Relationship Between Surface Dewpoint and Precipitable Water During the North American Monsoon

by

Paul T. Panhans

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved June 2017 by the
Graduate Supervisory Committee

Randall Cerveny, Chair
Robert Balling
Daniel Krahenbuhl

ARIZONA STATE UNIVERSITY

August 2017

ABSTRACT

The North American Monsoon (NAM) is a late summer increase in precipitation fundamentally caused by a wind shift that is evident in the southwestern United States and northwest Mexico from approximately June-August. Increased precipitation during these months bring an increased regional threat from heavy rains, blowing dust, and damaging storms. (Adams and Comrie 1997). Researchers in Phoenix, AZ theorized that using surface dewpoint measurements was an objective way to officially mark the start of the NAM in Phoenix, AZ (and Tucson, AZ). Specifically, they used three consecutive days at or above a certain dewpoint temperature (Franjevic 2017). The justification for this method was developed by Reitan (1957) who established that 25.4mm (1.00") of integrated precipitable water (IPW) was a sufficient threshold to create storm activity in the NAM region. He also determined (Reitan 1963) that a strong correlation existed between (IPW) and surface dewpoint (Td), whereas, Td could be used as a proxy to determine IPW.

I hypothesize that the correlation coefficients between IPW and Td will be greatest when using seasonal mean averages of IPW and Td, and they will decrease with shortened mean timescales (from seasonal to three-days). Second, I hypothesize that there is a unique relationship between IPW/Td that may signal monsoon onset. To conduct this study, I used the North American Regional Reanalysis (NARR) dataset (1979-2015). For ten locations in the Southwest, I conducted a series of statistical analyses between IPW, Td, and accumulated precipitation. I determined that there is a correlation between the two as set forth by Reitan (1963) as well as (Benwell 1965; Smith 1966; Ojo 1970). However, from the results I concluded this relationship is highly variable, spatially and

temporally. Additionally, when comparing the three-hour, three-day, and the weekly mean measurements, I can conclude that, for my study, timescale averaging did enhance the IPW/Td relationship from three-hour to weekly as expected. The temporal and spatial evolution of the IPW/Td correlation as presented in this thesis may provide a framework for future research that reevaluates the NAM's domain and the associated methods for determining its onset.

ACKNOWLEDGMENTS

I will like to thank my committee chair, Dr. Randall Cerveny for all his direction, patience, and mentorship throughout this process. Additionally, I would like to recognize my other committee members, Dr. Robert Balling and Dr. Daniel Krahenbuhl. Without Dr. Balling's statistics help and Dr. Krahenbuhl's programming expertise, this project would not have been possible. Finally, I would like to thank my loving wife, Dr. Allison Parker, without her love and encouragement I would not have succeeded in finishing this thesis.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1: Introduction

1.1 Overview

The North American Monsoon (NAM) is a potentially hazardous late summer increase in precipitation fundamentally caused by a wind shift. This wind shift is the result of a seasonal change in atmospheric circulation that occurs from mid-spring to early-fall (May-September) (Arias et al. 2015). This singularity is evident for the southwestern United States and northwest Mexico from approximately July-August as a regional threat from heavy rains, flash flooding, damaging winds, small tornadoes, lightning, and dust storms are largely concentrated during these few months (Adams and Comrie 1997). This is the result of increased moisture brought about by a shift from a dry westerly wind to a moist southeasterly wind (Douglas et al. 1993). This change in wind pulls moisture from the Gulf of Mexico, Gulf of California, and Pacific Ocean into the desert southwest of the United States and northwest Mexico. A dry continental air mass gives way to a tropical air mass that surges surface humidity levels, engenders atmospheric instability, and prompts storm activity (Brenner 1974).

While the occurrence of this summertime wind shift is climatologically reliable, the day-to-day weather is erratic and difficult to predict providing a challenge to local forecasters (Maddox et al. 1995). The results of these often spry and powerful storms can be detrimental, but there are positive consequences of the NAM. Every year the NAM provides ~50% of annual precipitation to central Arizona (Adams and Comrie 1997), which is an abundance of hydrologic resources to a dry region. As a result of the NAM's positive and negative implications, the public, weather forecasters, and government

agencies need to prepare for this seasonal change and in the mid-60s researchers developed the first method to determine a date for NAM onset. Increased surface moisture was used as a proxy for the NAM's fundamental seasonal wind shift. Early researchers thought that using a surface moisture measurement was an easy and objective way to officially mark the start of the NAM, specifically, using three consecutive days at or above a certain surface dewpoint temperature (Ellis et al. 2004). Dewpoint temperature, as defined by the American Meteorological Society's *Glossary of Meteorology* (AMS 2017), is the temperature at which saturation occurs for an air parcel held at a constant water vapor undergoing isobaric cooling. Therefore, the higher the surface dewpoint temperature, the greater the surface moisture. The National Weather Service Forecast Offices (NWSFO) in Phoenix, AZ and Tucson, AZ have long employed observed surface dewpoint temperature benchmarks to detect the arrival of the NAM (Skindlov 2007). For example, Phoenix NWSFO, until 2008, delineated the primary criterion for monsoon onset as three consecutive days of a dewpoint at or above 55°F (12.8°C) at Phoenix Sky Harbor Airport while Tucson NWSFO used a dewpoint of 54°F (12.2°C) or above (Ellis et al. 2004). These dewpoints were considered to be correlated with 25.4mm (1.00") of total condensed water vapor in a vertical column of the atmosphere (i.e., Integrated Precipitable Water (IPW)), which was thought to be sufficient to trigger storm activity (Reitan 1957, 1963).

The climatological significance of a three-day consistency of high dewpoints relates to the potential appearance of a pseudo-monsoon. The pseudo-monsoon is an anomalous early summer spike in dewpoint above 50°F (10°C) that does not linger long

enough to signal the full wind shift of the monsoon and produces little rainfall (Skindlov 2007; Franjevic 2017). This phenomenon happens when a mid-tropospheric dynamic system approaches Arizona and advects moisture into the region, which gives a false impression of monsoon moisture arrival. Three consecutive days of high dewpoints was determined sufficient as being a "true" indicator of the monsoon and indicative of the change to a tropical air mass (Franjevic 2017).

There have been concerns raised in applying this method to determine the onset of the monsoon (Ellis et al. 2004; Zhen and Lu 2004; Means 2013). The use of surface dew point as an indicator of moisture influx is variable based on elevation, and is the reason why the NWSFO in Tucson, AZ historically employed criterion of 54°F (12.2°C) dew point rather than 55°F (12.8°C) for the initiation of their monsoon period (Means 2013). Tucson, AZ is at an elevation of 806 meters while Phoenix, AZ elevation is at 346 meters mean sea-level (MSL). Additionally, this method may give a false positive if the moist layer is very shallow (Means 2013). Conceivably, this is the reason why a standard surface dewpoint onset benchmark has not been established for other areas in the NAM's region (e.g., El Paso, TX, Flagstaff, AZ, and Albuquerque, NM). Furthermore, a surface humidity target for Phoenix, AZ and Tucson, AZ only provides a way for determining NAM onset for two discrete locations and does nothing for the surrounding areas. Consequently, researchers have searched for additional ways to define the beginning of the NAM (e.g., Ellis et al. 2004; Zeng and Lu 2004; Lu et al. 2009; Means 2013), which steer away, at least partially, from localized surface humidity measurements. Further

explanation on these methods will be discussed in Chapter 2, which will outline how highly variable the start of the NAM can be.

In order to alleviate this ambiguity, in 2008, the National Weather Service combined with local media, established a firm start/end date for the monsoon (15 June - 30 September) (Haffer 2008). This was irrespective of any wind or dewpoint measurements. The NWS realized that seasonal variability, the range of differences in terms of potential threats, the development of regional benchmarks for the beginning and ending of the monsoon all have caused obscurity in the public's NAM preparedness. While setting a firm start and end date appeared to be the solution to the problem, there could be a complication for this in the future.

Arias et al. (2015) indicated that the possibility exists for an overall shortening of the monsoon season and a drop in rainfall intensity during the NAM, especially, on its northern flank (i.e., southwestern United States and northwestern Mexico). These researchers hypothesized that this might be attributed to rises in sea surface temperatures, a consequence of global warming, in conjunction with the influences of El Niño/Southern Oscillation and Atlantic Multidecadal Oscillation. Cook and Seger (2013) showed a possible delay in onset as well as retreat with a corresponding shift in increased precipitation to the later monsoon months (September–October) with a drop in precipitation intensity (June–July). Ultimately, this implies that a firm start/end date could, theoretically, be obsolete in the future. Because of the conceivable future limitations of a firm start/end date, combined with the shortcomings of other methods, there is a need to reassess the ways in which the beginning of the NAM can be

determined. Moisture is undoubtedly the necessary condition for observing monsoon onset, therefore, a possible link between integrated precipitable water (IPW) and its correlation to surface dewpoint (Td) may provide a scheme for defining the beginning of the NAM season.

1.2 Problem Statement and Hypotheses

The purpose of this thesis is to fundamentally examine the relationship between integrated precipitable water (IPW) and surface dewpoint (Td) during the NAM. Such an examination should 1) enhance knowledge about this relationship and assess the validity of cited literature on the topic and 2) examine if the relationship between IPW/Td is unique to cities in the NAM region such that it could be used as a means for evaluating monsoon onset. In order to properly make these assessments, this study must have an adequate spatial extent that includes cities inside and outside the NAM's domain. For this study, I selected ten cities (Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, and Yuma, AZ). Fig. 1.1 shows the spatial layout for these locations.

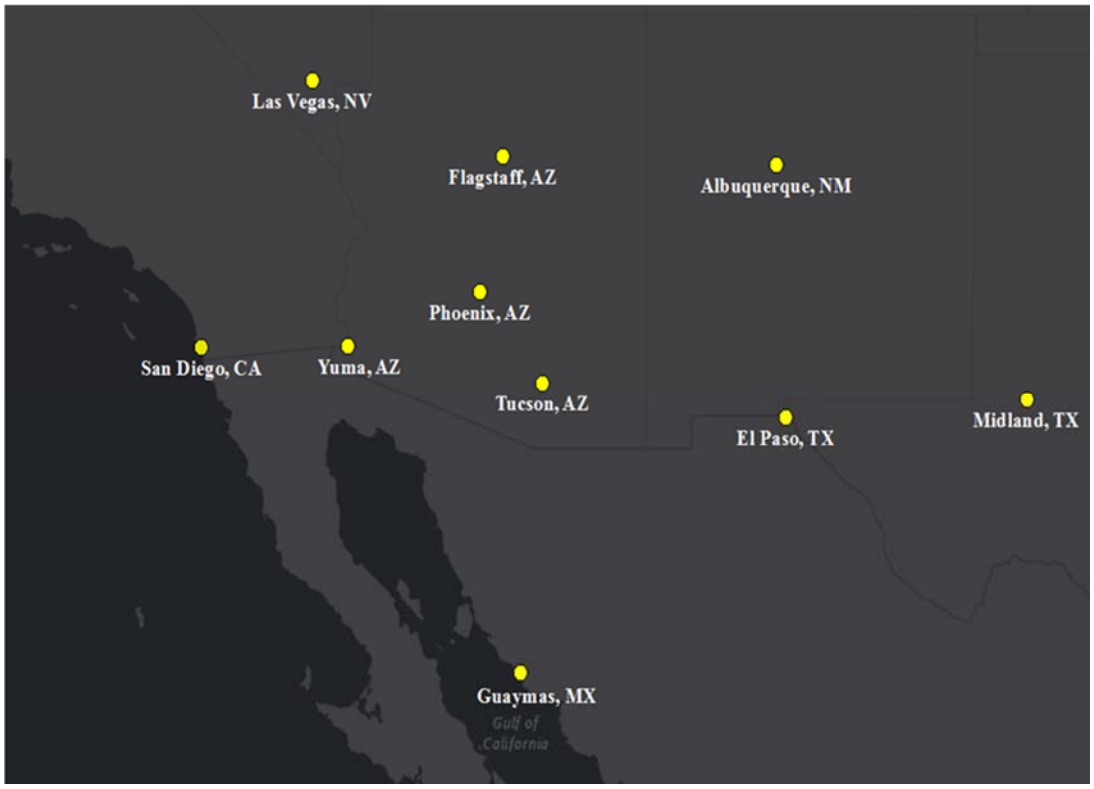


Fig. 1.1. The spatial extent of the study area showing the location of the ten selected cities by yellow dots. (Albuquerque, NM, EL Paso, TX, Flagstaff, AZ, Guyamas, MX, Las Vegas, NV, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ)

Using these locations, I investigate how the relationship between IPW/Td may vary by timescale. Understanding this relationship is important, as it is the critical assumption made by past researchers attempting to determine monsoon onset (Franjevic 2017). Their line of reasoning was that a certain surface dewpoint is indicative of a certain amount of IPW, and a certain IPW is linked to increased precipitation. Therefore, using transitive logic, a certain dewpoint should be linked to an associated level of precipitation. To assume this, understanding the methods in which Td is converted into IPW is necessary. Literature suggests that vertical mixing is needed to support a strong IPW/Td correlation (e.g., Chaboureau et al. 1998), thus, the correlation is time-dependent

and variable-based. If that is the case, then the IPW/Td correlation may be a proxy for the conflation of both monsoon moisture and uplift. This leads to the second research question and hypothesis.

Are there differences in the IPW/Td correlation between cities traditionally associated with the monsoon and those outside of the monsoon? Because the NAM is a spatially exclusive weather phenomena that is the consequence of seasonal spikes in moisture and atmospheric instability, then a correspondingly spatial exclusivity to the IPW/Td relationship may well exist. As result this could be an entirely new way to figure monsoon onset. Given these research questions, I have constructed two research hypotheses.

First, I hypothesize that because the IPW/Td relationship is likely dependent on vertical mixing, and vertical mixing variability increases towards a diurnal scale, the relationship of surface dewpoint and IPW will decrease as the timescale decreases. In other words, the correlation coefficients between IPW and Td will be greatest when using seasonal mean averages of IPW and Td, and will decrease with shortened mean timescales. Second, I hypothesize that there is a unique relationship between IPW/Td that may signal monsoon onset. Put differently, the IPW/Td relationship will have an amplified spatial/temporal distinctiveness for NAM cities versus non-NAM cities that will identify the start of the monsoon season.

1.3 Organization

In this thesis, I analyze the potential IPW/Td relationship both spatially and temporally. Geographically, I examine the correlation coefficient between IPW and Td in ten cities (Albuquerque, NM, El Paso, TX, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, and Yuma, AZ). Temporally, I examine the relationship for five different time intervals (three-hour non-averaged, three-day mean, weekly mean, monthly mean, and yearly mean). Using this, I will assess the changes in the IPW/Td associations across a differently averaged dataset and determine if there is a variability in IPW/Td correlation.

This thesis begins with a full literature review provided in Chapter 2. This will enhance the reasoning for the timescales which I selected, expand the logic of the cities selected, and provide greater justification for this project. In addition, Chapter 2 supplies the justification for use of the North American Regional Reanalysis (NARR) data for this study. Chapter 3 discusses the data used for this study, the associated descriptive statistics of the data for each city, and their associated histograms. Chapter 4 details the methods for the statistical testing both the technical aspects and the theoretical reasoning, and provides the results of the inferential testing. Finally, Chapter 5 is a discussion on the results, evaluation of the validity of my hypotheses, the implications of those hypotheses, and the possible usage of this research in future studies. To begin analysis of the monsoonal relationship between surface dewpoint and integrated precipitable water, an assessment of past research is warranted and that is given in the next chapter.

Chapter 2: Literature Review

2.1 Introduction

As discussed in the previous chapter, I am examining the relationship between surface dewpoint temperature and the depth of moisture (i.e., integrated precipitable water or IPW) throughout the atmosphere in United States' desert Southwest and northwest Mexico during the North American Monsoon (NAM). Through discussion and analysis of past literature, this chapter provides: 1) an understanding of the physical mechanisms driving the monsoon and some of the potential social hazards; 2) a review of the spatial extent of the NAM which gives credence to the selected data points that I use in my study; 3) a discussion of the history of surface dewpoint as an indicator of monsoon onset and examples of other methods for determining onset; 4) a discourse on the history of the "moisture source debate" as it relates to Gulf Surges and outline the importance of low-level moisture in storm genesis during the NAM, and 5) a discussion on the North American Regional Reanalysis (NARR) data that is used in the research questions. In summation, this literature review will aid in the theoretical justification for my selected research topic, data points used, and the research timeframe.

2.2 The North American Monsoon

The North American Monsoon (NAM) is a seasonally predictable but complicated weather phenomenon involving a seasonal wind shift (Adams and Comrie 1997). The shift occurs over the southwestern United States and northwest Mexico from mid-spring to early-fall (June -September) (Arias et al. 2015), and results in an influx of tropical moisture and increased storm activity from approximately July-September (Douglas et al.

1993). The invasion of moisture, combined with intense daytime surface heating and orographic lift creates an environment ripe for afternoon convective thunderstorms (Means 2013). In fact, central Arizona experiences a monsoonal nocturnal precipitation maximum (~2300 LST), which is in part a byproduct of storms forming in the higher terrain areas of Arizona (i.e., Mogollon Rim, White Mountains, Santa Catalina Mountains) and propagation downslope overnight (Balling and Brazel 1987; Watson et al. 1994b) and/or the northward movement of storms created at the US/Mexican border (King and Balling 1994). Inverted troughs, most notably detected on 500hPa maps, can track from the south into southern Arizona and assist in convective initiation in this area (Bieda et al. 2009)

These late afternoon/evening thunderstorms bring heavy rains to the arid/semi-arid desert southwest and creates a seasonal risk of flooding, damaging winds, and dust storms (Maddox et al. 1995; Skindlov 2007). Establishing the timing of the NAM and its spatial extent is the natural starting point for this discussion; however, both are presented as contentious issues in the literature (Adams and Comrie 1997). This section reviews pertinent research on the monsoon with direct relation to my central thesis questions involving the relationship of surface dewpoint and precipitable water and how that relationship may vary inside and outside of the NAM's territory. However, special attention needs to be given to the NAM's timing and its effects on Arizona.

Knowledge of the specific start and end dates of the monsoon in central/southcentral Arizona is important for several reasons. First, Phoenix, AZ and Tucson, AZ are the two largest metropolitans in the NAM's region of influence. Their

two counties alone (Maricopa, Pima) are home to ~5 million people (United States Census Bureau 2010). Second, Arizona is bisected by two major east/west freeways (I-10, I-40), which are major transit routes for ground shipping and personal commuting. Third, Phoenix's Sky Harbor airport is the nation's 10th busiest with respect to yearly takeoffs and landings (Federal Aviation Administration 2016). Lastly, the monsoon has massive hydrologic implications for this region vis-à-vis water storage (Diem 2005). Half of Arizona's yearly precipitation occurs during the monsoon (e.g., Jurwitz 1953; Higgins et al. 1997; Sheppard et al. 2002). And, 60% of southern Arizona's year precipitation occurs during the monsoon (Douglas et al. 1993). Consequently, careful considerations to water management during this season must be made in order to sustain life in the semi-arid Mohave and Sonoran Deserts. These factors combine into the National Weather Services' mission, "To provide weather, water, and climate data, forecasts and warnings for the protection of life and property, and enhancement of the national economy" (National Weather Service 2017). This is why considerable research has been dedicated to the study of the monsoon and its effects to Arizona, although, as discussed below, the spatial extent goes beyond the borders of Arizona.

2.2.1 Geographical Extent

Despite common misconceptions, the spatial extent of the NAM extends far beyond the borders of Arizona (Douglas et al. 1993; Adams and Comrie 1997; Means 2013). This is such a commonly held misconception that the North American Monsoon is often mistakenly called the "Arizona Summer Monsoon". In truth, the epicenter of the monsoon's influence is not even focused in the United States, rather, the monsoon pattern

is centered over the Sierra Madre Occidental (SMO) mountain range, which runs meridionally through west-central Mexico (Adams and Comrie 1997; Diem 2005). Douglas et al. (1993) made the point, that in addition to Arizona the spatial extent of the NAM also includes western New Mexico, southeastern California's Mojave Desert, southern Nevada, portions of west Texas, including the city of El Paso, TX and northwest Mexico, specifically the region west of the SMO. Douglas et al. (1993) also showed great similarities between rainfall patterns in Mexico west of the SMO and those of Bombay, India using mean monthly precipitation frequency tabulations. This suggests the rainfall seen in western Mexico is actually more of a traditional monsoon precipitation signature than that seen in the southwest United States. In their research, they stated that the terminology of the "'Arizona Monsoon'…reflects the geographical bias of the investigators of the United States" (p. 1667). This rebuke underscores the importance of researching the monsoon with a southerly extent well into Mexico (i.e., Guaymas, Sonora, MX). Moreover, the region west of SMO has the greatest amount of rainfall in the NAM's area of influence (Becker and Berbery 2008; Svoma, 2010). Fig. 2.1 from Douglas et al. (1993), shows the geographic extent of the monsoon and its temporal evolution from June through September.

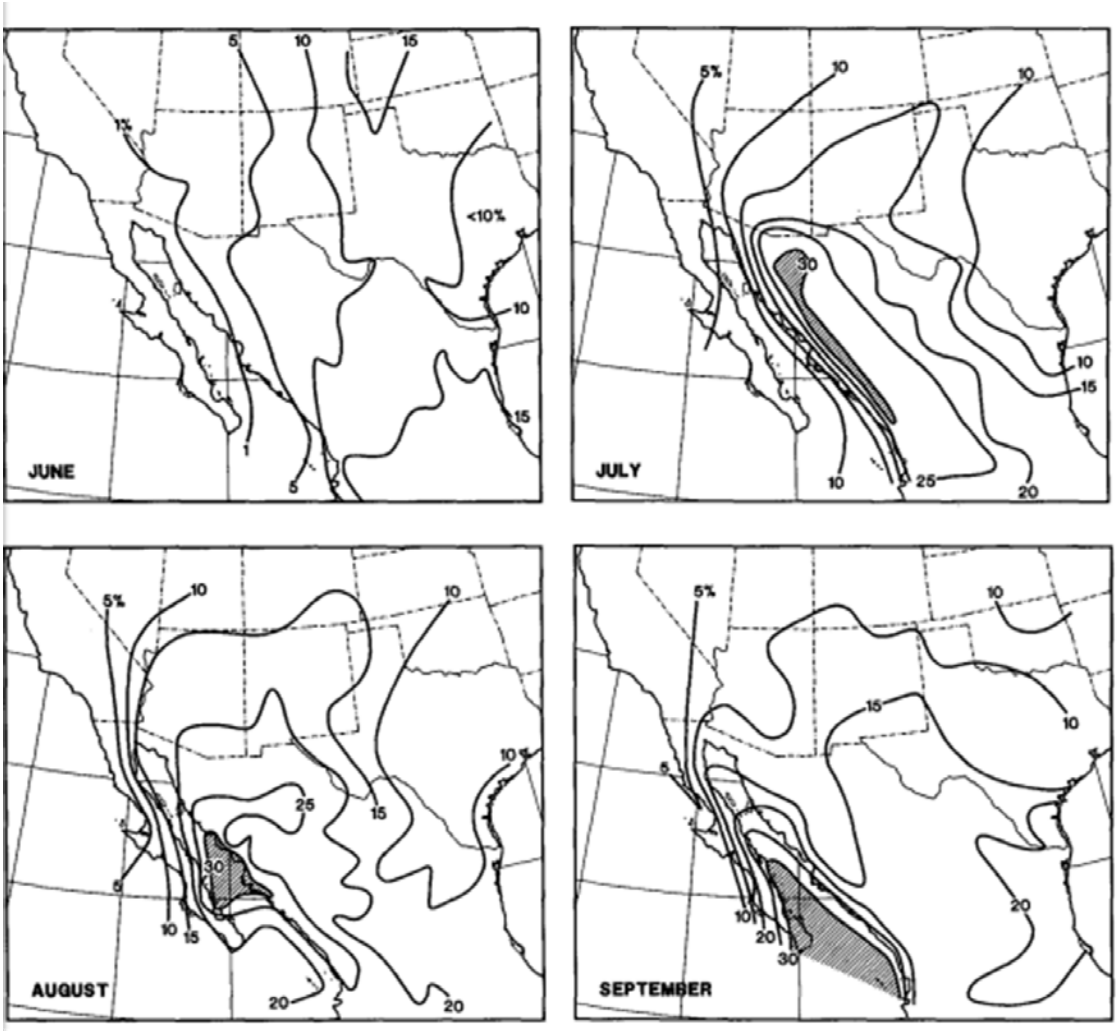


Fig. 2.1. The mean monthly rainfall and its contribution to annual rainfall (in percent) for the southwestern portion of the United States and northwest Mexico. Taken from Douglas et al. (1993).

This graphic illustrates that the monsoon is the primary contributor to annual precipitation in these regions, and the occurrence of this moisture is during mid-late boreal summer. This serves as a surrogate for identifying the spatial extent of the monsoon. That said, this is not to be taken as a strict translation of the NAM's geography as local effects can cause variability.

Changes in the large-scale general circulation patterns are the primary drivers of the NAM, however, changes in synoptic and mesoscale circulations combined with topographic influences can cause a seasonal variability in the spatial extend of the monsoon. Fig. 2.2 shows the variability of the monsoon spatially using time as the independent variable and precipitation as the dependent variable (Adams and Comrie 1997).

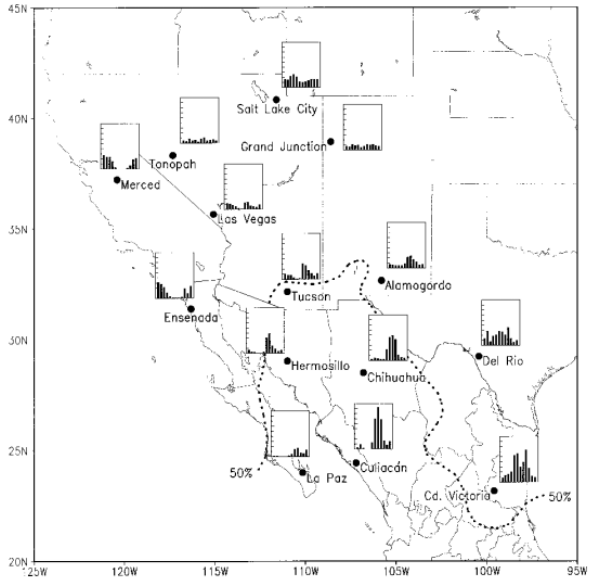


Fig. 2.2. Precipitation bar charts displaying the very prominent late summer spikes in precipitation throughout the southwestern United States and Mexico. Taken from Adams and Comrie (1997).

As shown in Fig. 2.2, there is an earlier peak to the precipitation histograms in Mexico versus the United States. Also, there is an overall decrease in precipitation intensity northward into the United States. The distribution of these histograms are bimodal, which is indicative of the change to a monsoon air mass. If these precipitation bar charts are used as a proxy for integrated precipitable water (IPW), then this would run counter to the typical lognormal distribution of IPW in nonmonsoon regions (Foster et al. 2006). In

summation, the findings of Adams and Comrie (1997) reiterates the importance of having a significant spatial extent when researching the monsoon in conjunction with an appropriate temporal extend as well. Additionally, this graphic fits with their criteria for the "true" monsoon circulation which includes:

1) Seasonal reversal of surface wind flow and pressure changes aided by a thermal low pressure developed in the lower Colorado-River valley.

2) Moisture advection into the region.

3) Two distinct precipitation seasons (wet and dry).

4) A temperature maximum before and after the wet season.

There is an important climatological distinction between "dry" monsoon and "wet" monsoon.

Dr. Ron Alberty in a letter dated May 27, 1986, wrote about the timing and distinction between the two sub-seasons of the monsoon. He explained, "Our dry monsoon is associated with hot dry westerly winds that blow on the north side of this high pressure belt…The dry monsoon is characterized by extreme heat…low relative humidity…and very little cloudiness." He went on to say, "The wet monsoon is associated with humid east to southeast winds on the southern side of the high pressure region. The growth of the high pressure belt causes the wind shift from west to east…and the change from dry to wet weather. This normally occurs in early July" (i.e., very little precipitation observed in May/June with noticeable precipitation increases in July/August/September on Fig. 2.2). The high pressure expansion that Alberty is referencing is that of the Bermuda High.

The Bermuda High is a semi-permanent high pressure system that is centered over Bermuda in the Atlantic Ocean during the summer months. Intense solar insolation during the summer months in the Northern Hemisphere causes the westward and northward expansion of the Bermuda High (Adams and Comrie 1997). The westward expansion of this high pressure causes mid-level anticyclonic rotation and gives the wind a south/southeast component on the eastern flank of the southwestern United States and northwest Mexico. This creates a mean wind shift (westerly-southeasterly) in the region. There is an evolution of this backing wind shift that generates the dry/wet season distinction. Fig. 2.3 shows this evolution of the Bermuda High's expansion in conjunction with the development of the Eastern Pacific Trough over the west coast of the United States. Assuming synoptic-scale geostrophic winds, the evolution of this counterclockwise wind shift is evident over the early summer months.

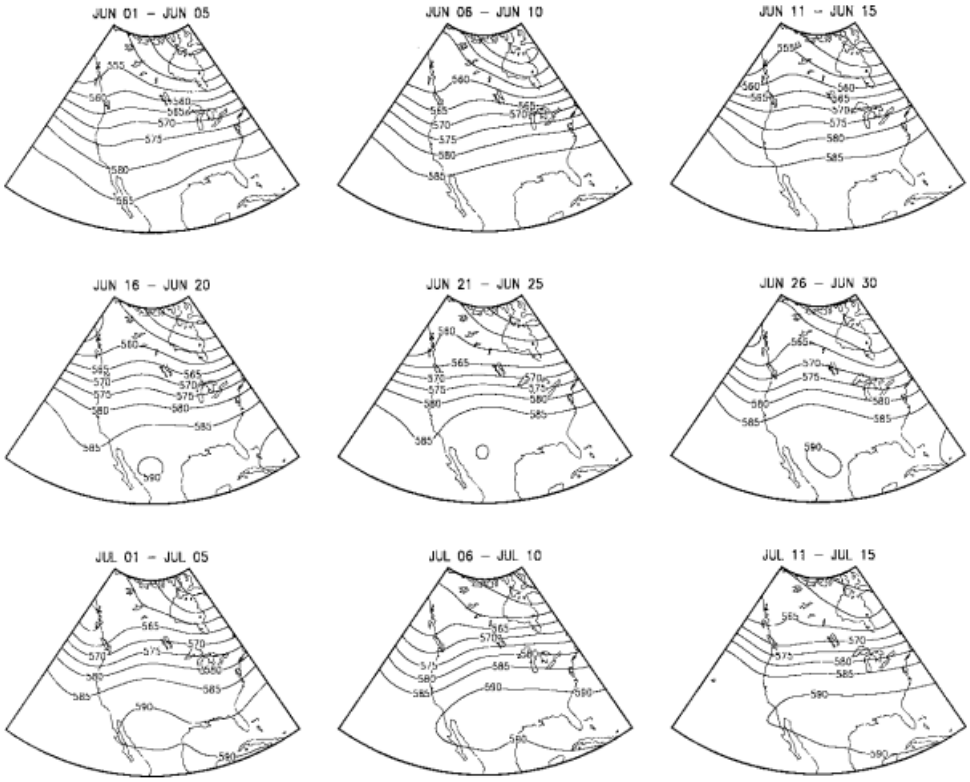


Fig. 2.3. 500-hPa height contours for North America from June 1st – July 15. The years are from 1978-1988. Taken from Adams and Comrie (1997).

According to Adams and Comrie (1997), Reed (1933) was the first to recognize this synoptic setup, 500hPa ridge axis over the top of the SMO, as the primary mechanism for moisture advection into southwest Arizona and northwest Mexico during the NAM. Erfani and Mitchell (2014), using reanalysis data, showed that the transport of humidity occurs in combination with a wind shift. Fig. 2.4 displays the time evolution of wind just to the west of the SMO (at 25ºN Latitude) and the associated transport of moisture.

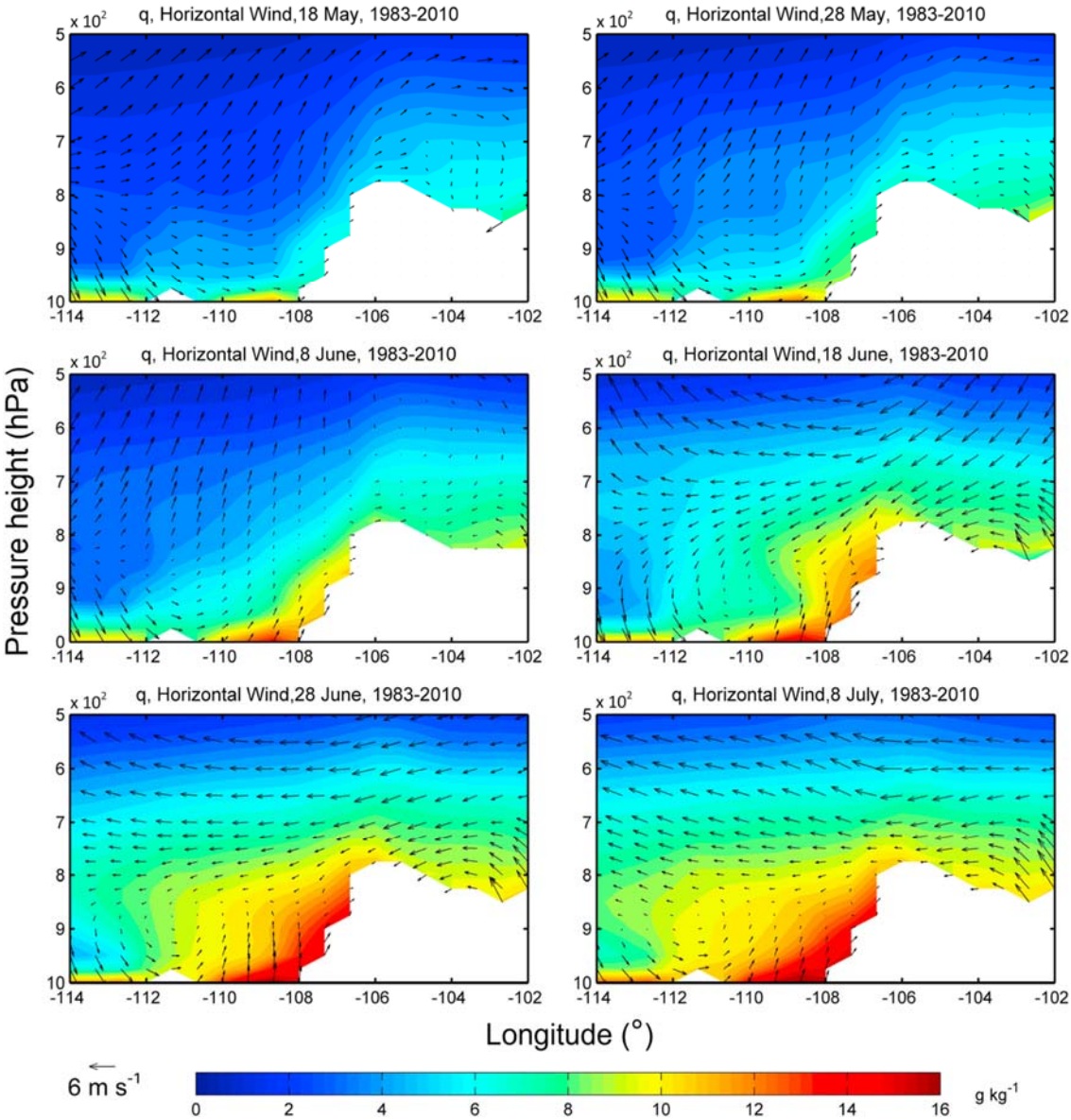


Fig. 2.4. Plotted horizontal wind vectors and moisture advection (color coded) averaged from 1983-2010 from 18 May – 8 July in 10 day intervals. The Sierra Madre Occidental (SMO) mountain range is shaded white. Taken from Erfani and Mitchell (2014).

Fig. 2.4 displays a longitudinal perspective on the NAM's wind shift and moisture advection while enhancing spatial situational awareness and thereby illustrates the critical shift in westerly to southerly wind on the western side of the SMO (Erfani and Mitchell

2014). Additionally, it suggests that the wind shift and moisture advection are related and provides some justification for the use of moisture as a proxy for the wind shift.

2.2.2 Moisture Criteria of the Monsoon

While climatological criteria for the NAM start have been developed for Tucson, AZ and Phoenix, AZ, no criteria for establishing monsoon onset has existed for Albuquerque, NM, El Paso, TX, nor Flagstaff, AZ offices (Ellis et al. 2004). In academic research, Guaymas, a city in southwest Sonora, Mexico has traditionally been the key location to identify the beginning of the NAM because of Guaymas' reliability in showing a quick rise in humidity with the associated monsoonal wind shift (Douglas et al. 1993). The origins of quantitative benchmarks for determining NAM onset in Arizona dates back to the mid-1960s.

In 1966, three meteorologists/climatologists in Phoenix, AZ, Messrs, Ingram and Kangieser, developed the dewpoint benchmark for identifying monsoon onset (Franjevic 2017). In doing so, they referenced a paper by Reitan (1963), which suggested that a 55°F (12.8°C) dewpoint was well-correlated with 1" (25.4 mm) of integrated precipitable water (IPW). According to the American Meteorological Society's *Glossary of Meteorology* (2017), "IPW is the total atmospheric water vapor contained in a vertical column of unit cross-sectional area extending between any two specified levels, commonly expressed in terms of the height to which that water substance would stand if completely condensed and collected in a vessel of the same unit cross section." The associated formula for IPW is:

$$IPW = \int_{p_1}^{p_2} \frac{1}{\rho_v g} x dp \tag{2.1}$$

Where,
g = acceleration do to gravity,
x = mixing ratio at pressure level p,
p_1 = surface pressure level,
p_2 = "top of atmosphere" pressure level,
ρ_v = density of water vapor

If the value of the IPW is close to 1" (25.4mm), in unification with the proper dynamic and thermodynamic forcing, significant moisture exists to initiate thunderstorm activity (Reitan 1957).

Reitan's (1963) paper laid the foundation for understanding the connection of surface dewpoint with IPW and the values associated with thunderstorm development during the NAM. Reitan used the assumption that vertical mixing of surface moisture would create a "uniform lapse rate" (p. 778) resulting in a monthly averaged constant decrease of moisture throughout a column of air. If this assumption is correct, then a surface measure of humidity (i.e., dewpoint) could be used to make an objective measurement of the IPW throughout the atmosphere without the need for radiosonde measurements. This is very useful in austere locations where upper-air data are limited, and direct IPW measurements cannot be made. Reitan's method included using mean-monthly dewpoints and IPW measurements at 15 locations across the United States each of which represent different climate regions to see if a correlation existed. The only station used in the NAM's region was Phoenix, AZ. The length of the study was three years (1954, 1955, 1956). The results showed an excellent linear correlation (r = [0.96 – 0.99]) for all stations with some mild standard deviation for some (Grand Junction, CO, Great Falls, ND, Phoenix, AZ). He adapted the fundamental definition of linear regression. Linear regression assumes a linear connection between two scalar variables.

Therefore, an equation can be developed and plotted over a scattergram that illustrates the degree of relationship between the two variables (Carruthers 1953). However, the linear relationship may only hold if one of the two variables is calculated using the natural logarithm. This is the statistical methodology employed by Reitan (1963) and his general equation is as follows:

$$\ln W(\text{cm}) = a + b\, T_d(\text{F}) \tag{2.2}$$

where,
$\ln W$(cm) = natural log of precipitable water in centimeters,
a, b = regression equation constants,
T_d(F) = dew point temperature in Fahrenheit

The regression results for Phoenix yielded the equation $\ln W(\text{cm}) = -0.981+0.0341 T_d(\text{F})$.

Using equation 2.2, an estimation of IPW can be made by measuring surface dew point. Reitan claimed that this equation was applicable for any place on Earth despite climate variability, however, he noted that "better equations for individual stations can be obtained…this regression equation can be used in all instances as an estimator of precipitable water from surface dew point when more precise equations are not available" (Reitan 1963 p. 778). The accuracy of this equation, according to Reitan, was excellent. The standard error of regression and the coefficient of variation of the standard error from regression was 0.18 and 10%, respectively. For Phoenix, AZ, the standard error of regression and coefficient of variation of the standard error from regression was 0.28 and 17%, respectively.

Shortly after Retian's (1963) work, Bolsenga (1964) used a similar linear regression relationship using mean daily measurements and hourly measurements. The

results displayed a very high correlation coefficient (r = 0.85 and r = 0.80, respectively). Smith (1966) modified Bolsenga's values by taking into account latitude and seasonality. By including latitude and seasonality, Smith showed that latitudinal differences in the moisture profile must play a role in the surface humidity and IPW relationship, and suggested the results of the Reitan and the Bolsenga regression equations may not be applicable in all situations and locations despite numerous studies, around the same time period, that have favored the results of Reitan and Bolsenga (e.g., Benwall 1965; Idso 1969). Moreover, Tuller (1977) identified that correlation coefficients between surface humidity parameters other than dew point (mixing ratio and water vapor pressure) and IPW were minimal. In summary, these initial studies indicate a consistent relationship between mean monthly surface humidity values and IPW.

These studies were fundamental in later studies abroad, where extensive upper-level data did not exist and, consequently, research relied on surface humidity measurements as the only measurement for IPW. Dew point/IPW relationships were successfully explored in West Africa (Ojo 1970; Anyadike 1979 Adedokun 1983; Adedokun 1986), over the Atlantic Ocean (Benwell 1965), using surface equivalent potential temperature in West Africa (Oduro-Afriyie 1992), India using 850hPa dew point temperature (Sinha and Sinha 1981), Nigeria and southern England (Abo 1975), Brazil using 850hPa dew point temperature (Rao et al. 1979) in central Saudi Arabia (Maghrabi and Al-Dajani 2013), Canada (Gueymard 1994), South Africa (McGee 1974), and Phoenix, AZ (Idso 1969; Skindlov 2007). These extensive studies building on Reitan's (1963) equation, appear to justify the 55°F (12.8°C) dew point NAM

benchmark. However, some studies have questioned the results of this linear relationship between surface humidity and IPW.

Schwarz (1968) used twice daily measurements (0000 UTC, 1200 UTC) of surface humidity and IPW for San Antonio, TX for a two-year period (1963-64). Schwartz used Smith's (1966) equation to estimate IPW which is defined as:

$$\ln U = \ln\left(\frac{\epsilon E_0}{g}\right) - \ln(\lambda + 1) + \left(\frac{\alpha t_d - \beta}{t_d + \gamma}\right) \ln 10 \qquad (2.3)$$

Where,
λ = Vertical moisture distribution based on latitude = 2.98,
ϵ = 0.622,
E_0 = 6108 (dyn cm-1),
g = gravity constant 979 (cm sec-2),
α = 7.5,
t_d = mean dew point temperature for 0000GMT 65.3°F,
= mean dew point temperature for 1200GMT 71.1°F,
β = 238.1°F,
γ = 395.1°F,
ln$U$ = precipitable water (cm)

The annual range of correlation coefficients for 0000 UTC measurements were (r=0.57 to 0.92) and (r=0.37 to 0.85) for 1963 and 1964, respectively. Interestingly, there was a large drop off for this relationship for 1200 UTC measurements. Negative correlations coefficients occurred in July 1963 (r = -0.18) and August 1964 (r = -0.02). Very low positive correlation coefficients occurred in June 1963 (r = 0.10), August 1963 (r = 0.37), July 1964 (r = 0.08), and September 1964 (r = 0.41). Schwarz (1968) attributed this variability to four different reasons:

1) Moisture flow is at low-levels.

2) A mechanism, either dynamic or thermodynamic, needs to be present for adequate vertical mixing. Dynamic system (i.e., fronts, low pressure systems) tend to remain north of San Antonio, TX during the summer, and solar heating has a greater impact on vertical moisture mixing at 0000 UTC than at 1200 UTC.

3) Lack of low-level moisture variability near San Antonio, TX during the summer months. (i.e., low horizontal moisture gradient).

4) Overnight cooling inhabits vertical transport of low-level moisture, therefore, this results in a shallow mixed layer and high surface moisture values that do not represent total moisture depth.

This finding, along with others (e.g., Chaboureau et al. 1998), illustrates the need for strong vertical mixing either thermodynamically or from a dynamic system in order for there to be a strong correlation between surface humidity measurements and IPW. Other studies have also determined that correlation coefficients between humidity and IPW that approach r = 1.00, regardless of location or seasons, are dependent on vertical mixing (Hay 1970; Karalis 1974; Tuller 1977; Revuelta et al. 1985). Altogether, this implies that Reitan's (1963) results may only apply in certain locations for certain times of the year under certain conditions. Other researchers question Reitan's (1963) results as well.

