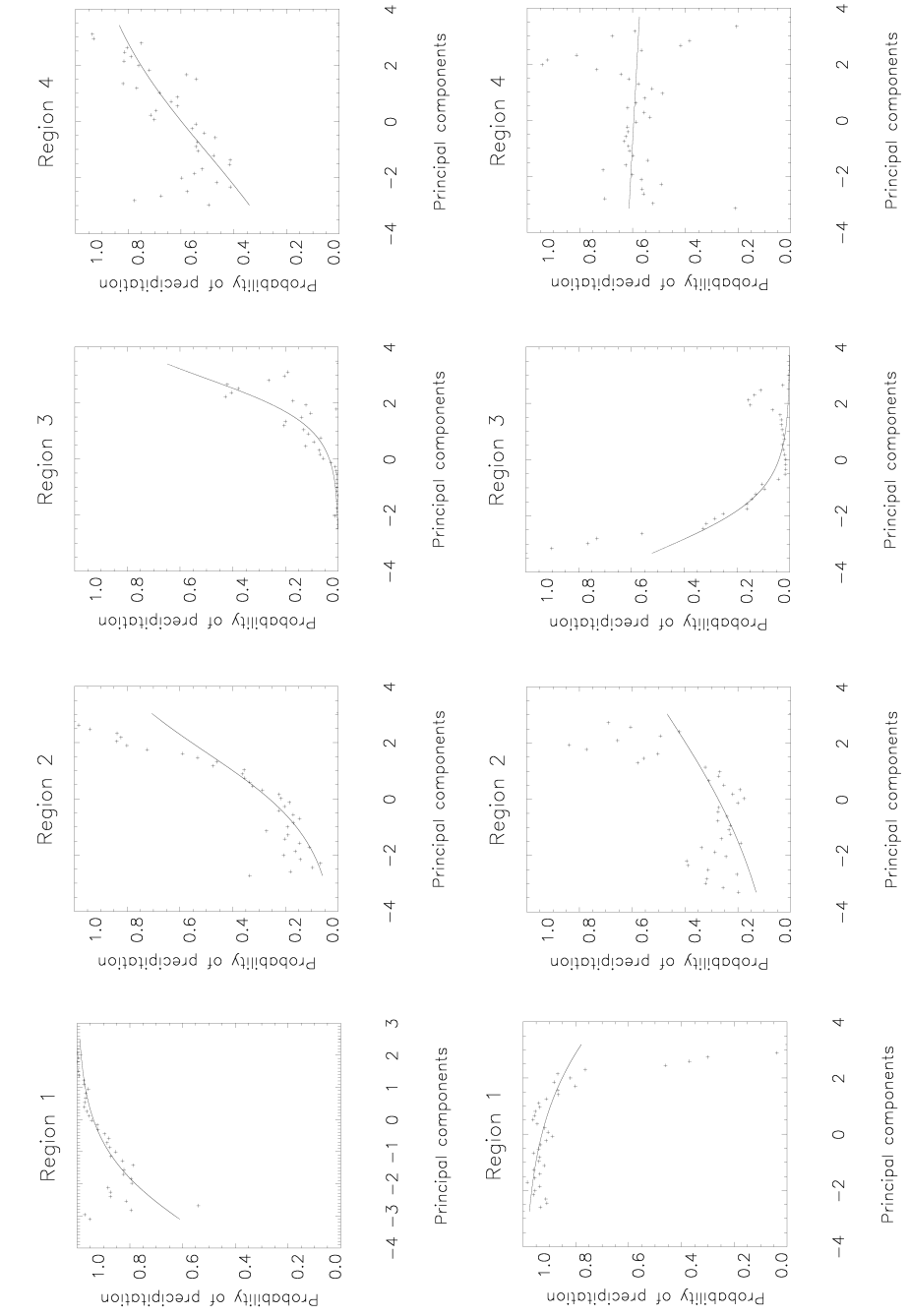

Fig. 25. Top: Probabilities, plus symbols, and logistic regression fits, solid line, from the principal component time series of the first eigenvector for the regions shown in Figure 9. Bottom: Same as above but for the second eigenvector.

