RAINFALL PATTERNS IN A SMALL-SCALE WATERSHED BASED ON NEXRAD AND GROUND-BASED DATASETS

BY

CONGYU HOU

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Agricultural and Biological Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Adviser:

Assistant Professor Maria L. Chu

Abstract

Quantifying and predicting the ecosystems responses to changes in natural and anthropogenic stressors using environmental models require a realistic representation of probable rainfall in its most sensible spatial-temporal dimensions matching that of the phenomenon under investigation. As one of the most critical inputs in environmental models, rainfall data can significantly change the quality and reliability of the model predictions. Due to the lack of ground-based measurements with high spatial and temporal resolution, other methods like radar, have been recently used as an alternative source of rainfall data. However, little research has been conducted to evaluate the possible tradeoffs in using radar generated rainfall data as oppose to ground-measured rainfall. The main objective of this study was to analyze the ability of radar estimates (NexRad N1P) in representing the rainfall patterns in a small-scale watershed and generate high temporal and spatial resolution rainfall data when rainfall pattern was given. To achieve this objective, first, we compared the precipitation from NexRad N1P and ground-based measurements in the Little Washita River Experimental Watershed in Oklahoma to quantify the differences in their patterns and distributions. And second, we tested the ability of a rainfall pattern simulator software "Zeus" to generate high temporal and spatial resolution rainfall data when the rainfall pattern was given. The generated rainfall data was compared with the original rainfall data from both NexRad and ground-based stations. The comparison of the NexRad and ground-based measured rainfall revealed that the mean rainfall from radar in March, April, May, June, and August is closer to the rainfall recorded from the ground-based station with an average difference between of less than 25%. Rainfall recorded in the other months can easily be affected by extreme rainfall events, and the difference in the mean monthly rainfall can be higher than 40%. Also, the analysis showed that the NexRad has a tendency of recording the heavy and intense rainfall higher than the ground-

based rainfall while the less intense rain less. The generated rainfall based on the NexRad data has less percent error than the ground-based when simulating the dry rainfall months (Jan, Feb, Mar, Apr, Sep, Oct, Nov and Dec). The results showed that significant differences were found between the NexRad and ground-based datasets that can significantly impact the response of environmental models that used them as inputs. This study enabled us to establish the rainfall patterns by using the NexRad when the density of ground-based stations is not sufficient to derive rainfall time series with high spatial and temporal resolution. Also, the "Zeus" software could help us generate the rainfall time series when other sources of high spatial and temporal resolution data are not available. The use of synthetically generated spatial-temporal rainfall patterns will enable us to explore the impacts of precipitation on hydrologic processes driven by changes in environmental stressors like land use and climate changes.

ACKNOWLEDGEMENTS

Firstly and most importantly, I would like to express my thanks to **Dr. Maria Chu and her husband, Dr. Jorge Guzman**, for the support in the last two years. During this time, they regarded me as their own son and provided me with care and advice in both academic and normal lives. Their kindness saved me from the bottom of my life during my graduate years and helped me finish my Master degree thesis.

Secondly, I would like to thank **Dr. Praveen Kumar and Dr. Rabin Bhattarai** for the time reading my thesis and providing valuable feedback especially in the busiest time of this semester. Your kind help did not only help me in finishing my Master thesis but also pointed out the future possible research for my PHD program.

Last but not the least, I would like to appreciate my **family and friends**. I could not survive and finish my thesis alone without your support. All of you are the most valuable fortune that I can ever have and enjoy.

Table of Contents

Chapter 1: Introduction 1

Chapter 2: Objectives 4

Chapter 3: Review of literature 5
- **3.1 The importance of high-quality rainfall data in environmental models** 5
- **3.2 The NexRad rainfall dataset** 6
- **3.3 The ground-based rainfall dataset** 7
- **3.4 Representing rainfall at watershed scale** 9
- **3.5 Rainfall patterns definition and generator** 10

Chapter 4: Methodology 12
- **4.1 Study area** 12
- **4.2 Precipitation stations** 14
- **4.3 Derivation of rainfall time series** 15
- **4.4 Comparison between ground-based and NexRad data** 21
- **4.5 Rainfall pattern definition and regeneration** 23
- **4.6 The comparison between observed rainfall data and generated rainfall data** 30

Chapter 5: Result and discussion 34
- **5.1 Deriving a rainfall map based on the NexRad rainfall data in the Little Washita watershed** 34
- **5.2 Comparisons of ground-based and NexRad data** 35
- **5.3 Simulated rainfall patterns** 42

Chapter 6: Conclusions and Future Work 55
- **6.1 The NexRad rainfall time series** 55
- **6.2 The Comparisons between NexRad rainfall data and ground-based rainfall data** 56
- **6.3 Rainfall pattern analysis** 57

Reference 60

APPENDIX A: The Locations of KFDR and KTLX stations 63

APPENDIX B: The cells and stations relationship 64

APPENDIX C: Double mass curve results for twelve months 66

APPENDIX D: R result of F-test of areal monthly rainfall between NexRad and ground-based dataset 71

APPENDIX E: R result of Kolmogorov-Smirnov test between KTLX and ground-based dataset 72

APPENDIX F: R result of Kolmogorov-Smirnov test between KFDR and ground-based dataset 73

APPENDIX G: R result of Kolmogorov-Smirnov test within the NexRad stations 74

APPENDIX H: The log-scale figures of ground-based dataset of twelve months 75

APPENDIX I: The log-scale figures of KFDR dataset of twelve months 88

APPENDIX J: The log-scale figures of KTLX dataset of twelve months 101

CHAPTER 1: INTRODUCTION

Quantifying and predicting the ecosystems responses to changes in natural and anthropogenic stresses using environmental models requires a realistic representation of probable rainfall in its most sensible spatial-temporal dimensions matching that of the phenomenon under investigation. As one of the most critical inputs in environmental models, rainfall data can significantly influence the quality and reliability of the model predictions. As a result, many models cannot reach the desired accuracy and efficiency because of the lack of high-resolution and precise rainfall data.

Rainfall, which affects both energy and water flow in the ecosystem, is one of the fundamental controlling ecosystem responses including runoff production, transport phenomena, and other environmental variables (Ivanov et al., 2007). As an example, distributed rainfall is used in real-time forecasting such as severe weather and flash flood warnings in which the adequate spatial and temporal resolution is critical (Hill. et al., 2011; Moser. et al., 2015; Kim & Valdés, 2003). Distributed rainfall is also fundamental when quantifying changes in contaminant movement (e.g., sediment, nutrients, and pesticides) in agriculturally dominated watersheds due to the implementation of land management practices. Moreover, the presence or absence as well as the spatiotemporal representation of rainfall also affect other hydrometric variables (Ivanov et al., 2007), and thus, the likelihood of the system response. The change in the rainfall distribution also affects the biotic and abiotic activities in the watershed ecosystems. In 2016, the shift of the rainfall levels has been proved associated with the amount of biomass in water (Badylak et al., 2016).

To represent the areal rainfall within a watershed, it is a common practice to use ground-based rainfall data. However, in some cases, ground-based monitoring networks may not exist or possess the adequate density to characterize rainfall variability properly across the watershed resulting in increased uncertainties. On the other hand, the Next Generation Weather Radar (NexRad) may

provide an alternative to ground-based observations providing sufficient spatiotemporal resolution within the contiguous U.S. However, it may also contain a high degree of uncertainty (Ciach et al., 2007) and high rate of missing data derived from both the rainfall estimation model (e.g., Z-R relationship) and radar equipment errors (Ciach et al., 2007). Thus, practitioners need to evaluate the tradeoffs from the undesired uncertainty and restrictions driven by limited ground-based data (rain-gauge networks) and radar-rainfall estimations when assessing environmental impacts. Ground-based and radar estimated rainfall (e.g., NexRad) represent the precipitation process based on different principles and therefore result in different rainfall datasets (e.g. different time series). The differences can be due to the measurement target, time, and distance. For example, the measurement target of a ground-based station is the amount of rain that hits the ground while for the NexRad is the reflexivity of the cloud cover which is then converted to equivalent rainfall. Ground-based stations are static and can measure the rainfall at any temporal resolution. In contrast, radars rotate every 3 to 5 minutes and hence can capture rainfall information only within this time interval. Ground-based precipitation is a representative of a point in space while radar generated rainfall can capture spatial variability. Quantifying the differences between them can help us understand the risk and tradeoffs that may result when using the datasets which is crucial especially when the rainfall datasets are used in simulating watershed processes that are in turn used to evaluate the consequences of environmental decisions and policies.

Assessing the impacts of climate change is simulated using projected future meteorological data in environmental models. The future meteorological data are commonly generated using General Circulation Models (GCMs) downscaled to a local or regional area. GCMs are series of three-dimensional mathematic models that simulate the climate based on some basic principles of mass, momentum, total energy, and water vapor to predict the climate changes in the future using

the surface hydrology, sea ice, cloudiness and other important climate factors (Grotch et al., 1991). The typical GCMs will have a resolution of between 250 km to 600 km for horizontal space which exceeded most of the size of a small watershed and the time interval is in years. Due to its low temporal and spatial resolution, it is nearly impossible to predict the climate changes in small areas that can capture the local climate regime or to be used in the environmental models focusing on the changes or behavior within a watershed. The process of downscaling the outputs of GCM models directly add another layer of uncertainties to the datasets such as the changeable relationship within the time series, and failure to capture most of the changes within the watershed (Wilby et al., 1997). Furthermore, most of the downscaling methods being used in most of the research are statistical and mathematical methods which might ignore the relationship between the hydrological and climate parts (Wilby et al., 1997), which played very important role in the water circulation in the ecosystem. Considering the important role that rainfall datasets play in developing environmental models that are then used to predict and forecast important events and to support the environmental decision and policies, a methodology to prepare them at the local scale based on some parameters which could preserve most information from available rainfall data is crucial and necessary.

CHAPTER 2: OBJECTIVES

Considering the importance of rainfall in evaluating the systemic responses of a watershed, understanding the properties and limitations of the different datasets in representing the areal watershed precipitation is crucial. The main objective of this study was to evaluate the tradeoffs between ground-based and NexRad radar data sets and to generate synthetic rainfall from both datasets to be used in scenario-based investigations. Specifically, this study was aimed to:

(1) Derive a rainfall time series from the Next Generation Radar (NexRad, NOAA WSR-88D Doppler radar) and ground-based weather stations;

(2) Compare NexRad and ground-based rainfall data and quantify the differences between these two rainfall measuring and recording methods in terms of rainfall patterns and the basic statistics of data;

(3) Generate synthetic rainfall data sets from ground-based and NexRad datasets using a rainfall generator model.

Understanding the differences between ground-based measured precipitation and NexRad will enable us to understand the risk and tradeoffs of using a specific data set to simulate hydrologic processes. Furthermore, by being able accurately to represent the temporal and spatial rainfall patterns peculiar to the watershed in question, we would be able to quantify accurately and predict the complex watershed responses (likelihood) within its probable behavioral spectrum under changing scenarios to predict the climate changes in small watershed or ecosystem area.

CHAPTER 3: REVIEW OF LITERATURE

3.1 The importance of high-quality rainfall data in environmental models

The quality of rainfall data referred to the characteristics of the rainfall data including and not limited to the density of the rainfall gauges (measured by the average area of each gauges covering), the reliability and catchments of the rainfall data, and the time interval of the rainfall time series. With the high-quality rainfall data, researchers can operate the analysis within the small watershed like flash flood and drought predictions, run-off models, non-point source pollution control, and land use analysis (Smith et al., 2007; Hansen et al., 1996; Jia et al., 2016; Aleman et al., 2016).

In the environmental models, as one of the most variable factors or input, poor-quality rainfall data would enlarge the spatial-temporal resolution and uncertainty in the results and conclusions. On the other side, high-quality rainfall data could also enable us to capture the variation within the watershed which is sometimes critical. As a very complicated weather situation, rainfall could be affected by many factors including land use, air pressure, cloud thickness and height, wind speed, and elevation. Much research has been conducted by many researchers to evaluate the effects the quality of rainfall data on the ability of environmental models to simulate the different hydrologic processes at or within the watershed scale. Hansen et al., (2001) found that the rainfall data set with low rain gauges density can result in a poor rainfall-runoff model performance while the rainfall data set with fair or good rain gauges density can largely improve the performance. However, the research also pointed that the catchment response dynamics could also have an impact on the performance (Hansen et al., 1996). Research by Duncan et al. indicated that high density of rainfall gauges has a strong and positive relationship with the accuracy of hydrograph

parameters. The accuracy within 5% of the hydrograph parameters could be achieved by the combination of the rainfall gauges data and the radar data which could provide the coverage in space (Duncan et al., 1993).

In conclusion, to get a reliable and high-quality environmental model result, it is of vital importance to acquire the rainfall data with both high resolution in space (4km*4km or less) and time (daily or less) (Zhong & Yang, 2015).

3.2 The NexRad rainfall dataset

The Next Generation Weather Radar (NexRad) program was developed in Oklahoma in 1988 with the aim of providing high accuracy and resolution weather for the models, weather forecasting, natural disaster warning, and other uses (Ulbrich & Lee, 1999). The radar system is called "WSR-88D" system which provided the Level-II (Base) data including the original reflectively data, wind velocity data, and Level-III products generated from the original data like rainfall, storm structure and other products. The NexRad project which covered large areas and provided various radar products can largely assist users in real-time weather monitoring and short-term or long-term weather forecast (Crum & Alberty, 1993), the detection of flood caused by the intense rainfall events (Seo et al., 2015), and calibration of the new rainfall detection method.

NexRad rainfall data was generated using the Precipitation Processing System (PPS) algorithm using the "Z-R" law in which the "Z" represents the radar reflectively factor and "R" represents the rainfall (Fournier, 1999). The NexRad data, from 1988, can be ordered and downloaded from the website (http://www.roc.noaa.gov/WSR88D/) and viewed or processed by the software downloaded from the site (https://www.ncdc.noaa.gov/data-access/radar-data/radar-display-tools). However, just like other measuring rainfall methods, the NexRad has its error issues. From the time NexRad was put into use, many types of research have been conducted to

evaluate the reliability and accuracy of the NexRad rainfall data product. In 1996, Hunter pointed out that error could happen due to many possible reasons like change in temperature and the reflective factor from different wavelength wave (Hunter, 1996). In 1999, Ulbrich & Lee conducted research about the reliability and accuracy of the "Z-R" law. The outcome showed that the equation above will get 33% more rainfall data during storm rain while 25% less rainfall during stratiform rain (Ulbrich & Lee, 1999). In 2011 and 2015, Seo et al. found that radar calibration errors are the main reason of rainfall error. However, all the results showed that the products from NexRad can provide the rainfall data for small-scale area environment modeling within an acceptable range when calibrated properly by other rainfall data like ground-based data measured and recorded by rainfall gauges (Seo et al., 2011&2015).

However, few studies have been focused on quantifying the difference between the NexRad N1P (Level-III) product and the ground-based rainfall data acquired from the Tipping-Bucket rainfall gauges in the small-scale of a watershed in a long time scale. The existing lag caused large uncertainty and error in the hydrological models when using the other source of data as alternative data source.

3.3 The ground-based rainfall dataset

The stations are equipped with Tipping-Bucket (TB) rainfall gauge which record data with a 5-minute interval. Much research has been operated by researchers about the errors and the source of the TB rainfall gauges measurement because of its widely usage region and temporal and spatial reliability. In 2003, Ciach found that the relative standard error of the TB rainfall gauge measurements in 5-min timescale was less than 5%. However, he also pointed out that this can be exacerbated by extreme rainfall events (e.g., short duration high-intensity rainfall or low-intensity long duration). Research operated by Habib et al. in 2001 found that the increase of the time scale

of the sampling could largely decrease the error of the TB rainfall gauge measurement. The error is not significant when the time scale is over 15 minutes while the error of 1-minute time-scale rainfall is significant especially at low rain rates (Habib et al., 2001). In conclusion, both types of research showed that the sampling mechanism and its inability in capturing the changes in small time-scale time series are the main sources of the error. In 2016, Hoffmann et al. pointed out that the height of the TB will also affect the outcome of the amount of rainfall recorded by the measurement system. The TB installed 1 m height will record 12.7% less rainfall than the lysimeters installed on the ground in the experiment. What's more, they also found that the TB system used in their study has a tendency to record less all the year while in winter the difference is more than in summer (Hoffmann et al., 2016). Apart from the factors listed above, Fankhauser in 1997 also found that the depth resolution, inaccurate rainfall depth per tip, exposure of the site, the wind and the uncertainty in rainfall distribution are all the factors which might cause errors in the TB rainfall gauges measurement. However, the depth resolution is not as important as the other factors (Fankhauser, 1998). Also, the location distribution of the rainfall gauges also affects the quality of the rainfall data set. Researchers operated by Seed and Austin in 1990 found that the regular network rain gauges could decrease the variable errors while the mean errors remain the same (Seed & Austin, 1990).

In conclusion, the rainfall time series recorded by the TB rainfall gauges could represent the rainfall level of the watershed when we compared with the NexRad time series. However, we should also be aware that the ground-based rainfall time series could also be inaccurate because of the factors that could lead to errors as reviewed.

3.4 Representing rainfall at watershed scale

The resolution of the rainfall data usually changes based on the area of the watershed and the resolution of the model used in the research. For example, the models which predicted the temperature, rainfall and, climate change in the global scale usually using the same rainfall value for one watershed or serial watersheds because the modification in one watershed or serial watersheds would not make a large difference (Hijmans, 2005). Most of the watershed with the area smaller than the cell in most of the models would also be regarded as one value for the whole watershed and ecosystem around like in GCMs. The mean value of the watershed can represent the general rainfall level and condition of the watershed in some extend and significantly simplify the environmental models when the model area is large and complex, however, this is not always useful and reasonable for the environmental models for small watersheds like regional runoff and non-point source pollution management (Smith et al., 2007; Hansen et al., 1996; Jia et al., 2016; Aleman et al., 2016). Furthermore, the rainfall data covering large area usually contained considerable error and uncertainty which would make the final result farther away from the real result.

To analyze the effect of the rainfall on other environmental factors within the watershed, it is of vital importance for us to capture the variation in the watershed which made a difference. For example, in the research from Zhong and Yang (2015), the distribution of the rainfall event in both spatial and temporal scale is analyzed in a relatively high resolution (4km*4km) which meant that the rainfall data set should be in a raster rainfall data map instead of one value for the entire watershed. The raster rainfall map and shapefiles rainfall map, which is both common source of rainfall data in two-dimension or multiple-dimension environmental models, are usually chosen to represent the rainfall variation within the watershed. The shapefiles is also the standard file type

which is used by the USGS data gateway as the source GIS data. In the following study, all the rainfall data maps are prepared in the raster and shapefiles.

In conclusion, it is of vital importance for us to generate the rainfall time series map for the small watershed with the high resolution to achieve the objective of quantifying the difference between the NexRad and the ground-based data and analyze and generate rainfall pattern for the small watershed shortly.

3.5 Rainfall patterns definition and generator

The rainfall pattern referred to the rainfall characteristics including the sum of rainfall in a certain interval of time, rainfall events probability distribution function, relationships among the mean, maximum, and standard deviation, and conditional probabilities of wet days (the day with rainfall events) and dry days (the days without rainfall events). As one of the factors that can represent general climate changes and local climate situation, many researchers operated research about rainfall pattern in different spatial and temporal scales. In 2015, Liuzzo et al. analyzed the rainfall pattern in both spatial and temporal trend in Sicily from 1921 to 2012 and found that the total annual rainfall increased from 1981 to 2012 which partial proved that the climate change would increase the rainfall in the short term of time. However, the research also found that the general trend of the rainfall is decreasing from 1921 to 2012 which should introduce further study (Liuzzo et al., 2015). In the same year, Shi et al. also analyzed the rainfall pattern in the Southwest China which found that the climate would result in more chance of drought and flood in a different area in the future (Shi et al., 2015). Furthermore, rainfall pattern, especially the long term rainfall pattern, is also subjective to the changes of land use, elevation, planting-covering-percentage, and other environmental factors.

As one of the important inputs of most environmental models, the rainfall time series is critical in predicting climate change, planning agricultural activities, and other hydrological related activities. However, little research was focused on the generation of the rainfall time series data based on the rainfall pattern in small watershed with a unique ecosystem based on the NexRad rainfall data nor the ground-based rainfall data with the high rainfall gauges the density of 31.25 km^2 per station (Hansen et al., 1996).

CHAPTER 4: METHODOLOGY

4.1 Study area

The Little Washita Experimental Watershed (LWREW) is located in south-central Oklahoma. It drains an approximate area of 625 km^2 (Figure 4.1). The elevation of the watershed is between 300m and 500m above mean sea level. The soil textures are from fine sand to silty loam while the exposed bedrock is sandstone dominant and Permian age sedimentary rocks. Average annual precipitation in the watershed is 760 mm while the mean temperature is 16°C. In this area, the land use includes range, pasture, forest, cropland, oil wasteland, quarries, urban/highways, and water (Elliott, R. L. et al., 1993).

The LWREW has been used as an experimental watershed since 1936 for soil erosion control studies. In 1961, the U.S. Department of Agriculture (USDA) Agricultural Research Service (ARS) set up monitoring stations (36 stations) in the watershed to collect rainfall data for flood control and monitoring studies (Elliott, R. L., et al., 1993) (Figure 4.1). In 1978, the USDA and the U.S. Environmental Protection Agency (EPA) started to use this experimental watershed as one of seven selected watersheds for national projects to study the effects of land conservation in water quality (Elliott, R. L. et al., 1993). Even with the reduced field measurements during 1985 to 1992, the stations from 1994 started to measure other environmental variables such as air and soil temperatures, relative humidity, and solar radiation apart from rainfall (Elliott, R. L., et al., 1993).

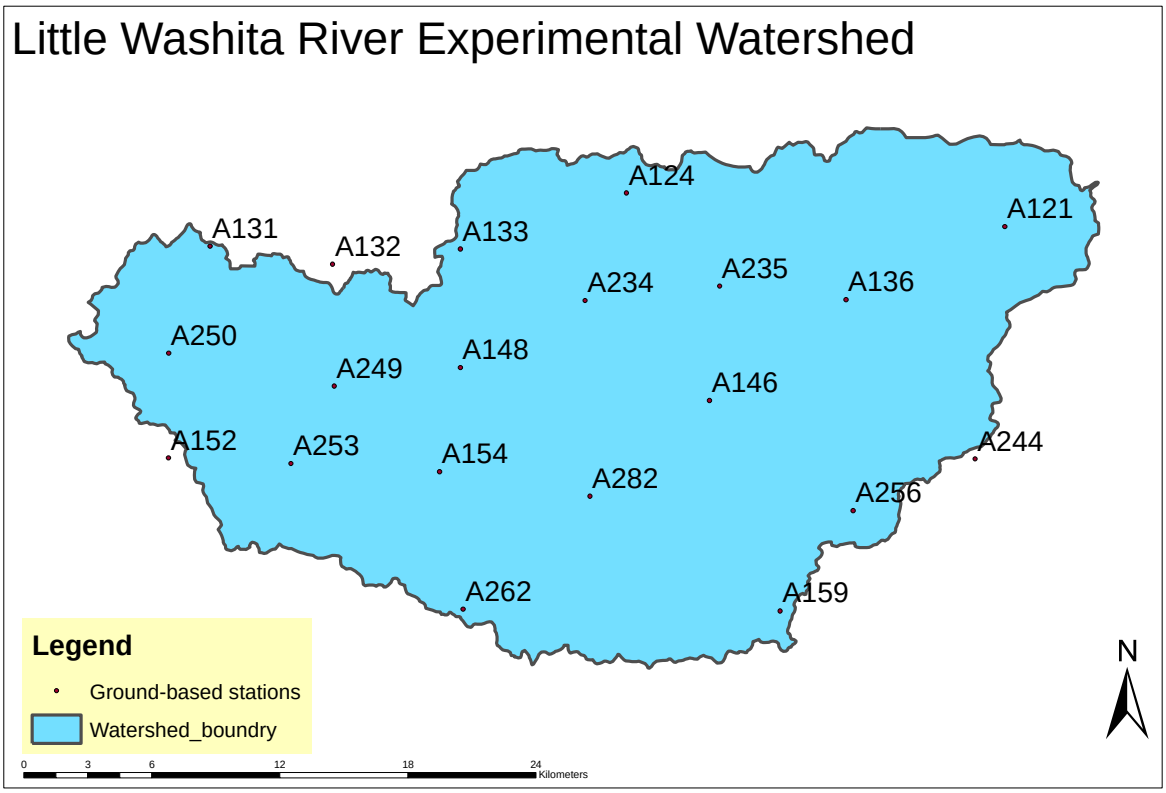


Figure 4.1. The Little Washita Experimental Watershed (LWREW) and ground-based stations

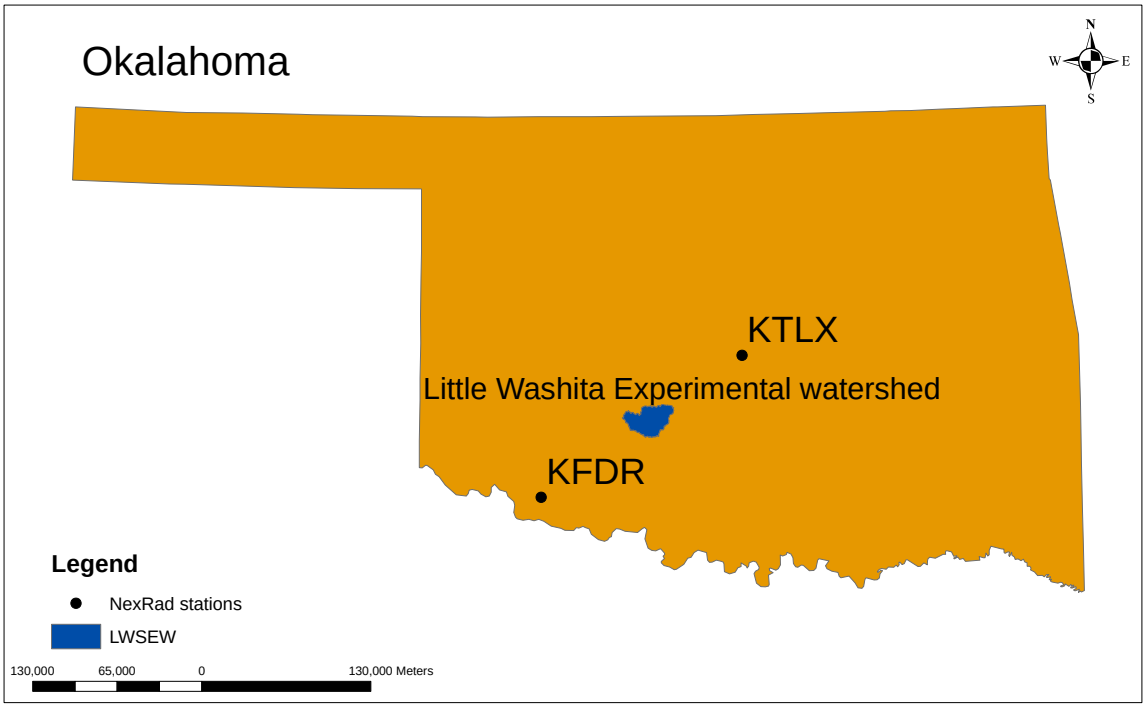


Figure 4.2. The locations of the NexRad stations and the watershed in Oklahoma

4.2 Precipitation stations

4.2.1 Ground-based stations

The data from the ground-based stations were collected from the Micronet and Mesonet networks which contained 20 ground-based stations (Figure 4.1). The stations are equipped with Tipping-Bucket (TB) rainfall gauge which records data at a 5-minute interval. The period of record of the datasets is from 1997 to 2014.

4.2.2 NexRad stations

The NexRad data used in this study was the Level-III radar product: “One-Hour Precipitation (N1P/78)”. This particular radar product represents the one-hour precipitation total ending at the volume scan time on a 1.1-nm x 1-degree grid (NOAA, 2016).

Two radars cover the study area, stations KTLX and KFDR (Locations seen in Appendix 1 and Figure 4.1). The period of record of the data sets is from 1994 to 2014 in station “LTLX” and 2002 to 2014 in station “KFDR”. Raw data were ordered and downloaded from the website in a 5-minute time scale. Using the NOAA's Weather and Climate Toolkit (WCT), the maps were exported to ASCII format.

4.3 Derivation of rainfall time series

4.3.1 NexRad data

Extracting rainfall data from NexRad was performed following three main steps (Figure 4.2). Raw radar data were obtained from the National Oceanic and Atmospheric Administration (NOAA). Algorithms and scripts were prepared in Python, ArcGIS, batch files, and Spellmap (Guzman et al., 2013) to facilitate data processing and management. Station-based rainfall was then derived at different time intervals such as 5-minute, one-hour and one-day following the steps described in Figure 4.3.

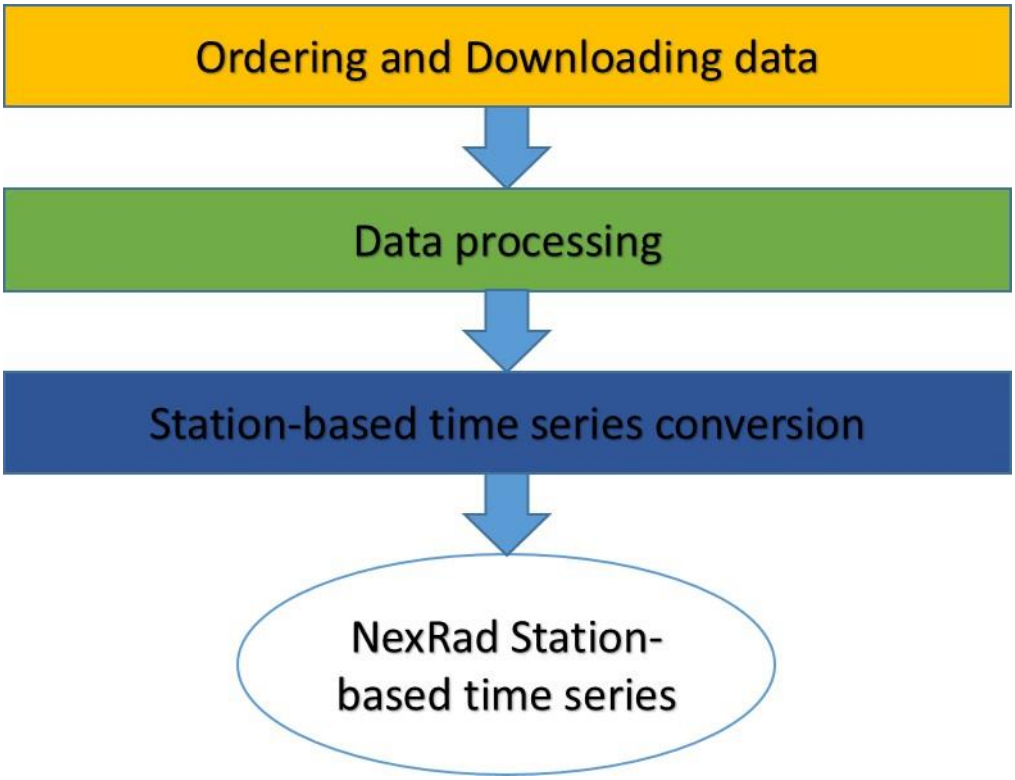


Figure 4.3. The working flow of rainfall time series data preparation.

Ordering and downloading of data: NexRad data were ordered and downloaded from NOAA National Centers for Environmental Information (http://www.ncdc.noaa.gov/has/HAS.FileAppRouter?datasetname=7000&subqueryby=STATION&applname=&outdest=FILE). The watershed was first located on the radar map and the radar stations that cover the area were identified (i.e., stations KTLX, KFDR). The data were then ordered corresponding to the period of interest (Figure 4.4). Ordered data were then downloaded from the received FTP link. The downloaded files were in a unique file format that could only be viewed using the NOAA toolkit. Each file represented the sum of rainfall in the past hour for the covered area. There is a 5-minute interval between files. One file in the dataset represented the sum of rainfall in the past one hour from the time indicated in the file for the radar covering the area with the five-minute interval between files.