Reber and Swope (1972) used monthly and annual surface absolute humidity and IPW calculations using the Solot (1939) method for three stations in Southern California (San Nicolas Island, Point Mugu, and China Lake) to examine the findings of Reitan (1963), Bolsenga (1964), and Benwell (1965), specifically. The Solot method required the use of three empirically determine humidity and IPW parameters:

$$\text{absolute humidity (gm m}^{-3}) = 217e/T \tag{2.4}$$

$$\text{specific humidity (gm kg}^{-1}) = 622e/(\text{P}-0.378e) \quad (2.5)$$

$$\text{precipitable water (mm)} = 0.00508(\Delta P)(SM_1+SM_2) \quad (2.6)$$

where,
e = water vapor pressure (hPa) calculated by using the Smithsonian Meteorological Tables (List 1963 p. 350),
P = pressure in (hPa),
T = Temperature in (K),
ΔP = difference in pressure between two neighboring pressure levels,
SM_1 & SM_2 = specific humidity's between two neighboring pressure levels

Reber and Swope (1972) took into account both day and night measurements and measurements from various months in the year (January, July, April, October), and found very little correlation between humidity and IPW. Their results varied greatly from the studies like Reitan's and the others that supported his surface humidity/IPW relationship, with monthly correlation coefficient ranging from (r = -0.29 to 0.83), and annual correlation coefficients (r = 0.53 to 0.66). The lowest correlation coefficients were found during the summertime for San Nicholas Island and Point Mugu, and a negative correlation coefficient during December for China Lake.

Reber and Swope (1972) attributed the divergence of results between Reitan's (1963) and their results to "data smoothing by computing the monthly averages" (p. 1324). Reber and Swope (1972) later stated "When radiosonde data are smoothed by monthly averaging to form mean monthly atmospheric models, the relationship between total precipitable water and surface absolute humidity is statistically enhanced; the resulting correlation coefficients, which approach 1.00, are artificial…estimates of precipitable water surface measurements, on an individual measurement basis, are not

sufficiently reliable to justify making surface measurements to infer existing precipitable water" (p. 1325). Reber and Swope's (1972) result combined with those of other studies (e.g., Schwarz 1968; Lowry and Glahn 1969; Hay 1970; Karalis 1974; Tuller 1977, Viswanadham 1981) suggests that Reitan's (1963), Bolsenga's (1964), and Smith's (1966) conclusions on the relationship of surface humidity and IPW are highly suspect. This is an important point because the theoretical justification for using a surface dew point of 55°F (12.8°C) for three days as the NAM's onset in Phoenix, AZ was largely based on Reitan's (1963) paper.

Using the Messrs, Ingram and Kangieser dew point criterion, the onset of the monsoon has been found to vary widely from year-to-year. Climatologically, the earliest onset for the monsoon in Phoenix, AZ has been 19 June with average start of 7 July. The latest onset occurring on 25 July (Ellis et al. 2004). Although, climatologically consistent (i.e., the monsoon occurs every year) the exact start and end dates, and spatial extent are highly variable (Higgins 1999).

Ellis et al. (2004) argued that it would be more reasonable to adopt a method for identifying the monsoon regionally rather than at a point location (i.e., Phoenix, AZ and Tucson) as this would better represent the synoptic-scale temporal and spatial effects of the monsoon. The method developed was a combination of three successive days of meeting thresholds of surface dew point 50°F (10°C) and at least 20% of the 193 weather stations located around Arizona, New Mexico, and west Texas measuring precipitation. Both of these benchmarks defined by Ellis et al. (2004), were calculated on a 52-year climatological average prior to the publication of this study. The average daily dew point

for this period of record from 15 June – 15 October for the NAM's region and the median of number of stations reporting precipitation were for the aforementioned benchmarks (i.e., 50°F and 20%). Although not used in the Ellis et al. (2004) calculation of onset, the median values of daily mean dew point for Tucson, AZ was 53.88°F (10.13°C) and 54.50°F (12.50°C) for Phoenix, AZ. Lu et al. (2009) provided further justification for this method.

Similarly, Lu et al. (2009) determined that moisture (i.e., dew point, IPW) increases from winter and summer naturally, so it is necessary to include precipitation when determining onset/demise. These researchers advanced the idea that IPW and precipitation are both needed to determine onset/demise. As a result of surface heating, precipitation increases in monsoon regions from winter to summer and can be used to signal monsoon onset and retreat. This pattern does not occur in nonmonsoon regions as yearly precipitation is more consistent. Conversely, IPW increases in both monsoon and nonmoonsson regions from winter to summer. Therefore, these researchers asserted that proper onset detection must include the conflation of precipitation measurements and IPW measurements.

These methods, however, can cause various problems because of measurement issues with these variables. For example, precipitation measurements for monsoon onset are impractical because of the impossibility of continuous spatial coverage of rain gauge measurements (Means 2013) and the inherent spatial variability of monsoonal precipitation (Diem and Brown 2006). Additionally, IPW values are problematic in that

they are susceptible to the same elevation-related problems as dew point (Zhen and Lu 2004; Means 2013).

Consequently, Zhen and Lu (2004), proposed the Normalized Precipitable Water Index (*NPWI*), which accounts for elevation in IPW calculations.

$$NPWI = \frac{PW - PW_{min}}{PW_{max} - PW_{min}} \tag{2.7}$$

Where,
PW = Precipitable water,
PW_{min} = Daily minimum averaged precipitable water at a location,
PW_{max} = Daily maximum averaged precipitable water at a location,

This method altogether eliminates topographic modifications on IPW. Means (2013), used the calculation of the *NPWI* in combination with Global Positioning System (GPS) derived IPW values to calculate onset of the monsoon for Southern California (across the Imperial Valley) and southern Nevada. The results showed the hallmark bimodal distribution of IPW of a monsoon region (Foster et al. 2006), with a range of onset (Julian) days between 188 – 191 (~7 July – 10 July) for the seven year period of study. This result is similar to that of the onset of the monsoon for Phoenix, AZ. However, none of these methods has been used as official forecasting tools by the NWS.

Similar controversy exists with the determination of for the onset of the Indian Monsoon. The start of the Indian Monsoon is determined subjectively by weather forecasters in Kerala, one of India's southern states. Once the start of the monsoon has been determined in this region, it is considered the official start of India's rainy season.

There are five subjective considerations taken by a local Kerala forecasters (Wang et al. 2009).

1) Widespread precipitation over Kerala and surrounding providences, with "large" amounts at individual stations
2) Persistent rainfall for several days
3) Low-level westerly winds
4) Deep moisture up to 500hPa

These subjective methods are still in use today (Noska and Misra 2016). However, these subjective methods are not without complications as seen in onset of the 2002 Indian Monsoon in which the anomalous rainfall caused a pseudo-monsoon that forecasters errantly declared the start of the rainy season. Once this rain passed, there was a two-week lull in precipitation before the onset of the true monsoon occurred (Flatau et al. 2003).

It is apparent that the establishment of onset/demise dates have been often confusing and highly variable. Discussions on quantitative methods for marking the start of the rainy season continue, however, the NWS continues to use predetermined set start/end dates of the monsoon. Consequently, NWS forecasters continue to monitor surface dew point and IPW in order to forecast storm activity (Favors and Abatzoglou 2013; Moore et al. 2015). The origins of this necessary moisture plays a major role in the monsoon, however, the source of monsoon moisture has historically been debatable.

2.2.3 Moisture sources/Gulf Surges

There has been a long-standing debate among researchers regarding the moisture source(s) of the NAM. Originally, the Gulf of Mexico was thought to be the primary source of all moisture for the NAM (e.g., Jurwitz 1953; Bryson and Lowry 1955; Reitan 1957) Paradoxically, Reitan (1957) asserted that most monsoonal moisture was in the lower atmosphere (50% below 800hPa and 86% below 600hPa). This does not seem feasible considering the moisture from the Gulf of Mexico would have to transverse up and over the ~10,000' (3048m) SMO without being subject to orographic precipitation on the windward side (Brenner 1974; Hales 1974).

Hales (1972, 1974) was the first to challenge this notion of the total Gulf of Mexico monsoonal moisture origination and advance the idea that the primary moisture source was the Gulf of California and that the Gulf of Mexico could play a role in providing upper-level moisture. Hales (1972) examined case studies of low-level moisture flows in cities east of the Gulf of California (i.e., Guaymas, MX Puerto Penasco, MX) into southern Arizona (i.e., Yuma, AZ, Tucson, AZ), southeastern California (i.e., Blythe, CA, Needles, CA), and Nevada (i.e., Las Vegas, NV). He discovered these anomalous events were associated with rapid spikes in dew point temperature and pressure, and drops in surface temperature. He also stated that these influxes of moisture were readily associated with convective thunderstorm activity in these areas. This suggests that the NAM is largely influenced by the Gulf of California and that the Gulf of Mexico is not the primary moisture source. According to Dixon (2005), researchers have since backed this idea of multiple moisture sources to also

include the Pacific Ocean (e.g., Brenner 1974; Adang and Gall 1989; Smith and Gall 1989; Carleton et al. 1990; Harington et al. 1992; Stensrud 1995; Schmitz and Mullen 1996; and Adams and Comrie 1997). This rapid influx of cool moist tropical air into southern Arizona along the Gulf of California is termed a "Gulf Surge" (Hales 1972; Brenner 1974; Hales 1974). Since the development of Gulf Surge theory, researchers have developed several methods for detection of these events.

Hales (1972) outlined the characteristics of surge events that were used as an early method for surge detection. The six characteristics are as follows:

1) Surges are most intense near the surface and intensity is lost with height A statewide (Arizona) drop in temperature accompanies a surge with the greatest cooling occurring along the Colorado River Valley.

2) 24 hour changes in temperature, pressure, and dew point are the only way to detect surges in northern Arizona.

3) Certain surge types will cause a statewide increase in storm activity.

4) The surge is strongest initially and then decreases in magnitude over time.

5) In the southern and western deserts of Arizona a maximum drop in temperature of 5°C is possible after a surge. Everywhere else in Arizona a drop in temperature of 3°C - 5°C is possible.

Using the foundation laid out by Hales (1972), other methods for identifying a surge event also have been established (e.g., Brenner 1974; Fuller and Stensrud 2000; Dixon 2005). For example, Brenner (1974), expanding on Hales' work, pointed out the surges in Yuma, AZ are often preceded by jumps in sea-level pressure, shifts to a southeasterly wind, reduced visibility from blowing dust, and jumps in dew point greater than 10°C in a half-hour as the surge passes. Fuller and Stensrud (2000) developed a detection method

in which the daily maximum surface dew point temperature for Yuma, AZ stayed above 60°F (15.7°C) in conjunction with a spike in surface wind speed greater than 4 ms^{-1}. Lastly, Dixon (2005) developed a methodology that he termed "Assessing Low-Level Atmospheric Moisture using Sounding (ALAMS)." This method defines a surge event such that, if over the course of a four-day period, the last two days show a one-kilometer-high dew point temperature increase of 4°C or more in comparison to the preceding two days, then that would indicate a surge. Having a proper detection method for surge events is vital, as this influx of low-level moisture, generally within 700hPa (Adams and Comrie 1997), promotes atmospheric instability and is a precursor to severe weather events in Arizona during the NAM (McCollum et al. 1995).

The importance of Gulf Surges with respect to precipitation during the NAM cannot be overstated. Monsoon precipitation has been shown to have a large dependence on Gulf Surges (e.g., Berbery and Fox-Rabinovitz 2003; Dixon 2005; Becker and Berry 2008; Favors and Abatzoglou 2012). Svoma (2010) found that this dependence was more prominent for drier monsoon areas around central and southern Arizona, including Phoenix, AZ, and was less apparent in the traditionally wetter eastern portions of Arizona. Higgins et al. (2004) showed an increase in wetness for southern Arizona and northwest Mexico in the aftermath of Gulf Surge, where surges detected in Yuma, AZ were associated with a 66% of rainfall in Arizona and western New Mexico while surges in Tucson, AZ were related to 38%. Pascale and Bordoni (2016) found higher percentages, stating that 70% - 80% of the mean rainfall in Arizona and western New

Mexico during the NAM was in direct correlation to Gulf Surge events, and influenced 18 precipitation events annually.

There are a variety of reasons for why these surge events occur. Hales (1974) believed that a large-scale sea breeze, caused by differential heating of the Pacific Ocean and Gulf of California compared to the surrounding desert, drove southerly winds and pushed moisture into southern Arizona. This phenomenon was not observed by Hales during the winter months. Brenner (1974) suggested that a surge may be caused by an outflow boundary moving up the Gulf of California. The development of a thermal-low in southeastern Arizona is thought to add in drawing up moisture Additional work by Strensrud et al. (1997), examined the role of mid-latitude trough passage preceding an easterly wave passage on Gulf Surges. They found that if these two troughs are out of phase, "a strong surge is likely… However, if the two troughs pass through at the same time, then the effects of the mid-latitude system likely can stop the surge from reaching Arizona and may even prevent a surge from developing at all." (p. 435). Fig. 2.5 shows exactly this setup by Hales (1974) and that of Strensrud et al. (1997) both may trigger a surge event. Adams and Comrie (1997) also stated that minor surges could be caused by backdoor cold fronts, tropical cyclones passing the mouth of the Gulf, and the presence of Mesoscale Convective Complexes west of the SMO.

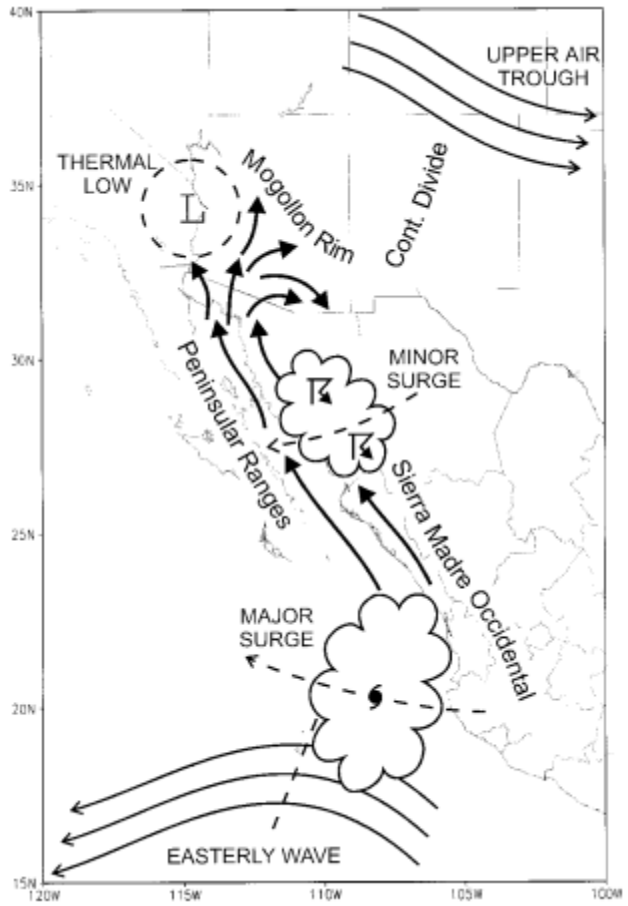


Fig. 2.5. illustrates the various mechanisms that cause major/minor Gulf Surges. Taken from Adams and Comrie (1997).

Recent research has also suggested additional mechanisms by which surge events in the Gulf of California are enhanced. Low-level jets are thought to advect moisture from the eastern Pacific and the Gulf of Mexico northward into Arizona (Anderson et al. 2002). For example, Rogers and Johnson (2007) described the role of nocturnal low-level jets (LLJ) over the Gulf of California as surge amplifiers in the presence of surges induced by Mesoscale Convective System (MCS) outflows. Similarly, Mejia et al. (2016) expanded on this work by identifying that the mere presence of MCS activity in the

southern portion of the Gulf of California and Eastern Pacific can enhance Gulf Surges into Arizona irrespective of the synoptic perpetrator of the surge.

Since the early 2000s, many researchers have identified various synoptic features that can spawn a Gulf Surge. Fuller and Stensrud (2000) suggest that 75% of Gulf Surges may be the result of a passing tropical easterly wave (TEW), and subsequently when other mechanisms do create the surge, the presence of TEW can enhance the surge and have increased intensity on rainfall in southern Arizona and northwest Mexico west of SMO (Adams and Stensrud 2007). Eastern Pacific tropical cyclones (EPTC) are also consider to be a Gulf Surge source (Higgins and Shi 2005), with reported spikes in summertime precipitation spanning from central New Mexico, southern California, Arizona, southern Nevada, and Baja California as a result of this moisture advection and lingering low-level vorticity from remnant (EPTC) creating a deep convective environment (Corbosiero et al. 2009).

These low-level surge events are major components to the NAM and are directly related to precipitation throughout the core regions of the NAM (Favors and Abatzoglou 2012). These surges can spawn spikes in dew point as well as precipitation. However, because these are low-level events, surges can raise surface dew points while simultaneously not providing significant moisture in terms of IPW. Vertical mixing of surface moisture can be accomplished in two ways mechanically or thermodynamically with more intense insolation causing a greater depth in the mixed layer (Stull 2009). Consequently, Schwarz (1968) showed that in the case of San Antonio, TX, a moist boundary layer in the absence of any dynamic forcing, created an environment in which

surface dew point did not correlate well with IPW. Additionally, in the case of Reber and Swope (1972) where temperature inversions likely inhibited deep mixing, the correlation between surface dew point and IPW breaks down. Coincidently, desert boundary layers in mountainous areas (e.g., Arizona, New Mexico, Mexico, southeastern California, Nevada) can become detached from the surface as a result of an elevated mixed layer that is pushed off a high terrain that forms a stable environment and inhibits thermodynamic mixing of surface moisture (Whiteman 2000; Warner 2004; Stull 2009). Similarly, these capping inversions can occur over the Gulf of California when sea-surface temperatures are below 29°C. (Erfani and Mitchell 2014). The relationship likely holds only in idealized conditions and may not be applicable to the desert Southwest and northwest Mexico during the NAM. In order to examine this relationship appropriately, North American Regional Reanalysis (NARR) data is the most appropriate data source.

2.3 North American Regional Reanalysis (NARR)

The North American Regional Reanalysis (NARR) project was undertaken as a follow-on to the National Centers for Environmental Protection Global Reanalysis 1&2 (NCEP-GR, NCEP-GR2) model (Kistler et al. 2001). The NARR is a continuously updated system (initial completion in 2004), which provides a source of reliable weather data for North America beginning in 1979 (Mesinger et al. 2006). The purpose of this project was meant to, "Help answer questions about the variability of water in weather and climate, in particular as it concerns U.S. precipitation patters… (with) a good representation of extreme events, such as floods and droughts" (Mesinger et al. 2006 p. 344). The NARR inputs are a conflation of various data sources to include: Eta model

outputs, rawinsodes, pilot balloons, dropsondes, aircraft data, surface observations and cloud drift. (Shafran et al. 2004). This source can be preferable to direct rawisounde/radiosonde data because these data sets historically are convoluted with measurement practice modification, site-relocations, and balloon drift that can create inherent inconsistencies in the data (Bosart 1990; McGrath 2006). Ideally, NARR would help elevate these problems. Altogether this dynamic, high-resolution reanalysis model, was shown to be a major improvement over its predecessors.

The NARR's performance in comparison to previous reanalysis models was initially studied by Mesinger et al. (2006) by describing the fundamental differences between models. First, the NARR has finer grid spacing than its NCEP-GR predecessor (32-km versus 80-km spacing, respectively). Secondly, the NARR has a greater temporal resolution (3-hr versus 6-hr, respectively). Lastly, the NARR develops data through 29 vertical pressure levels in the atmosphere in (hPa): specifically, 1000, 975, 950, 925, 900, 875, 850, 825, 800, 775, 750, 725, 700, 650, 600, 550, 500, 450, 400, 350, 300, 275, 250, 225, 200, 175, 150, 125, 100) (NOAA.gov 2017).

Mesinger et al. (2006) analyzed model performance between the NARR and NCEP-GR/GR2 with comparisons to real world observations for various variables. Fig. 2.6 represents the results of the NARR in model output to observed monthly precipitation values.

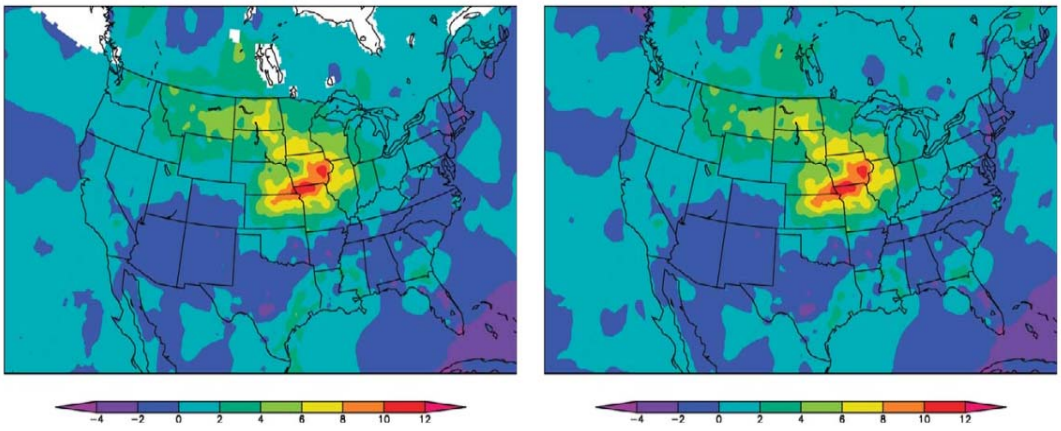


Fig. 2.6. On the left the observed monthly precipitation value differences for 1988 and 1993 for the months of June-July compared to the NARR model output on the right forthe same time period units are (month^{-1}). Taken from Mesinger et al. (2006).

The model output captures precipitation, especially, for the NAMs region of influence. The high accuracy of the model is likely a result of using latent heat flux rather than direct measurement of precipitation in data assimilation (Lin et al. 1999). This was important in resolving issues with matching NARR model output to observed rainfall in the geographically complex Western United States (Mesinger et al. 2006). Therefore, these researchers conclude that precipitation data through reanalysis is high in the NAMs domain and useful to helpful in answering my central thesis question.

The usefulness of NARR precipitation data when compared to other reanalysis models (e.g., CPC, NCEP-DOE, ERA-40) was also found (Burkovsky and Karoly 2007). However, there are some caveats. Burkovsky and Karoly (2007) found that the inaccuracies arouse with NARR properly representing spatial precipitation patterns on the fringes of the United States. Fig. 2.7 is an example of this occurring during a monsoon

storm event in 1999.

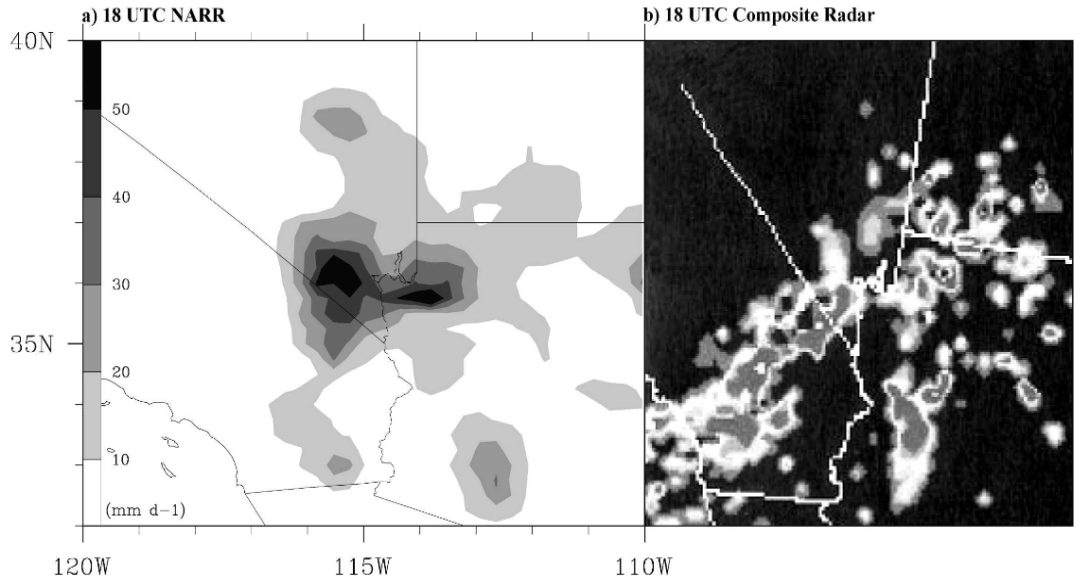


Fig. 2.7. Depiction of an extreme 1800 UTC weather event 8 July 1999 using NARR and Composite Radar. The left shows the 3-hour average precipitation rate derived from the NARR for the storm event. The left graphic is the composite radar (right) of that same event. Graphic taken from Burkovsky and Karoly (2007).

Fig. 2.7 shows the NARR did not capture the exact location of this particular storm event. This phenomenon is considered to be by-product of limited data assimilation in Mexico, Canada, and the Oceans. The other two variables that are important to my study are surface dew point and IPW. The NARR relies on proper hindcasting of rawinsondes as a function of pressure as a means of calculating these two variables.

Mesinger and colleagues (2006) compared the NARR to NCEP-GR to compare the ability of the model to yield rawinsondes data. Fig. 2.8 shows the aspect of this comparison.

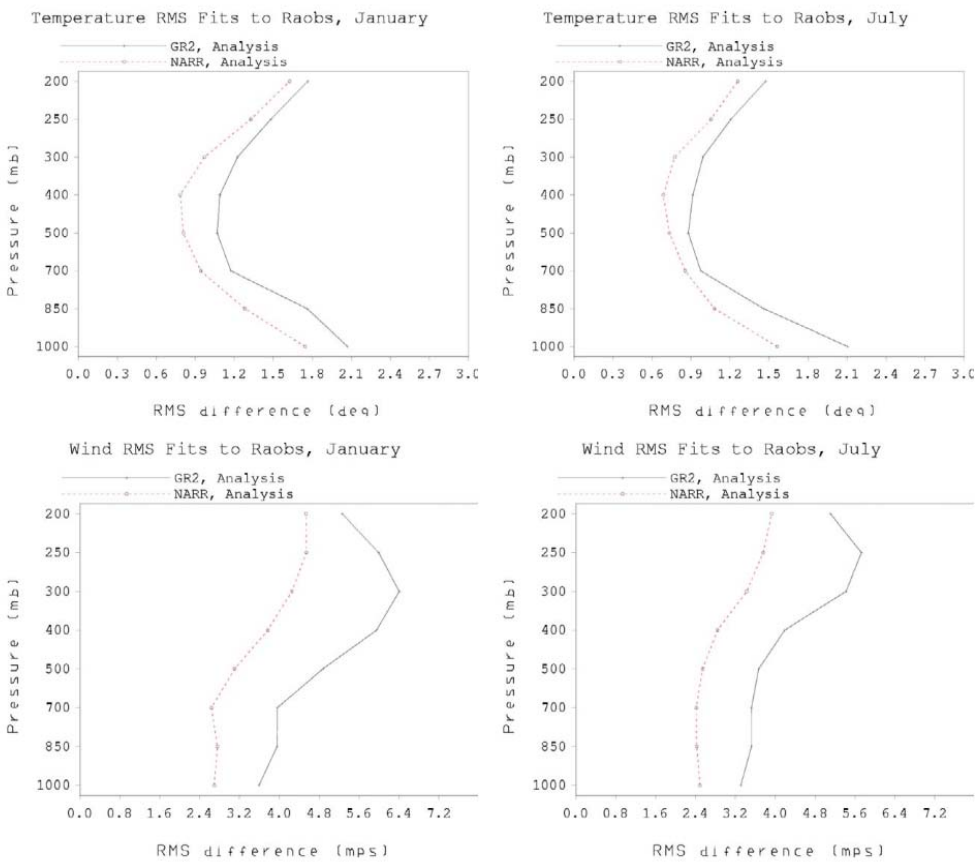


Fig. 2.8. An analysis of rawinsondes of both temperature (Top) and wind vector (Bottom) as a function of pressure versus Root Mean Square (RMS). The NARR is represented by the dashed line and the NCEP-GR is represented by the solid line. Taken from Mesinger et al. (2006).

Mesinger et al. (2006) showed that the NARR was better at reproducing of a rawinsonde than GR. As this is important in the development of producing IPW and surface dew point data, the NARR appears to be a reasonable choice to use in comparing surface dew point and integrated precipitable water. However, Mesinger et al. (2006) stated, "There have also been a few weaknesses found that require understanding their origin. The most conspicuous of these is the systematic excessive strength of the Gulf of California low-level jet in summer (Mo et al. 2005a), with large differences compared to various observational evidence over the northern Gulf of California" (p. 357). As the

low-level jet, located near the Gulf of California, is heavily associated with Gulf Surges into the NAMs domain, this may call into question the efficacy of using the NARR for my research project and needs further discussion (Kanamitsu and Mo 2003; Tian et al. 2004)

The findings of Mo et al. (2005a) showed that there is an overestimation of the low-level jet over California, specifically, located on the northern portion of the Gulf of California. Their study was a comparison between reanalysis model output and satellite derived observations. As a result of their comparisons, they discovered that reanalysis data "systematically" (p. 727) displayed an overestimation in the low-level jet especially in the meridional wind component, which is the main component for moisture advection from the Gulf of California into the NAMs area. The result of this is an overestimation of water vapor transport into the region (Fig. 2.9).

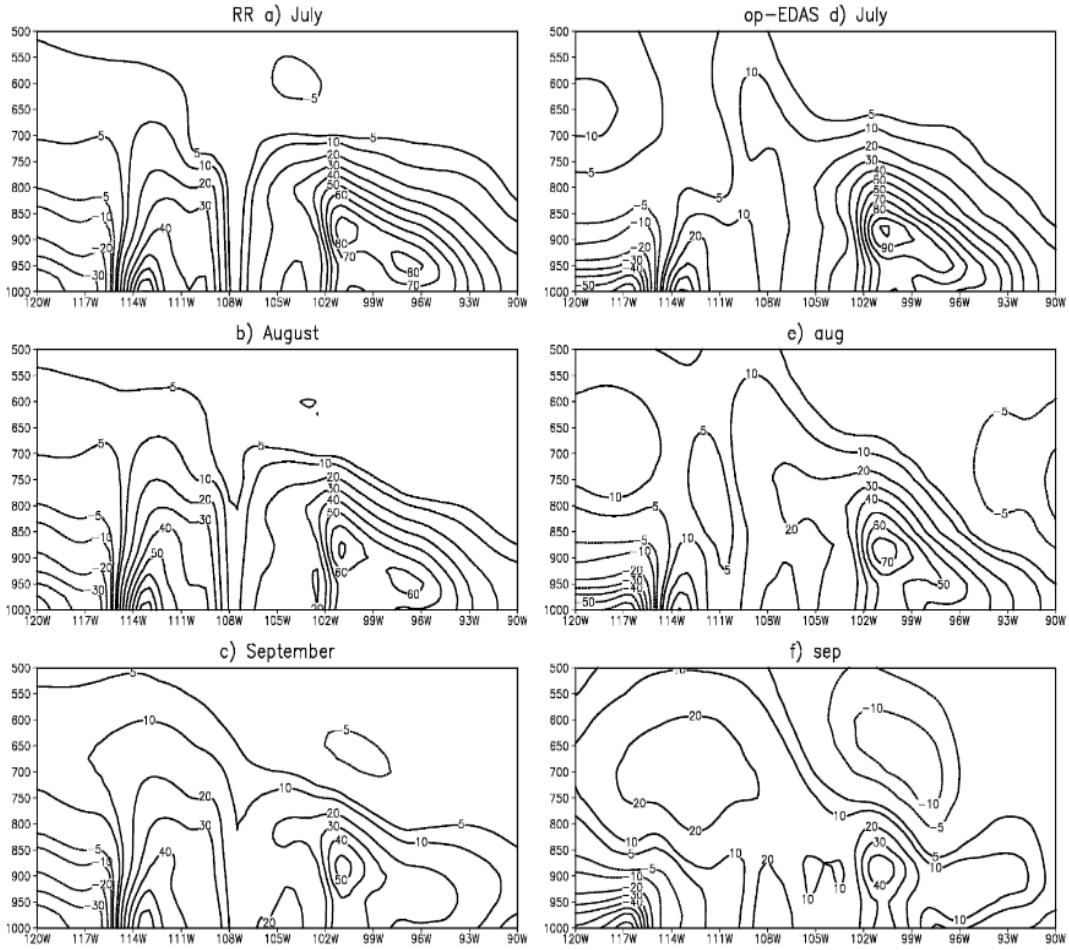


Fig. 2.9 shows a comparison between horizontal moisture flux reanalysis data (left) and observation (right). The contour lines are of moisture flux in units of 10 $gkg^{-1}ms^{-1}$. Taken from Mo et al. (2005a).

The differences between these are fairly noticeable horizontal moisture flux reanalysis data and observation are fairly noticeable in Fig. 2.9 First, there is an increase at the surface moisture advection located at 108°W (roughly the longitude for Tucson) and 114°W (roughly the longitude for Yuma) for all three months when compared to observation. Later, research created a similar finding to this study, but the effects are negligible of longer timescales (Becker and Berbery 2008; Ruane 2010; Radhakrisha et al. 2015). Mo et al. (2005a) explain these inaccuracies as the limitation of surface and

upper-data in this region that is not being assimilated into the reanalysis data. Similar results showed that the NARR represented low-level jets well outside of the NAMs region, and it did not misrrepresent their frequency nor moisture transport (Berg et al. 2015). This means that the use of NARR in the context of moisture advection is better captured away from the boundaries of the United States in less topographically challenging regions. This could be problematic for using reanalysis data.

In summation, according to Schiffer and Nesbitt (2011), "NARR moisture budgets may suffer because of 1) overactive modeled convection leading to overestimates of MFC over the core monsoon region in Mexico (Becker and Berbery 2008), 2) differences between model- resolved terrain and real terrain where assimilated precipitation is observed (Ruane 2010), and 3) discontinuities in the precipitation datasets used in the assimilation scheme along the U.S.–Mexico border (Mo et al. 2005a)" (p. 4220). In their study, they examined using NARR data for Gulf Surge identification. In their conclusion they stated, "These results come with the caveat that the NARR tends to overestimate the up-gulf moisture flux related to the Gulf of California low-level jet (Mo et al. 2005b)" (p. 4239). Nevertheless, (Radhakrishna et al. 2015) have shown that IPW over the NAM during the summer on a diurnal scale is not reliable, but there is no current research to suggests its unreliability on the sub-weekly scale or greater.

2.4 Summary and Discussion

Weather during the North American Monsoon can be hazardous and pose numerous threats. Seasonal afternoon thunderstorms, in Arizona, southern Nevada, west Texas, New Mexico, southeastern California, and northwest Mexico can bring flash

flooding, damaging winds, lightning, and hail (Adams and Comrie 1997). The storms are the byproduct of a wind shift that occurs in mid-late summer. This wind shift is formed due to summertime solar heating leading to the westward and northward expansion of the anti-cyclonically rotating Bermuda High and the simultaneous creation of a cyclonically rotating thermal low in the Colorado River Valley. Together, these shift the winds from a dry westerly wind to a moist southeasterly wind that promotes moisture movement into this otherwise dry area. This new moisture creates atmospheric instability and is the catalysis for thunderstorm genesis during the NAM. While these storms can pose numerous societal threats, they are vital to sustain life as these areas are typically dependent on monsoon moisture for yearly water needs. Much of the research has been centered on Arizona as a result of the large population centers and the economic ramifications of severe weather in that area, however, as discussed the NAM's area spans across all the states in the southwestern United States and northwest Mexico.

For nearly 40 years, forecasters established monsoon onset in central/southern Arizona by having three consecutive days at or above a certain surface dew point threshold. This was based, primarily, on a study that linked surface dew point to integrated precipitable water (IPW) whereas 1" (24.5 mm) of IPW was considered sufficient for storm formation. While substantial research supported this linkage, other research called it into question. These researchers stated that only in the presence of certain conditions that promoted vertical mixing of low-level moisture, may surface humidity parameters be used as a substitution for IPW. Fundamentally, the literature suggests that such an accepted relationship, long used to forecast severe monsoon

weather, should be reexamined to see if there is a correlation between surface dew point and IPW to see if the proper vertical mixing exists to well mix the low-level moisture fed into the region predominantly by the Gulf of California.

NAM researchers believed that the Gulf of Mexico was the primary moisture source. That contention has since been disputed with the earliest work arising in the 1970s and further supported by contemporary researchers. Today, researchers assert that the primary moisture sources for the NAM include the Gulf of California, Pacific Ocean, and the Gulf of Mexico. Establishing the monsoon's moisture source is vital to my study because low-level moisture, such as from the Gulf of California, is more difficult to vertically mix. Consequently, even in the presence of thermodynamic and dynamic forcing, the relationship of IPW and surface humidity could break down in the NAM regions.

The North American Regional Reanalysis (NARR) dataset provides an extensive library of data that can be spatially and temporally analyzed for this study. Research has shown this to be a very useful dataset even in comparison to direct measurement. The research regarding the NARR showed that it represented precipitation patterns and IPW well. Some limitations in its daily representation of low-level moisture exists for the NAM region, although, through timescale averaging any misrepresentations of the low-level moisture should be smoothed out, while validity testing will be conducted in order to ensure the usefulness of this data. As a result, the literature provided reasoning for using the NARR data set for this study and provided a rationalization for the selection of certain data locations.

This chapter has aided in the theoretical justification for my selected research topic, validation for the data points used, and the research timeframe. Specifically, this review has established the need for good quality weather data over a long time period from which I can examine the IPW/dew point relationship and the domain of the NAM area. As detailed in the next chapter, I will use the North American Regional Reanalysis (NARR) dataset as a source of establishing a quality network of data points. For my analysis, I have selected specific locations including Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Guaymas, MX Las Vegas, NV, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ, that I will use to compare surface dew point in these locations to IPW from 1979-2015.

Chapter 3: Data

3.1 Introduction

In order to examine my fundamental research question involving the relationship between integrated precipitable water (IPW) and surface dew point, I have established in Chapter 2 the need for good quality weather data over a long time period for the domain of the North American Monsoon (NAM). I have also shown that the North American Regional Reanalysis (NARR) dataset has been used in various other NAM studies that demonstrate the data yielded by the NARR are high quality. Additionally, in studies that compare the accuracy and reliability of NARR data to other reanalysis datasets, the NARR has proven superior (e.g., Burkovsky and Karlovy 2007). Furthermore, the NARR has shown advantages over direct measurements; direct radiosonde data over very long timescales has been questionable as changes in measurement protocols occurred over time, thus, creating reliability issues (Bosart 1990). In this chapter, I will use the NARR dataset as my primary source of establishing a quality network of data points. However, as I discussed in Chapter 2, there are some problems with the use of the NARR.

The NARR is a reanalysis dataset and is not composed of direct measurements. Reanalysis data are created essentially by a climate model that integrates observational data from a variety of sources, gridded to a uniform network and run for a $t = 0$ simulation. As a result, certain problems have arisen using NARR data, specific to the NAM's domain. For example, Mo et al. (2005a) showed the embellishment of low-level moisture during Gulf Surges by the NARR. Additionally, the NARR also has poor simulated moisture over the western United States during the summer months on very

short timescales (Radhakrisha et al. 2015). While Burkovsky and Karlovy (2007) described a high accuracy by the NARR when illustrating precipitation patterns in the central portions of the United States, they also documented cases of minor spatial variances in composite reflectivity with actual Doppler-radar. This reiterates the limitations of the NARR, which is dependent on having an extensive network of surface and upper-air data that does not exist on the fringes of the Unites States. While these issues are problematic, especially with respect moisture flux convergence during diurnal timescales in the NAM region, they are considered negligible over long time periods (i.e., monthly/sub-seasonal/seasonal) (Ruane 2010; Radhakrishna et al. 2015). Studies have shown that IPW over the desert Southwest during the summer on a diurnal scale is not reliable (e.g., Radhakrishna et al. 2015), but there is nothing to suggest that from a sub-weekly scale and beyond the NARR provides significant inconsistencies in IPW. Therefore, I believe that I am able to mitigate some of these difficulties for my study by employing a well-extended spatial extent and multiple temporal analyses of surface dew point and IPW.

3.2 Geographic and Temporal Domains

3.2.1 Geographical Domain

For my analysis from the North American Regional Reanalysis (NARR) dataset, I have selected ten specific locations, explicitly Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix AZ, San Diego, CA, Tucson AZ, and Yuma AZ that I will use to compare surface dew point in these locations to IPW. Fig. 1.1 illustrates the location of these cities while further geographic data of the ten

locations are listed in Table 3.1. These data include latitude/longitude information, population, and elevation above mean seal-level (MSL).

Table 3.1. Geographic coordinates, elevation (meters), and population information for the ten study locations. Population values are from the US Census Bureau (2017) for the US cities and from the Instituto Nacional de Estadística y Geografía (INEGI) (2017) for the Mexican city of Guaymas.

City Name	Latitude	Longitude	Elevation (Mean Seal-Level) (meters)	Population (July 2015 estimates)
Albuquerque, NM	35.09° N	106.6° W	1632	559,121
El Paso, TX	31.76° N	106.49° W	1206	681,124
Flagstaff, AZ	35.19° N	111.65° W	2138	70,320
Guaymas, MX	27.92° N	110.91° W	18	149,299 (2010)
Las Vegas, NV	36.17° N	115.14° W	665	623,747
Midland, TX	32.00° N	102.07° W	848	132,950
Phoenix, AZ	33.4° N	112.1° W	346	1,563,025
San Diego, CA	32.7° N	117.16° W	5	1,394,928
Tucson, AZ	32.22° N	110.92° W	806	531,641
Yuma, AZ	32.64° N	114.64° W	65	94,139

These locations were selected in order to encapsulate the majority of the NAM's domain for the purposes of having a quality study. In choosing the locations for this study, I also considered the following factors: 1) having an extensive spatial scope, 2) selecting sites that appeared in the past literature associated with monsoon research, 3) having urban areas that experience the greatest impacts from monsoon rainfall (i.e., personal injury, financial loss or hydrologic management), and 4) selecting locations with variable climate and topography.

These cities, with the exception of Las Vegas, NV, Midland, TX and San Diego, CA are the major cities in the NAM's domain that receive at least half of their yearly rainfall from the NAM (Douglas et al. 1993). My primary NAM cities are A) Guaymas, Mexico, a city that provides adequate southerly extent to this study. It is routinely used

as a location for identifying the beginnings of a Gulf Surge (Brenner 1974; Hales 1974; Adams and Comrie 1997), and it is close to the epicenter of the NAM's influence receiving ~65% of its annual rainfall from June-September. B) Yuma, AZ is the main location used in identifying Gulf Surges in the United States (Brenner 1974; Adams and Comrie 1997; Higgins et al. 2004; Wu et al. 2009). (C and D) Phoenix, AZ and Tucson, AZ, aside from their large populations, are critical points for this study because they are the only two locations to have employed a surface dew point threshold to determine monsoonal onset. From a theoretical standpoint, these locations, both situated in valley topography, are susceptible to capping inversions that could deteriorate the surface dew point and IPW relationship (as discussed in the previous chapter). E) Flagstaff, AZ, a city at an elevation of over 2000 meters, is unique because it should not have the same tendency for capping inversions that Phoenix, AZ and Tucson, AZ have. F) Albuquerque, NM is another high altitude city (1632m MSL), which is also not likely to experience capping inversions. Finally, G) El Paso, TX is a medium sized metropolitan on the far eastern extent of the NAM region that receives just under 100mm of rainfall from July through August (Hales 1974), and the city receives over half of its yearly rainfall June – September (Douglas et al. 1993).

I have also added the cities of Las Vegas, NV, Midland, TX, and San Diego CA to my study area for the following reasons. First, Las Vegas, NV is a city that receives approximately 20% of its yearly rainfall from the NAM, which is less than the other cities, but it has a sizeable financial concern (i.e., tourism). In addition, the city provides coverage of the far northwest boundary of appreciable monsoon rainfall. Midland, TX is

the first major city to the east of El Paso, TX and considered outside the NAM region. This is important as it will provide a way of comparing any results between NAM cities and non-NAM cities. Correspondingly, San Diego, CA is also a city beyond the extent of the NAM's expanse but is placed into this research as a means for far western extent of spatial coverage, and is the location used by Reber and Swope (1972) for their study in comparing surface humidity and IPW. Given that the Reber and Swope (1972) study is a major counter to the findings of Reitan (1963), including this city in my analysis will provide a comparison point to the nearby monsoonal cities.

There are, however, a couple of issues with these locations for this study. First, all these cities are located far removed from the central United States (with the exception of Midland, TX), which is problematic because the NARR can have issues with accurately representing data in these locations on very short time scales (Burkovsky and Karlovy 2007). Second, the measurements of these ten locations are inherently discrete; in other words, fine details could be missed because of the lack of continuous coverage. However, the positive points raised beyond, in my opinion, negate these possible limitations.

3.2.2 Temporal Domain

For the temporal coverage, I needed to select both the seasonal timeframe and the long-term length-of-record. Using the information presented in the last chapter from previous research, I am defining the seasonal period of study to be slightly longer than the accepted period for the NAM since my study is attempting to ascertain the best "start" time for the NAM. Consequently, I have selected the seasonal time scale of 0000 UTC 2 June to 2100 UTC 1 October. By using this seasonal timeframe, I should have ample

coverage of the entire NAM season as well as displaying the evolution of the changing dry air mass to the moist air mass that perpetrates the start of the monsoon season.