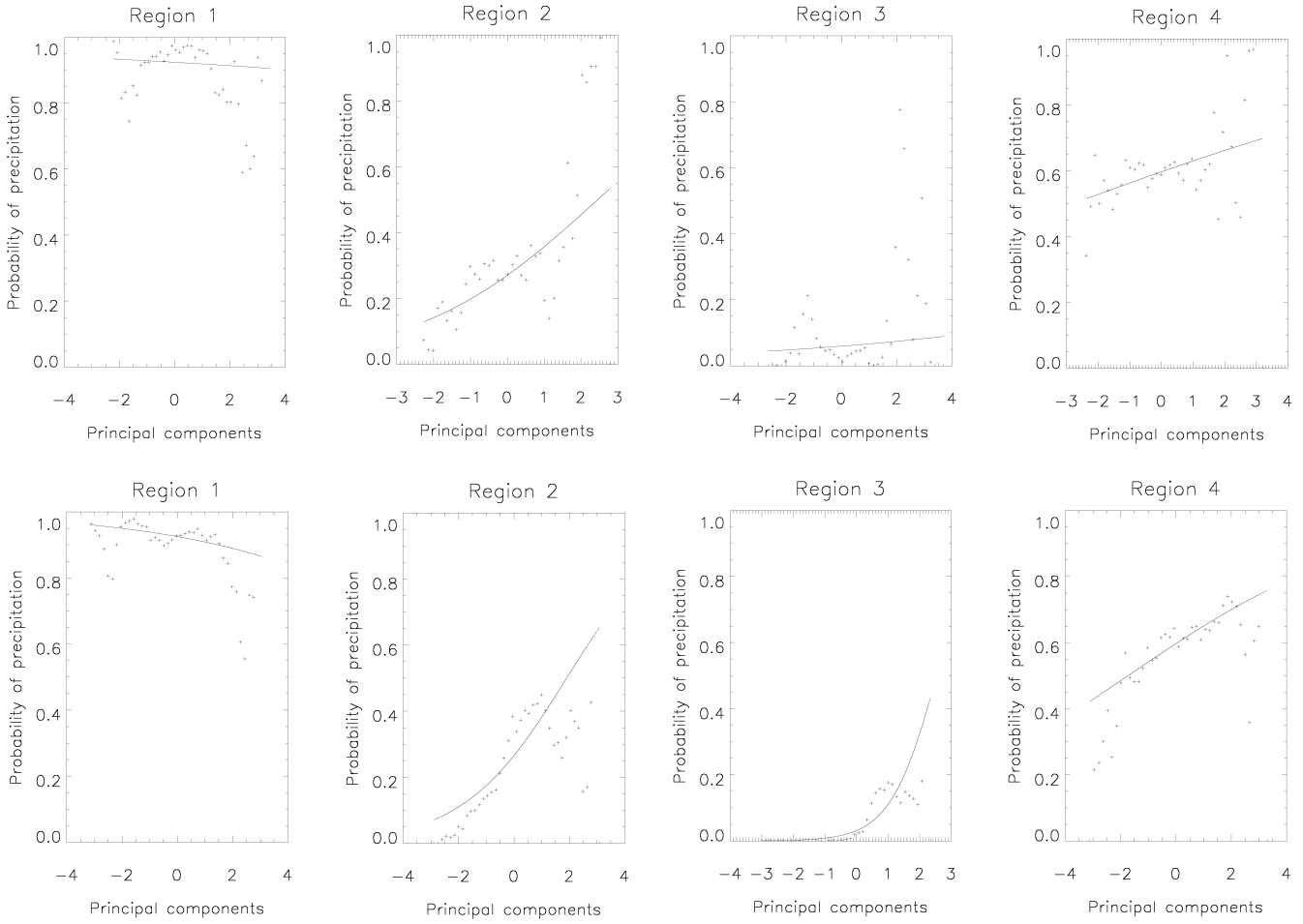

Fig. 26. Top: Probabilities, plus symbols, and logistic regression fits, solid line, from the principal component time series of the third eigenvector for the regions shown in Figure 9. Bottom: Same as above but for the fourth eigenvector.