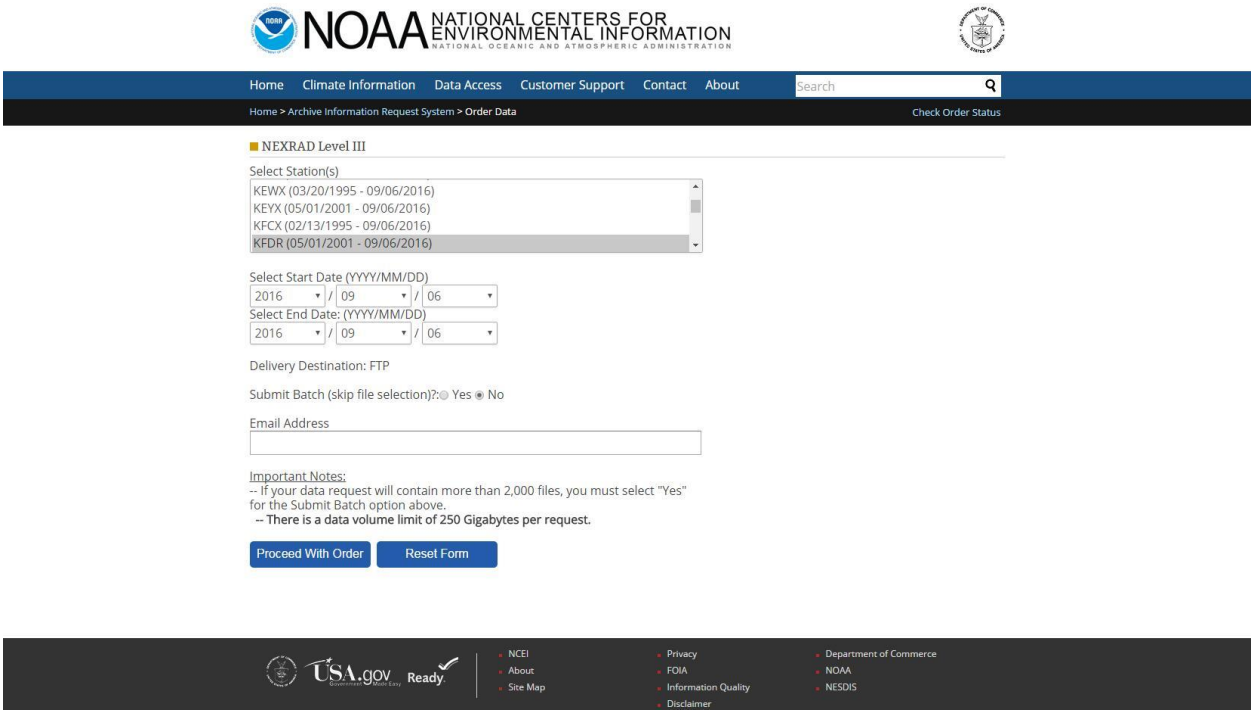


Figure 4.4. For the LWREW, data were ordered and downloaded from NOAA National Centers for Environmental Information.

Data processing: The downloaded data (e.g. Figure 4.5) were first converted to ASCI format to facilitate further processing. The NOAA toolkit (http://www.ncdc.noaa.gov/wct/install.php) was used to export the N1P data to ASCI file using the batch file processing capabilities of the toolkit. The exported data were now in the form of gridded maps with the spatial resolution of 4 km and temporal resolution of 5 minutes. There were more than 6000 files in one month in the time period of 12 years (more than 864000 files) from radar station KFDR and 20 years (more than 1440000 files) from radar station KTLX with the total size of more than 15 GB in memory.

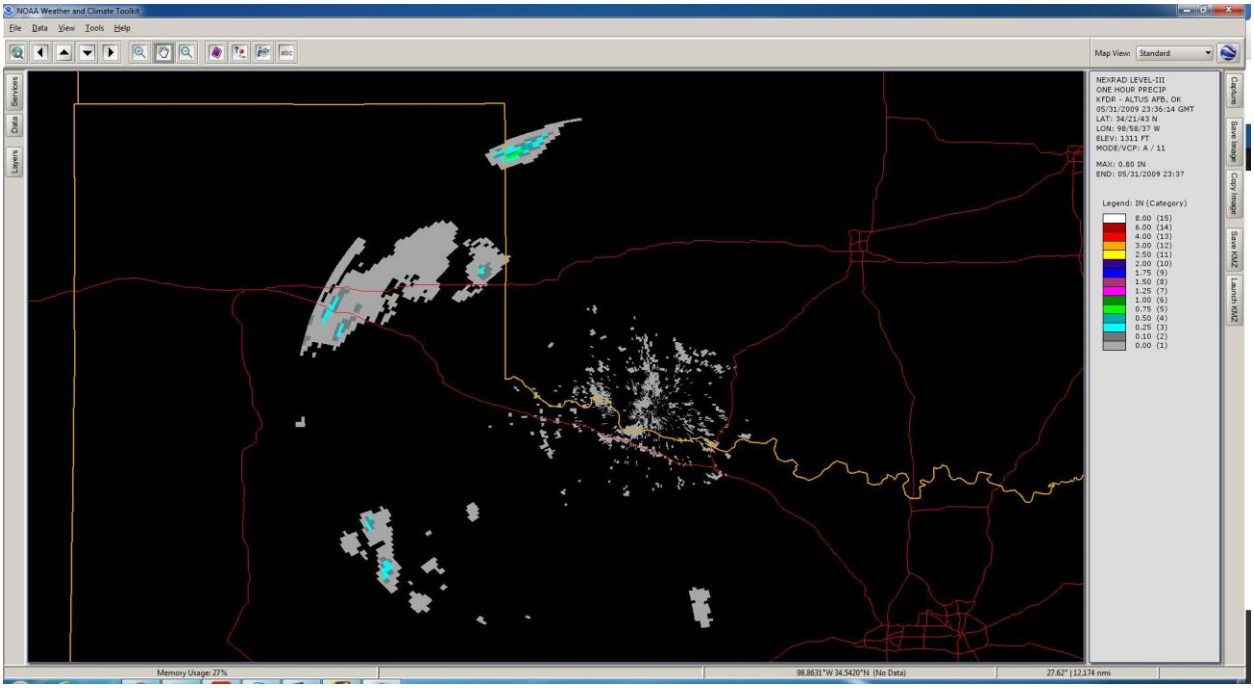


Figure 4.5. The NOAA toolkit was used to convert the N1P data to gridded maps (ASCI format) containing the rainfall depth (mm) at 4 km 4 km resolution.

Gridded rainfall data were then projected from "GCS_North_American_1983" to "Universal Transverse Mercator (UTM) 14N" using ArcGIS. Scripts were prepared in Python to facilitate the batch projection of the maps. Once projected, the maps were clipped to the watershed boundaries using Spellmap (Guzman et al., 2013). Finally, the rainfall data in the clipped maps were converted to a time series format at 5-minute time step (Figure 4.6).

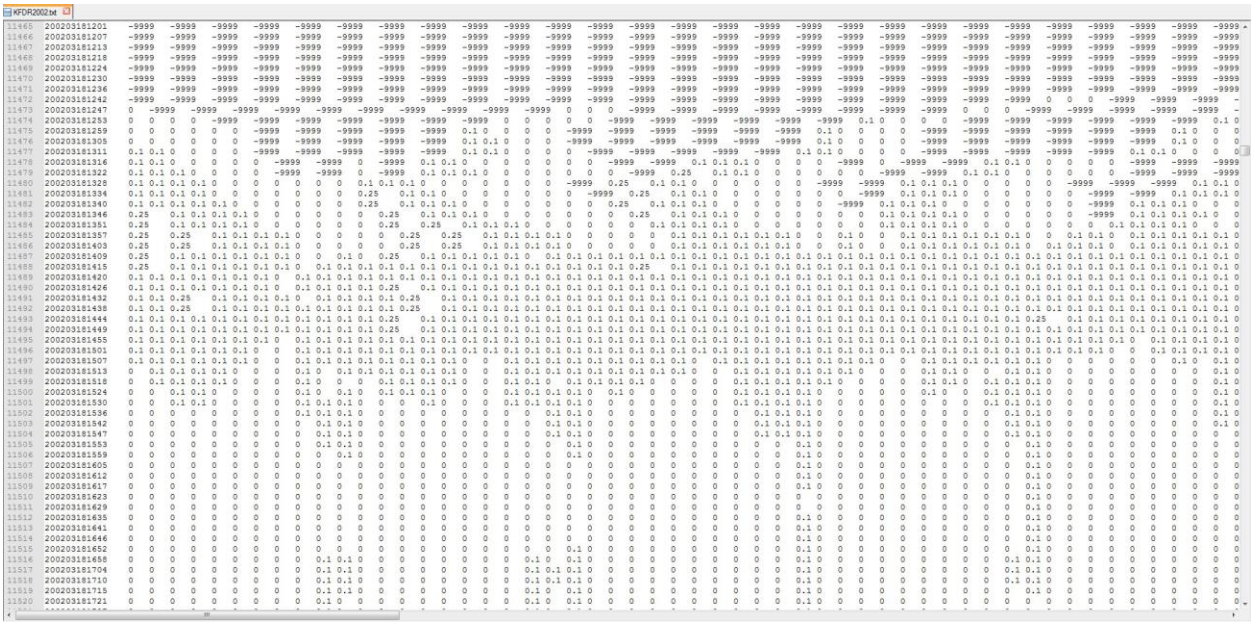


Figure 4.6. An example of the final product of data processing. The first column is the date (in string format of) and the succeeding columns are rainfall measurements for each 4 km x 4 km cell. The corresponding number for the cell in the TXT file could be found in Figure 4.7.

Station-based time series conversion: The cell-based 5-minute NexRad time series were converted to daily station-based rainfall time-series to facilitate the comparison of the NexRad and ground-based data. The location of the cells relative to the ground-based stations was used as the basis for the conversion. For example, for cells close to the ground-based stations (Figure 4.7) were assigned to that stations. If two or more cells were located at the same distance to the station, the average rainfall from these cells was used for that station. The process of deriving the station-based NexRad data is summarized as follows:

(1) For stations close to the middle of the cells, the rainfall data of these stations is the rainfall of these cells;

(2) For stations on the border of two cells, the rainfall data of these stations is the average rainfall of these cells when both cells are not "-9999" (i.e., missing); if one of the cells is "-9999", the data in this station is that of the other cell;

For the stations in the middle of more than two stations, the rainfall data of these stations is the average of rainfall data of the cells which are not "-9999".

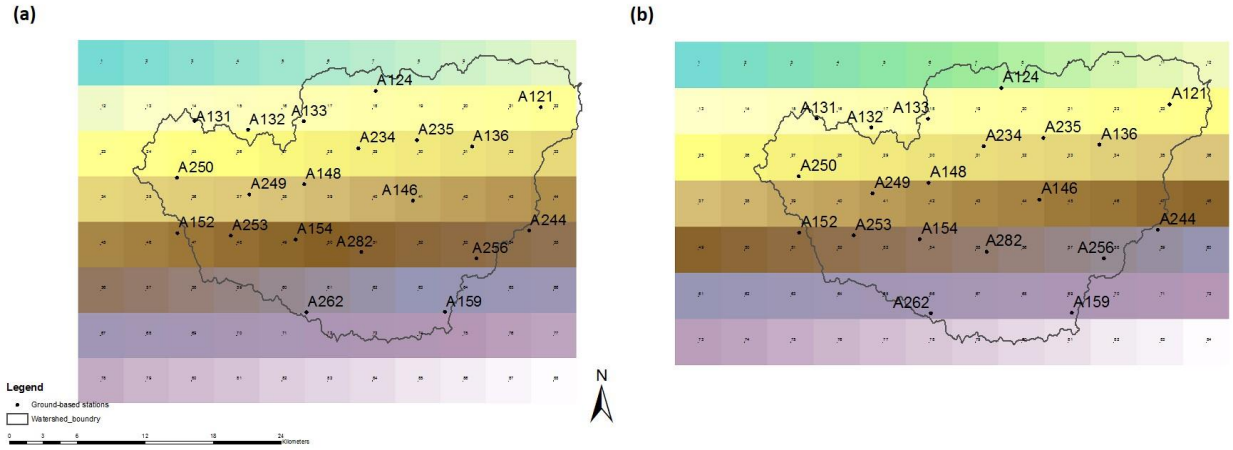


Figure 4.7. The ground-based stations (dots) and radar cells map (colored grids) for (a) KFDR station and (b) KTLX station

The table containing the cell-station relationship is shown in Appendix 2. After the temporal and spatial processing, the final format was a 4 x 4 km grid with each grid containing a daily rainfall time series. The NexRad rainfall data was also processed to obtain the *total daily rainfall* which is the total rainfall in one day in this watershed. It was computed as follows:

$$Daily\ total\ rainfall = \sum Rainfall_{Hourly} \qquad \text{(Equation 4.1)}$$

where $Rainfall_{Hourly}$ is the hourly rainfall obtained from the 20 stations in the watershed. The monthly daily time series were also obtained by organizing all the days in each month for the entire period of record (e.g. all the days in January from 1997-2012).

Both the ground-based and NexRad daily entire time series were further processed to obtain the areal monthly rainfall. The areal monthly rainfall is one of the most important factors that we have to consider in comparing the NexRad and ground-based data because it reflects the seasonality of the time series. The areal monthly rainfall was calculated as follows:

$$R_M = D * \bar{R} \qquad \text{(Equation 4.2)}$$

where R_M is the areal monthly rainfall, D is the number of days in the particular month, $\bar{R}$ is the average daily rainfall of all the days in this month when taking the "no-data" ('-9999' in the dataset) as 0. Table 4.1 summarizes the time series used in this study.

Table 4.1. Daily and monthly daily total time series rainfall for each station (20 stations) for both the ground-based and NexRad datasets.

TIME SERIES	TIME STEP	PERIOD		
		Ground-based	KTLX	KFDR
DAILY	daily	1997-2012	1994-2014	2002-2014
JANUARY	daily	1997-2012	1995-2014	2002-2014
FEBRUARY	daily	1997-2012	1994-2014	2002-2006
MARCH	daily	1997-2012	1994-2014	2002-2014
APRIL	daily	1997-2012	1994-2014	2002-2014
MAY	daily	1997-2012	1995-2014	2002-2014
JUNE	daily	1997-2012	1994-2014	2002-2014
JULY	daily	1997-2012	1994-2014	2002-2014
AUGUST	daily	1997-2012	1994-2014	2002-2014
SEPTEMBER	daily	1997-2012	1995-2014	2002-2014
OCTOBER	daily	1997-2012	1995-2014	2002-2014
NOVEMBER	daily	1997-2012	1995-2014	2002-2014
DECEMBER	daily	1997-2012	1995-2014	2002-2014
MONTHLY AVERAGE	monthly	1997-2012	1994-2014	2002-2014

4.4 Comparison between ground-based and NexRad data

As one of the methods to record rainfall, it is of vital importance to compare the two approaches when we plan to use the other kind of data if the necessary data is not available. The comparison could quantify the difference and provide the possible error range between each type of data. The comparison of the ground-based and NexRad data was conducted using the following metrics:

Basic statistics of data: The primary statistics of the ground-based and NexRad datasets were compared using the one-way ANOVA test for the areal monthly rainfall and the two-sample

Kolmogorov-Smirnov test for the daily time series for each month. The former compares the mean of the two datasets while the latter compares their probability distribution function (pdf).

4.4.1 Basic statistics of data

The objective of this comparison is to determine how differently, the rainfall data in this watershed area is represented from different sources. The areal monthly rainfall from the ground-based data of the watershed is calculated using all the rainfall data recorded in the watershed instead of using only the rainfall larger than 2.54 mm in one day, which is the same as the NexRad data, to get all rainfall data from the ground-based stations.

The NexRad data of the watershed was calculated by averaging the rainfall data from the two stations, KFDR, and KTLX, to get the general relationship between the NexRad data and ground-based data.

Apart from the areal monthly rainfall of the watershed, double mass analysis based on the areal monthly rainfall from NexRad and ground-based stations was also performed to find the possible relationship between the NexRad data and ground-based data. The double mass plot is the cumulative value in one station (e.g. ground-based) against the cumulative value in another station (e.g. NexRad). The annual residual mass, which is defined as a curve of accumulative departures from the standard value, was constructed using:

$$M = \sum_1^{12} Y_j - \left(\frac{\sum Y}{\sum X}\right) * \sum_1^{12} X_j \qquad \text{(Equation 4.3)}$$

where M is the annual residual mass of ground-based stations; $\sum Y$ is the accumulated annual rainfall of ground-based data; $\sum X$ is the accumulated annual rainfall of the mean of two NexRad stations; Y_j is the mean rainfall of all the ground-based data in month j; X_j is the mean rainfall for month j of all the NexRad data in month j.

4.4.2 Probability distribution functions (pdfs)

The pdf of the ground-based and NexRad datasets was compared to determine if they came from the same distribution. If the two methods of measuring rainfall data have the same pdf, then we can conclude that they are sampled from the same rainfall population in this watershed area.

To test whether the two pdfs are the same, a two-sample Kolmogorov-Smirnov test was applied to the PDFs. The two-sample Kolmogorov-Smirnov test is one of the most widely used tests to analyze whether or not two pdfs come from the same distribution. The test statistic used in Kolmogorov-Smirnov test is given by:

$$D_n = sup_x|F_{n1}(x) - F_{n2}(x)| \qquad \text{(Equation 4.4)}$$

where sup_x is the supremum function; $F_{n1}(x)$ and $F_{n2}(x)$ are the empirical distribution functions of the first and the second sample. Dn is the cumulative difference of the two PDFs.

However, when testing the NexRad data in two-sample Kolmogorov-Smirnov test, the probability of "no-data" was found to be relatively high which meant the "no-data" cannot be simply deleted. To rectify this, "no data" days in NexRad were compared with ground-based data to determine whether they are in fact missing or "0."

From the probabilities, we can calculate that the 95% confidence interval for KFDR is (0.887, 0.909) and KTLX is (0.893, 0.913) where P is the probability of the days when no rainfall happened in the days that were recorded as 'no data.' As a result, we assumed that 90% of the "no-data" days we have in NexRad are days without rainfall while the other 10% was replaced by the average value of the rainfall in that area.

4.5 Rainfall pattern definition and regeneration

Rainfall, being the most important driver of runoff, sediment transport, and other environmental processes have to be adequately represented in time and space to properly simulate

the response of the hydrologic system. However, rainfall data with high spatial and temporal resolution is not always available. As a result, the impacts of rainfall variability, especially in small watersheds, is not taken into consideration when evaluating the impacts of environmental stressors (e.g., climate change and land use change) on the systemic responses. Downscaling rainfall datasets to account for both the temporal and spatial variability in a small watershed is therefore an important task in environmental modeling. This study defined the rainfall patterns in the study area based on measured rainfall and then synthetically generate probable rainfall datasets (Figure 4.8).

Using the observed daily rainfall from three datasets (ground-based and 2 NexRad stations), the average, maximum, and standard deviation of all the cells (stations) were computed (Figure 4.8). For the little Washita watershed, rainfall pattern described the characteristics of rainfall in time and space. In this study, rainfall patterns were defined by:

(1) The sequence of days with rain and without rain, (i.e., the conditional probability of days with rain or without rain);

(2) Daily rainfall probability distribution function; and

(3) The relationship boundaries between areal maximum, mean, and standard deviation in log-scale figures;

(4) Auto-correlation of daily time series.

Once the rainfall pattern was established, the rainfall generator, "Zeus", was used to generate synthetic rainfall time series at a daily time step.

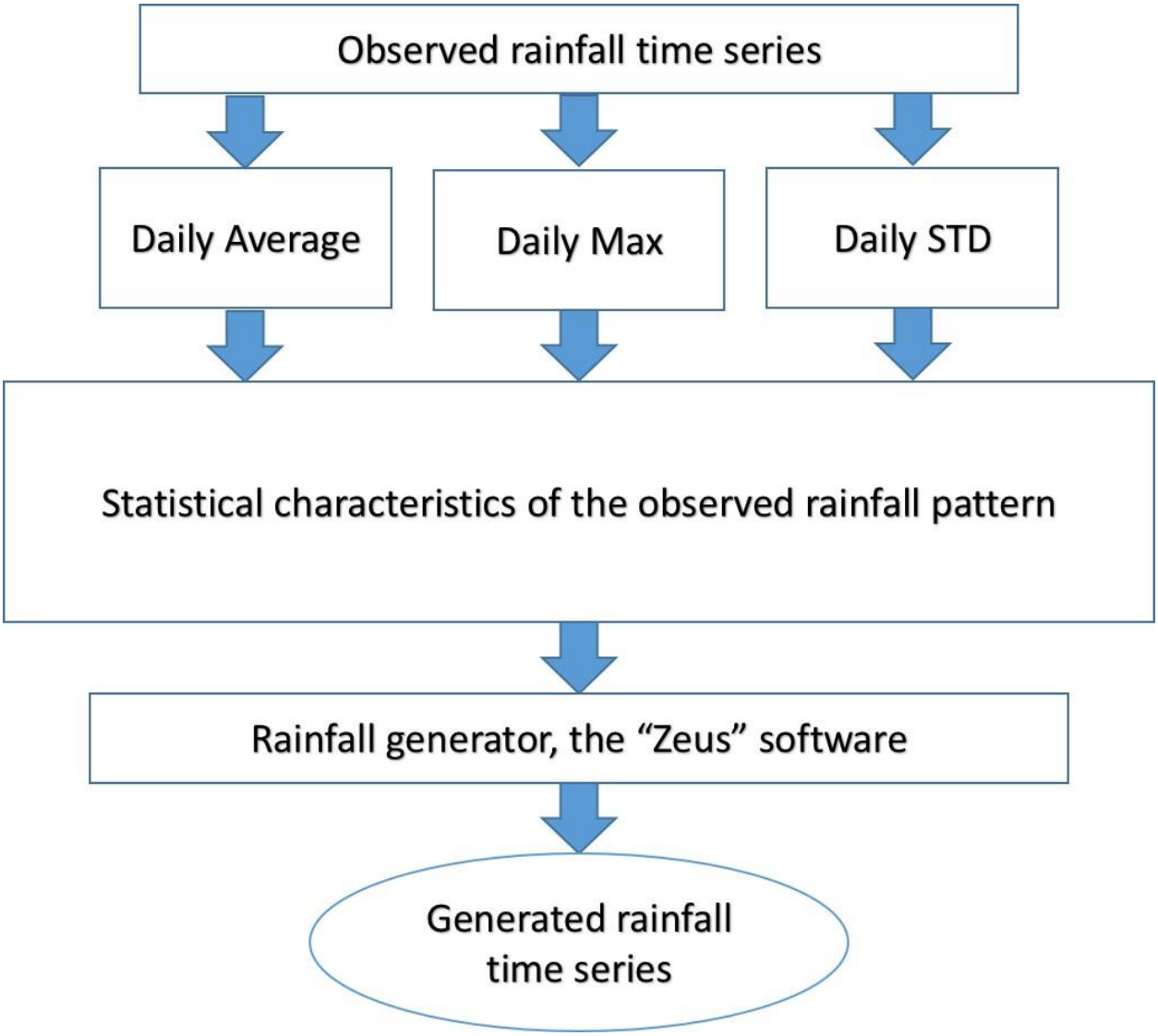


Figure 4.8. The working flow of the analysis and regeneration statistical characteristics of the observed rainfall pattern for the Little Washita watershed

4.5.1 Defining rainfall pattern in LWREW

4.5.1.1 Daily conditional probabilities of wet days and dry days

The occurrence of rain (or no rain) was described in terms of the conditional probabilities of wet and dry days in a given month. In general, there are four conditional possibilities that rain (or no rain) can occur in a given month and are summarized in Table 4.2. The computation of the conditional probabilities is further illustrated in Figure 4.9. For example, P_RR is the probability

that today is a rainy day when yesterday was also a rainy day and P_RN is the probability that today was a no-rain day when yesterday was a rainy day. The conditional probabilities also satisfied the following equations:

$$P\_R+P\_N=1 \quad \text{(Equation 4.5)}$$

$$P\_RR+P\_RN=1 \quad \text{(Equation 4.6)}$$

$$P\_NR+P\_NN=1 \quad \text{(Equation 4.7)}$$

Table 4.2. Computations of the conditional probabilities describing the rainfall pattern in LWREW.

PROBABILITY	DESCRIPTION	CALCULATION EQUATION
P_R	Probability of rainy days in all days in a particular month	$\frac{D\_R}{D\_T}$
P_N	Probability of dry days in all days in a particular month	$\frac{D\_N}{D\_T}$
P_RR	Probability of a rainy day when last day is rainy in a particular month	$\frac{D\_RR}{D\_R}$
P_RN	Probability of a dry day when last day is rainy in a particular month	$\frac{D\_RN}{D\_R}$
P_NR	Probability of a rainy day when last day is dry in a particular month	$\frac{D\_NR}{D\_N}$
P_NN	Probability of a dry day when last day is dry in a particular month	$\frac{D\_NN}{D\_N}$

where D_R is the number of rainy days in a particular month; D_N is the number of dry days in a particular month; D_T is the total number of days in a particular month; D_RR is the number of rainy days with rainy last day; D_RN is the number of dry days with rainy last day; D_NR is the number of dry days with dry last day; D_NN is the number of rainy days with dry last day.

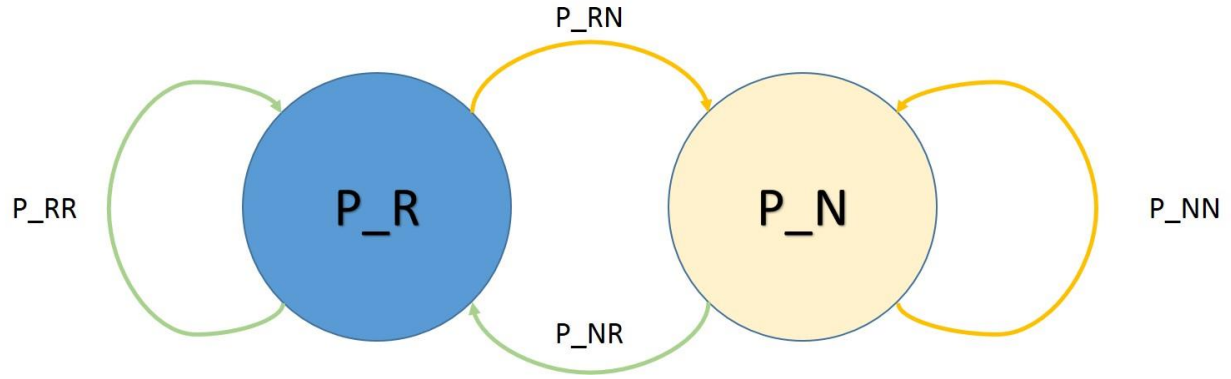


Figure 4.9. The conditional probability relationship figure.

4.5.1.2 Autocorrelation of the area rainfall time series

Autocorrelation is a variable that evaluates the statistical dependence of the time series to itself at different time lags. The autocorrelation is usually a number between -1 and 1 where "1" defines perfect correlation, "0" defines no correlation and "-1" defines perfect anti-correlation.

The autocorrelation was calculated using:

$$R(\tau) = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2} \quad \text{(Equation 4.8)}$$

where $R(\tau)$ is the autocorrelation with the time interval of τ days ($\tau \geq 0$); E is the expectation function; X_t is the value at time t; $X_{t+\tau}$ is the value at time $t + \tau$; and σ^2 is the variance of the time series. In this research, the autocorrelation was computed using Spellmap with the time interval of 0 day to 24 days for the three datasets (KFDR, KTLX, and ground-based) from 2002 to 2012 (Guzman, J. A., et al. 2013).

4.5.1.3 PDFs of the maximum, minimum, standard deviation and pixel rainfall of rainy days

The probability distribution functions (PDF) of all the rainy days from 2002 to 2012 of three rainfall data sets (KTLX, KFDR, and ground-based) were constructed for the maximum and minimum rainfall and the standard deviation. The PDFs were constructed using ten classes defined as follows:

$$Class\ Interval = \frac{Max\ of\ the\ parameters - Min\ of\ the\ parameters}{10} \quad \text{(Equation 4.9)}$$

where the Class interval is the length of the value interval of each class; Max and Min of the parameters are the maximum value and minimum values, respectively, of all the rainy days in a particular month from 2002 to 2012 from each of the three datasets.

After the calculation of the class intervals, all the boundary values for each class were calculated as follows:

$$Low\ interval\ value = Min\ of\ parameters + (n-1) * \ Class\ Interval$$

(Equation 4.10)

$$High\ interval\ value = Min\ of\ parameters + n * \ Class\ Interval$$

(Equation 4.11)

where Low interval value and High interval value are the two border values of class n; Min of parameters is minimum value in the parameters of all the rainy days in particular month from 2002 to 2012 from each of the three data sets; Class interval is the length of the interval calculated in the equation 4.(9).

With the border values for each class, the PDFs were constructed using Python as follows:

$$Probability\ for\ particular\ interval = \frac{The\ days\ with\ value\ in\ the\ particular\ interval}{All\ the\ rainy\ days}$$

(Equation 4.12)

where Probability for a particular interval is the PDF values for this particular interval; The days with value in the particular interval is the number of days with the parameter value larger than the low interval value and less or equal than the high interval value.

4.5.1.4 The boundary conditions of the rainfall events

The relationship between the maximum, mean, standard deviation, and cell daily rainfall data was determined by the log-scale plot of all the rainy days in the time series. The plots used daily maximum cell rainfall in the x-axis while the other statistical characteristics (mean and standard deviation) on the y-axis (e.g., Figure 4.10). Based on the log-scale plots, upper and lower boundaries of the rainfall statistical characteristics were determined (Figure 4.10). These boundaries were used in constraining the rainfall generated by Zeus.

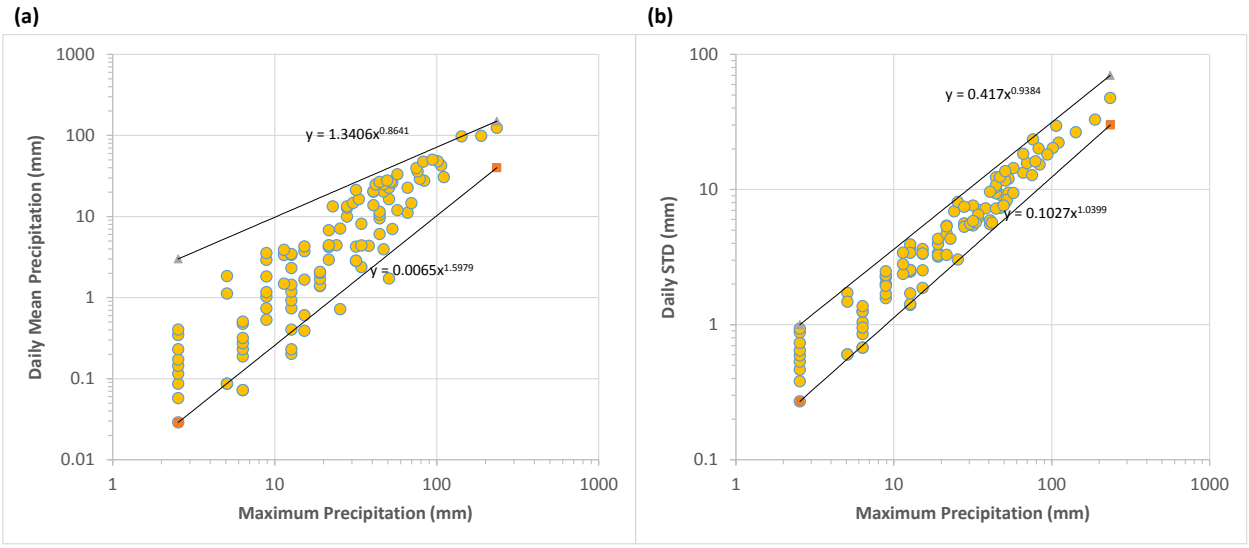


Figure 4.10. Examples of the log-scale figures used to find the possible boundary relationships: (a) The log-scale figure between the daily maximum precipitation and the daily mean of May from 2002 to 2012 from the KFDR data set and the upper and lower boundary; (b) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in May from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

4.5.2 The downscaling and generating rainfall data using "Zeus" software

After establishing the rainfall pattern in the experimental watershed, a rainfall generator software was used to synthetically generate precipitation time series that mimic the patterns of the observed data. An assessment of the generated precipitation was conducted to evaluate how well the generated rainfall represented the observed rainfall in time and space. The rainfall patterns were generated using the "Zeus" software. "Zeus" is a software designed to generate downscaled rainfall data time series using the parameters listed in Table 4.3. By generating multiple time series, the average of all the time series is usually considered as the rainfall time series which is used in analysis.

Inputs to Zeus were contained in an XML (Table 4.4). For each type of rainfall datasets, an XML file was prepared to define the rainfall pattern acquired from each dataset. The KTLX and KFDR datasets were from 2002 to 2012 with a temporal resolution of 5 minutes and a spatial resolution of 4 x 4 km. The ground-based dataset is from 2002 to 2012 with a time interval of 5

minutes. The software generates "n" number of rainfall time series for every station in the watershed. In this study, 1000 daily time series for the 20 ground-based station were generated for every dataset.

Table 4.3. Inputs of the model contained in the XML files

ITEMS NAME	DESCRIPTION
Years(Scenarios)	Years of simulation
Realizations(Scenarios)	Number of time series generated
Stations(Network)	The station locations
Automaton(Events Sequence)	The conditional probabilities of wet and dry days
Residual Serial Correlation(Events Sequence)	Autocorrelation at different lags
Max Precipitation (pdf)	The max PDFs of daily precipitation
Event Mean Precipitation (pdf)	The mean PDFs of daily precipitation
Event STD Precipitation (pdf)	The STD PDFs of daily precipitation
Pixel Precipitation (pdf)	The cells (stations) PDFs of daily precipitation
Max To Mean (Boundaries)	The upper and lower boundaries: daily maximum rainfall vs daily mean rainfall
Max To STD (Boundaries)	The upper and lower boundaries: daily maximum rainfall vs daily standard deviation
Active (Solver)	Generating a pdf using two Gamma functions by the solver to replace the original pdf from the observed data set.
Screening Iterations (Solver)	The times to fit the pdf within the target error.
Target Error (Solver)	The tolerance error in the fitting functions.
Error In Mean (Solver)	The tolerance error of the mean in the fitting functions.
Minimum Value (Solver)	The minimum daily rainfall the rainfall generator generated

4.6 The comparison between observed rainfall data and generated rainfall data

The generated rainfall time series was compared to the observed rainfall time series, which is the typical downscaled datasets used for climate change studies. The comparison included:

(1) The area average monthly rainfall from all 20 stations from observed rainfall from three datasets (KTLX, KFDR and ground-based), and the re-generated rainfall data from 2002 to 2012;

(2) The pdf of the maximums of all the 20 stations from observed rainfall from three datasets (KTLX, KFDR and ground-based), and the re-generated rainfall data from 2002 to 2012 for each month.