Reber and Swope's (1972) major criticism is that Retain (1963) values were "artificial" as a result of using mean monthly data. For my study, while a comparison on a diurnal timescale may not produce useful results because of the shortcomings of the NARR representing moisture convergence on daily timescales, I have selected a variety of measures, specifically, 1) non-averaged three-hour measurements for length-of-record, 2) a three-day average, 3) a weekly average, 4) a mean-monthly average, and 5) a seasonal average. The NARR includes a data value for every three hours (0000 UTC, 0300 UTC, 0600 UTC, 0900 UTC, 1200 UTC, 1500 UTC, 1800 UTC, and 2100 UTC). By using the five different timescales, I will better represent the evolution of the surface dew point and the IPW relationship. Moreover, as I state in my problem statement, if Reber and Swope (1972) were correct in their assessment of Reitan (1963), then the correlation between surface dew point and IPW should improve as the averaged time increases.

For the long-term length-of-record, I am limited to the time frame used with the NARR dataset. Consequently, I have selected from the period from 1 January 1979 – 31 December 2015. Of course, this limits the temporal coverage of my study. By using a period of record that only goes back to 1979, I am not able to include any data from the time period for which Reitan (1963) conducted his study. Additionally, any variation in the NAM prior 1979 cannot be accessed from my study. However, because the period-of-record is 37 years with eight data points produced by the NARR every 24 hours that

provides a large enough sample size to conduct statistical significance testing. Specifically, the N-size for the three-hour, three-day, weekly average, monthly average, and seasonal period for each location is 36,112, 1,505, 645, 174, and 37, respectively.

3.3 Meteorological Variables

Two specific meteorological variables are needed to examine my fundamental research question involving the relationship between surface dew point temperature and the depth of moisture throughout the atmosphere in United States' desert Southwest and northwest Mexico during the North American Monsoon (NAM). These are 1) surface dew point (measured in °C) and 2) integrated precipitable water (IPW) (measured in mm). The NARR dataset produces these values in format of time in UTC, on a given day in a day-month-year format. The three variables of interests, surface dew point (Td), accumulated precipitation (ACPC), IPW, and are computed in the NARR in units of Kelvin (K) and kgm^{-2}, respectively. Conveniently, kgm^{-2} is equivalent to mm. My analysis matrix consists of ten files (one for each location) of 36,112 observations by 5 values (time, day, Td, ACPC, IPW). That said, I need to create at least one calculated columns in the statistical software that will convert from K to °C to conduct the assessment. Table 3.2 below is the descriptive analysis of the ten study locations and their central tendency, measures of dispersion, and symmetrical distribution values.

Table 3.2. Basic statistical information of the meteorological variables (Dew Point, ºC; IPW mm), size, maximum value, minimum value, range, mean, median, standard deviation, skewness, and kurtosis for, Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix AZ, San Diego, CA, Tucson, AZ, Yuma, AZ from 1979-2015 for the season period 1 June to 1 October.

Variable	Sample Size	Max.	Min.	Range	Mean	Med.	Standard Deviation	Skewness	Kurtosis
ALBUQUERQUE, NM									
Dew Point (°C)	36112	17.30	-26.26	43.56	4.74	5.86	5.36	-0.78	0.25
IPW (mm)	36112	37.90	1.33	36.57	17.83	18.47	5.90	-0.22	-0.70
EL PASO, TX									
Dew Point (°C)	36112	21.41	-17.37	38.78	8.59	9.76	5.22	-1.04	0.97
IPW (mm)	36112	48.29	2.47	45.82	24.39	25.48	7.38	-0.04	-0.51
FLAGSTAFF, AZ									
Dew Point (°C)	36112	16.93	-20.04	36.97	4.59	5.45	5.62	-0.53	-0.28
IPW (mm)	36112	34.25	1.27	32.98	15.00	15.13	6.10	0.03	-0.99
GUAYMAS, MX									
Dew Point (°C)	36112	28.54	-7.38	25.92	21.16	22.20	4.18	-1.66	3.62
IPW (mm)	36112	70.23	8.63	61.60	42.19	45.20	11.10	-0.77	-0.30
LAS VEGAS, NV									
Dew Point (°C)	36112	16.05	-29.44	45.48	-.21	-.24	6.06	-0.20	-0.17
IPW (mm)	36112	38.80	1.15	37.66	14.35	13.04	6.54	0.59	-0.43
MIDLAND, TX									
Dew Point (°C)	36112	24.19	-9.87	34.06	14.16	14.76	3.61	-1.26	2.74
IPW (mm)	36112	54.82	5.39	49.43	29.60	30.11	7.09	-0.35	~0.00
PHOENIX, AZ									
Dew Point (°C)	36112	22.84	-24.16	47.00	9.78	10.86	6.15	-0.56	-0.37
IPW (mm)	36112	58.29	2.84	55.45	29.32	29.99	11.25	-0.08	-1.10

SAN DIEGO, CA									
Dew Point (°C)	36112	21.61	-9.53	31.15	15.50	15.56	1.73	-0.61	5.10
IPW (mm)	36112	56.80	3.81	53.00	24.85	22.91	8.60	0.77	-0.01
TUCSON, AZ									
Dew Point (°C)	36112	21.37	-16.19	37.56	9.30	10.67	6.12	-0.68	-0.33
IPW (mm)	36112	47.67	3.22	44.45	24.42	23.54	8.95	-0.22	-1.03
YUMA, AZ									
Dew Point (°C)	36112	25.18	-24.24	49.24	11.92	13.22	6.56	-0.81	0.38
IPW (mm)	36112	65.15	1.95	62.21	27.26	26.45	11.32	0.22	-0.90

Guaymas, MX has the highest recorded Td and IPW (28.54°C, 70.23mm) maximum throughout the study period. Likewise, it has the highest minimum Td and IPW recorded (-7.38°C, 8.63mm). This is in contrast to Las Vegas, NV, which recorded the lowest maximum Td of all the data locations and the 3rd lowest IPW (16.05°C, 38.80mm) while recording the greatest minimum values of (-29.44°C, 1.15mm). Looking at the limits of the measured maximum/minimum values, Guaymas, MX and Las Vegas, NV represent the locations with the greatest Td and IPW extremes. These extremes are apparent when looking at the mean and median of Td and IPW measurements for Guaymas, MX (mean = 21.16°C and 41.19mm, median = 22.20°C and 45.20mm) and Las Vegas, NV (mean = -0.21°C and 14.35mm, median = -0.24°C and 13.04mm) as well. Guaymas, MX, Yuma, AZ, Midland, TX, Phoenix, AZ, San Diego, CA, El Paso, TX, and Tucson, AZ are, in order, the ranking of the top 7 location for maximum Td measurements. However, this stratification does not hold up for the top 7 IPW measurements as Midland, TX, Phoenix, AZ, and San Diego, CA shift positioning. After Tucson, AZ, the three cities with the

lowest maximum Td and IPW are Albuquerque, NM, Flagstaff, AZ, and Las Vegas, NV. There is some incidental ordinal variability in minimum Td and IPW measurements when comparing these locations. This variability is carried over when looking at dispersion.

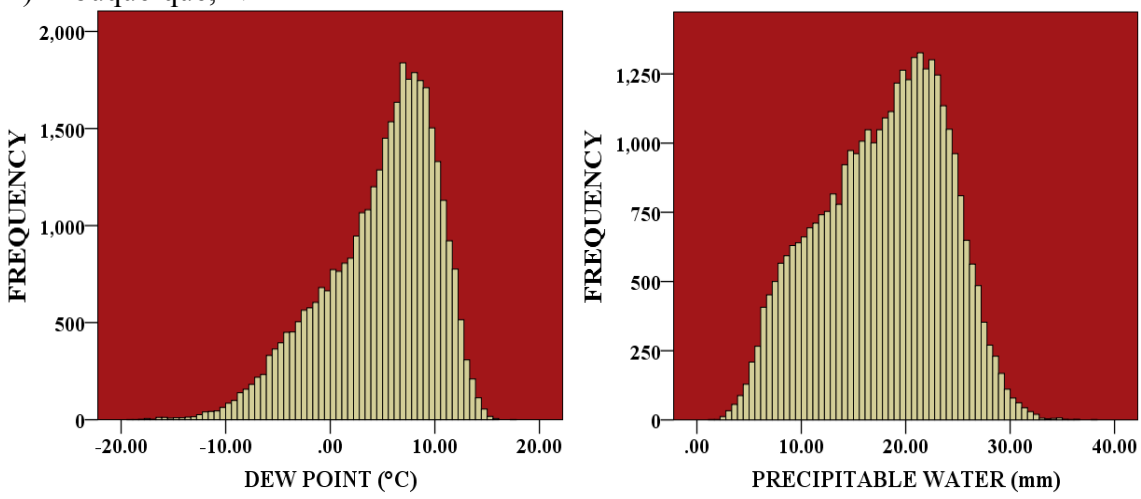
The dispersion measurements for this study are range and standard deviation. The city with the greatest range in Td and IPW is Yuma, AZ (49.24°C, 62.21mm). Guaymas, MX has the smallest Td range (25.92°C), but has the 2nd greatest IPW range (61.60mm). Yuma, AZ has the greatest standard deviation for both Td (6.50°C) and IPW (11.32mm). This suggests that for both Td and IPW, Yuma, AZ, when compared to the other cities, is the most prone to extreme values. Guaymas, MX has a relatively low standard deviation for dew point (4.18°C), but has the 3rd highest standard deviation for IPW (11.10mm), which could mean that Guaymas, MX is subject to extreme events in the case of IPW but not with respect to Td. Additionally, it is worth noting that Phoenix, AZ ranked 2nd amongst the cities for Td range (47.00°C), Td standard deviation (6.15°C), IPW standard deviation (11.25mm), and 3rd for range Td (55.45mm). This puts forth the idea that Phoenix, AZ is relatively susceptible to large variations in both Td and IPW during the NAM. Altogether, all the cities in the study have a range of Td at least of 31.15°C and an IPW range 32.98mm, which indicates large variability in surface moisture and moisture aloft during the NAM.

The Td skewness histograms for all locations are negatively skewed with Las Vegas, NV having the least skew (-0.20) and Guaymas, MX have the largest skew (-1.66). Three cities, San Diego, CA (0.77), Las Vegas, NV (0.59), and Yuma, AZ (0.22) exhibit positive skewness for IPW. Two locations are very close to zero skew for IPW,

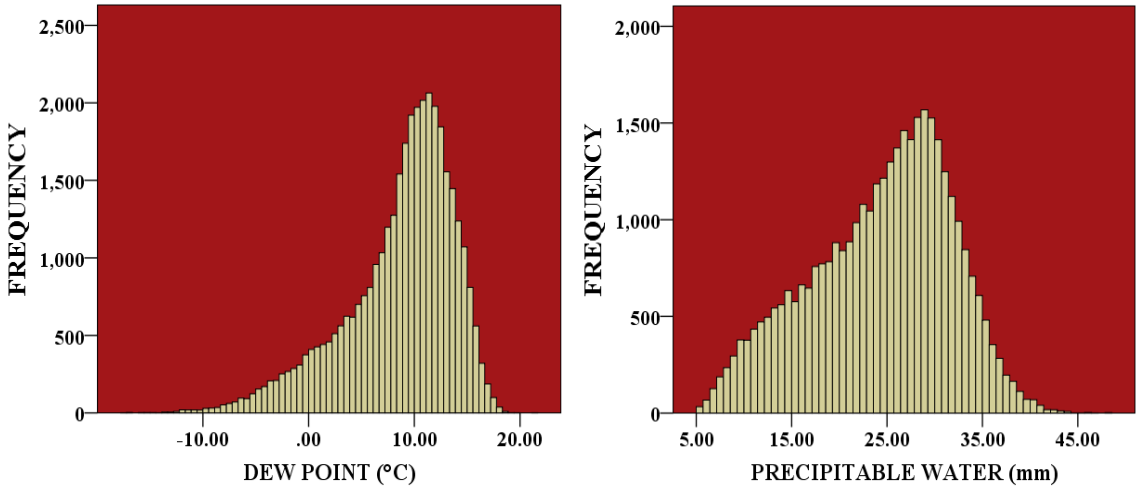
specifically Flagstaff, AZ (0.04) and Phoenix, AZ (-0.08). Meanwhile, the remaining cities of Albuquerque, NM, Tucson, AZ, El Paso, TX, Midland, TX, and Guaymas, MX all show negative skewness at -0.22, -0.22, -0.35, -0.37, -0.77, respectively.

These cities show both positive and negative kurtosis with Td values. San Diego, CA is the most leptokurtic (5.10). Las Vegas, NV is closest to being mesokurtic (-0.17) while Phoenix, AZ is the most platykurtic (-0.37). For IPW measurements, Midland, TX and San Diego, CA are very close to mesokurtosis (-0.003 and -0.01) with all the cities becoming increasing platykurtic. Phoenix, AZ is the most platykurtic (-1.10). Below is Fig. 3.1 which provides the graphical representation of skewness and kurtosis for both Td and IPW.

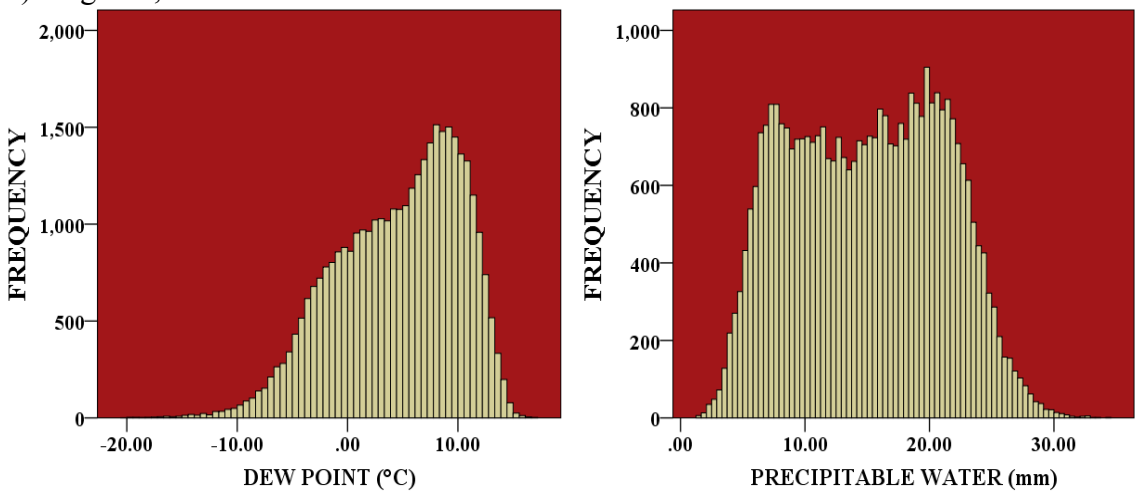
(A) Albuquerque, NM



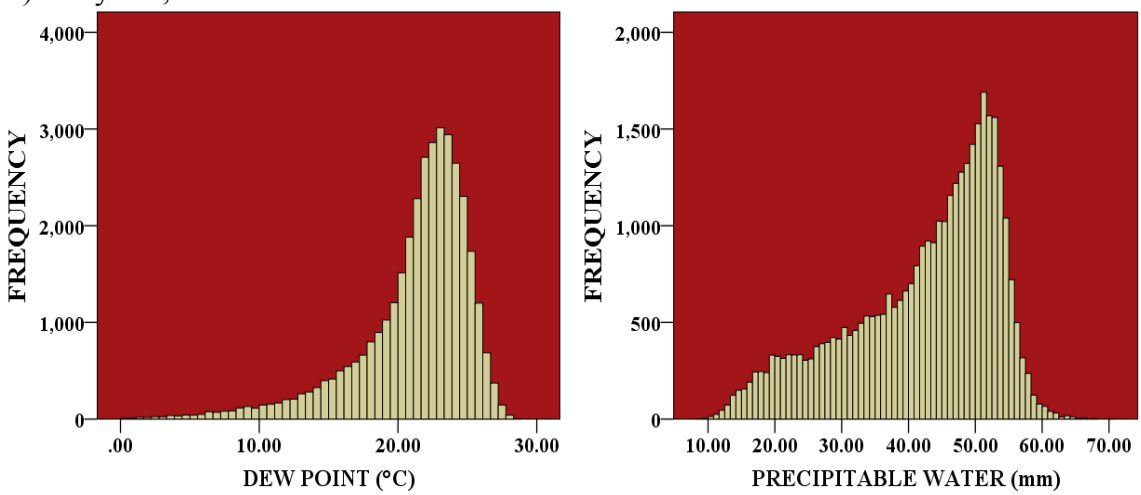
(B) El Paso, TX



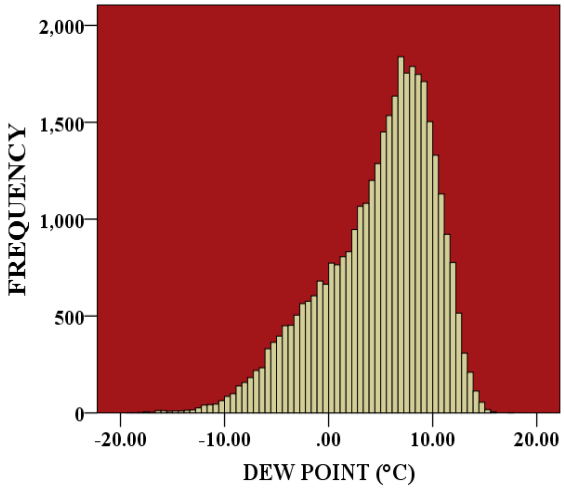
(C) Flagstaff, AZ

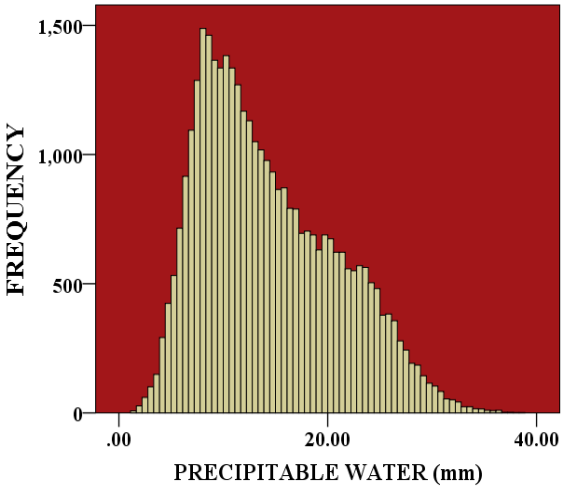


(D) Guaymas, MX

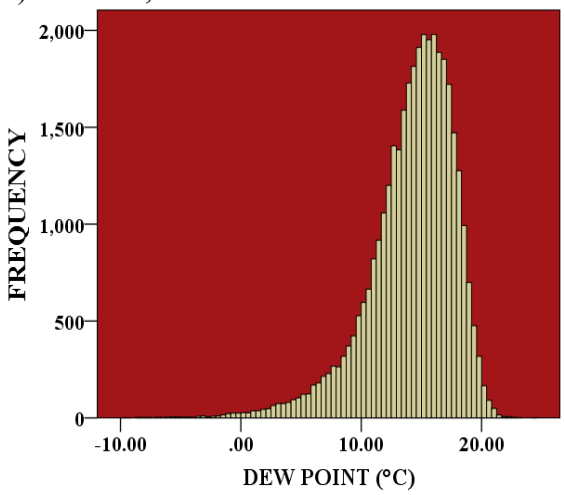


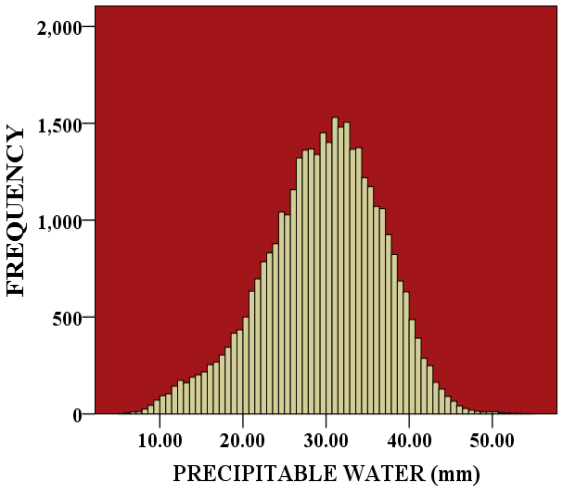
(E) Las Vegas, NV



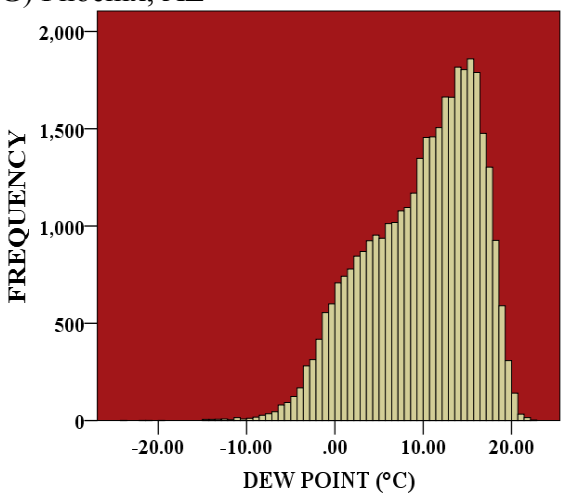


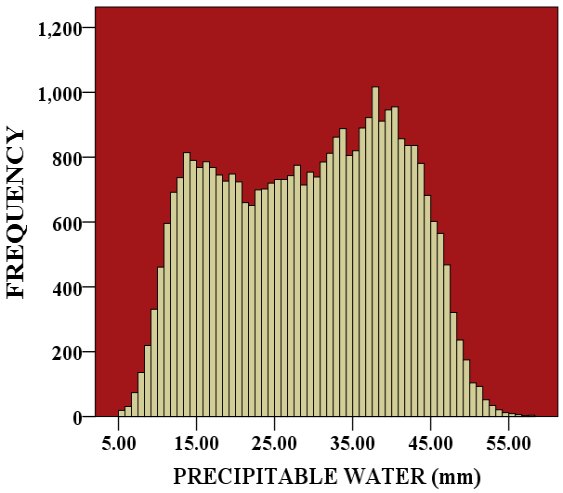
(F) Midland, TX



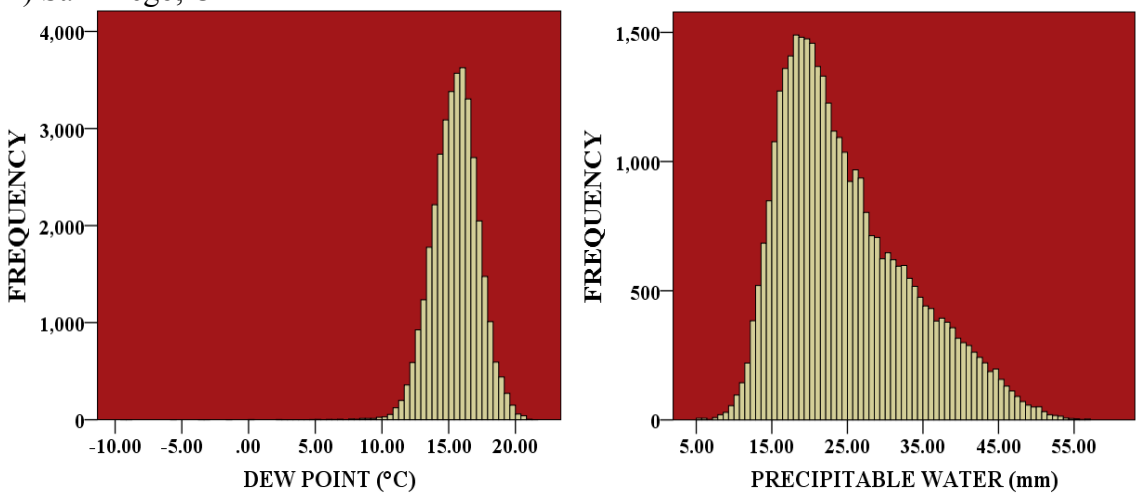


(G) Phoenix, AZ

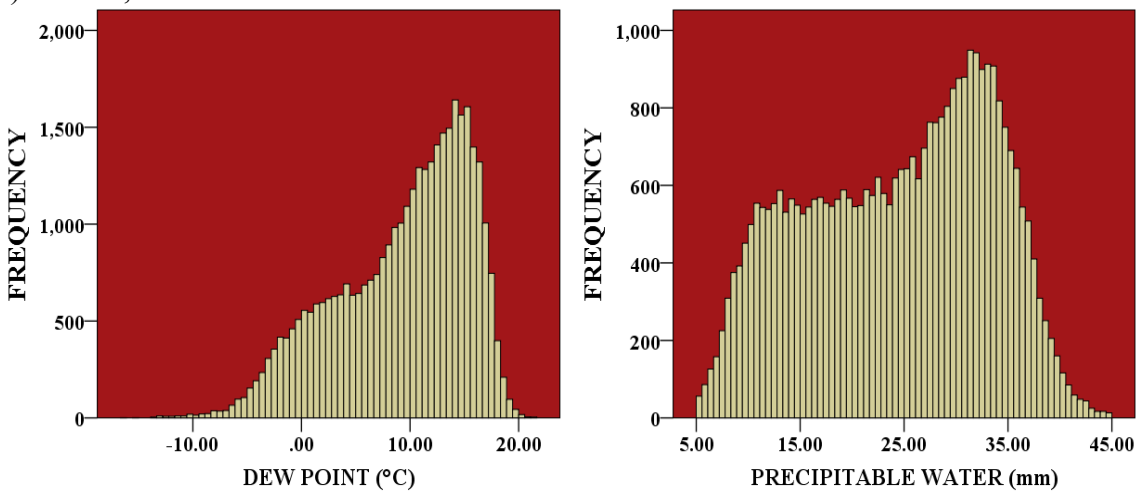




(H) San Diego, CA



(I) Tucson, AZ



(J) Yuma, AZ

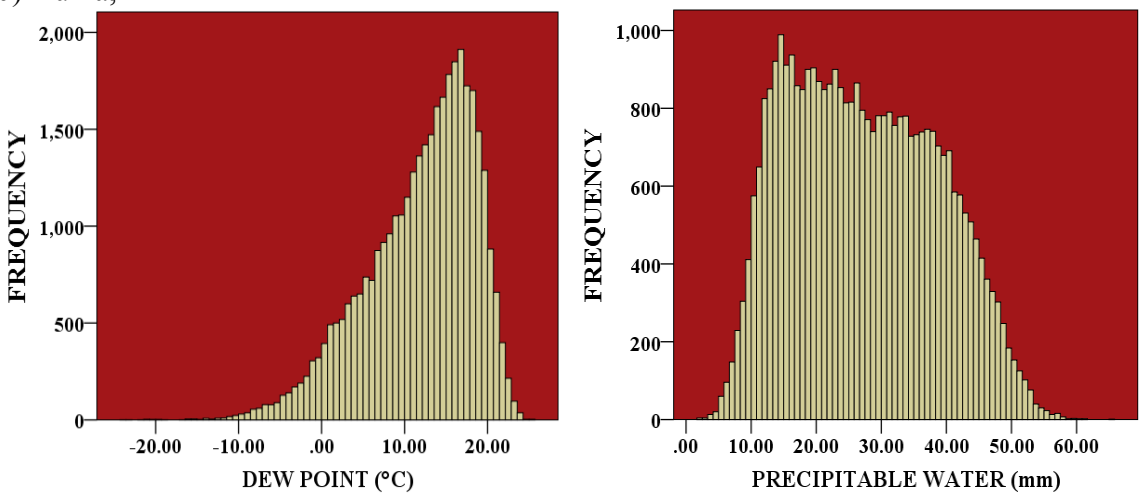


Fig. 3.1. Surface Td (°C) and IPW (mm) histograms for each of the ten study locations for the monsoonal season (1979-2015). A) Albuquerque, NM, B) El Paso, TX, C) Flagstaff, AZ, D) Guaymas, MX, E) Las Vegas, NV, F) Midland, TX, G) Phoenix, AZ, H) San Diego, CA, I) Tucson, AZ, J) Yuma, AZ.

3.4 Summary

The purpose of my thesis is to examine the relationship between surface dew point to IPW and relate that to a possible way to determine NAM onset. To conduct this study, I am using data from the North American Regional Reanalysis (NARR) data, For my analysis from the NARR dataset, I have selected ten specific locations, explicitly including Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix AZ, San Diego, CA, Tucson AZ, and Yuma AZ. I have selected the seasonal time scale of 0000 UTC 2 June to 2100 UTC 1 October. I have selected a variety of short-term measures, specifically, 1) a non-averaged three-hour measurement for the length-or-record, 2) a three-day average, 3) a weekly average, 4) a mean-monthly average, and 5) a seasonal average. The NARR produces includes a data value for every three hours (0000 UTC, 0300 UTC, 0600 UTC, 0900 UTC, 1200 UTC, 1500 UTC, 1800 UTC, and 2100 UTC).

The variables that I have selected for study are 1) surface dew point (measured in °C) and 2) integrated precipitable water (IPW) (measured in mm). When I applied basic statistical analysis of these variables for the ten locations, the important points that I observed are: 1) the regions with the greatest moisture (Guaymas, MX) and least moisture (Las Vegas, NV) during the NAM, 2) the range and standard deviations, which illustrate the large temporal variability of moisture over certain locations during monsoon season, and 3) the skewness and kurtosis, which are vital in normality testing, determining which inferential statistical tests are appropriate, and significance testing. Given this dataset, I am now able to conduct a statistical analysis of the relationship between surface dew point and integrated precipitable water. The details of that analysis are given in the next chapter.

Chapter 4: Methods and Results

4.1 Introduction

In order to continue my research of the relationship of integrated precipitable water (IPW) and surface dew point (Td), I must analyze the data from the North American Regional Reanalysis (NARR), discussed in Chapter 3, and conduct extensive statistical testing. All NARR data were made available online through esrl.noaa.gov. In order to extract the data from this site, a python code was written(Appendix 1). All computations were conducted using the Statistical Package for the Social Sciences (SPSS) software. The statistical tests for this investigation are as follows: 1) comparing the validity of NARR data to direct measurement via radiosondes from the University of Wyoming's archived site, using a combination of the Kolmogorov-Smirnov test (KS) and the Shapiro-Wilk test (SW) for normality testing of all datasets, 2) employing the proper correlation testing, based on normality, between IPW and Td, and finally 3) examining the evolution of those correlations over a variety of timescales. Results from all of these tests provides an understanding of the IPW and Td relationship for the southwest United States and northwest Mexico during the North American Monsoon (NAM) from 1979 to 2015.

Once the relevant data were extracted, they were converted from a .txt format and uploaded into matrices on SPSS. A separate matrix was developed for each of the ten study locations. The dimensions of the matrices were 36,112 x 6. Four of the six columns were the date (yyyy-mm-dd hh:mm:ss), dew point in kelvin, accumulated precipitation (mm), and precipitable water (mm). Two additional calculated columns were created to

convert dew point into °C and precipitable water into the natural logarithm (Ln) of precipitable water (mm). Each row represented the associated three-hour measurement of each variable for the entire length-of- record. Additional calculated columns, and modifications to the matrices occurred as needed to conduct the necessary tests of the variable at different averaged timescales.

4.2 Validity of data source

Prior to correlation analysis, it is necessary to briefly review the NARR's performance compared to direct radiosonde measurement as initially discussed in Chapter 2. Past research has shown that the NARR has shortcomings when representing low-level moisture during the monsoon. It has a tendency to over represent low-level moisture during surge events, areas in complex terrain, and regions that boarder the ocean or outside the United States (Mo et al. 2005a; Mesinger et al. 2006). Since a major moisture source of the NAM comes from Gulf Surges, I sought to compare the NARR's Td and IPW values to direct radiosonde measurements. The purpose of this was to verify the adequacy of using the NARR data with respect to this study.

In order to conduct this testing, I selected five out the ten cities. These cities were selected because they were on the fringes of my study area or located in a region of complex terrain. Guaymas, MX and Yuma, AZ would have also been good choices for this data efficacy analysis, but reliable radiosonde data for these stations were not obtainable. Consequently, the five cities I selected were: El Paso, TX, Flagstaff, AZ, Midland, TX, San Diego, CA, and Tucson, AZ. The specifics of these cities' analyses are outlined in the Table 4.1.

Table 4.1. The city, month, year, and N-size used for the five locations (El Paso, TX, Flagstaff, AZ, Midland, TX, San Diego, CA, and Tucson, AZ) used in verification testing. N-Size is variable across cities as a limited number of direct measurements were missing from the record.

City	Month	Year	N-Size
El Paso, TX	August	1996	56
Flagstaff, AZ	June	2000	56
Midland, TX	August	1997	56
San Diego, CA	September	2004	60
Tucson, AZ	July	2009	62

The month (either June, July, August, September) and year (1979-2015) for this data testing were selected randomly and compared to direct 0000 UTC and 1200 UTC measurements archived by the University of Wyoming (weather.uwyo.edu). To conduct the test, specific 0000 UTC and 1200 UTC measurements were imputed, by hand, into SPSS and correlated with their respective NARR counterpart in two different ways: 1) A correlation between direct measurement (sonde) Td and IPW, NARR Td and IPW, and, 2) a correlation between Td (sonde)/Td(NARR) and IPW(sonde)/IPW(NARR). Either Spearman's rank-order correlation (ρ) or Pearson's correlation (r) were used based on the normality of the distribution. The results of this testing are presented in Table 4.2

Table 4.2. Mean monthly value comparisons of Td (°C) and IPW (mm) for NARR and sonde for the five locations (El Paso, TX, Flagstaff, AZ, Midland, TX, San Diego, CA, and Tucson, AZ) used in verification tests. All data were non-normally distributed with the exception of Midland, TX.

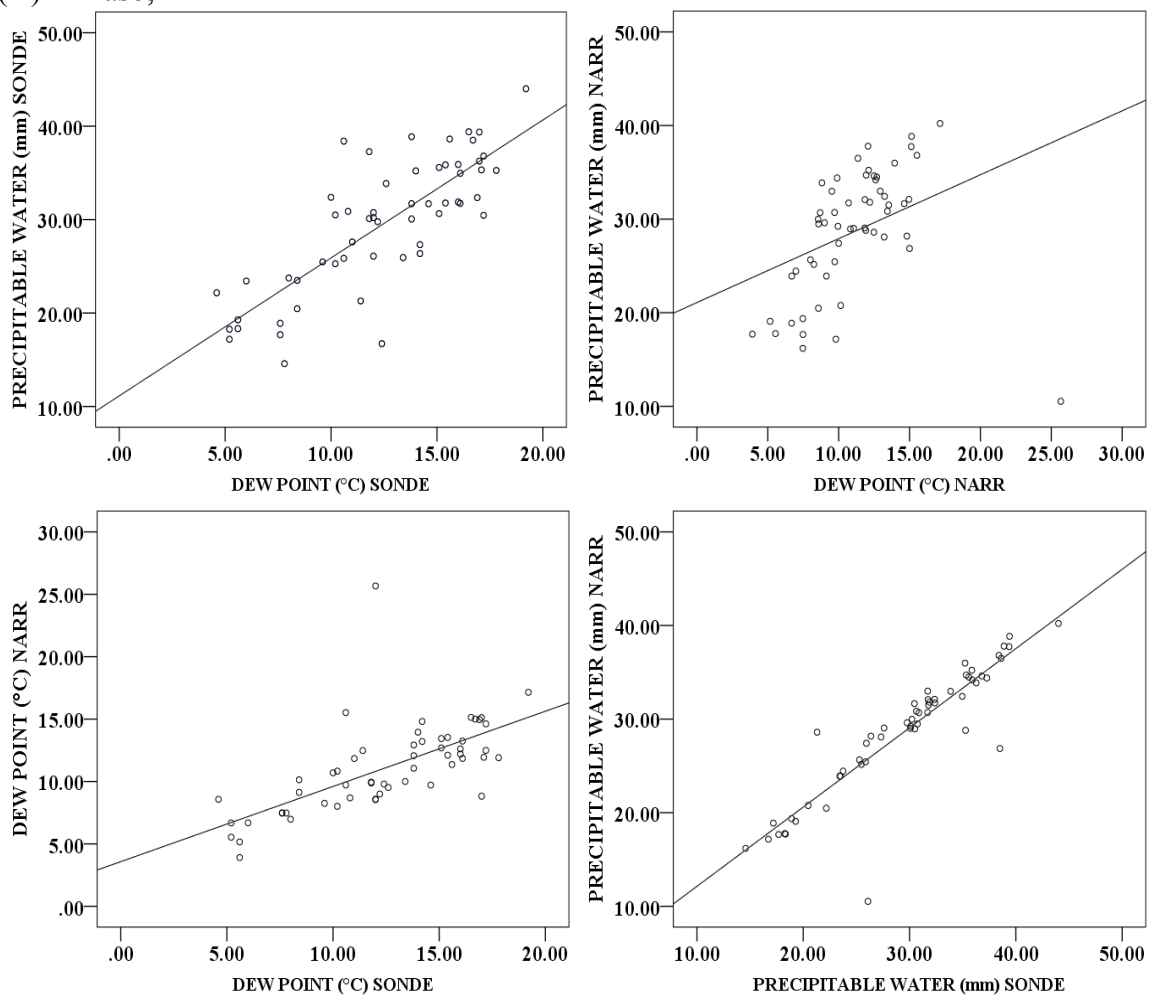
City	Mean Td Sonde (°C)	Mean IPW Sonde (mm)	Mean Td NARR (°C)	Mean IPW NARR (mm)	(ρ) Td Sonde & IPW Sonde	(ρ) Td NARR & IPW NARR	(ρ) Td Sonde & Td NARR	(ρ) IPW Sonde & IPW NARR
El Paso, TX	12.44	29.49	11.07	28.64	0.768	0.603	0.738	0.918
Flagstaff, AZ	-2.08	10.94	1.39	11.74	0.843	0.817	0.830	0.987

Midland, TX	15.92	32.05	15.72	32.83	(ρ) = 0.635 (r) = 0.633	(ρ) = 0.268 (r) = 0.315	(ρ) = 0.606 (r) = 0.625	(ρ) = 0.973 (r) = 0.980
San Diego, CA	12.52	18.67	14.99	19.95	0.595	0.785	0.607	0.956
Tucson, AZ	13.27	31.19	11.05	28.07	0.719	0.670	0.543	0.978

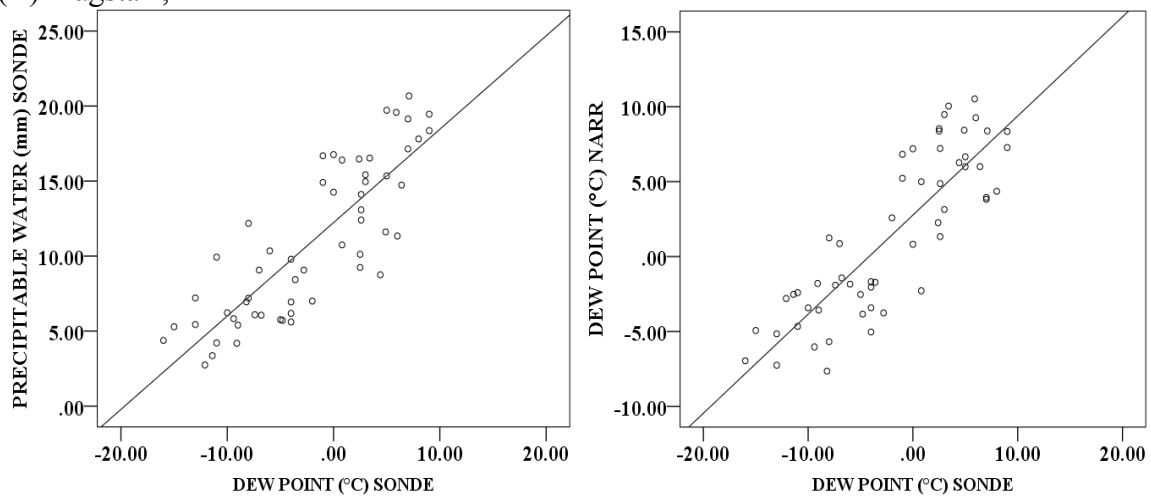
The correlation results show that the NARR and Sonde data for mean monthly values of IPW and Td are very close. IPW is very well represented for all locations with all $\rho \geq 0.918$. However, there is some variablity between Td Sonde and Td NARR correlations. The range of values are ($0.543 \leq \rho \leq 0.830$). There is no real spatial pattern apparent with this variablity as Flagstaff, AZ, the region with the most complex terrain, represented mositure the best of the five. Tucson, AZ, a location that is suceptable to moisture from Gulf Surges, showed the weakest correlation.

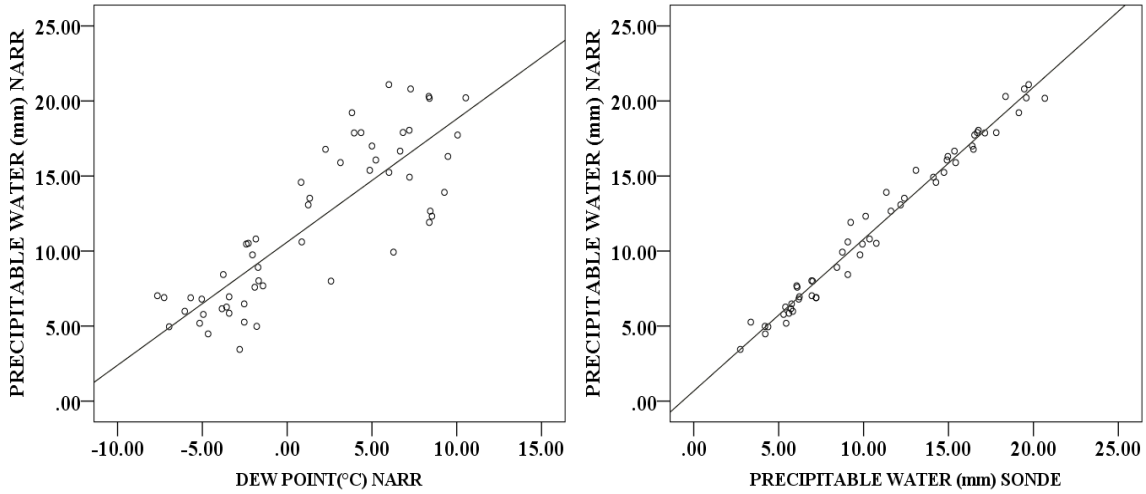
Technically, this could be problematic for the study, except any issues with the representation of low-level moisture likely are smoothed out for longer timescales as outlined in the literature (e.g., Ruane 2010; Radhakrisha et al. 2015). A visual representation of these relationships is presented in Fig. 4.1.