Table III. Monthly mean vertical temperature principal component and precipitation logistic regression fit parameters. All parameters are statistically significant at the 99% level.

	Region 1		Region 2		Region 3		Region 4	
EOF #	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
1	2.72	0.73	-1.04	0.63	-3.36	1.17	0.40	0.35
2	2.56	-0.41	-0.97	0.27	-3.16	-0.98	0.38	-0.02

D. Summary

Logistic regression has large β_1 values for NCEP surface temperature and precipitation on monthly time scales. Figure 24 shows a clear thresholded relationship is associated with these parameter values. Region 1 in the Western Pacific shows the largest change in 'slope' parameter between analysis for the months of January and July. Principal component analysis has smaller fit parameters on monthly time scales. Parameters for principal component relationships on pentad and daily time scales are much smaller. In all of the analysis region 3 has the most non-significant fits.

CHAPTER VI

CONCLUSIONS

In order to search for statistical relationships between precipitation and the atmospheric circulation on finer spatial scales, we have calculated correlations between TRMM precipitation and NCEP re-analysis surface temperatures, 500 mb omega, moist static energy and QuikSCAT divergence in the Tropics. One emphasis of this research was how these relationships change on several different time scales. We find that monthly correlations are strongest and that correlation coefficients decrease on finer time scales. The finest spatial and temporal scales are comprised mostly of noise except for very strong correlation features.

In addition to these correlations, we calculate empirical orthogonal functions of vertical temperature, from satellite retrievals, as a proxy for convective instability. We find the shape of the first few EOFs is robust for both re-analysis and satellite data in the Tropical Pacific. The principal components for the first EOF are positively correlated with rainfall structures in the Tropical Pacific on monthly time scales but fade on finer time scales. We have also applied logistic regression analysis techniques to this problem with mixed results. This analysis shows the thresholded behavior of SST – precipitation relationship well, but is unable to provide much information on the relationship between precipitation and principal components of vertical temperature EOFs.

Surface temperature and precipitation correlations, in January, have a coherent structure of positive correlations in the Tropical Pacific near the equator. This structure maintains shape on all timescales, but the correlation values decrease on finer

timescales. There are consistent negative correlations with 500 mb omega across the Tropics on all timescales. Moist static energy correlations with precipitation in the month of January show a consistent positive structure in the Western Pacific and a large negative structure in the Central and Eastern Pacific. Many of these correlation structures agree with each other. We expect there to be upward air motion associated with precipitation which can be seen in the 500 mb omega correlations.

The surface temperature correlation structures appear in areas with high climatological rainfall for that time of year, such as along the ITCZ and in the Western Pacific. The correlations are much noisier in areas with low climatological rainfall like descending branches of the Hadley cell. These low rainfall areas may be clearly seen in Figure 1. One exception to this is the large area of negative correlations that dominates the Central and Eastern Pacific MSE – precipitation correlation maps. This one area maintains a consistent structure in a climatologically low rainfall location. In most of the maps shown the structures have greater spatial extent in January than July. QuikSCAT divergence correlations with precipitation are weak and noisy. High SSTs and moist static energy as well as upward motion of air are found to correlate well with precipitation in the same locations. The more complete re-analysis variables tend to show more structure, as opposed to the QuikSCAT relationships. This is not a surprising result due to the spotty nature of satellite data. It also means the structures we see must be treated with caution because re-analysis variables have undergone more processing than the satellite measurements.