The PDF of the maximum area daily rainfall was used as one of the comparison metrics because the maximum of the daily area rainfall was the first generated rainfall data from the "Zeus" software while the mean and standard deviation were generated based on the log-scale relationship inputted from the XML file. As a result, the PDFs of daily area maximum rainfall can represent how close the generated rainfall pattern was to the original rainfall dataset which measured the ability to simulate the rainfall pattern of the "Zeus" software.

The maximums of all the daily rainfall were classified into 10 intervals to determine the PDF of the maximums of the daily rainfall. The nth interval in the PDF of the maximum of the daily rainfall included the days with daily maximum rainfall R_n satisfied the equation 4.13.

$$0.1 * (n - 2) * Max < Rn \leq 0.1 * (n - 1) * Max \qquad \text{(Equation 4.13)}$$

where Max is the maximum of the daily maximum rainfall in this area among all the original data; n is the class number; R_n is the maximum rainfall of that day.

The Probability Difference (D_n) for the nth interval was calculated using the Equation 4.14.

$$Dn = |Fsn - Fon| \qquad \text{(Equation 4.14)}$$

where F_{sn} is the frequency of the days in class n of all the simulated rainfall data; F_{on} is the frequency of the days in class n of all the original rainfall data.

The comparison of the PDF analysis included:

(1) The average of the absolute PDF difference between the daily maximum value of original and simulated rainfall was calculated from all the 3000 simulations to get the general performance of the "Zeus" software in generating rainfall pattern;

(2) The average of the absolute PDF difference between the daily maximum value of original and simulated rainfall was calculated based on three rainfall datasets including two

NexRad datasets (KTLX and KFDR) and ground-based data sets by each month to get the possible difference between the performance for each kind of datasets and months.

In order to understand the differences between the generated and measured rainfall, it is necessary to understand the process that generates rainfall. An observed rainfall time series can be written as Equation 4.15.

$$R(t) = S(t) + L(t) \quad \text{(Equation 4.15)}$$

where R (t) is the amount of rainfall detected and recorded for a particular period of time in a dataset (KTLX, KFDR and ground-based stations in this research). The length of the period can be a year, a month or a day; S (t) is the amount of rainfall for a particular period of time which was defined in the rainfall pattern analysis; L (t) is the amount of rainfall for a particular period of time which was not included using the rainfall pattern analysis we operated. S (t) is represented using the parameters we generated from different factors that made up the rainfall pattern (Table 4.4)

Table 4.4. The names of the rainfall pattern parameters and the affected time scale

NAMES OF THE RAINFALL PATTERN PARAMETERS	PATTERN TIME SCALE
Autocorrelation of the area rainfall time series	Daily
PDFs of the area maximum, minimum, standard deviation and pixel rainfall of all the rainy days	Daily and monthly
The boundary conditions of the rainfall events	Daily and monthly

The unrepresented part of rainfall, L (t), on the other hand can be written as:

$$L(t) = F(t) + Er(t) \quad \text{(Equation 4.16)}$$

where F (t) is the amount of rainfall which has a temporal pattern but cannot be summarized using the parameters listed in Table 4.4; Er (t) is the amount of rainfall caused by random environmental factors. As the rainfall generator only used the parameters listed in Table 4.4, the rainfall time series generated could be written as follows:.

$$G(t) = S(t) + Ep(t) + Eg(t) \quad \text{(Equation 4.17)}$$

where G (t) is the amount of rainfall which was generated by the rainfall generator; Ep (t) is the error resulting from generating method which satisfied$|Ep(t)| \geq 0$; Eg (t) is the random error in the rainfall generating process which satisfied$|Eg(t)| = 0$.

As a result, the difference between the generated rainfall time series and the observed rainfall time series can be written as the Equation 4.18.

$$D(t) = F(t) + Er(t) + Ep(t) + Eg(t) \quad \text{(Equation 4.18)}$$

where D (t) is the difference between the generated rainfall time series and the observed rainfall time series.

As the average of 1000 simulations was used in the comparison, $Eg(t) = 0$ in Equation 4.18, the difference of area monthly average rainfall between the generated rainfall time series and the observed rainfall time series is composed of three components:

(1) F (t), the amount of rainfall which has a pattern but could not be summarized using the parameters listed in Table 4.4. This factor is one of the two main reasons for the monthly differences. The difference caused by this factor could be reduced by adding or changing rainfall pattern parameters;

(2) Er (t), the amount of rainfall caused by random and unexpected rainfall events without any temporal patterns which could be summarized. This factor is the main reason for the extreme rainfall events in the observed time series which would also cause the difference;

Ep (t), the error caused by improper regenerating methods used in the software. This factor is the other main reason for the monthly difference. The difference caused by this factor could be reduced by improving regenerating methods and adding or changing rainfall pattern parameters.

CHAPTER 5: RESULT AND DISCUSSION

5.1 Deriving a rainfall map based on the NexRad rainfall data in the Little Washita watershed

The time series of rainfall derived from the NexRad data set was cell-based covering the watershed with 84 cells for radar station "KTLX" and 88 cells for "KFDR". Figure 5.1 as an example of the average rainfall map (4 x 4 km) for the month of June which can be used as the rainfall data map for GIS application or other models where high resolution and a large area of rainfall data is needed.

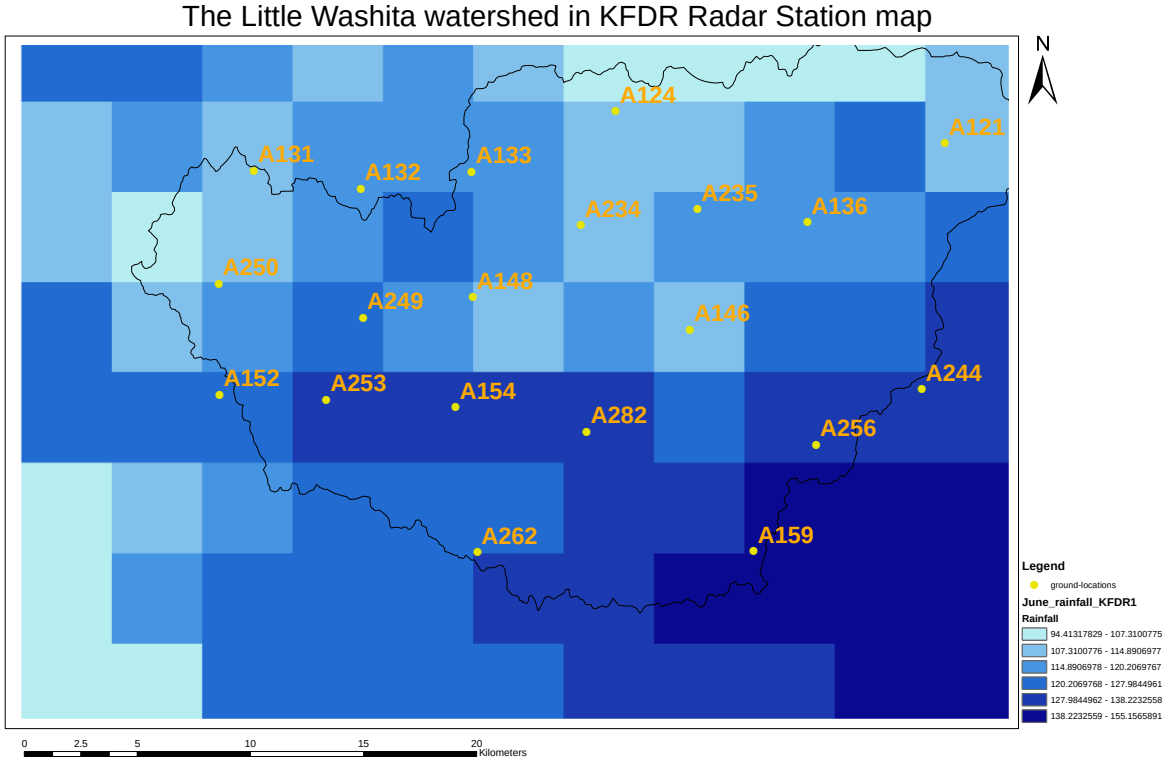


Figure 5.1. The rainfall map of the Little Washita river watershed for the June for KFRD station.

From the figure, we found that the distribution of the monthly rainfall in the watershed was not the same from the two NexRad stations. Using the monthly rainfall sum of 20 stations from

2002 to 2012, the one-way ANOVA test was conducted using R. Based on the R results, we found that apart from the rainfall data in December, the monthly sum of rainfall in other 11 months was all significantly different from each other with the p-value of 0.05. In order to present the rainfall level recorded by the NexRad stations in the term of monthly sum of rainfall, all the NexRad areal monthly rainfall used in the comparison between NexRad and ground-based data is the average of the two stations.

5.2 Comparisons of ground-based and NexRad data

The comparisons of ground-based and NexRad rainfall datasets were conducted to quantify the differences and find a possible relationship between the two datasets. The results of the two comparisons, basic statistics of data and the PDFs, were shown in section 5.2.1 and 5.2.2.

5.2.1 Basic statistics of data

The means of the two NexRad stations were calculated which were further used to calculate the sum and annual cumulative sum of the NexRad dataset. The results of all the monthly mean values of ground-based data and NexRad data are shown in Figure 5.2.

(a)

Month	Ground	NexRad	Difference	Difference percentage
Jan	28.39	15.35	13.04	84.94%
Feb	36.45	14.00	22.45	160.39%
Mar	51.90	43.62	8.28	18.98%
Apr	87.49	74.57	12.92	17.33%
May	90.09	111.07	-20.98	-18.89%
Jun	109.71	103.42	6.30	6.09%
Jul	68.39	49.73	18.66	37.52%
Aug	83.93	71.59	12.33	17.23%
Sep	62.24	36.00	26.24	72.90%
Oct	76.78	54.39	22.39	41.17%
Nov	33.00	22.91	10.08	44.00%
Dec	28.37	15.50	12.87	82.99%

(b)

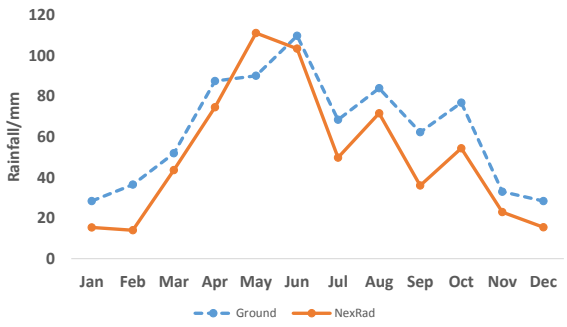


Figure 5.2. (a) Tabulated mean monthly total rainfall of the 20 stations in the NexRad dataset and the ground-based dataset from 2002 to 2012. Highlighted values in green are differences lower than 25% while highlighted in yellow is the month when the ground-based data is lower than the NexRad. (b) Plotted differences between the NexRad dataset and the ground-based dataset. The X-axis is the month and the Y-axis is the amount of rainfall.

The one-way ANOVA tests were performed on the values of areal monthly rainfall from NexRad and ground-based rainfall data. According to the p values from the 12 tests corresponding to the 12 months, the mean values of rainfall data for each month are not the same with the NexRad rainfall data for all the months. The differences between the NexRad and ground-based data was further investigated as a function of the NexRad data (Figure 5.3).

From the regression relationship, we can conclude that the differences decreases as the NexRad rainfall increases. The reasons for this might be due to the following: (1) the detection threshold of NexRad radar is 2.54 mm (The minimum rainfall value in the data set is 0.1 inch) while the ground-based is 0.254 mm. This means that the ground-based data include the rainfall amounts that the NexRad couldn't record in the range of 0.254 to 2.54 mm in a day. This differences is more significant in the months when the rainfall events have a shorter duration and lower intensity, for example in January and February. In the months when the rainfall events have a longer duration and higher intensity (June and August), the undetected rainfall events (0.254 to 2.54 mm) may not matter as much as they do in dry months; (2) NexRad rainfall data was generated from the radar scolding the atmosphere at 360 degrees every 5 minutes. Rainfall events that occur between the 5-minute intervals will not be captured by the radar. Therefore, even if the rainfall intensity exceeds the detection threshold of 2.54 mm, the rainfall will not be detected and recorded with the duration less than 5 minutes.

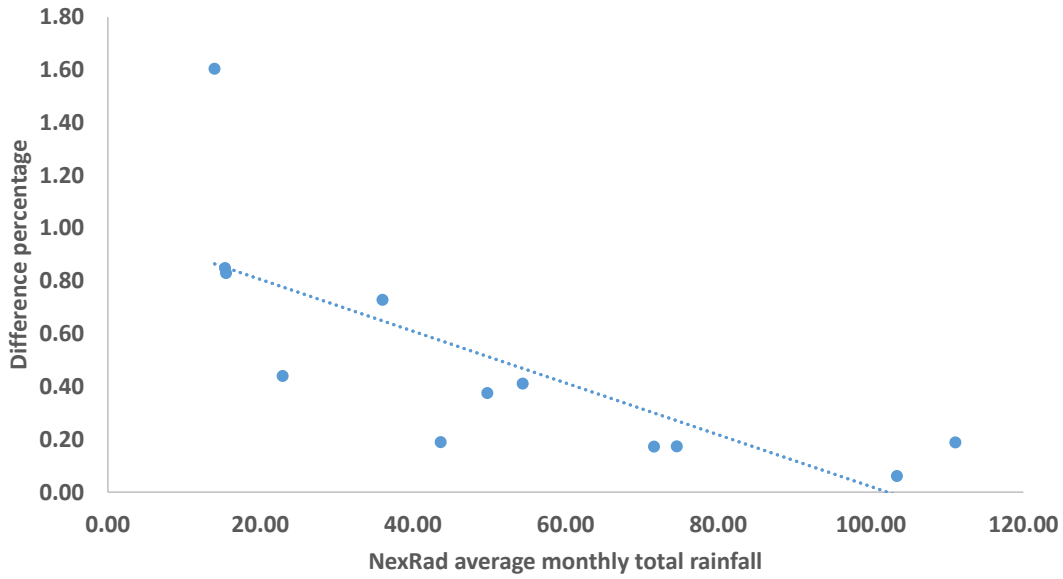


Figure 5.3. The relationship between the differences between ground-based and NexRad rainfall and the average NexRad monthly total rainfall. The X-axis is the mean of NexRad monthly rainfall and the Y-axis is the differences between the mean of NexRad monthly rainfall and mean of ground-based monthly rainfall.

Furthermore, the methods to handle the extreme rainfall events in the NexRad rainfall time series were also further discussed. Based on the rainfall analysis result about the total monthly rainfall between 2002 and 2012, the only month that the rainfall recorded by the NexRad is higher than the ground-based data is May. The reasons for this unusual result may be because there are some days when the amount of rainfall is extremely higher than other days. The days in NexRad with extremely high rainfall is shown in Table 5.1.

Table 5.1. The days with high rainfall amount in May in the NexRad dataset

DATE	KFDR	KTLX	AVERAGE	GROUND-BASED	DIFFERENCES
20120520	128.89	107.36	118.12	44.69	164.31%
20110520	110.64	107.36	109.00	42.80	154.69%

From the table above, we found that these days were all recorded as the relatively high amount of rainfall in all of the three methods. In addition, the average rainfall amount recorded from the two NexRad stations are both over 150% higher than the ground-based rainfall data. Considering the average total amount of rainfall recorded by the NexRad stations was 114.35 mm in May from

2002 to 2012, the days above were modified and changed to the data recorded in the ground-based dataset. Then the monthly total averages rainfall from 2002 to 2012 were recalculated. The results are shown in Figure 5.4.

(a)

Month	Ground	NexRad	Difference	Difference percentage
Jan	28.39	15.35	13.04	84.94%
Feb	36.45	14.00	22.45	160.39%
Mar	51.90	43.62	8.28	18.98%
Apr	87.49	74.57	12.92	17.33%
May	90.09	98.23	-8.15	-8.30%
Jun	109.71	103.42	6.30	6.09%
Jul	68.39	49.73	18.66	37.52%
Aug	83.93	71.59	12.33	17.23%
Sep	62.24	36.00	26.24	72.90%
Oct	76.78	54.39	22.39	41.17%
Nov	33.00	22.91	10.08	44.00%
Dec	28.37	15.50	12.87	82.99%

(b)

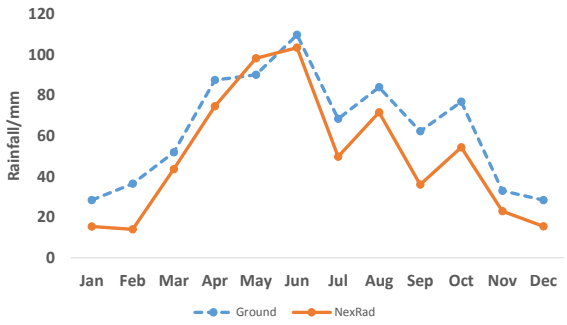


Figure 5.4. (a) Tabulated mean monthly total rainfall of the 20 stations in the modified NexRad dataset and the ground-based dataset from 2002 to 2012. Highlighted values in green are differences lower than 25% while highlighted in yellow is the month when the ground-based data is lower than the NexRad. (b) Plotted differences between the modified NexRad dataset and the ground-based dataset. The X-axis is the month and the Y-axis is the amount of rainfall.

As shown in Figure 5.4, we found that the absolute differences in May changed from 23.29% to 9.05% which meant there was more than 10% differences resulted from the two days with high rainfall listed in Table 3. Further investigation showed that the reason for the extremely high rainfall recorded might result from the "Z-R" relationship equation, which was used in calculating the rainfall based on the reflexivity of the clouds, might not be accurate in the extreme weather conditions.

As a result, when we are dealing with the rainfall data from the NexRad, the extremely large values (Over 100mm in one day) should be handled and referenced by the other source of data such as ground-based data or satellite rainfall data. However, we should also be aware that the ground-based rainfall data only measures and records the rainfall happens to the points where the rainfall gauges locate. This means that the rainfall can happen in the area where the rainfall gauges cannot detect or record.

In conclusion, the extremely large rainfall events recorded in the NexRad data set can be up to 150% more than the real rainfall event amount happened that time in some particular locations. However, the extremely large rainfall events in the NexRad dataset represent that there should be heavy rainfall events happened in that particular region at that particular time which is of vital importance in prediction the possible natural disasters such as flooding or mud-rock flow.

5.2.2 Double mass analysis

The results of the double mass analysis were represented as follow. In Figure 5.2, the NexRad recorded consistently less rainfall in all the months than the ground-based data except May which is usually the wettest month in the watershed. These results are in agreement to what Ulbrich and Lee, (1999) reported that NexRad had a tendency to record higher rainfall during the wet seasons and less in the dry seasons.

The comparisons between the NexRad and Ground-Based data were also performed using double-mass curve and residual mass curve (Figure 5.5).

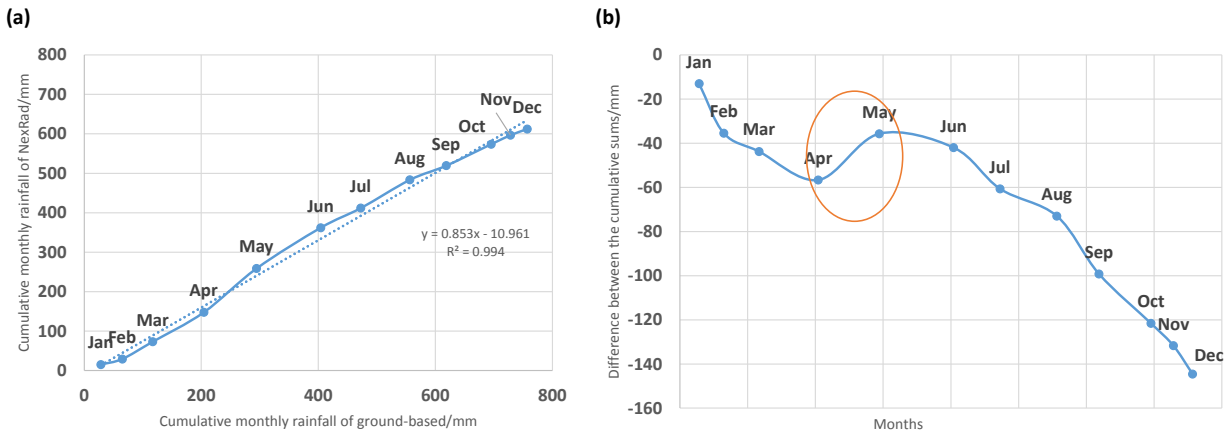


Figure 5.5. (a) Plots of the cumulative mean monthly rainfall between NexRad and ground-based data from 2002 to 2012. The X-axis is the cumulative monthly rainfall of ground-based dataset and the Y-axis is the cumulative monthly rainfall of NexRad dataset. (b) Plots of the Residual Mass Curve between residual of cumulative rainfall and cumulative monthly rainfall of ground-based. The X-axis is the month and the Y-axis is the differences between the cumulative monthly rainfall of NexRad dataset and ground-based dataset.

From the Figure 5.5(a), the tendency line with an R^2=0.994 indicated that the NexRad and the ground-based dataset were strongly correlated which could be used as the prediction model when we only have the NexRad rainfall data within this watershed. The regression equation is shown in Equation 5.1.

$$Y = 0.853X - 10.961 \quad \text{(Equation 5.1)}$$

where Y is the cumulative annual rainfall of the mean of two NexRad stations; X is the cumulative annual rainfall of ground-based dataset.

From the Figure 5.5(b), we found that the residual was decreasing except in May (shown in the orange circles in the figure) which indicated that the ground-based stations recorded more rainfall than the NexRad except in May. This result is also in agreement with the results from Ulbrich and Lee, (1999) that the NexRad has a possibility of recording more rain in the rainy seasons while recording less in the dry seasons. In addition, as we also found that the annual sum of NexRad would become farther away from the ground-based data more as rainfall data from more months were added to the sum which indicated that the NexRad data needs modification if we want to use NexRad as the data for building the prediction model designed for the ground-based dataset.

As the relationship between NexRad and ground-based data changes according to the month, based on the monthly data for each station, the cumulative sums of stations were also calculated to get the particular relationship between the NexRad and Ground-Based data for each month in this watershed.

5.2.3 Probability distribution functions (pdf)

The Kolmogorov-Smirnov tests were conducted both between the two NexRad stations and between each station and the ground-based data. The time series used in the test was the daily total

rainfall. Before the test, the NexRad time series was modified based on the conclusion we drew in the previous section by the following methods: In every 10 days with 'no-data' ('-9999' in the data set), one day is changed to the average daily rainfall in that particular month and the other 9 days are recorded as '0' which indicated no rainfall in that day.

The results of the tests between the two NexRad stations showed that the data from each month in both stations were sampled from the same population. This results proved that different NexRad stations would record the same rainfall events even with different mean values. As a result, if the data from one NexRad station is missing, the data from the other stations covering the same area can be used to replace the missing data.

However, the results of the tests between the two NexRad stations and the ground-based showed that the daily rainfall data was not sampled from the population with the same PDFs in all the months except July. Further analysis showed that the highest difference between both of the NexRad and the ground-based dataset was due to the fact that the ground-based dataset has more days with little rainfall (less than 1/10 of the maximum daily rain in that month) than the NexRad data. These results conclude that the NexRad would record some of the not-intense rainfall as '0'.

The possible reasons for the differences in the PDFs in the area of not-intense rainfall are:

(1) About 90% of the rainy days are the days with relatively little rainfall and lasting in a short time. These rainfall events as above are not easy to be detected and recorded by the NexRad because the time interval between each could of a certain area is 5 minutes which could be longer than the rainfall events lasting time. As a result, the NexRad will record the rainfall events as no rain while the ground-station will manage to record the rain;

(2) The NexRad uses the "Z~R" relationship to calculate the possible rainfall in the certain area. As a result, the reflexivity of the not-intense rain might not be recorded by the

radar because it might not reach the detectable level to be considered related with possible rainfall;

(3) The minimum value of the rainfall recorded by the NexRad is 2.54 mm which means that the rainfall events less than that will be recorded as '0'. This may cause that the number of rainfall events from the NexRad data set is much less than the number of rainfall events from the ground-based data set;

(4) The ground-based rainfall data using the rainfall gauges recorded the rainfall data in the particular points while the NexRad recorded the average rainfall in each resolution cell. The differences in the measuring targets might also lead to this differences.

In conclusion, the NexRad could detect and record the days with relatively intense rainfall more efficiently and closer to ground-based dataset than the days with little rainfall (less than 1/10 of the maximum daily rain in that month).

5.3 Simulated rainfall patterns

Using the "Zeus" software and the rainfall patterns defined from the KTLX, KFDR, and ground-based datasets, 1000 simulations were performed to evaluate the ability of the software to generate rainfall time series. The results of the comparisons between the observed dataset and simulated dataset were conducted using the following method:

5.3.1 Analysis of monthly average rainfall data

The results of each of the 1000 rainfall time series simulations based on the rainfall patterns defined from the KTLX, KFDR, and ground-based datasets are shown in Figure 5.6 to 5.8. All the months with a differences from the measured rainfall of is less than 25% are highlighted in green. The blue lines in the figures are the mean of the 1000 simulations while the orange lines are from the observed dataset.

(a)

Month	Simulated ground	Original ground	Difference	Difference percentage
Jan	10.16	28.39	18.23	64.21%
Feb	21.7	36.45	14.75	40.47%
Mar	19.57	51.9	32.33	62.29%
Apr	26.43	87.49	61.06	69.79%
May	71.03	90.09	19.06	21.16%
Jun	103.56	109.71	6.15	5.61%
Jul	51.86	68.39	16.53	24.17%
Aug	75.4	83.93	8.53	10.16%
Sep	35.44	62.24	26.8	43.06%
Oct	44.45	76.78	32.33	42.11%
Nov	24.32	33	8.68	26.30%
Dec	12.61	28.37	15.76	55.55%

(b)

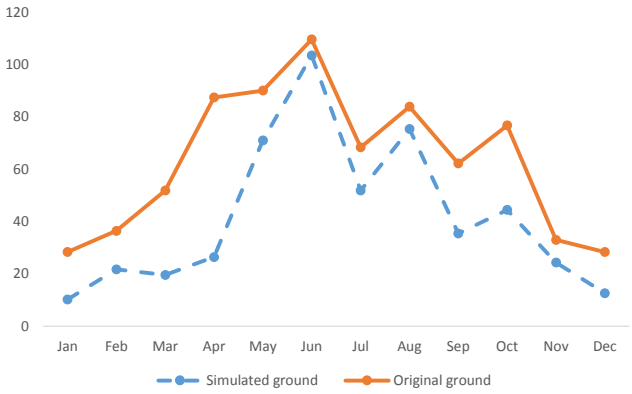


Figure 5.6. (a) Tabulated mean monthly total rainfall of the 20 stations in the observed ground-based dataset and the simulated ground-based dataset from 2002 to 2012. Highlighted values in green are differences lower than 25%. (b) Plotted differences between the observed ground-based dataset and the simulated ground-based dataset. The X-axis is the month and the Y-axis is the amount of rainfall.

(a)

Month	Simulated KFDR	Original KFDR	Difference	Difference percentage
Jan	19.60	19.06	0.54	2.83%
Feb	12.02	11.81	0.22	1.83%
Mar	71.57	51.68	19.89	38.49%
Apr	89.56	85.54	4.02	4.70%
May	140.18	126.80	13.38	10.55%
Jun	154.49	122.68	31.81	25.93%
Jul	72.66	56.31	16.36	29.05%
Aug	123.50	84.25	39.26	46.60%
Sep	58.59	42.61	15.98	37.49%
Oct	96.94	65.81	31.13	47.31%
Nov	39.33	30.84	8.49	27.52%
Dec	14.09	15.66	1.57	10.02%

(b)

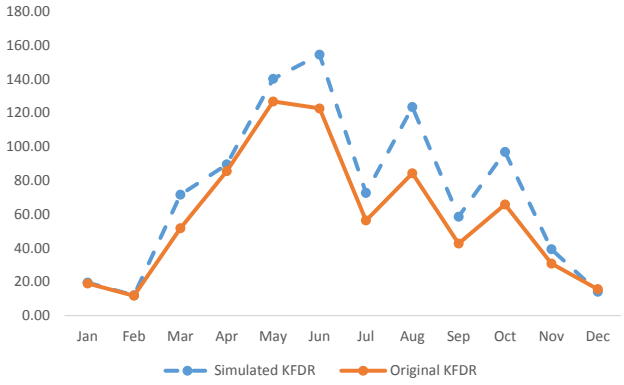


Figure 5.7. (a) Tabulated mean monthly total rainfall of the 20 stations in the observed KFDR dataset and the simulated KFDR dataset from 2002 to 2012. Highlighted values in green are differences lower than 25%. (b) Plotted differences between the observed KFDR dataset and the simulated KFDR dataset. The X-axis is the month and the Y-axis is the amount of rainfall.

(a)

Month	Simulated KTLX	Original KTLX	Difference	Difference percentage
Jan	18.07	11.64	6.43	55.23%
Feb	19.88	16.19	3.69	22.81%
Mar	39.13	35.56	3.57	10.04%
Apr	79.69	63.59	16.10	25.32%
May	106.32	95.33	10.99	11.52%
Jun	130.16	84.16	46.00	54.66%
Jul	56.99	43.16	13.83	32.05%
Aug	88.46	58.93	29.52	50.10%
Sep	40.39	29.39	11.00	37.43%
Oct	54.77	42.96	11.80	27.47%
Nov	21.68	14.98	6.70	44.73%
Dec	16.69	15.34	1.35	8.78%

(b)

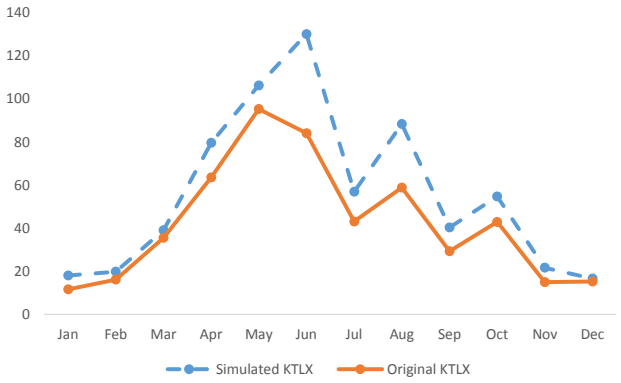


Figure 5.8. (a) Tabulated mean monthly total rainfall of the 20 stations in the observed KTLX dataset and the simulated KTLX dataset from 2002 to 2012. Highlighted values in green are differences lower than 25%. (b) Plotted differences between the observed KTLX dataset and the simulated KTLX dataset. The X-axis is the month and the Y-axis is the amount of rainfall.

By averaging the monthly differences (simulated minus observed) in three datasets, we found that the average percentage of two NexRad methods was 27.60% and the average percentage of ground-based was 38.74%. As the software ran 1000 times for each dataset, the differences could represent the ability of the software to regenerate the rainfall time series based on the rainfall patterns defined from the observed dataset. We conclude that the Zeus software can simulate the NexRad rainfall data more closely than the ground-based data set.

In addition, from figure 5.7 and 5.8, we found that the "Zeus" software generated the same increasing and decreasing tendency especially during the dry seasons including Jan, Feb, Mar, Apr and, Dec as the observed NexRad data. However, in the rainy months (e.g., May, June, July, and August), the ground-based data were more closely represented (Figure 5.6).

The differences between the simulations and the observed were further investigated by considering the processes generating the rainfall patterns. Based on the Equation 4.18, the average of 1000 simulations could make the expectation of the random error in the rainfall generating process zero. As a result, the differences between the generated rainfall time series and the observed rainfall time series can be written as follows:

$$D(t) = F(t) + Er(t) + Ep(t) \quad \text{(Equation 5.3)}$$

where D (t) is the differences between the generated rainfall time series and the observed rainfall time series; F (t) is the amount of rainfall which has a temporal pattern but could not be defined using the parameters listed in Table 4.5; Er (t) is the amount of rainfall caused by random environmental factors; Ep (t) is the error resulting from generating method which satisfied$|Ep(t)| \geq 0$.

As the rainfall pattern parameters generating methods and the rainfall time series generating processes were identical for all the three process, we conclude that:

$$Ep(t)_{NexRad} = Ep(t)_{Ground} \quad \text{(Equation 5.4)}$$

Based on Equation 5.3 and 5.4, we drew the conclusion that the differences in the performance between the NexRad dataset and the Ground-based dataset came from F (t) and Er (t). The relationship was shown in the Equation 5.5:

$$\Delta D(t) = \Delta F(t) + \Delta Er(t) \quad \text{(Equation 5.5)}$$

where $\Delta D(t)$ is the differences in performance of the rainfall re-generation process between the two datasets; $\Delta F(t)$ is the differences of the defined rainfall patterns between the two datasets using the parameters in the Table 4.4; $\Delta Er(t)$ is the differences of the random rainfall events between the two datasets.

The reasons for this differences might come from the differences in the threshold of minimum detection rainfall. As the detection range of the NexRad data set is 2.54 mm while the ground-based data set is 0.254 mm, the ground-based data can detect and record daily rainfall less than 2.54 mm, which suggests that the ground-based will record more rainfall events than the NexRad. This would result in differences in the amount of rainfall recorded in NexRad dataset and ground-based dataset as follow:

$$R(t)_{Ground} > R(t)_{NexRad} \quad \text{(Equation 5.6)}$$

where $R(t)_{Ground}$ is the recorded rainfall patterns in the ground-based rainfall time series; $R(t)_{NexRad}$ is the recorded rainfall patterns in the NexRad rainfall time series

However, as same methods and same parameters were used in the rainfall patterns defining process which were shown in section 4.5.1 but the observed datasets were different, we drew the Equation as follow:

$$\Delta F(t) > 0 \quad \text{(Equation 5.7)}$$

where $\Delta F(t)$ is the differences between the rainfall patterns defined in ground-based dataset and NexRad dataset.