(A) El Paso, TX

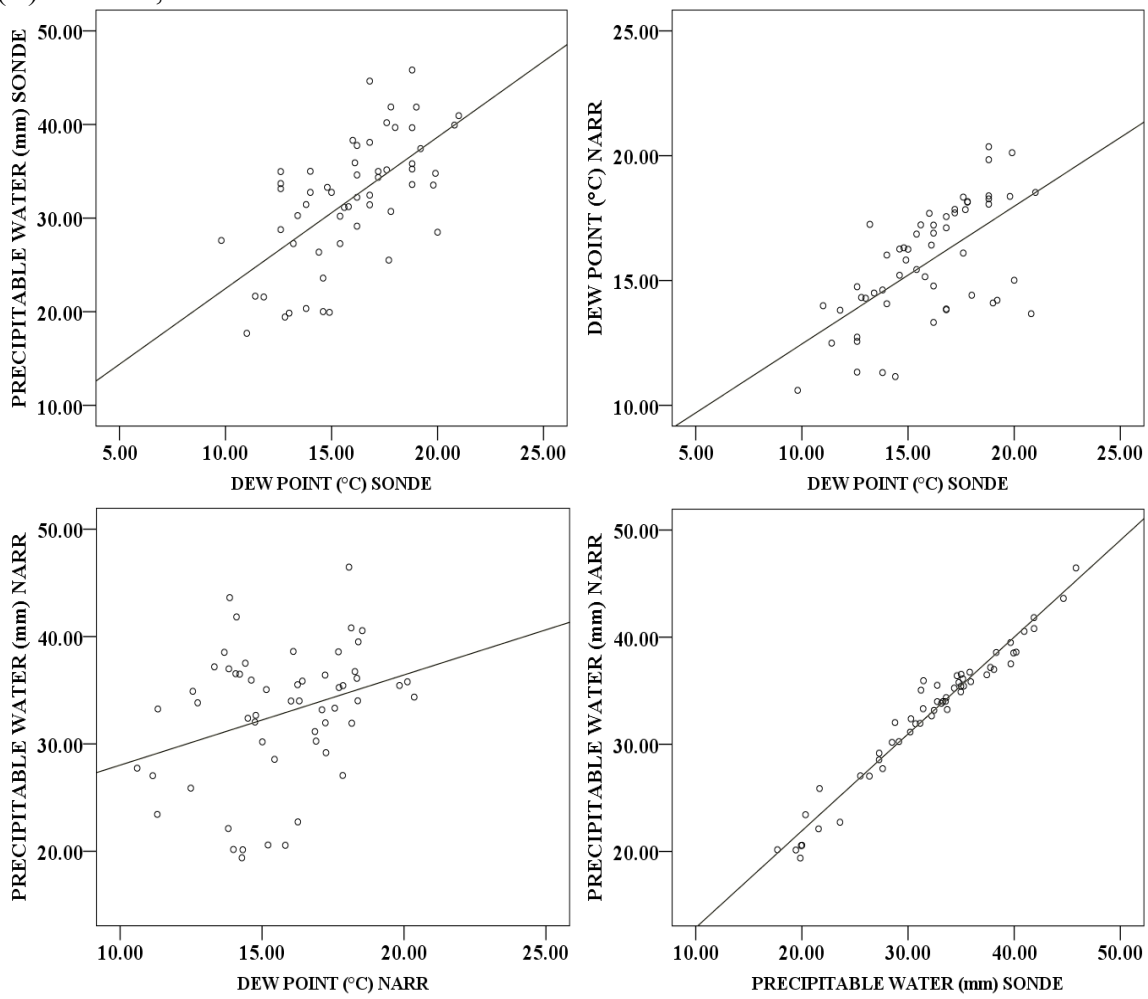


(B) Flagstaff, AZ

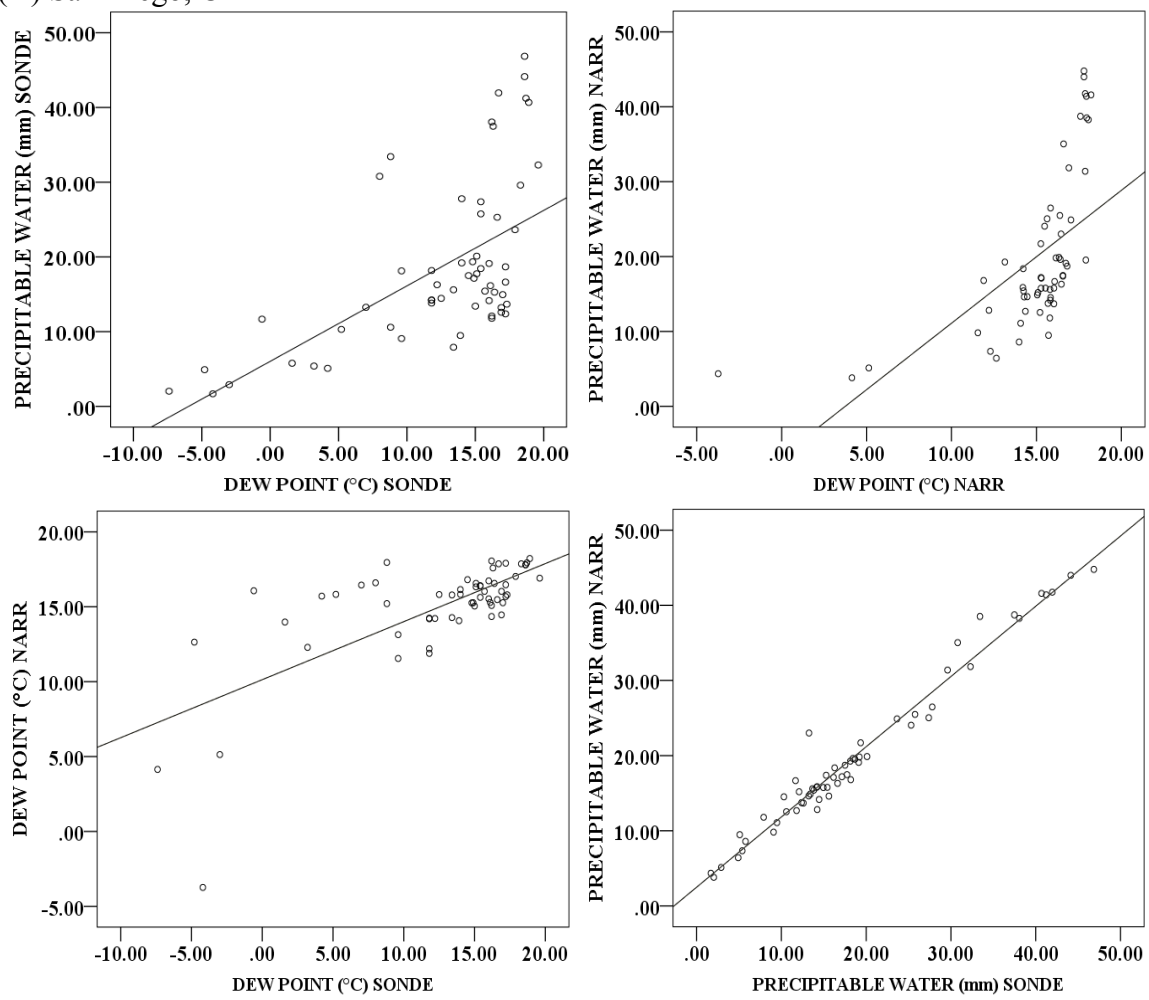



PRECIPITABLE WATER (mm) NARR
25.00
20.00
15.00
10.00
5.00
.00
-10.00
-5.00
.00
5.00
10.00
15.00
DEW POINT(°C) NARR
PRECIPITABLE WATER (mm) NARR
25.00
20.00
15.00
10.00
5.00
.00
.00
5.00
10.00
15.00
20.00
25.00
PRECIPITABLE WATER (mm) SONDE

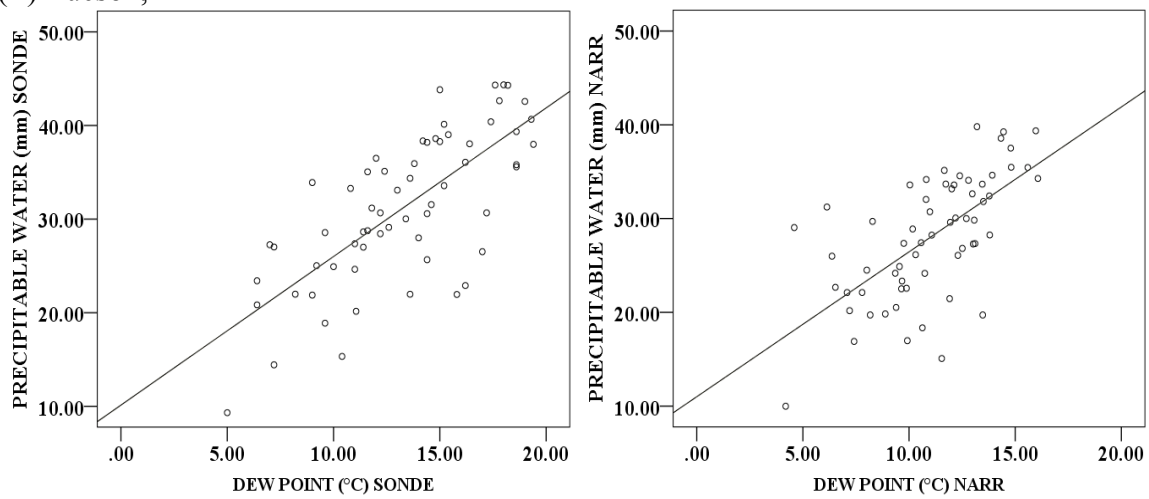
(C) Midland, TX


PRECIPITABLE WATER (mm) SONDE
50.00
40.00
30.00
20.00
10.00
5.00
10.00
15.00
20.00
25.00
DEW POINT (°C) SONDE
DEW POINT (°C) NARR
25.00
20.00
15.00
10.00
5.00
10.00
15.00
20.00
25.00
DEW POINT (°C) SONDE
PRECIPITABLE WATER (mm) NARR
50.00
40.00
30.00
20.00
10.00
15.00
20.00
25.00
DEW POINT (°C) NARR
PRECIPITABLE WATER (mm) NARR
50.00
40.00
30.00
20.00
10.00
20.00
30.00
40.00
50.00
PRECIPITABLE WATER (mm) SONDE

(D) San Diego, CA



(E) Tucson, AZ



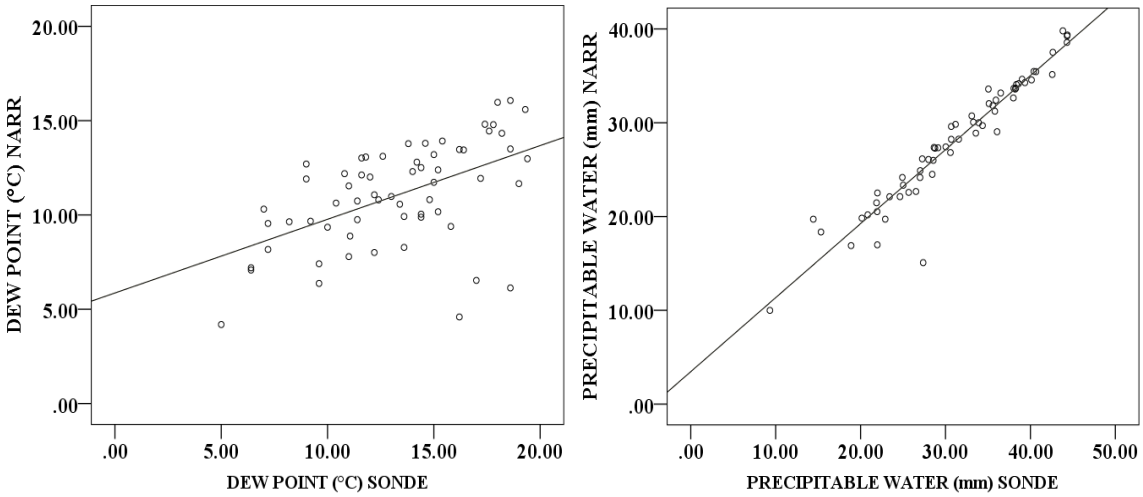


Fig. 4.1. Scattergrams of dew point (Td in °C) and precipitable water (IPW in mm) of the five locations for the NARR results versus observed rawinsonde observations. Four scattergrams IPW(sonde) vs. Td(sonde), IPW(NARR) vs. Td(NARR), Td(NARR) vs. Td(sonde) and IPW(NARR) vs. IPW(sonde), and the associated best-fit trend lines are presented for each of the five sample locations for the NAM (1979-2015). A) El Paso, TX, B) Flagstaff, AZ, C) Midland, TX, D) San Diego, CA, E) Tucson, AZ.

Additionally, because Gulf Surges are short-term and periodic (Hales 1972), theoretically as N-size increases, similar smoothing should take effect. Nevertheless, even with this shortcoming, the results do not eliminate the NARR data from this type of analysis as any limitations can be resolved to some degree with increased timescales and increased N-sizes. Furthermore, direct measurement data includes flagged data, missing data, and changes of measurement protocols (Bosart 1990) that could make any attempt at comparing long-term direct measurement data incompatible. Moreover, three additional considerations favor the use of NARR data. First, Guaymas, MX does not have readily available radiosonde data; second, most cities have missing radiosonde data while the NARR has no missing values, and, third, IPW is well represented by the NARR dataset. Consequently, I selected the NARR reanalysis as most appropriate for this study. The

next step was to determine the normality of IPW and Td for all of the cities at all the different time scales as this determines the appropriate correlation test.

4.3 Normality testing

Normality testing is necessary in order to determine the proper correlation testing to be used. Commonly, Pearson's r is the standard correlation test, however, it requires an assumption of normality (Wilks 2011). In contrast, Spearman's rank-order correlation (ρ) is a derivative of Pearson's r but does not require a normal distribution (Wilks 2011). Thus, it is vital to determine the distribution of each of the variables to be examined as this will determine, which statistical test can best be applied. Two different tests, the Kolmogorov–Smirnov (KS) and Shapiro–Wilk (SW) tests, were employed for normality determination while using the same null hypothesis (H_0) and alternative hypothesis (H_a). H_0 was that the population was normally distributed while H_a was that the population was not normally distributed. Both normality tests were automatically calculated through SPSS, however, further discussion on their usage, assumptions, and interpretation of output is needed.

The KS test is one of the ways to determine normality for a given sampling of data so long as the sample size is greater than 25. Theoretically, the KS test examines the cumulative distribution function (CDF) of a sampled set of data and compares it to the expected CDF of a normally distributed data set (Wilks 2011). The equation for KS is as follows:

$$D_n = \max|F_n(X) - F(X)| \quad (4.1)$$

Where,
D_n = test statistic, max = is greatest difference between the empirical and theoretical CDF,
$F_n(X)$ = is the empirical cumulative probability,
$F(X)$ = is the theoretical cumulative distribution function evaluated at X.

The test statistic must be compared to a critical value (C_α).

(C_α), at the 5% confidence level, used for each N-size was figured by the equation from (Wilks 2011):

$$C\alpha = \frac{0.886}{\sqrt{N}} \quad (4.2)$$

Where,
N= the number of samples in the dataset

If the test statistic (Dn), is greater than or equal to the critical value, then the null hypothesis is rejected, and the data are assumed to be non-normally distributed. Although useful in determining normality in larger samples, KS is not considered as robust as SW (Wilks 2011). Thus, it is important to also apply the SW test.

The Shapiro-Wilk test is a more robust means for determining normality of a data set. It works by comparing the correlation between the empirical data and the ranks of the normally distributed quantiles functions (i.e., it runs a correlation test for a Q-Q plot) (Wilks 2011). This test is considered to be very good at determining normality, however, it is limited to N-sizes lower than 1000 (Wilks 2011). This test is also be run through SPSS. Analysis of the SW test output works in the *opposite* fashion of the KS test. There is an established critical value for the 5% confidence interval level based on a given N-size that is provided in a table by Wilks (2011). However, if the test statistic is *less than or equal to* the critical value, then the null hypothesis is rejected and the data is

considered to be non-normal (Wilks 2011). The formula for this test is the identical formula of Pearson's (r), which is discussed in detail below. In summary, KS is considered to be an improvement over other methods for determining normality (e.g., chi-squared), but SW is thought to be an improvement over KS, despite its limitations with sample size (Wilks 2011). Two datasets of this study exceed this N-size for SW, so the need to implement two different tests for normality is justified.

In order to test normality for each location at various timescales in SPSS, the aggregate data operation was used. This function allows a SPSS user to separate data according to any parameter; in this case the data were aggregated according to every 24 consecutive measurements, which gives a three-day average (eight measurements a day for three days), a weekly average defined as 56 consecutive measurements (eight measurements a day for seven days), by each month, and by each year. Once separated, basic descriptive statistics were applied to these data and associated normality testing computed. These descriptive statistics were presented in Chapter 3. Table 4.3 gives the results of only the normality testing for the ten cities, including the associated test statistic and critical values. If any datasets contain one normally distributed variable while another is non-normal, the assumption of normality is still not applicable and non-parametric inferential statistics must be applied (e.g., yearly mean measurements for Albuquerque, NM).

Table 4.3. Normality test results (N-size, type of normality test (KS or SW), test statistic, critical value and evaluation of normality) for the ten study locations (Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ) as classified by timescale (Three-hour, three-day, weekly, monthly, yearly).

Albuquerque, NM					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurements					
Td	36112	KS	0.049	0.005	N
IPW	36112	KS	0.084	0.005	N
Ln(IPW)	36112	KS	0.105	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.110	0.023	N
IPW	1505	KS	0.067	0.023	N
Ln(IPW)	1505	KS	0.109	0.023	N
Weekly Mean Measurements					
Td	645	SW	0.972	0.997	N
IPW	645	SW	0.949	0.997	N
Ln(IPW)	645	SW	0.940	0.997	N
Monthly Mean Measurements					
Td	174	SW	0.974	0.993	N
IPW	174	SW	0.953	0.993	N
Ln(IPW)	174	SW	0.937	0.993	N
Yearly Mean Measurements					
Td	37	SW	0.973	0.972	Y
IPW	37	SW	0.932	0.972	N
Ln(IPW)	37	SW	0.952	0.972	N

El Paso, TX					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurements					
Td	36112	KS	0.102	0.005	N
IPW	36112	KS	0.060	0.005	N
Ln(IPW)	36112	KS	0.118	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.076	0.023	N
IPW	1505	KS	0.112	0.023	N
Ln(IPW)	1505	KS	0.120	0.023	N
Weekly Measurements					
Td	645	SW	0.923	0.997	N

IPW	645	SW	0.968	0.997	N
Ln(IPW)	645	SW	0.915	0.997	N
Monthly Measurements					
Td	174	SW	0.935	0.993	N
IPW	174	SW	0.962	0.993	N
Ln(IPW)	174	SW	0.918	0.993	N
Yearly Measurements					
Td	37	SW	0.972	0.972	Y
IPW	37	SW	0.979	0.972	Y
Ln(IPW)	37	SW	0.974	0.972	Y

Flagstaff, AZ					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurements					
Td	36112	KS	0.072	0.005	N
IPW	36112	KS	0.055	0.005	N
Ln(IPW)	36112	KS	0.095	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.092	0.023	N
IPW	1505	KS	0.071	0.023	N
Ln(IPW)	1505	KS	0.095	0.023	N
Weekly Measurements					
Td	645	SW	0.958	0.997	N
IPW	645	SW	0.966	0.997	N
Ln(IPW)	645	SW	0.951	0.997	N
Monthly Measurements					
Td	174	SW	0.969	0.993	N
IPW	174	SW	0.964	0.993	N
Ln(IPW)	174	SW	0.921	0.993	N
Yearly Measurements					
Td	37	SW	0.979	0.972	Y
IPW	37	SW	0.975	0.972	Y
Ln(IPW)	37	SW	0.970	0.972	N

Guaymas, MX					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurements					
Td	36112	KS	0.132	0.005	N
IPW	36112	KS	0.108	0.005	N
Ln(IPW)	36112	KS	0.151	0.005	N

Three-Day Mean Measurements					
Td	1505	KS	0.154	0.023	N
IPW	1505	KS	0.129	0.023	N
Ln(IPW)	1505	KS	0.158	0.023	N
Weekly Measurements					
Td	645	SW	0.865	0.997	N
IPW	645	SW	0.905	0.997	N
Ln(IPW)	645	SW	0.849	0.997	N
Monthly Measurements					
Td	174	SW	0.838	0.993	N
IPW	174	SW	0.908	0.993	N
Ln(IPW)	174	SW	0.863	0.993	N
Yearly Measurements					
Td	37	SW	0.974	0.972	Y
IPW	37	SW	0.986	0.972	Y
Ln(IPW)	37	SW	0.983	0.972	Y

Las Vegas, NV					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurements					
Td	36112	KS	0.026	0.005	N
IPW	36112	KS	0.083	0.005	N
Ln(IPW)	36112	KS	0.042	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.038	0.023	N
IPW	1505	KS	0.084	0.023	N
Ln(IPW)	1505	KS	0.040	0.023	N
Weekly Measurements					
Td	645	SW	0.961	0.997	N
IPW	645	SW	0.994	0.997	N
Ln(IPW)	645	SW	0.990	0.997	N
Monthly Measurements					
Td	174	SW	0.980	0.993	N
IPW	174	SW	0.992	0.993	N
Ln(IPW)	174	SW	0.969	0.993	N
Yearly Measurements					
Td	37	SW	0.974	0.972	Y
IPW	37	SW	0.973	0.972	Y
Ln(IPW)	37	SW	0.983	0.972	Y

Midland, TX					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurement					
Td	36112	KS	0.075	0.005	N
IPW	36112	KS	0.030	0.005	N
Ln(IPW)	36112	KS	0.083	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.153	0.023	N
IPW	1505	KS	0.098	0.023	N
Ln(IPW)	1505	KS	0.133	0.023	N
Weekly Measurements					
Td	645	SW	0.939	0.997	N
IPW	645	SW	0.988	0.997	N
Ln(IPW)	645	SW	0.940	0.997	N
Monthly Measurements					
Td	174	SW	0.836	0.993	N
IPW	174	SW	0.901	0.993	N
Ln(IPW)	174	SW	0.809	0.993	N
Yearly Measurements					
Td	37	SW	0.972	0.972	Y
IPW	37	SW	0.977	0.972	Y
Ln(IPW)	37	SW	0.970	0.972	N

Phoenix, AZ					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurements					
Td	36112	KS	0.076	0.005	N
IPW	36112	KS	0.061	0.005	N
Ln(IPW)	36112	KS	0.103	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.080	0.023	N
IPW	1505	KS	0.072	0.023	N
Ln(IPW)	1505	KS	0.107	0.023	N
Weekly Measurements					
Td	645	SW	0.958	0.997	N
IPW	645	SW	0.961	0.997	N
Ln(IPW)	645	SW	0.936	0.997	N
Monthly Measurements					

Td	174	SW	0.951	0.993	N
IPW	174	SW	0.960	0.993	N
Ln(IPW)	174	SW	0.935	0.993	N
Yearly Measurements					
Td	37	SW	0.980	0.972	Y
IPW	37	SW	0.990	0.972	Y
Ln(IPW)	37	SW	0.991	0.972	Y

San Diego, CA					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurement					
Td	36112	KS	0.017	0.005	N
IPW	36112	KS	0.090	0.005	N
Ln(IPW)	36112	KS	0.033	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.025	0.023	N
IPW	1505	KS	0.104	0.023	N
Ln(IPW)	1505	KS	0.060	0.023	N
Weekly Measurements					
Td	645	SW	0.997	0.997	Y
IPW	645	SW	0.951	0.997	N
Ln(IPW)	645	SW	0.986	0.997	N
Monthly Measurements					
Td	174	SW	0.996	0.993	Y
IPW	174	SW	0.993	0.993	Y
Ln(IPW)	174	SW	0.994	0.993	Y
Yearly Measurements					
Td	37	SW	0.973	0.972	Y
IPW	37	SW	0.979	0.972	Y
Ln(IPW)	37	SW	0.979	0.972	Y

Tucson, AZ					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurements					
Td	36112	KS	0.090	0.005	N
IPW	36112	KS	0.072	0.005	N
Ln(IPW)	36112	KS	0.117	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.103	0.023	N
IPW	1505	KS	0.085	0.023	N

Ln(IPW)	1505	KS	0.129	0.023	N
Weekly Measurements					
Td	645	SW	0.936	0.997	N
IPW	645	SW	0.951	0.997	N
Ln(IPW)	645	SW	0.920	0.997	N
Monthly Measurements					
Td	174	SW	0.934	0.993	N
IPW	174	SW	0.954	0.993	N
Ln(IPW)	174	SW	0.934	0.993	N
Yearly Measurements					
Td	37	SW	0.976	0.972	Y
IPW	37	SW	0.980	0.972	Y
Ln(IPW)	37	SW	0.985	0.972	Y

Yuma, AZ					
Variable	N-Size	Normality Test	Test Statistic	Critical Value	Normal (Y/N)
Three-Hour Measurements					
Td	36112	KS	0.080	0.005	N
IPW	36112	KS	0.057	0.005	N
Ln(IPW)	36112	KS	0.071	0.005	N
Three-Day Mean Measurements					
Td	1505	KS	0.085	0.023	N
IPW	1505	KS	0.061	0.023	N
Ln(IPW)	1505	KS	0.073	0.023	N
Weekly Measurements					
Td	645	SW	0.958	0.997	N
IPW	645	SW	0.974	0.997	N
Ln(IPW)	645	SW	0.964	0.997	N
Monthly Measurements					
Td	174	SW	0.889	0.993	N
IPW	174	SW	0.967	0.993	N
Ln(IPW)	174	SW	0.932	0.993	N
Yearly Measurements					
Td	37	SW	0.975	0.972	Y
IPW	37	SW	0.976	0.972	Y
Ln(IPW)	37	SW	0.979	0.972	Y

With the exception of San Diego, CA, data from all ten cities of this study showed non-normal distributions for all timescales, up to and including: three-hour, three-day,

weekly, monthly , but were all normally distributed during the yearly mean measurements. In other words, IPW and Td were normally distributed across all locations if all values were averaged for a given NAM season. Sub-annual variables of the monsoon can vary extensively, but the annual monsoon displays a more uniform pattern.

A normal distribution would imply that there is a steady increase in surface moisture and moisture aloft followed by a steady dissipation of that moisture, which is the understood migratory pattern of NAM moisture. However, as the timescale is increasingly restricted, this normal distribution fails to exist as would be expected. For example, if there was a persistent presence of moisture in one month compared to another month, (e.g., June vs. July) this will result in a skewed and kurtotic distribution. Consequently, these results identify which statistical test is best suited for this study. There are two tests that I am conducting: a) for all data sets that are non-normal, I will apply Spearman's rank-order correlation (ρ) analysis, and (b) for datasets that showed a normal distribution, I will apply Pearson's (r) correlation analysis.

Pearson's (r) correlation is used to examine the association between an independent and dependent variable. It is a widely recognized statistic and has been often used in monsoonal research (as cited in chapter 2, e.g., Reitan 1963; Bolsenga 1964; Benwell 1965). The output is bounded by a range from $-1 \leq r \leq 1$. At $r = 1$ the data show a perfect positive correlation between two variables while at $r = -1$ the data show a perfect negative correlation between the two variables, and at $r = 0$, there is no correlation. There are certain thresholds set forth to determine the strength of any correlation. If $0 \leq |r| \leq 0.3$ that is considered a weak correlation, while $0.3 < |r| \leq 0.7$ is

considered a moderate correlation, and |r| > 0.7 is considered a strong correlation (Wilks 2011). One implication of the Pearson's r is that the "square of the Pearson correlation…specifies the proportion of the variability of one of either x or y that is…the variance of one variable 'explained' by the other" (Wilks 2011 p. 52). It is important to note that this variability is not necessarily the result of any causative relationship as any number of different physical mechanisms could be driving their variability. Below is the equation for Pearson's (r).

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}[(x_i-\bar{x})(y_i-\bar{y})]}{[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2]^{\frac{1}{2}}[\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2]^{\frac{1}{2}}} \quad (4.3)$$

Where,
x and y refer to the two independent variables and n is the sample size.

Eqn. 4.3 is simply the covariance of the two variables (x,y) divided by the product of the standard deviation of x and the standard deviation of y. However, there are a few assumptions that need to be made in order for this test to be used (Wilks 2011). The first assumption is linearity and the second is that there are no outliers. If either of these are not met, then the output from the test can be misleading. Thus, any data in my study determined to be non-normally distributed or with excessive outliers cannot be analyzed with this test and another correlation test must be used.

Spearman's rank-order (ρ) test does not require these assumptions and only requires that the data be ordinal (put into a rank) and monotonic (as x increases/decreases, y increases/decreases correspondingly) (Wilks 2011). The test is basically conducted in the same manner as Pearson's (r), but computed using the ranks of

the data. It does this by transforming the data by ranking the variables before conducting the correlation test. The range of ρ is the same as the range of r, and the correlation thresholds are the same. The equation for Spearman's (ρ) is below:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)} \quad (4.4)$$

Where,
D_i^2= is the difference in ranks between the *i*th pair of data values and n is the sample size.

The Spearman rank-order (ρ) test is considered to be "robust and resistant" as it does not require normality, linearity, and adjusts for outliers (Wilks 2011). Consequently, this test will be applied to the majority of datasets in my study as they are mostly non-normally distributed.

4.4 Data analysis for the study

Using correlation analysis, I examined the relationship of IPW/Td across five different time scales for my ten separate city locations. The rationale for city selection was discussed in Chapter 2 while the primary reason for the use of five different timescales comes from Reber and Swope's (1972) criticisms of Reitan (1963) in using mean monthly averages of IPW and Td to enhance the correlation between the two. Reitan (1963) had obtained correlation coefficients that approached r = 1. Consequently, I will evaluate this relationship not only spatially but across various averaged time scales. If Reber and Swope (1972) were correct in their criticisms, the high correlation between IPW and Td should breakdown as the timescale average decreases. In other words, the diurnal relationship between IPW and Td should be more variable than then the relationship of these variables averaged across the entire NAM season.

The first timescale analysis was the non-averaged three-hour measurements provided by the NARR. These data were uploaded into SPSS, and determined to be non-normally distributed, so I applied Spearman's ρ correlation test. The correlation results for the ten locations yielded a range of values ($0.626 \leq \rho \leq 0.926$) for the IPW and Td relationship (with identical results for the Td and Ln(IPW) relationship) with $p < 0.001$ for all results. Consequently, there is a shared variance of between 39 to 86 percent.

Interestingly, there seems to be a variability in the correlation between cities on the perimeter of the study region compared to the inside. The range for cities on the fringes of my study's domain (i.e., Guaymas, MX, Las Vegas, NV, Midland, TX, and San Diego, CA) was $0.625 \leq \rho \leq 0.914$. The internal cities of my study's domain (i.e., Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Phoenix, AZ, Tucson, AZ, Yuma, AZ) had a much smaller range $0.801 \leq \rho \leq 0.926$. This may suggest that there is much smaller variability in the IPW/Td relationship for cities that are largely associated with the monsoon as opposed to those either outside of the NAM's region of influence (e.g., San Diego, CA and Midland, TX) or on the very fringes of it (e.g., Las Vegas, NV). The results are presented in Table 4.4.

Table 4.4. Spearman's rank-order (ρ) correlation, shared variance (%), and significant values, for three-hour measurements for the length-of-record (1979-2015) for each of the ten location (Albuquerque, NM, El Paso, TX Flagstaff, AZ, Guaymas, MX, Las Vegas, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ)

City	N-Size	(ρ) Td & IPW	ρ^2 (%)	Sig (2-tail)	(ρ) Td & Ln(IPW)	ρ^2 (%)	Sig (2-tail)
Albuquerque, NM	36112	0.872	76	p < 0.001	0.872	76	p < 0.001
El Paso, TX	36112	0.863	74	p < 0.001	0.863	74	p < 0.001

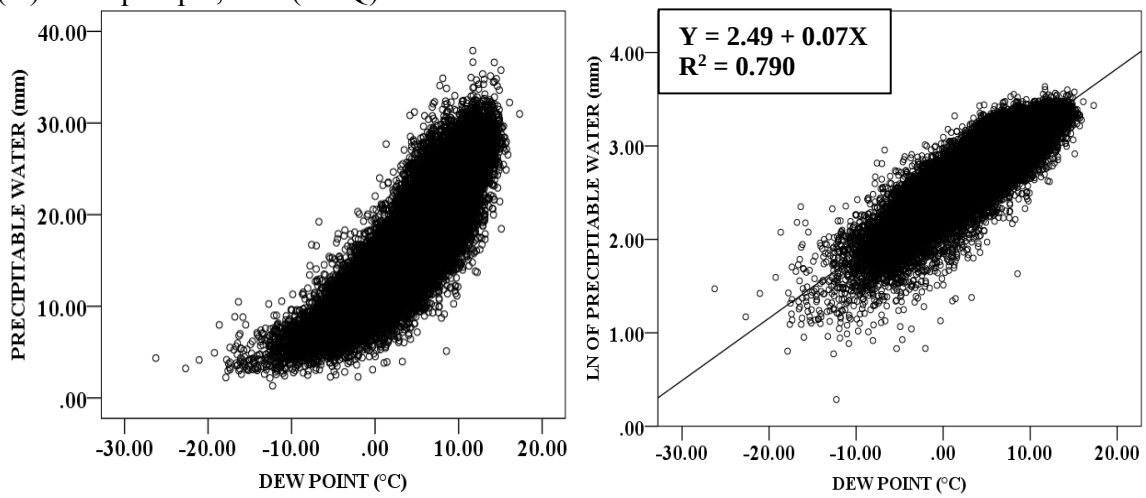
Flagstaff, AZ	36112	0.881	77	p < 0.001	0.881	77	p < 0.001
Guaymas, MX	36112	0.646	42	p < 0.001	0.646	42	p < 0.001
Las Vegas, NV	36112	0.914	85	p < 0.001	0.914	85	p < 0.001
Midland, TX	36112	0.625	39	p < 0.001	0.625	39	p < 0.001
Phoenix, AZ	36112	0.926	86	p < 0.001	0.926	86	p < 0.001
San Diego, CA	36112	0.692	48	p < 0.001	0.692	48	p < 0.001
Tucson, AZ	36112	0.893	80	p < 0.001	0.893	80	p < 0.001
Yuma, AZ	36112	0.801	64	p < 0.001	0.801	64	p < 0.001

The three cities with the lowest IPW/Td relationship for this timescale were Midland, TX, Guaymas, MX, and San Diego, CA with ρ-values of 0.625, 0.646, 0.692, respectively. Two of these cities are located next to a large body water. Consequently, their relatively low correlation is likely the result of three things: 1.) Marine layers that cause a spike in surface Td point but are low-level and do not provide much in terms of IPW, 2.) Cold sea-surface temperatures can inhibit IPW irrespective of any existing forcing (Chaboureau et al. 1998), 3.) Absence of dynamic forcing in these areas (Reber and Swope 1972). Conversely, Midland, TX is largely removed from any source of water, but it experiences an absence of summer-time dynamic systems as they typically remain north of this area (Schwartz 1968).

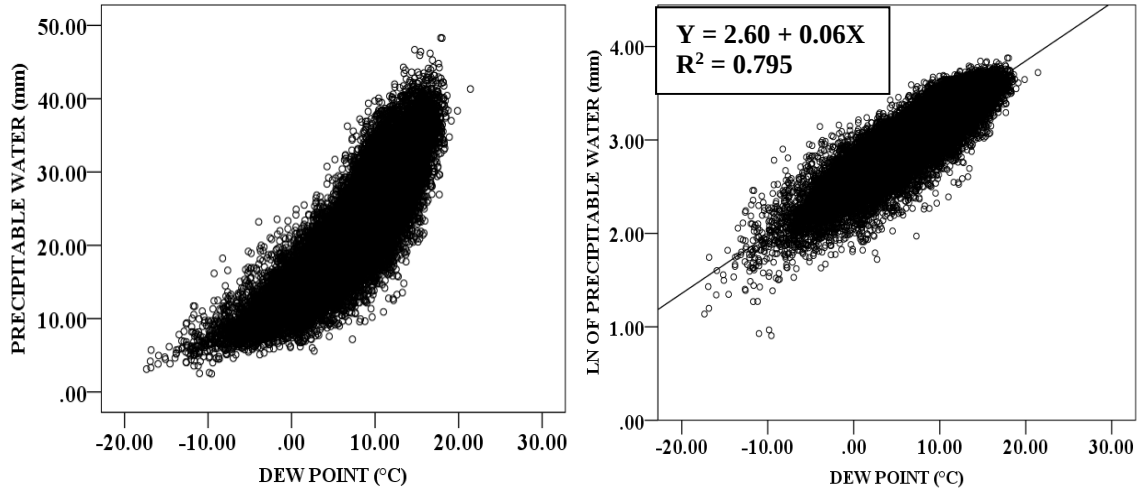
Collectively, it appears that the regions more closely associated with the NAM provided the best relationship of IPW/Td for three-hour measurements. Of these cities, Phoenix, AZ, Las Vegas, NV, and Tucson, AZ showed the greatest associations. The

correlations for these three cities were 0.926, 0.914, and 0.893, respectively. This indicates good vertical transport of surface moisture throughout the total atmosphere for three-hour measurements from 1979-2015. Furthermore, these results show the potential usefulness of using Td as a proxy for IPW apropos monsoon onset and thunderstorm genesis for Phoenix, AZ and Tucson, AZ. However, because this relationship was the cornerstone of the aforementioned NAM onset, I computed a linear regression for this set of data and used it to recalculate the associated Td and IPW values for all of the cities in my study. Fig. 4.2 is the scatterplots for all ten cities and the associated best-fit linear regression equations.

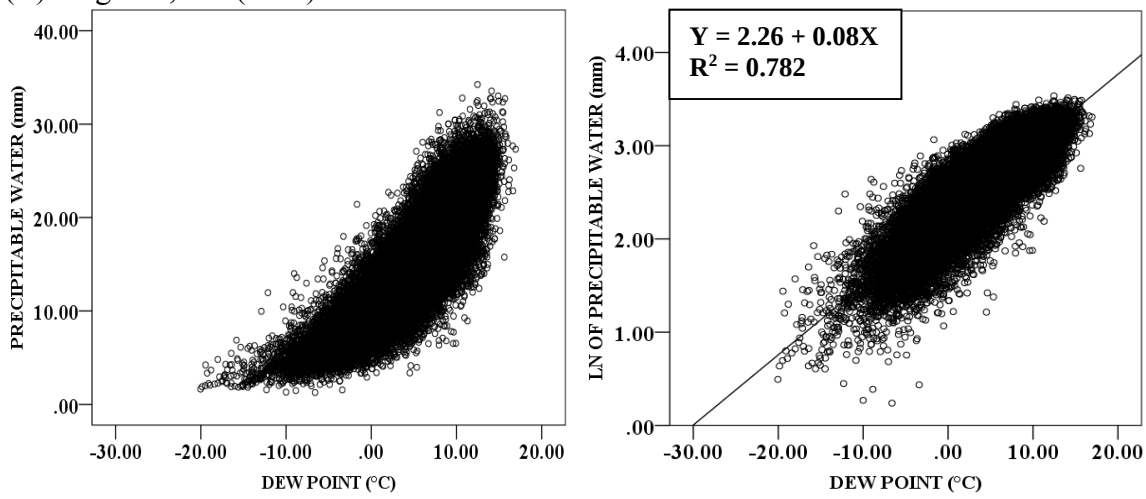
(A) Albuquerque, NM (ABQ)



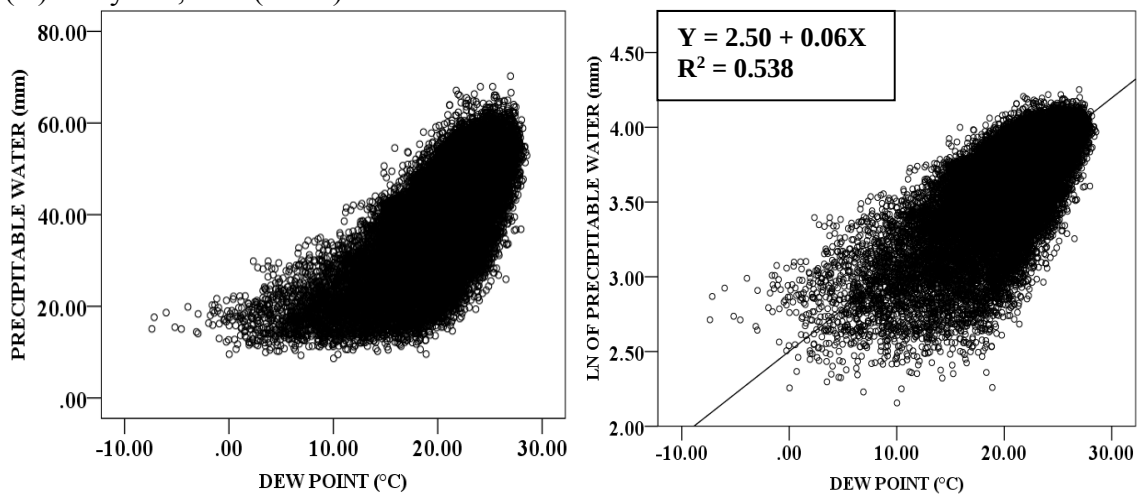
(B) El Paso, TX (ELP)



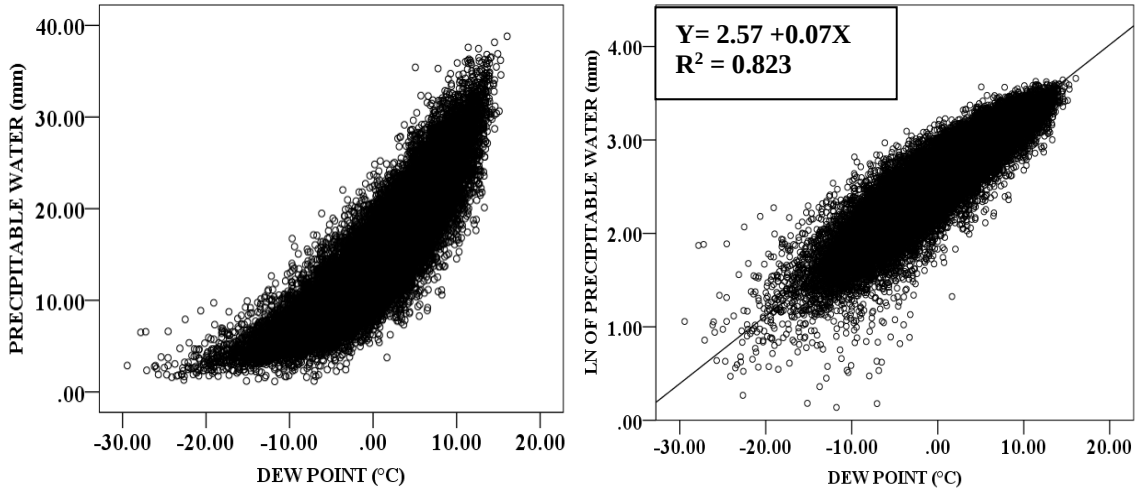
(C) Flagstaff, AZ (FLG)



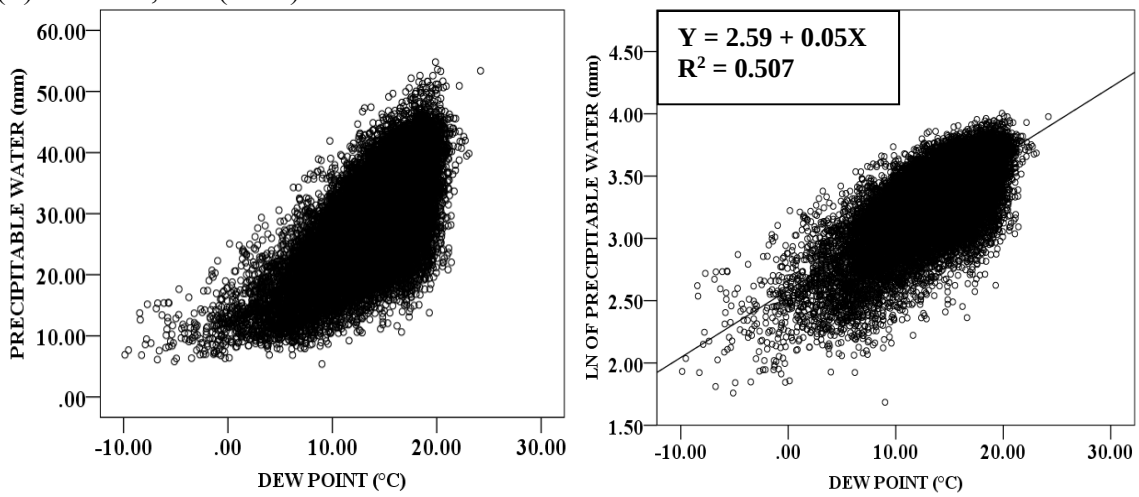
(D) Guaymas, MX (GUY)



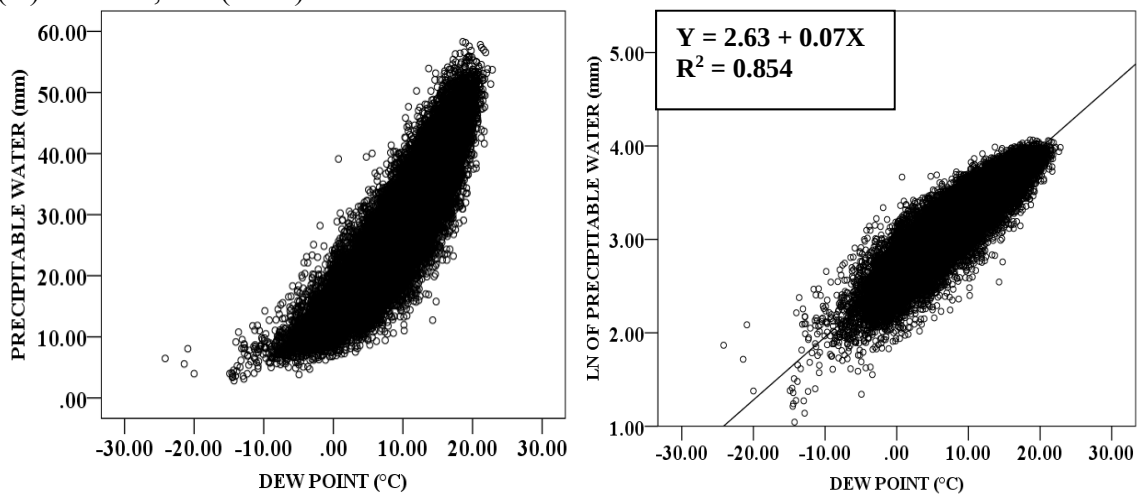
(E) Las Vegas, NV (LAS)



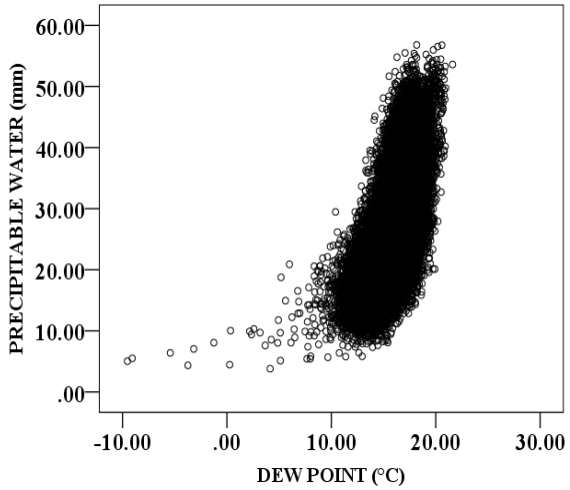
(F) Midland, TX (MID)

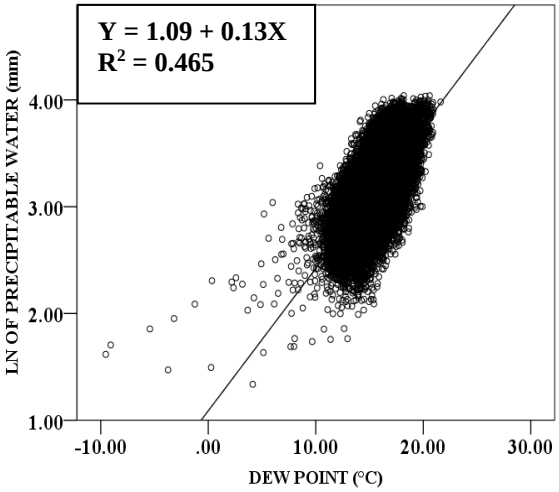


(G) Phoenix, AZ (PHX)

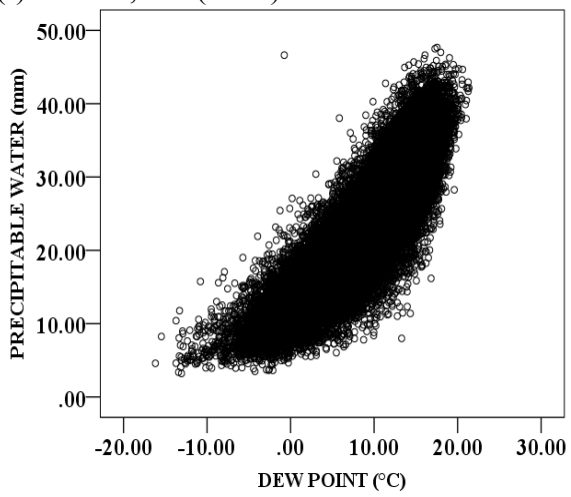


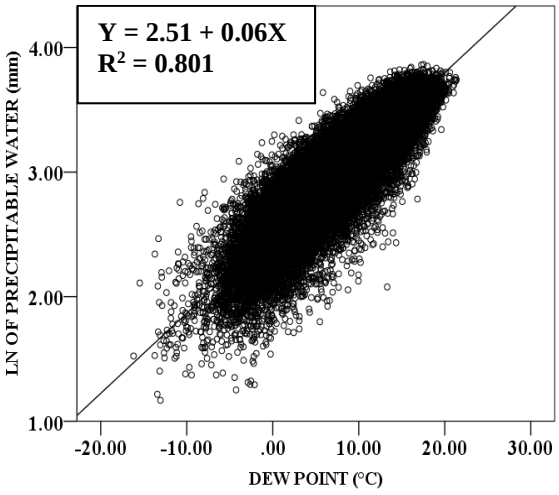
(H) San Diego, CA (SAN)



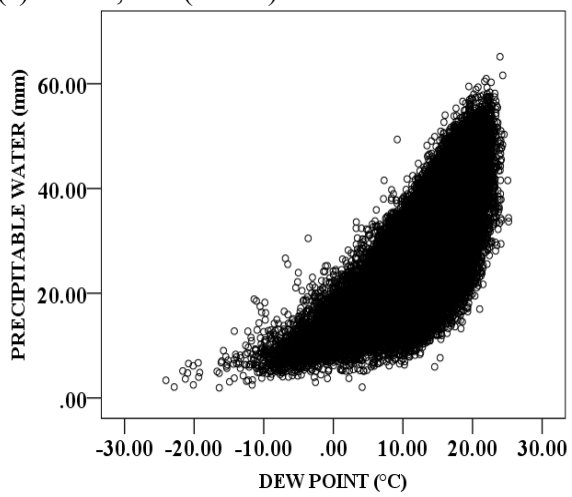


(I) Tucson, AZ (TUC)





(J) Yuma, AZ (YUM)



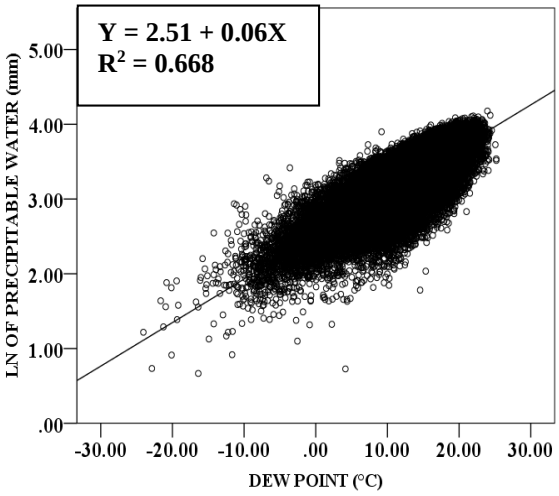


Fig. 4.2. Scatterplots of the three-hour averaged IPW (in mm) /Td (in °C) and Ln(IPW) (in ln(mm)) / Td (°C) for the ten study cities. A) Albuquerque, NM, B) EL Paso, TX, C) Flagstaff, AZ, D) Guaymas, MX, E) Las Vegas, NV, F) Midland, TX, G) Phoenix, AZ, H) San Diego, CA, I) Tucson, AZ, J) Yuma, AZ. Ln(IPW)/Td shows the associated best-fit trend line, the calculated linear regression equation, and R^2 value.

