A QUALITY ASSURANCE/QUALITY CONTROL ANALYSIS OF RAINFALL DATA COLLECTED BY VOLUNTEERS IN TUCSON, ARIZONA FOR THE RAINLOG.ORG PROGRAM

by

Candice Lea Rupprecht

A Thesis Submitted to the Faculty of the

DEPARTMENT OF HYDROLOGY AND WATER RESOURCES

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE
WITH A MAJOR IN HYDROLOGY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2009

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Candice Lea Rupprecht

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

________________________________ 11/4/09
Gary C. Woodard
Associate Director of SAHRA

________________________________ 11/4/09
Thomas Maddock, III
Prof. and Head of the Dept. of Hydrology & Water Resources

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF FIGURES

LIST OF FIGURES - CONTINUED

LIST OF TABLES

LIST OF EQUATIONS

ABSTRACT

Scientists now recognize how quickly environmental conditions are changing, yet to monitor and understand these spatially distributed changes more dispersed quantitative and qualitative data are needed than ever before. The need for more comprehensive and robust data has created the burgeoning field of citizen science, which engages volunteers to monitor environmental changes and report this information to scientists. Precipitation monitoring networks like RainLog.org are considered one of the oldest types of citizen science with many networks in existence for over 100 years. RainLog.org is a more modern version of these original networks and was developed in response to a need to better characterize precipitation events and provide stakeholders with more robust precipitation totals and distributions throughout Arizona.

RainLog.org is a statewide precipitation monitoring network that relies on volunteers across Arizona to report daily precipitation into an online reporting system. To ensure that these data are reliable, a quality assurance and quality control analysis (QA/QC) was completed on a subset of gauges in the Tucson area. Results indicate that although there are many errors inherent with any precipitation network, whether volunteer or scientist driven, these errors are for the most part identified using basic interpolation methods. This paper analyzes a range of user reporting and gauge type errors, discusses the significance of each error type and provides recommendations for mitigating reporting errors in any citizen science network.

1. INTRODUCTION

1.1 Background

Volunteer citizen scientist networks are becoming increasingly popular as earth scientists recognize a need for more densely and spatially distributed data collection. Additionally, qualitative observations of changing environmental conditions which require human observers to collect and report data are also increasing. Citizen science networks offer a low-cost alternative for researchers seeking large quantities of data that may be otherwise cost- or time-prohibitive.

The scientists that utilize volunteer data to support research goals directly benefit from volunteer-collected data, but they are not the only beneficiaries. Stakeholders that indirectly benefit from citizen science projects by utilizing this data in decision-making and public outreach include land managers, educators, water managers, health departments, water departments and weather forecasters and reporters. The direct benefit of citizens participating in the scientific process is thought to increase the level of knowledge in participants about earth science systems through experiential education. This collaboration encourages citizens and scientists to communicate at a mutual level of understanding and may ultimately engage citizens in addressing natural resource issues in their local community. In addition to learning scientific techniques and developing a greater appreciation for their local environment, citizen scientists may become advocates for research once they understand how their own data can be used to support large-scale projects; providing citizens with opportunities to engage in research at local and individual levels may engage them in community-wide discussions of the important role of science in our society, from individual daily behaviors up to decision-making that influences policy.

Now, more than ever, our society has the ability to communicate information instantaneously, which is beneficial to the data sharing process of citizen science. Previous volunteer networks relied on observers to complete monthly forms, such as the National Weather Service B-91 forms, and mail them in. Data entry and processing took 45-60 days from the end of the data month (Shein and Owen, 2006). Precipitation data collected this way was useful for creating long-term weather and climate records, but did not provide participants or researchers with any real-time data to aid in short-term planning and forecasting efforts.

Unlike older volunteer monitoring systems that require participants to fill out data cards and send in to a human processor, RainLog.org utilizes new technology that allows data to be made available to the public in less than a minute after it is reported; assuming volunteers report their daily data shortly after they record it at 7:00 a.m., RainLog.org can provide regional rainfall totals only hours after the previous day's storm event. This near-instant feedback keeps participants engaged and allows interested parties to view graphical rainfall reports of localized events that demonstrate spatial variability and support short and long-term trend analysis. These near real-time data are especially important for weather forecasters wanting to track storm systems and show images of rainfall patterns; spatially distributed data displayed on a map allow them to do this through a user-friendly graphic interface. Additionally, flood control managers want to develop the predictive capability of preparing for streamflow and stormwater responses from rainfall events; spatially distributed data help them develop these relationships without depending on their own staff to maintain equipment.

Citizen science initiatives like RainLog.org provide citizens with a hands-on, reoccurring task that engages participants in a collective approach to individual data collection. This collective approach is critical to scientists who need data from densely monitored systems, but do not have the time and/or resources to setup and maintain entire monitoring systems. In addition to precipitation data being collected by citizen scientists, researchers are now asking the general public to engage in collecting qualitative and quantitative data to support drought monitoring and climate change reserach. Through Arizona Drought Watch (http://azdroughtwatch.org/), citizen scientists may report observations like supplemental watering and irrigation, increased erosion due to vegetation loss and livestock health impacts as a result of drought conditions in their local area. Once entered into an online system these reports of cumulative impacts help land and water managers develop a more comprehensive management plan for protecting ecosystems and human needs. While these impacts may be seen over the course of months, some climate-related changes may take years or decades to establish a trend. The National Phenology Network (www.usanpn.org/) has engaged citizens throughout the U.S. to monitor the annual life cycle events of plants to help determine whether the timing of these events is shifting. These data will provide researchers with on-the-ground evidence of where we are seeing the most significant climate change impacts and will help support targeted policies to address solutions. Citizen science is an exciting nexus between science and policy that engages citizens at a grassroots level to improve the understanding of our earth systems and may influence science-based decision-making at local and national levels that balances environmental and anthropogenic needs.

A National Science Foundation (NSF) survey about science and technology issues conducted in 2001 found that 45% of people polled were very interested in new scientific discoveries and the use of new inventions and technologies, 45% were moderately interested and only 10% of people were not interested at all. These findings suggest that people have a relatively high interest in science and technology and that participation and education involving science and technology may result in a better-informed public. However, the level of interest in science and technology declined between 1996 and 2002 and despite a professed interest in science and technology; only 15% of respondents in the same 2001 NSF survey described themselves as well-informed, while 30% considered themselves poorly informed (National Science Board, 2004). Not only do many people consider themselves not well informed about science and technology issues, scientific literacy, defined as knowing basic facts and concepts about science and having an understanding of how science works, is also fairly low in the United States. Two-thirds of Americans do not understand the scientific process and in the 2001 NSF survey of fundamental scientific terms and concepts, the responses to half of the questions were answered correctly less than 60% of the time. A technology-driven society needs citizens to have a basic understanding of scientific principles and the scientific process in order to help individuals better understand news related to science and technology, make fact-based decisions and participate in public discourse that may affect national policy.

To determine the most effective mechanism for educating the public about science concepts and issues, ascertaining information sources that are utilized by the public is important. Not surprisingly, television is still the most popular source of information about science and technology, accounting for 44% of individuals. In this 2001 NSF survey, only 9% of people got their general science and technology information from the Internet. However, people interested in specific scientific issues used the Internet 44% of the time and television 6% of the time, suggesting that a majority of people who want real-time, content-specific knowledge will turn to the Internet to find answers. An example of this trend is evidenced by the fact that people using the Internet to get weather news increased from 47% in 1996 to 70% in 2002 (National Science Board, 2004).

In a 2002 survey of experienced Internet users, 28%, which was the largest group of respondents, said “the ability to get information quickly” was the main reason for initially using the Internet (National Science Board, 2004). The increasing interest and utilization of the Internet as the primary source of issue-specific science information, coupled with television as the consistently most popular source of general science and technology information, presents a need to link real-time science and technology information on the Internet and on television programs. RainLog.org is trying to meet this need by providing publicly available, near real-time data online, while working with a local news station to integrate these data into weather forecasts and reports.

In all citizen science projects, the main goal is to collect data, either qualitative or quantitative, to support a larger scientific research question. A secondary goal of citizen-science projects that is not always as well defined, and one that was not discussed at the inception of the RainLog.org project, is increasing participants' knowledge about related scientific concepts, increasing understanding of the scientific process and improving attitudes toward science. These secondary goals are assumed to be met merely through participation in citizen science projects, but research suggests that understanding of the scientific process and increased knowledge of how data support a bigger scientific question are dependent on how transparent the goals of the research project are to the participants and whether they are given specific training (Brossard et al., 2005).

Results of a study on The Birdhouse Network (TBN), in which participants put up a nesting box and then observe box inhabitants, provide evidence that while most citizen science participants are already highly engaged in scientific issues, participation in this project did not change their attitude toward science and technology in general. Additionally, participants did not increase their understanding of the scientific process, although researchers offer the disclaimer that engaging participants in the scientific process was not the focus of the bird study and the primary motivation for participating in the TBN was an interest in birds, not the scientific process. Evaluation of the TBN survey indicates a statistically significant increase in knowledge of bird biology, supporting the theory that through experiential education, participants increased their knowledge of the subject they were actually engaged in observing. This finding suggests that researchers must make the goals of the project very clear to participants and that focusing efforts on increasing knowledge and understanding of science concepts most relevant to the project will have the greatest chance of success. In the case of

RainLog.org, this study implies that learning objectives should focus on: accurately reading and reporting precipitation measurements; understanding the spatial variability of rainfall; understanding the need for denser networks of rainfall measurements; and increasing awareness of RainLog.org data applications.

In addition to increasing participants' knowledge about scientific processes, citizen science also plays a role in developing human behavior responses to nature and the environment. This paradigm shift of opening science to the public raises the issue of volunteer motivation; the question becomes why individuals choose to participate in citizen science projects and at what level. Little comprehensive research has been done to answer these questions; however a pool of literature addressing different types of participation and strategies for engaging public audiences exists. Research that elicits information about volunteer motivation is usually limited to specific projects that identify motivation primarily as a concern for an individual species (Bradford and Israel, 2004). Lawrence (2009) suggests that motivation is driven not only by the collection of data, but also by the integration or application of this data, indicating a need for projects that collect and use data in meaningful ways. Citizen science, often through anecdotal evidence, also points to the emotional experiences of participation, which affect personal values and feelings that can shape perspectives on policy (Lawrence, 2006; Ellis and Waterton, 2004; Milton, 2002). Therefore, engaging individuals in citizen science projects helps foster better decision-making through increased data availability and influences participants' ideals that are linked to community attitudes and values.

1.2 Other Community Precipitation Networks

Using volunteers to collect precipitation data over a large area and enter it into a central database is not a new concept. Many community-based precipitation networks exist across the U.S., ranging from local irrigation districts supported by farmers mailing in postcards once per month to state and national networks including thousands of volunteer monitors that report daily precipitation online. Programs vary in the type of rain gauges they expect participants to use, as well as the purpose of collecting data and thus how data are reported. However, data collected by a disparate group of people, recently coined as “citizen scientists” is becoming increasingly popular as budgets tighten and the demand for more densely gauged systems increases. A major goal of RainLog.org has been to assimilate these smaller networks by offering our online capabilities. In fact, many smaller networks desire online reporting with real-time or near real-time results and RainLog.org is able to provide this service, which is especially popular with county flood control districts that need precipitation data in a timely manner. Both the Maricopa County and Pima County Flood Control Districts are working with RainLog.org to share data and increase their volunteer observer base, which Maricopa County says will “continue to help the District improve the safety and welfare of County residents through sophisticated flood control and floodplain management programs” (2007). Two of the largest volunteer precipitation networks are discussed below as a comparison to RainLog.org; these conglomerate networks have taken a similar approach by recruiting participants from existing volunteer precipitation networks.

1.2.1 National Weather Service Cooperative Observer Program (COOP)

The NWS COOP is the oldest existing citizen science program in the U.S and it has relied on volunteer observations of daily precipitation since 1890 when it was enacted under the Congressional Organic Act as part of the newly established Weather Bureau (U.S. NWS, 1989). Since its inception, the COOP has grown to over 12,000 volunteers that record daily precipitation and air temperature and some station observers that also record soil temperature and evaporation (NWS, 1989). In addition, COOP stations may be established in conjunction with hydrometerological stations that report river stage, wind movement, agricultural data, atmospheric data such as hail and thunderstorms, flash flooding and road hazards (NWS, 2007). Gauges used for recording precipitation include 4-inch and 8-inch nonrecording gages, both of which have a funnel catch and overflow container, and Belfort (Fischer & Porter) and Universal recording (weighing type) gages, which the NWS recommends checking weekly to ensure proper function.

Because the NWS depends on the COOP network to observe and describe the climate of the U.S., NWS personnel work with the volunteer observers to setup, train, calibrate, inspect and service gauges for reporting (NWS, 2008). The NWS takes its role as a community partner very seriously and acknowledges its role to support these on-the-ground observers, so that accurate data are collected and reported in a timely matter. Although COOP volunteers are not compensated, they are expected to collect and report data daily and it is through the established relationships between COOP observers and the NWS that COOP stations can be maintained for long-term periods. A spatial analysis of the COOP data to produce interpolation maps of the conterminous United States found that the patterns were highly discontinuous; these patterns were a result of individual observer bias caused by underreporting and a tendency to report amounts that were divisible by 5 and/or 10 (Daly et al., 2007). These rounding errors raise questions about whether RainLog.org data is subject to these same observer biases and whether the biases are related to the type of gauge used by the COOP, or whether these are errors common to all volunteer precipitation networks.

1.2.2 Community (originally Colorado) Collaborative Rain, Hail and Snow (CoCoRAHS) Network

CoCoRAHS was started in 1998 by scientists at Colorado State University who had identified the extreme spatial variability of precipitation in Northern Colorado and decided that volunteers can play an important role in collecting these data and increasing awareness of local rainfall patterns and storm events. Volunteers wanting to participate in CoCoRAHS are asked to purchase a 4-inch diameter, high-capacity plastic rain gauge, which has a funnel catch with overflow design (similar to rain gauge #4 shown in Figure 2.1). Participants are provided with basic instructions on how to collect data, either in a group training session or online training and after training, participants are expected to report precipitation daily via phone or the online CoCoRAHS website: www.cocorahs.org (Reges et al., 2006). Currently the CoCoRAHS network has over 9,000 reporters in 32 states with a goal of 20,000 participants by 2010 (CoCoRAHS, 2008).