The first few EOFs, of vertical temperature profiles in the Tropical Pacific, from NCEP re-analysis, ERA-40 and AIRS observations are very similar in structure. Further examination shows that the observations between NCEP and AIRS are similar

(Figure 19). One explanation for this is the assimilation of IR satellite retrievals from TOVs in the NCEP re-analysis. We see that near the surface AIRS has a bias to larger temperatures while higher up in the atmosphere the NCEP data is biased to larger temperatures. An examination of AIRS EOFs shows they maintain similar shape on monthly, pentad and daily time scales. The amount of variance explained by the first AIRS EOF decreases on finer time scales. One of the more striking features of these EOFs is the sign for the first EOF remains the same from the surface up to around 200mb. Correlations of the principal components for this first EOF and precipitation in the month of January show a band of positive correlations across the Tropical Pacific. The structure degrades on pentad time scales and dissolves on a daily time average. Monthly averaged principal components and precipitation correlations in July show positive correlations in the Western Pacific and negative correlations in the Eastern Pacific. This structure fades on pentad and daily averaged data. As in the correlation analysis described in the third chapter these correlations become more noisy on finer timescales.

Logistic regression is the final statistical technique we use to examine these data. We calculate logistic regression fits between TRMM precipitation, NCEP surface temperature and principal components of vertical temperature EOFs. Logistic regression has large β_1 values for NCEP surface temperature and precipitation on monthly time scales. The slope regression coefficients, β_1, are larger for calculations including the entire year as opposed to a single month. Larger β_1 values indicate a stronger thresholded relationship, i.e. a fit line that comes closer to resembling a step function. The seasonal cycle of rainfall most likely has a strong influence on the larger values occurring for the entire year as opposed to a single month.

Principal component analysis has smaller fit parameters on monthly time scales than surface temperature. Parameters for principal component relationships on pentad and daily time scales are much smaller. Figures 25 and 26 show the importance of rainfall for these calculations. Region 1 in the Western Pacific has high annual rainfall, this leads to lower slope coefficients because of the lack of sufficient zero rain rates. Fewer zero rain rates makes it harder for the regression algorithm to find a fit resembling a step function and thus lower slope values. Region 3, in contrast, has climatological rain rates close to zero, with very few non-zero rain rates. This region appears to have consistently higher slope coefficients than the first region. This is most likely due to the rain threshold of 1.0 mm/day we established at the start. A higher threshold would cause more points in region 1 to be considered zeros and likely increase the slope parameters. This brings up some important considerations for logistic regression analysis.

First, it is hard to establish a suitable threshold for rain rate. One possible solution to this is to perform a multiple logistic regression. This effectively performs a logistic regression against many different thresholds. Although this method will still require us to establish new thresholds based on subjective choices. A comparison of the slope parameters between SST and principal component logistic regression also hints at another difficulty. Logistic regression parameters are not found explicitly, they are found using an iterative computational procedure. This procedure effectively assumes an initial fit, then slowly corrects itself until a convergence criteria is met. On finer time scales, such as daily and near instantaneous, we found that, typically, the convergence criteria would not be met and fits were not found. Finally the logistic regression fits are required to resemble cdf type curve depending on the link function, logit in this case. This immediately assumes there is some kind of

thresholded relationship between the two variables. While this works fairly well for SST and precipitation, most atmospheric variables are not so well related. In light of these facts, logistic regression can be a valuable tool, such as for examining SST – precipitation thresholds, but must be implemented with caution.

In summary, we have found correlations between TRMM precipitation and NCEP re-analysis surface temperature, 500 mb omega, moist static energy have consistent patterns. The correlation coefficients are strongest on monthly time scales and decrease on pentad, daily and near instantaneous. There is significant non-linearity in the precipitation data and extremes greatly affect the correlations. While not present in out time period, a strong ENSO signal dominates these correlations. QuikSCAT divergence correlations with precipitation are weak and noisy. EOFs of vertical temperature profiles in the Tropical Pacific are not sensitive to the time period and have similar structure for both satellite and re-analysis data.