Furthermore, the lower detection limit made the ground-based dataset record more random rainfall events than the NexRad dataset which meant that:

$$\Delta Er(t) > 0 \quad \text{(Equation 5.8)}$$

In conclusion, the generated rainfall time series is closer to the observed rainfall time series based on the NexRad dataset in general. However, some of the relationships may be different from the general equations in the wet and dry months.

To understand and explain the differences between the performance of NexRad and ground-based datasets in dry and wet months, the rainfall patterns in both of the datasets were further investigated. All the rainfall patterns in the dataset, including the patterns defined and undefined in the parameters used, could be written as:

$$P(t) = S(t) + F(t) = A1(t) + A2(t) + A3(t) + \cdots + Ax(t) \quad \text{(Equation 5.9)}$$

where $P(t)$ is the sum of all the rainfall patterns which could be defined; $An(t)$s are the rainfall patterns with different magnitudes. All the $An(t)$s satisfied the equations:

$$|Ax| \geq 2.54 \qquad for\ x \leq l_1 \quad \text{(Equation 5.10.1)}$$

$$0.254 \leq |Ax| < 2.54 \qquad for\ l_1 \leq x < l_2 \qquad \text{(Equation 5.10.2)}$$

$$|Ax| < 0.254 \qquad for\ x > l_2 \qquad \text{(Equation 5.10.3)}$$

$$|Ax| \geq |Ay| \qquad for\ 1 \leq x \leq y \leq n \qquad \text{(Equation 5.10.4)}$$

As the detection limits are different in the NexRad and Ground-based datasets, $S(t)$ could be written as Equation 5.11 in NexRad dataset and 5.12 in ground-based dataset.

$$S(t)_{NexRad} = A1(t) + A2(t) + \cdots + An(t) \qquad n = l_1 \qquad \text{n n} \qquad \text{(Equation 5.11)}$$

$$S(t)_{ground\_based} = A1(t) + A2(t) + \cdots + An(t) + \cdots + Am(t) \qquad m = l_2 \text{ (Equation 5.12)}$$

The rainfall patterns in dry months (e.g., Jan, Feb, Mar, Apr, and Dec) usually have more rainfall events but with lower intensity and duration. As a result, the rainfall that NexRad recorded is, in general, larger than or equal to 2.54 mm. On the other hand, the rainfall time series that the ground-based dataset recorded includes the rainfall events with lower magnitudes which can result in a larger differences between the observed dataset and simulated dataset. However, in the wet months (e.g., May, June, July, and August) the rainfall events usually have higher intensity and/or longer duration. As a result, the rainfall time series that NexRad recorded does not include the rainfall events with magnitudes lower than 2.54 mm which can result in incompleteness and errors in the defined rainfall patterns. On the other hand, the rainfall that Ground-based dataset can record a complete rainfall patterns which can be defined and re-generated more easily compared with the rainfall patterns recorded in the NexRad dataset.

Apart from the percent differences between dry months and wet months, the relationship between the generated dataset and observed dataset also have some patterns. All the simulated averages of NexRad monthly rainfall were larger than the observed data as shown in Figure 5.7 and 5.8, while the generated ground-based data were less than the observed data as shown in Figure 5.6. The reasons for this differences might include the following:

(1) The rainfall patterns defined from the NexRad dataset has the minimum threshold of 2.54 mm while the rainfall generator generated rainfall time series with the minimum value of 0.254 mm. As the result of this, the rainfall generator will generate some rainfall events which were not detected and recorded in the NexRad dataset; On the other hand, the rainfall pattern generated from the ground-based dataset has the threshold of 0.254mm while the rainfall generator generated the rainfall time series with the same threshold. However, as a highly complicated variable, the rainfall pattern defined can only represent and preserved part of the actual rainfall patterns. As a consequence, more patterns were generated in the simulated rainfall which caused the differences;

(2) The differences in the detection limits also affected the differences in the random rainfall events which means that the more random events would be detected and recorded in the Ground-based dataset than the NexRad dataset. However, the methods used in the rainfall patterns definition process did not include the random events. As the result of this, we should expect that the generated rainfall would be less than the observed rainfall.

However, as the differences and magnitude was the combinations of both factors, and the rainfall generated from the differences in the detection limits was more significant than the differences caused by the random differences in NexRad, the simulated NexRad data was still higher in magnitude than the observed NexRad data.

5.3.2 Distribution of the simulated maximum rainfall

The averages of the absolute PDFs differences between the daily maximum values of the observed and the simulated rainfall time series were calculated for all the 3000 simulations in KTLX, KFDR, and ground-based datasets to get the general performance of the “Zeus” software in generating rainfall pattern using the Equation 4.14. The results are shown in the Figure 5.9

where the X axis are the classes corresponding to the maximum rainfall intervals given in Table 5.2.

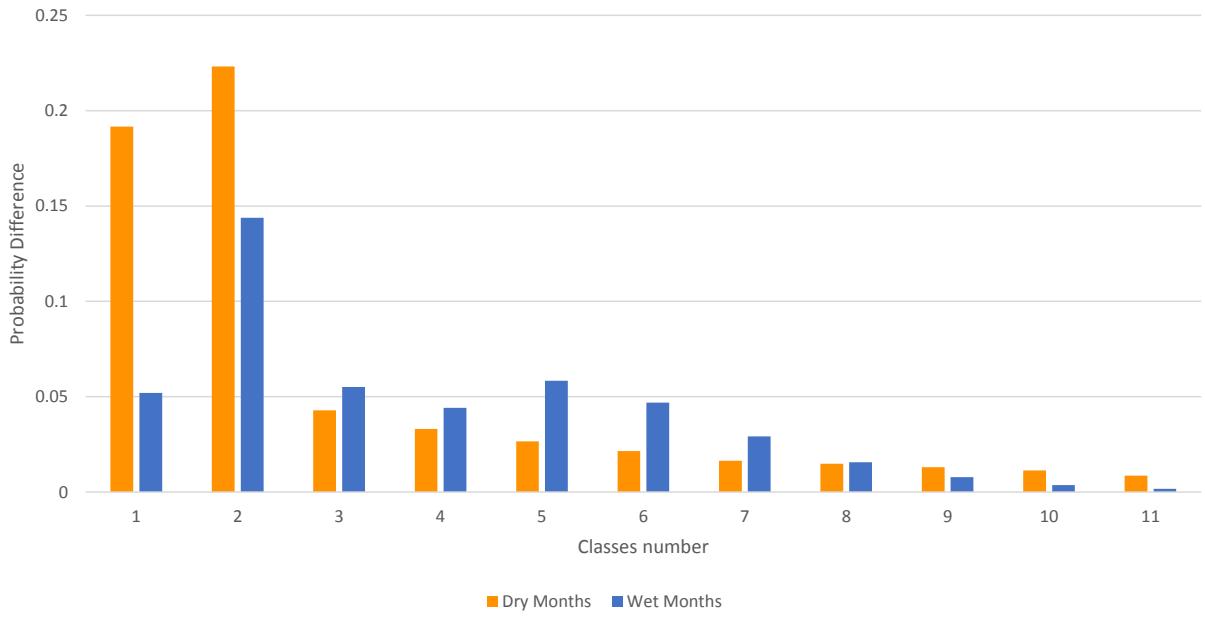


Figure 5.9. The X-axis is the classes of the daily area maximum rainfall classes with the meaning shown in Table 5.2; The Y-axis is the average absolute value of the probability differences between observed and simulated rainfall; the orange bars are the average differences in dry months; the blue bars are the differences in wet months.

Table 5.2. The amount of rainfall in the days contained in each class number n.

CLASS NUMBER N	THE RAINFALL AMOUNTS IN THE CORRESPONDING CLASS
1	0
2	(0,0.1Rm*]
3	(0.1Rm,0.2Rm]
4	(0.2Rm,0.3Rm]
5	(0.3Rm,0.4Rm]
6	(0.4Rm,0.5Rm]
7	(0.5Rm,0.6Rm]
8	(0.6Rm,0.7Rm]
9	(0.7Rm,0.8Rm]
10	(0.8Rm,0.9Rm]
11	(0.9Rm,Rm]

*Rm is the maximum rainfall in that month.

In Figure 5.9, the averages of the absolute PDF differences between the daily maximum value of observed and simulated rainfall in the wet months were lower in general with an average of 4.17%, compared with the average of 5.48% in dry months. Furthermore, the differences in the days without rainfall was about 5% in the wet months while the differences in dry months is about

19%. As the first class probability is the frequency of dry days in the time series, the "Zeus" software can generate the time series closer to the observed time series considering the patterns of rainy days and dry days.

Apart from the averages absolute differences, the averages of each class in dry months (Jan, Feb, Mar, Apr, Sep, Oct, Nov, and Dec) and wet months (May, Jun, Jul, and Aug) in both ground-based dataset and NexRad data were also calculated and compared based on data sources with the result shown in Figure 5.10 and 5.11.

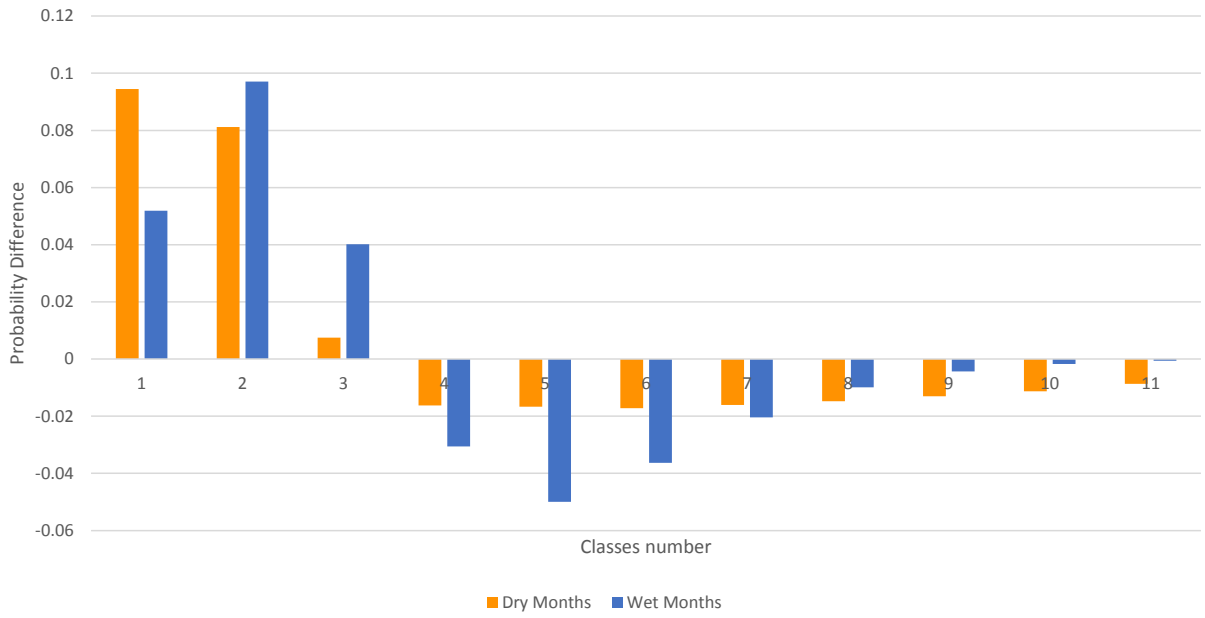


Figure 5.10. Average of the PDF differences between the daily maximum value of observed and simulated rainfall in the NexRad dataset. The X-axis is the classes of the daily area maximum rainfall classes with the meaning shown in Table 5.2; The Y-axis is the averages of the probability differences between observed and simulated rainfall in the NexRad dataset. The orange bars represented dry months and the blue bars represented wet months.

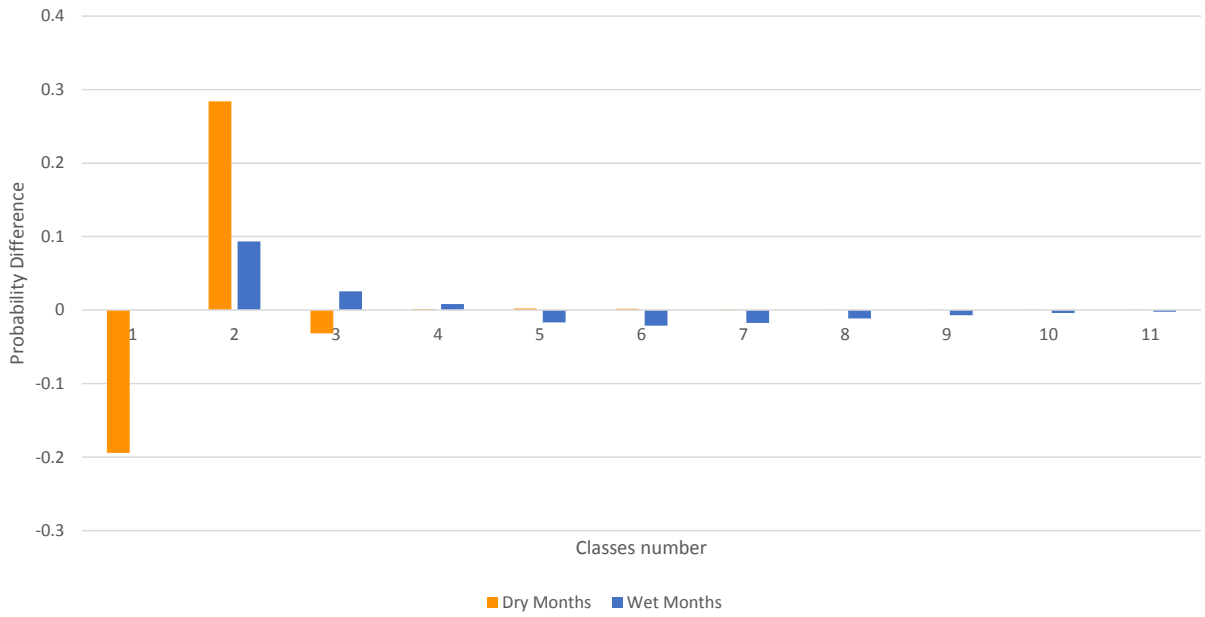


Figure 5.11. Average of the PDF differences between the daily maximum value of observed and simulated rainfall in the Ground-based dataset. The X-axis is the classes of the daily area maximum rainfall classes with the meaning shown in Table 5.2; The Y-axis is the averages of the probability differences between observed and simulated rainfall in the ground-based dataset. The orange bars represented dry months and the blue bars represented wet months.

For further analysis of the different performance of the software between the NexRad dataset and ground-based dataset, maximum PDF differences in both datasets based on dry or wet months were also compared in Figure 5.12 and 5.13.

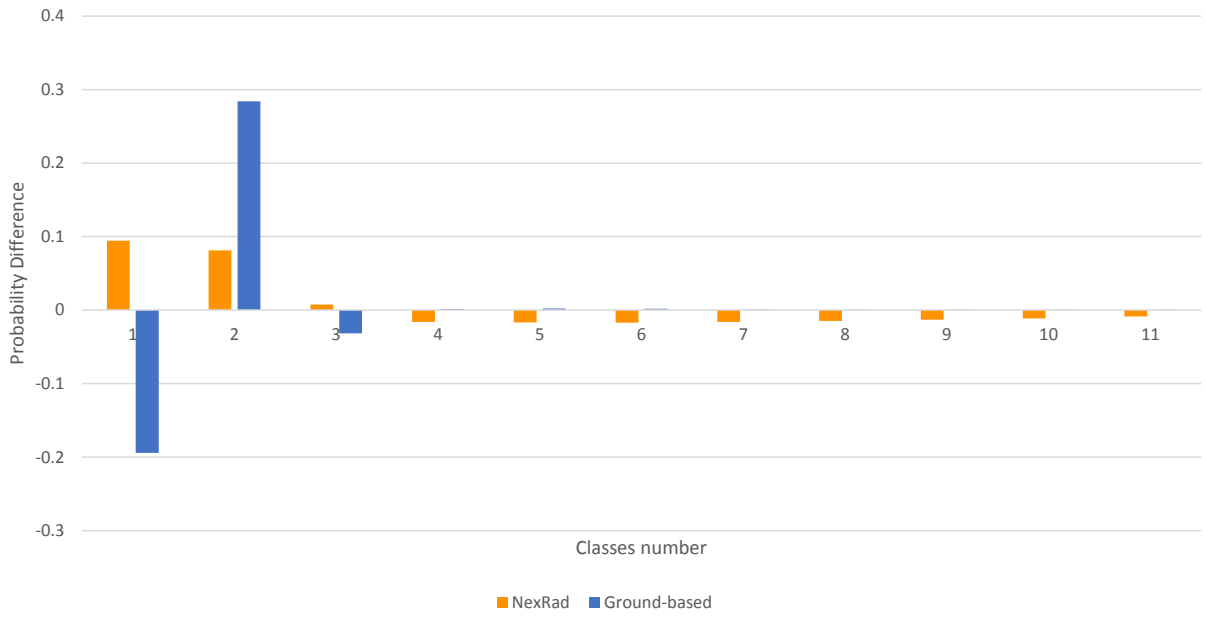


Figure 5.12. Average of the PDF differences between the daily maximum value of observed and simulated rainfall in dry months. The X-axis is the classes of the daily area maximum rainfall classes with the meaning shown in Table 5.2; the Y-axis is the averages of the probability differences between observed and simulated rainfall in the ground-based dataset. The orange bars represented NexRad and the blue bars represented ground-based.

Figure 5.13. Average of the PDF differences between the daily maximum value of observed and simulated rainfall in wet months. The X-axis is the classes of the daily area maximum rainfall classes with the meaning shown in Table 5.2; The Y-axis is the averages of the probability differences between observed and simulated rainfall in the ground-based dataset. The orange bars represented NexRad and the blue bars represented ground-based.

In Figure 5.10 to 5.13, all the differences were calculated by the Equation 5.13:

$$D(n,t)_{PDF} = F(n,t)_{Simulated} - F(n,t)_{Observed} \quad \text{(Equation 5.13)}$$

where $D(n,t)_{PDF}$ is the differences in the PDF of the nth class in the month t between the simulated dataset and observed dataset. The meanings of class number n are shown in Table 5.2; $F(n,t)_{Simulated}$ is the frequency of the nth class in the month t in the simulated dataset; $F(n,t)_{Observed}$ is the frequency of the nth class in the month t in the observed dataset.

Based on the explanations of the Equation 5.13, a positive $D(n,t)_{PDF}$ means that more days in the nth class in the month t from the simulated dataset than in observed dataset while a negative $D(n,t)_{PDF}$ means that less days in the nth class in the month t from the simulated dataset than in observed dataset.

From Figure 5.10, all the PDF differences of all the classes were less than 10% in general which showed the software could generate rainfall time series better based on the NexRad dataset. In addition, the software would generate more days with no or little rain (≤0.1*maximum daily rainfall) and fewer days with larger rainfall (>0.3*maximum daily rainfall) than the observed dataset in both dry and wet months. From Figure 5.11, all the PDF differences in the wet months were smaller in absolute value which demonstrated again that the software could simulate rainfall time series closer to the observed data in wet months in the ground-based dataset. In the dry months, the software would generate fewer days without rainfall but more days with little rainfall (≤0.1*maximum daily rainfall). From Figure 5.12, the simulated dataset was closer to the observed dataset in the NexRad dataset than the ground-based dataset. From Figure 5.13, the small value in the first class in both datasets showed that the rainfall generator could simulate the pattern of rain and no rain very close to the observed dataset in terms of the rainy days and dry days. However, both NexRad and ground-based datasets showed more days were generated with little daily rainfall than the observed datasets.

Based on the conclusions, the rainfall patterns defining the ability of the parameters used in the rainfall generator and the rainfall generator simulation ability changes according to different rainfall patterns. The ability improves when the rainfall level is extremely high or low, for example dry months in the NexRad dataset when dry days outnumber the wet days. This could also explain the relatively good performance in the wet months in the ground-based dataset when rainfall is rich compared with other months.

CHAPTER 6: CONCLUSIONS AND FUTURE WORK

6.1 The NexRad rainfall time series

The issue of high uncertainty and error in environmental models caused by the absence of high temporal and spatial resolution rainfall data in a small-scale watershed can be addressed by NexRad rainfall dataset which is a Level-III radar product generated from the Level-II radar data. Compared with the time series generated from the ground-based rainfall gages, the NexRad rainfall time series has advantages over the following aspects:

(1) Large coverage area for the U.S: the NexRad covers nearly all the mainland U.S.A. including areas in which ground-based rainfall gages are hard to set-up(e.g. regions in deep forests, high altitude area);

(2) High temporal and spatial resolution: the highest temporal resolution is 5 minutes and the highest spatial resolution is 250 meters;

(3) Multiple stations covering the same area for calibration: due to large number of stations and large coverage area of each station, there are usually more than one NexRad stations covering every watershed in the U.S;

(4) Historic rainfall data availability: the historic rainfall data is available since 1988 when NexRad project was set up;

In conclusion, NexRad is a useful resource for rainfall data that can provide both high temporal and spatial resolution.

6.2 The Comparisons between NexRad rainfall data and ground-based rainfall data

The comparison between the NexRad dataset and the ground-based dataset revealed some similarities and differences between these two dataset.

(1) NexRad has a tendency of recording higher rainfall when events have high intensity and long duration. As a result, the rainfall recorded in the months of March, April, May, June, and August are more close to the rainfall recorded from the ground-based station with the average different percentage of less than 25%. Rainfall recorded in the other months can easily be affected by extreme rainfall events and the difference is usually higher than 40%.

(2) The PDFs of the NexRad rainfall time series are not the same as the PDFs of the ground-based rainfall time series. Further investigation revealed the main reason for the differences was that NexRad could not detect the rainfall events less than 2.54mm in the time series. The other possible reasons can be the "Z-R" relationship used in the model is not effective in measuring short-lasting rainfall events and differences in the detecting targets between NexRad and ground-based gages.

(3) The all-time high-resolution NexRad rainfall is very useful in capturing heavy and intense rainfall events and can be used to model processes that result from such events (e.g., flooding, erosion). However, it might not deliver realistic results as the ground-based datasets for low intensity or short duration rainfall events. Furthermore, it should be noted that the NexRad measures a bigger spatial area than the ground-based station. The ground-based station, although measuring a smaller spatial area, measures the actual rainfall reaching the ground as opposed to the NexRad which is based on reflectivity information from the clouds. These differences should be taken into consideration when deciding which datasets can best describe the processes to be modeled.

In conclusion, NexRad rainfall time series is different from the ground-based rainfall time series in monthly mean rainfall values and PDFs of daily rainfall which can result in uncertainty and errors in the outcomes when the other rainfall time series is used in the environmental models. However, both rainfall measuring and recording method can record the same increasing or decreasing tendency in general.

6.3 Rainfall pattern analysis

By comparing the re-generated rainfall time series based on three datasets (KFDR, KTLX, and ground-based) with their observed datasets, the "Zeus" software, which uses the parameters listed in Table 4.5 as rainfall pattern inputs, proves to be a useful software to generating and downscale the rainfall time series. The results of the comparisons revealed following conclusions:

(1) All the re-generated rainfall time series showed the same increasing and decreasing tendency in monthly scale. Furthermore, the rainfall generator could re-generate rainfall time series within 10% error in the monthly rainfall in NexRad dataset and wet months (May, June, July and August) in the ground-based dataset in the process.

(2) The PDFs of the daily maximum rainfall time series also showed similarities in some particular months. The result showed that in the dry months (January, February, March, April, September, October, November, and December) in NexRad dataset and wet months in the ground-based dataset, all differences within all classes listed in Table 5.2 were smaller than 10%. However, the analysis also showed significant differences between simulated rainfall time series and observed time series in terms of total monthly rainfall and PDFs of daily maximum rainfall. Possible reasons include the difference between the thresholds of the observed datasets, incomplete rainfall pattern defining, and errors in re-generating process.

In conclusion, the method of defining rainfall patterns based on the parameters in Table 4.5 and rainfall time series generating based on these parameters is useful when rainfall time series data is not available. However, further improvements must be made to decrease the errors and differences between the simulated time series and observed time series in both rainfall pattern defining the process and the software.

Overall, our results evidence the following points:

(1) NexRad is a useful and reliable source of rainfall time series in environmental research for most of the area in the U.S with high temporal and spatial resolution;

(2) The NexRad rainfall time series is different from the ground-based rainfall time series in mean value and PDFs. However, both of them showed the same tendency in monthly and yearly scale which means the NexRad can be used as one substitute dataset for the ground-based dataset;

(3) The rainfall pattern defining method used in the thesis and the rainfall generator can generate the basic increasing and decreasing tendency in monthly scale. However, further improvements on both the rainfall pattern defining methods and the software are necessary before reliable rainfall time series can be generated.

The results of this study are not limited to a comparison between different rainfall datasets but have implications for better hydrologic process understanding in small-scale watershed. For example, by analyzing the difference in rainfall inputs and their impacts on the model outputs, we can better understand how the different rainfall patterns affect the hydrologic processes and consequently the ecosystem. This will also facilitate understanding of the impacts of climate changes on the environment. By being able to define the rainfall pattern accurately and completely in small watersheds will lead to better representation of the response of the environmental system

and will therefore result in less complexity and better performance in environmental models, simplified rainfall dataset, and a better understanding of the relationships between different environmental factors in a small-scale watershed.

REFERENCE

Aleman, J. C., Blarquez, O., & Staver, C. A. (2016). Land-use change outweighs projected effects of changing rainfall on tree cover in sub-Saharan Africa. Global change biology.

Badylak, S., Phlips, E., Dix, N., Hart, J., Srifa, A., Haunert, D., ... & Yang, Y. (2016). Phytoplankton dynamics in a subtropical tidal creek: influences of rainfall and water residence time on composition and biomass. Marine and Freshwater Research, 67(4), 466-482.

Ciach, G. J. (2003). Local random errors in tipping-bucket rain gauge measurements. Journal of Atmospheric and Oceanic Technology, 20(5), 752-759.

Crum, T. D., & Alberty, R. L. (1993). The WSR-88D and the WSR-88D operational support facility. Bulletin of the American Meteorological Society,74(9), 1669-1687.

Duncan, M. R., Austin, B., Fabry, F., & Austin, G. L. (1993). The effect of gauge sampling density on the accuracy of streamflow prediction for rural catchments. Journal of Hydrology, 142(1), 445-476.

Elliott, R. L., Schiebe, F. R., Crawford, K. C., Peter, K. D., & Puckett, W. E. (1993, December). A unique data capability for natural resources studies. In International Winter Meeting of the American Society of Agricultural Engineers, Chicago, IL (pp. 14-17).

Fabry, F., Bellon, A., Duncan, M. R., & Austin, G. L. (1994). High resolution rainfall measurements by radar for very small basins: the sampling problem reexamined. Journal of Hydrology, 161(1), 415-428.

Fankhauser, R. "Measurement properties of tipping bucket rain gauges and their influence on brain runoff simulation." Water Science and Technology 36, no. 8 (1997): 7-12.

Fournier, J. D. (1999). Reflectivity-rainfall rate relationships in operational meteorology. National Weather Service Technical Memo, National Weather Service, Tallahassee, FL.

Grotch, S. L., & MacCracken, M. C. (1991). The use of general circulation models to predict regional climatic change. Journal of Climate, 4(3), 286-303.

Guzman, J. A., Moriasi, D. N., Chu, M. L., Starks, P. J., Steiner, J. L., & Gowda, P. H. (2013). A tool for mapping and spatio-temporal analysis of hydrological data. Environmental modelling & software, 48, 163-170.

Habib, E., Krajewski, W. F., & Kruger, A. (2001). Sampling errors of tipping-bucket rain gauge measurements. Journal of Hydrologic Engineering, 6(2), 159-166.

Hansen, D. P., Ye, W., Jakeman, A. J., Cooke, R., & Sharma, P. (1996). Analysis of the effect of rainfall and streamflow data quality and catchment dynamics on streamflow prediction using the rainfall-runoff model IHACRES. Environmental Software, 11(1), 193-202.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International journal of climatology, 25(15), 1965-1978.

Hill, D. J., Liu, Y., Marini, L., Kooper, R., Rodriguez, A., Futrelle, J., ... & McLaren, T. (2011). A virtual sensor system for user-generated, real-time environmental data products. Environmental Modelling & Software, 26(12), 1710-1724.

Hoffmann, M., Schwartengräber, R., Wessolek, G., & Peters, A. (2016). Comparison of simple rain gauge measurements with precision lysimeter data. Atmospheric Research, 174, 120-123.

Hunter, S. M. (1996). WSR-88D radar rainfall estimation: capabilities, limitations and potential improvements. Natl. Wea. Dig, 20(4), 26-38.

Jia, Z., Tang, S., Luo, W., & Hai, Y. (2016). Water quality improvement through five constructed serial wetland cells and its implication on nonpoint-source pollution control. Hydrological Sciences Journal.

Kim, T. W., & Valdés, J. B. (2003). Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. Journal of Hydrologic Engineering, 8(6), 319-328.

Liuzzo, L., Bono, E., Sammartano, V., & Freni, G. (2015). Analysis of spatial and temporal rainfall trends in Sicily during the 1921–2012 period. Theoretical and Applied Climatology, 1-17.

Moser, B. A., Gallus Jr, W. A., & Mantilla, R. (2015). An Initial Assessment of Radar Data Assimilation on Warm Season Rainfall Forecasts for Use in Hydrologic Models. Weather and Forecasting, 30(6), 1491-1520.

Seed, A. W., & Austin, G. L. (1990). Sampling errors for raingauge-derived mean areal daily and monthly rainfall. Journal of Hydrology, 118(1), 163-173.

Seo, B. C., Dolan, B., Krajewski, W. F., Rutledge, S. A., & Petersen, W. (2015). Comparison of Single-and Dual-Polarization–Based Rainfall Estimates Using NEXRAD Data for the NASA Iowa Flood Studies Project.Journal of Hydrometeorology, 16(4), 1658-1675.

Seo, B. C., Krajewski, W. F., Kruger, A., Domaszczynski, P., Smith, J. A., & Steiner, M. (2011). Radar-rainfall estimation algorithms of Hydro-NEXRAD.Journal of Hydroinformatics, 13(2), 277-291.

Shi, P., Wu, M., Qu, S., Jiang, P., Qiao, X., Chen, X., ... & Zhang, Z. (2015). Spatial distribution and temporal trends in precipitation concentration indices for the southwest China. Water Resources Management, 29(11), 3941-3955.

Smith, J. A., Baeck, M. L., Meierdiercks, K. L., Miller, A. J., & Krajewski, W. F. (2007). Radar rainfall estimation for flash flood forecasting in small urban watersheds. Advances in Water Resources, 30(10), 2087-2097.

Ulbrich, C. W., & Lee, L. G. (1999). Rainfall measurement error by WSR-88D radars due to variations in ZR law parameters and the radar constant. Journal of Atmospheric and Oceanic Technology, 16(8), 1017-1024.

Wetterhall, F., Halldin, S., & Xu, C. Y. (2005). Statistical precipitation downscaling in central Sweden with the analogue method. Journal of Hydrology, 306(1), 174-190.

Wilby, R. L., Hassan, H., & Hanaki, K. (1998). Statistical downscaling of hydrometeorological variables using general circulation model output. Journal of Hydrology, 205(1), 1-19.

Wilby, R. L., Hay, L. E., & Leavesley, G. H. (1999). A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado. Journal of Hydrology, 225(1), 67-91.

Zhong, S., & Yang, X. Q. (2015). Ensemble simulations of the urban effect on a summer rainfall event in the Great Beijing Metropolitan Area. Atmospheric Research, 153, 318-334.