The main difference between RainLog.org and the CoCoRAHS network is that CoCoRAHS participants are asked to purchase a specific rain gauge, known to be accurate based on previous comparative studies between the 4-inch plastic gauge and the 8-inch metal standard gauge used by the National Weather Service (Doesken, 2005). Specifying a certain gauge to be used system-wide decreases the amount of data variability within the system due to the variation in precision of different types and brands of gauges. The CoCoRAHS website has a variety of searchable reports and it allows one to create several different types of maps that may be useful for displaying particular rainfall, snow or hail storm events (Figure 1.1). However, the map interface has no land surface reference points and is searchable only by county. Users cannot zoom or define the map area, nor can they look at an individual station's historical data. Also problematic for real-time reporting is the lengthy delay between when users enter data and when data is displayed on the map. RainLog.org, created after CoCoRAHS, improved on many of CoCoRAHS's cumbersome functions to provide a smoother interface.

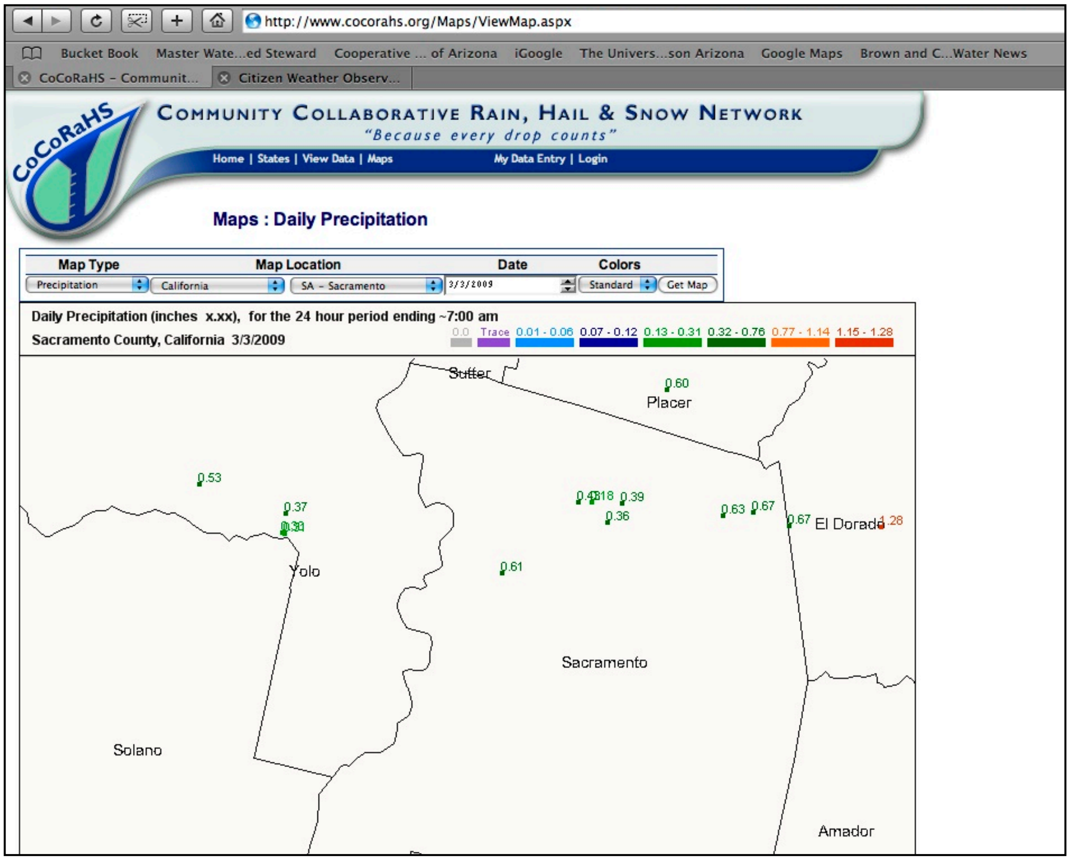

Figure 1.1 - Screencapture of CoCoRAHS map. Notice that rainfall values are color-coded by value range, however unlike the Google Maps™ interface, there are no geographic markers and the only political markers are county boundaries, which are not identifiable on the landscape.

1.3 RainLog.org Overview

RainLog.org is a statewide precipitation monitoring network developed for Arizona and it aims to provide a more comprehensive picture of the state's spatial and temporal rainfall variability; this includes better characterizing summer rainfall patterns. With the increasing need to observe large-scale systems and resolve uncertainties in our physical earth systems, the benefit of recruiting volunteers to monitor daily precipitation quantities is two-fold. By relying on volunteers to report data, the significant time and expense invested in setting up a precipitation network is avoided. Manual rain gauges must be read and reported on a daily basis, which is not an efficient use of time for scientists and agencies. Alternatively, tipping bucket gauges can be set up to transmit data electronically, but a single electronic gauge costs several hundred to thousands of dollars, which would be cost prohibitive for a larger network.

Additionally, rainfall reporting provides a non-technical, yet scientific opportunity for citizens to be involved in current research. While monitoring precipitation requires a high level of accuracy and a daily time commitment to read gauges, it does not require extensive hands-on training that is often a hurdle in encouraging citizen participation in scientific research. As a citizen science initiative, the goals of RainLog.org are: to provide localized rainfall data to residents to help improve irrigation schedules; increase participants' knowledge about meteorology, climatology and the scientific process; and increase participants' understanding of and interest in local water resource issues.

1.4 RainLog.org Development

In 2005, RainLog.org was developed as a cooperative rainfall network by researchers at the University of Arizona to address local community needs to better gauge rainfall events. Like many arid and semi-arid regions, Arizona is subject to significant climatic extremes, expressed as spatial and temporal variability (Adams and Comrie, 1997). This variability, for interested citizens and scientists alike, caused regional weather reports to be inadequate. To better understand climate patterns in Arizona and to address this lack of local data, the NSF SAHRA Center (Sustainability of semi-Arid Hydrology and Riparian Areas) and Arizona Cooperative Extension partnered staff and resources to develop the RainLog.org framework. The goal of this project was originally to provide researchers, drought monitors, irrigation schedulers, weather reporters, educators and interested citizens with more accurate precipitation data gathered from a large number of semi-clustered gauges, reported by volunteers who agreed to monitor daily rainfall accruals in a precipitation gauge located at their home or work.

Significant time and software development capacity were invested in RainLog.org to create a functional, online website and reporting system. This system requires each volunteer reporting rainfall to register and create a unique user account that is password-restricted. Once logged-in to the system, users are provided with a daily reporting form with unique cells for four rainfall parameters including: quality of reading, rainfall amount, reading type and reading date; the form also has space for additional comments. Only about three percent of all data entries include comments, but these comments can be helpful in understanding storm characteristics and potential reporting errors. Many comments included information about the type of precipitation, such as snow and hail; other weather conditions affecting gauge performance such as windspeed, dew, frost or frozen gauges; storm timing and intensity; and periods of absence, which results in no data or data totaled for multiple days. If data recorded in the rainfall amount category is questionable, the comments may provide insight as to whether data accurately reflect a single day precipitation value. Once these daily entries are submitted into the database, they are then stored in two locations. Currently all data are stored within a large database accessible only by database mangers and upon request, but individual reports can also be viewed in table and bar graph formats linked to the online map on the website.

1.4.1 Online Data Access

Precipitation data collected through the RainLog.org network is openly available online using a Google Maps™ satellite map interface. This feature is user-friendly due to the colored text boxes that are a visual indicator of approximate rainfall amount (Figure 1.2); the Google Maps™ satellite imagery provides a visual reference, yet the exact location of reported rainfall values cannot be ascertained because this map interface does not have georeferenced locations available. While exact locations based on latitude and longitude are stored in the RainLog.org database, they do not appear on the Google Maps™ interface.

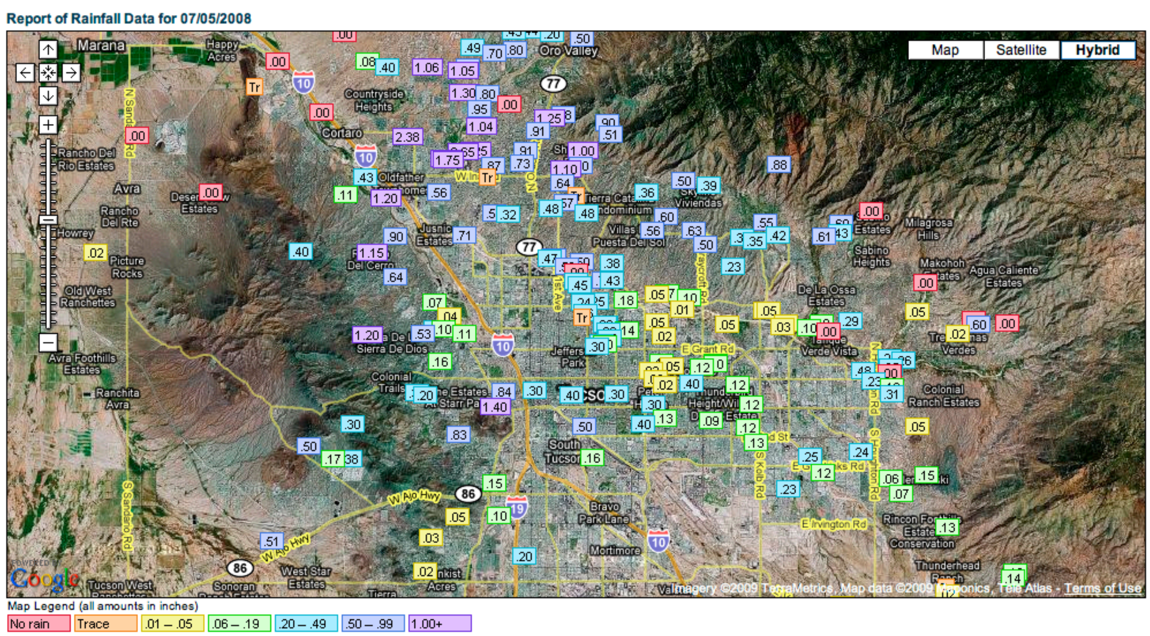

Figure 1.2 - Precipitation values are displayed in colored text boxes on the map based on precipitation value ranges, that allows users to discern rainfall patterns and anomalies.

Users that need georeferenced points for the rainfall data have the ability to download any precipitation data on record. Any individual's reported rainfall values can be downloaded and imported into Excel; this feature allows the user to generate his/her own graphs, correlate this precipitation data with other data sets and perform statistical analyses on this data. To georeference reported rainfall locations users can opt to import a data file into Google Earth™. Users can then match up certain data sets with their UTM or Latitude/Longitude location.

1.4.2 Rain Gauge Selection

The selection of rain gauges for the RainLog.org network was a three-fold decision, based on cost, accuracy and precision, and ease of use. To recruit a maximum number of participants with minimal time and money invested, recommending an affordable gauge that was easy to use without providing extensive training was important, as was allowing other gauge types to be used if users already owned one (Table 1.3). In addition, ensuring that the selected gauge would collect accurate data for small quantity and intense precipitation events, both characteristics of storm events in Arizona was also important.

Gauge Brand	**Gauge Type**
Acu-Rite	Simple Cylinder or Rectangular Catch
All Weather	Funnel Catch with overflow
Davis	Tipping Bucket/Electronic Gauge
Honeywell	Tipping Bucket/Electronic Gauge
La Crosse	Tipping Bucket/Electronic Gauge
NWS	Funnel Catch with overflow
Oregon Scientific	Tipping Bucket/Electronic Gauge
Productive Alternatives	Funnel Catch with overflow
Springfield	Simple Cylinder or Rectangular Catch
Stratus	Funnel Catch with overflow
Taylor	Wedge
Tenite	Funnel Catch with overflow
Thermor	Tipping Bucket/Electronic Gauge
Torrent	Tipping Bucket/Electronic Gauge
Tru-Chek	Wedge

Table 1.3 - Brands and associated gauge types of all known RainLog.org rain gauges.

A study comparing the National Weather Service (NWS) Standard Rain Gauge (SRG) to a 4” diameter plastic gauge found that over a 10-year period, the monthly averages for the 4” gauge equaled or exceeded the NWS SRG for all months, resulting in a 3% difference per year (Doesken, 2005). The study attributes this difference to the type of material and design of the different gauges. The SRG has a metallic surface, which is predicted to evaporate water more quickly than a plastic surface. This finding suggests that in a semi-arid region like Arizona, plastic gauges are more accurate since reference evapotranspiration exceeds precipitation at weather stations across Arizona. However, plastic gauges will need to be replaced more often due to ultraviolet exposure (AZMET, 2008). In addition, the study found that most of the gauge differences of 0.05” or greater occurred during snow, heavy rain, hail and high wind storm events, resulting in a slightly higher annual average for the 4” gauges than the SRG 8” gauges. Since heavy rain, high wind and sometimes hail, characterize Arizona’s summer precipitation events, and summer precipitation accounts for half of Arizona’s total precipitation (Sheppard et al., 2002), selecting a gauge that captured rainfall events as accurately as possible was important.

The findings of the Doesken study support the decision to recommend a simple, plastic rain gauge, although preference is given to the wedge-shaped gauge because smaller amounts are easier to read on this type of gauge. The Tru-Chek gauge is now the standard for RainLog.org because of its reasonable cost ($12, as of July 2009) relative to other gauge options on the market; to date 60% of RainLog.org participants have registered a Tru-Chek gauge. To help guide participants in selecting and reading their gauge, an online handbook, including graphic and text-based information, was developed to explain: the differences among gauge types to aid gauge selection; how to read gauges with rainfall and snow data; how to report data online; and answers to frequently asked questions. Although reading through the online learning modules is encouraged, there is no requirement to read this training information before registering on RainLog.org and entering data.

1.4.3 Rain Gauge Siting

Rain gauges are widely used to estimate how much precipitation falls on the land surface in a certain location. Due to the earth's non-homogenous land surface, environmental interferences can affect gauge accuracy. The World Meterological Organization (WMO) recommends that rain gauges be buried even with the land surface to minimize environmental interferences (e.g., air turbulence). However, for ease of training, reading and maintenance, gauges are often mounted above the land surface to reduce debris collection in the gauge shaft and to prevent flooding of the gauge. RainLog.org recommends that gauges be sited at the top of a fence post in an open area, free from building and vegetation interference. Based on WMO recommendations, gauges should be twice as far from interfering objects as the height of the gauge (1996).