Using the linear regression equation, I calculated the computed Td for 25.4mm (1.00") of IPW and the IPW values for past monsoonal criteria of 12.78°C(55°F) and 12.22°C(54°F) Td thresholds once used for Phoenix, AZ and Tucson, AZ, respectively. While these standards were never applied to the other cities in my study, it is interesting to see the implications of those thresholds. Table 4.5 shows the results of this examination.

Table 4.5. The computed value of either IPW (MM) or Td (°C and °F) for all ten cities (Albuquerque, NM (ABQ), El Paso, TX (ELP), Flagstaff, AZ (FLG), Guaymas, MX (GUY), Las Vegas, NV (LAS), Midland, TX (MID), Phoenix, AZ (PHX), San Diego, CA (SAN), Tucson, AZ (TUC), Yuma, AZ (YUM)) using 25.4mm (IPW), 12.78°C (Td), and 12.22°C (Td).

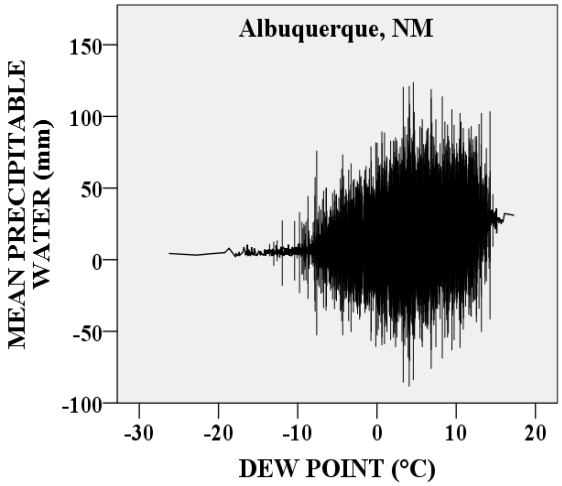
	ABQ	ELP	FLG	GUY	LAS	MID	PHX	SAN	TUC	YUM
IPW	Td °C (°F)	Td °C (°F)	Td °C (°F)	Td °C (°F)	Td °C (°F)	Td °C (°F)	Td °C (°F)	Td °C (°F)	Td °C (°F)	Td °C (°F)
25.4mm (1.00")	10.64 (51.15)	10.58 (51.04)	12.18 (53.93)	12.25 (54.05)	9.50 (49.1)	12.89 (55.20)	8.64 (47.55)	16.50 (61.7)	12.08 (53.74)	12.08 53.74)
Td	IPW mm	IPW Mm	IPW mm	IPW mm	IPW mm	IPW mm	IPW mm	IPW mm	IPW mm	IPW mm
12.78°C (55°F)	29.51 (1.16")	28.99 (1.14")	26.64 (1.04")	26.23 (1.03")	31.96 (1.26")	25.25 (1.00")	33.94 (1.33")	15.66 (0.62")	26.49 (1.04")	26.49 (1.04")
12.22°C (54°F)	28.37 (1.12")	28.03 (1.10")	25.47 (1.00")	25.36 (1.00")	30.73 (1.21")	24.56 (0.97")	32.64 (1.29")	14.56 (0.57")	25.62 (1.00")	25.62 (1.00")

Not surprisingly, the dew point temperature is highly variable compared to the IPW based on location (Fig. 4.2). The dew point threshold for Phoenix, AZ (12.78°C) does not correspond to 25.4mm of IPW, rather it is closer to 33.94mm (1.33"). That is near the result determined by Skindlov (2007) using only data from 2005-2006. In actuality, Phoenix, AZ only requires a Td of ~ 8.64°C to correspond to 1" of precipitable

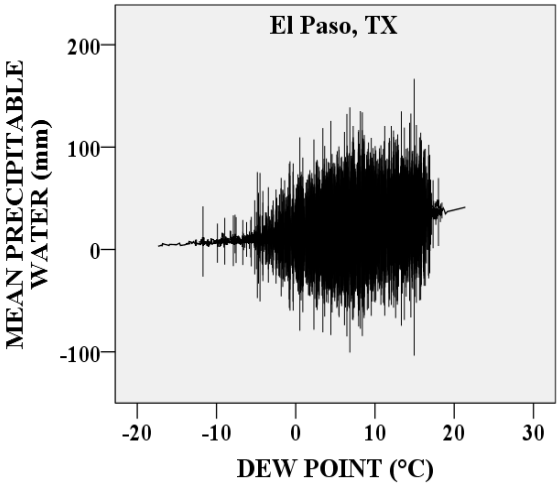
water. Tucson, AZ, the only other city with a past surface dew point threshold to mark the start of the monsoon season, is well-matched with 25.62mm (1.00”) of IPW. The variability between other cities is largely the byproduct of different availability of uplift and differences of elevation on surface dew point and IPW measurements.

The shape of the IPW/Td scatterplots (Fig. 4.2) is identical to a visually apparent logarithmic distribution that could be transformed into a linear distribution. However, there is a sharp upper-limit edge on the top of the IPW/Td graph (Fig. 4.2). This implies that there is a marked upper-boundary to the amount of precipitable water possible given a particular surface dew point. For example, Phoenix, AZ (g) at a 5°C Td, using the regression equation yields an IPW of 19.70mm. However, examining 5°C graphically, an IPW up to 31.00mm is possible but not more. In the entire NARR record from 1979-2015, an IPW greater than 31.00mm has never occurred for a 5°C Td in Phoenix, AZ. This shows there is apparently an upper-limit to the IPW based on Td. Additionally, there appears to be a wide variability in the possible IPW outcomes with increased surface Td values. To better represent this finding, Fig. 4.3 shows the relationship between Td and the mean IPW with the respective standard error bars at the 5% confidence interval.

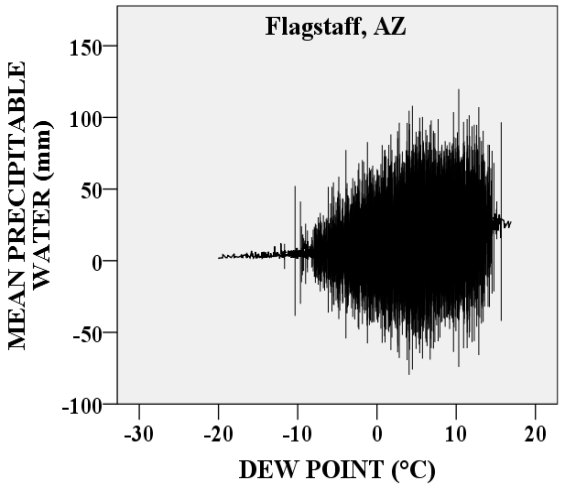
(A) Albuquerque, NM



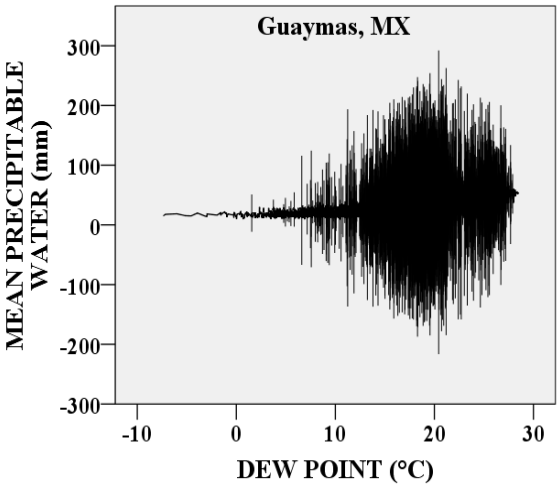
(B) El Paso, TX



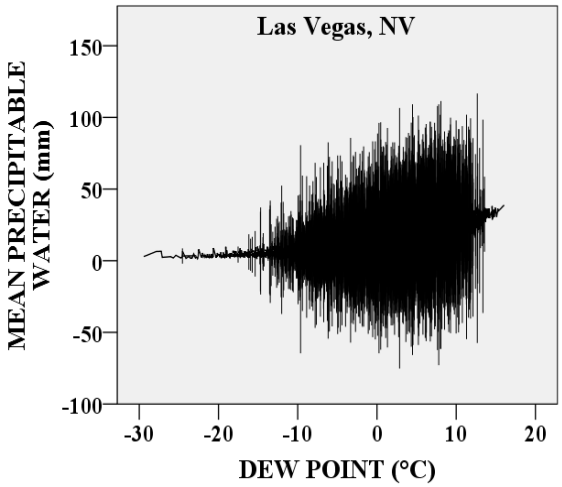
(C) Flagstaff, AZ



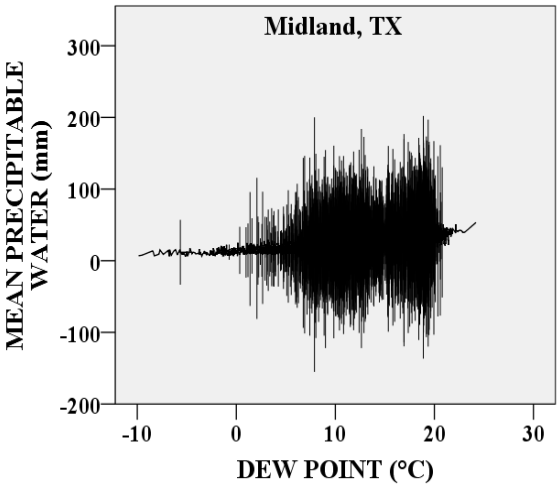
(D) Guaymas, MX



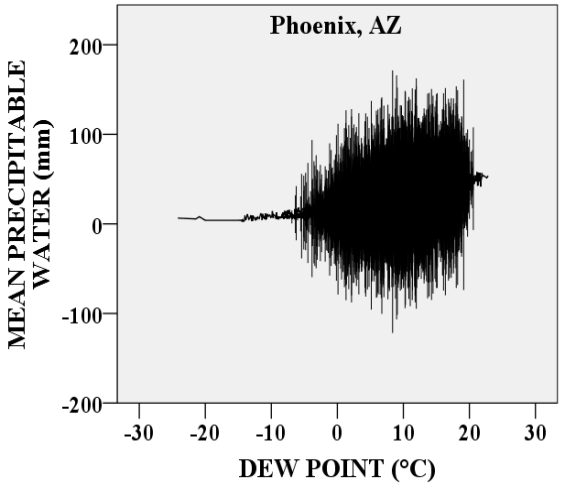
(E) Las Vegas, NV



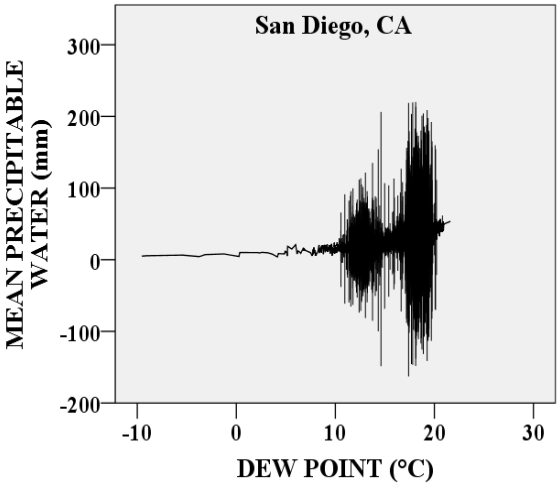
(F) Midland, TX



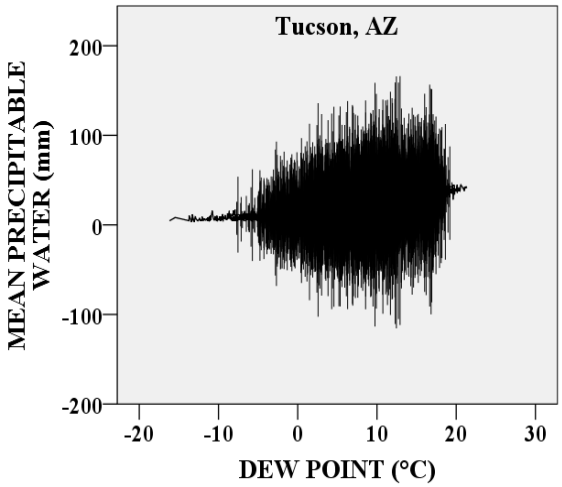
(G) Phoenix, AZ



(H) San Diego, CA



(I) Tucson, AZ



(J) Yuma, AZ

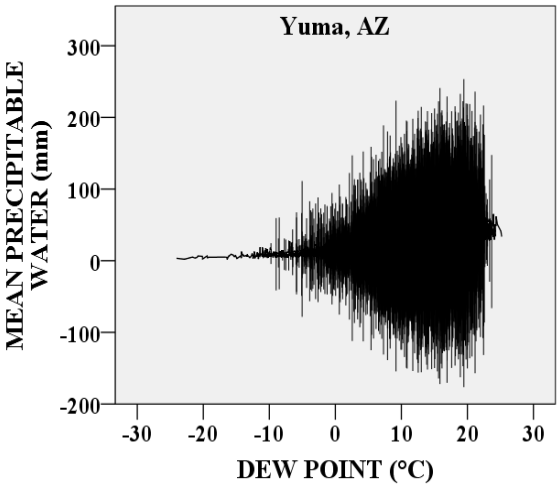


Fig. 4.3. The three-hour measurements for Td (in °C) against mean IPW (in mm) with standard error bars at a 5% confidence interval for the entire length of record (1979-2015) for A) Albuquerque, NM, B) El Paso, TX, C) Flagstaff, AZ, D) Guaymas, MX, E) Las Vegas, NV, F) Midland, TX, G) Phoenix, AZ, H) San Diego, CA, I) Tucson, AZ, J) Yuma, AZ).

Not surprisingly, for each of the locations, the mean IPW changes as Td changes (Fig. 4.3). There is a clear middle-ground Td temperature in which a very wide range of IPW outcomes exists. This likely suggests that certain physical mechanisms, such as vertical mixing, drive the IPW/Td relationship. In essence, as surface Td increases, the *potential* for a higher value of IPW correspondingly increases. However, in order for Td

to highly relate to a higher IPW for the whole atmosphere, there must be much greater vertical mixing of this low-level moisture. Consequently, the presence of a high Td does not guarantee a high IPW; rather, it only guarantees the *potential* for one, resulting in the large variability in the relationship. If uplift in a given region of high surface humidity is lacking, there will not be a correspondingly high IPW. Furthermore, there appears to be a median Td where the range of IPW begins to slightly narrow. For example in Tucson, AZ, at a Td ~14°C, there is a drop in the range of possible IPW outcomes as compared to dew points of 11°C/12°C. This trend is also subtly observed in Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Las Vegas, NV, Phoenix, AZ, and Yuma, AZ. The three "outlying cities" (Guaymas, MX, Midland, TX, and San Diego, CA) take on a completely different shape (similar to an hour glass turned sideways) that represents a narrowing of the range of mean IPW values near the median Td followed by a subsequent widening of values.

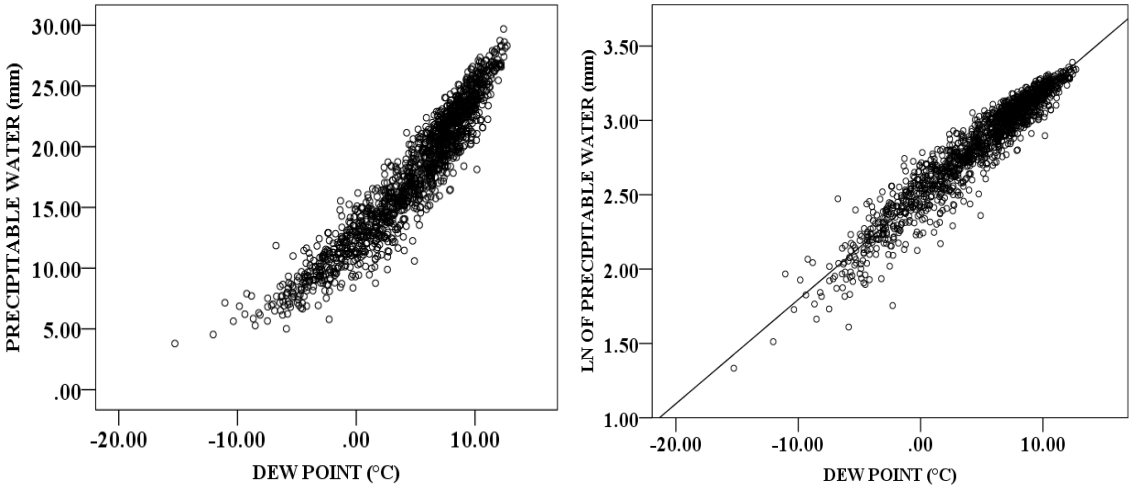
This variability shift is likely a byproduct of low-level moisture boosting surface based instability (i.e., CAPE). In other words, at some point, the presence of increased surface moisture simultaneously increases the atmosphere's ability to thermodynamically mix, thereby providing the IPW/Td relationship with a positive feedback mechanism. However, since this relationship for determining monsoon onset is dependent on three-days of measurements, a three day correlation analysis of Td/IPW relationship is appropriate. Table 4.6 illustrates the results of the correlation testing for this averaged timescale.

Table 4.6 Spearman's rank-order (ρ) correlation, shared variance (%), and significance values, for three-day mean measurements for the length-of-record (1979-2015) for each of the ten location (Albuquerque, NM, El Paso, TX Flagstaff, AZ, Guaymas, MX, Las Vegas, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ)

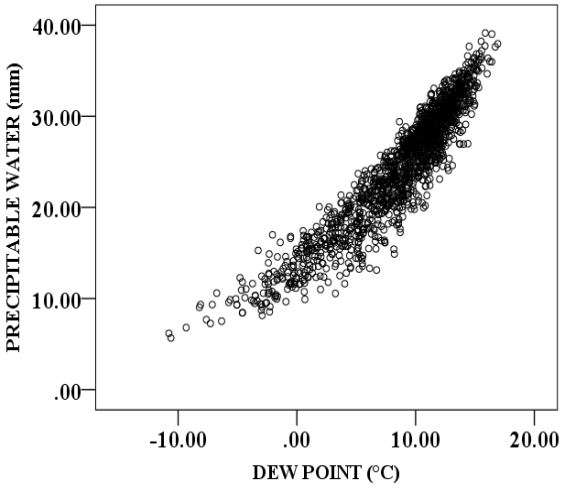
City	N-Size	(ρ) Td & IPW	ρ^2 (%)	Sig(2-tail)	(ρ) Td & Ln(IPW)	ρ^2 (%)	Sig(2-tail)
Albuquerque, NM	1505	0.953	91	p < 0.001	0.952	91	p < 0.001
El Paso, TX	1505	0.933	87	p < 0.001	0.932	87	p < 0.001
Flagstaff, AZ	1505	0.936	87	p < 0.001	0.935	87	p < 0.001
Guaymas, MX	1505	0.797	64	p < 0.001	0.796	63	p < 0.001
Midland, TX	1505	0.775	60	p < 0.001	0.772	60	p < 0.001
Las Vegas, NV	1505	0.952	91	p < 0.001	0.953	91	p < 0.001
Phoenix, AZ	1505	0.953	91	p < 0.001	0.953	91	p < 0.001
San Diego, CA	1505	0.773	60	p < 0.001	0.774	60	p < 0.001
Tucson, AZ	1505	0.935	87	p < 0.001	0.934	87	p < 0.001
Yuma, AZ	1505	0.909	83	p < 0.001	0.910	83	p < 0.001

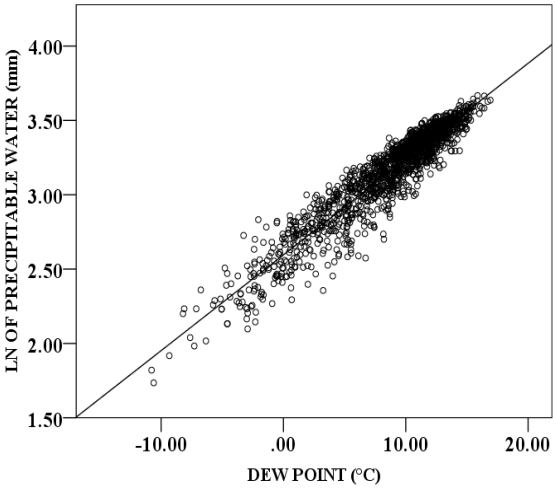
When compared to the results of the three-hour measurement correlation analysis, the ρ-values markedly increase, which aligns with the assertions of Reber and Swope (1972). The range of values for these data were $0.773 \leq \rho \leq 0.953$. The cities on the fringes of the study still showed lower correlations (e.g. Midland, TX $\rho = 0.775$) than the monsoonal interior (e.g., Flagstaff, AZ $\rho = 0.936$), but all values were greater than the 0.70 benchmark to consider this relationship strongly correlated.

(A) Albuquerque, NM

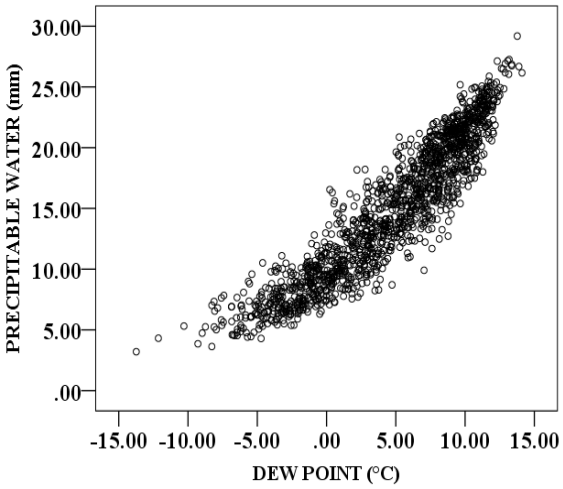


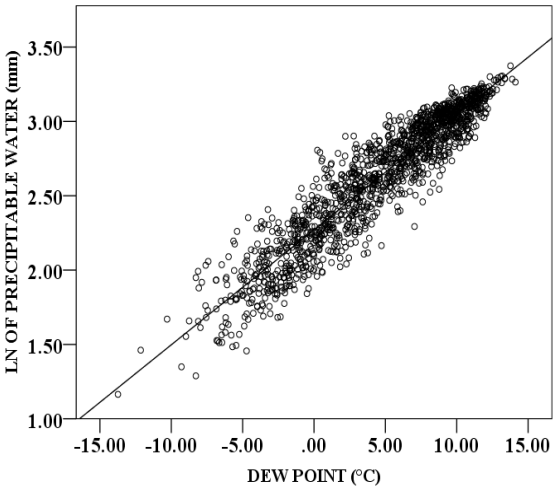
(B) El Paso, TX



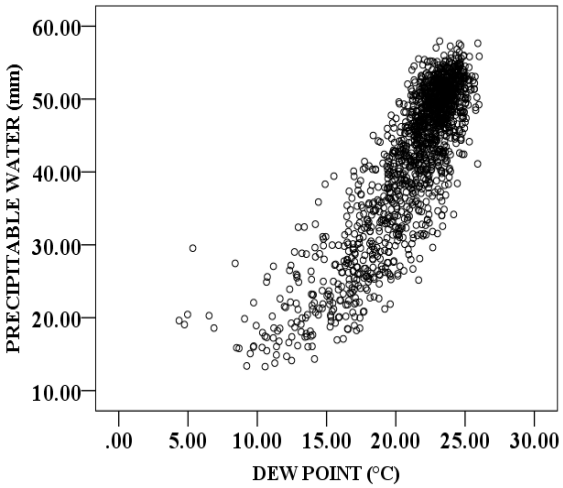


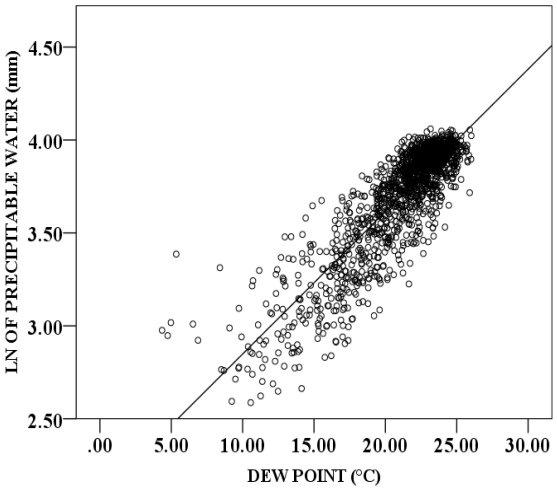
(C) Flagstaff, AZ



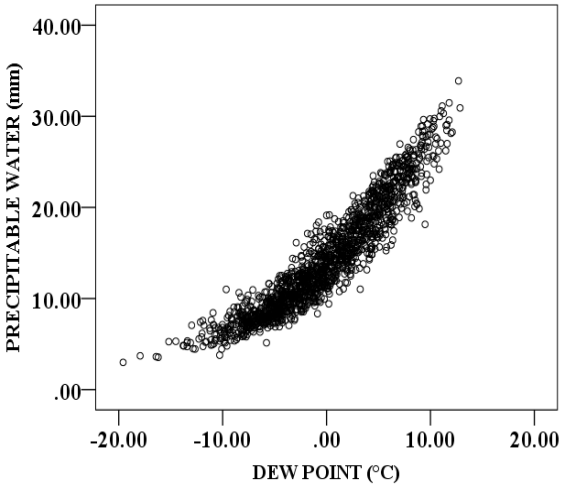


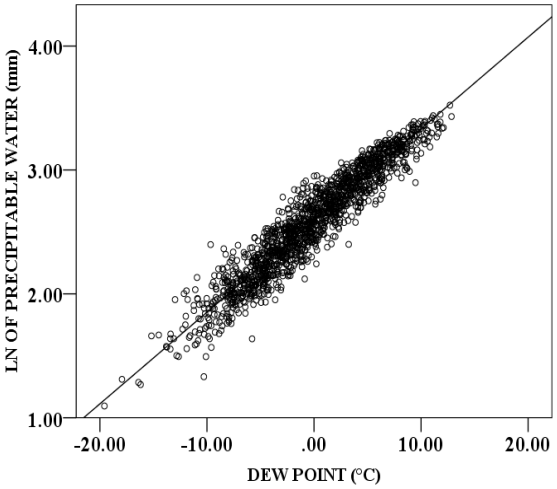
(D) Guaymas, MX



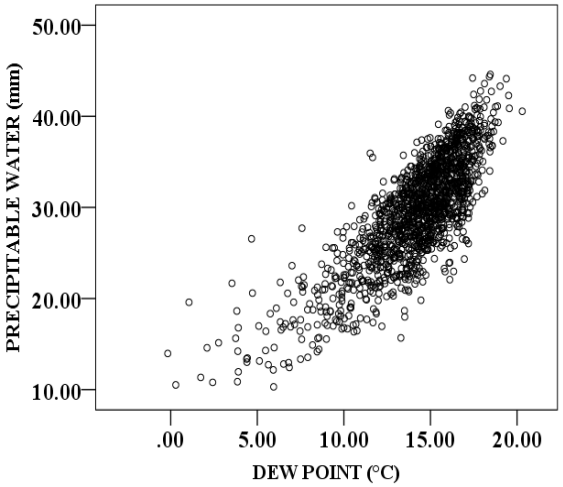


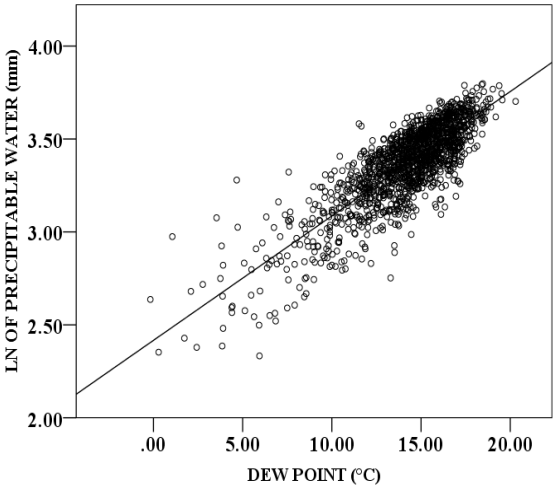
(E) Las Vegas, NV



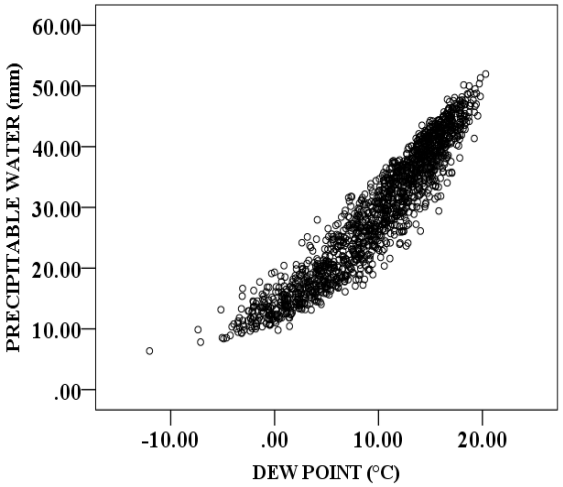


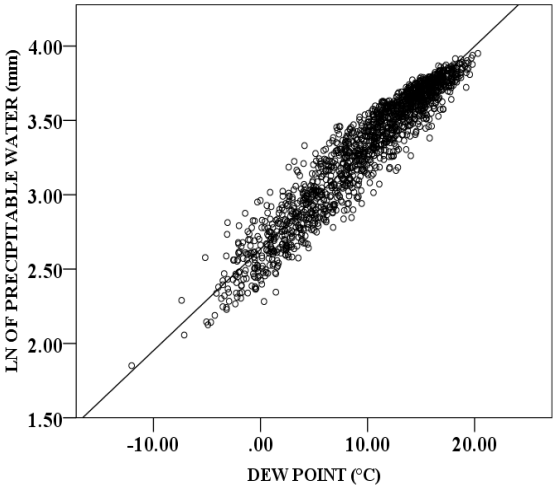
(F) Midland, TX



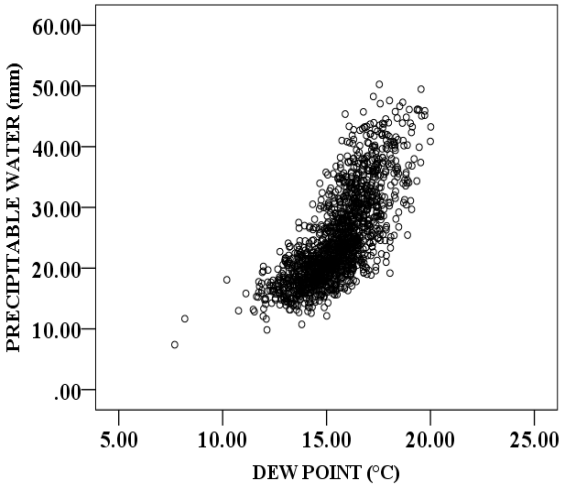


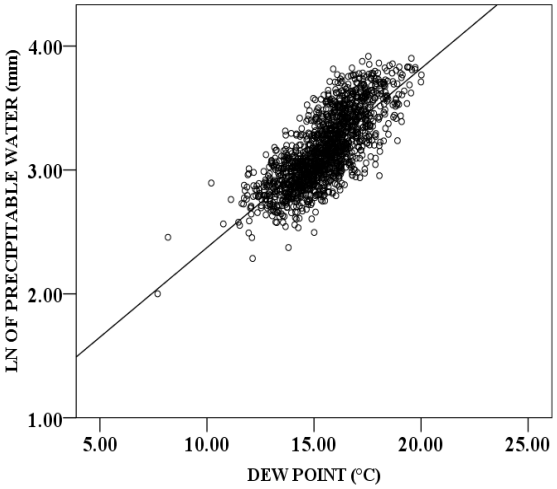
(G) Phoenix, AZ



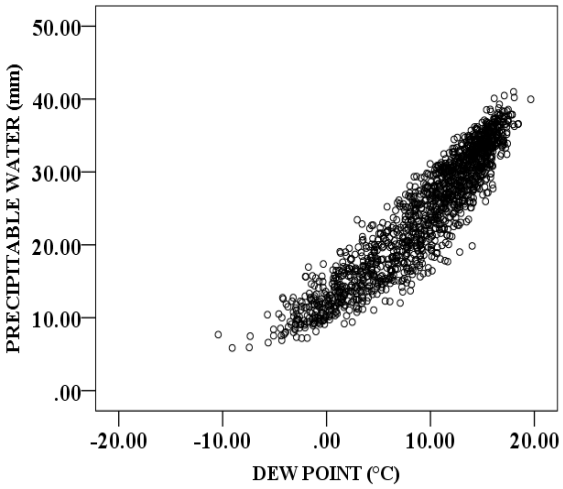


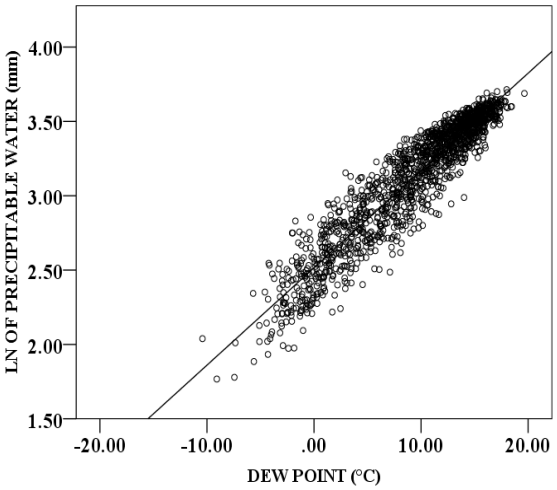
(H) San Diego, CA



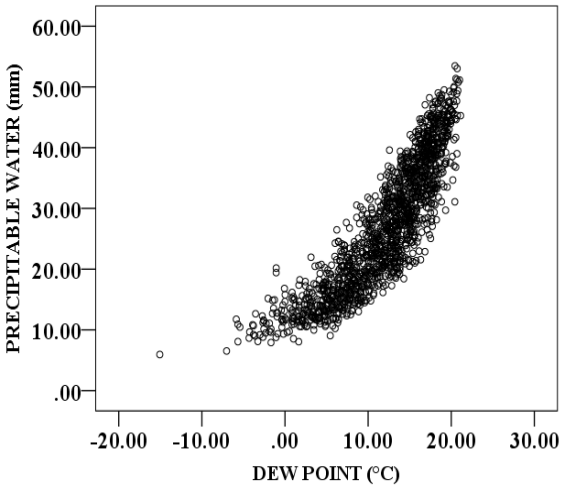


(I) Tucson, AZ





(J) Yuma, AZ



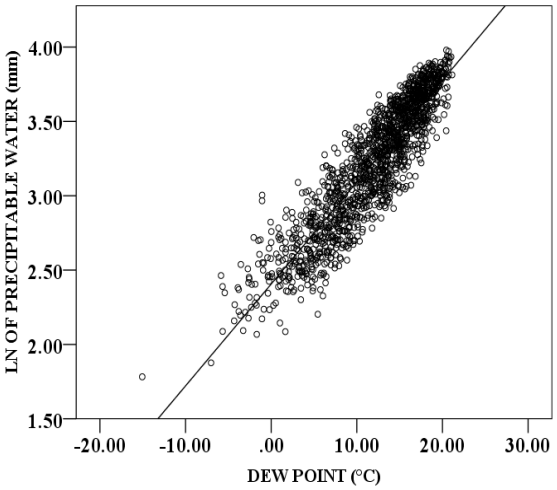


Fig. 4.4. Scatterplots for the three-day averaged values of IPW (in mm) /Td (in °C) and Ln(IPW) (in Ln(mm)) / Td (°C) for the ten study cities. A) Albuquerque, NM, B) El Paso, TX, C) Flagstaff, AZ, D) Guaymas, MX, E) Las Vegas, NV, F) Midland, TX, G) Phoenix, AZ, H) San Diego, CA, I) Tucson, AZ, J) Yuma, AZ. Ln(IPW)/Td shows the associated best-fit trend line.

Fig. 4.4 compares relatively well with the overall shape of scattergrams for the three-hour measurements. An interesting aspect of Fig. 4.4 is that the range of values decreases as Td increases. This is in comparison to the observation made in Fig 4.3 where the reverse held true. Thus, over the course of a three-day average for surface dew point, Td is more indicative of the IPW in the atmosphere and is not quite as susceptible to the large variability observed in the three-hour measurements. Similar findings are apparent in the weekly timescale (Table 4.7).

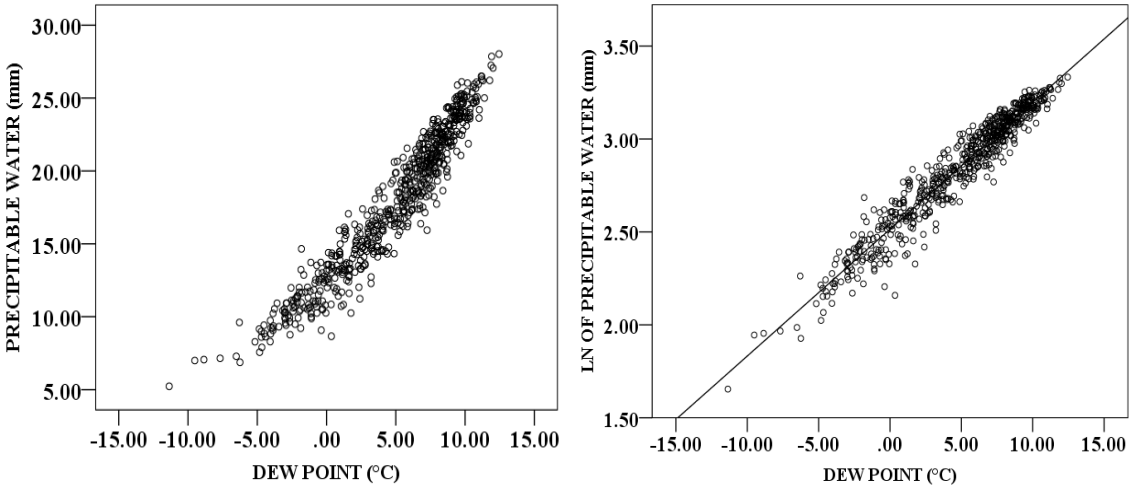
Table 4.7 Spearman's ρ correlation (ρ), shared variance (%), and significant values, for weekly mean measurements for the length-of-record (1979-2015) for each of the ten location (Albuquerque, NM, El Paso, TX Flagstaff, AZ, Guaymas, MX, Las Vegas, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ)

City	N-Size	(ρ) Td & IPW	ρ^2 (%)	Sig(2-tail)	(ρ) Td & Ln(IPW)	ρ^2 (%)	Sig(2-tail)
Albuquerque, NM	645	0.963	93	$p < 0.001$	0.961	92	$p < 0.001$
El Paso, TX	645	0.943	89	$p < 0.001$	0.940	89	$p < 0.001$
Flagstaff, AZ	645	0.949	90	$p < 0.001$	0.949	90	$p < 0.001$
Guaymas, MX	645	0.845	71	$p < 0.001$	0.844	71	$p < 0.001$
Las Vegas, NV	645	0.960	92	$p < 0.001$	0.962	93	$p < 0.001$
Midland, TX	645	0.830	69	$p < 0.001$	0.841	69	$p < 0.001$
Phoenix, AZ	645	0.961	92	$p < 0.001$	0.960	92	$p < 0.001$
San Diego, CA	645	0.803	64	$p < 0.001$	0.805	65	$p < 0.001$
Tucson, AZ	645	0.944	89	$p < 0.001$	0.943	89	$p < 0.001$
Yuma, AZ	645	0.928	87	$p < 0.001$	0.930	86	$p < 0.001$

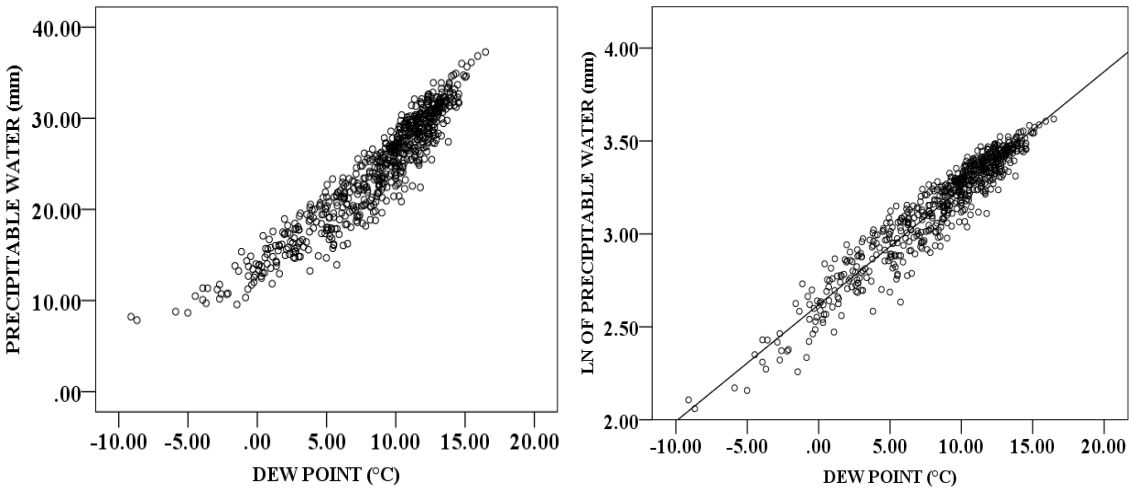
The trend that was observed from three-hour measurements to three-day averaged measurements continues in the weekly measurements (Fig. 4.4). The weekly mean range for ρ-values was less than that seen in the three-day measurement values. The range was

$0.803 \leq \rho \leq 0.963$. This once again identifies a reduced variability in Td/IPW correlation as a result of averaging. All locations showed a very strong correlation between surface Td/IPW irrespective of their location within this study. The scattergrams in Fig. 4.5 illustrate the trend in the data as the averaged timescale increases.