Logistic regression analysis does well in characterizing the thresholded relationship between precipitation and SST, but does not perform as well for principal component – precipitation relationships. This statistical technique may be useful in examining thresholded behavior between precipitation, at various levels, and circulation variables. Future directions for this work include creating a simple statistical model for precipitation based on the relationships we have found and further examination of the vertical temperature structures. Other flow variables such as convectively available potential energy (CAPE) should be include in both correlation and logistic regression analysis. The logistic regression analysis may be expanded to include different indicator variables and multinomial logistic regression. We were unable to find a strong enough relationship, using simple statistical relationships, to merit a

statistical parameterization in climate models. We did find that logistic regression has promise as a non-linear statistic relating precipitation with flow variables such as surface temperature.

REFERENCES

Back, L. E., and C. S. Bretherton, 2006: Geographic variability in the export of moist static energy and vertical motion profiles in the tropical Pacific. *Geophys. Res. Letters*, **33,** L17810.

——, and C. S. Bretherton, 2008a: A simple model of climatological rainfall and vertical motion patterns over the tropical oceans. *J. Climate, accepted.*

——, and C. S. Bretherton, 2008b: On the relationship between SST gradients, boundary layer winds and convergence over the tropical oceans. *J. Climate, accepted.*

Bielli, S., and D. L. Hartmann, 2004: On wind, and SST variations in the North-eastern tropical Pacific associated with the Madden-Julian oscillation. *J. Climate*, **17,** 4080–4088.

Bowman, K., 2004: Comparison of TRMM precipitation retrievals with rain gauge data from ocean buoys. *J. Climate*, **18,** 178–190.

Bretherton, C. S., M. E. Peters, and L. E. Back, 2004: Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, **17,** 1517–1528.

Fu, R., Del Genio, A. D., and Rossow, W. B., 1993: Influence of ocean surface conditions on atmospheric vertical thermodynamic structure and deep convection. *J. Climate*, **7,** 1092–1108.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77,** 437–471.

Madden, R. A., and J. Williams, 1978: The correlation between temperature and precipitation in the United States and Europe. *Mon. Wea. Rev.*, **106,** 142–147.

Neelin, J. D., and I. M. Held, 1987: Modeling tropical convergence based on the moist static energy budget. *Mon. Wea. Rev.*, **115,** 3–12.

Sherwood, S. C., 1998: Convective precursors and predictability in the tropical western Pacific. *Mon. Wea. Rev.*, **127,** 2977–2991.

Simpson, J., R. F. Adler, and G. R. North, 1998: A proposed tropical rainfall measuring mission (TRMM) satellite. *Bull. Amer. Meteor. Soc.*, **69,** 278–295.

Sobel, A. H., and J. D. Neelin, 2006: The boundary layer contribution to intertropical convergence zones in the quasi-equilibrium tropical circulation model framework. *Theor. Comp. Fluid Dynam.*, **20,** 323–350.

Trenberth, K. E., and D. J. Shea, 2005: Relationships between precipitation and surface temperature. *Geophys. Res. Letters*, **32,** 27–45.

Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., Da Costa Bechtold, V., et al., 2005: The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.*, **131,** 2961–3012.

Zhang, C., 1993: Large-scale variability of atmospheric deep convection in relation to sea surface temperature in the tropics. *J. Climate*, **6,** 1898–1913.

Zhao, W., and M. A. K. Khalil, 1993: The relationship between precipitation and temperature over the contiguous United States. *J. Climate*, **6,** 1232–1236.

SUPPLEMENTAL SOURCES

AIRS data available at: http://disc.sci.gsfc.nasa.gov/AIRS/AIRS/AIRS/overview

ERA40 Re-analysis found at: http://dss.ucar.edu/pub/era40/

NCEP/NCAR Re-analysis available from: http://www.cdc.noaa.gov/

Quikscat data found at: http://podaac.jpl.nasa.gov/

TRMM data website: http://daac.gsfc.nasa.gov/data/datapool/TRMM/index.html

VITA

Kyle J. Borg received his Bachelor of Arts degree in physics from Austin College in 2006. He entered the Atmospheric Sciences program at Texas A&M in September 2006 and received his Master of Science in May 2010. His address is Department of Atmospheric Sciences Texas A&M University 3150 TAMU College Station, Texas 77843-3150. His email is: borgkj@gmail.com