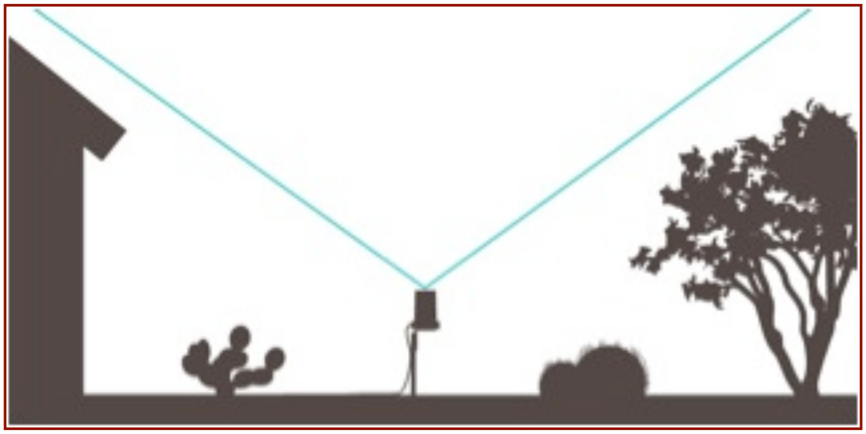

Figure 1.4 - Image created for the RainLog.org website illustrating recommendation for placing rain gauges in an open, unobstructed area, at least twice the horizontal distance from overhead objects (e.g. trees, roofs) the vertical distance between the object and the gauge. An acceptable alternative is to place the gauge on top of a fence in an open area.

1.4.4 Rain Gauge Reporting

RainLog.org asks people to report the previous day's precipitation accumulation at 7:00 a.m. each morning. This 7:00 a.m.-7:00 a.m. period is based on NWS COOP Observer reporting practices and is the standard for daily precipitation reporting (NWS, 1989). The assumption behind this standard is that since most volunteers within the network do not have electronic gauges, people are more willing and able to take a daily recording at 7:00 a.m. than at midnight. Research on human's natural sleep cycles and residential water use patterns indicate that 7:00 a.m. is a time when most people are awake, but still at home, making it an ideal time to report rainfall on a daily basis (Figure 1.5) (Vitaterna et. al, 2009; Mayer and DeOreo, 1999). Most individuals are usually up at 7:00 a.m. and can take measurements before they leave for work, unlike at midnight when most individuals are asleep and darkness would complicate taking a reading. Also, avoiding taking a measurement during a storm event when precipitation could be lost ensures a more accurate reading; since convective storm events do not usually develop until the afternoon when daily air temperature is highest, it is best to take measurements in the morning when it is least likely to rain.

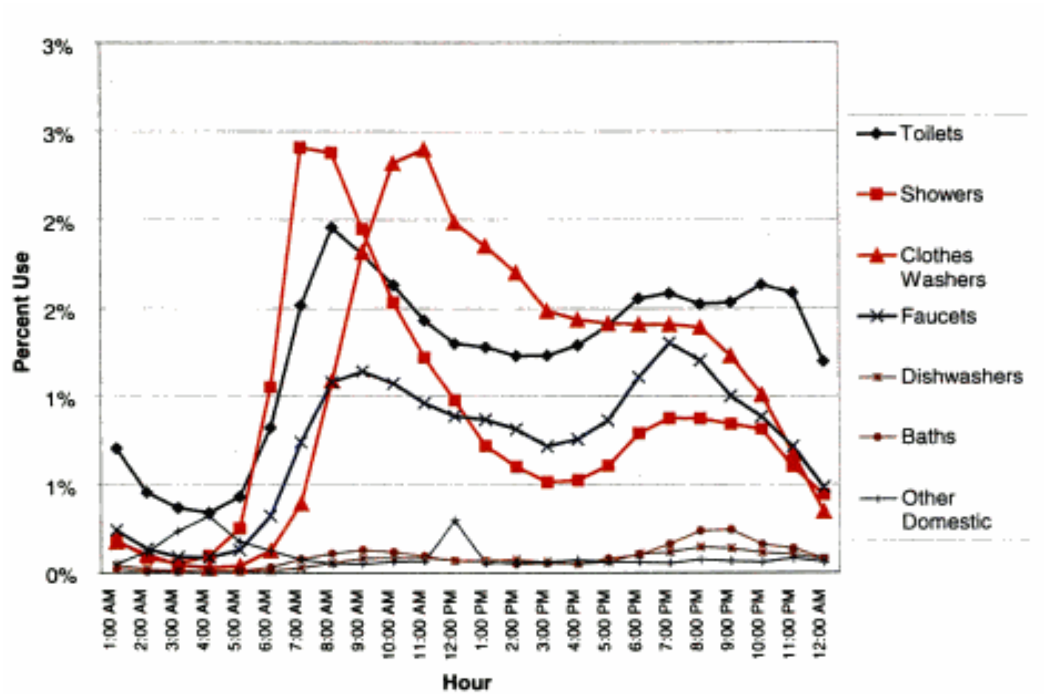

Figure 1.5 - Hourly residential water use patterns of different indoor water fixtures. People are at home when water uses are highest, supporting the 7:00 a.m. reporting period for Rainlog.org and other volunteer networks (Mayer and DeOreo, 1999).

Because RainLog.org is a volunteer network, it was important to develop procedures that were convenient and minimized people's inability to consistently measure and report data. While this 24-hour reporting period is not completely representative of a single date's precipitation, it minimizes the potential error incurred when people report too early, too late or create their own 24-hour reporting window without informing the system administrator. Essentially, this 7:00 - 7:00 window captures 17 hours of the previous day's rainfall and 7 hours of the current day's rainfall. Volunteers may be confused by this 24-hr reporting window and instead of reporting for the previous day, will report precipitation for the same day. This error can happen just on occasion or daily if the volunteer does not have a clear understanding of our 7 a.m. reporting time. The problem may be exacerbated by individuals who do not report data on a daily basis, but instead enter all values into the database at the end of each month. Depending on each individual's method for recording data before it is entered in the database, he/she may forget that there is a one-day lag time in entering data.

1.5 RainLog.org Volunteers

Due to the nature of this program, RainLog.org managers have relied on word-of-mouth communication, county-based presentations by University of Arizona faculty and press coverage to attract volunteer participants. RainLog.org coordinators have given many presentations about the volunteer network throughout Arizona to agency and volunteer groups alike, which include local flood control districts, irrigation districts, Master Gardener and Master Watershed Steward groups. Often, this leads to an increase in word-of-mouth registrations in addition to the targeted participants that are recruited by partner agencies. In some cases, local, state and federal agencies have sponsored gauges to be purchased for a specific city or neighborhood to increase community awareness and to provide more localized precipitation data. As of July 2009, 1,794 users had created accounts and reported rainfall measurements in the RainLog.org system. These participants are located throughout Arizona, as well as in 26 other states across the United States.

RainLog.org invites anyone interested in climatological and hydrological data and willing to collect and report data on a daily basis to participate in the RainLog.org network. To meet the RainLog.org goal of creating as dense of volunteer network as possible, the requirements to join and participate (discussed in Section 1.4) are minimal. The expectations of all RainLog.org volunteers are basic, but critical to the success and validity of the network. Ranked from highest to lowest importance volunteers should:

1. Report their data (in a timely manner if possible).
2. Report their data on the correct day.
3. Report their data accurately.
4. Report their data with precision.

Each of these expectations includes avoiding unique ways in which errors can enter the data. While virtually all errors are unintentional, discerning the types of errors that are associated with each of our expectations is important, so that we can target solutions to alter behaviors or correct the system data if necessary. Errors associated with each expectation are discussed in detail in Chapter 2 and an analysis of how significant these errors are will be presented in Chapters 3 and 4.

1.6 Thesis Project Objectives

Our project aims to evaluate data collected and input into the RainLog.org system, from the perspective that it is a citizen science volunteer monitoring network. We need to better understand the constraints of these data and find solutions for ensuring that these data meet the needs of scientists and water managers who will use these data. The focus of this project is to design effective strategies for dealing with Quality Assurance and Quality Control (QA/QC) issues in volunteer data and developing methodologies for ongoing assessment. The research questions to be answered in the scope of this project are:

1. What are the most likely causes of error?
2. How can we identify systematic and random errors?
3. How can the identified errors be reduced or eliminated?
4. How "good" are the RainLog.org data?
5. To what degree can we trust these data?

To answer these questions we researched and analyzed the variables associated with the RainLog.org network including gauge type, gauge location and potential data reporting errors based on data input by users in the Tucson area over the period from June 15, 2006 to September 30, 2008. This time period is characterized by three monsoon seasons (June 15 – September 30) and two frontal seasons (October 1 – June 14). During this period, a significant effort occurred on the part of the developers to expand the network; as of June 15, 2006 there were 108 participants in the entire RainLog.org network and 66 in the Tucson area. At the conclusion of the study period, there were over 1,500 participants in the entire RainLog.org network and 425 participants in the Tucson area.

2. QA/QC OF VOLUNTEER DATA

The nature of a volunteer data collection network inherently allows for various participant and equipment errors; however, we believe that quantifying and correcting these errors yields a more robust data set that can be used for multiple purposes. The goal of deploying a spatially distributed network of volunteers to collect daily precipitation data is to develop a database of point-based precipitation that is representative of a larger area. With all point-based measurements intended to represent an area greater than the point, the choice of site, as well as the systematic measurement errors of the network instruments, is important in understanding the accuracy of data collected (WMO, 1983; WMO, 1982; Allerup and Madsen, 1979). To better understand the potential errors that exist throughout the RainLog.org system, we conducted a thorough analysis of the different data issues that exist within the system and divided them into four basic types of potential errors:

1. Gauge precision and accuracy – different gauges have different levels of precision and accuracy; what is the system-wide affect of these different gauge types?
2. Gauge location – gauges are set up in a variety of locations; do these different locations have interferences that can affect reporting?
3. Human reporting errors – data can be entered incorrectly; can we automate the system to identify when this happens? What information mechanisms (e.g., email, online learning module) mitigate this problem?
4. Network variability – gauges are not always evenly distributed; does this cause system biases?

This chapter explores the four different types of errors that can potentially affect the quality of the data in the RainLog.org system and a discussion of possible solutions is offered for each error type that may improve the data by reducing bias in the system. Chapter 4 offers an analysis of implementing some of the solutions discussed for each of the errors in this Chapter.

2.1 Gauge Precision and Accuracy Issues

Although gauge accuracy and precision are similar, they reflect important differences in how gauges function and how users report data. Accuracy reflects the average difference between the gauge reading and the actual value, while gauge precision reflects the repeatability of measurements reported by the gauge. Except with electronic gauges, humans are ultimately in charge of insuring precision, but this is helped when a gauge is easy to read. Gauge accuracy is a measure of exactness and is mostly influenced by design, because it is up to the manufacturer to make sure that the 0.5 inch line on the gauge is exactly 0.5 inches. Ultimately, accuracy is more important than precision Because we are not measuring the exact same quantity during each storm event, we have to rely on the equipment and the user to ensure that they are both as accurate and precise as possible. This also ensures that variability within the network is attributed to the spatial variability of rainfall and not of different gauges or humans.

RainLog.org was designed as an open-access online volunteer network that accepts data from anyone with a rain gauge. Unlike other cooperative rainfall monitoring networks that require the purchase and/or use of a specific rain gauge, RainLog.org was designed to be a user-friendly system that anyone could join with minimal investment.

While RainLog.org was able to recruit over 1,300 participants across Arizona in less than two years, potential data errors exist due to the varying precision and accuracy of various gauge types. Before data corrections are applied to data entered using certain gauge types, we must first know what the exact levels of accuracy and precision are for the major gauge types and what type of gauge each participant uses to report rainfall measurements.

Based on World Meteorological Organization studies (1983; 2006), the six potential sources of error that are systematic and can affect individual gauges, depending on gauge type, are listed below in order of largest potential errors:

a) Error due to systematic wind-field deformation above the gauge orifice: typically 2 to 10 percent for rain and 10 to 50 percent for snow;
b) Error due to the wetting loss on the internal walls of the collector;
c) Error due to the wetting loss in the container when it is emptied: typically 2 to 15 percent in summer and 1 to 8 percent in winter, for (b) and (c) together;
d) Error due to evaporation from the container (most important in hot climates): 0 to 4 percent;
e) Error due to blowing and drifting snow;
f) Error due to the in- and out-splashing of water: 1 to 2 percent.

Not all gauges experience all of these errors, however the significance of these potential errors in the RainLog.org network is that we do not have any data collected from which we can infer how the various errors may affect certain gauges because we did not do any independent research. The first four errors (a-d) are dependent on which type of gauge the participant chose to purchase and where he/she chose to site the gauge, but will almost always result in underreporting errors. The net error due to blowing and drifting snow and to in- and out-splashing of water can be either negative or positive, although these are difficult to quantify and generally of less concern for the RainLog.org system because our climate is such that most of Arizona does not receive large quantities of snow and we do not recommend that participants install gauges in a pit, which can result in the in- and out-splashing of water from the ground surface. Arizona's arid climate causes evaporation rates to be on the high end of average errors, which increases underreporting. Hail is another potential underreporting error not listed by the WMO that results from hail bouncing out of the gauge. Precipitation often falls as hail during monsoon storm events, so underreporting due to hail should be added to the WMO list of potential errors.

Classifying all gauges within the system provides a clear picture of the total system errors that may exist due to various gauge types. There are many different types and brands of gauges, but an analysis of the database revealed that there are five basic gauge types used by RainLog.org participants, so we categorized all existing gauges within the system into one of five categories; if we could not discern which category a gauge belonged to, it was classified as "unknown". The online handbook was updated to provide an illustration and explanation of each of these gauge types to help users classify their gauges; the five different gauge type categories are:

1. Tru-Check/Wedge Type - rectangular opening and tapered base; easier to read for smaller amounts of precipitation, which is useful in Arizona. *(RainLog.org recommended gauge)*
2. Simple Cylinder or Rectangular Catch – straight sides; these have basic construction and are inexpensive.
3. Funnel Catch - similar to simple cylinder, but top opening has funnel catch and sides are scaled.
4. Funnel Catch with overflow – small tube with funnel catch that is housed in a larger overflow cylinder; very useful for monitoring individual precipitation events under 1", but the overflow may evaporate over time.
5. Tipping Bucket/Electronic Gauge - electronically record rainfall based on pre-determined weight or quantity intervals that report data to a base station.