APPENDIX A: THE LOCATIONS OF KFDR AND KTLX STATIONS

Table A.1. The locations of KFDR and KTLX stations

	LONGITUDE	LATITUDE
KFDR	-98.97611	34.36222
KTLX	-97.2775	35.33306

APPENDIX B: THE CELLS AND STATIONS RELATIONSHIP

Table B.1. The cells and station relationship of KTLX station

STATION NAME	NUMBER OF CELL NEEDED	CELL NUMBER
A124	4	7,8,19,20
N1NN	2	10,22
A121	1	23
A133	1	18
A132	1	17
A131	2	15,16
APAC	1	26
A250	2	27,39
A249	1	41
A148	2	30,42
A234	1	31
A235	4	32,33,20,21
A136	1	34
A146	1	44
A152	2	39,51
A253	2	52,53
A154	1	54
A282	1	55
ACME	2	68,69
A159	2	69,81
A256	1	58
A244	2	47,59
A262	2	66,78

Table B.2. The cells and station relationship of KFDR station

STATION NAME	NUMBER OF CELL NEEDED	CELL NUMBER
A154	1	49
A124	2	7,18
N1NN	2	20,21
A121	1	22
A131	1	14
A132	2	15,26
A133	2	16,17
APAC	1	24
A250	2	25,36
A249	1	37
A148	2	38,39
A234	1	29
A235	1	30
A136	1	31
A146	1	41
A152	1	47
A253	1	48
A282	1	51
ACME	1	63
A256	1	53
A244	2	54,55
A159	4	63,64,74,75
A262	4	60,61,71,72

APPENDIX C: DOUBLE MASS CURVE RESULTS FOR TWELVE MONTHS

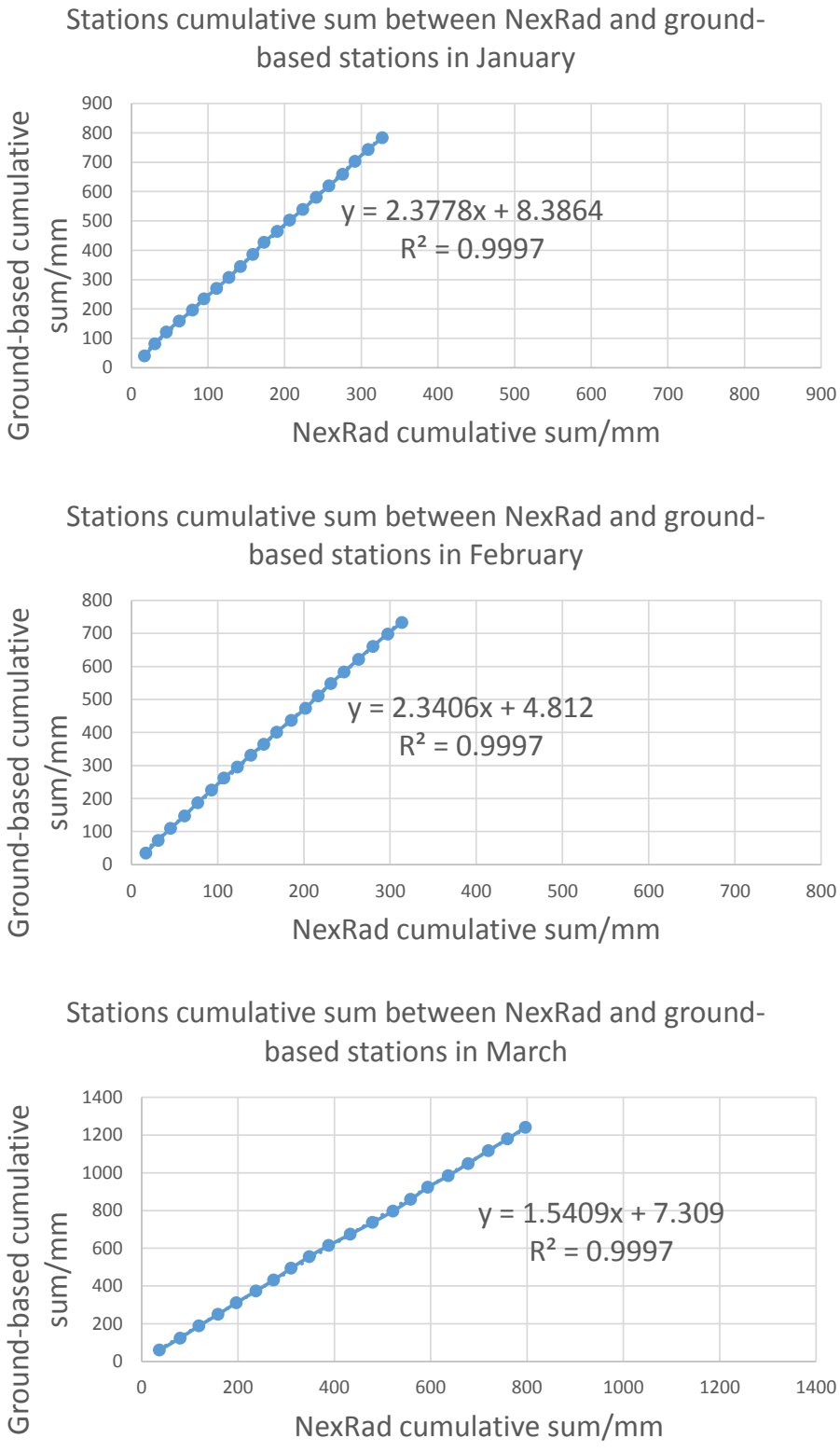


Figure C.1. The double mass analysis between the ground-based dataset and NexRad dataset in January, February, and March.

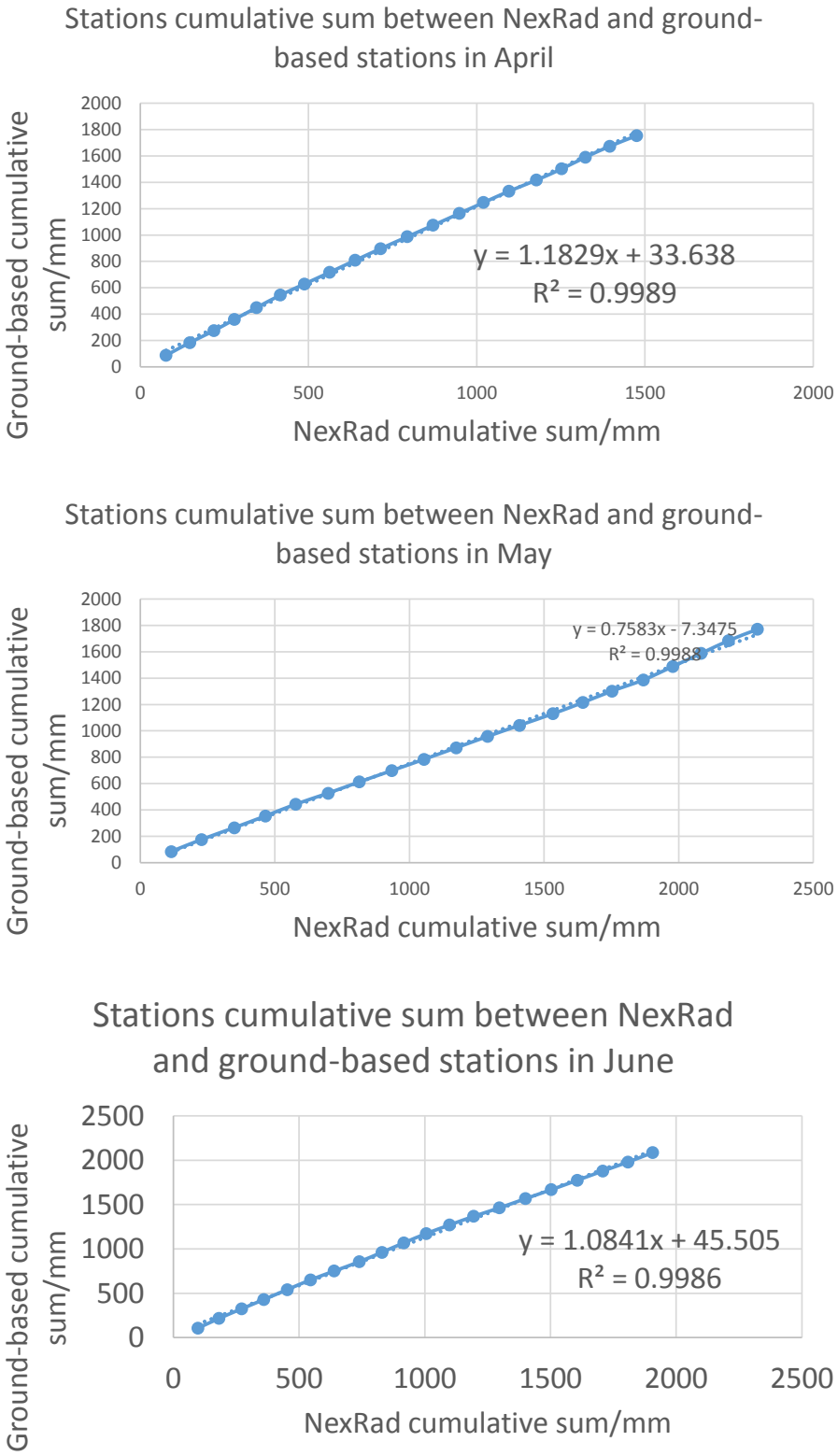
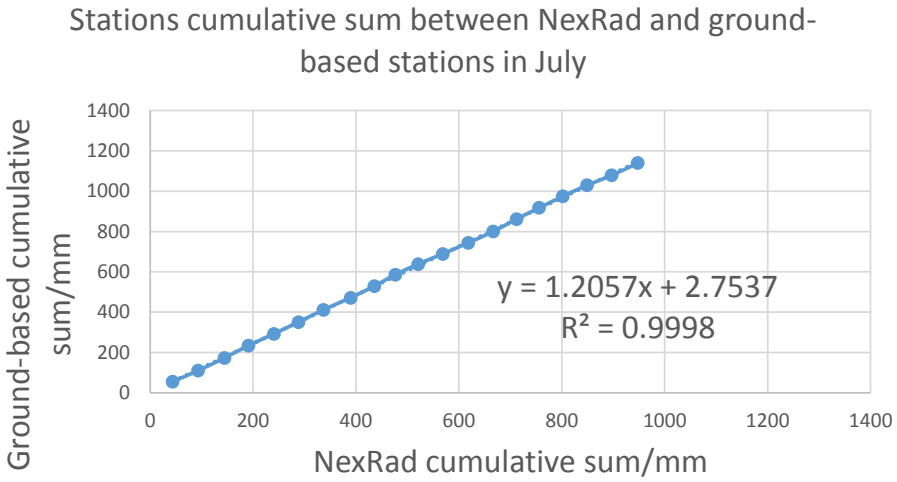
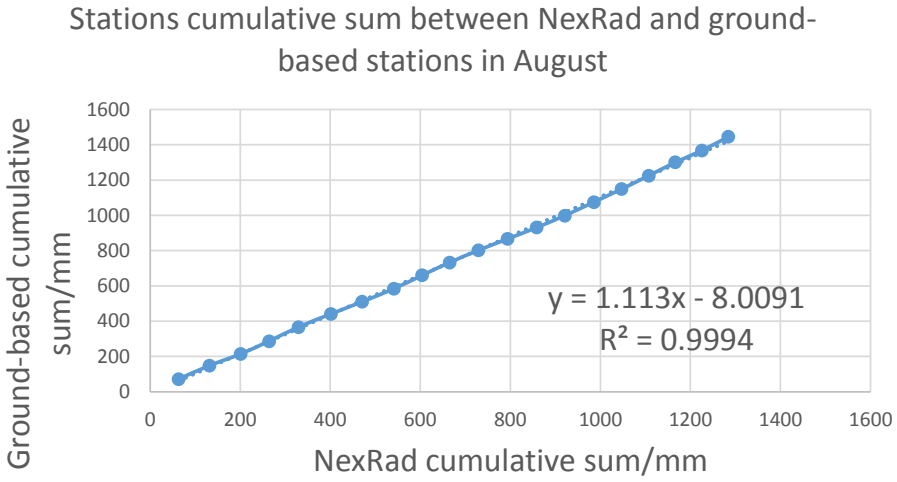


Figure C.2. The double mass analysis between the ground-based dataset and NexRad dataset in April, May, and June.





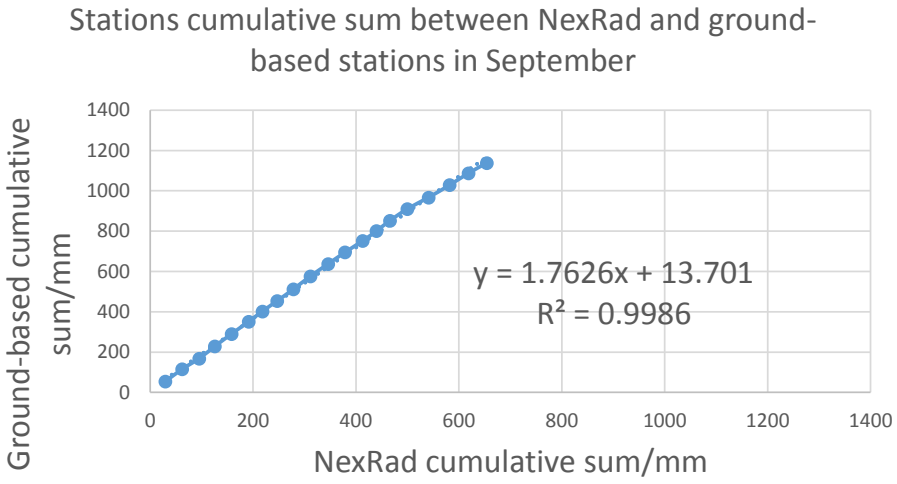
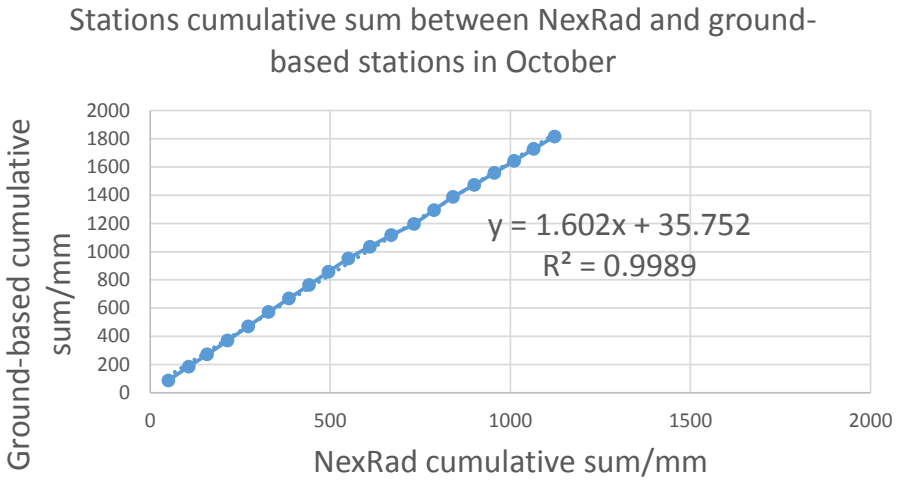
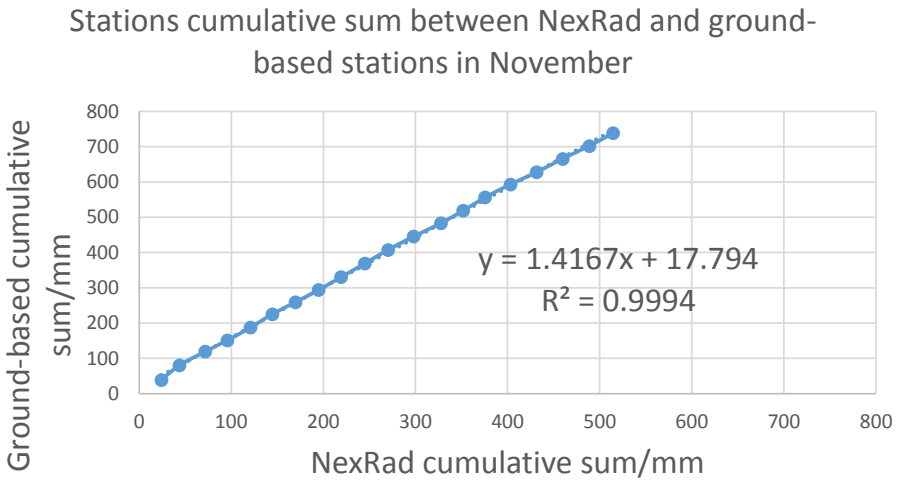


Figure C.3. The double mass analysis between the ground-based dataset and NexRad dataset in July, August, and September.





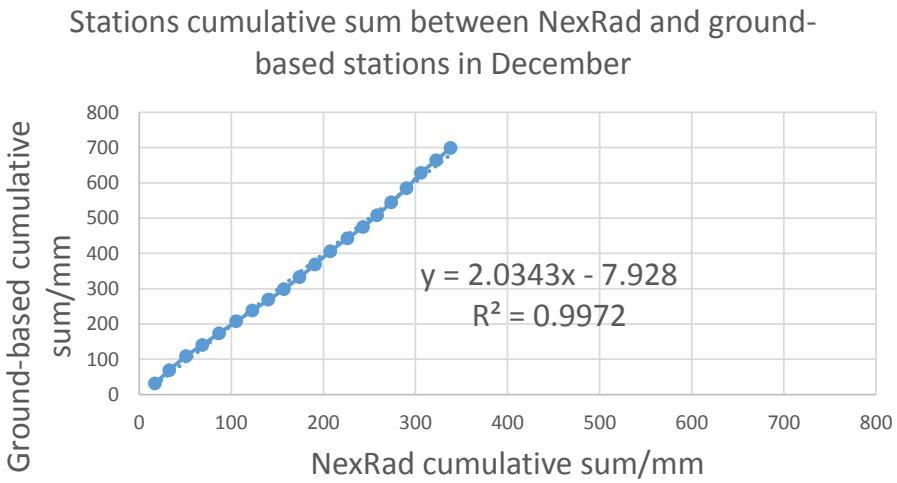


Figure C.4. The double mass analysis between the ground-based dataset and NexRad dataset in October, November, and December.

APPENDIX D: R RESULT OF F-TEST OF AREAL MONTHLY RAINFALL BETWEEN NEXRAD AND GROUND-BASED DATASET

Table D.1. The R result of F-test of areal monthly rainfall between NexRad and ground-based dataset

TIME SERIES MONTH	F VALUE	DEGREES OF FREEDOMS	P VALUE
JAN	1850.2	1	< 2.2e-16
FEB	2152.7	1	< 2.2e-16
MAR	657.82	1	< 2.2e-16
APR	102.73	1	3.549e-12
MAY	219.01	1	< 2.2e-16
JUN	27.791	1	6.509e-06
JUL	85.226	1	5.046e-11
AUG	35.877	1	6.529e-07
SEP	318.07	1	< 2.2e-16
OCT	428.32	1	< 2.2e-16
NOV	243.94	1	< 2.2e-16
DEC	460.65	1	2.937e-16

APPENDIX E: R RESULT OF KOLMOGOROV-SMIRNOV TEST BETWEEN KTLX AND GROUND-BASED DATASET

Table E.1. The R result of Kolmogorov-Smirnov test between KTLX and ground-based dataset

TIME SERIES MONTH	D VALUE	ALTERNATIVE HYPOTHESIS	P VALUE
JAN	0.2117	Two-sided	3.546e-10
FEB	0.2490	Two-sided	5.784e-13
MAR	0.1580	Two-sided	3.933e-06
APR	0.1352	Two-sided	0.0001908
MAY	0.1500	Two-sided	1.082e-05
JUN	0.1649	Two-sided	5.166e-07
JUL	0.0389	Two-sided	0.7824
AUG	0.0831	Two-sided	0.04106
SEP	0.1348	Two-sided	0.0001204
OCT	0.1924	Two-sided	5.047e-09
NOV	0.1867	Two-sided	4.781e-08
DEC	0.2556	Two-sided	6.55e-15

APPENDIX F: R RESULT OF KOLMOGOROV-SMIRNOV TEST BETWEEN KFDR AND GROUND-BASED DATASET

Table F.1. The R result of Kolmogorov-Smirnov test between KFDR and ground-based dataset

TIME SERIES MONTH	D VALUE	ALTERNATIVE HYPOTHESIS	P VALUE
JAN	0.2327	Two-sided	7.449e-11
FEB	0.2737	Two-sided	7.046e-07
MAR	0.1808	Two-sided	1.149e-06
APR	0.1548	Two-sided	7.186e-05
MAY	0.1479	Two-sided	0.0001335
JUN	0.1422	Two-sided	0.0002999
JUL	0.0343	Two-sided	0.9525
AUG	0.0735	Two-sided	0.1721
SEP	0.1232	Two-sided	0.002618
OCT	0.1985	Two-sided	1.093e-07
NOV	0.1895	Two-sided	5.539e-07
DEC	0.2760	Two-sided	1.199e-14

APPENDIX G: R RESULT OF KOLMOGOROV-SMIRNOV TEST WITHIN THE NEXRAD STATIONS

Table G.1. The R result of Kolmogorov-Smirnov test within the NexRad stations

TIME SERIES MONTH	D VALUE	ALTERNATIVE HYPOTHESIS	P VALUE
JAN	0.0225	Two-sided	0.9999
FEB	0.0308	Two-sided	1.0000
MAR	0.0269	Two-sided	0.9961
APR	0.0300	Two-sided	0.9879
MAY	0.0309	Two-sided	0.9776
JUN	0.0640	Two-sided	0.2797
JUL	0.0268	Two-sided	0.9962
AUG	0.0208	Two-sided	1.0000
SEP	0.0297	Two-sided	0.9880
OCT	0.0211	Two-sided	1.0000
NOV	0.0277	Two-sided	0.9972
DEC	0.0204	Two-sided	1.0000

APPENDIX H: THE LOG-SCALE FIGURES OF GROUND-BASED DATASET OF TWELVE MONTHS

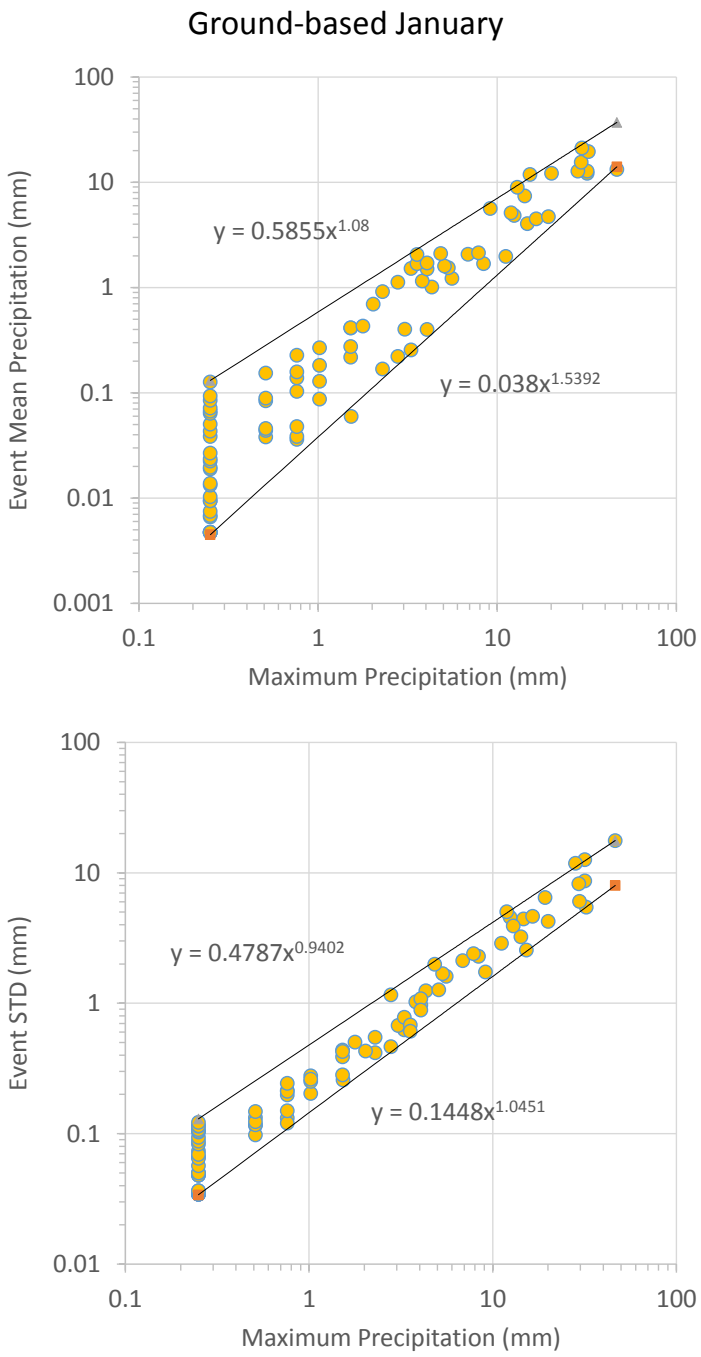


Figure H.1. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in January from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in January from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

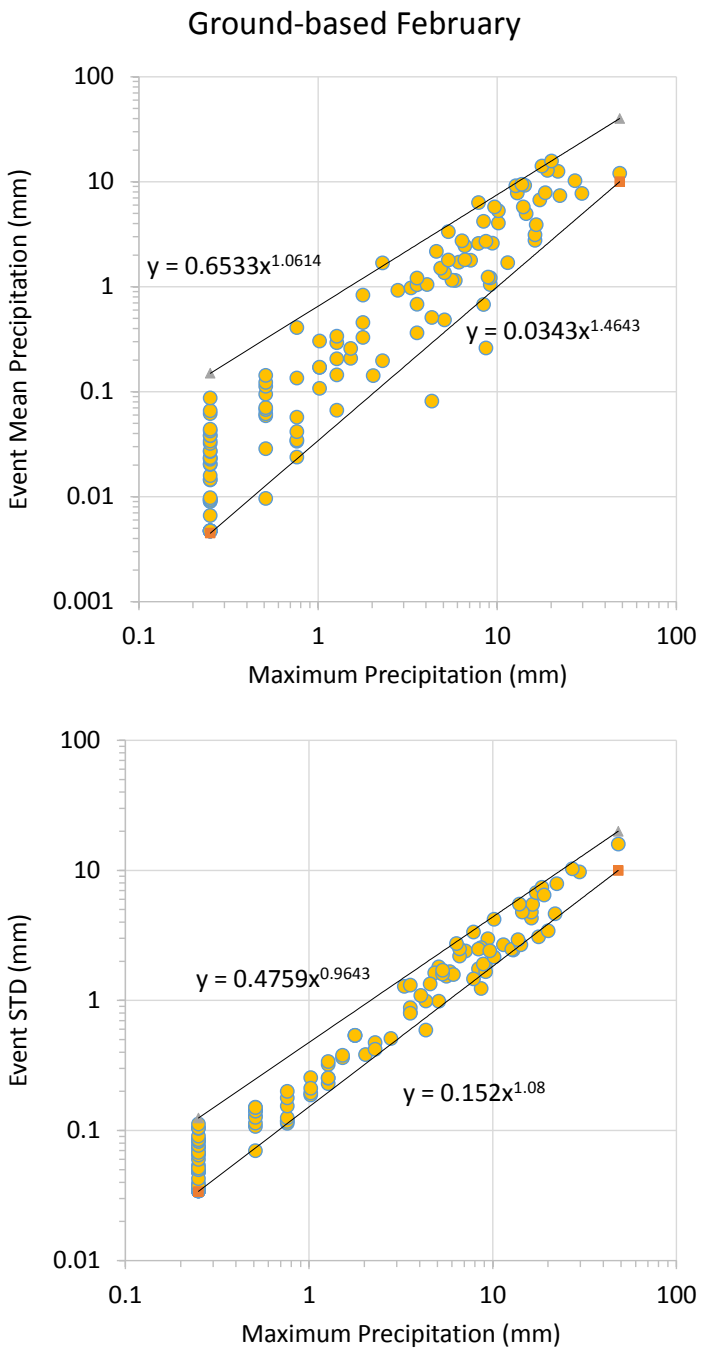


Figure H.2. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in February from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in February from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

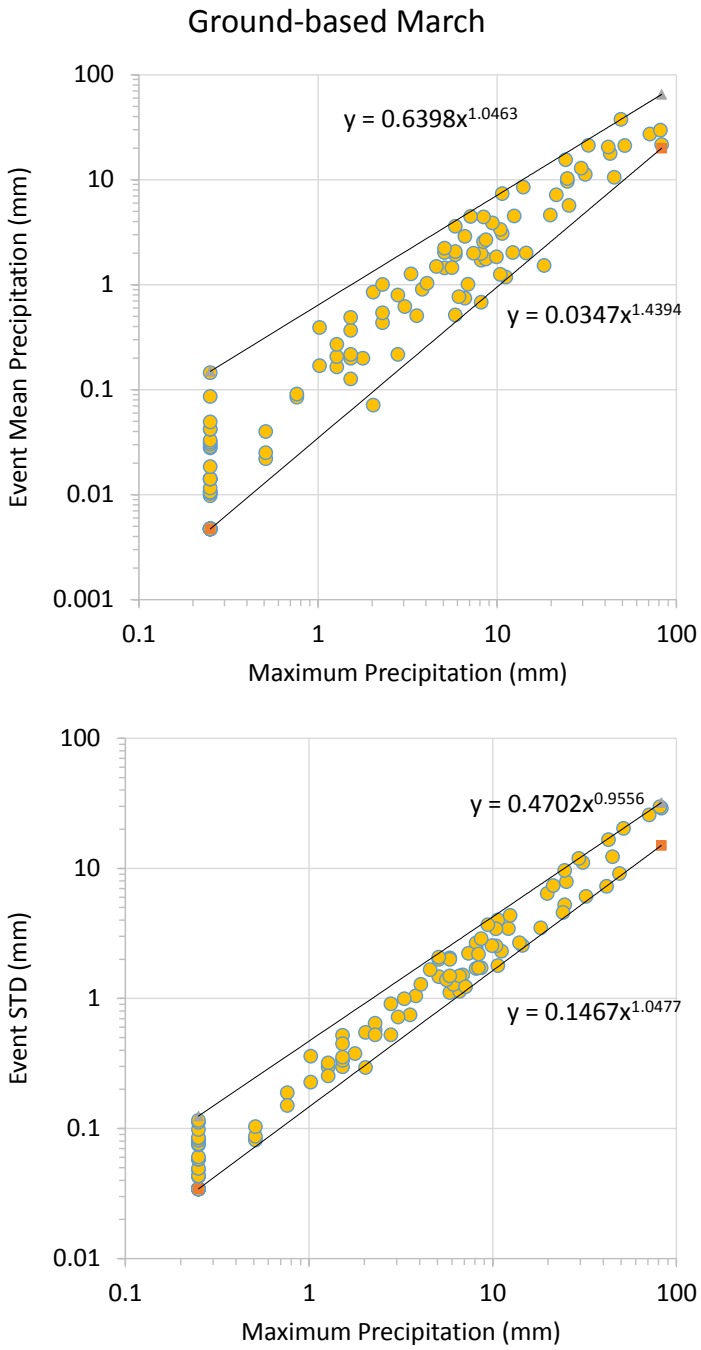


Figure H.3. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in March from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in March from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

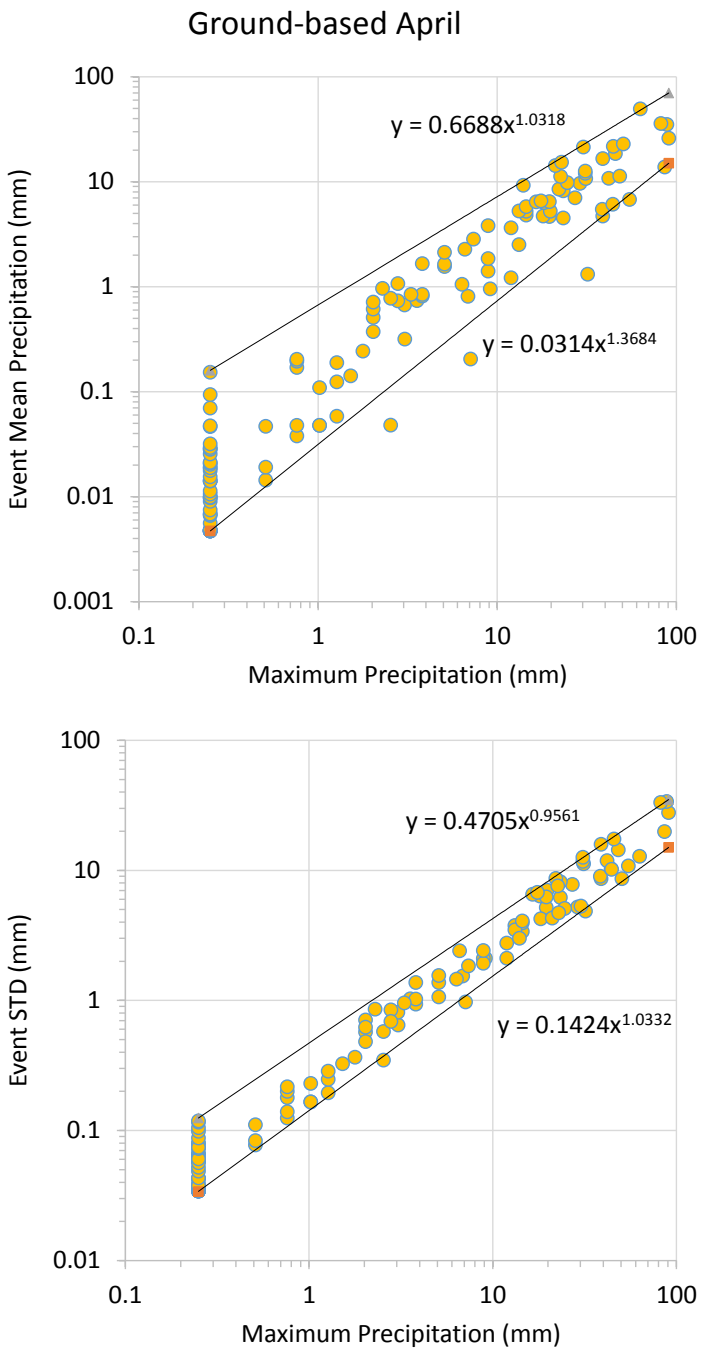


Figure H.4. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in April from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in April from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