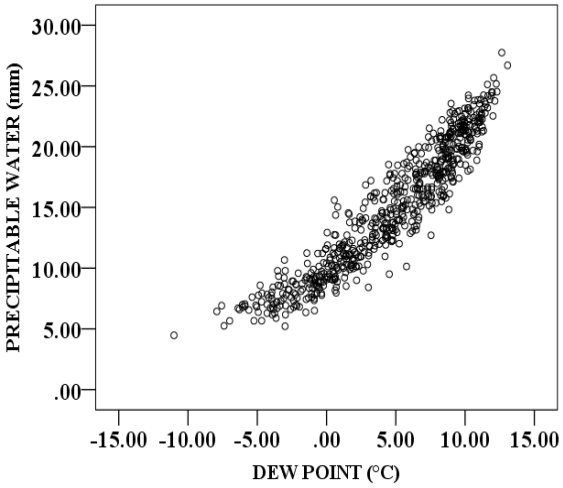
(A) Albuquerque, NM

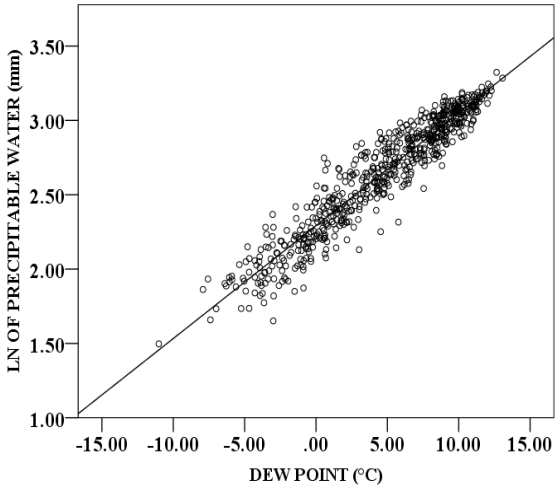


(B) El Paso, TX

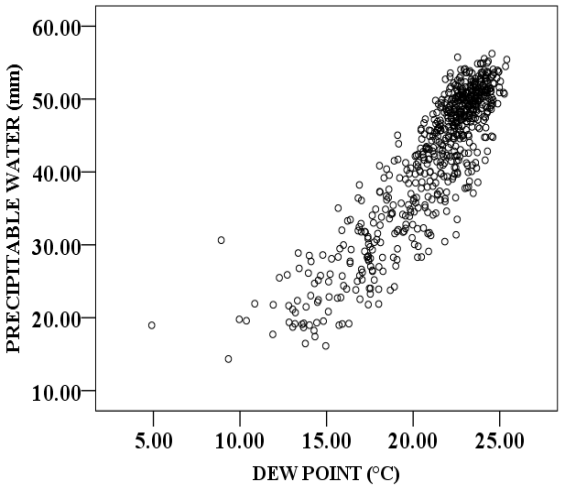


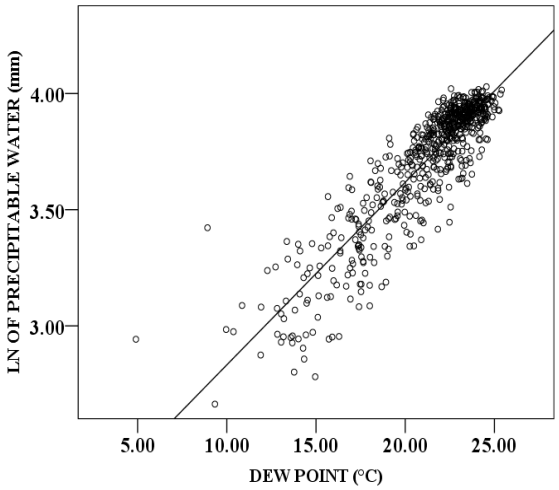
(C) Flagstaff, AZ



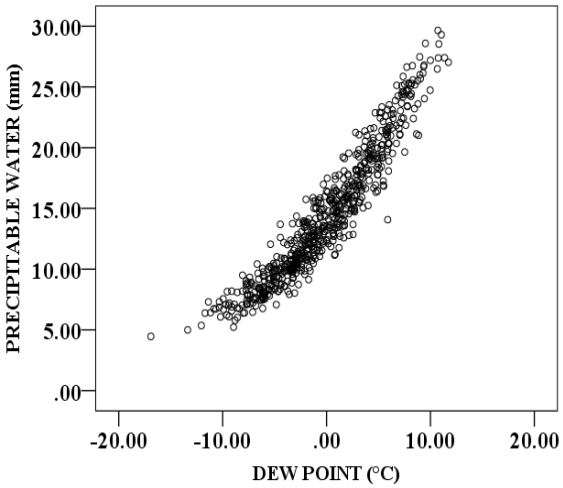


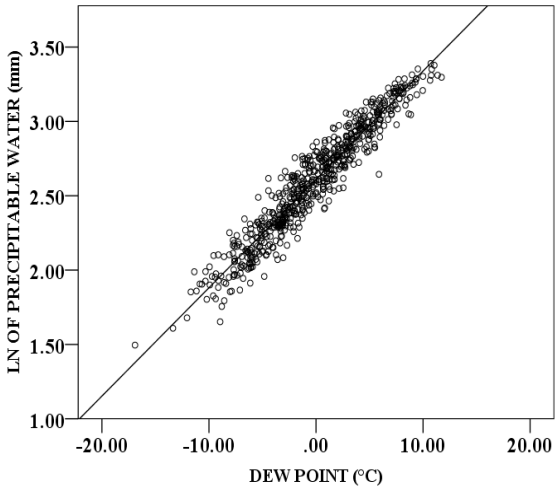
(D) Guaymas, MX



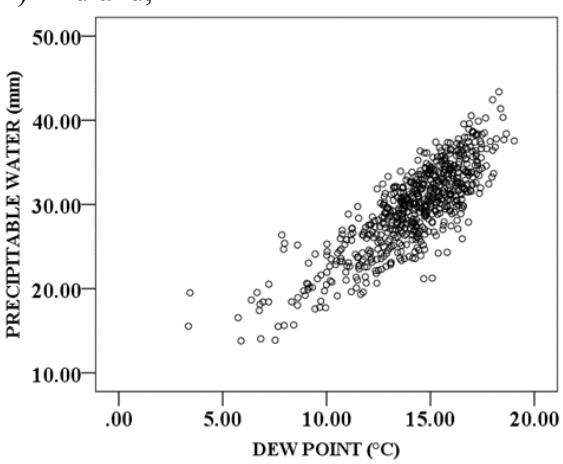


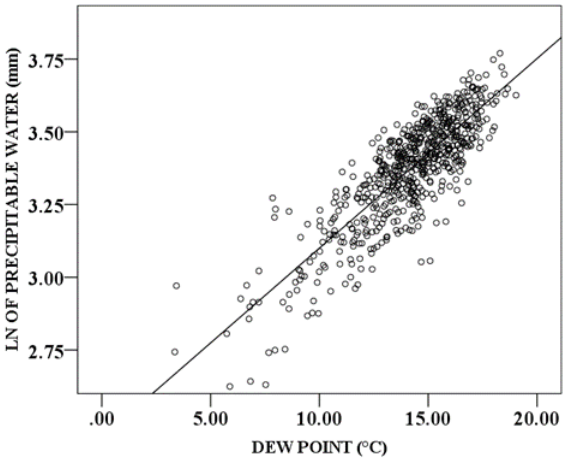
(E) Las Vegas, NV



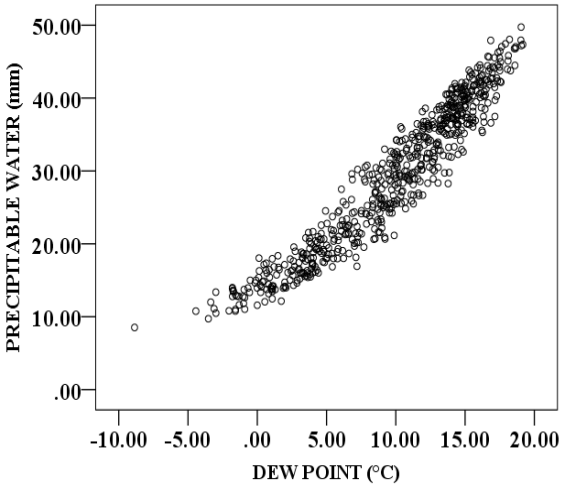


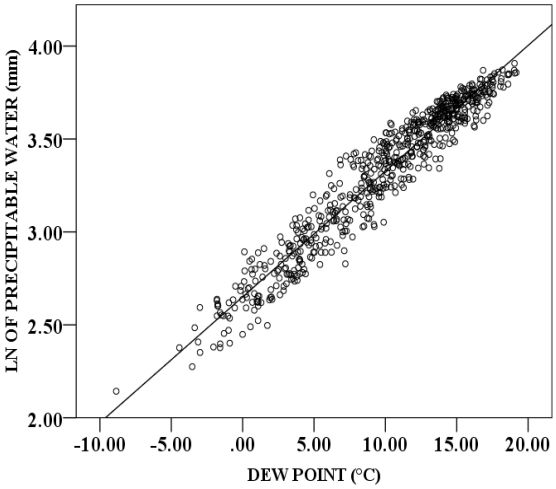
(F) Midland, TX



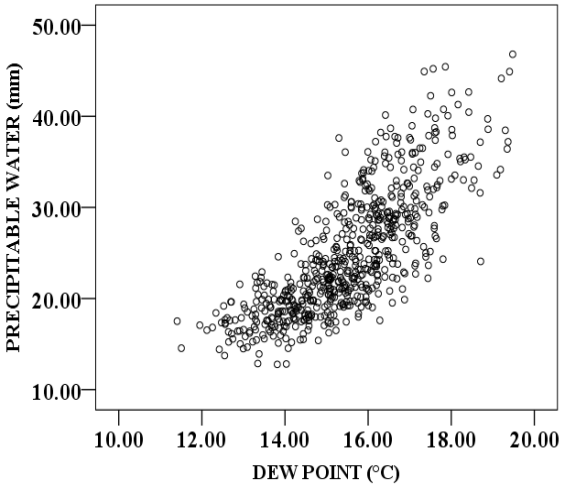


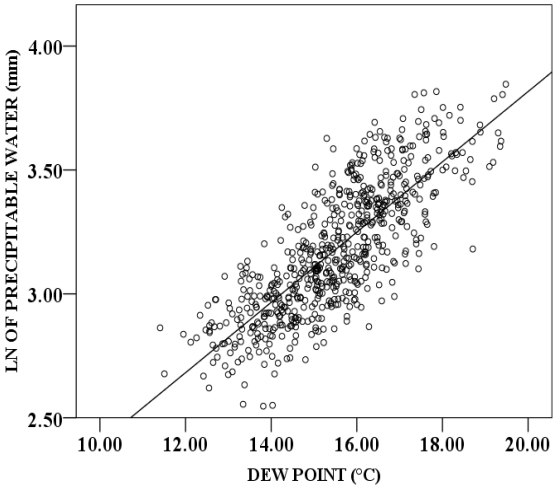
(G) Phoenix, AZ



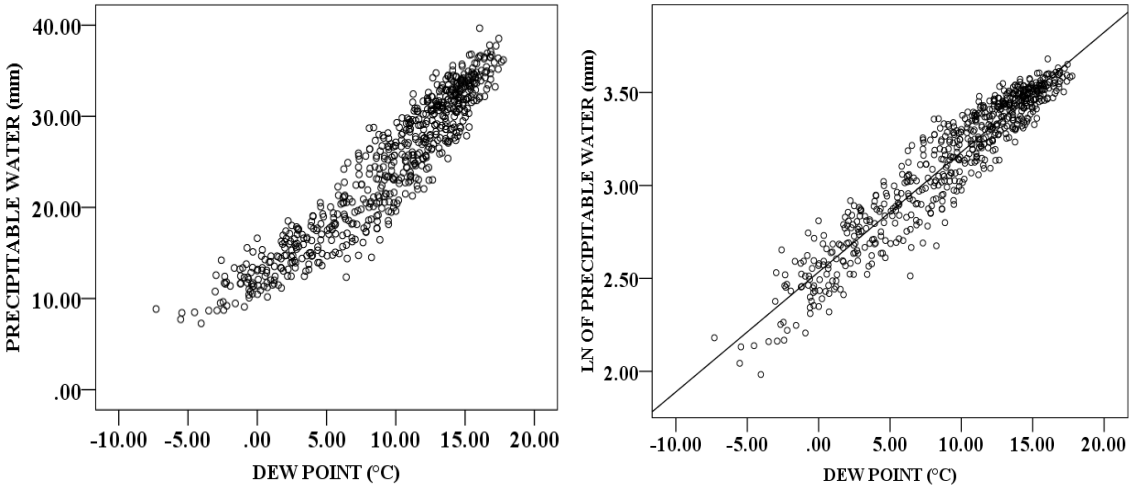


(H) San Diego, CA





(I) Tucson, AZ



(J) Yuma, AZ

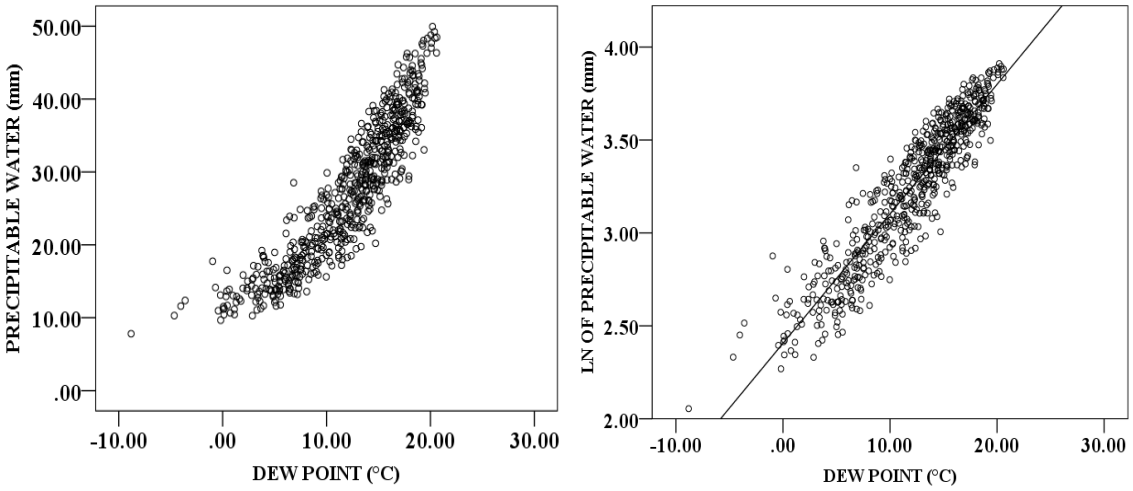


Fig. 4.5. Scatterplots for the weekly averaged values of IPW (in mm) /Td (in °C) and Ln(IPW) (in ln(mm)) / Td (°C) for the ten study cities. A) Albuquerque, NM, B) El Paso, TX, C) Flagstaff, AZ, D) Guaymas, MX, E) Las Vegas, NV, F) Midland, TX, G) Phoenix, AZ, H) I) Tucson, AZ, J) Yuma, AZ. Ln(IPW)/Td shows the associated best-fit trend line.

Fig. 4.5 shows variability in IPW output from a certain Td input in comparison to both Fig. 4.3 and Fig 4.4. This is represented by the more obvious scattering of plot points for each of the cities implying the IPW/Td relationship is decreasing in correlation for each location as the timescale increases. However, the scattergram appears more

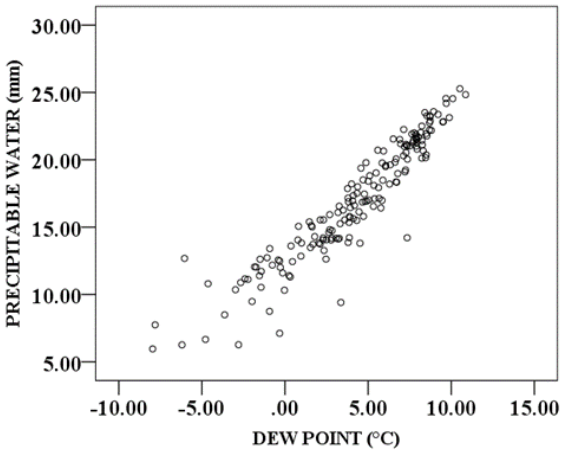
linear with each increase of timescale. This suggests that the atmosphere is tending towards a less logarithmic distribution of IPW values over an averaged time.

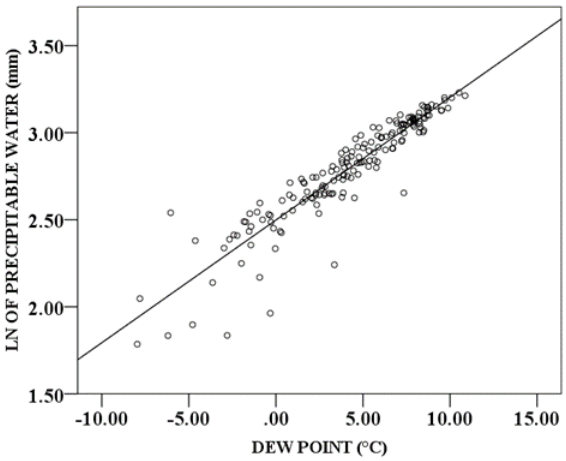
Table 4.8 illustrates the correlation values for the mean-monthly averaged timescale. The variability of IPW/Td drops, slightly, for this timescale, which interestingly is not the expected trend. Although, the range of values $0.779 \leq \rho \leq 0.955$ is still small and the correlation between IPW/Td is still considered strong, the overall trend is not consistent. It should be the case that the range continues to decrease despite the increasing timescale. This could be the byproduct of a shrinking N-Size. As the N-Size decreases, the smoothing effect from timescale averaging may be negated. So the relationship may still hold, but it is not represented well in this dataset.

Table 4.8 Spearman's (ρ) correlation, shared variance (%), and significant values, for monthly-mean measurements for the length-of-record (1979-2015) for each of the ten location (Albuquerque, NM, El Paso, TX Flagstaff, AZ, Guaymas, MX, Las Vegas, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ)

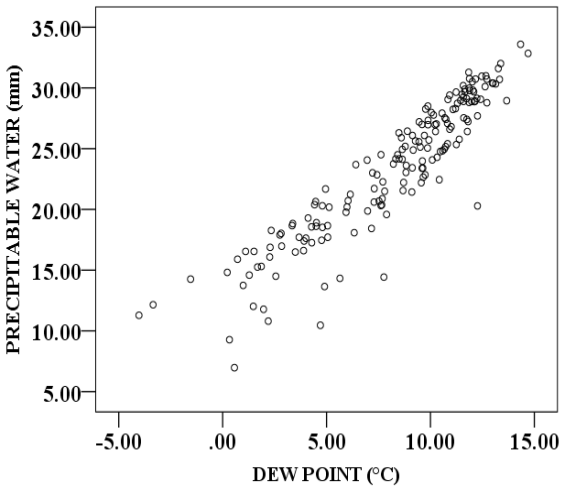
City	N-Size	(ρ)Td & IPW	ρ^2 (%)	Sig(2-tail)	(ρ) Td & Ln(IPW)	ρ^2 (%)	Sig(2-tail)
Albuquerque, NM	174	0.955	91	$p < 0.001$	0.959	92	$p < 0.001$
El Paso, TX	174	0.932	87	$p < 0.001$	0.932	87	$p < 0.001$
Flagstaff, AZ	174	0.947	90	$p < 0.001$	0.953	91	$p < 0.001$
Guaymas, MX	174	0.825	68	$p < 0.001$	0.826	68	$p < 0.001$
Las Vegas, NV	174	0.935	87	$p < 0.001$	0.935	87	$p < 0.001$
Midland, TX	174	0.787	62	$p < 0.001$	0.787	62	$p < 0.001$
Phoenix, AZ	174	0.958	92	$p < 0.001$	0.959	92	$p < 0.001$
San Diego, CA	174	0.779	61	$p < 0.001$	0.784	61	$p < 0.001$
Tucson, AZ	174	0.939	88	$p < 0.001$	0.943	90	$p < 0.001$
Yuma, AZ	174	0.913	83	$p < 0.001$	0.918	84	$p < 0.001$

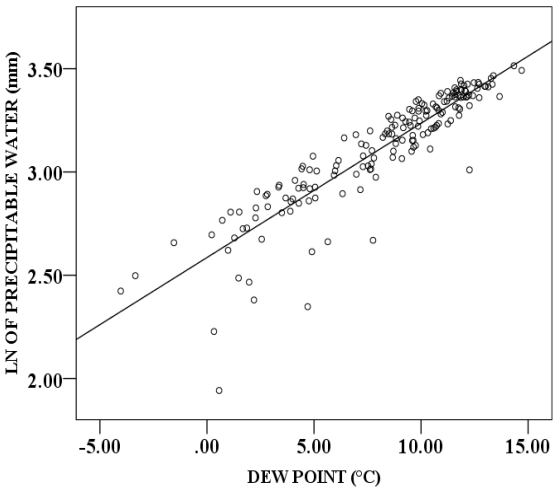
(A) Albuquerque, NM



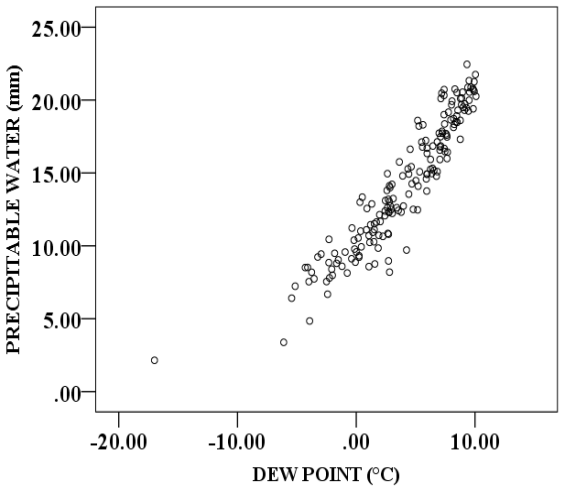


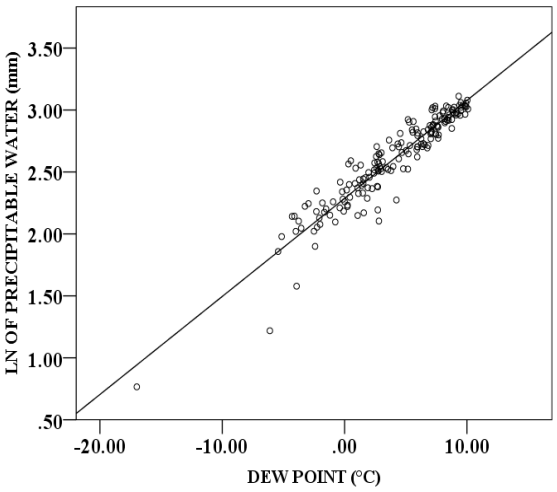
(B) El Paso, TX



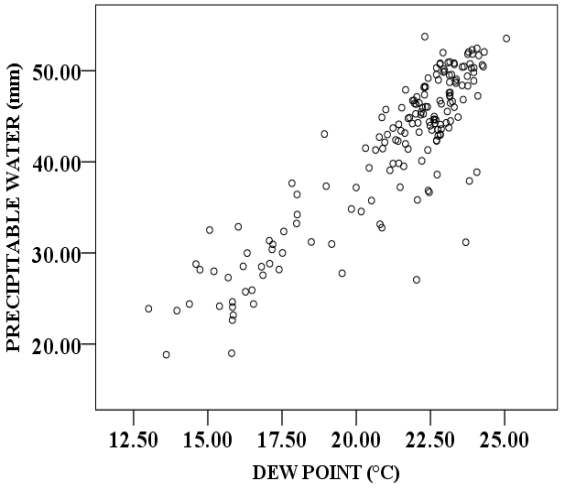


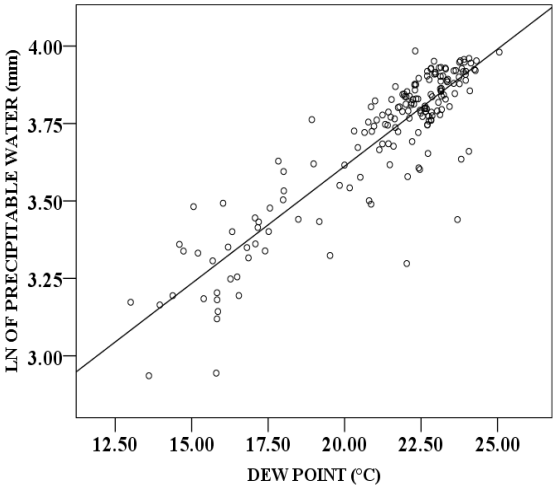
(C) Flagstaff, AZ



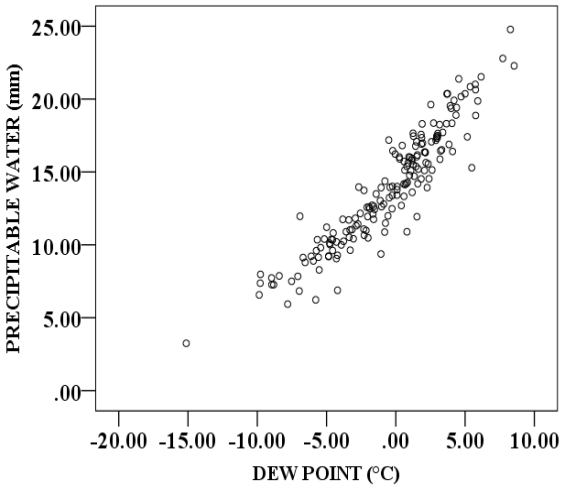


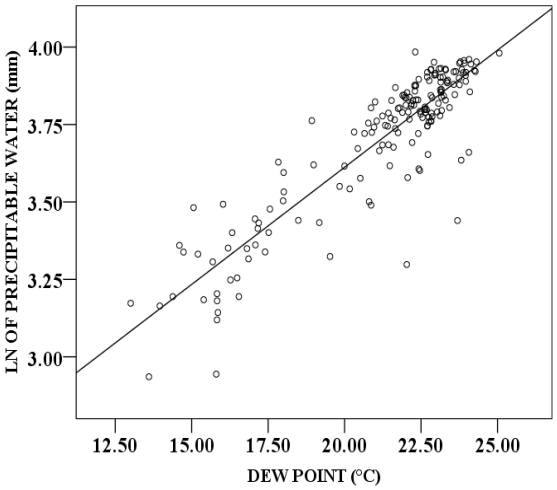
(D) Guaymas, MX



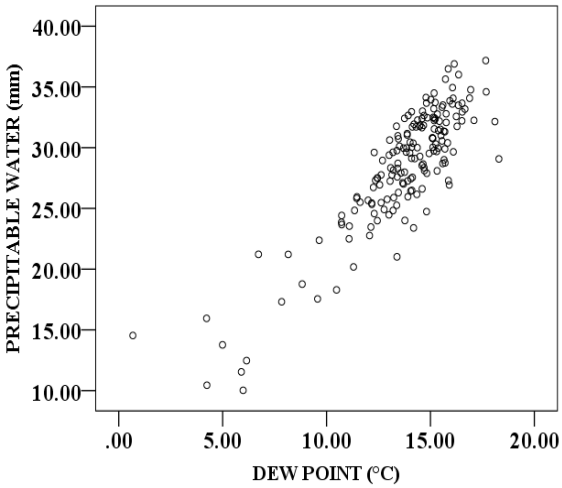


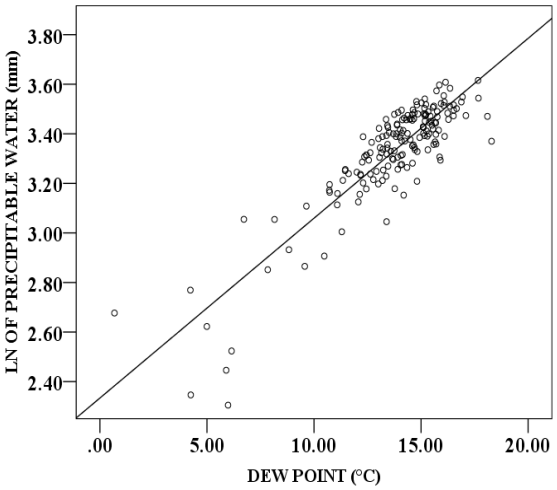
(E) Las Vegas, NV



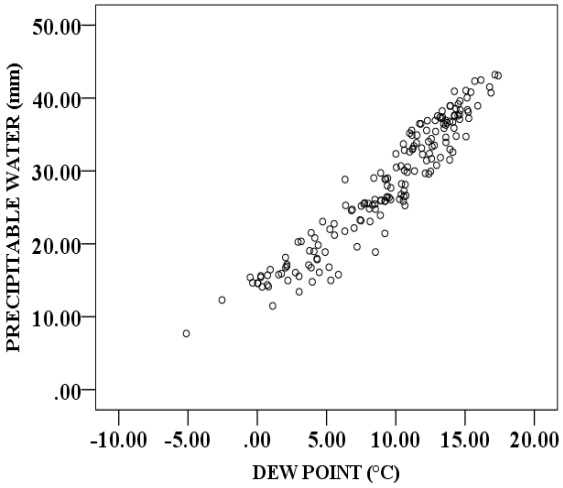


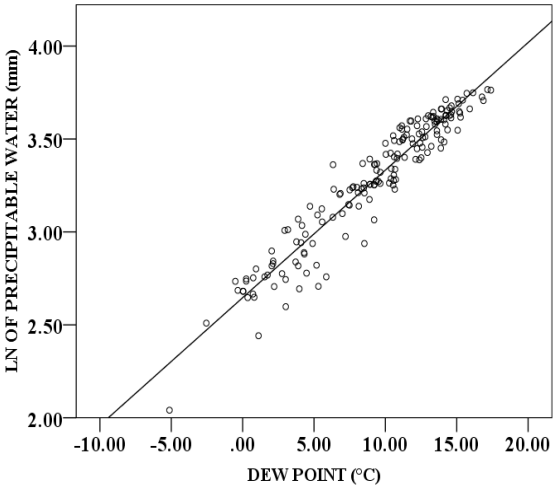
(F) Midland, TX



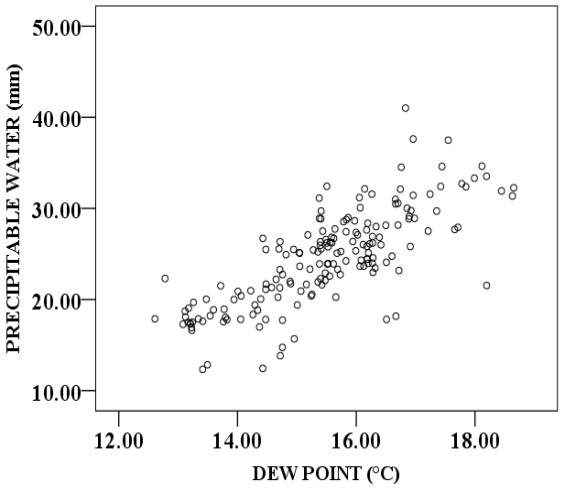


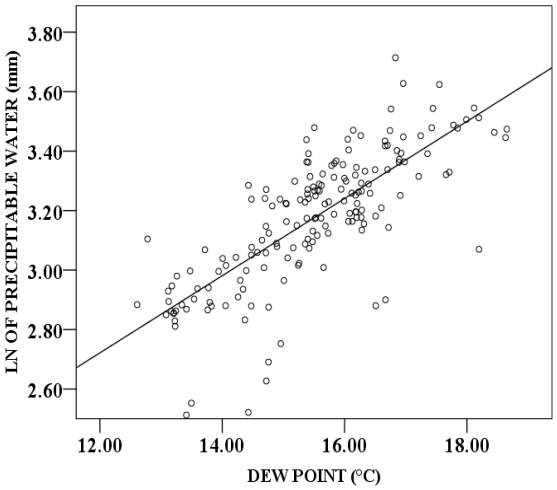
(G) Phoenix, AZ



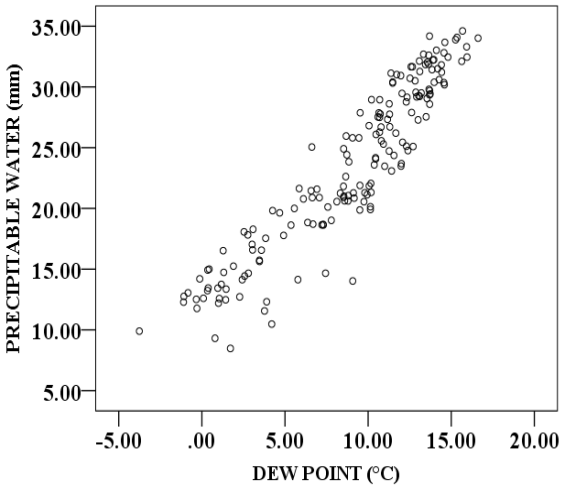


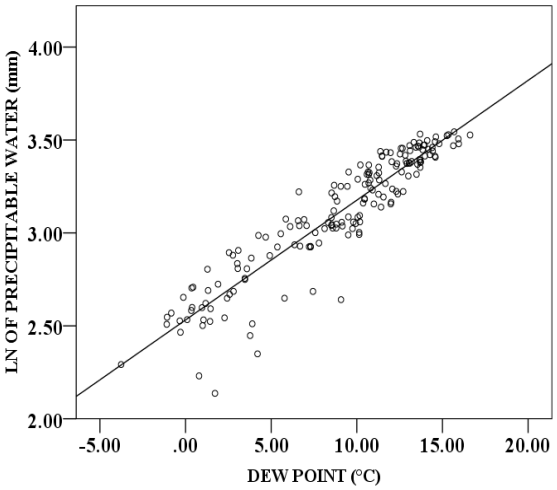
(H) San Diego, CA





(I) Tucson, AZ





(J) Yuma, AZ

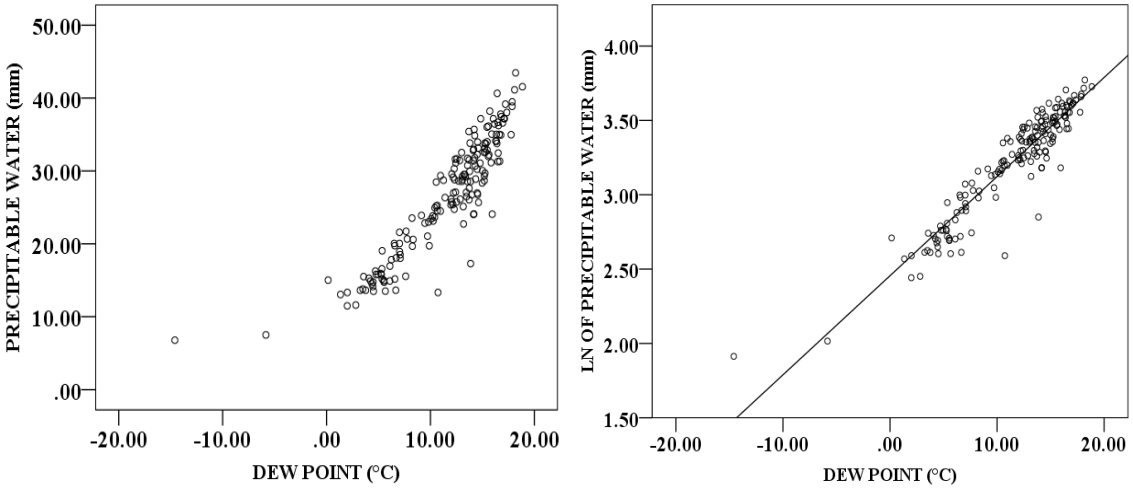


Fig. 4.6. Scatterplots for the monthly averaged values of IPW (in mm) /Td (in °C) and Ln(IPW) (in ln(mm)) / Td (°C) for the ten study cities. A) Albuquerque, NM, B) El Paso, TX, C) Flagstaff, AZ, D) Guaymas, MX, E) Las Vegas, NV, F) Midland, TX, G) Phoenix, AZ, H) San Diego, CA, I) Tucson, AZ, J) Yuma, AZ. Ln(IPW)/Td shows the associated best-fit trend line.

Even though the correlations are showing a relatively weaker relationship at the monthly timescale, the overall shape of the scattergrams continue to trend more linearly (Fig 4.6). Additionally, there does appear to be slightly more variability in IPW vs. Td values compared to the weekly average analysis, especially near coastal cities. Altogether, this corresponds with Reitan's (1963) correlations for IPW/Td for mean-monthly data, it may also link to why Reber and Swope's (1972) correlations for IPW/Td were much lower for their study in southern California.

The decline in the relationship between IPW/Td continues with yearly-mean averaged data. Table 4.9 is the correlated data analysis for this timescale. Because this timescale showed normality with the data, I used Pearson's correlation and provided Spearman's rank-order simply for comparison purposes. There is little variability between the two correlations.

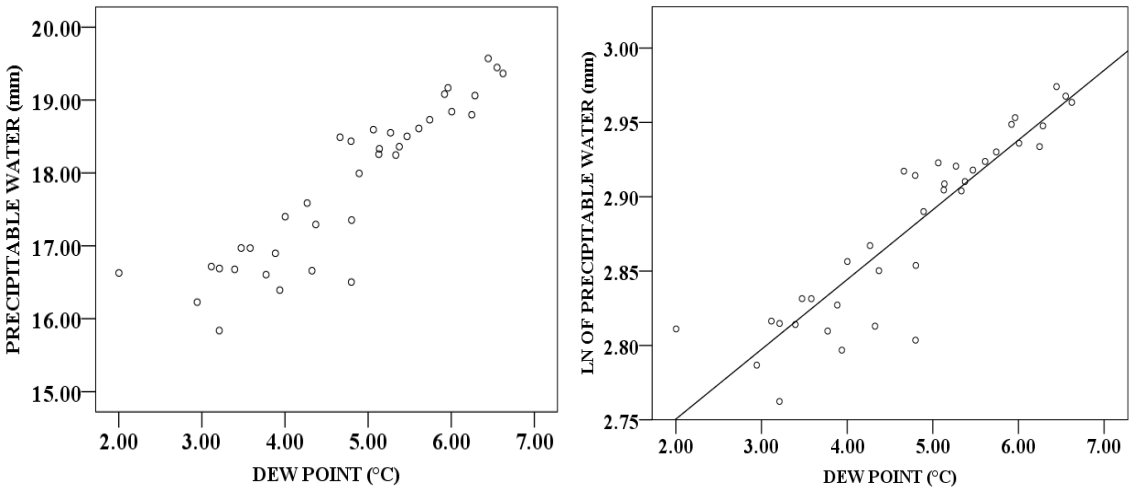
Table 4.9 Spearman's rho (ρ) correlation, Pearson's r correlation (r), shared variance (%), and significant values, for yearly mean measurements for the length-of-record (1979-2015) for each of the ten location (Albuquerque, NM, El Paso, TX Flagstaff, AZ, Guaymas, MX, Las Vegas, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ)

City	N	(ρ) Td & IPW	ρ^2 (%)	Sig (2-tail)	(r) Td & IPW	r^2 (%)	(ρ) Td & Ln(IPW)	ρ^2 (%)	(r) Td & Ln(IPW)	r^2 (%)	Sig (2-tail)
ABQ	37	0.908	82	p < 0.001	0.909	83	0.957	92	0.938	88	p < 0.001
ELP	37	0.927	86	p < 0.001	0.922	85	0.906	82	0.930	86	p < 0.001
FLG	37	0.756	57	p < 0.001	0.762	58	0.849	72	0.841	71	p < 0.001
GUY	37	0.831	69	p < 0.001	0.832	69	0.792	63	0.814	66	p < 0.001
LAS	37	0.837	70	p < 0.001	0.899	81	0.922	85	0.944	89	p < 0.001
MID	37	0.722	52	p < 0.001	0.842	71	0.732	54	0.861	74	p < 0.001
PHX	37	0.854	73	p < 0.001	0.902	81	0.894	80	0.923	87	p < 0.001
SAN	37	0.624	39	p < 0.001	0.698	48	0.671	45	0.706	50	p < 0.001
TUC	37	0.851	73	p < 0.001	0.873	76	0.884	78	0.897	80	p < 0.001
YUM	37	0.819	67	p < 0.001	0.825	68	0.842	71	0.837	70	p < 0.001

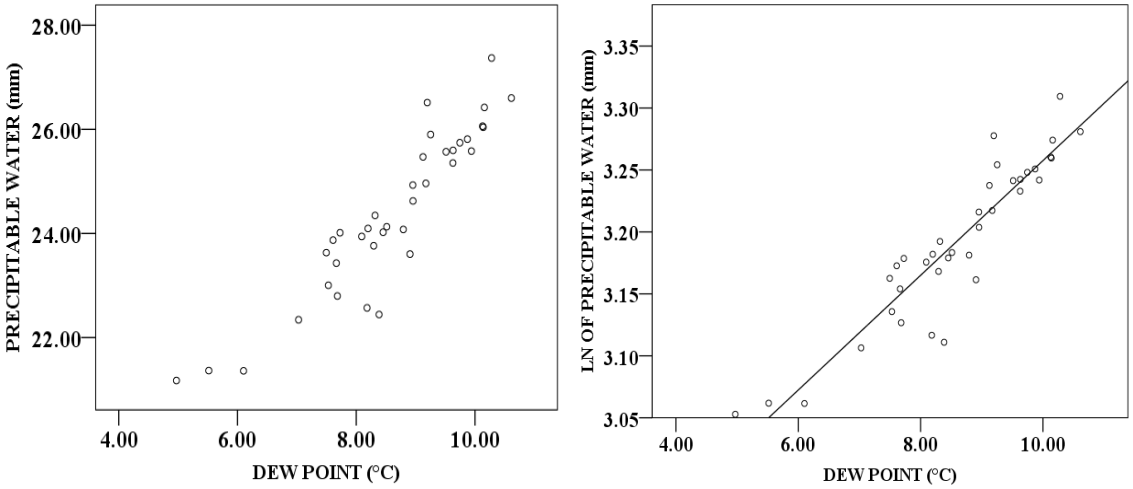
The IPW/Td weakened for most cities as the time average went from monthly to yearly averaged data. Albuquerque, NM, El Paso, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, and Yuma, AZ all saw a slight drop off in the correlation. These cities with a slight drop off may be a byproduct of a decreasing sample size. Flagstaff, AZ showed a relatively large drop in the correlation from 0.947(monthly) to 0.762(yearly). This suggests that Flagstaff, AZ is more sensitive to seasonally influenced atmospheric teleconnections during the NAM as opposed to the other cities in the study region. Guaymas, MX and Midland, TX displayed an increase in correlations from monthly to yearly. Guaymas, MX went from 0.825(monthly) to 0.832(yearly), while Midland, TX went from 0.787(monthly) to 0.842(yearly). This fits with the overall theme of this study

in that the IPW/Td relationship for cities outside of the NAM's domain act atypically in comparison to those inside the NAM's domain. Overall, the range of values were $0.698 \leq r \leq 0.957$ showing that the correlations were for all cities was strong.