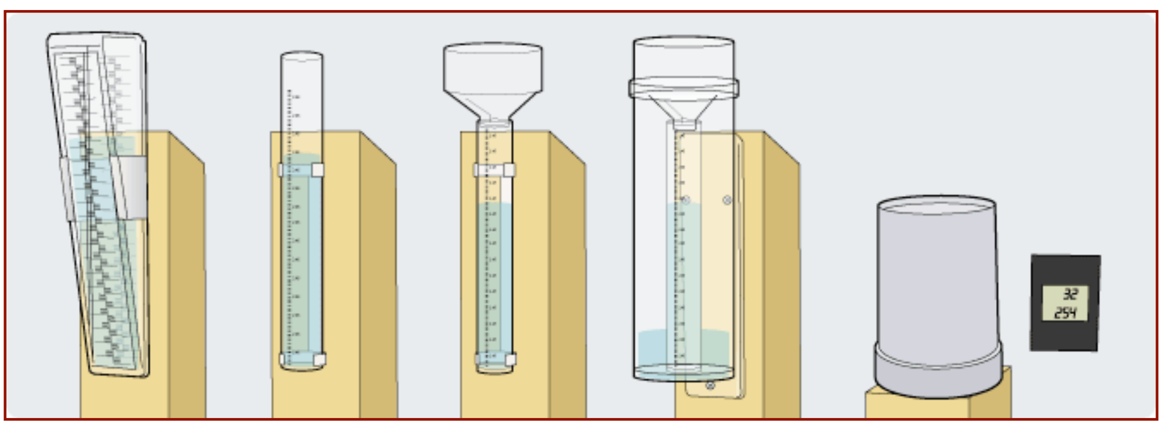

Figure 2.1 - Illustration of the five gauge type categories used in the RainLog.org network. From left to right, 1. Tru-Check/Wedge, 2. Simple Cylinder, 3. Funnel Catch, 4. Funnel Catch with overflow and 5. Tipping Bucket/Electronic Gauge.

To insure system-wide data accuracy by correlating data bias to gauge type, a review of existing research on gauge accuracy and precision was conducted. Based on available literature, the following reporting errors are associated with each type of gauge:

1. Tru-Check/Wedge Type:
    a. Wedge-shaped gauges have a tendency to get clogged with debris in the bottom point that is difficult to remove without scrubbing or high water pressure.
    b. Subject to evaporation from exposure to atmosphere.
2. Simple Cylinder or Rectangular Catch:
    a. Subject to evaporation from exposure to atmosphere.
    b. Less precise for smaller events due to design.
3. Funnel Catch:
    a. The funnel surface area at the top of the gauge is usually larger than on wedge or simple cylinder gauges and may result in more evaporation due to adhesion of water droplets to gauge funnel.
4. Funnel Catch with overflow:
    a. The funnel surface area at the top of the gauge is usually larger than on wedge or simple cylinder gauges and may result in more evaporation due to adhesion water droplets to gauge funnel.
    b. In larger rainfall events where water spills into overflow, some water may be lost due to evaporation or spillage.

c. Less precise for larger events when water spills into overflow.

5. Tipping Bucket/Electronic Gauge:

a. Frequency of tips causes precision errors, especially in low or high-volume and long-duration events.

i. In low-volume events, especially in semi-arid climates like Arizona, some precipitation may never be reported because it will evaporate between tips.

ii. In extremely high rainfall events, the gauge may underestimate due to water loss while the gauge is tipping and if a small bucket is overfilled. Underestimation of extremely high rainfall events has been documented for several gauge brands and has been estimated at 15% underestimation for Belfort gauges (Nystuen et al., 1996).

iii. Long-duration events may result in overestimation because a small amount of water will stay in the buckets after each tip, causing the bucket to tip before it has accumulated a full bucket of precipitation (Chvila et al., 2005).

iv. Larger tipping buckets may underreport if water is left in the bucket after rainfall stops and the bucket does not tip.

v. Small tipping buckets may underreport due to wind blowing water out of the tipping bucket before it tips.

b. Various gauge brands have different tipping rates, which makes comparing the precision of tipping bucket gauges difficult.

c. Failure to regularly calibrate the gauge may result in reporting errors.

d. Wind-induced errors resulting in underreporting may be most significant with electronic gauges due to thicker orifice rims that increase wind speed above the gauge. Gauge losses due to wind interference may amount to 2-15% (Sevruk, 1996).

e. A failure to regularly visually inspect the gauge may result in gauge failure due to animals (e.g., frogs, insects) in the gauge (Nystuen et al., 1996).

In a study conducted in the Goodwin Creek watershed in northern Mississippi to obtain accurate point rainfall measurements, none of the Agricultural Research Service or Texas Electronics Inc. tipping bucket gauges used met the manufacturer's specifications for tipping rate (Sieck et al., 2007). All gauges were tested by running a known amount of water through the instruments for various rainfall rates. Both gauge brands recorded fewer tips that expected and more importantly, the tip rate (number of tips per mm) decreased as the rainfall rate increased. This indicates that tipping bucket gauges report the highest errors in the largest storms, which are the predominant type of weather Arizona experiences during its summer rainy season. Without proper and routine calibration (e.g., monthly or quarterly) these gauges reported rainfall observations 7% to 15% less than actual rainfall accumulation.

2.2 Gauge Location Issues

Because the earth's surface is heterogenous, many research studies have been conducted to identify potential gauge siting errors associated with landscape discontinuity. To ensure that the gauges accurately report rainfall and therefore system data can be trusted, potential siting problems must be identified and addressed to avoid observation errors due to environmental interferences. As a reference standard for rain gauge measurements, the WMO recommends that gauges be installed in pits deep enough so the gauge rim is level with the ground surface and have a shield to protect the gauge from splash off the ground (Sevruk and Nespor, 1998). In the Goodwin Creek study, the incidence of 'Trace' rainfall reports resulted in less total precipitation accumulation for above-ground, unshielded gauges than for buried gauges (Sieck et al., 2007). This result supports the WMO recommendation that to accurately capture rainfall, gauges should be placed below ground, with the orifice even with the ground surface.

However, gauges are typically mounted above the ground surface to minimize debris and animals entering the gauge and to avoid flooding of the pit, both of which can cause inaccurate readings. Above-ground gauges are common throughout the U.S. and are recommended by RainLog.org, but are subject to their own types of errors.

The placement of gauges in urban and residential settings, which is where most RainLog.org participants report from, is complicated by structures and vegetation that can interfere with precipitation accumulating in the gauge. In fact, these structures may cause even more of an interference when gauges are flush with the ground surface instead of elevated, as RainLog.org recommends. RainLog.org provides directions for ideal gauge placement, although environmental conditions are not always optimal due to permanent structures, and the size of an individual's property. Various problems associated with the placement of rain gauges are discussed in the following section.

2.2.1 Placement under buildings and/or trees

Overhead structures including buildings and trees can shield gauges, inhibiting precipitation accumulation and resulting in reported data that are lower than actual accumulated rainfall. Gauges also may be subject to underreporting if there is wind interference from nearby structures that prevents rainfall from entering the gauge. Conversely, overhead structures can also cause excess runoff due to power lines, roofs and trees dripping into gauges, which produces artificially high total precipitation amounts. Trees can also cause over-reporting by clogging gauges if debris falls into the gauge and is not removed.

2.2.2 Placement in open areas

Gauges may also experience underreporting if they are placed above the ground in exposed, open spaces or on elevated sites, such as open fields and rooftops. A self-standing gauge is subject to localized turbulence as air moves around the gauge and this disturbance may cause some of the smaller raindrops to miss the gauge orifice (Robinson and Rodda, 1969; Constantinescu et al., 2007). This happens because elevated gauges distort the wind field above the gauge orifice, which forces the wind speed to increase over the gauge orifice, moving smaller raindrops up and over the gauge, known as the blocking effect (Sevruk et al., 1989; Nespor and Sevruk, 1999).

2.2.3 Off-level placement

Gauges that are not installed on a level surface may be subject to over or underreporting depending on which direction the rain usually falls and whether the gauge wall is tipped toward or away from the rainfall. If the gauge is tipped toward the rainfall the gauge may accumulate more rainfall that the surrounding area and if the gauge is tipped away from the rainfall it may deflect raindrops and collect less rain than the surrounding area.

2.2.4 Wind interference

Wind interference of rain accumulation in gauges happens at different scales and can result in two to 10 percent reporting errors throughout the network (WMO, 2006). Generally these errors tend to yield less rain in gauges than actually fell; however, determining the precise amount of under or overreporting for each gauge is complicated by several factors including whether the gauge is shielded or unshielded, whether the gauge is in an open or vegetated location, the intensity and normal patterns of the wind and what other objects are nearby. As investigated by Sieck et al. (2007), all of these factors can cause turbulence problems that prevent rain from entering the gauge. However, unlike research gauges, most volunteer gauges do not have shields that may cause interference from localized air turbulence. To mitigate this issue, shields are not recommended for use by RainLog.org volunteers.

Studies conducted to evaluate the accuracy of gauge collection at different wind speeds suggest that manual gauges, such as the NWS gauge, report less precipitation with increasing wind speed, for all types of precipitation (Yang et al., 1998). Using the NWS gauge to compare catch from different precipitation types, the same study also found that for the same wind speed, snow undercatch was always greater than for rain or mixed storm events. Unlike many climates where annual precipitation accumulation is dominated by winter precipitation patterns in the form of snow, a smaller percentage of the total annual precipitation throughout Arizona is attributed to snow. To determine the effect of snowfall accumulation on the accuracy of annual reported precipitation in Arizona, data from the National Oceanic and Atmospheric Administration (NOAA) NWS Cooperative Observer Network gauges obtained from the Western Regional Climate Center was analyzed. For the Tucson area, three gauges with the longest records and most complete data sets were selected. Based on data weighted and averaged from the three gauge records in Tucson, the average annual precipitation total is 11.43 inches (slightly less than the reported annual average of 12.17 inches by the NWS during the same time period) and average annual snowfall is 0.69 inches. Based on a mean national snow water equivalent (SWE) ratio of 15.6:1, meaning that 1.56 cm of fresh snow is equivalent to 1.0 mm of water, the average annual precipitation from snowfall in Tucson is 0.044 inches or 0.37% of total precipitation (Roebber et al., 2003). In the history of recorded precipitation data collected through the NWS Cooperative Observer Network, snowfall in the Tucson region effectively contributes a negligible amount of water to the annual precipitation total.

An analysis of annual snowfall accumulation in Flagstaff based on data from two NWS Cooperative Observers indicates that while this area of the state receives more precipitation than the Tucson area, snowfall still contributes only a small percentage of water to the total average annual precipitation of 21.1 inches (WRCC, 2008). The weighted average snowfall of the two NWS Cooperative Observer Network gauges is 87.3 inches, which based on the 15.6:1 SWE ratio equates to 5.60 inches of water.

Therefore, in the Flagstaff region, snow accumulation yields 26.5% of the yearly average precipitation, which compares well with a study done by Serreze et al. (1999), estimating that on average, 9.49 inches, or 39% of annual precipitation falls as snowfall in Arizona and New Mexico. These low annual percentages are a contrast from much of the western United States where snowfall accounts for 60-70% of annual precipitation (Serreze et al., 1999). Because wind-induced losses can be ten times greater for snowfall than rain, the potential errors associated with wind-induced losses during snowfall are minimal in Arizona due to the low snow depths and low percentages of snowfall that contribute to total precipitation (Sevruk, 2008).

2.3 Data Reading and Entry Issues

All volunteer data collection networks face the challenge of their participants having various levels of experience and skill in recording data. To minimize problems that may arise from individuals with varying abilities reporting the data differently, RainLog.org has created an online handbook to help participants effectively record and report daily precipitation values. While RainLog.org recommends that all new participants read through the online information, RainLog.org administrators do not require that volunteers read through this material. This model creates a flexible, accommodating system, although reporting errors may still be present in the system due to minimal training, varying degrees of initial knowledge and skill, and RainLog.org being a volunteer commitment, not a job requirement. The most common data reporting errors are discussed in the following sub-sections.

2.3.1 Missing Data

As was discussed in Section 1.4.4, volunteers are asked to report daily precipitation the following morning at 7:00 a.m. However, commonly when there is no precipitation the day before, volunteers may make a note of this but do not immediately enter their reported 0.00 inches into the RainLog.org system. We can usually assume that they did not receive precipitation, but no data could also indicate that they are out of town or that they forgot to check their gauge. Whether or not precipitation was collected in their gauge, some volunteers may not report any data until the end of the month anyway or may fill in missing values after they receive the monthly newsletter reminding them to do so.

On most days of the year it does not matter whether or not volunteers enter data because it did not rain anywhere in Tucson; however during the monsoon season and after large frontal storm events, county flood control districts and weather reporters want to know whether or not rain fell in a certain location and how much it rained in different locations. Currently, the number of data entries increases for days that storm events are reported (Figure 4.9); however the data would be more reliable if all users consistently reported daily values so we could determine with confidence whether it rained in a certain location. Unfortunately, users are not required to enter data on a daily basis, so the easiest way to fill in missing values is to do a statistical interpolation that would estimate a rainfall value for a given location based on the values reported from the nearest gauges. This method provides a more complete map than one with several missing values. A fuller discussion about uses of these interpolations is in Section 3.6.2.

2.3.2 Numeric Estimation

As discussed in Section 1.2.1, COOP, the NWS volunteer observer network has experienced data errors due to observer bias, namely from underreporting daily precipitation amounts less than 0.05 inches and overreporting precipitation amounts divisible by 5 and/or 10, or rounding. A recent study by Daly et al. (2007) found that these two types of errors were usually related and that they could be identified by doing ratio tests and frequency analysis of complete data sets. Only 2,807 of the 12,439 COOP stations were analyzed because only these 25% passed the data completeness test, meaning data were reported for 85% of all days each year and that 26 of the years between 1971-2000 had complete records. Of the stations with complete records, 75% of these stations indicated some sort of observer bias, either underreporting values less than 0.05 inches, overreporting values divisible by 5 and/or 10, or both. Although the results cannot be extrapolated for the entire system, it is likely that stations without complete data records are likely subject to observer bias as well, since 75% of the analyzed gauges did not pass one or both of the tests. This high rate of failure suggests a lack of understanding about COOP measurement procedures and/or a lack or commitment to the program, which may be a result of not understanding how their data are used (Daly et al., 2007). This paper will try to resolve whether these findings apply to all volunteer precipitation networks, or whether the results are specific to the gauges used by COOP observers.