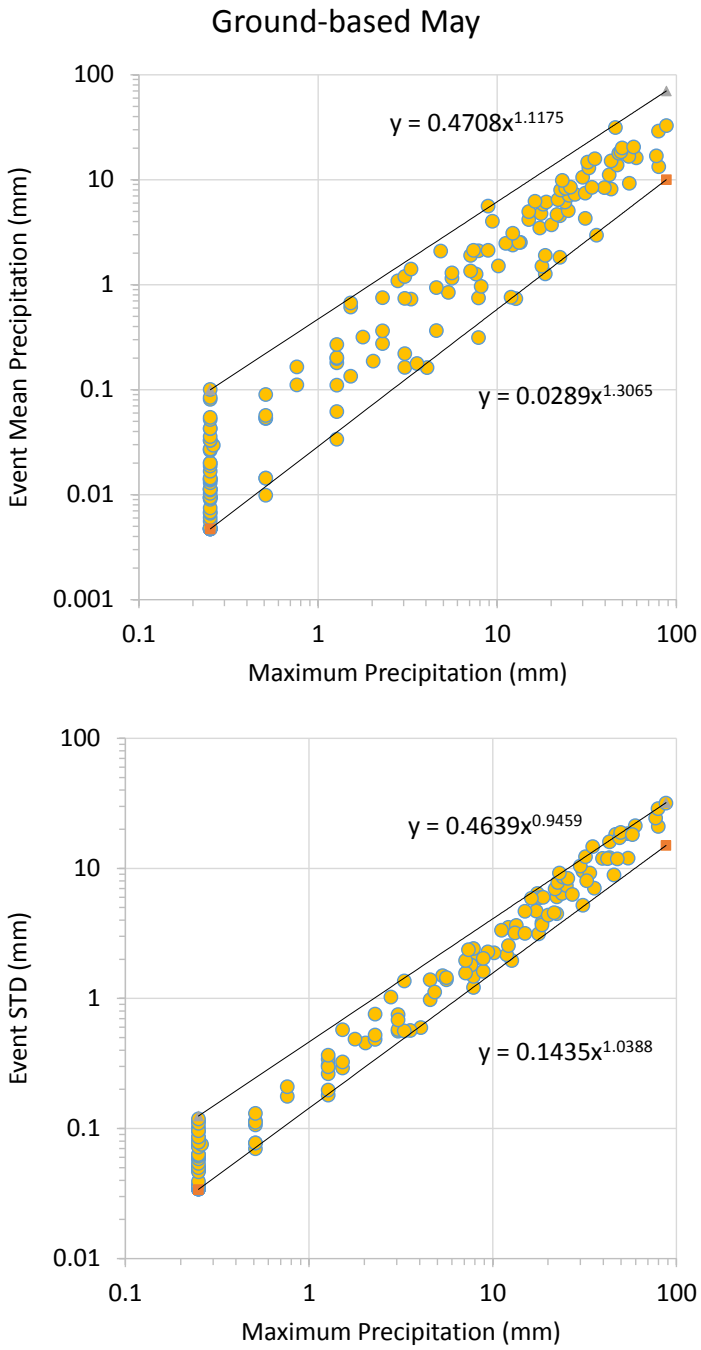


Figure H.5. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in May from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in May from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

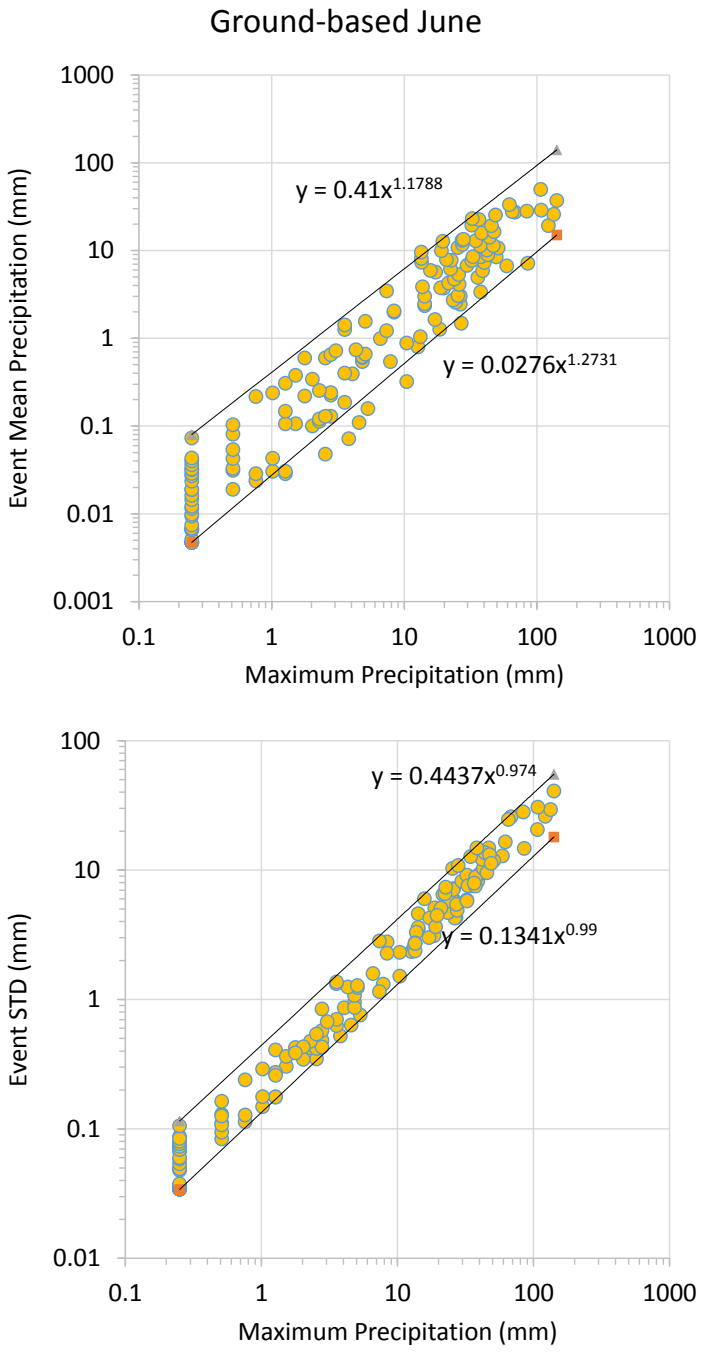


Figure H.6. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in June from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in June from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

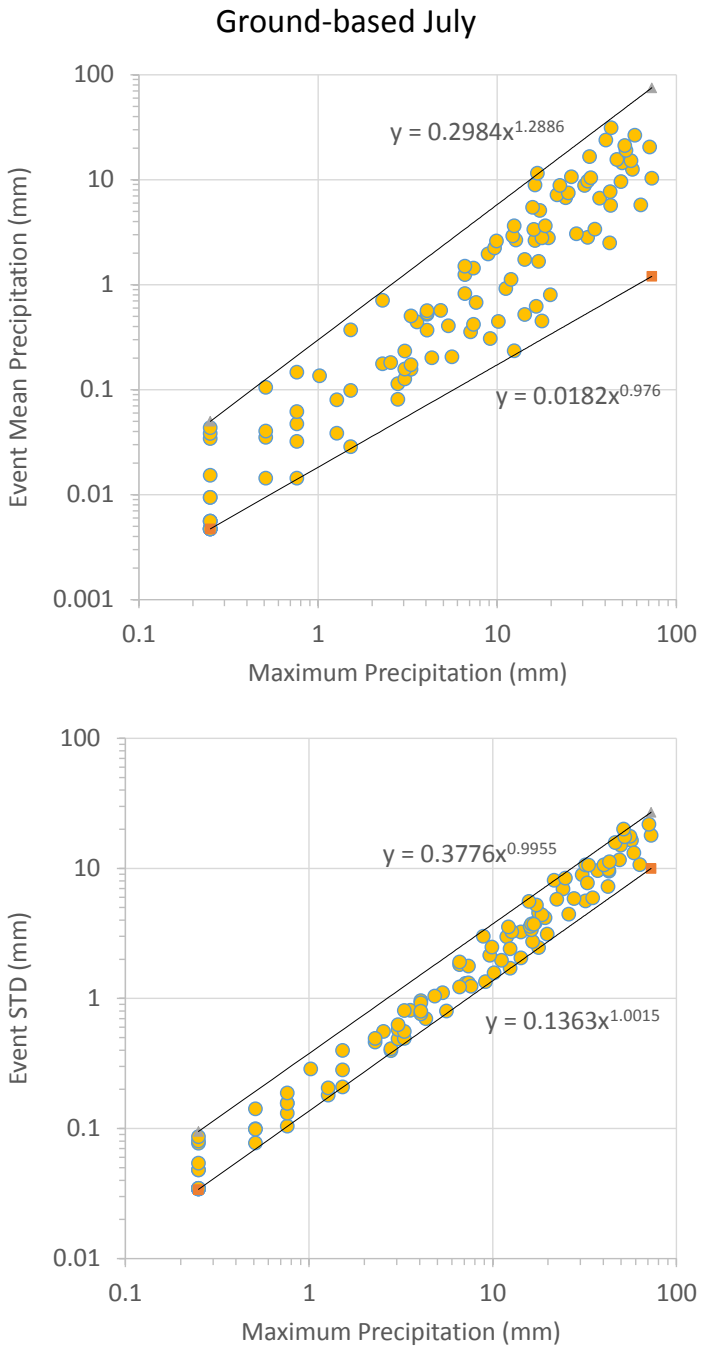


Figure H.7. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in July from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in July from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

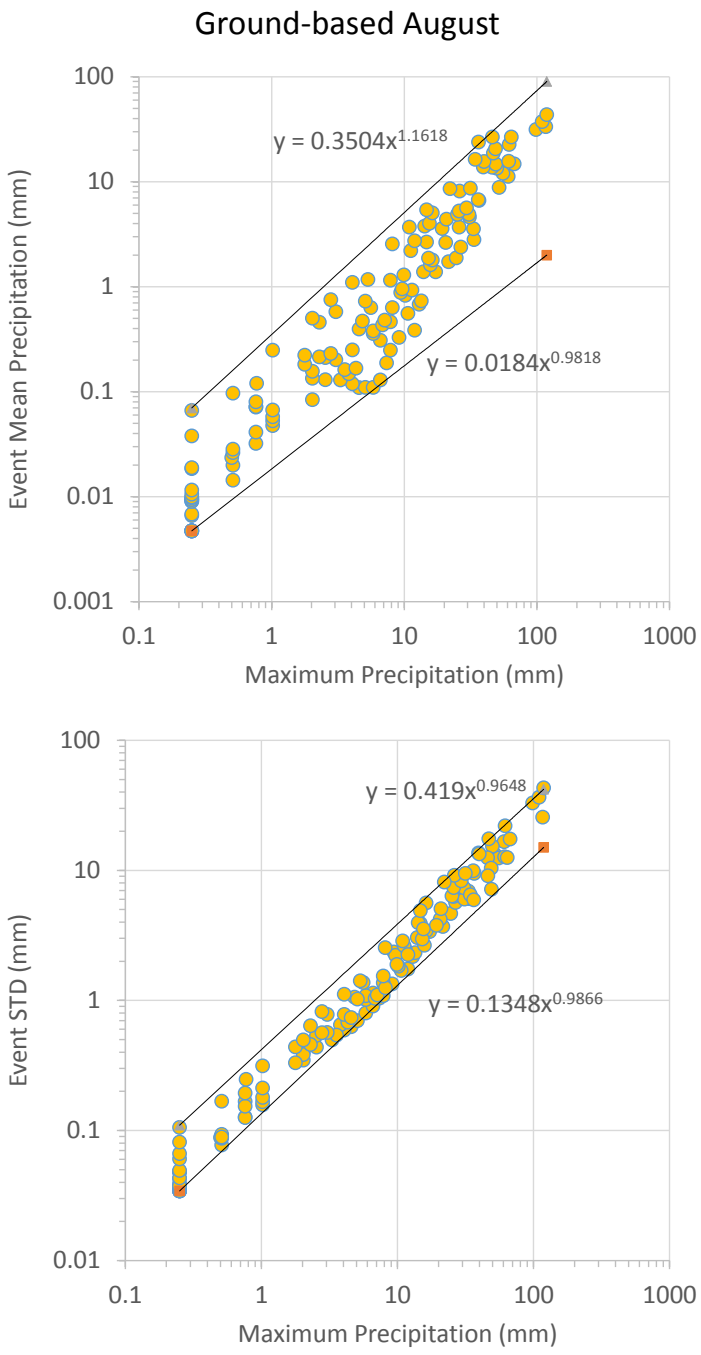


Figure H.8. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in August from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in August from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

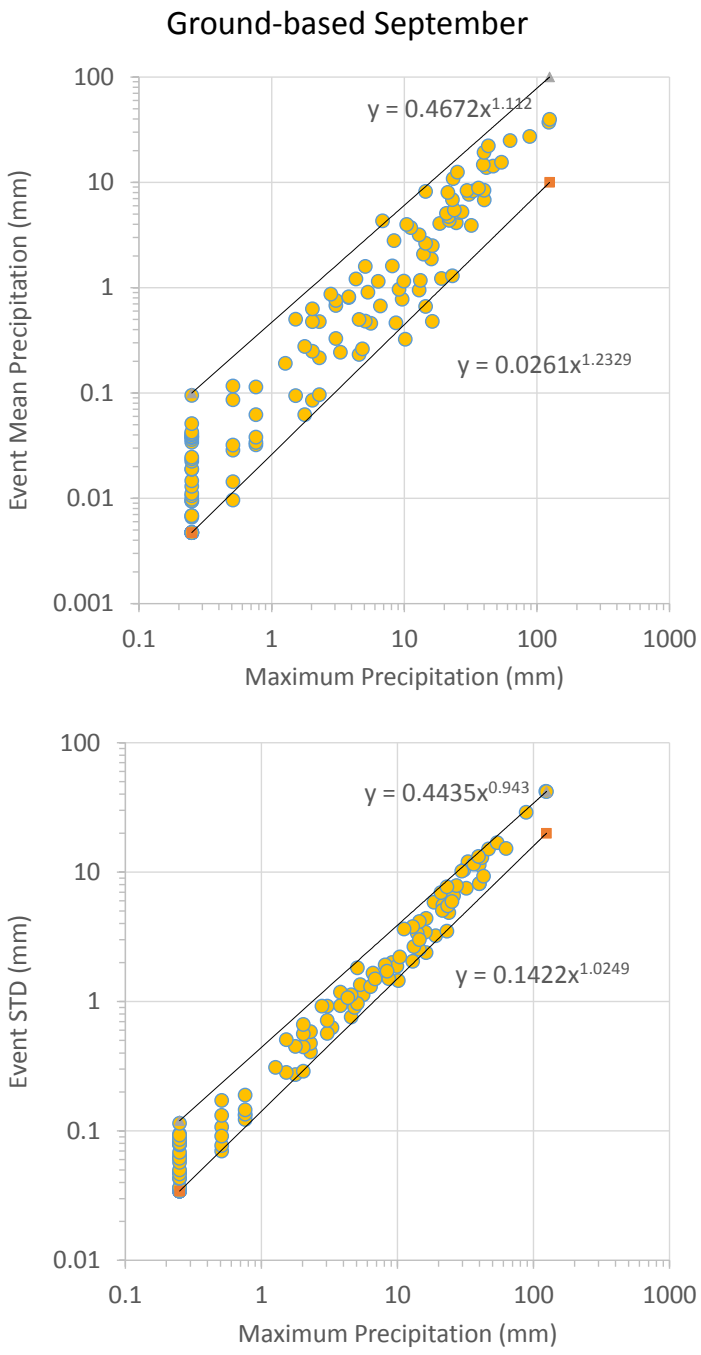


Figure H.9. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in September from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in September from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

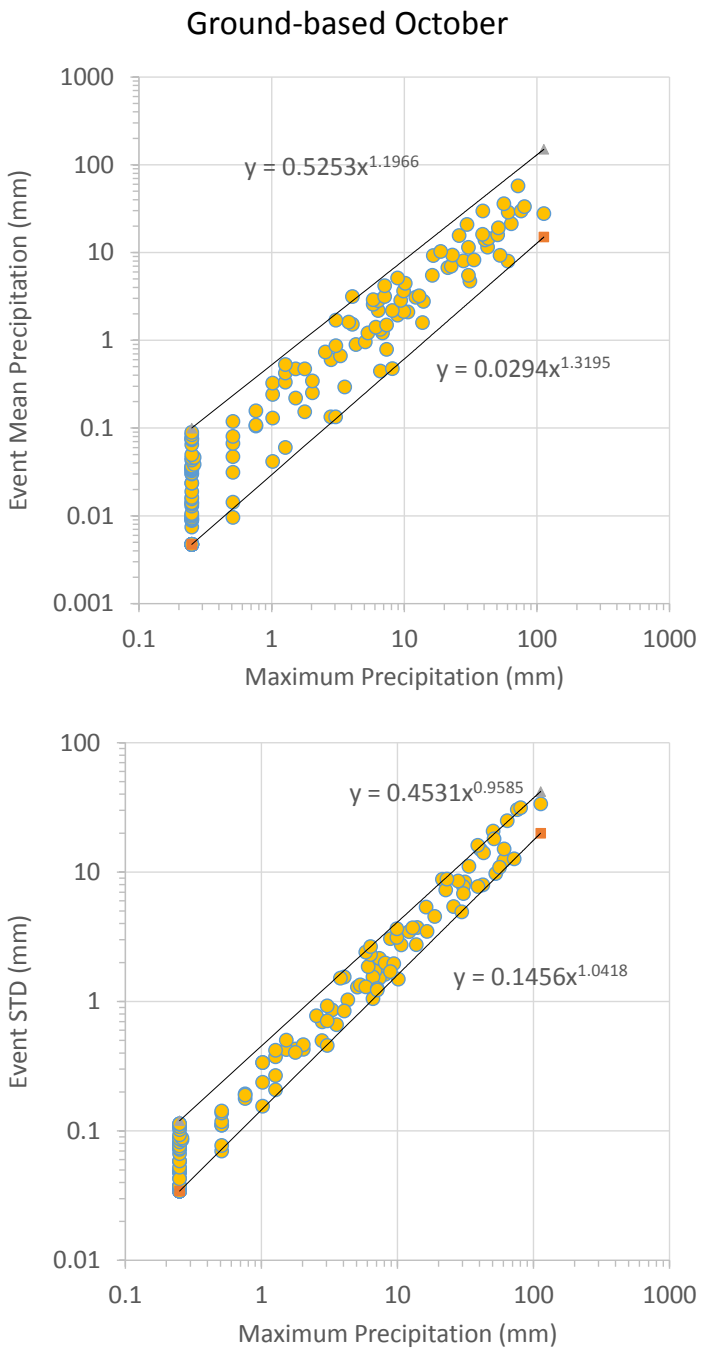


Figure H.10. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in October from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in October from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

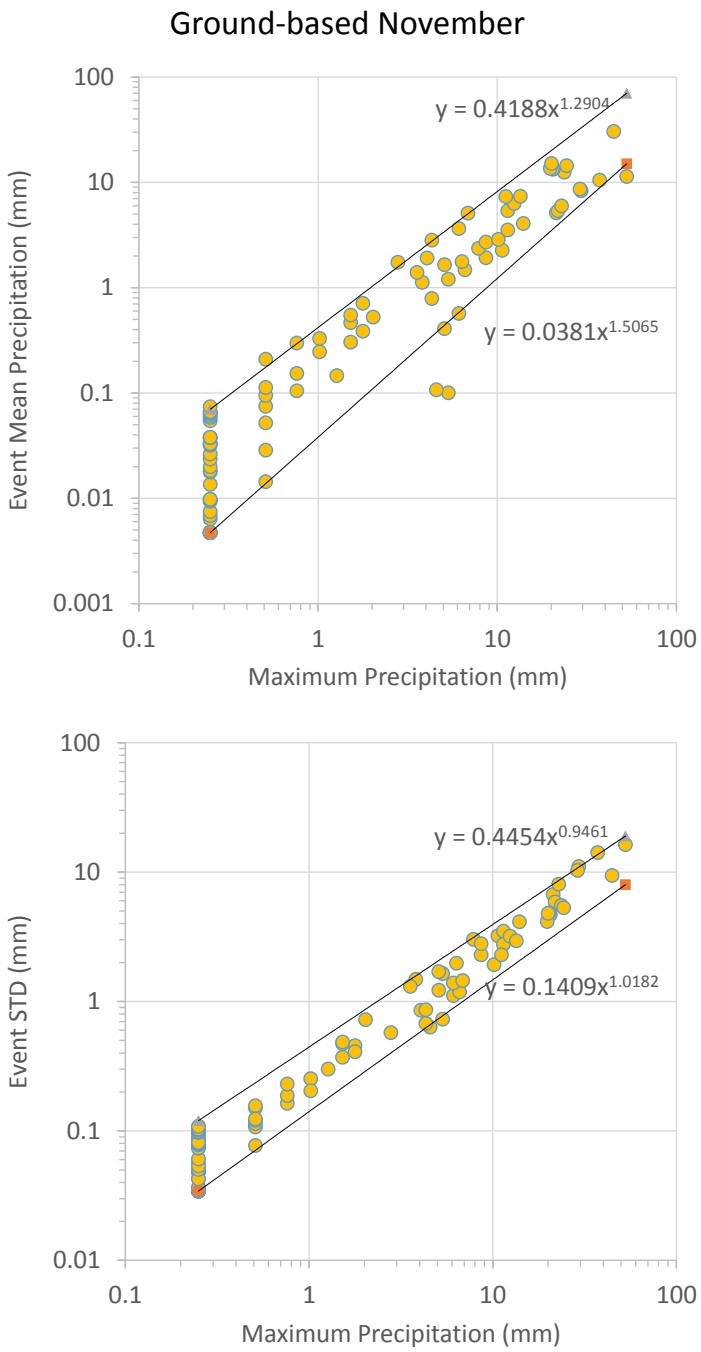


Figure H.11. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in November from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in November from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

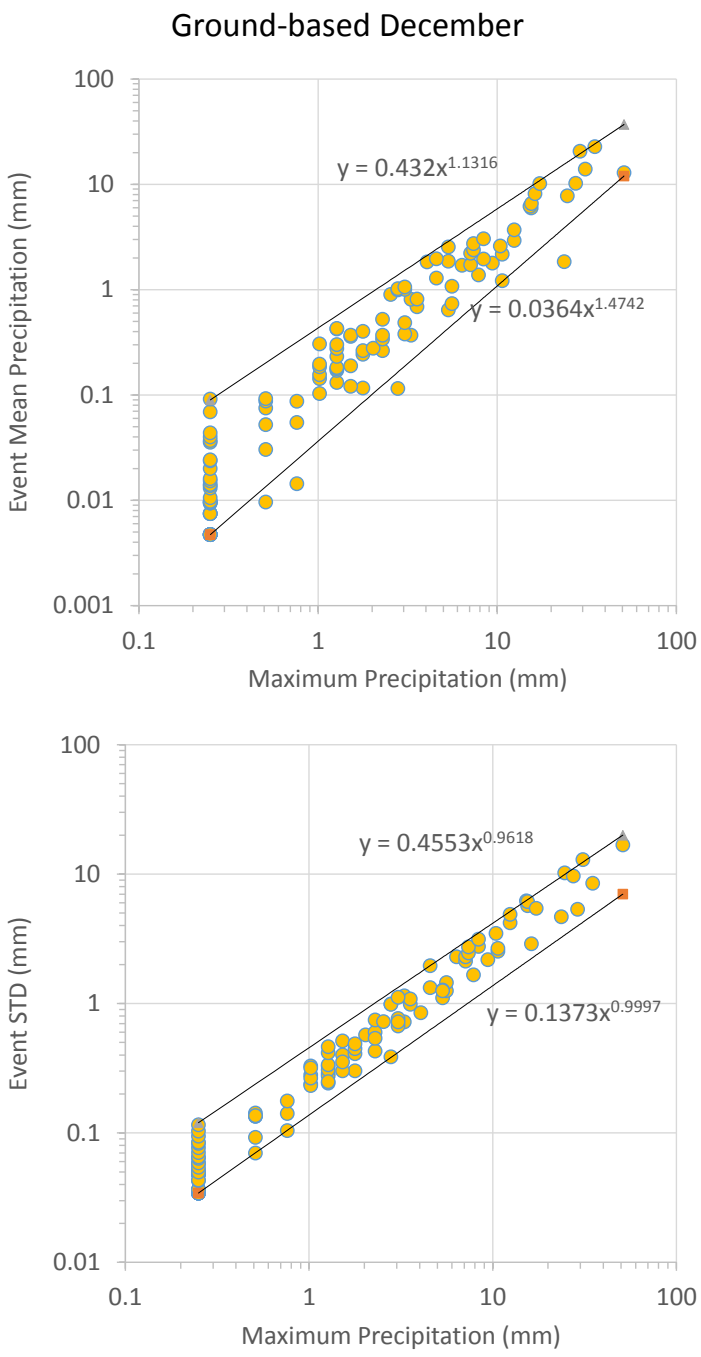


Figure H.12. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in December from 2002 to 2012 from ground-based data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in December from 2002 to 2012 from the ground-based data set and the upper and lower boundary.

APPENDIX I: THE LOG-SCALE FIGURES OF KFDR DATASET OF TWELVE MONTHS

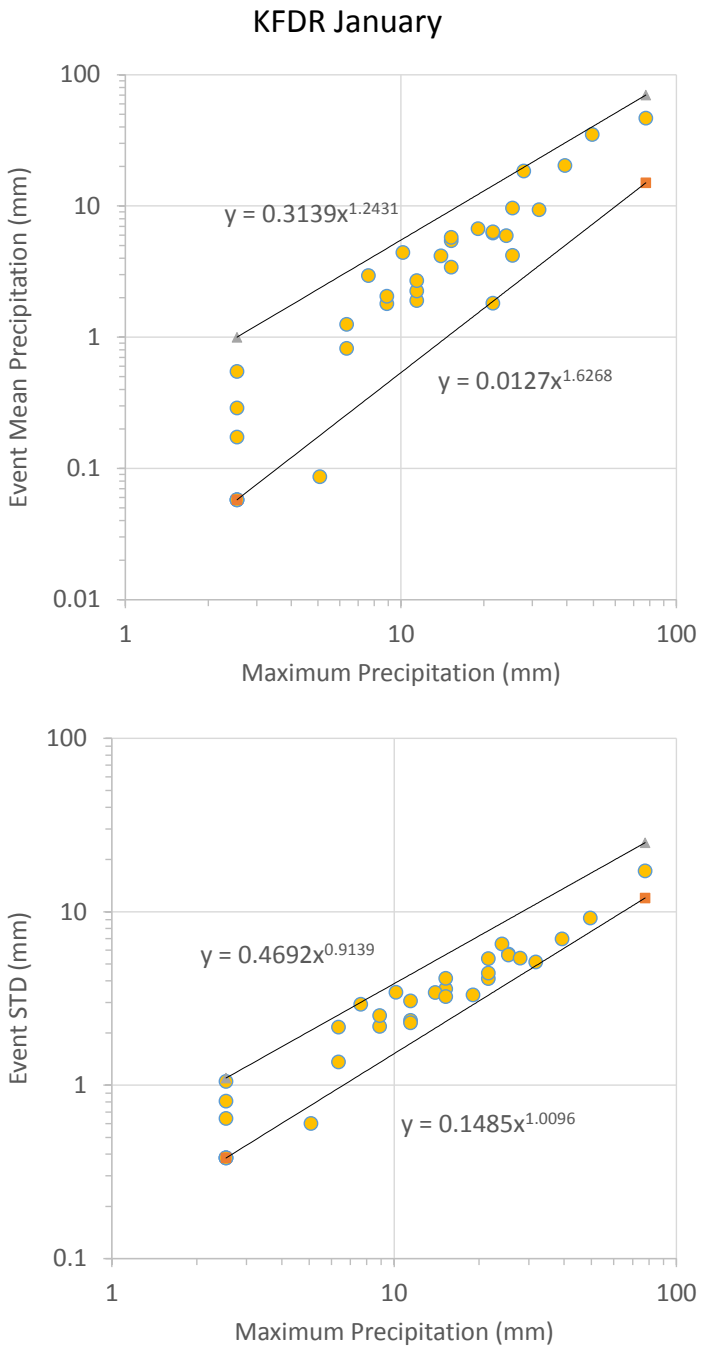


Figure I.1. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in January from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in January from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

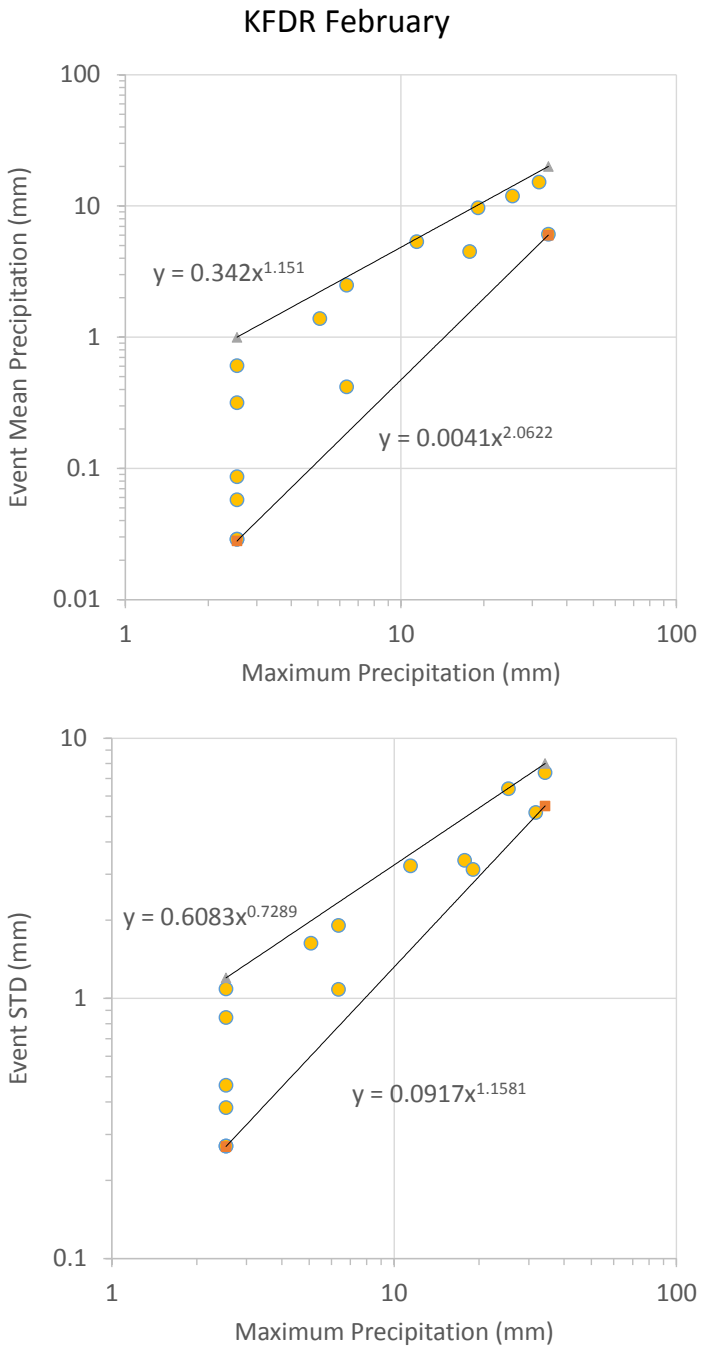


Figure I.2. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in February from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in February from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

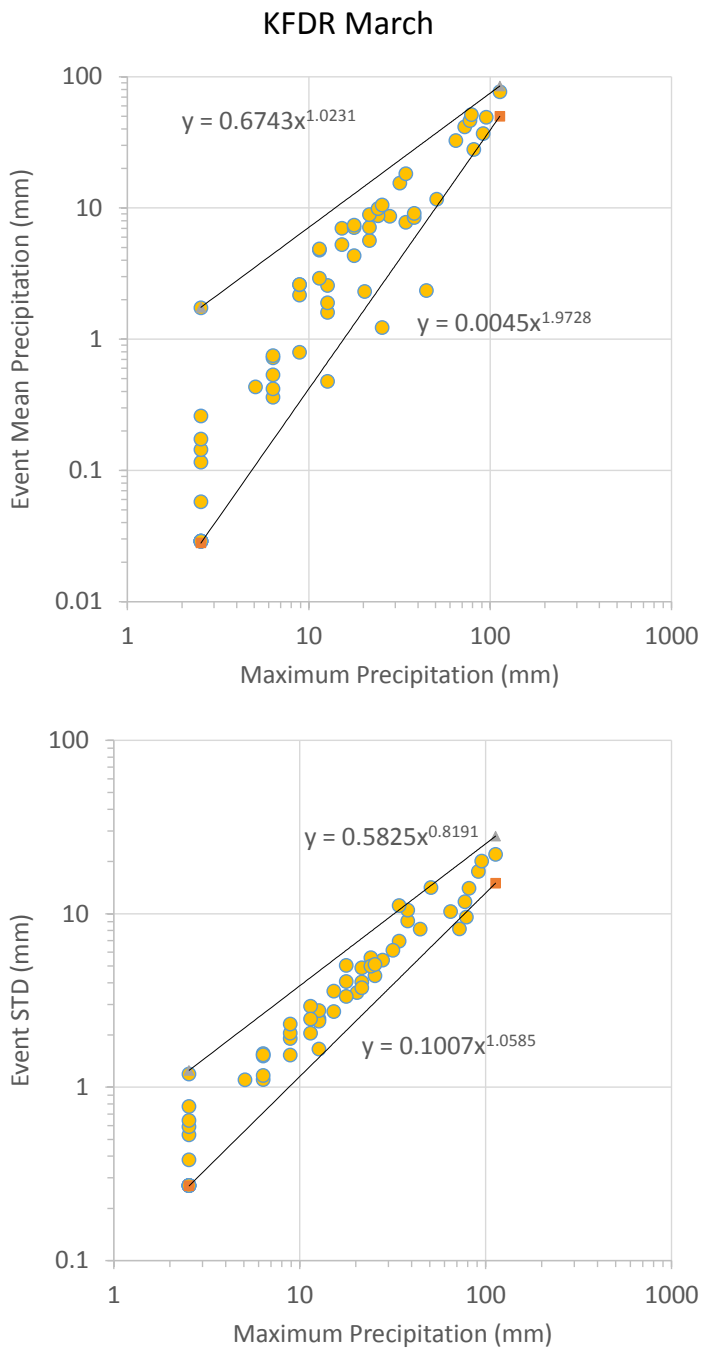


Figure I.3. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in March from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in March from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