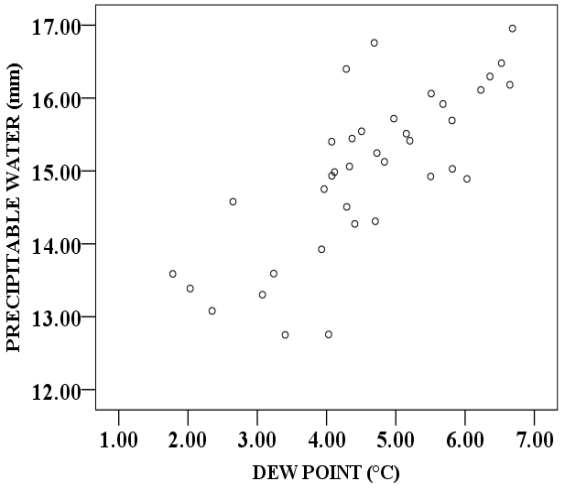
(A) Albuquerque, NM

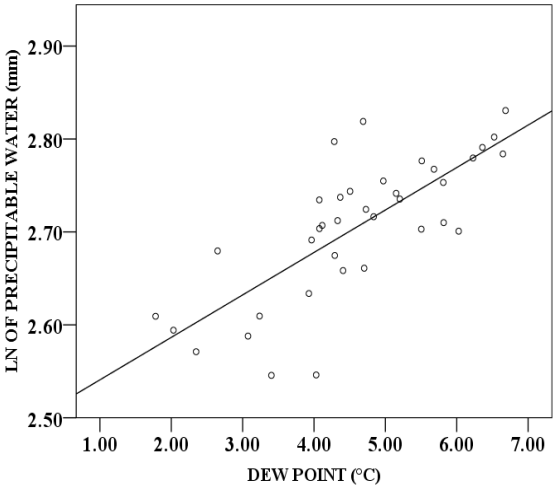


(B) El Paso, TX

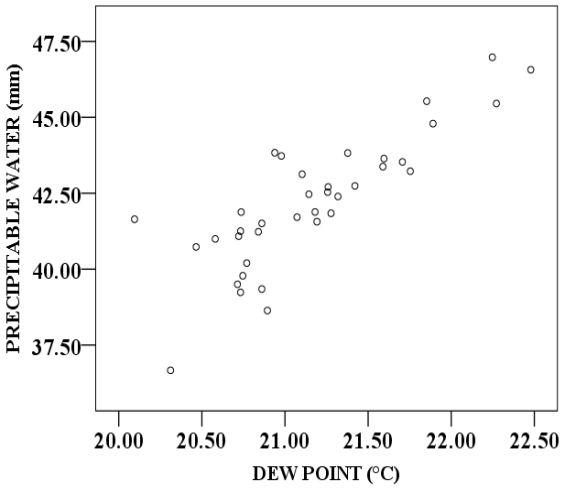


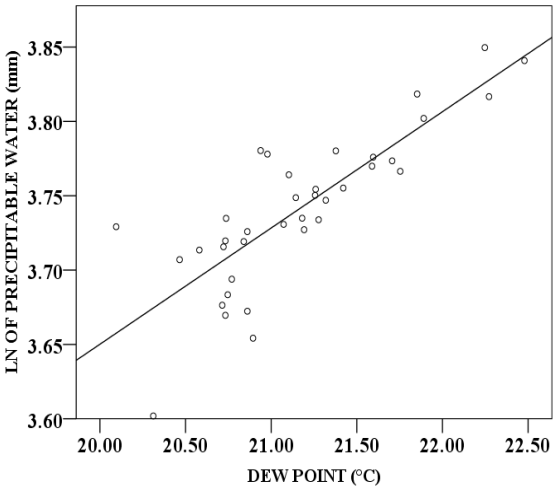
(C) Flagstaff, AZ



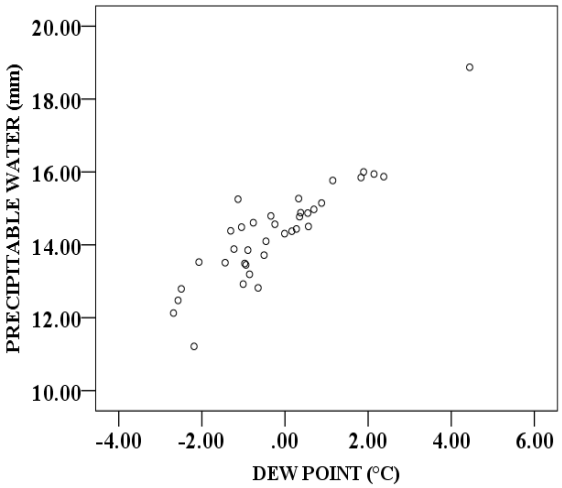


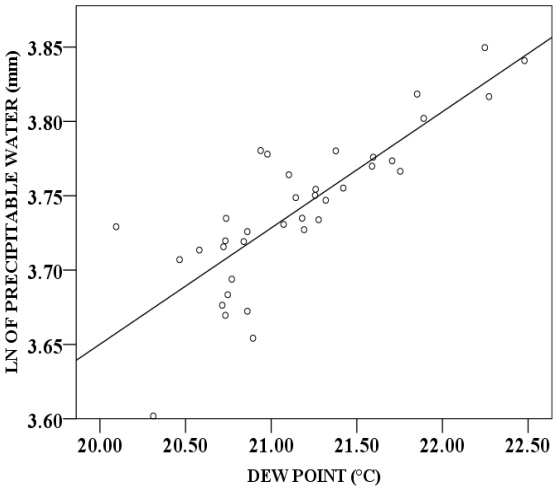
(D) Guaymas, MX



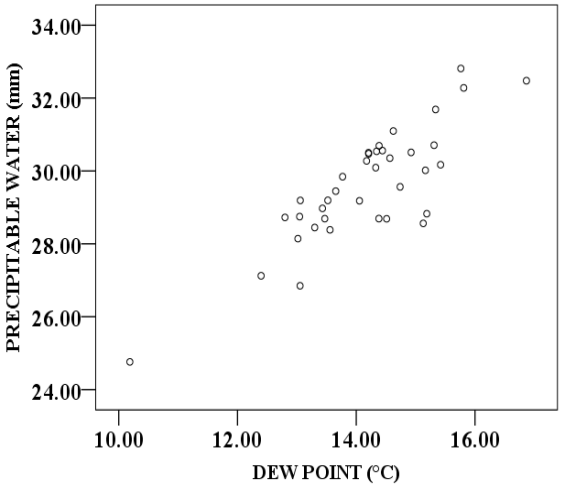


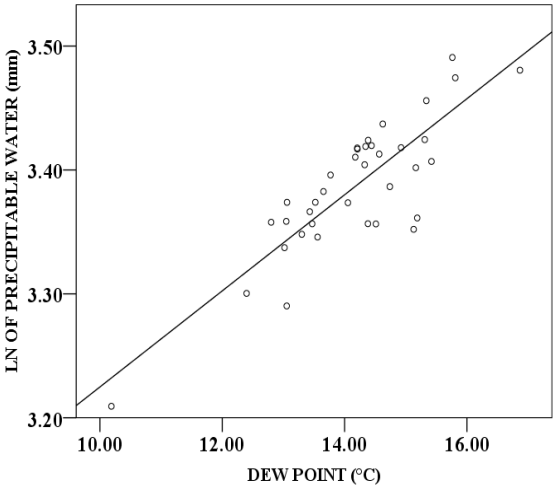
(E) Las Vegas, NV



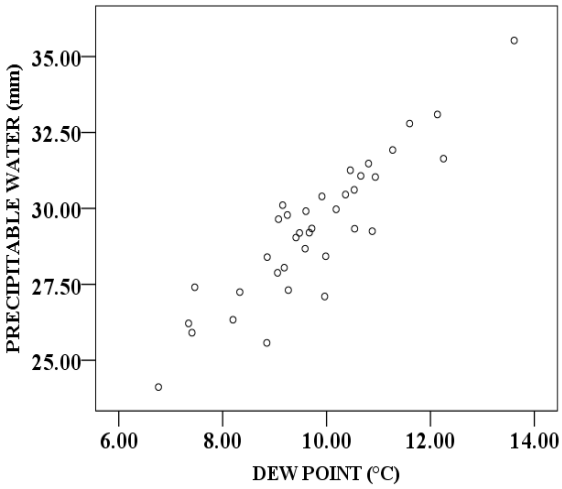


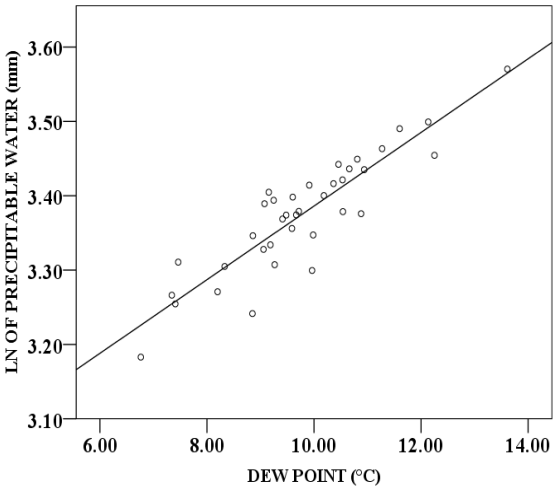
(F) Midland, TX



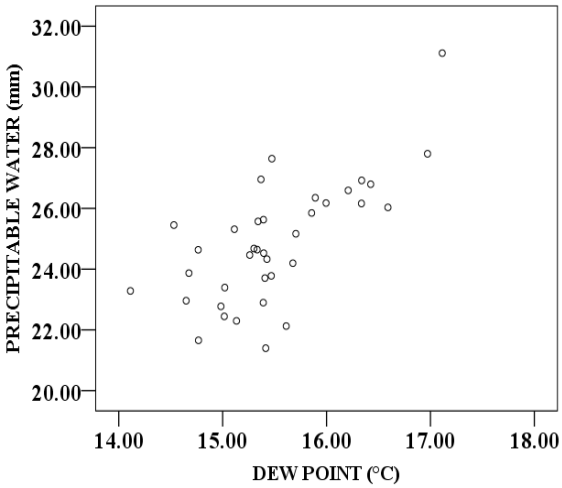


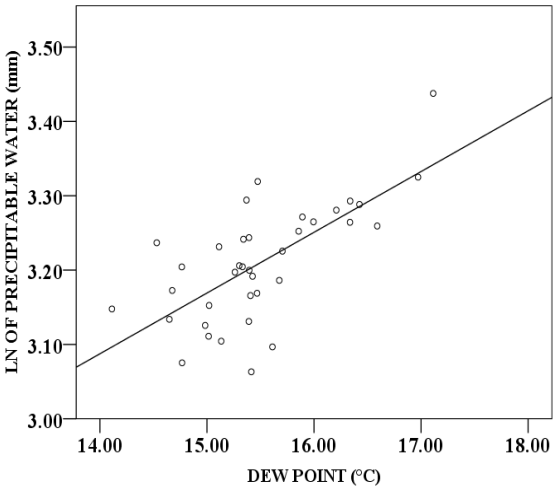
(G) Phoenix, AZ



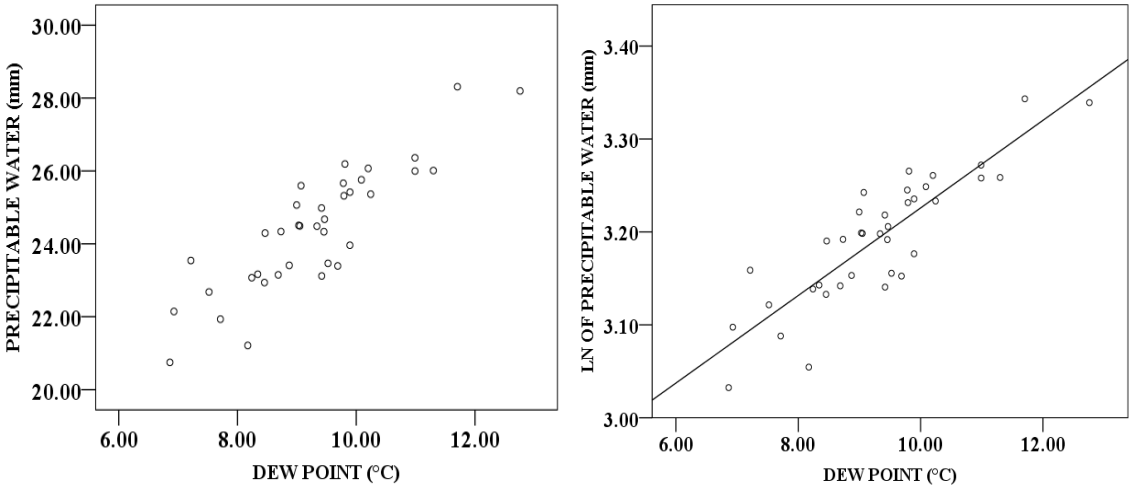


(H) San Diego, CA





(I) Tucson, AZ



(J) Yuma, AZ

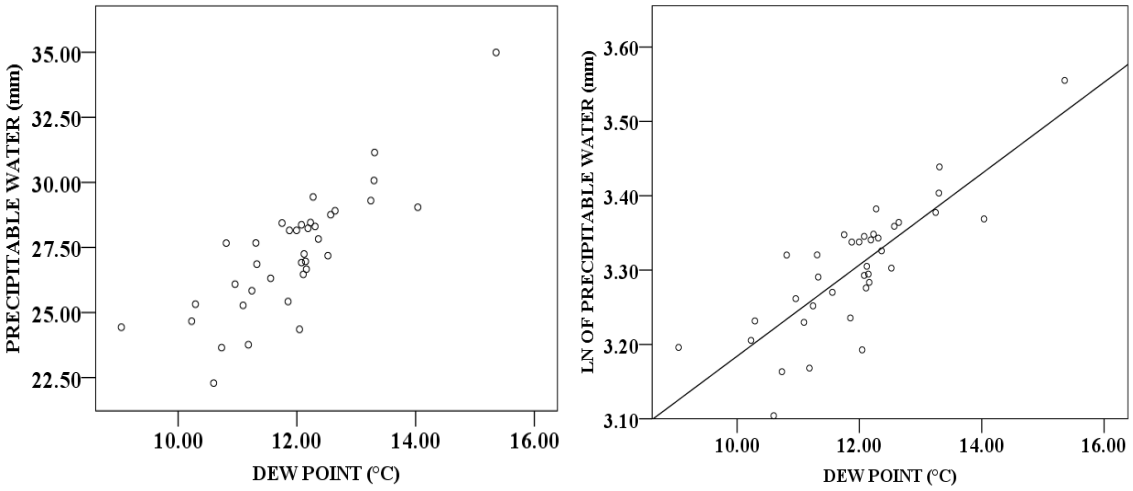


Fig. 4.7. Scatterplots for the yearly averaged values of IPW (in mm) /Td (in °C) and Ln(IPW) (in Ln(mm)) / Td (°C) for the ten study cities. A) Albuquerque, NM, B) El Paso, TX, C) Flagstaff, AZ, D) Guaymas, MX, E) Las Vegas, NV, F) Midland, TX, G) Phoenix, AZ, H) San Diego, CA, I) Tucson, AZ, J) Yuma, AZ. Ln(IPW)/Td shows the associated best-fit trend line.

The scattergrams for annual conditions (Fig. 4.7) continue to show that the data are becoming more linearly as opposed to the earliest timescale represented. In addition, the variability appears to have increased in a similar fashion to that of the mean monthly timescale. This could be the result of yearly climatological influences that may affect seasonal rainfall (e.g., El Nino, La Nina, and drought). If a season is greatly affect by one

of these atmospheric teleconnections, then the results will not be consistent season by season thus modifying the correlation coefficient.

This section has examined the relationship of IPW/Td both spatially and temporally. The relationship between IPW/Td is apparent, however, there is a caveat to this. The relationship is highly variable. Variability with this relationship was observed as the time-average values were changed. A potential upper-limit IPW exists for a given surface dew point value, however, the likelihood that this IPW is realized is depended on additional conditions (e.g., uplift). That said, this variability is variable spatially. Cities inside the NAM's domain vary different than those outside of the NAM's domain both respect to time and the potential output of IPW based on a given Td. One question that remains unaddressed is whether or not these results are consistent throughout the time domain of the analysis (1979 - 2015). The next section examines the variability in these results as function of the specific years of observations.

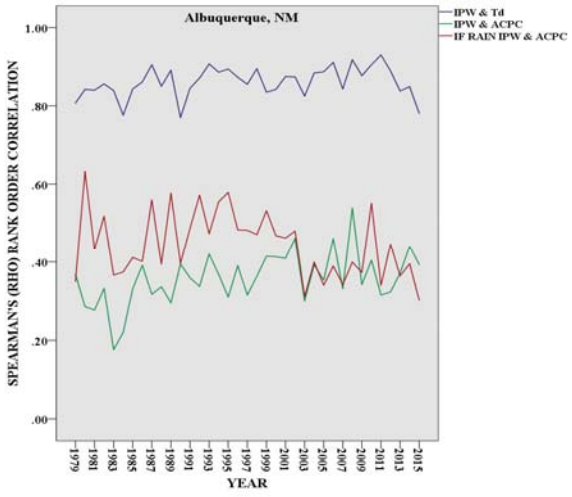
4.5 Correlation vs. various time scales

In this section, I produce three different analyses that examine the temporal progression of the relationship between IPW/Td. This is to explore any additional differences in the IPW/Td relationship that may exist between cities on the border of the study's region and those within. The first analysis shows the evolution of the IPW/Td correlation in comparison to the correlation of IPW and accumulated precipitation (ACPC) in each of the ten locations. The second analysis is a plot of the Spearman's rank-order (ρ) values between weekly averaged IPW and Td. The reason weekly averaged values where chosen for this is that this timescale yielded the greatest

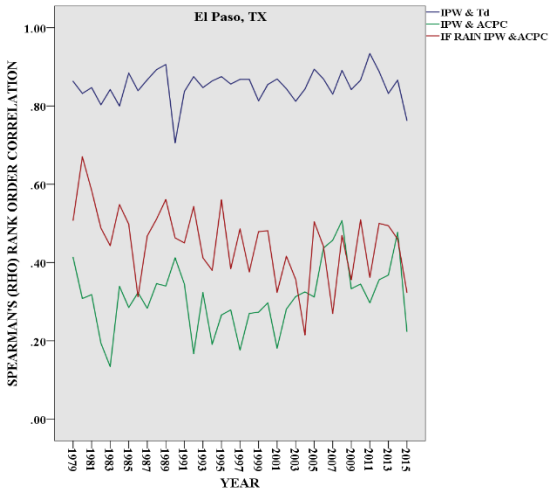
correlations (Table 4.7) when match with all other timescales. The last analysis is a figure showing a three-day averaged IPW/Td correlation for all ten cities for the entire length-of-record (1979-2015).

Examining the IPW-ACPC relationship is important over the temporal domain of the dataset because, as outlined in Chapter 2, the surface dew point-IPW relationship was established because a surface threshold was thought to translate into 25.4mm of IPW, which was thought to be enough to initiate thunderstorm activity (Reitan 1957, 1963). Because the NARR data were aggregated into three-hour measurements, there were numerous zero precipitation values in the dataset, which could theoretically inhibit the relationship. Therefore, I eliminated all samples containing a zero value for ACPC and ran the Spearman's ρ correlation against the remaining associated IPW values.

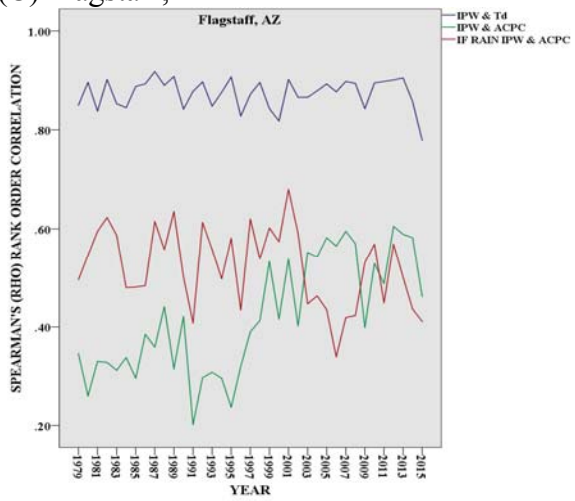
(A) Albuquerque, NM



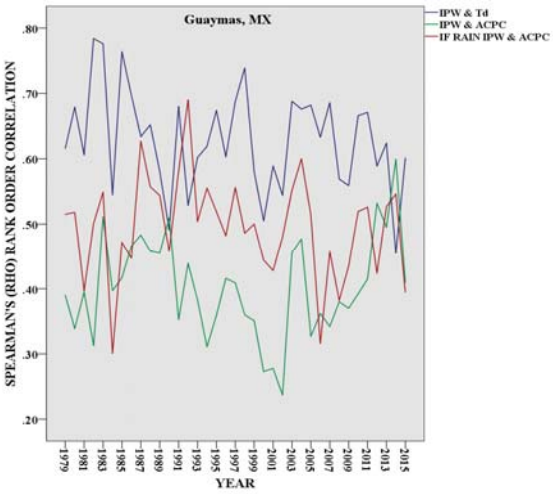
(B) El Paso, TX



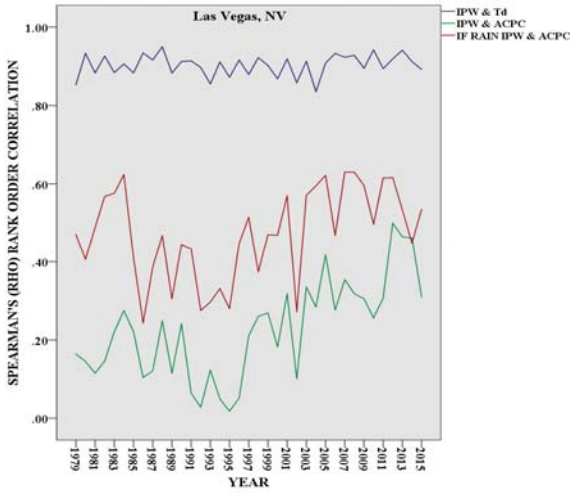
(C) Flagstaff, AZ



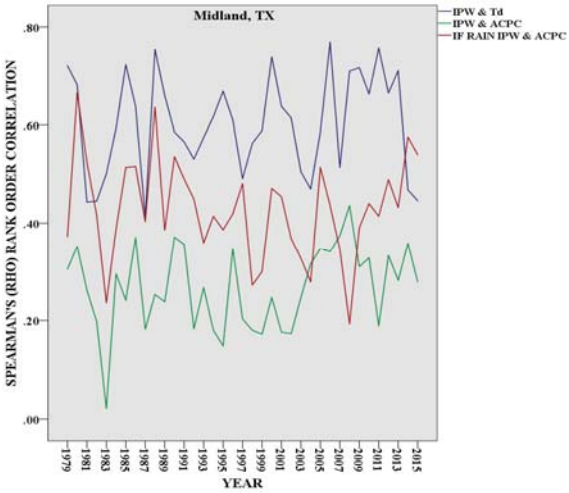
(D) Guaymas, MX



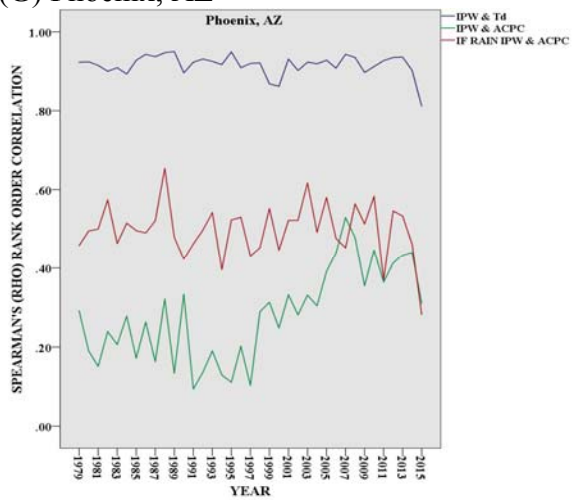
(E) Las Vegas, NV



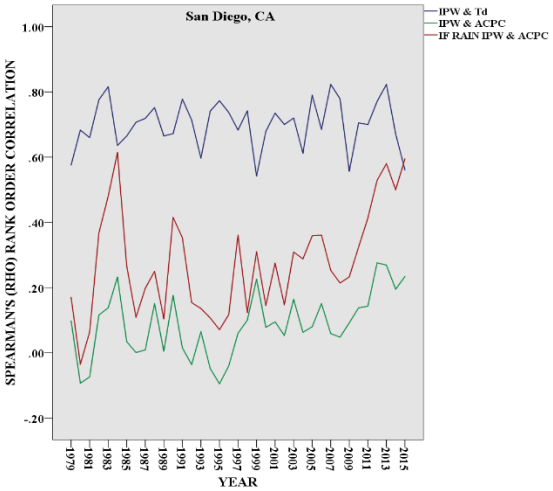
(F) Midland, TX



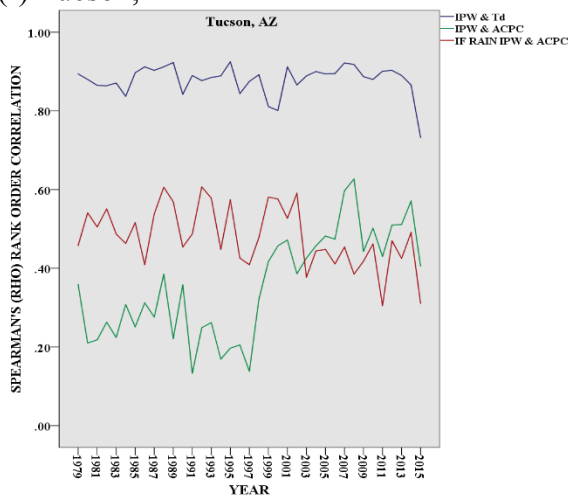
(G) Phoenix, AZ



(H) San Diego, CA



(I) Tucson, AZ



(J) Yuma, AZ

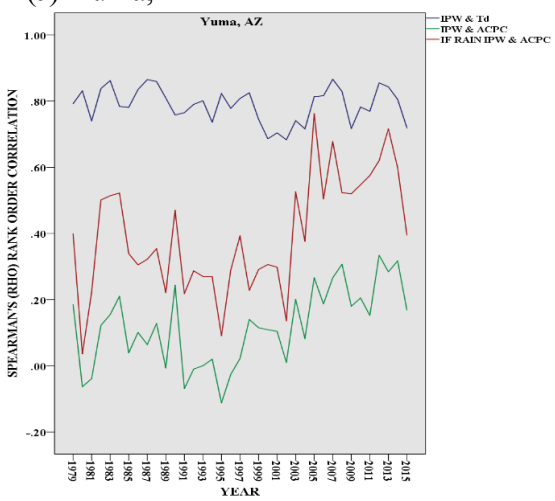


Fig. 4.8. Graphs of the yearly averaged IPW/Td correlation (ρ) values (blue line), the yearly averaged (integrated precipiaable water) IPW/ACPC (accumulation precipitation) correlation (ρ) values (green line), and the yearly averaged IPW/ACPC correlation (ρ) values with all zero ACPC values removed from the dataset (i.e., there was measurable precipitation) (red line) for all ten cities. A) Albuquerque, NM, B) El Paso, TX, C) Flagstaff, AZ, D) Guaymas, MX, E) Las Vegas, NV, F) Midland, TX, G) Phoenix, AZ, H) San Diego, CA, I) Tucson, AZ, J) Yuma, AZ). These data were averaged for the entire length-of-record (1979–2015).

First, the variability of the correlations between IPW/Td for cities traditionally associated with the monsoon versus those on the peripheries of this study is markedly different (Fig. 4.8). The traditional monsoon cities (Albuquerque, NM, El Paso, TX, Las Vegas, NV, Flagstaff, AZ, Phoenix, AZ, Tucson, AZ, and Yuma, AZ) all have consistent correlations between IPW/Td for any given year when seasonally averaged. Las Vegas, NV shows the most consistent IPW/Td correlation for every year of the study with $\rho \cong 0.90$. Conversely, Guaymas, MX, Midland, TX, and San Diego, CA all exhibited pronounced seasonal variability with respect to their correlation coefficients. For example, Guaymas, MX, showed that some years the IPW/Td association was at $\rho \cong 0.80$ while for some years $\rho < 0.50$. This was similar to Midland, TX. However, of all the

"fringe" locations, San Diego, CA was slightly more modulated than the other cities. This was likely a byproduct of the consistent low-level moisture from the ocean marine layer despite the lack of forcing. The other seven cities have a reliable influx of moisture with predictable increases in thermodynamic, topographic, and the intermittent passing of inverted dynamic systems that all increase instability. The outlier cities are not as predictable with respect to seasonal moisture nor with respect to uplift.

Precipitable water (IPW) was not well-correlated with ACPC despite IPW being considered the greatest single variable that corresponds to ACPC (Skindlov 2007). IPW and ACPC in all cities rarely exceeded 0.50 for any of the years. IPW-Precipitation correlations reached above 0.50 for Albuquerque, NM in 2008, El Paso, TX in 2009, Flagstaff, AZ 12 times from 1999-2015, Guaymas, MX from 2012-2015, Phoenix, AZ in 2007, and eight times during the years from 2001-2015 in Tucson, AZ. However, those correlations never reached 0.50 for Las Vegas, NV, Midland, TX, San Diego, CA, and Yuma, AZ for the entire 37 year study. This may produce a lack of confidence in IPW as a sound singular forecasting tool for precipitation. However, removal of all the ACPC zeros from the dataset may enhance the usability of this relationship.

Assuming that days with zero ACPC could create a lower correlation between IPW/ACPC, I eliminated those days from the sampling and re-ran the correlation with only days in which ACPC was non-zero. I then compared the values against the associated IPW for that time measurement. This, for the most part, increased the correlation for IPW/ACPC. Nearly all the cities showed a closely related range of correlations between the two measurements. However, a correlation above the

benchmark $\rho > 0.70$ was never achieved in any of the cities for any time period. Furthermore, there was a lot of variability with this correlation. In summation, IPW appears not to be a strong indicator of precipitation. These results suggest the need for additional atmospheric ingredients (i.e., thermodynamics/kinematics) to better correlate to precipitation events. However, the correlation between IPW/Td may yield a means for determining monsoon onset not previously considered.

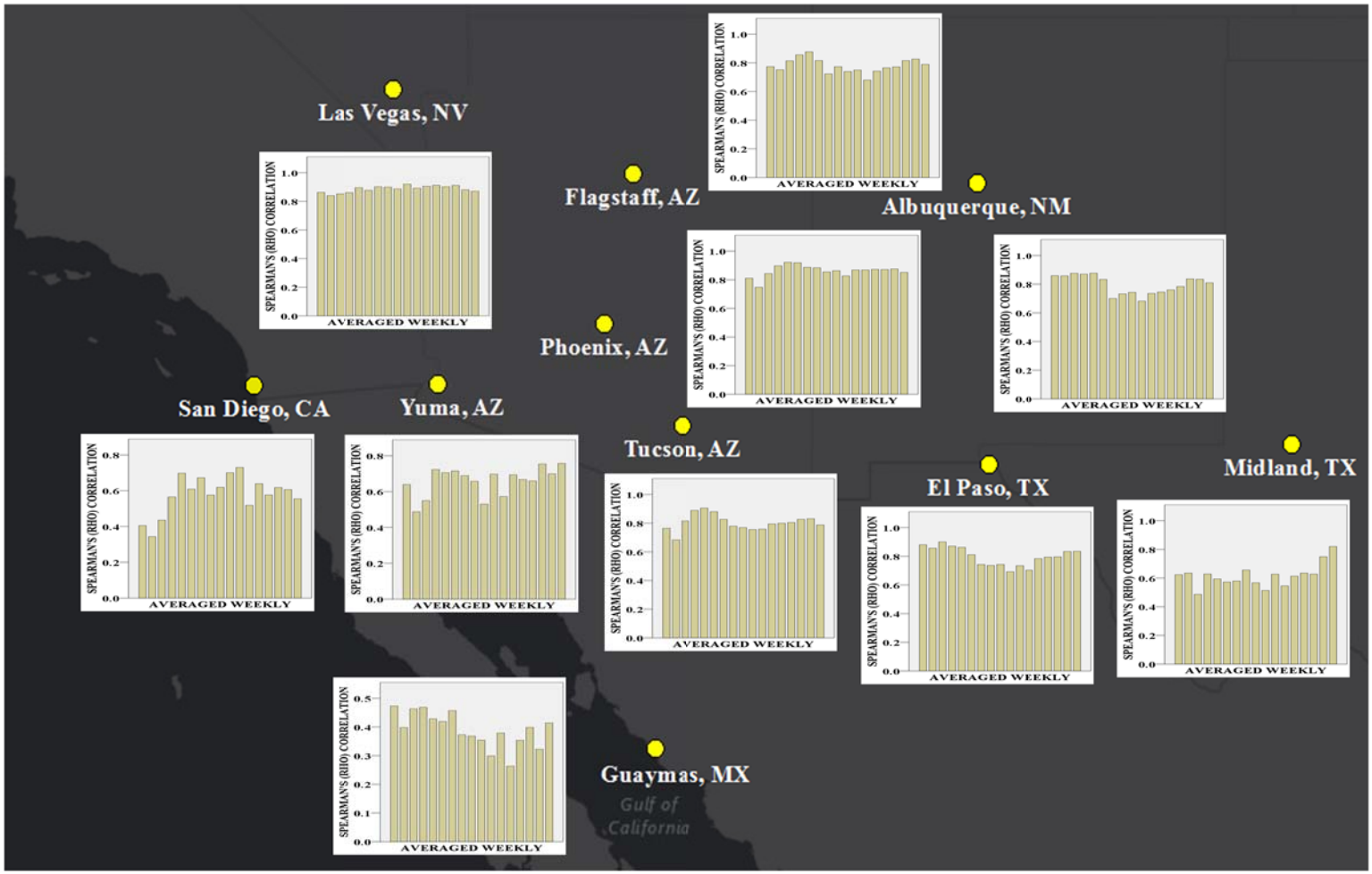


Fig. 4.9. Spatial distribution of the weekly averaged IPW/Td values vs. the associated Spearman's correlation (ρ) values for Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ. These data were averaged for the entire length-of-record (1979–2015).

With the exception of Yuma, AZ, all of the traditional monsoon cities (Albuquerque NM,l Paso TX, Flagstaff, AZ Las Vegas, NV Phoenix, AZ and Tucson, AZ) display a visually apparent spike in the correlation between IPW around 4th/5th week of the monsoon seasonal period (~25 June- 8 July). This is near the time of the average onset date for the meteorological monsoon in Phoenix, AZ (7 July) (Ellis et al. 2004). This peak is followed by a reduction of the correlation between IPW and Td, which would roughly be in mid-July followed by another peak occurring roughly at week 13-14 of the study (~27 August 27 – 9 September). This relationship could be illustrating the monsoon burst/break phenomenon outlined by Adams and Comrie (1997). They discussed that periods of increased precipitation can be followed by decreased times of monsoonal activity. This bimodal relationship could be a proxy for the averaged onset for NAM's burst and breaks.

In contrast, San Diego, CA, Yuma, AZ, Midland, TX, and Guaymas, MX illustrate a more parabolic or "U-shape" to their respective histograms of correlation over the course of the monsoon season. This indicates that there is not a boreal summertime (JJAS) IPW/Td correlation maximum, which may be further evidence for the aforementioned bimodality to be a monsoon signature. The reason for the reductions in correlation is likely the result of greater low-level atmospheric moisture with co-absence of vertical lifting from dynamic and topographic sources. Because the influx of monsoon moisture into the traditional monsoon region is advected from its three primary sources long distance, it is likely to encounter vertical mixing inbound. This helps to further facilitate the IPW/Td relationship and may suggest a possibility for the uniqueness seen

in the histograms for the monsoon region. This relationship can be examined further comparing the evolution of correlations coefficients between IPW/Td for all the regions in the monsoon.

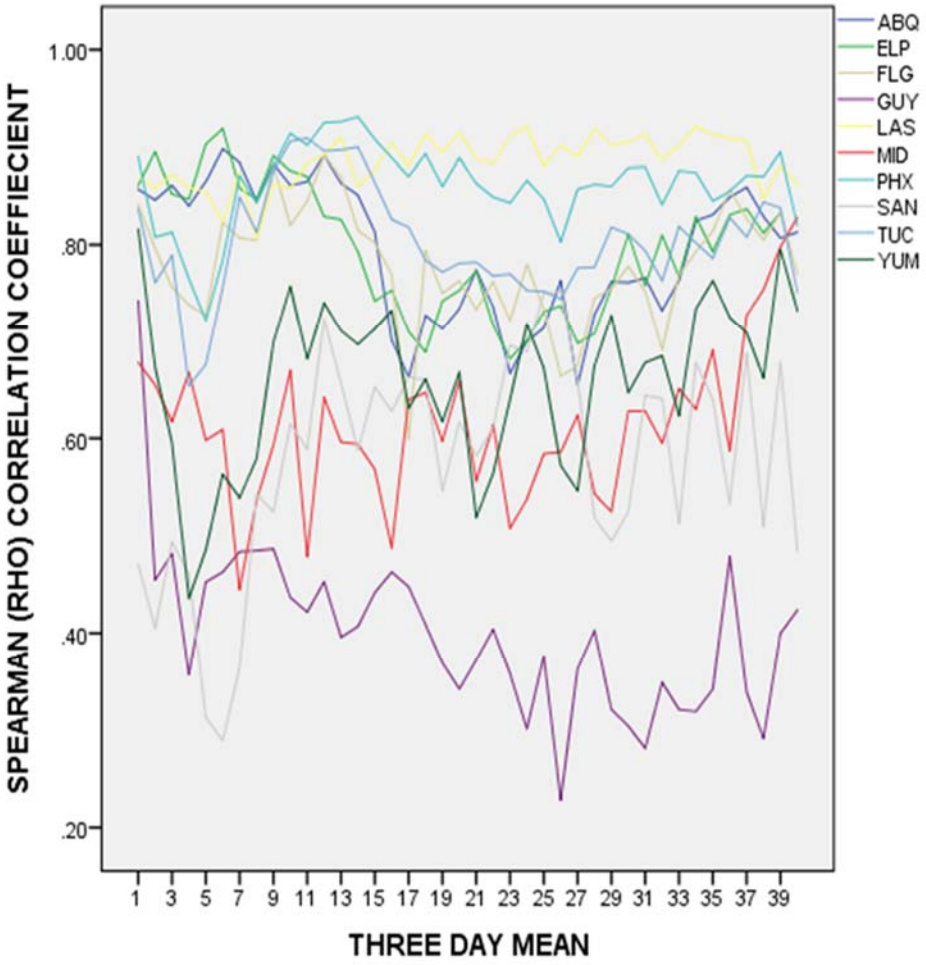


Fig. 4.10. Three-day averaged IPW/Td values vs. the associated Spearman's correlation (ρ) values for Albuquerque, NM (blue line), El Paso, TX (green line), Flagstaff, AZ, (tan line) Guaymas, MX, (purple line) Las Vegas, NV, (yellow line) Midland, TX, (red line) Phoenix, AZ, (aqua blue line) San Diego, CA, (gray line) Tucson, AZ, (light blue) Yuma, AZ (dark green line). These data were averaged for the entire length-of-record (1979–2015).

The three-day averaged IPW/Td correlations for the entire period of record for all ten study locations show a possible monsoon onset signature (Fig. 4.10). All the cities in

the graph show a drop in the IPW/Td correlation. The cities of Albuquerque NM, El Paso, TX, Flagstaff, AZ, Las Vegas, NV, Phoenix, AZ, and Tucson, AZ collectively have a rebound in this relationship that starts ~5th three-day average (13 June – 15 June). Additionally, they clustered together at the 9th three-day average (25 June – 27 June) with a small range of high correlations all near $\rho = 0.88$. The other four cities continue to drop in the IPW/Td for this same period. On the 9th three-day average, these four cities show a range of correlations at $0.43 < \rho < 0.70$. Phoenix, AZ had the highest correlation observed for this analysis, which was at the 14th three-day average (10 July–12 July). The correlation was $\rho \cong 0.93$. This is roughly the average monsoon onset date for Phoenix, AZ.

The rank of the cities with the highest correlations and their respective three-day period in which it occurred is as follow: El Paso, TX ($\rho \cong 0.92$, 5), Tucson, AZ ($\rho \cong 0.91$, 11), Las Vegas, NV ($\rho \cong 0.90$, 13), Albuquerque, NM ($\rho \cong 0.89$, 6), Flagstaff, AZ ($\rho \cong 0.88$, 12). These cities all reached their respective maximums between the three-day averaged period of (5 – 14) which translates into (13 June – 12 July). There was a decline in the relationship in the early portion of the study for these cities (three-day period 1-5), which translates to (1 June – 15 June), followed by a collective rebound that could be indicative of the encroaching monsoon season. The drop off is likely the result of high-pressure building over the Southwest during the hot and dry early summer. The occurrence of high-pressure preceding the monsoon was described in chapter 2 (see Fig. 2.3). This high pressure system aloft creates widespread subsidence that potentially decouples the IPW/Td correlation. However, as the monsoon progresses, that high-

pressure system, eventually moves to the east and transports low-level moisture into the monsoon region thereby initiating summertime thunderstorm activity. There seems to be a monsoon onset signature associated with these data in that there is a rebound of the IPW/Td relationship in mid-late-June with a correlation maximum in early July. In mid-July through mid-September there is a lull seen in this data set as well with an apparent peak with all the monsoon cities roughly around three-day period 36 (26 September-28 September). This could be indicative of a generally more active time for the monsoon season. This relationship was not perceived with the four "fringe" cities

The rank of the remaining four cities with the lowest correlation and their respective three-day period in which it occurred is as follows: Yuma, AZ ($\rho \cong 0.82$, 1), Midland, TX ($\rho \cong 0.82$, 39), Guaymas, MX ($\rho \cong 0.77$, 1), San Diego, CA ($\rho \cong 0.70$, 12). These cities reached their respective maximums between (1 June – 1 October). This spans the entire seasonal study period. These four "fringe" cities showed a huge reduction in the correlation between IPW/Td at the onset of boreal summer that lasted until the middle of July. San Diego, CA, Midland, TX, and Yuma, AZ showed a steady increase in the correlation between Td and IPW as the summer progressed. Guaymas, MX showed a steady decline in the relationship throughout the entire summer. The reasoning for these cities' declines is likely localized. For San Diego, CA, as the summer progresses, the Pacific Ocean slowly rises in temperature, leading to more evaporation and greater atmospheric instability. Midland, TX, during the mid-summer does not see a large amount of dynamic systems, thus losing the ability to vertically mix any moisture. Guaymas, MX, a city that experiences the greatest monsoon influence, with respect to

yearly precipitation percentage, is also a region with the lowest correlations. This may be a byproduct of too much low-level moisture for the atmosphere to efficiently vertically mix. Finally, Yuma, AZ, a city that is close enough to the Gulf of California, like Guaymas, MX, may have the same pitfalls with intense moisture surges spiking low-level moisture, but the atmosphere not having the strength to vertically mix, and thus the relationship between IPW/Td breaks down. Altogether, it appears there is unique relationship with the correlations between IPW/Td with respect to the traditional monsoon cities that is simply not seen in the surrounding areas.

4.6 Summary

This thesis explores the relationship between integrated preciptiable water (IPW) and surface dew point (Td) in the North American Monsoon (NAM) region. This relationship has a long history in monsoonal forecasting. Specifically, surface Td has long been used as an indicator of moisture advection into southwest United States, which starts the monsoon season. This study reevaluates the validity of this relationship by examining it across various timescales (three-hour, three-day, weekly, monthly, yearly), across ten cities (Albuquerque, NM, El Paso, TX, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, Yuma, AZ), and evaluating the relationship of IPW and accumulated precipitation (ACPC) in order to delineate any differences between this relationship for cities inside the (NAM) region or on its periphery.

Through the course of my analysis, ten important conclusions regarding these relationships and their temporal and spatial evolutions can be extracted.

1) The IPW and Td values were largely non-normal distributions for most timescales. Non-normal distributions occurred for the three-hour, three-day, weekly, and monthly (with the exception of San Diego, CA) average of these data. However, IPW and Td were normally distributed for the yearly mean measurements for all the cities. This shows that for sub-seasonal timescales the reliability of IPW and Td to be normally distributed was very limited, and reiterates the predictable ebb and flow of moisture into and out of the southwest United States and northwest Mexico during the boreal summer.

2) The three-hour IPW/Td correlation results for the ten study locations produces a wide range of values ($0.626 \leq \rho \leq 0.926$). There was a shared variance of between 39 to 86 percent. Greater variability was observed in the correlation between cities on the perimeter of the study region compared to the inside. The perimeter cities like Guaymas, MX, Las Vegas, NV, Midland, TX, and San Diego, CA had a range of $0.625 \leq \rho \leq 0.914$. The internal cities Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Phoenix, AZ, Tucson, AZ, and Yuma, AZ had a much smaller range $0.801 \leq \rho \leq 0.926$. This was important because Guaymas, MX, Midland, TX, and San Diego, CA showed a marked difference in the way Td was related to IPW for cities on the inside of my study area versus outside on sub-daily timescale.

3) The lowest correlations for the three-hour study were observed in Guaymas, MX, Midland, TX, and San Diego, CA. For San Diego, CA, the low correlation was likely the result of three things a) Low-Level marine layers b.) Cold sea-surface temperatures that can inhibit IPW irrespective of any existing forcing (Chaboureau et al. 1998), c.) Absence of dynamic systems in these areas (Reber and Swope 1972). For Guaymas, MX marine layers and lack of dynamic forcing are involved as well. For Midland, TX, which is far removed from any large source of water, it experiences a lack of summer-time dynamic systems that typically remain north of this area (Schwartz 1968). The regions more closely associated with the NAM provide the best relationship of IPW/Td for three-hour measurements.