A similar type of rounding error known to exist in the RainLog.org system is the issue of users entering values with false precision. Most gauges are accurate to hundredths of an inch, although some users, especially with small rainfall amounts, will enter a value more precise than can actually be read by a gauge. For example, Tru-Chek gauges can only be read in 0.01 increments, but some example data entries are 0.012 and 0.115. With Tru-Chek gauges it is reasonable to estimate the thousandth decimal place for volumes less than 0.05 inches since the lines are farther apart, but it is not good practice to resolve numbers with a greater precision than the design of the gauge allows. These extra numbers are essentially meaningless in calculating daily and annual precipitation totals because thousandths of an inch are such a small value compared to tenths and hundredths of inches.

2.3.3 Wrong Day Reporting

RainLog.org participants are expected to report their precipitation values on a daily basis. However, for a variety of reasons data may get entered into the database on the wrong day. When entering data, participants may inadvertently enter a record on the wrong day, causing a random data error. Anyone, including official weather gauge readers, can occasionally make data entry errors; however systematic errors that occur when data are continuously entered on the wrong day are problematic and need to be identified. Although the daily reporting period is clearly explained in the RainLog.org online handbook, participants may be confused about which day to report rainfall for since the reporting period ends at 7:00 a.m. the following day. For example, if there is a storm event on July 2nd, the user will read their gauge on the morning of July 3rd and should report data online for July 2nd; occasionally this precipitation is instead reported for July 3rd, which is incorrect. Unfortunately, some users do not understand the reporting time period and may consistently enter data one day late. Also, since it rains only 15% of the days in a year, participants may wait to enter data until RainLog.org administrators send out a reminder email at the end of each month (WRCC, 2008).

Both random and systematic errors are hard to detect when looking at the entire data set, but can be identified by analyzing daily data entered by all users in the same geographic area. In populated areas with large numbers of RainLog.org participants, the complete data set is robust and fairly accurately reports when an area received precipitation and when it did not. If we find a single user that reported precipitation on a day when no one else is in the same geographic area did, we can assume that this datum was entered incorrectly (Figure 2.2). The difficulty then becomes determining whether this misreporting was a one-time incidence (random) or if the user always reports his/her data on the wrong day (systematic).

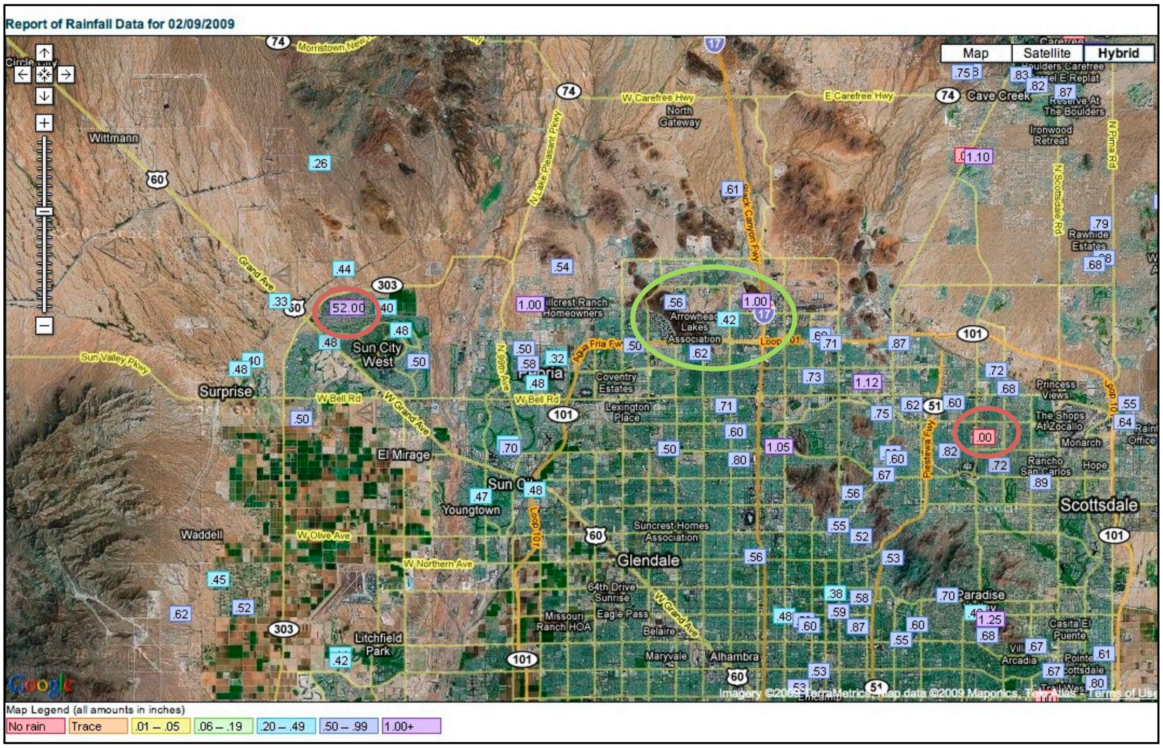

Figure 2.2 - Screencapture created of west Phoenix area on 2/9/09 illustrating the two biggest data errors related to human mistakes. User 724 in Sun City reported 52.00 inches, while the nearest neighbors reported between 0.33-0.50 inches. 0.52 inches would be a likely report, meaning that the user put the decimal point in the wrong location and is now two orders of magnitude higher. As of July 2009 the 52.00 inch value remains. User 3373 reported 0.00 inches of precipitation on the same day, even though the nearest neighbors to this gauge reported between 0.60-0.89 inches. This error is an indication that the user is reporting rain on the wrong date or has not entered any data for the month of February 2009. This user reported 0.00 inches every day from January 1, 2009 through February 21, 2009 when they stopped reporting completely. The green circle illustrates the spatial variability of rainfall entries.

2.3.4 Orders of Magnitude

In addition to reporting data for the wrong day, users may also inadvertently enter a decimal point in the wrong place, which will cause the data to be reported at an order of magnitude larger or smaller than the actual value (see Figure 2.2). For example, if a participant read his/her rain gauge at 0.45 inches, they might accidently enter 4.5 inches or 0.045 inches. Due to the spatial variability of rainfall in Arizona, this error can be hard to catch because a region could potentially get rainfall in one area that was 0.5 inches and nearby, 1.0 inches. Due to the low annual average precipitation across much of Arizona, these errors may significantly affect annual totals, especially if data is being over-reported.

2.3.5 Trace Reports

Due to the spatial variability of rainfall in Arizona, it is not uncommon for participants to receive measurable rainfall, while others within the same city or watershed report only a 'Trace.' Although this trace cannot be quantified by most gauges used in the RainLog.org network because they are not precise enough to measure such small amounts of rainfall, this information helps illustrate the movement of a storm over a region. In addition, many trace reports could add up to measurable rainfall, although quantifying this information is beyond the scope of this study. It is also unclear what environmental benefits trace rainfall provides compared to quantifiable rainfall and how these different rainfall amounts affect aquifer recharge rates, surface water storage and ecosystems including plant and animal water needs.

2.3.6 Gauge Relocation

Data entry errors may also occur if users change their reporting location without updating their profile online to reflect the location change. This can happen if users move to a new address or move the gauge to a new location on their property, but is difficult to identify in a data record. Moving across town or to a new city would have significant implications on the amount and timing of data being reported in this new location. Even relocating a gauge on a user's property could have significant implications on the accuracy of reporting if the microclimate surrounding the gauge changes.

2.4 Spatial Variability Issues

Because people are distributed unequally across a watershed, recruiting an evenly distributed network of volunteer gauges that accurately captures rainfall patterns is challenging. This challenge is exacerbated in Arizona by the state's limited and highly spatially variable rainfall (Sheppard et al., 2002). Any data collected by volunteer networks is subject to misrepresentation without a dense gauge network and RainLog.org must continuously recruit participants and partner with existing networks to enhance its database capabilities.

2.4.1 Gauge Distribution Issues

Previous research has concluded that the standard error of precipitation data in an observation network is dependent on the configuration of gauges within the network. An equidistant, geometrical distribution provides the most accurate spatial interpolation results because gauge spacing is uniform (Morrissey et al., 1995). However, rain gauges are not equally distributed across the land surface because people are unequally distributed and precipitation networks do not have the resources or manpower to maintain equidistant gauges. Simply put, rain gauges are located primarily where humans live, work and recreate and are often co-located with other types of gauging stations, so higher network densities exist in more populated areas and along waterways. For example, the Tucson and Phoenix metropolitan areas account for over half of the total RainLog.org users in Arizona (RainLog.org, 2009). While Maricopa and Pima counties, which house Phoenix and Tucson, respectively, account for nearly 80% of the state's population, these two counties combined only account for 16% of the total land area (U.S. Census Bureau, 2007; National Association of Counties, 2005). The location of these gauges results in three distribution issues:

1. Clusters – usually occur in populated areas where several gauges are located within close proximity (within a mile to a few miles) to each other.
2. Holes – usually occur in unpopulated areas, often correlated with mountainous terrain, where gauges are located very far apart.
3. Filaments – strings of gauges co-located with other monitoring equipment, usually along watercourses and small-scale transects in urban corridors.

Clustered networks raise concerns about the accuracy of interpolating precipitation measurements, even over short distances. When using inverse distance weighting (IDW) or similar equations, the direction of the distances relative to the interpolated gauge value is not considered (see Section 3.6.2 for a longer discussion of IDW). When interpolating values from convective storms that move and change structure on the order of minutes, the interpolated value may not reflect the dynamic nature of the storm and may thus under or over-estimate the value. Although the interpolation may not take storm direction into consideration, the greater benefit of clustered networks, especially with three or more gauges, is that they provide a cross-check for each other. Even a rough interpolation will identify users that have entered data significantly different than nearby users, which is useful for determining data entry problems discussed in Section 2.3. Unlike networks maintained by paid staff, clustered volunteer networks, although not evenly distributed, provide a free self-check mechanism that would otherwise be cost-prohibitive. Using interpolation methods, these clusters will help us find bad data and certainly, the more users that report within a clustered area, the more accurate our results will be.

These densely populated areas have provided RainLog.org with an opportunity to create a densely-gauged network, although gauge distribution varies widely even within cities. O'Connell and Todini (1996) and Chaubey et al. (1999) recommend that to offset spatial rainfall variability errors the use of dense rain gauge networks should be coupled with radar data. These additional data will help researchers better understand rainfall patterns and the significance of precipitation variability in the hydrologic cycle and within basin and catchment-scale water budgets. There are some dense rain gauge networks throughout the country that cover small watersheds, such as the Walnut Gulch Experimental Watershed in southeastern Arizona. While these highly-gauged watersheds provide a good platform for research, they are often in remote locations and do not provide the benefit of observing rainfall patterns in highly populated areas; dense networks in urban areas serve not only as a good source of general data collection, but these data can also be used to study rainfall impacts on land use changes as well as flood control patterns.

Filament networks are convenient because precipitation gauges are co-located with other monitoring equipment like streamflow gauges, air and soil temperature, soil moisture sensors and air quality stations. These gauges are not usually evenly distributed over the land surface either, although they are often located along a linear transect or along a stream reach that allows incremental changes to be seen. Aside from convenience, gauges are co-located to determine whether correlations can be made between different data sets. For example, tracking a stream's response to a storm is aided by knowing how much rain fell at upstream and downstream locations; study sites like the Walnut Gulch Experimental Watershed with 88 rain gauges covering 149 km^2 (58 mi^2), some of which are located at flumes in the stream channels, are designed to collect these high volumes of data (Goodrich et al., 2008). Filament networks certainly provide convenient data and a mutually beneficial partnership when data can be shared among stakeholders, however they do not necessarily provide a complete picture of rainfall patterns. These shortcomings support RainLog.org efforts to assimilate smaller networks to create a larger, more robust network, as well as to continue interpolation work so that managers and users of these smaller networks have a better understanding of what is happening in the geographic area between their gauges.

The converse issue with these cluster and filament networks is the data gaps or holes that result from large, ungauged areas. Unlike populated areas that can use real-time precipitation data for emergency management planning, remote areas do not have the short-term data needs; yet having spatially distributed, long-term data to help predict seasonal water availability and drought regimes is beneficial. The dependence on volunteer participation limits the spatial extent to which data can be densely collected, so development efforts should focus on recruiting volunteers in ungauged locations and improving interpolations in clustered and co-located networks.

2.4.2 Precipitation Variability Issues

The problem of highly variable precipitation patterns adds to the complexity of spatial variability in all precipitation networks, but especially in volunteer networks because of the initial lack of trust inherent with volunteer-collected data. Because of Arizona's unique geography and climate, it is subject to exceptional precipitation variability, both spatially and temporally (Adams and Comrie, 1997), which can result in nearby gauges reporting values that differ by an order of magnitude. Hydrologic processes, including precipitation, occur over a wide range of temporal and geographic scales, however the characteristics of convective storms, which dominate Arizona's summer monsoon can vary on the order of 1 km and within minutes, unlike frontal storm systems that vary on the order of thousands of kilometers and over several hours or even days (see Figure **2.3**) (Blöschl and Sivapalan, 1995). These discrepancies initially raise red flags about the validity of volunteer data, but several researchers have noted the difficulty of quantifying this variability with our existing sparse gauge networks (Germann and Joss, 2001).

Assuming volunteers have followed installation and reporting data guidelines posted at the RainLog.org website, any errors from the type of equipment they use would be minimal compared to potential errors resulting from averaging rainfall over a large area. In fact, uncertainty is often observed in models that predict hydrologic and water quality responses of watersheds due to the input error from the spatial variability of rainfall (Chaubey, 1999). This study showed that using rainfall reported by a single gauge to estimate hydrologic and water quality parameters exceeded the rainfall measurement error, which illustrates the need for a highly gauged land surface. Nijssen and Lettenmaier (2004) also concluded that although fewer errors in hydrologic modeling occur as temporal and areal parameters increase because models better predict at coarser scales, precipitation errors still exhibit the largest spatial correlation errors; therefore efforts should continue to refine gauge network density as well as satellite-based precipitation measurements.