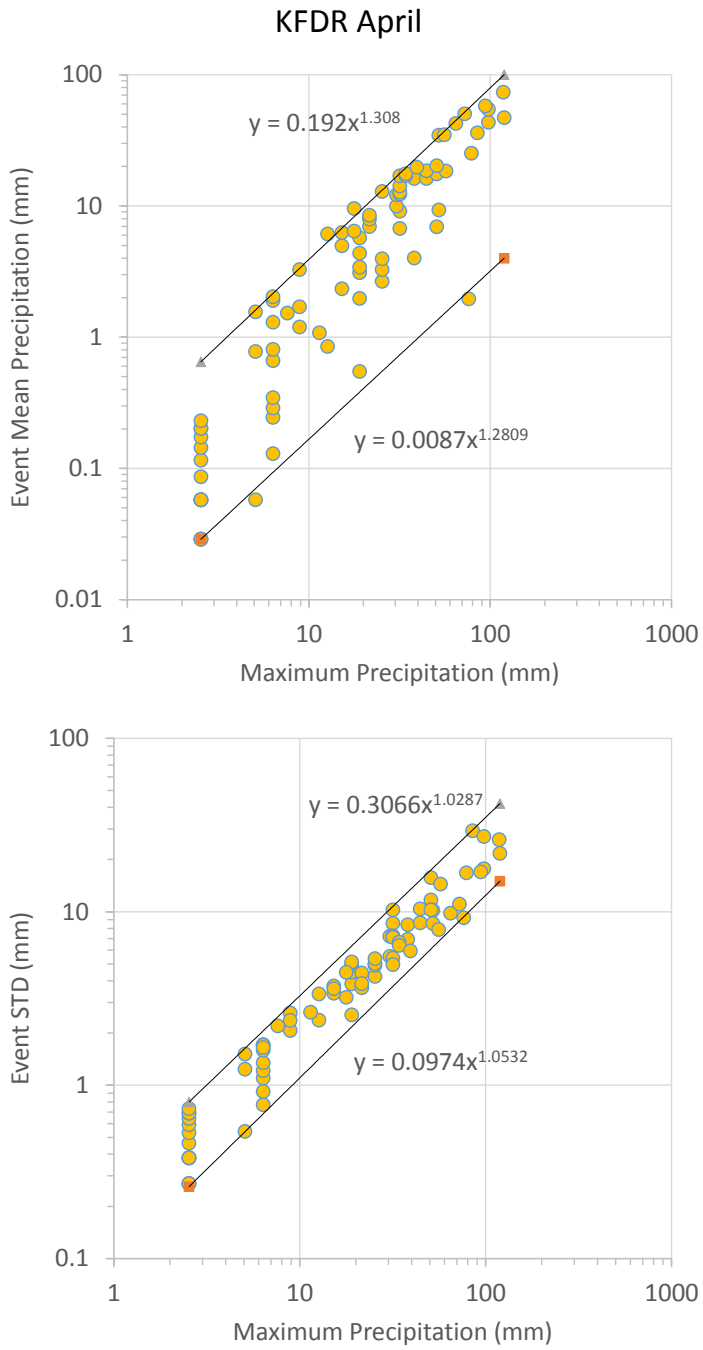


Figure I.4. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in April from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in April from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

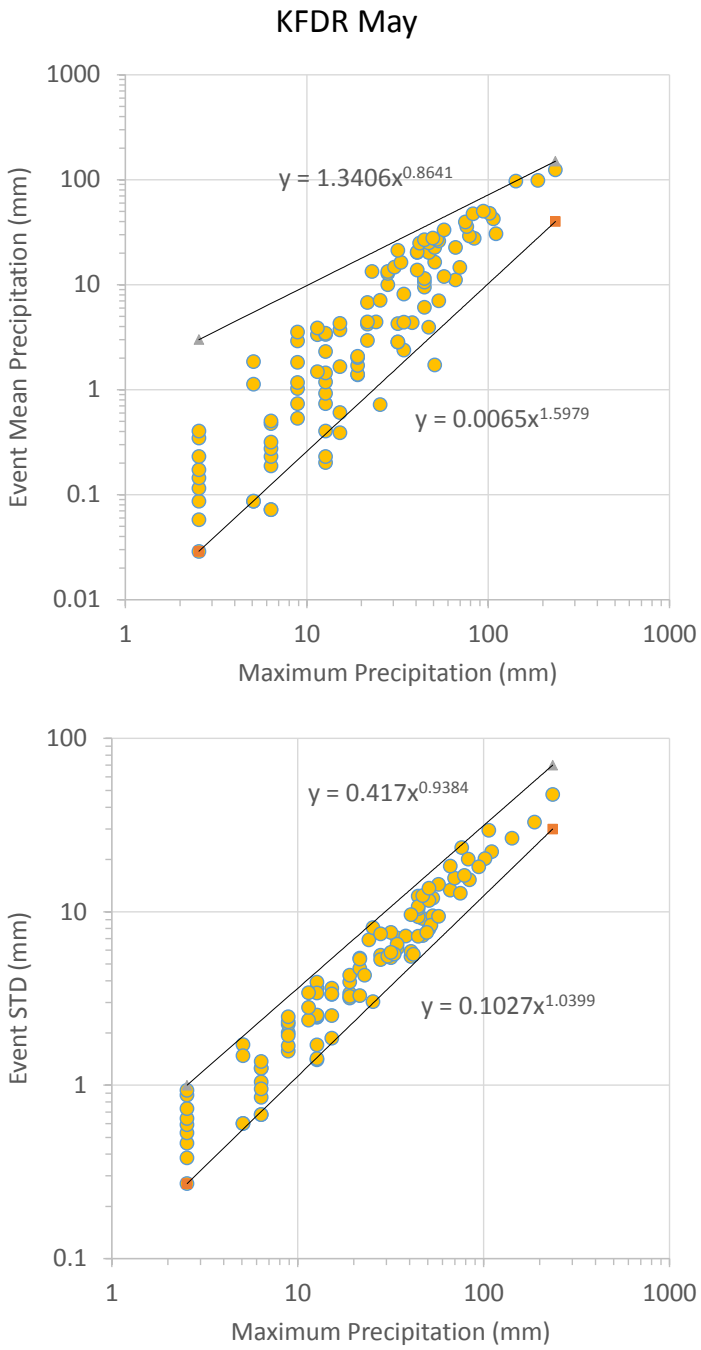


Figure I.5. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in May from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in May from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

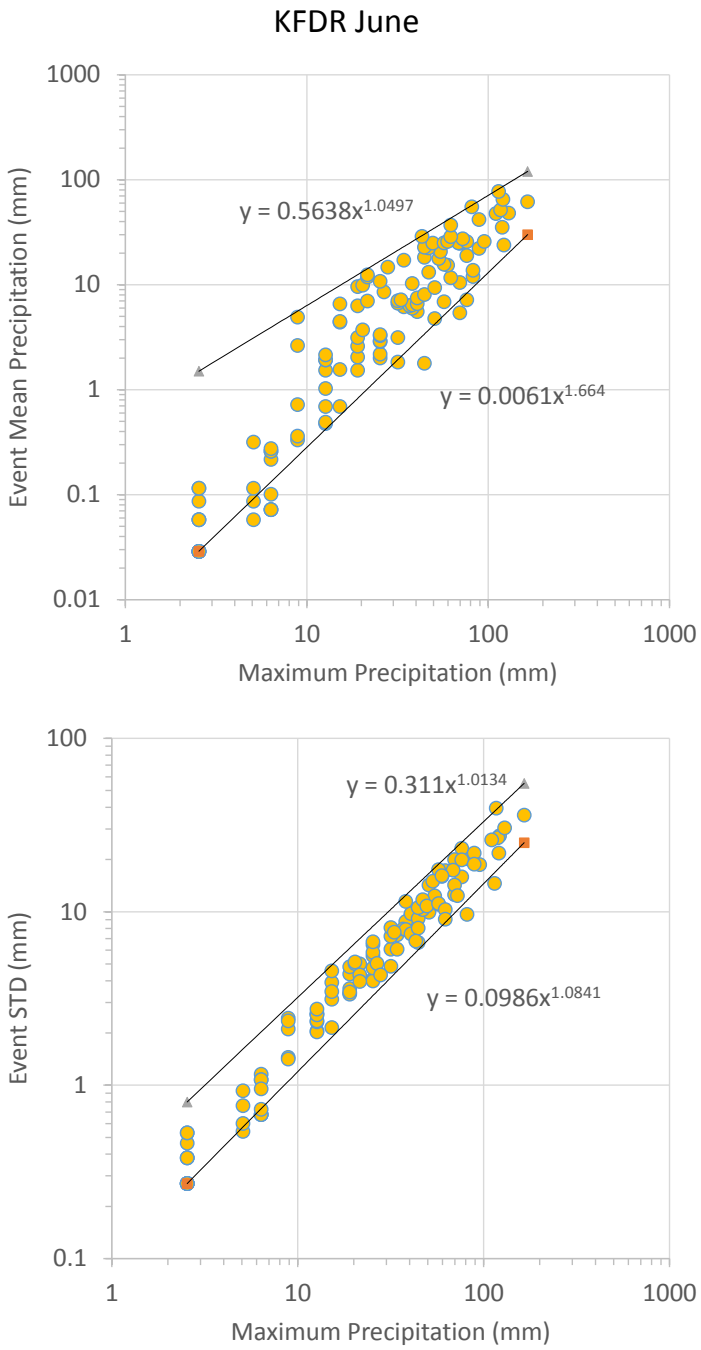


Figure I.6. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in June from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in June from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

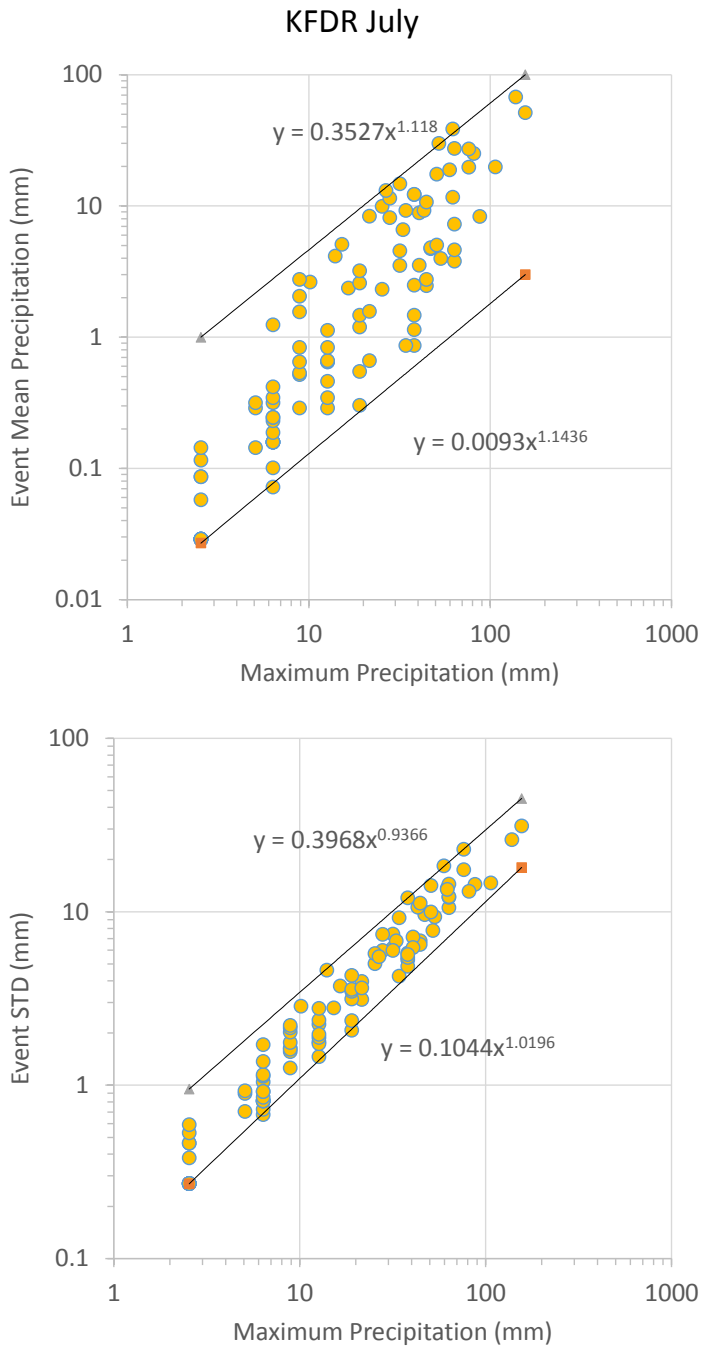


Figure I.7. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in July from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in July from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

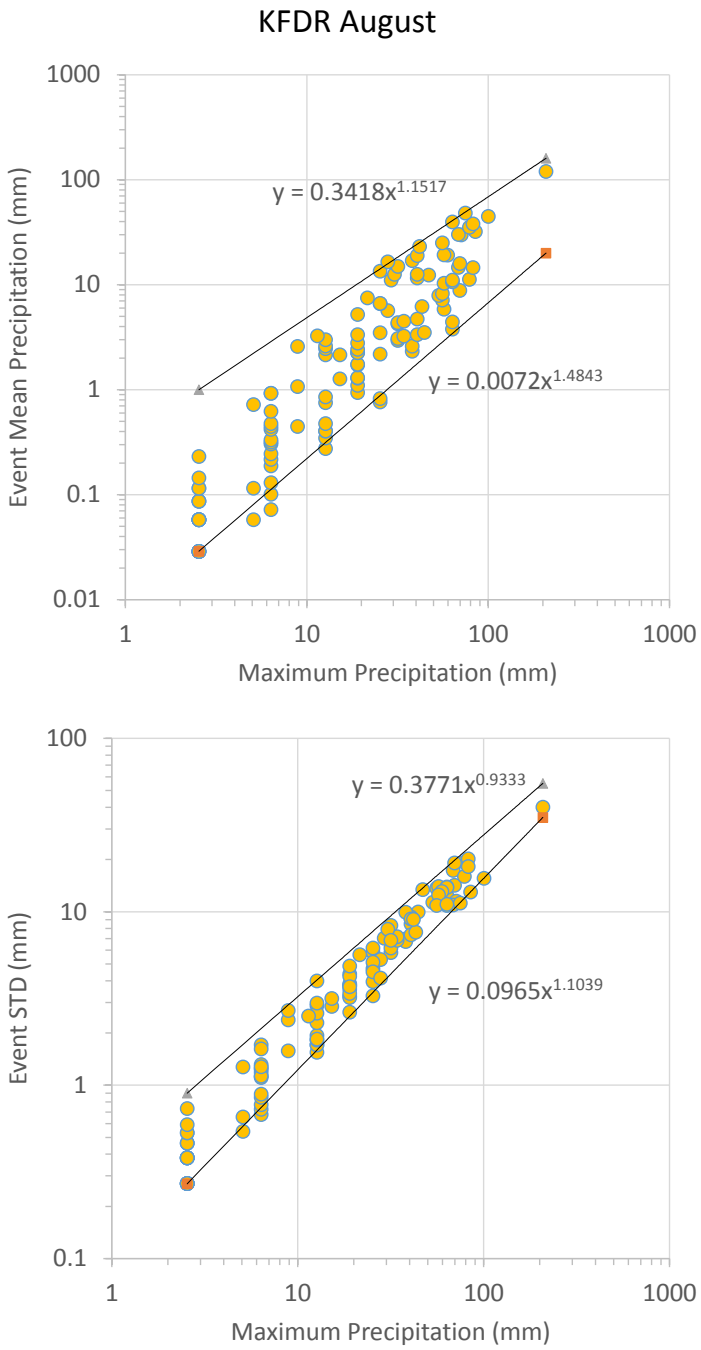


Figure I.8. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in August from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in August from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

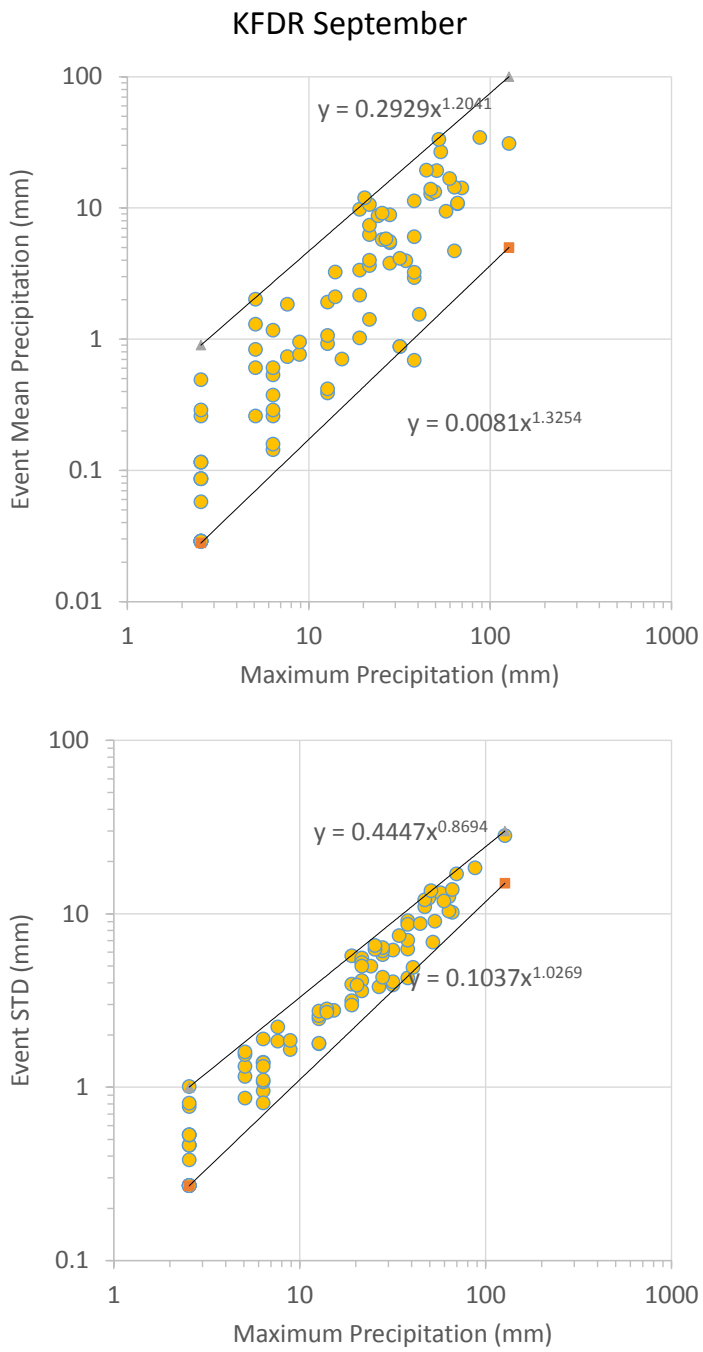


Figure I.9. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in September from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in September from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

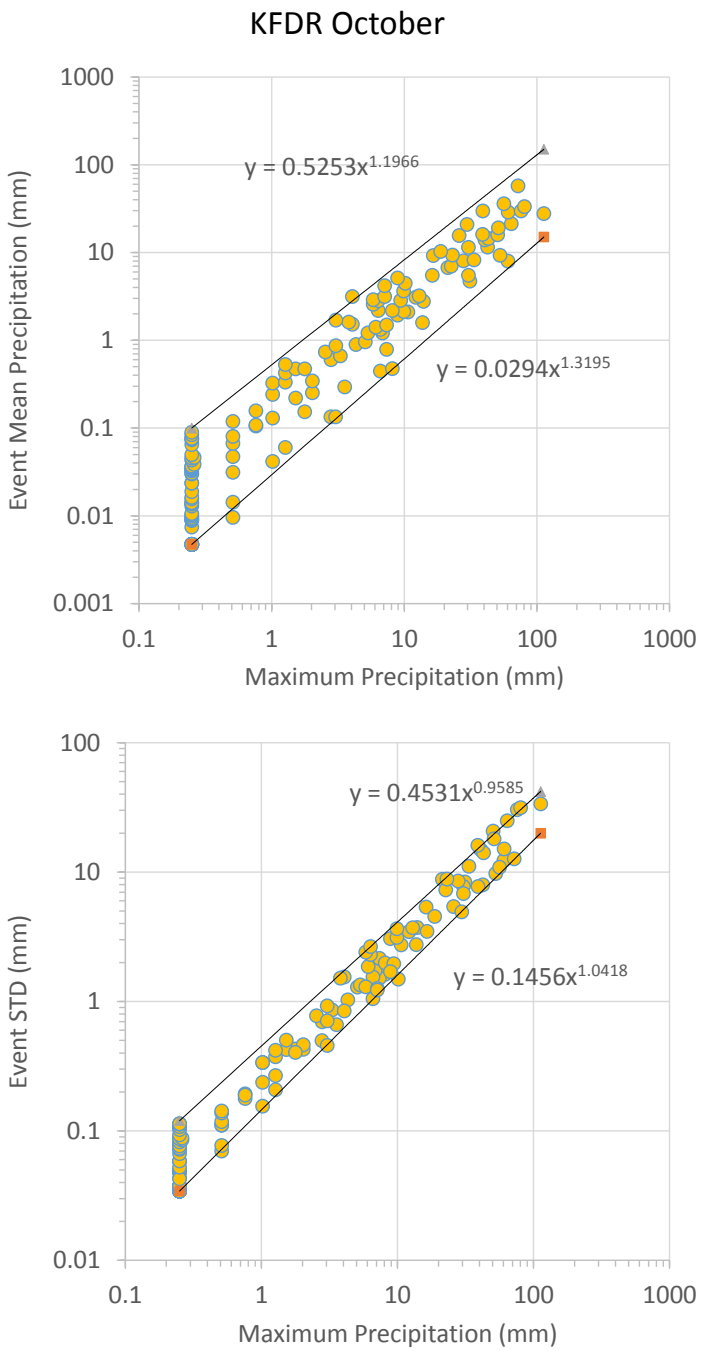


Figure I.10. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in October from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in October from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

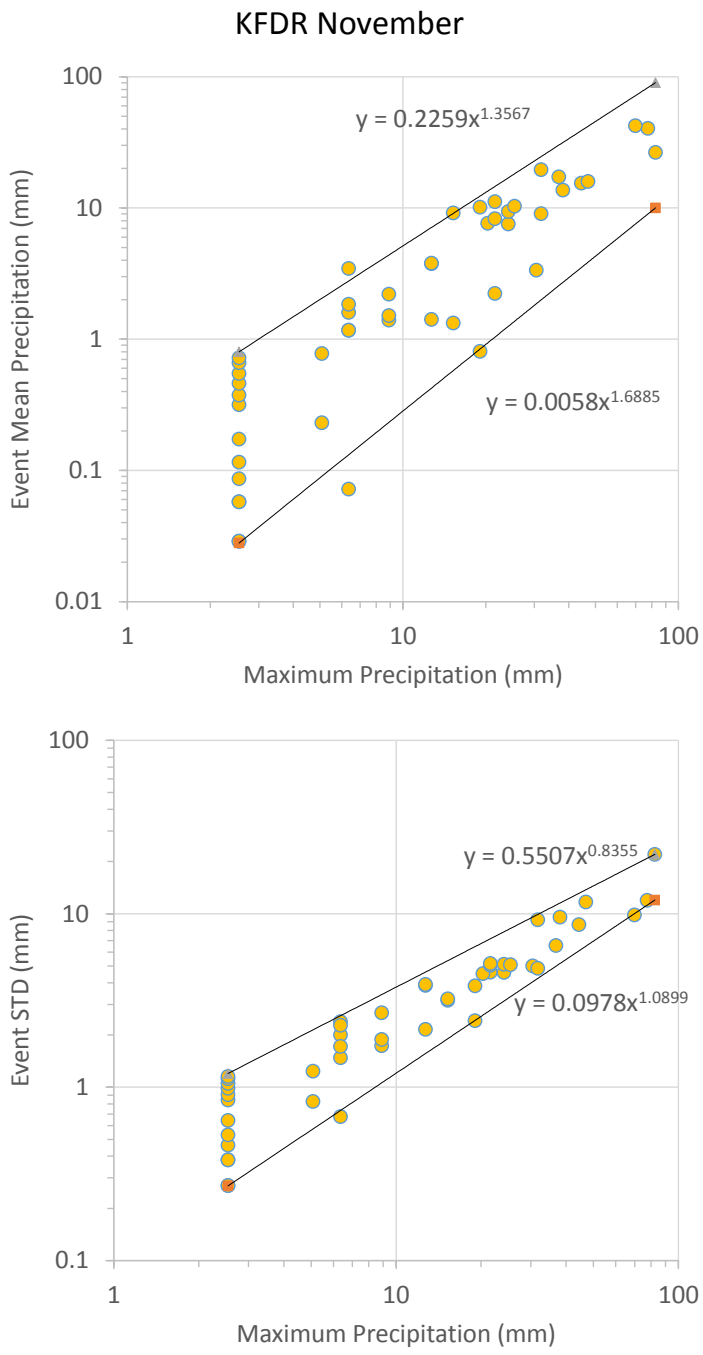


Figure I.11. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in November from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in November from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

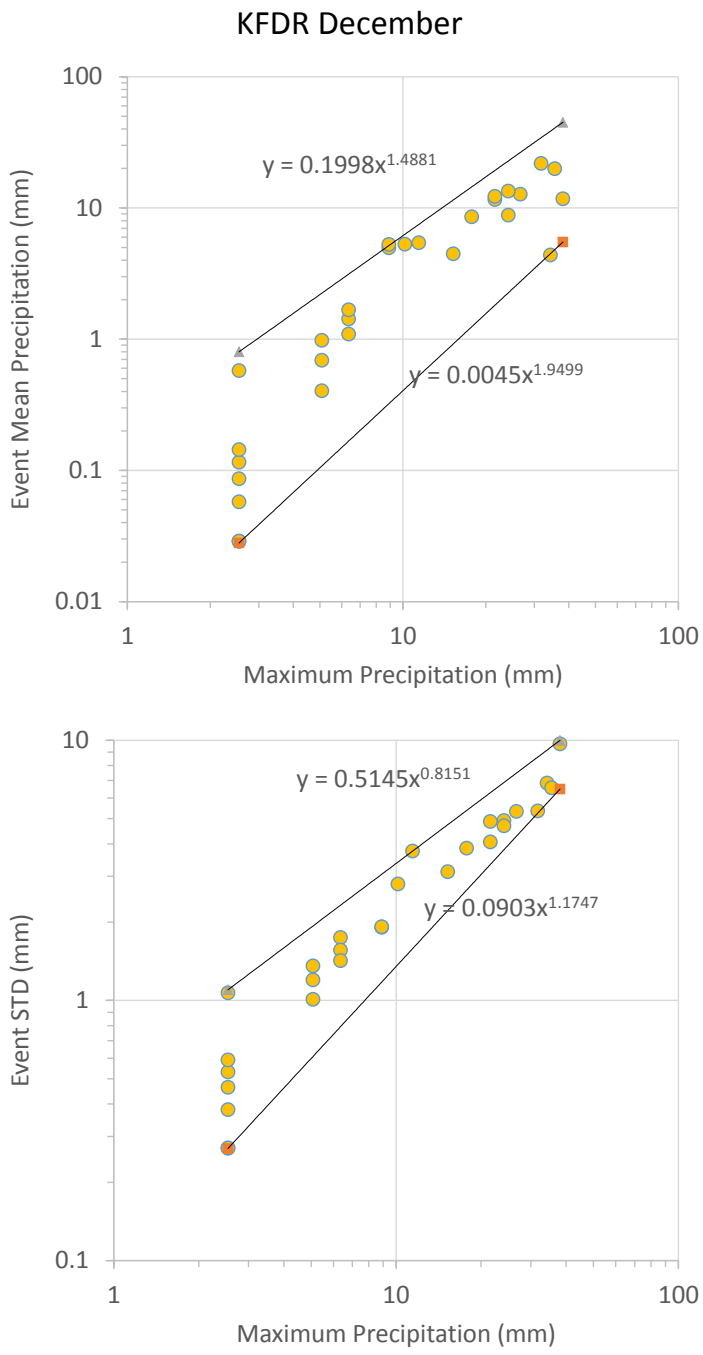


Figure I.12. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in December from 2002 to 2012 from KFDR data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in December from 2002 to 2012 from the KFDR data set and the upper and lower boundary.