4) The dew point threshold once held to indicate monsoon onset for Phoenix, AZ (12.78°C) did not correspond to 25.4mm of IPW. It was closer to 33.94mm (1.33”). Phoenix, AZ only needed a Td of ~ 8.64°C to correspond to 1” of IPW. Tucson, AZ, was well-matched with its once held monsoon onset threshold (12.22°C) as this was closely related to 25.62mm (1.00”) of IPW. This meant that the Td/IPW relationship over Phoenix, AZ has not been well constructed and was in need of reevaluation. The IPW/Td values for Phoenix, AZ/Tucson, AZ came from Reitan’s (1963) equation. Moreover, this conclusion reiterated how IPW/Td measurements can vary spatially as has been documented in the literature (e.g., Zhen and Lu 2004).

5) Large variability exists for possible IPW outcomes given a certain Td. That suggests that strong vertical mixing was likely a necessary condition needed to sustain the IPW/Td relationship, therefore, as surface Td increased, only the *potential* for a high IPW correspondingly increased. The presence of a high Td did not guarantee a high IPW.

6) The size of variability for IPW outcomes based on a specific Td changes as the Td increased. In other words, for roughly a median Td value in a given area, this was where the greatest variability was for a given IPW. Variability started to shrink at a certain Td threshold. This was likely a result of low-level moisture boosting surface based instability. Therefore at certain benchmark, increased surface moisture simultaneously increased the atmosphere's ability to thermodynamically mix, thereby providing the IPW/Td relationship with a positive feedback mechanism.

7) The correlations between IPW/Td, for all cities in the study, increased from the three-hour measurements to the three-day mean measurements. The same trend continued from three-day mean averaged data to weekly averaged data. Thus, Td became more indicative of the IPW in the atmosphere as a result of timescale averaging irrespective of study location until mean-monthly averages. At the mean-monthly average and into the yearly mean average, there were decreases in the IPW/Td correlation, which may be the result of a shrinking N-size, given the aggregation of data into the longer time intervals.

8) When comparing IPW/Accumulated Precipitation (ACPC), no city reached a correlation above $\rho > 0.70$, and there was substantial variability with this correlation. When ACPC rainless (zeroes) events were removed, the correlation improved but not enough to show that IPW was a strong indicator of precipitation. These results suggested the need for additional atmospheric ingredients (i.e., thermodynamics/kinematics) to better anticipate precipitation events.

9) Examination of weekly averaged correlations of IPW/Td spatially (Fig. 4.9) revealed that most of the traditional monsoon cities in this study (Albuquerque NM, El Paso TX, Flagstaff, AZ Las Vegas, NV Phoenix, AZ and Tucson, AZ) showed a jump in correlation between IPW around (~25 June- 8 July). This peak is followed by a reduction of the correlation between IPW and Td, (roughly at ~27 August 27 – 9 September). This relationship likely was illustrating the monsoon burst/break phenomenon outlined by Adams and Comrie (1997). In contrast, San Diego, CA, Yuma, AZ, Midland, TX, and Guaymas, MX illustrated a more parabolic or "U-shape" to their respective histograms which indicated there was not an obvious summertime IPW/Td correlation weekly maximum. The drops in correlation may be the result of increased low-level atmospheric moisture with the co-absence of vertical lifting from dynamic and topographic sources.

This implied a uniqueness in the IPW/Td correction in the NAM region compared to areas outside the NAM.

10) Changes in successive three-day averaged IPW/Td correlations may were used to identify a possible monsoon onset signature (Fig. 4.10). All the cities in the graph showed a drop in the IPW/Td correlation. The cities of Albuquerque NM, El Paso, TX, Flagstaff, AZ, Las Vegas, NV, Phoenix, AZ, and Tucson, AZ showed maximum correlation values of $\rho \cong 0.89$, $\rho \cong 0.92$, $\rho \cong 0.88$, $\rho \cong 0.90$, $\rho \cong 0.93$, $\rho \cong 0.91$, respectively in which the period these cities reached their respective maximums from 13 June to 12 July. The four other study cities Guaymas, MX, Midland, TX, San Diego, CA, and Yuma, AZ showed maximum correlation values of $\rho \cong 0.77$, $\rho \cong 0.82$, $\rho \cong 0.70$, $\rho \cong 0.82$, respectively. These cities obtained their maximum correlation between (1 June – 1 October). All cities started the analysis off with a drop in IPW/Td correlation, which may be the result of high-pressure building in the early summer. However, as the typically synoptic NAM pattern emerged in mid-late June, the associated correlations spiked at nearly the same time. This was not the case with the other four cities despite Yuma, AZ widely considered a NAM city. Various localized reasons may account for the variability of the IPW/Td correlation in these other cities.

In the next chapter, I will extract the fundamental results from these analyses to access whether or not my two research hypotheses discussed in Chapter 1 are valid. I will then address the underlying significance of this research and place these results into an applied context in relationship to monsoon forecasting and overall preparedness for the NAM. In addition, I will include how these results may lead to future work with IPW/Td correlations.

Chapter 5: Summary and Conclusions

5.1 Summary of Research

The North American Monsoon (NAM) is a seasonal shift in atmospheric circulation that creates a summertime precipitation maximum for the southwestern United States and northwest Mexico. Movement of a large-scale anti-cyclone east of Arizona causes the wind to shift from a westerly to a southeasterly that promotes moisture advection and instability in the region. These seasonal changes in weather are potential hazardous while ironically necessary in order to sustain life by providing water into an arid region. The need to anticipate the timing of these changes is crucial for public and government preparedness, and much research effort has been put into objective methods to affirm the start of the NAM season. Currently, the start of the NAM season is not based on any meteorological measurements (e.g., humidity levels, precipitation, or wind changes); rather, it is declared by firm start and end dates (15 June – 30 September). This may pose future problems as climate change could modify the timing and intensity of the monsoon season, and the need to figure the official start date of the monsoon may be better served by going back to a method that monitors changes in atmospheric conditions.

The first purpose of this thesis was to reexamine the critical assumption of the long-held method for determining NAM onset in the southwestern United States. This technique relied on the supposition that there is a consistent relationship between surface dew point (Td) and integrated precipiatble water (IPW). Past research has suggested that an IPW of 25.4mm (1.00”) was a moisture threshold sufficient to create thunderstorm activity during the monsoon season in the southwestern United States and northwest

Mexico so long as proper forcing existed (Reitan 1957). Reitan (1963) discovered a correlation between IPW and surface humidity, whereas, the need for taking vertical measurements of water vapor in the atmosphere were not necessary to determine IPW; instead, Td could stand in as an alternative measurement. This laid the foundation for National Weather Service meteorologists in Arizona to incorporate this IPW/Td relationship as a means for determining the start of the NAM for Phoenix, AZ and Tucson, AZ (Franjevic 2017).

Many researchers have explored the relationship throughout the world with some mixed results. Studies in west Africa (e.g., Adedokun 1983; Oduro-Afriyie 1992), Canada (Hay 1970), over the ocean (Benwell 1965) yielded results consistent with those of Reitan (1963). On the contrary, studies by Lowry and Glahn (1969), Schwarz (1968), Reber and Swope (1972), and others brought into question the strength of the correlation. The research appeared to support an IPW/Td relationship but only under assured conditions. This uncertainly prompted me to reevaluate the IPW/Td relationship for the NAM region. Although this method for determining the NAM season was replaced in 2008 (Haffer 2008), the IPW/Td relationship is still measure that all forecasters consider when creating their daily monsoonal forecasts, especially in the context of flash flood forecasting (Moore et al. 2015). Therefore, this reevaluation of the IPW/Td relationship provided a necessary revisionist perspective on a key part of the NAM's history, and may have yielded a potential new method for determining monsoon onset.

The second purpose of this thesis was to explore the IPW/Td correlation as a possible new method for determining the start of the monsoon season. I questioned

whether there was a uniqueness to the IPW/Td relationship for NAM cities versus non-NAM cities and hypothesized that there was. I based this on previous literature that suggested the need for robust vertical mixing to support a strong IPW/Td relationship (Stull 2009). I suspected that early summer high-pressure (i.e., subsidence) over the NAM region just prior to the start of the monsoon season would degrade the IPW/Td relationship followed by a rebound in the correlation coefficient between IPW/Td from monsoon moisture and instability. A rebound in the correlation would preempt the start of the NAM and thereby provide an objective way to signal its start. In order to answer these research questions, I developed a study that features cities located inside and outside of the NAM's region of influence.

For this study, I selected ten cities (Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Guaymas, MX, Las Vegas, NV, Midland, TX, Phoenix, AZ, San Diego, CA, Tucson, AZ, and Yuma, AZ), which all the cities played a specific role in the study either providing climatological difference, topographic complexity, or a comparison location outside of the NAM's domain with a length-of-record spanning 37 years, from 1979-2015. This provided a lengthy extent to the study in combination with enough data points to conduct statistical significance testing. In addition, the length of this study was longer than other related research regarding the IPW/Td relationship and the NAM (e.g., Skindlov 2007; Means 2013).

Furthermore, because critics of Reitan's (1963) conclusion cited "enhancement" in the IPW/Td correlation as a result of data smoothing from using longer timescales (Reber and Swope 1972), I sought to examine this relationship not only spatially, but

temporally as well. The selected timescales included: three-hour non-averaged, three-day mean, weekly mean, monthly mean, and yearly mean measurements. A more thorough justification for the selected sites and selected timescales was discussed in Chapter 3.

The methods for employing my statistical testing required the use of data from the North American Regional Reanalysis (NARR) (esrl.noaa.gov), selected direct measurements from the University of Wyoming's archived radiosonde site (weather.uwyo.edu) in combination with the Statistical Package for the Social Sciences (SPSS) software. Because of potential shortcomings with using NARR data for this study, a comparison analysis between how well the NARR and direct measurements represent both IPW and Td was conducted for five randomly selected cities using measurements taken at 0000 UTC and 1200 UTC. The results showed some shortcomings, but the NARR still appeared as a very useful dataset in order to conduct this study.

The next step in my research was to analyze all the data across all locations for all timescales using descriptive statistics and proper normality testing. The results of these tests informed the appropriate correlation testing to determine if an association existed for IPW/Td in the NAM region and the selected periphery cities. Most datasets showed non-normal distributions, which required the use of Spearman's rank-order correlation (ρ), however, yearly averaged dataset for certain cities were normally distributed making it appropriate to use Pearson's correlation test (r) for the data in that timescale. All results were presented in Chapter 4.

Fundamentally, I identified ten important results from this testing.

1) IPW and Td were non-normal distributions for four out of five the timescales (three-hour, three-day, weekly, and monthly (with the exception of San Diego, CA). IPW and Td were normally distributed for the yearly mean measurements only. This showed a seasonal increase in moisture during the summer in accordance with results found by Lu et al. (2009), but it suggested sub-seasonal variability.

2) The three-hour IPW/Td correlation results produced a large range of values ($0.626 \leq \rho \leq 0.926$). Cities on the fringes of my study (Guaymas, MX, Las Vegas, NV, Midland, TX, and San Diego, CA had a larger correlation range ($0.625 \leq \rho \leq 0.914$) then those on the inside of the study (Albuquerque, NM, El Paso, TX, Flagstaff, AZ, Phoenix, AZ, Tucson, AZ, and Yuma, AZ) ($0.801 \leq \rho \leq 0.926$).

3) The lowest correlations for the three-hour study were observed in Guaymas, MX, Midland, TX, and San Diego, CA. For San Diego, CA, the low correlation was likely the result of three things 1.) Low-Level marine layers 2.) Cold sea-surface temperatures that can inhibit vertical mixing (Chaboureau et al. 1998), 3.) Lack of upper-level dynamics as a result of poleward migration of the summertime subtropical jet (Arias et al. 2015). For Guaymas, MX marine layers and lack of dynamic forcing are involved as well. For Midland, TX, which is far removed from any large source of water, it experiences a lack of summer-time dynamic systems that typically remain north of this area (Schwarz 1968). The regions more closely associated with the NAM provide the best relationship of IPW/Td for three-hour measurements.

4) Surface Td of (12.78°C) in Phoenix, AZ did not correspond to 25.4mm of IPW. It was closer to 33.94mm (1.33”); rather, Phoenix needs a surface Td of ~ 8.64°C for 1” of IPW. Tucson, AZ, was well-matched with its once held monsoon onset threshold (12.22°C). This Td in Tucson, AZ was closely related to 25.62mm (1.00”) of IPW. This analysis provided an additional benefit. The results of Phoenix, AZ are similar to those found by (Skindlov 2007) and for Tucson, AZ are similar to those found using Reitan’s (1963) regression equation. A concern with using NARR data on very short time scales was described in Chapters 2 and 3, however, by showing that IPW/Td is correlated to previously determined values in other studies that used direct measurements, this provides some reassurance that the NARR yielded accurate three-hour non averaged moisture data despite the cited concerns.

5) In order to support a high correlated IPW/Td relationship, vertical mixing appears to be a necessary ingredient. The IPW/Td relationship can break down in the absence of vertical motion. The results of my study illustrated this. For my primary monsoonal cities, as surface Td increased, only the potential for a high IPW increased. The presence of a high Td did not guarantee a high IPW.

6) Variability between Td input and a specific IPW output started to shrink at a certain Td threshold (specific to each location). Increase low-level moisture means that extensive overturning of the atmosphere is needed to vertically transport moisture aloft and to obtain a strong IPW/Td relationship. However, if surface Td increases, correspondingly, surface instability as increases especially in the NAM region (McCollum et al. 1995; Adams and Comrie 1997). Therefore, increased surface moisture simultaneously increases the atmosphere's ability to thermodynamically mix, thereby providing the IPW/Td relationship with a positive feedback mechanism.

7) The correlations between IPW/Td increased from the three-hour measurements to three-day mean measurements and again from three-day mean averaged data to weekly averaged data. Td became more associated with IPW as a result of timescale averaging. However, there was a drop in correlations from weekly to monthly averaged and from monthly to seasonally averaged. This may be the result of a decreasing sample size.

8) When comparing IPW/ACPC, no city reached a correlation between above $\rho > 0.70$. These results suggested the need for the presence of additional thermodynamics/kinematics parameters to better correlate IPW to precipitation.

9) There was a distinct difference in weekly averaged correlation histogram shapes between the typical monsoon cities in this study (Albuquerque NM, El Paso TX, Flagstaff AZ, Las Vegas NV, Phoenix, AZ and Tucson, AZ) in contrast to the remaining cities (San Diego, CA, Yuma, AZ, Midland, TX, and Guaymas, MX) (Fig. 4.9). The typical monsoon cities had a bimodal trough/peak shape while the other cites had a more parabolic or "U-shape" to their respective histograms. This shows a uniqueness in the IPW/Td correction in the NAM region compared to areas outside the NAM.

10) Successive three-day averaged IPW/Td correlations showed a possible monsoon onset signature (Fig. 4.10). All cities began the monsoonal season with a drop in IPW/Td correlation, which may be a result of high-pressure building in the early summer. As the typically synoptic NAM pattern emergences in mid-late June, the associated correlations of the some monsoon cities (Albuquerque NM, El Paso TX, Flagstaff AZ, Las Vegas NV, Phoenix, AZ and Tucson AZ) peaked at nearly the same time (Fig. 4.10). This was not the case with the other four cities despite Yuma, AZ being widely considered a core NAM city. Various localized reasons may account for the variability all related to a lack of consistent dynamic forcing outside the NAM's domain.

5.2 Implications of this Study

The purpose of this thesis has been to examine the relationship between integrated precipitable water (IPW) and surface dew point (Td) in the North American Monsoon (NAM). Essentially, this study has provided a closer insight into the overarching relationship between IPW/Td for Southwest North America. There is a correlation between the two as set forth by Reitan (1963) as well (Benwell 1965; Smith 1966; Ojo 1970). However, from the results I concluded this relationship is highly variable, spatially and temporally.

I hypothesized that as the timescale decreased towards a diurnal scale, the relationship of surface dew point and IPW decreased correspondingly. Through the statistical analyses of ten locations over five different time intervals using data from the NARR from 1979 to 2015, I have shown that this was the case for all cities when comparing the three-hour, three day, and the weekly mean measurements. Conversely, there was a decrease in the correlation as the timescale moved from weekly to seasonally averaged data. This may be the byproduct of a shrinking N-size, but I cannot unconditionally confirm my hypothesis. A strong conclusion is that, for my study, timescale averaging did enhance the IPW/Td relationship from three-hour to weekly as expected. That said, the core point is that this relationship is temporally variable and broad assumptions about the IPW/Td relationship cannot be made simply from direct surface measurements. But, what does it suggest about the state of the atmosphere when the correlation is strong?

Reber and Swope (1972) were critical of the overall strength of this relationship and my study does verify their final conclusion, specifically, how timescale averaging can enhance the IPW/Td relationship. However, their criticisms of the IPW/Td relationship may have been somewhat fortuitous. Their study was confined to southern California, which is an area that does not have an abundance of uplift mechanisms to vertically mix low-level moisture and even cited that capping inversions from cold coastal waters may have caused their results. Again, showing vertical mixing as a necessary condition to the IPW/Td relationship as was seen in my results and other cited literature (e.g., Schwarz 1968; Lowry and Glahn 1969 Hay 1970; Karalis 1974; Tuller 1977; Revuelta et al. 1985; Chaboureau et al. 1998). Considering this, in conjunction with the known seasonal jumps of instability in the desert Southwest and northwest Mexico, I hypothesized that there was a unique relationship between the IPW/Td for NAM cities that may signal monsoon onset.

Cities on the interior of NAM's domain, areas considered to have predictable moisture and uplift during the monsoon months, had stronger correlations then those cities on the perimeter of the NAM region (with the exception of Yuma, AZ). All NAM cities in the study showed a reduction in positive correlations between IPW/Td during the early parts of the boreal summer, as expected from synoptic conditions causing widespread stability, with a visibly apparent increase in correlations just prior to the average monsoon onset time. This rebound may show the conflation of the two fundamental ingredients in monsoon storm genesis (moisture and uplift) all represented in this single correlation. In addition, the cities' correlations may also show a weakening

of the high-pressure subsidence just prior to the influx of surface moisture and instability as the NAM season commences. This may suggest that this relationship could be useful in diagnosing NAM onset that has not been previously mentioned in literature.

The cities on the NAM's perimeter are more variable in their low-level moisture availability and forcing during summer months. The correlations for these cities markedly contrasted those for NAM cities. According to my research, there is a mid-June spike in correlations between IPW/Td in which six of the cities (ABQ, ELP, FLAG, PHX, TUC, LAS) correlations all converge at nearly the same three-day time period (period 9). Yuma, AZ showed a similar pattern, but failed to achieve the same level of correlation. Guaymas, MX, also considered to be part of the NAM region, related more closely in its correlation coefficient evolution to San Diego, CA, and Midland, TX. As a result, if the correlation coefficient between IPW/Td is a measure that can identify monsoon onset, it would not apply very well to Yuma, AZ and Guaymas, MX. I suspect that these poor correlations are the result of low-level moisture, and capping inversions that can occur near the Gulf of California when the sea surface temperature is below 29°C (Erfani and Mitchell 2004). Yuma, AZ may be additionally affected by capping inversions that can be created from decoupled elevated mixing layers in mountainous regions (Whitman 2000; Warner 2004; Stull 2009).

In general, Yuma, AZ is an outlier city in this study. While Guaymas, MX is a NAM city, Guaymas is located in the subtropics and is relatively far removed from regional influences in the southwestern United States. In contrast, Yuma, AZ has a well-cited history of intense low-level moisture surges (Hales 1972; Brenner 1974; Dixon

2005), but that does not translate into greater rainfall by comparison with regions in eastern Arizona that have lower dew point temperatures (Adams and Comrie 1997; Higgins et al. 2004). Adams and Comrie (1997) go on to explain that precipitation variability is closely likened to the moisture availability in "interior" NAM regions citing (Hales 1972; Brenner 1974), which is analogous to the results of my study. Fundamentally, the use of IPW/Td correlations as a regional NAM onset indicator may be justified, so long as the caveat is stated that Yuma, AZ is an atypical NAM city.

5.3 Recommendations for Future Research

The North American Monsoon (NAM) is a vital component to sustaining life in the desert Southwest of the Unted States and northwest Mexico. Consequently, any intended or unintended modification to monsoon onset/demise or intensity could have significant effects on life and property. These results may help to facilitate research into revised objective methods for the start of the monsoon season that supersede the current method.

I would suggest that the next step of this research is to continue the spatial and temporal evaluation of the IPW/Td relationship across the southwest United States and northwest Mexico. For my study, I only selected ten discrete locations, whereas with enough resources the NARR could significantly enhance the spatial resolution of a similar study. The examination of IPW/Td could be redone with thousands of grid points from 1979-present. Geographically speaking, this would enhance any spatial variation of the IPW/Td relationship. I purported that Yuma, AZ did not exhibit similar IPW/Td

correlations as other cities in the NAM, it would be interesting to explore at which location between Yuma, AZ and Tucson, AZ does the IPW/Td relationship breakdown.

Because my study only included 37 years of data (1979-2015), any trends in the NAM prior to this were not included in this study. There could be vital information lost as a result of the temporal limitation of the NARR dataset. Consequently, conducting a study that uses the entire radiosonde length-of-record, despite the cited pitfalls with those data (Bosart 1990), could be useful. If rebound of IPW/Td correlation is observed in data pre-1979 this would add further credibility to using the IPW/Td correlation as a monsoon onset variable.

5.4 Significance of this Study

The NWS has shifted from employing a surface humidity parameter (e.g., surface Td) as an indicator of monsoon onset to a pre-set start and end date to signal the beginning of the season (Haffer 2008). While this may be beneficial for public readiness, there has been research suggesting that the NAM climatological onset/demise and intensity may shift as result of climate change (e.g., Cook and Seger 2013; Arias et al. 2015). There is uncertainty in the current research regarding the exact reasons for potential changes to NAM onset/demise, and to what extent, but anthropogenic climate change is likely a factor (Arias et al. 2012; Arias et al. 2015). Therefore, a prescribed start date may be a flawed means to establish the start of the monsoon season. Consequently, the temporal and spatial evolution of the IPW/Td correlation as presented in this thesis may provide a framework for future research that reevaluates the NAM's domain and the associated methods for determining its onset.

REFERENCES

Abo, T. O., 1975: Analysis of data on surface and tropospheric water vapour. *Journal of Atmospheric and Terrestrial Physics*, **38,** 565-571.

Adams, D. K., and A. C. Comrie, 1997: The North American Monsoon. *Bull. Amer. Meteor. Soc.,* **78**, 2197-2213.

Adams, J. L., and D. J. Stensrud, 2007: Impact of tropical easterly waves on the North American Monsoon, *J. Climate*, **20,** 1219-1238.

Adang, T. C., and R. Gall, 1989: Structure and dynamics of the Arizona monsoon boundary. *Mon. Wea. Rev.*, **117,** 1423-1438.

Adedokun, J. A., 1983: Intra-Layer (Low-Level/Mid-Tropospheric) precipitable water vapour relations and precipitation in West Africa. *Archives for Meteorology, Geophysics, and Bioclimatology*. **33,** 117-130.

Adedokun, J. A., 1986: On a relationship for stimulating precipitable water vapour aloft from surface humidity over West-Africa. *Journal of Climatology,* **6,** 161-172.

Alberty, R., 1986: Letter to NWS staff. Clarification of the term "Arizona Monsoon", 1 pp.

AMS, 2017: *Glossary of Meteorology*. Accessed 5 May 2017. [Available online at http://glossary.ametsoc.org/wiki/Main_Page].

Anderson, B. T., J. O. Roads, S. C. Chen, and H. M. H. Juang, 2001: Model dynamics of summertime low-level jets over northwestern Mexico. *Journal of Geophysical Research,* **106,** D4.

Anyadike, R. N. C., 1979: The content of water vapour in the atmosphere over West Africa. *Archives for Meteorology, Geophysics, and Bioclimatology.* **28,** 245-254.

Arias, P. A., R. Fu, and K. Mo, 2012: Changes in monsoon regime over northwestern Mexico in recent decades and its potential causes. *J. Climate*. **25,** 4258-4274.

Arias, P. A., R. Fu, C. Vera, and M. Rojas, 2015: A correlated shortening of the North and South American monsoon seasons in the past few decades. *Climate Dynamics.* **45,** 3183-3203.

Balling, R. C., Jr., and S. W. Brazel, 1987: Diurnal variations in Arizona monsoon precipitation frequencies. *Mon. Wea. Rev.*, **115,** 342-346.

Becker, E. J., and E. H. Berbery 2008: The diurnal cycle of precipitation over the North American monsoon region during the NAME 2004 field campaign. *J. Climate,* **21,**771-787.

Benwell, G. R. R., 1965: Estimation and variability of precipitable water. *Meteorology Magazine*, **94**, 319-327.

Berbery, E. H., and M. S. Fox-Rabinovitz, 2003: Multiscale diagnosis of the North American monsoon system using a variable resolution GCM. *J. Climate*, **16,** 1929-1947.

Berg, L. K., L. D Rihimaki, Y. Qian, H. Yan, and M. Huang, 2015: The Low-Level jet over the southern Great Plains determined from observations and reanalyses and its impact on moisture transport. *J. Climate*, **28,** 6682-6706.

Bieda, S. W., C. L. Castro, S. L. Mullen, A. C. Comrie, and E. Pytlak, 2009: The relationship of transient upper-level troughs to variability of the North American monsoon system. *J. Climate*, **22**, 4213-4227.

Bolsenga, S. J., 1965: The relationship between total atmospheric water vapor and surface dew point on a mean daily and hourly basis. *J. Appl. Meteor*., **4**, 430-432.

Bosart, L. F., 1990: Degradation of the North American radiosonde network. *Wea. Forecasting,* **5,** 680-690.

Brenner, I. S., 1974: A surge of maritime tropical air--Gulf of California to the Southwestern United States. *Mon. Wea. Rev*., **102,** 375-389.

Bryson, R. A., and W. P. Lowry, 1955: The synoptic climatology of the Arizona summer precipitation singularity. *Bull. Amer. Meteor. Soc*., **36**, 329-339.

Burkovsky, M. S., and D. J. Karoly, 2007: A brief evaluation of precipitation from the North American Regional Reanalysis. *J. Hydrometeor.* **8,** 837-846.

Carelton, A. M., D. A. Carpenter, and P. J. Weber, 1990: Mechanisms of interannual variability of the southwest United States summer rainfall maximum. *J. Climate*, **3,** 999-105

Carruthers, B., 1953: *Handbook of Statistical Methods in Meteorology*. AMS Press, 412 pp.

Chaboureau, J. P., A. Chedin, and N. A. Scott, 1998: Relationship between sea surface temperature, vertical dynamics, and the vertical distribution of atmospheric water vapor inferred from TOVS oversvations. *Journal of Geophysical Research,* **103,** D18.

Cook, B. I., and R. Seger, 2013: The response of the North American Monsoon to increased greenhouse gas forcing. *Journal of Geophysical Research: Atmosphere,* **118,** 1690-1699

Corbosiero, K. L., M. J. Dickinson, and L. F. Bosart, 2009: The contribution of eastern North Pacific tropical cylones to the rainfall climatology of the southwest United States. *Mon. Wea. Rev*., **137**, 2415-2435.

Diem, J. E., 2005: Northward extension of intense monsoonal activity into the southwestern United States, *Geophysical Research Letters,* **32,** L14702.

Diem, J. E., and D. P. Brown, 2006: Tropospheric moisture and monsoonal rainfall over the southwestern United States. *Journal of Geophysical Research,* **111,** D16112.

Dixon, P. G., 2005: Using sounding data to detect gulf surges during the North American monsoon. *Mon. Wea. Rev*., **133**, 3047-3052.

Douglas, M. W., R. A. Maddox, K. Howard, and S. Reyes, 1993: The Mexican monsoon. *J. Climate*, **6,** 1665-1677.

Douglas, M. W., and J. C. Leal, 2003: Summertime surges over the Gulf of California: Aspects of their climatology, mean sounding, and evolution from radiosonde, NCEP reanalysis, and rainfall data*. Wea. Forecasting*, **18**, 55-74.

Ellis, A. W., E. M. Saffell, and T. W. Hawkins, 2004: A method for defining monsoon onset and demise in the southwestern USA. *International Journal of Climatology,* **24,** 247-265.

Erfani, E., and D. Mitchell, 2014: A partial mechanistic understanding of the North American monsoon. *Journal of Geophysical Research: Atmospheres,* **119,** 13,096-13,115.

FAA, 2017: Commercial service enplanements. Accessed 5 May 2017. [Available at https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/media/cy15-commercial-service-enplanements.pdf]

Favors, J. E., and J. T. Abatzoglou, 2013: Regional surges of monsoonal moisture into the southwestern United States. *Mon. Wea. Rev.,* **141,** 182-191.

Flatau, M. K., P. J. Flatau, J. Schmidt, and G. N. Kiladis, 2003: Delayed onset of the 2002 Indian monsoon. *Geophysical Research Letters*, **30,** 1-4.

Foster, J., M. Bevis, and W. Raymond, 2006: Precipitable water and the lognormal distribution. *Journal of Geophysical Research,* **111.** D15102.

Franjevic, M. W., 2017: 55°F *Dewpoint – History and Investigation*. Undated, but apparently written in the late 1990's. Available at the National Weather Service Forcast Office, Phoenix, AZ. Accessed 15 May 2017, 2 pp.

Fuller, R. D., and D. J. Stensrud, 2000: The relationship between tropical easterly waves and surges over the Gulf of California during the North American monsoon. *Mon. Wea. Rev*., **128**, 2983-2989.

Gueymard C., 1994: Analysis of monthly average atmospheric precipitable water and turbidity in Canada and the northern United States. *Solar Energy*, **53**, 57-71.

Haffer, T., 2008: *National Weather Service Explains the Arizona Monsoon*. Public Information Statement NWSFO Phoenix, AZ, 1 pp. [Available at Iowa State NWS Text Archive AFOS Product finder https://mesonet.agron.iastate.edu/wx/afos/p.php?pil=PNSPSR&e=200803112201]

Hales, J. E., Jr. 1972: Surges of maritime tropical air northward over the Gulf of California. *Mon. Wea. Rev*., **100,** 298-306.

Hales, J. E., Jr. 1974: The southwestern United States summer monsoon source---Gulf of Mexico or Pacific Ocean? *J. Appl. Meteor*., **13,** 331-342.

Harington, J. A., Jr. R. S. Cerveny, and R. C., Balling., 1992: Impact of the southern oscillation on the North American Southwest Monsoon. *Physical Geogr*aphy, **13,** 318-330.

Hay J. E., 1970: Precipitable water over Canada: I computation. *Atmosphere*, **8,** 128-143.

Higgins. R. W., Y. Yao, and X. Wang, 1997: Influence on the North American monsoon system on the U.S. summer precipitation regime. *J. Climate*, **10,** 2600-2622.

Higgins, R. W., Y. Chen, and A. V. Douglas 1999: Interannual variability of the North American warm season precipitation regime, *J. Climate,* **12,** 653-680.

Higgins, R. W., W. Shi, and C. Hain, 2004: Relationships between Gulf of California moisture surges and precipitation in the southwestern United States. *J. Climate,* **17,** 2983-2997.

Higgins, R. W., and W. Shi, 2005: Relationships between Gulf of California moisture surges and tropical cyclones in the eastern pacific basin. *J. Climate*, **18,** 4601-4620.

Idso, S. B., 1969: Atmospheric Attenuation of Solar Radiation. *J. Atmos. Sci.,* **26,** 1088-1095.

INEGI, 2017. Instituto Nacional de Estadistica y Geografia. Accessed 5 May 2017. [Available at http://www.inegi.org.mx]

Jurwitz, L. R., 1953: Arizona's two-season rainfall pattern. *Weatherwise*, **6**, 96-99.

Kanamitsu, M., and K. C. Mo, 2003: Dyanmical effect of land surface processes on summer precipitation over the southwestern United States, *J. Climate,* **16,** 494-509.

Karalis, J. 1974: Precipitable water and its relationship to surface dew point and a vapor pressure in Athens. *J. Appl. Meteor.* **13,** 760-766.

King, T. S., and R. C. Balling, 1994: Diurnal variations in Arizona monsoon lightning data. *Mon. Wea. Rev*., **122,** 1659-1664.

Kistler, R., and Coauthors, 2001: The NCEP-NCAR 50 year reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.,* **82,** 247-267.

Lin, Y., K. E. Mitchell, E. Rogers, M. E. Baldwin, and G. J. DiMego, 1999: Test assimilations of the real-time, multi-sensor hourly precipitation analysis into the NCEP Eta. Model. Preprints, *8th Conference on Mesocscale Meteorology*, Boulder, CO, AMS., 341-344.

List, R. J., 1963: *Smithsonian Meteorological Tables, 6th rev. ed.* Washington, D. C., Smithsonian Institution.

Lowry, D. A., and H. R. Glann, 1969: Relationship between integrated atmospheric moisture and surface water. *J. Appl. Meteor*., **8**, 762-768.

Lu, E. R., Z. Zeng, Z. Jiang, Y. Wang, and Q. Zhang, 2009: Precipitation and precipitable water: Their temporal-spatial behaviors and use in determining monsoon onset/retreat and monsoon regions. *Journal of Geophysical Research*, **114,** D23105.

Maddox, R. A., D. M. McCollum, and K. W. Howard 1995: Large-scale patterns associated with severe summertime thunderstorms over central Arizona. *Wea. Forecasting*, **10,** 763-778.

Maghrabi, A., and H. M. Al-Dajani, 2013: Estimation of precipitable water vapour using vapour pressure and air temperature in an arid region in central Saudi Arabia. *Journal of the Association of Arab Universities for Basic and Applied Science,* **14,** 1-8.

McCollum, D. M., R. A. Maddox, and K. W. Howard, 1995: Case study of a severe mesoscale convective system in central Arizona. *Wea. Forecasting*, **10,** 643-665.

McGee, O. S., 1974: A surface dewpoint-precipitable water vapour relationship for South Africa. *South African Journal of Science,* **70,** 119-120.

McGrath R., T. Semmler, C, Sweeney, and S. Wang, 2006: Impact of balloon drift errors in radiosonde data on climate statistics. *J. Climate*. **19,** 3430-3442.

Means, J. D., 2013: GPS precipitable water as a diagnostic of the North American Monsoon in California and Nevada. *J. Climate*, **26**, 1432-1444.

Mejia, J. R., M. W. Douglas, and P. J. Lamb, 2016: Observational investigation of relationship between moisture surges and mesoscale- to large-scale convection during the North American monsoon. *International Journal of Climatology,* **36,** 2555-2569.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc*., **87,** 343-360.

Mo, K. C., M. Chelliah, M. L. Carrera, R. W. Higgins, and W. Ebisuzaki, 2005a: Atmospheric moisture transport over the United States and Mexico as evaluated in the NCEP regional reanalysis. *J. Hydrometeor.,* **6,** 710-728.

Mo, K. C., J. K. Schemm, H. M. H. Juang, R. W. Higgins, and Y. Song, 2005b: Impact of model resolution on the prediction of summer precipitation over the United States and Mexico. *J. Climate,* **18,** 3910-3827.

Moore, A. W., and Coauthors, 2015: National Weather Service forecasters use GPS precipitable water vapor for enhanced situation awareness during the southern California summer monsoon. *Bull. Amer. Meteor. Soc.,* **96,** 1967-1877.

NCEP, 2017: *North American Regional Reanalysis*. Accessed 25 March 2017. [Available at http://www.esrl.noaa.gov/psd/].

Noska, R., and V. Misra, 2016: Characterizing the onset and demise of the Indian summer monsoon. *Geophysical Research Letters*, **43,** 4547-4554.

NWS, 2017: National Weather Service mission statement. Accessed 6 May 2017. [Available at http://www.nws.noaa.gov/mission.php].

Oduro-Afriyie, K., 1992: Estimating precipitable water over West Africa from surface equivalent potential temperature and related potential instability indicies. *Atmospheric Research,* **28,** 11-20.

Ojo, O., 1970: The distribution of mean monthly precipitable water vapor and annual precipitation efficiency in Nigeria. *Archives for Meteorology, Geophysics, and Bioclimatology,* **18,** 221-238.

Pascale, S., and S. Bordoni, 2016: Tropical and extratropical controls of Gulf of California surges and summertime precipitation over the Southwestern United States. *Mon. Wea. Rev.,* **144,** 2695-2718.

Radhakrisha, B., F. Fabry, J. J. Braun, and T. Van Hove, 2015: Precipitable water from GPS over the continental United states: diurnal cycle, intercomparisons with NARR, and link with convective initiation. *J. Climate*. **28,** 2584-2599.

Rao, N. J. M., Y. Viswandham, and T. V. Ramana Rao, 1979: A preliminary study of precipitable water over Brazil. *Pure and Applied Geophysics,* **117**, 883-890.

Rao, N. J. M., Y. Viswanadham, and G. S. S. Nunes, 1980: Moisture relationships for the Southern Hemisphere. *Pure and Applied Geophysics,* **118,** 1076-1089.

Reber, E., and J. Swope, 1972: On the correlation of the total precipitable water in a vertical column and absolute humidity at the surface. *J. Appl. Meteor*., **11**, 1322-1325.

Reed, T. R., 1933: The North American high level anticyclone. *Mon. Wea. Rev*., **61,** 321-325.

Reitan, C. H., 1957: The role of precipitable water vapor in Arizona's summer rains. Tech. Rep. on the Meteorology and Climatology of Arid Regions 2, The Institute of Atmospheric Physics, The University of Arizona, Tucson, 19 pp.

Reitan, C. H., 1963: Surface dew point and water vapor aloft. *J. Appl. Meteor*., **2**, 776-779.

Revuelta, A., C. Rodriguez, J. Mateos, and J. Garmedia, 1985: A model for the estimation of precipitable water. *Tellus*, **37B,** 210-215.

Rogers, P. J., and R. H. Johnson, 2007: Analysis of the 13-14 July gulf surge event during the 2004 North American Monsoon Experiment. *Mon. Wea. Rev.,* **135,** 3098-3117.

Ruane, A. C., 2010: NARR's atmospheric water cycle components. Part I: 20-Year mean and annual interactions. *J. Hydrometeor.,* **11,** 1205-1219.

Schiffer, N. J., and S. W. Nesbit, 2012: Flow, moisture, and thermodynamic variability associated with Gulf of California surges within the North American Monsoon. *J. Climate*, **25**, 4220-4241.

Schmitz, T. J., and S. L. Mullen, 1996: Water vapor transport associated with the summertime North American monsoon as depicted by ECMWF analyses. *J. Climate*, **9,** 1621-1633.

Schwarz, F. K., 1968: Comments on note on the relationship between total precipitable water and surface dew point. *J. Appl. Meteor*., **7**, 509-510.

Shafran, P., J. Woollen, W. Ebisuzaki, W. Shi, Y. Fan, R. W. Grumbine, and M. Fennessy, 2004: Observational data used for assimilation in the NCEP North American Regional Reanalysis, 14th Conference on Applied Climatology, Seatthle, WA, AMS. [Available at http://ams.confex.com/ams/84Annual/techprogram/paper_71689.htm.].

Sheppard, P. R., A. C. Comrie, G. D. Packin, K. Angersbach, and M. K. Hughes, 2002: The climate of the US Southwest. *Climate Research*, **21,** 219-238.

Sinha, S., and S. K. Sinha, 1981: Precipitable water and dew point temperature relationship during the Indian summer monsoon. *Pure and Applied Geophysics*, **119,** 913-921.

Skindlov, J., 2007: The "true" monsoon: The history and use of the 55°F dew point temperature criteria to define monsoon onset and retreat in Phoenix. Proc. Fourth Symp. On Southwest Hydrometeorology, Tucson, AZ, NWS. [Available at www.atmo.arizona.edu/~swhs/Presentations/FriAM2/].

Smith, W. L., 1966: Note on the relationship between total precipitable water and surface dew point. *J. Appl. Meteor*., **5**, 726-727.

Smith, W. P., and R. L. Gall, 1989: Tropical squall lines of the Arizona monsoon. *Mon. Wea. Rev*., **117,** 1553-1569.

Solot, S. B., 1939: Computation of depth of precipitable water in a column of air. *Mon Wea. Rev.*, **67**, 100-103.

Stensrud, D. J., R. L. Gall and K. W. Howard, 1995: Model Climatology of the Mexican Monsoon. *J. Climate,* **8,** 1775-1794.

Stensrud, D. J., R. L. Gall, and M. K. Nordquist, 1997: Surges over the Gulf of California during the Mexican monsoon. *Mon. Wea. Rev*, **125,** 417-437.

Stull, R. B., 2009: *An Introduction to Boundary Layer Meteorology*. Springer Science, 670 pp.

Svoma, B. M., 2010: The influence of monsoonal gulf surges on precipitation and diurnal precipitation patterns in central Arizona. *Wea. Forecasting*, **25,** 281-289.

Tian, B., B. J. Soden, and X. Wu, 2004: Diurnal cycle of convection, clouds, and water vapor in the tropical upper troposphere: Satellites versus a general circulation model. *J. Geophys. Res.,* **109.**

Tuller, S. 1977: The relationship between precipitable water vapor and surface humidity in New Zealand. *Archives for Meteorology, Geophysics, and Bioclimatology,* **26,** 197-212.

USCB, 2017: Interactive population search. Accessed on 5 May 2017. [Available at https://www.census.gov/2010census/popmap/ipmtext.php?fl=04].

Viswanadham, Y. 1981: The relationship between total precipitable water and surface dewpoint. *J. Appl. Meteor*. **20,** 3-8.

Warner, T. T., 2004: *Desert Meteorology*. Cambridge University Press, 595 pp.

Wang, B., Q. Ding, and P. V. Joseph, 2009: Objective definition of the Indian summer monsoon onset, *J. Climate,* **22,** 3303- 3316.

Watson, A. I., R. L. Holle, and R. E. Lopez, 1994b: Cloud-to-ground lightning patterns in AZ during the southwest monsoon. Mon. Wea. Rev., **122,** 1716-1725.

UWYO, 2017: Atmospheric Soundings. Accessed 6 May 2017. [Available at http://weather.uwyo.edu/upperair/sounding.html].

Whiteman, C. D., 2000: *Mountain Meteorology: Fundamental and Applications*. Oxford University Press, 355 pp.

Wilks D. S., 2011: *Statistical Methods in the Atmospheric Sciences: Third Edition.* Academic Press, 676 pp.

Wu, M. L. C., S. D. Schubert, M. J. Suarez, and N. E. Huang, 2009: An analysis of moisture fluxes into the Gulf of California. *J. Climate*, **22,** 2216-2239.

Zeng, X. B., and E. Lu, 2004: Globally unified monsoon onset and retreat indexes. *J. Climate*, **17,** 2241-2248.

APPENDIX A

PYTHON PROGRAM USED TO EXTRACT NARR DATA

```
1  from scipy import *
2  import netCDF4
3  import ephem
4
5  # lets explore the 2-meter dew point temperature
6  dptpath = '/users/cole/dropbox (asu)/monsoon/narr data/dpt.2m.1980.nc' # define the
7  path
8  dptdata = netCDF4.Dataset(dptpath)  # load the dataset using the netCDF4 library
9  keys = dptdata.variables.keys()  # lets see the dataset variable keys -- it acts like a
10
11
12 lat = dptdata.variables['lat'][:]
13 lon = dptdata.variables['lon'][:]
14
15 print dptdata.variables['time'].units #hours since 1800-1-1 00:00:0.0
16
17 # Deal with NARR's time index
18 time = dptdata.variables['time'][:]
19 dates = array([ephem.date('1800/1/1 00:00') + ephem.hour*t for t in time])
20 T = [ephem.date(d).datetime() for d in dates]
21
22 dpt = dptdata.variables['dpt'][0,:,:]
```

```
23 dpt = dptdata.variables['dpt'][0,:,:]
24 contourf(dpt, 100)
25
26
27 X = dptdata.variables['dpt'][:,84,159]
28
29
30
31 prpath = '/users/cole/dropbox (asu)/monsoon/narr data/pr_wtr.1980.nc'
32 prdata = netCDF4.Dataset(prpath)
33 keys = prdata.variables.keys()
34 pr_wtr = prdata.variables['pr_wtr'][:,84,159]
35
36
37
38 def find_nearest(a,b, a0,b0):
39    "Element in nd array `a` closest to the scalar value `a0`"
40    adx = np.abs(a - a0)
41    bdx = np.abs(b - b0)
42 c = (adx + bdx)
43 mbool = c.min() == c
44 #print lat[bool],lon[bool]
```

```
45 return mbool
46
47
48 np.where(find_nearest(-112,33,lon,lat)==True)
49 # (array([84], dtype=int64), array([159], dtype=int64))
50
51 Y = prdata.variables['pr_wtr'][:,84,159]
52
53
54 def KtoF(t):
55   # function to convert Kelvin to Faheinheit
56   return  (t*(9./5)) - 459.67
57
58
59 XX = X[1216:2192]
60 YY = Y[1216:2192]
61
62
63 from scipy import stats
64
65 r,p = stats.pearsonr(XX,YY) # basic corr test
66
```

```
67
68 import seaborn as sb
69
70 sb.jointplot(XX,YY)
71
72
73
74
75
76 from scipy import *
77 import netCDF4
78 import ephem
79
80
81 variables = ['dpt.2m', 'apcp','pr_wtr']
82 years = range(1979,2015)
```