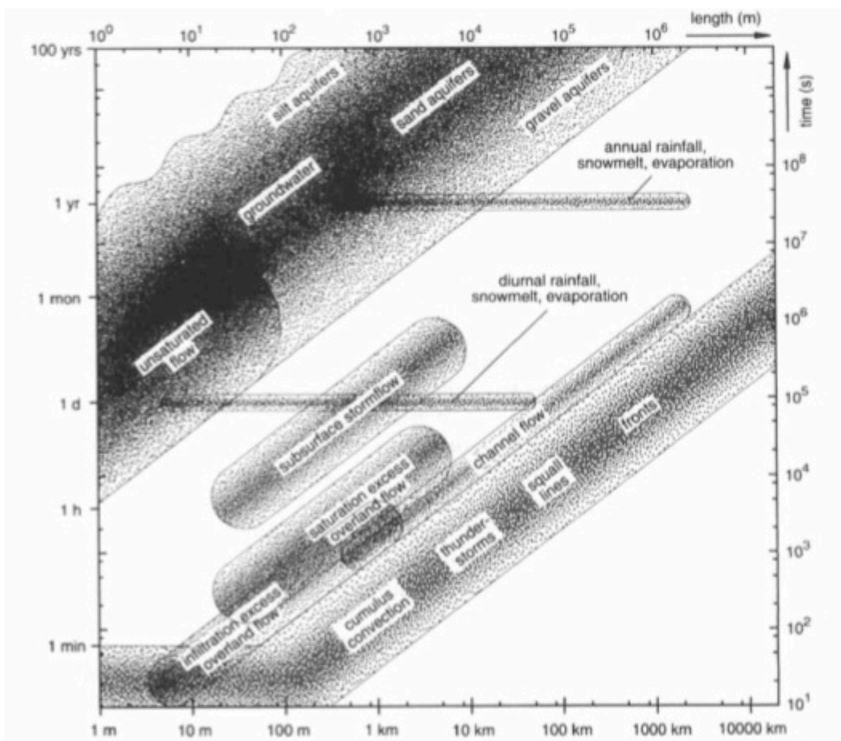

Figure 2.3 - Graphic illustrating time-space scales of hydrologic processes, including the dynamics of convective and frontal storms (Blöschl and Sivapalan, 1995).

Unlike rain gauges, weather radar provides a detailed representation of the spatial variability and temporal evolution of precipitation events over large area (Sinclair and Pegram, 2005). However, radar does not always replicate ground surface activity, particularly in mountainous terrain and semi-arid conditions, and therefore may underestimate rainfall compared to gauge values (Stellman et al., 2001). Combined however, rain gauges can validate radar estimates; once these radar estimates reflect surface conditions they yield an increased understanding of precipitation variability.

3. QA/QC SOLUTIONS AND PROJECT METHODOLOGY

This chapter presents solutions for each of the potential errors highlighted in Chapter 2. Solutions to correct these errors must be uniquely designed, so solutions to each to each error type are discussed and addressed separately. Additionally, this chapter also discusses the data extraction and preparation process and the various analysis methodologies.

3.1 Gauge Type Solutions

Adjusting estimates of reporting errors by gauge type results in a coarse error adjustment approximation because numerous brands are grouped within each classification type. To better resolve reporting errors attributed to various rain gauge brands, a literature review was conducted to determine estimated errors associated with several brands of rain gauges that are used by RainLog.org participants. A product review in Southwest Hydrology (2005) compared three digital rain gauges (LaCrosse, Oregon Scientific and Torrent) and a Tru-Chek gauge and found that their precisions ranged from 0.001 to 0.04 inches. The Oregon Scientific gauge did not list an accuracy, the LaCrosse stated ±3% and the Torrent ±5%. The Torrent consistently reported the largest amount of rainfall, probably due to the greater resolution. All three digital gauges reported more rainfall than the Tru-Chek, which is attributable to evaporation. These variances indicate that reported precipitation is subject to gauge bias, yet if gauges are maintained these differences should be negligible. Table 3.1 highlights one of the brands represented by each different gauge type and indicates that the accuracy of these different gauge types ranges from 0.01 to 0.1 inches, although most publish an accuracy of +/-0.01 in.

During small rainfall events, these accuracy ranges account for a large percentage of the total rainfall. Of the 19,991 rainfall entries during the study period, 7.2% (1,433) were less than or equal to 0.01 in., meaning that for these entries, determining their accuracy would be difficult since the values are only as large as the reported accuracy range. Rainfall entries greater than 0.01 in. and less than or equal to 0.1 in. account for 33.1% (6,618) of all entries. Estimated gauge accuracy accounts for a large percentage of the total rainfall for all entries less than or equal to 0.1 in., which may render these values inaccurate. However, most gauge types, with the exception of straight cylinders, report very similar accuracies, so most rainfall values in this range will result in few detectable biases due to gauge accuracy. To maintain consistency, all RainLog.org gauge types should have a minimum accuracy of 0.01 in.

For all rainfall entries greater than 0.1 in., with a 0.01 in. reporting accuracy, the error percentages would be less than 10%, which is within the same order of magnitude as other potential errors discussed in Section 2.1, so there would be little benefit to resolving gauge accuracy errors. It is also important to note that while the WMO list of errors may compound to 10-25% underreporting, in rainfall values less than 0.1 in. these error percentages are no greater than the accuracy of rain gauges used within the RainLog.org network and the variability of rainfall patterns. This finding supports the fact that error, especially within a volunteer network is unavoidable and often unquantifiable. This analysis suggests that although small rainfall event biases are often negligible, because these small events account for over 40% of total rainfall entries, RainLog.org should recommend gauges with accuracy values of at least 0.01 in. to ensure that data errors due to gauge type are minimal.

Gauge Type	**Company**	**Model**	**Precision**	**Accuracy**	**Price**
Cylinder	Acu-Rite	850		± 0.1"	$7
Electronic	La Crosse	WS-7038UF	0.0105"/tip	± 0.02" (±3%)	$80
Wedge	Tru-Chek			± 0.01"	$12
Funnel Catch w/ overflow	Productive Alternatives	RG202		± 0.01"	$40

Table 3.1 - This chart lists the major gauges types used within the RainLog.org system and illustrates the different precisions, accuracies and prices of a specific brand and model representative of each gauge type.

To address the potential errors resulting from the different gauge types in the RainLog.org network, the factory specifications were collected for representative models from each of the different gauge type categories. Arizona's arid climate and intense storm patterns may result in an underreporting or underestimation of rainfall. The three most common scenarios that would cause underestimation of rainfall in gauges are:

1. evaporation during low-volume events,
2. water loss during high-volume events,
3. and water left in the bucket after a storm event.

The second and third errors are unique to tipping buckets and are addressed separately. If some water is evaporated during every storm event the cumulative total of these underreporting losses could result in a significant percent of the total annual precipitation. Evaporation losses for gauges without funnel devices are 0.8mm/day (0.032 in.) during the late spring in mid-latitudes, which includes Arizona. For winter losses, evaporation is on average 0.1-0.2 mm/day (WMO, 2006). Assuming that Tucson's precipitation is fairly evenly distributed in a bimodal rainfall pattern, the average evaporation from gauges would be 0.05 mm/day (0.012 in.). Based on 84 storm events per year (WRCC, 2008), this would be an annual loss of 4.2 cm (1.65 in.) or 13.6% of the annual total precipitation. Therefore, evaporation is not insignificant when compared to gauge accuracy errors and spatial variability of convective storms, so to minimize these losses, the WMO (1996) and RainLog.org recommend putting mineral oil in the gauge if the gauge cannot be read on a regular basis, which will form a thin film that will help suppress evaporation.

There is no method for analyzing the exact amount of water unaccounted for with tipping bucket gauges and new research in this area is out of the scope of this study, but we will use the LaCrosse tipping bucket as an example of the cumulative effects of gauge underreporting due to environmental factors on gauge design. These effects include water loss during high-volume events due to an inadequate tipping rate that does not tip frequently enough to keep up with rainfall intensity from some storm events and water left in the tipping bucket after a storm event that evaporates before the next rainfall event. The LaCrosse brand gauge (wireless receiver model #: WS-7047U and bucket model#: TX5U) has been purchased for 35 RainLog.org participants to establish a wireless network of tipping bucket gauges to gather near real-time data on local storm events. The factory specifications on the LaCrosse gauge indicate that the bucket tip rate is 0.01 inch/tip. The water lost during each storm event due to the three factors discussed above is always greater than 0.00 mm (in), but most likely less than 1 full tip. To estimate annual losses, we will assume ½ tip average for the amount of rainfall lost from tipping bucket rain gauges after each storm event. In the case of the La Crosse gauge, 0.005 in. is lost per event, so Tucson gauges will lose 0.42 in. per year based on an average of 84 storm events each year (WRCC, 2008). The total average annual rainfall in Tucson is 12.17 inches (NWS, 2009), so the underreporting of precipitation by tipping bucket gauges results in a 3.5% underestimation of rainfall in Tucson each year. If however, the tip rate is decreased to 0.04 in. or 1 mm/tip, which is the standard for many other tipping bucket gauges on the market such as the Oregon Scientific and Taylor models, the underestimation by these gauges would result in a shortfall of 1.68 in. per year or a 13.8% underestimation of Tucson's annual precipitation.

While there is no current standard for systematically correcting the bias on rainfall data due to tipping bucket rates, analyzing this potential bias in future studies should be a priority, since it only affects a certain population within the RainLog.org network that may skew a subset of the volunteer network. However, this type of analysis requires more specific user information to be collected on gauge type and gauge brand, so that gauge specifications can be researched and tested. At a minimum, tipping buckets should be calibrated seasonally and if possible, co-located with a non-recording gauge.

3.1.1 Gauge Type Distribution Solutions

To improve our understanding of system-wide impacts of different gauges being used to collect precipitation data, we first had to develop a classification scheme for the main types of gauges used in the RainLog.org system and then we needed to know how many of each type of gauge were being used to determine whether certain gauge type categories would significantly affect RainLog.org data.

Based on existing literature and an analysis of the RainLog.org database, the five most common gauges types are Funnel Catch, Funnel Catch with Overflow, Simple Cylinder or Rectangular Catch, Tipping Bucket/Electronic Gauge and Tru-Chek or Wedge gauges (Figure 2.1). Although new users have always been asked what type of gauge they have, there was no requirement to provide an answer to this question during the registration process and many users left this information blank.

The first analysis of the database indicated that 22% of all user gauges (n=1,364) were either unknown or unreported, which means that the data entered by these unknown gauges cannot be verified because we do not know the errors associated with those gauges.

Gauge Type	**Count**	**% Total**
Funnel Catch	1	0
Funnel Catch with overflow	52	4
Simple Cylinder or Rectangular Catch	17	1
Tipping Bucket/Electronic Gauge	218	16
Tru-Chek or Wedge	778	57
Unknown	298	22
Total	1,364	100

Table 3.2 - Gauge type distribution of the RainLog.org network as of November 1, 2007 before contacting users with Unknown gauges to update user profiles.

To ascertain gauge types and brands of unknown gauges, emails were sent to each user with an "unknown" gauge. The email sent on January 18, 2008 explained that understanding the total gauge demographic of the RainLog.org system will allow our volunteer data to be more readily used by researchers and water managers. The 22% of users with unknown gauges represented 299 individuals, which represented the second largest gauge category in the RainLog.org system. Several email addresses were no longer valid, so the email was received by only 285 users.

After a two-month response period to gather more information about unknown gauge types, the Gauge Brand and Comments columns were analyzed to determine the correct gauge type for all still “unknown” gauges. If there was enough information to correctly determine the gauge type for a particular user, the Gauge Type column was manually updated with one of the five gauge types; if there was not enough information “unknown” was retained in the Gauge Type column. Concurrently, the database interface was updated to require that all new registering participants select one of the five gauge category options. When users enter their profile information, a description of each gauge type is accompanied by an illustration of the corresponding gauge to ensure that registrants can easily identify the gauge type that best represents their personal gauge.

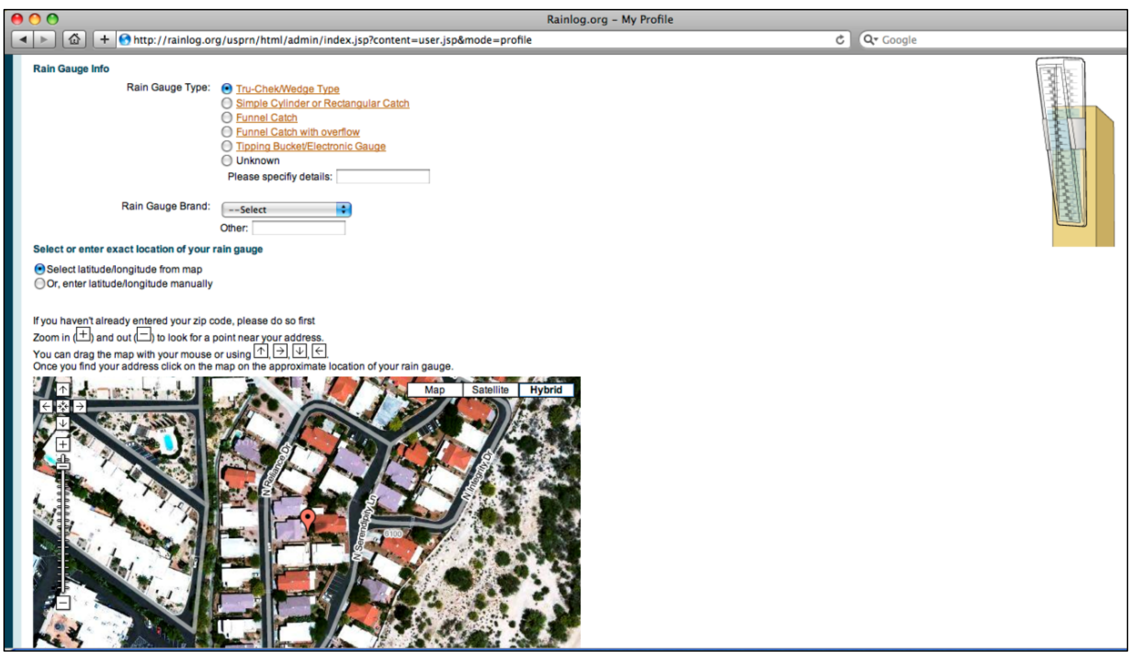

Figure 3.3 - Screen capture image of part of the RainLog.org registration page showing the rain gauge and location selectors. Both have graphic interfaces to help users select the correct gauge and site location.