APPENDIX J: THE LOG-SCALE FIGURES OF KTLX DATASET OF TWELVE MONTHS

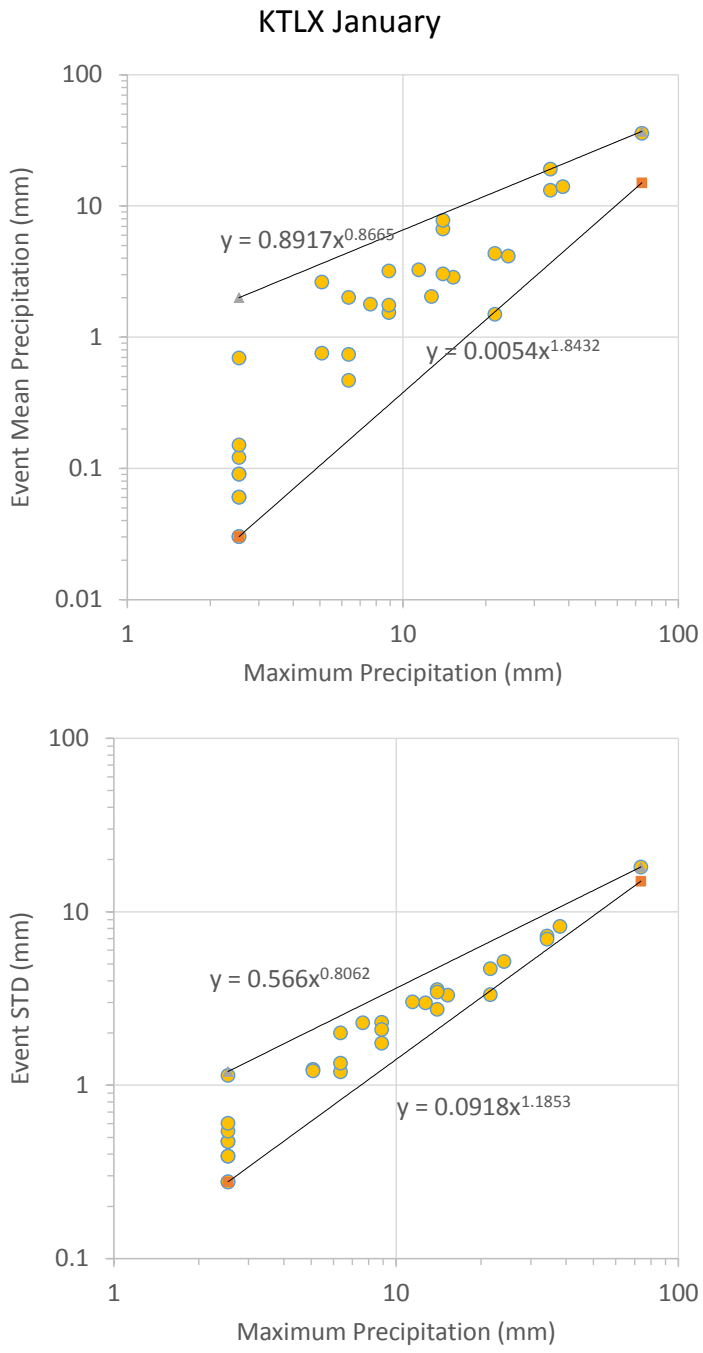


Figure J.1. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in January from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in January from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

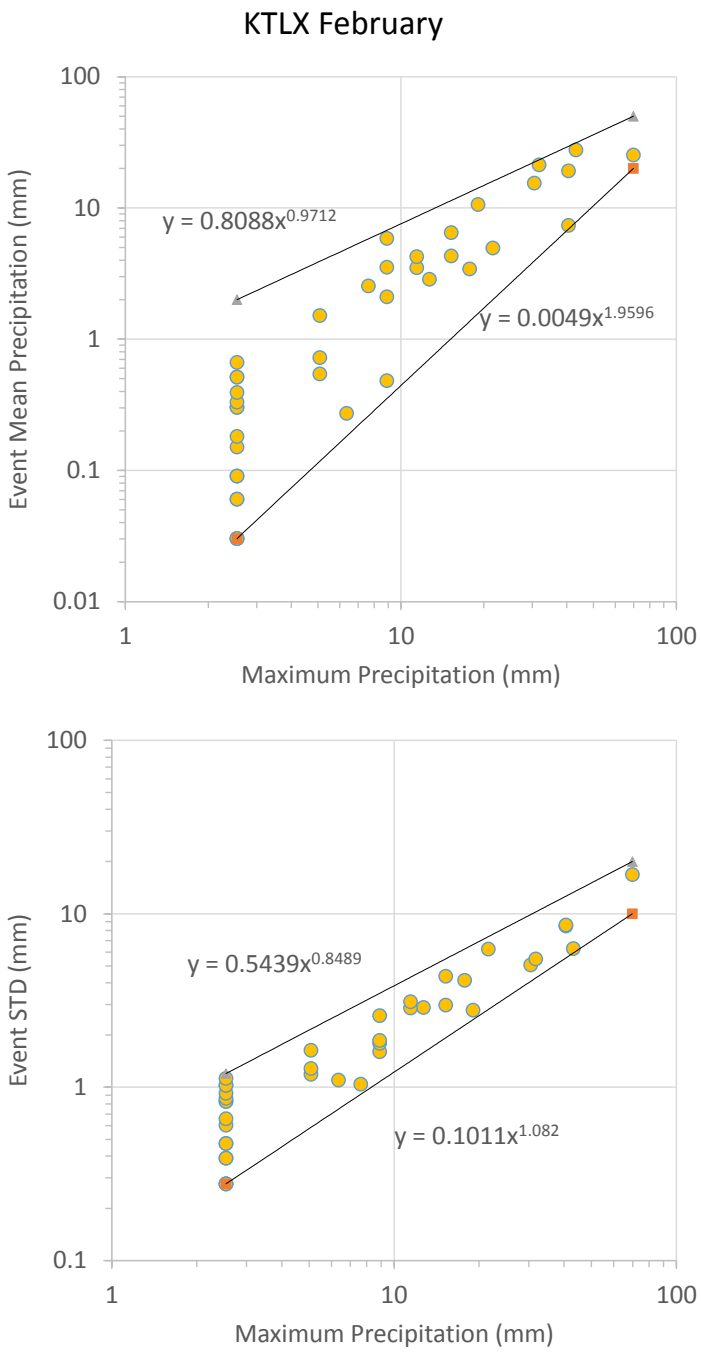


Figure J.2. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in February from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in February from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

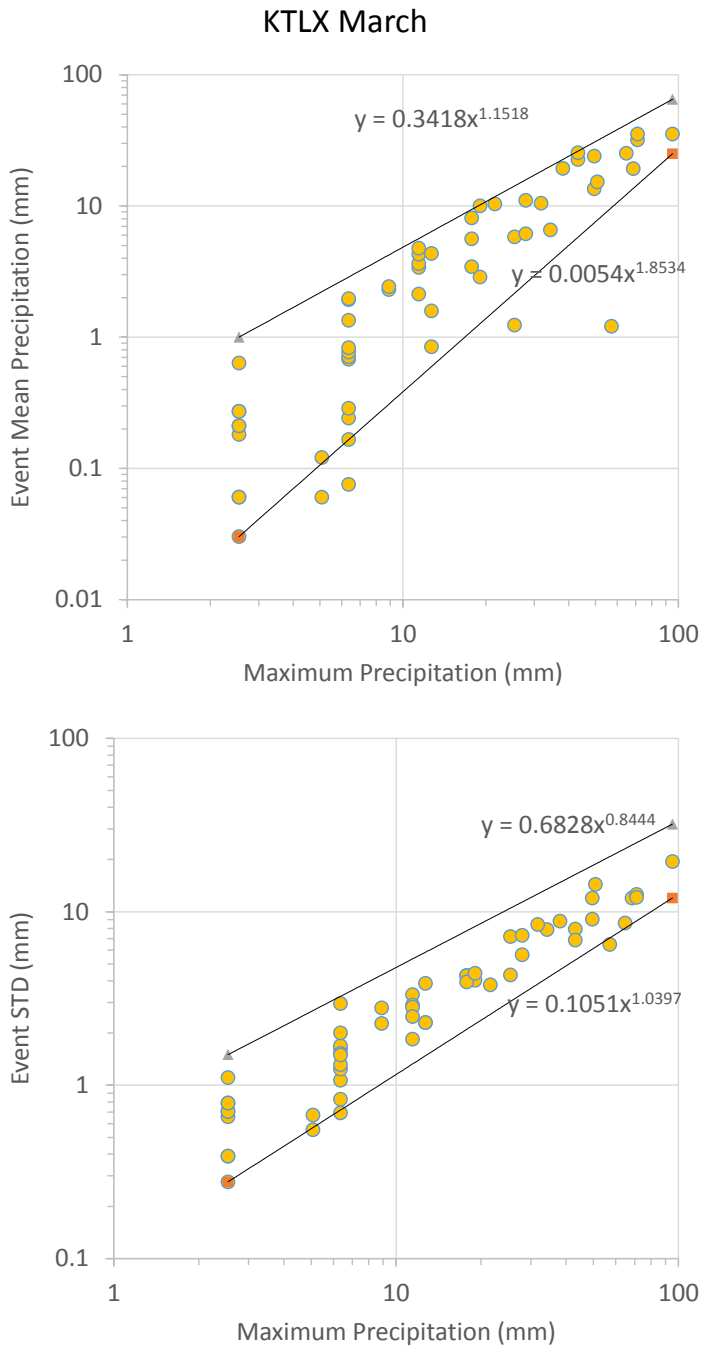


Figure J.3. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in March from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in March from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

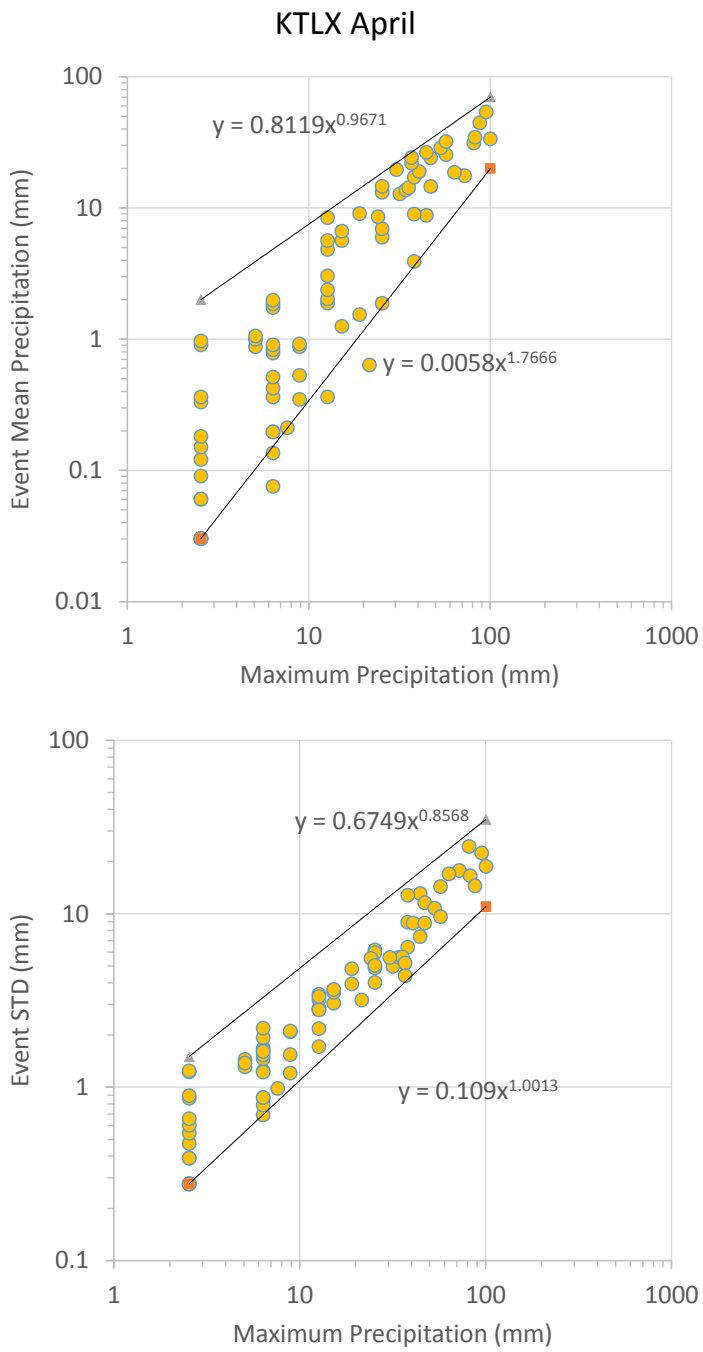


Figure J.4. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in April from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in April from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

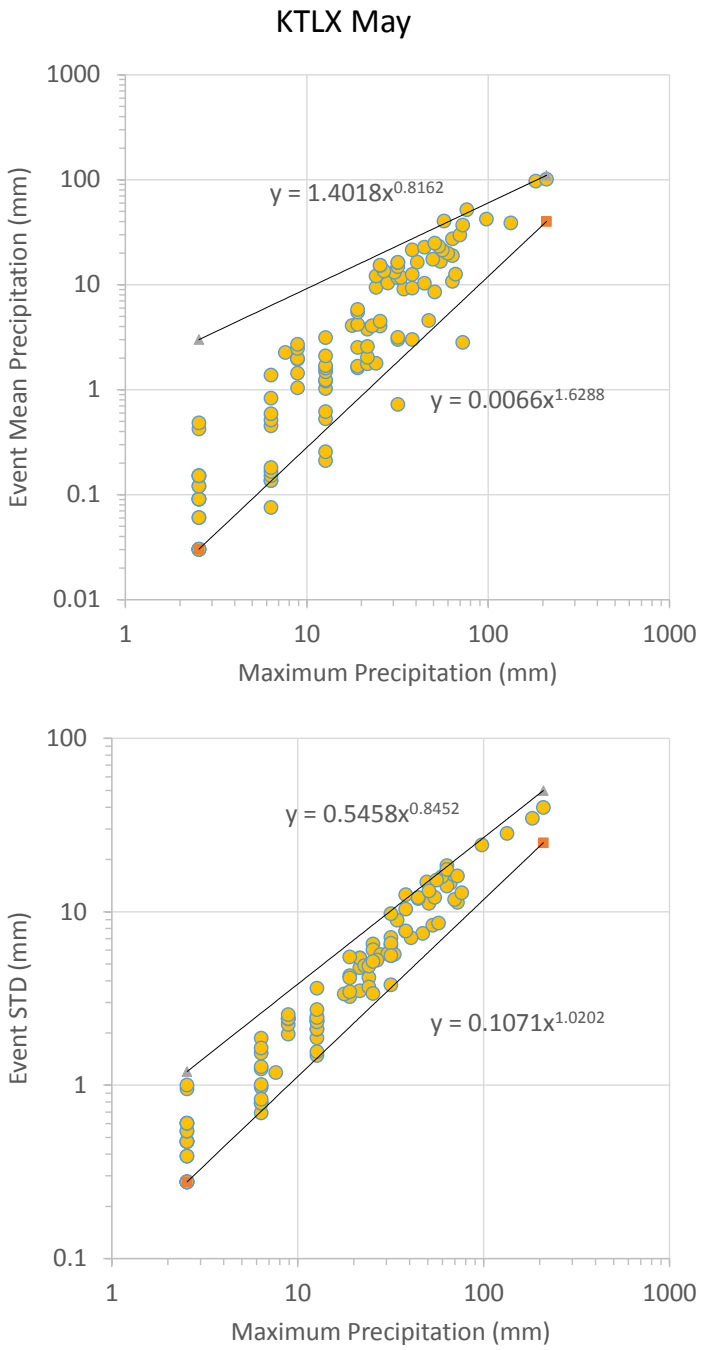


Figure J.5. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in May from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in May from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

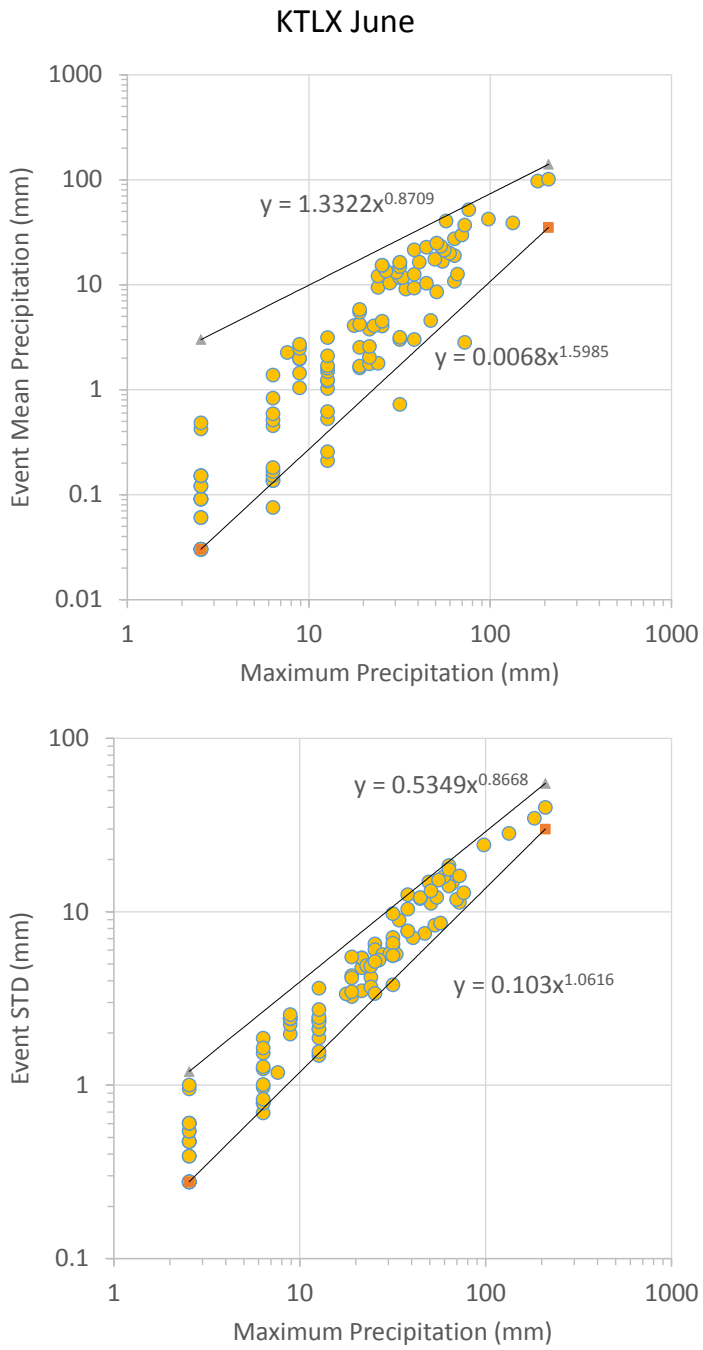


Figure J.6. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in June from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in June from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

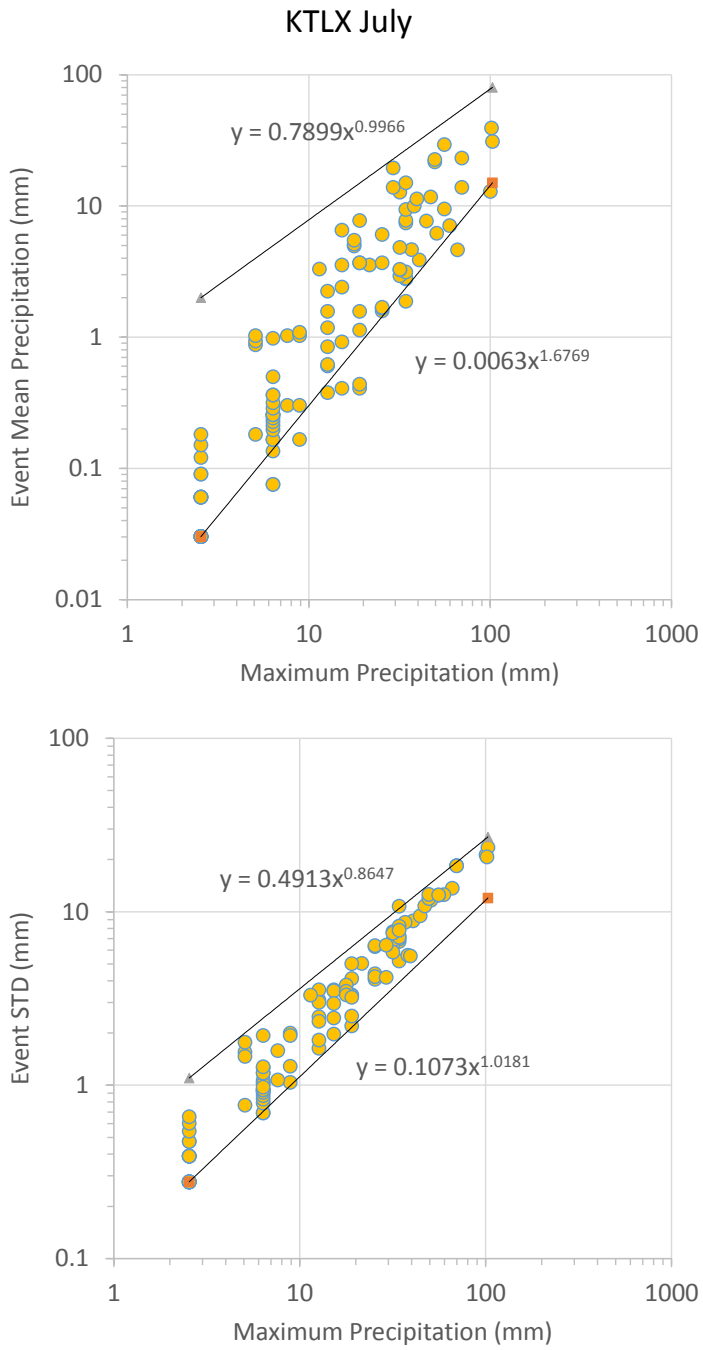


Figure J.7. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in July from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in July from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

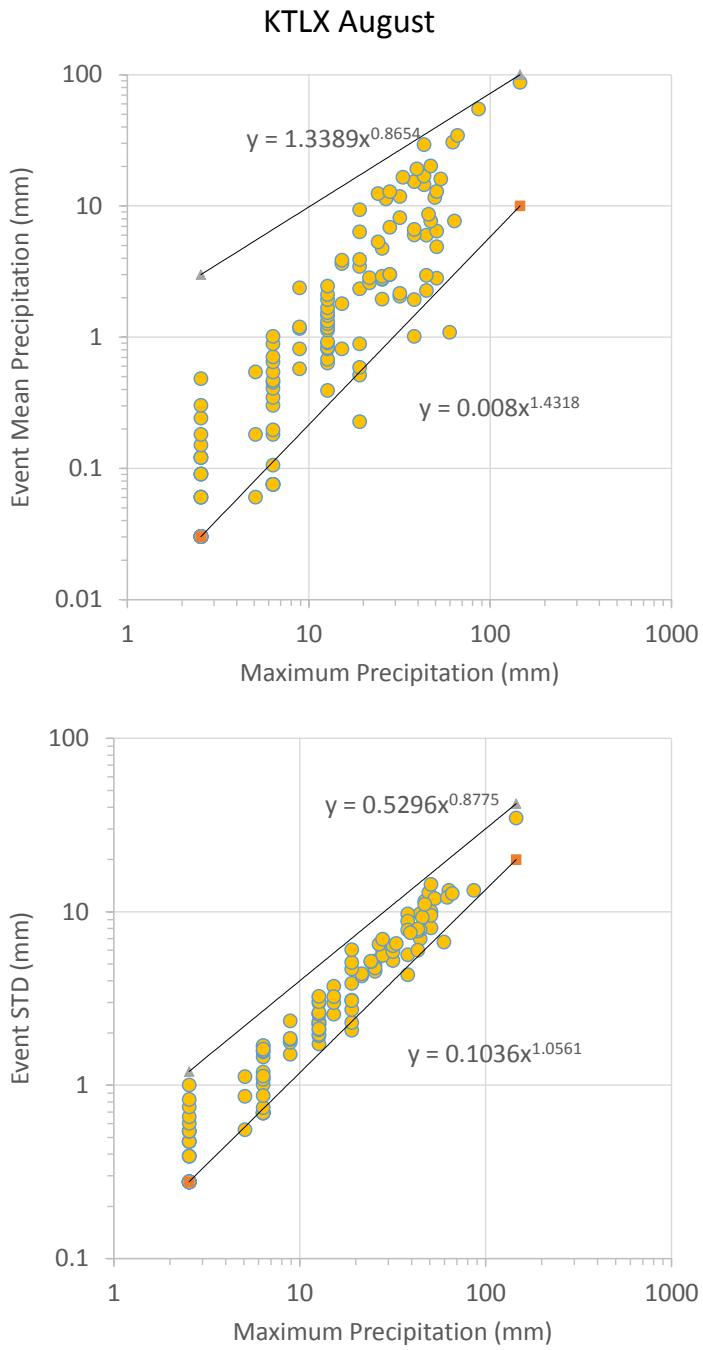


Figure J.8. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in August from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in August from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

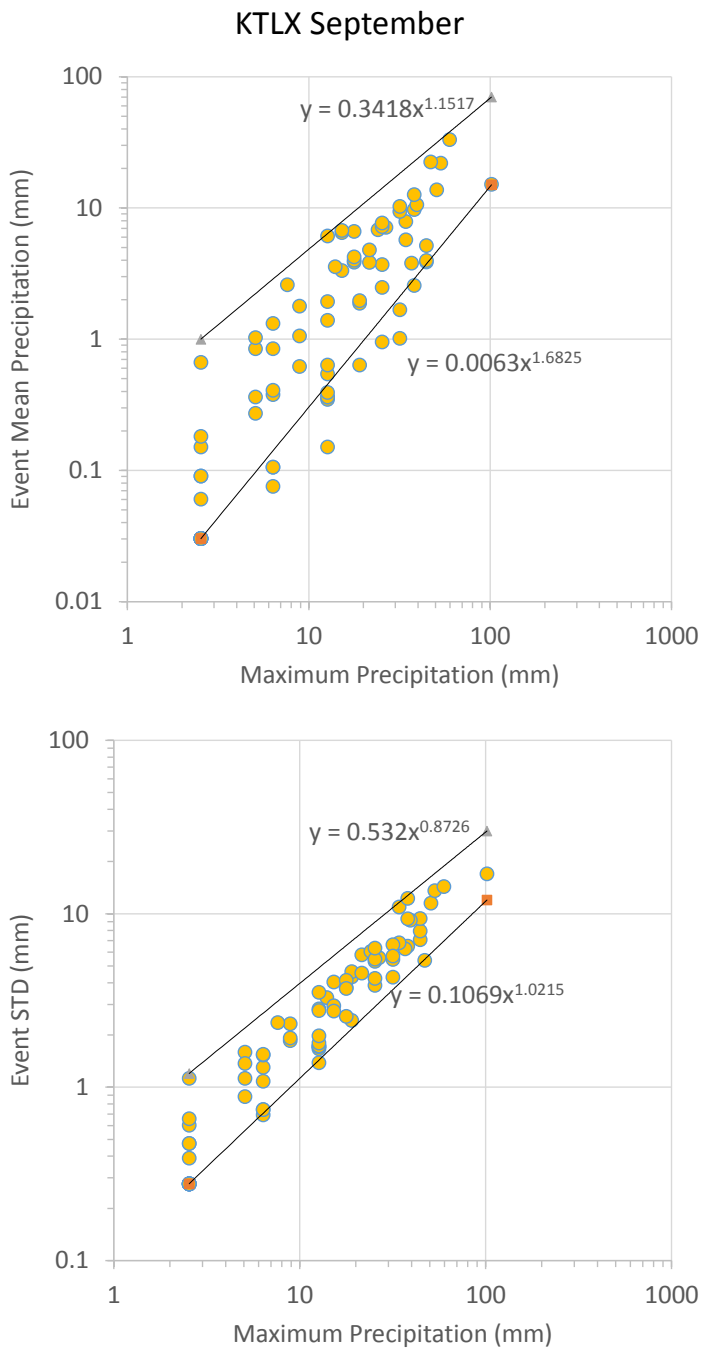


Figure J.9. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in September from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in September from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

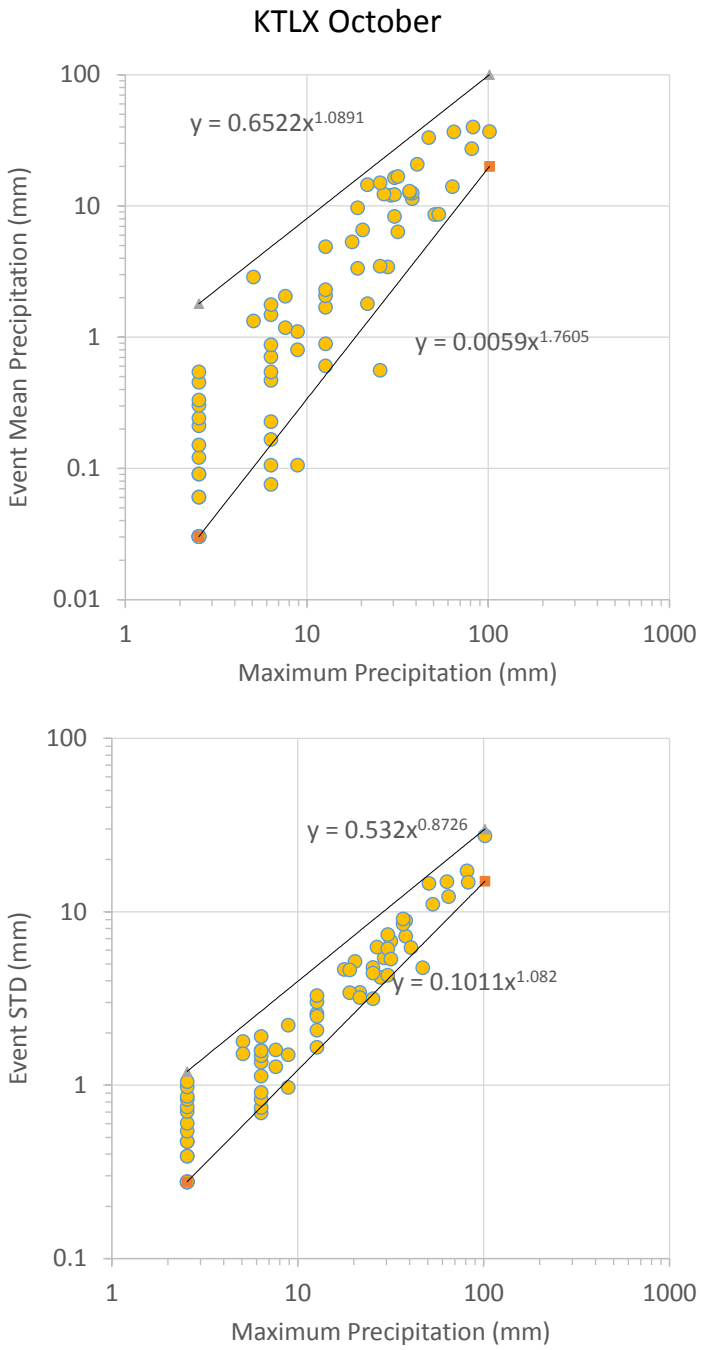


Figure J.10. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in October from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in October from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

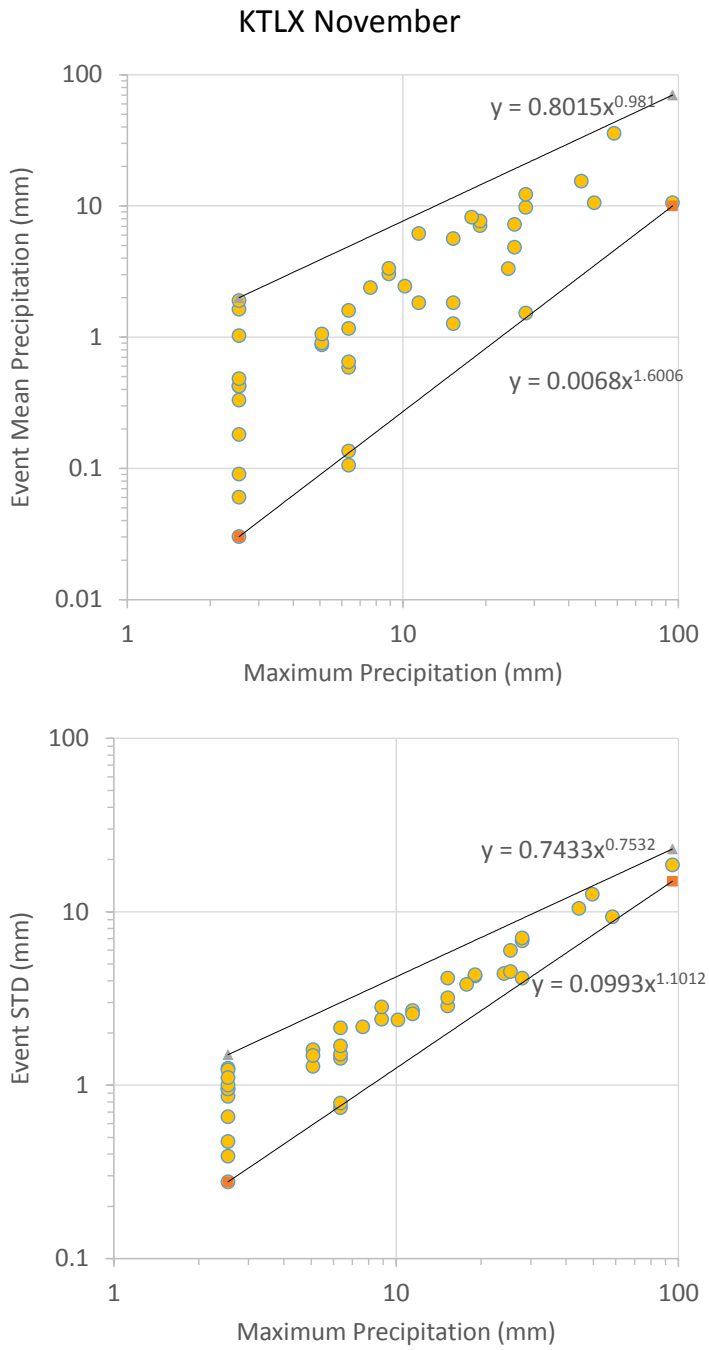


Figure J.11. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in November from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in November from 2002 to 2012 from the KTLX data set and the upper and lower boundary.

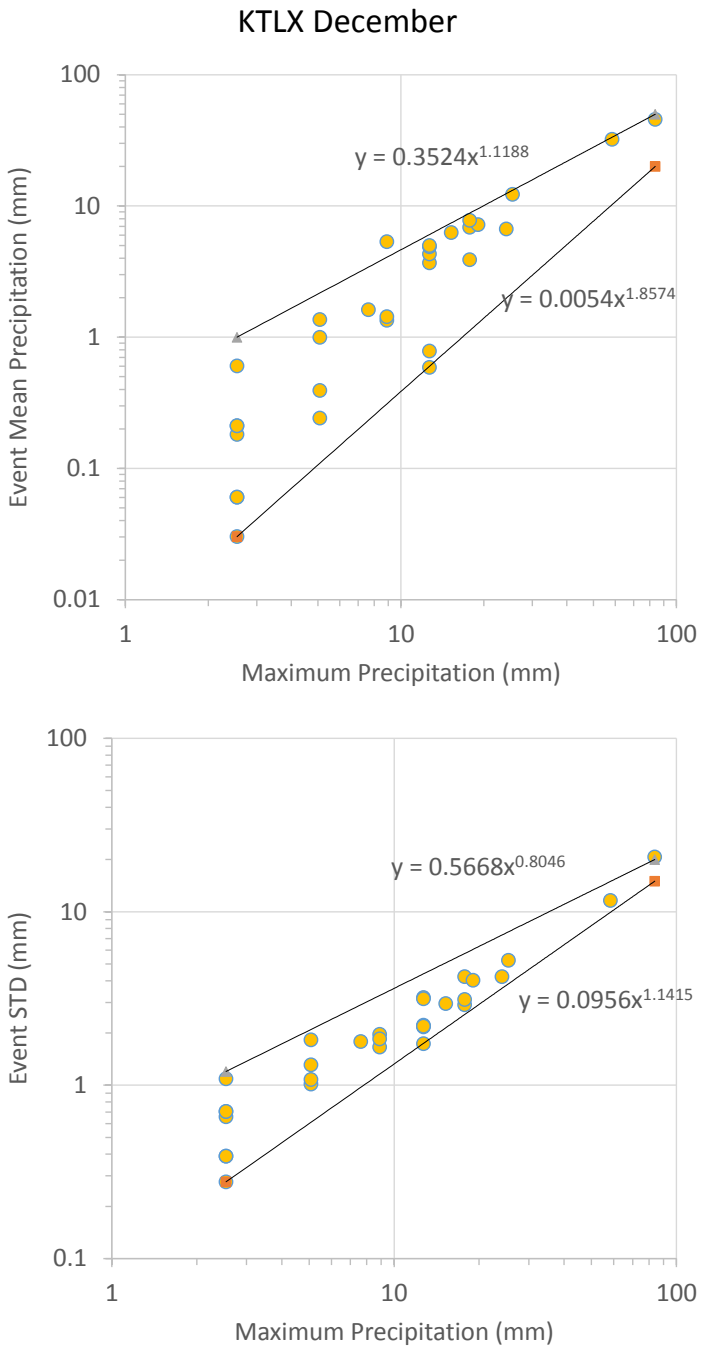


Figure J.12. The log-scale figures used to find the possible boundary relationships: (1) The log-scale figure between the daily maximum precipitation and the daily mean of all the cells in December from 2002 to 2012 from KTLX data set and the upper and lower boundary; (2) The log-scale figure between the daily maximum precipitation and the daily standard deviation of all the cells in December from 2002 to 2012 from the KTLX data set and the upper and lower boundary.