As of April 4, 2008, 150 of these 285 participants had updated their user profile, resulting in a high response rate of 53%. Although some users still maintained that their gauge type was "unknown/other", most of the people that updated their profile did in fact update their gauge type. In addition to user responses, the Gauge Type column in the user database was manually updated if there was enough information in the Gauge Brand or Comments columns to determine the type of gauge being used. Responses from this email request to update user profiles and manual database updates reduced the percent of unknown gauges to from 22% to 8%, leaving 105 still unknown gauges (Table 3.4). Continued email requests to the remaining "unknown" gauge type users would not likely result in a significantly higher response rate, so these unclassified gauges will remain unknown and people who use this precipitation data will need to understand that if they want data by gauge type this information may not exist. However, the non-response from many individuals may be the result of inactive participants who are no longer reporting data and therefore actual precipitation data will not be skewed by the unknown gauges.

Gauge Type	**Count**	**% Total**
Funnel Catch	16	1
Funnel Catch with overflow	92	7
Simple Cylinder or Rectangular Catch	56	4
Tipping Bucket/Electronic Gauge	234	17
Tru-Chek/Wedge	861	63
Unknown	105	8
Total	1,364	100

Table 3.4 - Gauge type distribution of the RainLog.org network as of April 4, 2008 after 150 of the 285 Unknown gauge users updated their user profiles and Unknown gauges decreased from 22% to 8%.

3.2 Gauge Location Solutions

Of the three potential errors associated with gauge siting (splash, wind and placement under objects), only wind interference can be corrected for if wind speed data are available. However applying a wind speed correction factor to the data would require that we know exactly how high above the ground each gauge is and an anemometer would need to be installed near each gauge. Unfortunately, the RainLog.org system does not ask participants to report how high above the ground surface they installed their gauge, nor do most participants have a way to collect continuous wind speed data. Additionally, the RainLog.org system is set up to record only one data value for every 24 hour period, so the system does not provide enough detail to correlate precipitation recorded during a particular storm event with the average wind speed during that event. The results of the Yang et al. study (1998) and the frequency of data collection within the RainLog.org network indicate the need for very clear gauge placement instructions.

To ensure that a gauge is placed in the best place possible, participants should try to follow these guidelines, in order of importance for ensuring accurate rainfall measurements (see Figure 1.4):

1. The distance from the gauge to any object (e.g., buildings, trees) should be at least two times the height of the respective object above the rain gauge (WMO, 2006; WMO 1983).
    a. For example, if a house is 20 feet tall and the rain gauge is mounted flush on a 5-foot fence, the height difference is 15 feet, so the rain gauge should be placed 30 feet away from the house.
2. Place the gauge in an open space with uniform protection in each direction. Windshields should not exceed twice the distance from the gauge (NWS, 2007).
3. If open space is not an option, the height of the surrounding vegetation should be kept at the same level as the gauge rim (WMO, 2006).

a. Maintaining a constant height around the gauge to avoid wind disturbance may also be achieved by using fence structures or by using a shield around the gauge (WMO, 2006).

Although solutions for some of these problems are offered, RainLog.org does not have any control over how gauges are sited, nor do we have a landscape or backyard view of where the gauges are sited. Therefore no correction factors are applied to the data based on improper siting of the gauges. Additionally, most gauges used by researchers have adjoining shields to protect the gauges from some of the wind effects; however, most RainLog.org participants have unshielded gauges posted on fences and we do not ask whether a wind shield is present or how they installed the gauges, so we assume that wind shields are not present and that most gauges are installed on top of fences. As is listed in the WMO guidelines, installing a gauge on a fence so that the gauge rim is parallel with the top of the fence helps reduce some wind interference and provides an easy platform for mounting the gauge without digging a pit to bury the gauge or installing a special post on which to mount the gauge.

The most important action to ensure that gauges are properly sited by the participant is to make all online directions very clear, using written and visual instructions and real-life examples and photographs. Research in online education suggests that information should be presented in different modes or styles and that learners best retain information when it is contextualized (Ally, 2004), so including good and bad gauge siting examples will help participants understand how and where to install their rain gauges. Before participants install their gauges they should do a site evaluation to find the best location on their property that minimizes siting problems discussed in Section 2.2 and is easily accessible so that participants will continuously take daily readings. As vegetation grows participants need to reevaluate the gauge location every few years; trees especially, may become too tall, breaking the 2:1 rule (see Figure 1.4) and shrubs may encroach on the gauge and cause wind disturbances. If vegetation becomes too much of an interference, participants may consider re-siting their gauge in a more open area of their property. Often in urban areas where people tend to have smaller properties, there may not be any other siting options, but it may be possible to trim some of the vegetation and reduce some of the wind interference.

3.3 Data Reporting Solutions

Solutions for improving the consistency of reporting focus on human error checks, both at the front and back end of the RainLog.org database. Additional training materials may help make some concepts more clear, however time is probably better spent putting warnings on the online reporting form that warns users when data are entered out of the normal range. Phone assistance may also aide users that would prefer direct communication instead of emails. Monthly emails can also include a reminder of how to report data or a request to update profiles if people have moved or changed gauges. On the back end after data are already in the database, mathematical interpolations can help identify data reported on the wrong day, missing data and numerically estimated data.

3.3.1 Missing Data Solutions

The problem of missing values makes a good case for doing interpolations to calculate an assumed value for gauges without data reported. Although a variety of interpolation methods have been used to estimate rainfall patterns over large areas, one of the more popular methods is the inverse distance weighting (IDW) method that gives preferential weighting to the values of gauges nearest to the location being interpolated. The IDW method uses a simple equation that allows for consistent operational usage and is not computationally intensive.

3.3.2 Numeric Estimation Solutions

Based on work done by Daly et al. (2007) that looked at the preponderance of data that were underreported and biased with respect to rainfall values divisible by 5 and/or 10, we decided to perform similar analyses on RainLog.org data to determine whether this bias exists in all volunteer precipitation networks or whether it is directly related to the NWS COOP instrumentation. Although there is no existing solution for mitigating this data recording issue, developing training materials and more consistent communication with users that specifically addresses these issues can mitigate these errors. Providing participants with timely information on how their data are being used as well as stressing the importance of reporting accurate measurements instead of assuming the rainfall was inconsequential or rounding to a number ending in a 5 or 0.

False precision reporting, or entering numbers with more decimal places than a gauge can actually read, may also be a problem within the RainLog.org network. This error could be attributed to entering data with the wrong order of magnitude, which is discussed in Section 2.3.4 or it could occur when volunteers try to resolve data more precisely than is possible. All gauges can measure rainfall to the tenth-inch and many can measure rainfall to the hundredth-inch, but most commonly available gauges do not measure rainfall to the thousandth-inch. Data reported to the thousandth-inch indicates that the user incorrectly entered or read the data. A solution to prevent this from happening would be to have an automated pop-up warning anytime a user tries to enter data with three decimal places.

3.3.3 Wrong Day Reporting Solutions

Finding data entered on the wrong day, either a day early or a day late, is best solved by doing jackknife (basic) interpolations to calculate an estimated value for gauges based on the nearest neighbor values reported. Once the interpolated value is calculated the difference between actual and interpolated values, the residual, can be solved. A large variance indicates that either the user or one of the nearest neighbors has reported on the wrong day. Once these variances are identified, the data can be refined to determine whether users consistently report on the wrong day or whether it was a sporadic event; however, identifying this issue is often easier with frontal storms than convective storms because the extent of frontal storms cause precipitation throughout an entire region. Initial analysis to identify wrong day reporting should focus on months that receive frontal storms. Either way, the user should be flagged until the extent of their wrong day reporting problem is uncovered and communication has been made to address this problem. As was discussed in Section 3.3.1, one of the more popular methods of interpolation is the inverse distance weighting (IDW) method that gives preferential weighting to the values of gauges that are nearest to the location being interpolated. The IDW method uses a simple equation that allows for consistent operational usage and is not computationally intensive, which is discussed in Chapter 4.

In addition to interpolations, the RainLog.org map can also be viewed to look for users that have entered data on the wrong day (see Figure 2.2). While this is an easy method to use for looking one day at a time because the value ranges are color-coded and outliers are therefore easily indentified manually, analyzing months or years of data is cumbersome. Manual analysis is complicated by the fact that two consecutive days have to be viewed at the same time, and the anomaly is visible only if a rainfall event was isolated, so no rain fell the day before or the day after. The monsoon often brings several consecutive days of storms, so accurately identifying users that reported values on the wrong day by looking at the RainLog.org is not a very effective method. Also, since many users do not enter data until the end of each month, this method is not timely for identifying wrong day reporting.

3.3.4 Orders of Magnitude Solutions

Finding data entered with the decimal in the wrong place, which causes the value to be one or two orders of magnitude higher or lower than the accurate value, is best solved by doing interpolations to calculate an assumed value for gauges based on the nearest neighbor values reported. Once the interpolated value is calculated the difference between actual and interpolated values, the residual, can be solved and a large variance may be an indicator that either the user or one of the nearest neighbors has reported data of the wrong magnitude. Once these variances are identified, the user should be flagged and contacted to make them aware of the inaccurate value they entered. They should correct the value if they are confident that they just accidently put the decimal place in the wrong place or they should leave the cell blank if they are not confident that it was merely an order of magnitude problem.

In addition to doing interpolations, the RainLog.org map can also be viewed to look for users that have entered data of the wrong magnitude (see Figure 2.2). While this is an easy method to use for looking one day at a time because the value ranges are color-coded and outliers are therefore easily indentified, it is a cumbersome method for analyzing months or years of data. A front-end solution that may decrease this reporting problem would be to limit the range of accepted values or to prompt the user to confirm that their value is correct if they input a value that is greater than 1.00 inches or less than 0.01 inches. Entries greater than 1.00 in. account for 7.4% of all entries, while 0.054% of all entries are reported less than 0.01 in. so this check would prevent nearly 8% of the most extreme values from being entered. This still leaves the possibility of users reporting wrong order of magnitude values, such as 0.4 inches instead of 0.04 inches, but this input check will help prevent extreme values from being entered into the database.

3.3.5 Trace Report Solutions

Although the value of individual trace entries is negligible, Tucson has 84 storm events on average per year and a Trace is an amount of rainfall less than 0.01 inches, so on average we can say that each trace event is 0.005 inches. If RainLog.org participants in this study report on average, 16.4% of these events, or 14 events, as trace events, annual participation may be underreported by 0.07 inches. Although this not a large number, there is a need to better quantify how cumulative trace reports affect annual precipitation totals for different places around the state. Although trace events may not significantly impact a water budget, they may be crucial to ecosystem health and can offer benefits to plants and some arthropods (Malek et al., 1999), so at least understanding how frequently Trace events occur is important.

3.3.6 Gauge Relocation Solutions

Without a statistical analysis on a long-term data set with more than 10 years of records (Daly et al., 2007), determining a perceptible shift in the rainfall-reporting pattern that may be attributable to gauge relocation would be extremely difficult. This challenge is exacerbated by the fact that RainLog.org is still a fairly young program and most participants only have data starting from when they joined the network, which would be 2005 at the earliest. To avoid undetectable relocation problems, users should be reminded in a monthly reporting email to update their personal profile if they have moved, relocated the gauge or purchased a new gauge. Additionally, there could be a reminder upon login that asks users to confirm their correct personal information. These measures do not ensure that people would update their information, but if people still want to participate in RainLog.org after moving, they would likely update their gauge location, especially after an email reminder.

3.4 Spatial Variability Solutions

The best solution to offset potential uncertainty and interpolation problems when using RainLog.org data is to acquire and develop as dense of a gauge network as possible because denser networks yield better interpolation results that help identify some of the data reporting errors described in Section 2.3. Addressing the issues of holes and filaments within the network may be difficult because holes are attributable to a lack of population and filaments already exist as sub-networks within a larger network. Efforts should focus on ensuring that clustered data are accurate, which can be done while doing interpolation calculations to address solutions described previously in this section.

3.5 Data Extraction Method

The user and precipitation data analyzed for this project were extracted from the RainLog.org database, which is stored in Oracle. That dataset is structured with a unique row for each entry recorded by individual user for each date. To organize the data it was imported into Microsoft Access to create permanent tables that could then be imported to Microsoft Excel for data analysis. Some of the data entries were incomplete because people did not fill in the city or zip code on the registration form. Since my research focuses on the Tucson area, I needed a complete regional dataset and therefore determining unknown cities was important. If city information was missing but zip codes provided Google Earth was used to georeference the city and the Access database was updated. Once all the records were updated, two new tables were created, one that included all the daily rainfall data and a second table that had all the information about each user. The User ID was designated as the Primary Key, or linking field, between the two main tables, allowing relationships to be made among datasets, which was used for data analysis. Unlike Excel, Access data are harder to manipulate because there are no analysis tools, however it does provide a safe location to store data in a relational database without accidentally scrambling or deleting data in Excel. A query in Access was performed to filter only cities in the Tucson metropolitan geographic area. The cities included: Tucson, Marana, Oro Valley, Catalina, Oracle, Corona de Tucson, Vail, Saddlebrooke and Sahuarita. After the query was run and duplicates were removed, 425 gauges remained. Once the permanent database was organized into the correct tables, data were exported into Excel for analysis.

To summarize rainfall patterns over a region, one must know not only how much rain fell in each location, but also how many locations receive precipitation during a given rainfall event. Two columns were added to the spreadsheet, 'Trace' and '>0'. To calculate how many people entered a value for precipitation greater than zero the function =IF(x>0,1,0) was used, where 1 indicates yes and 0 indicates no. In the database, Trace entries are coded as -1, so to calculate how many people reported Trace amounts of rainfall the function =IF(x=-1,1,0) was used. Both of these functions were calculated for every record in the spreadsheet to determine whether they had reported a numeric value or a trace amount of precipitation for each date. Once these values were assigned 1 and 0 values, all records for the one date were collapsed using the Pivot Table function, which performs summary statistics on large data sets and characterizes all data records for the same date or same user into one succinct row or column of data. The Pivot Table function created tables that summed for each date, the total number of users, the number of entries >0 and the number of Trace entries. Through the manipulation of raw data and the creation of summary spreadsheets, the necessary analyses could be performed.

The inverse distance weighting (IDW) analysis required finding the 10 nearest neighbor gauges for each gauge in the Tucson area. The RainLog.org database manager ran a query to find the 10 nearest neighbors, their distance from the actual gauge and their reported rainfall value for each date. The result of this query is a separate spreadsheet for each month in the study period that has a unique row for each date that each gauge reported data, and allows an independent interpolation calculation for each gauge, for each date. Each spreadsheet for a single month of data in the Tucson area contains over 8,000 rows of data, so although not computationally intensive, manipulation of the data is time consuming.

3.6 Data Analysis Methodology

3.6.1 Study Design Methodology

Although RainLog.org was just started in 2005, the database contains precipitation data dating back to 1983 from a few individuals who historically measured daily rainfall and recently entered the values into RainLog.org. While these data provide long-term, historic data sets, the goals of RainLog.org and this project include analyzing a robust set of volunteer data to determine its usefulness for a variety of applications. Since RainLog.org was not developed until 2005, the data set is not really of use to determine user and system errors until the beginning of 2006. For consistency in reporting and analysis, the study period for this project is June 15, 2006 through September 30, 2008. Because we have a bimodal precipitation distribution in Arizona characterized by very different weather patterns, analyzing these precipitation patterns independently is important (Sheppard et al., 2002). These dates were chosen to provide three full years of monsoon-driven, convective storm patterns and two years of frontal storms that occur between September 30 and June 14. The determination of the monsoon is based on the new National Weather Service (NWS) North American Monsoon season from June 15 – September 30 (Poole, June 13, 2008) that has been changed to reflect improvements in technology that allow detectable variations in dew point across regions, making an exact 55 degrees for three days difficult to define on a large scale. As the science behind the North American Monsoon advanced, the Arizona National Weather Service took the lead in redefining the monsoon season in an effort to increase public awareness and knowledge of severe weather by using set calendar dates (Sampson and Pytlak, 2009). RainLog.org supports the mission of increasing the general public's knowledge of climate and weather issues, so for the purposes of this study we will use the same monsoon season definition as the NWS.

The thesis questions laid out in Section 1 include identifying: the most likely causes of error; systematic and random errors; solutions for reducing or eliminating errors; and evaluating the overall data quality. To answer these questions, the following analyses have been performed:

- Gauge type analysis
- Precision estimation analysis

- 5/10 rounding bias analysis
- Trace event analysis
- User reliability analysis
- Seasonal reporting analysis
- IDW interpolation for wrong day reporting
- IDW interpolation for missing data
- IDW interpolation for wrong order of magnitude data

Results and discussion of these analyses are presented in Chapter 4.

3.6.2 Spatial Interpolation (Inverse Distance Weighting) Methodology

To determine which users are entering data on the wrong date and are entering incorrect data, I interpolated an estimated value for each rainfall record based on the inverse distance weighting (IDW) method. For each analysis, a unique reported rain gauge value is compared to 10 nearest neighbor values for each date during the study period and then an inverse distance weighting (IDW) interpolation is used to reduce bias in these comparisons. This method inversely weights data points based on their distance to the interpolated point, so that the nearest points are weighted most heavily and the farther points are weighted less. Recent research suggests that IDW interpolations only provide course estimates of reported rainfall values and should not be used for determining actual interpolated values over large areas with few gauges (Garcia et al., 2008). However, for these analyses, IDW was selected and $d^2$was chosen as the denominator because the goal was to identify potential errors and not to fix errors and fill in missing values based on the interpolated values. As the denominator exponent increases it weighs more heavily the nearest gauges, which would make results too sharp or peaked with only 10 gauges.

To perform this interpolation 10 nearest neighbor gauges and their associated reported rainfall values were extracted for every user in the Tucson area, for every date within the study period (June 15, 2006 – September 30, 2008) and calculated the distance in meters between each user and its 10 nearest neighbors. Each of

Equation 3.5 - The equation for the IDW interpolation, where x_i is the interpolated value, d_i is the distance from each nearest neighbor gauge to the interpolated gauge point and z_i is the actual precipitation value at each nearest neighbor gauge.

$$x_i = \left(\frac{\sum \frac{z_i}{d_i^2}}{\sum \frac{1}{d_i^2}} \right)$$

the 28 months during the study period were exported into different Excel workbooks for easier manipulation. The data were organized in rows by date, which allowed a separate interpolation to be performed on each user for each date, so that individual user errors could be identified. To perform the interpolations (see Equation 3.5) in Excel this calculation had to be divided into four steps so that the numerator and the denominator for each nearest neighbor could be calculated separately. Some nearest neighbor distances were not available, so IF statements were required to ensure that a blank cell calculated as a zero instead of resulting in an error. Without the IF statement the calculation cell would return '#DIV/0!', which would invalidate the entire interpolation.

$$= IF\left(d_i ="",0,\left(\frac{z_i}{d_i^2}\right)\right)$$

Equation 3.6 - The numerator equation where " " indicates a blank cell for the distance between points; therefore if a reference cell is blank the numerator will automatically be zero, otherwise the equation will be calculated.

$$= IF\left(z_i ="",0,\left(\frac{1}{d_i^2}\right)\right)$$

Equation 3.7 - The denominator equation where " " indicates a blank cell for the rainfall value of the nearest neighbor. Therefore, if a reference cell is blank the distance between points will not be included.

The first step of the interpolation is to separately calculate the numerator and denominator for each nearest neighbor, so that 20 Excel columns are added to the existing data, one numerator column and one denominator column for each nearest neighbor. The numerators are calculated first (see Equation 3.6), followed by the denominators (see Equation 3.7). The IF statement calculations are important only in instances where there are not enough nearest neighbors, which results in blank cells, but the same equation is applied to all numerator calculations. If the value z_i is zero, then the numerator equals zero. Because the numerator and denominator are calculated separately, the denominator should also equal zero if the numerator is zero. Otherwise, the denominator might be added into the summary calculation without the associated numerator, which would skew the interpolation. Because the numerator is calculated from a precipitation value and a distance value and the denominator is calculated based on a distance value this problem occurs anytime rainfall is not reported. Therefore, the second step of the interpolation is to calculate the denominator for each nearest neighbor. If the associated rainfall value cell is blank, this statement assumes a zero and this function will apply a zero value to the denominator cell and it is not included in the summary calculation. The denominator equation IF statement is contingent on the rainfall value because if no rainfall data are entered, it should not be factored into the interpolation. This should not be confused with zero values that are actually reported, as they will be included in the overall calculation because these IF statements only exclude blank cells.

The third step of the interpolation separately sums all of the numerator values and all of the denominator values for the 10 nearest neighbors. Once these separate components are totaled, the numerator is divided by the denominator. This fourth step calculated an overall interpolation of one rainfall value for one day based on the 10 nearest neighbor rainfall values. To analyze the accuracy of the data in question, the interpolated value was subtracted from the actual value in a new column. The smaller this difference, or residual, the more accuracy there is among reported values. Highly negative or highly positive residuals indicate outliers that need to be investigated because they may indicate a reporting error. A highly negative residual (< -1) indicates an unusually large value from one or more of the nearest neighbors; if the number is highly positive (>1) it indicates an unusually large value from the original gauge.

4. DATA ANALYSIS AND IMPLICATIONS

4.1 Data Reporting Analysis

This chapter presents the results of the various analyses completed to determine the significance of user and systematic errors within the RainLog.org system. These errors and reporting characteristics are explored to determine the overall impact on the accuracy of RainLog.org data, followed by recommendations to address associated errors. Jackknife interpolation, which compares estimates against measured values for a set of locations different from those used as input data, should reveal all off the reporting problems discussed in Chapter 3.

Once the daily user residuals were calculated, an overall residual was calculated to test whether the data, in aggregate, accurately represent precipitation events and storm distributions. The overall average residual for the 28-month study period is 0.0040 inches and the overall residual for the frontal storm season is 0.0031 inches. These small overall residuals indicate that interpolated data accurately reflect actual reported rainfall values and therefore IDW and similar analyses are useful tools for studying local rainfall patterns. The frontal storm residual is less than the overall residual, suggesting that there is slightly more consistency in reporting during these winter months. This finding supports the knowledge that convective storms are more spatially variable than frontal storms. A positive result also supports the conclusion that overall, reported data is reliable, accurate data; otherwise a negative residual would indicate systematic user reporting errors. However, data may be slightly skewed in the positive direction for two reasons. On any given day, one or more nearest neighbors report zero, while other nearest neighbors report rainfall values greater than zero, which will decrease the daily interpolated value and n some instances, only the nearest neighbors farthest from the gauge will report data and since these nearest neighbors are weighted less the interpolated value will reflect these values with less affect on the overall interpolation.

Analyzing the average residual for each user helps identify users that may be subject to significant reporting errors discussed in the following sub-sections and plotting the average residual for each user indicates that the average residual tends to increase as the number of overall daily entries, or the user reliability (discussed further in Section 4.2.1) decreases (see Figure 4.1). Therefore, encouraging daily reporting improves overall data accuracy, and extreme residual values can be analyzed further to determine which reporting problems they are subject to.

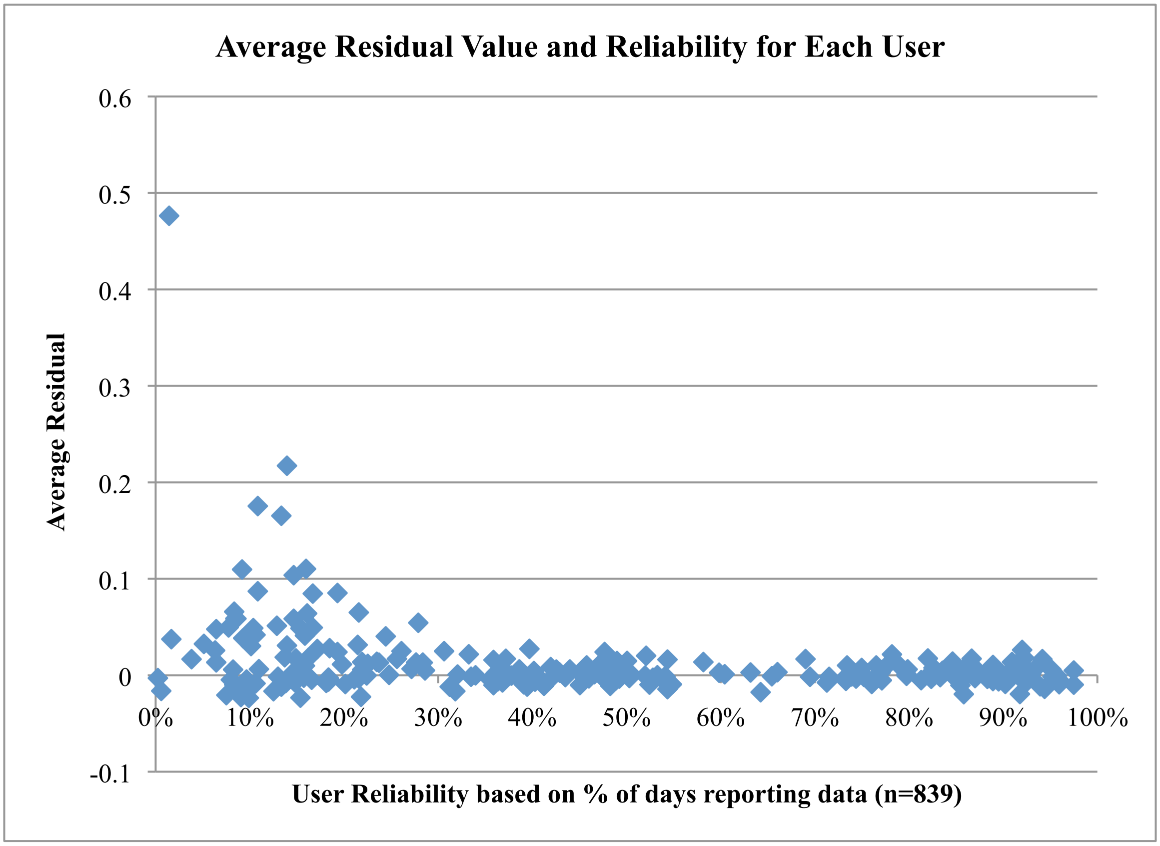

Figure 4.1 - Chart comparing user reliability (based on percentage of days reporting) and average residual value.

4.1.1 Wrong Day Reporting

To identify users that entered data on the wrong day, two analyses were performed. First, the percentage of entries reporting rain for each date was calculated. Storm events, especially, summer monsoon storms, are highly spatially variable, so not all users within the 400 mi^2 Tucson metropolitan region will experience precipitation during each storm event (PAG, 2008). Yet we assume that a significant number of RainLog.org participants would receive some amount of precipitation in a densely gauged network like Tucson. To identify users that reported on the wrong date, a pivot table was created to sum the total number of users reporting Trace and greater than 0.00 in. values for each date. This analysis indicates many instances where only one or a few users reporting rainfall on a given date, suggesting these users probably reported data on the wrong day (see Figure 4.2). The chart also indicate that the ratio of Trace reports to greater than 0.00 in. reports is consistent because the percentage of days with reported rainfall is similar across the entire x-axis. Therefore, trace reports do not skew overall rainfall reporting patterns and do not increase the number of days rainfall is reported by only a few users. This finding supported the decision to exclude trace reports from additional analyses for this study. Without comparing these daily reports to the days preceding and following a uniquely low report, determining a wrong day report would be difficult. Additionally, the spatial component of a rainfall network requires that the wrong day report be compared with its nearest neighbors instead of other random network users. This first analysis provides a helpful summary of user reporting behaviors, but does not provide in-depth information on where these questionable reports are located within the region. Possibly only one user was home to witness the storm event, so he/she reported a Trace even if there was no visible precipitation in the gauge or he/she checked the gauge immediately following the event before the water evaporated. If surrounding users did not check their gauge until the next morning and were not home during the event, they would have no evidence of rainfall. While this is technically a user reporting error, there is no solution because these behavior reporting characteristics cannot be corrected due to the volunteer nature of this network.

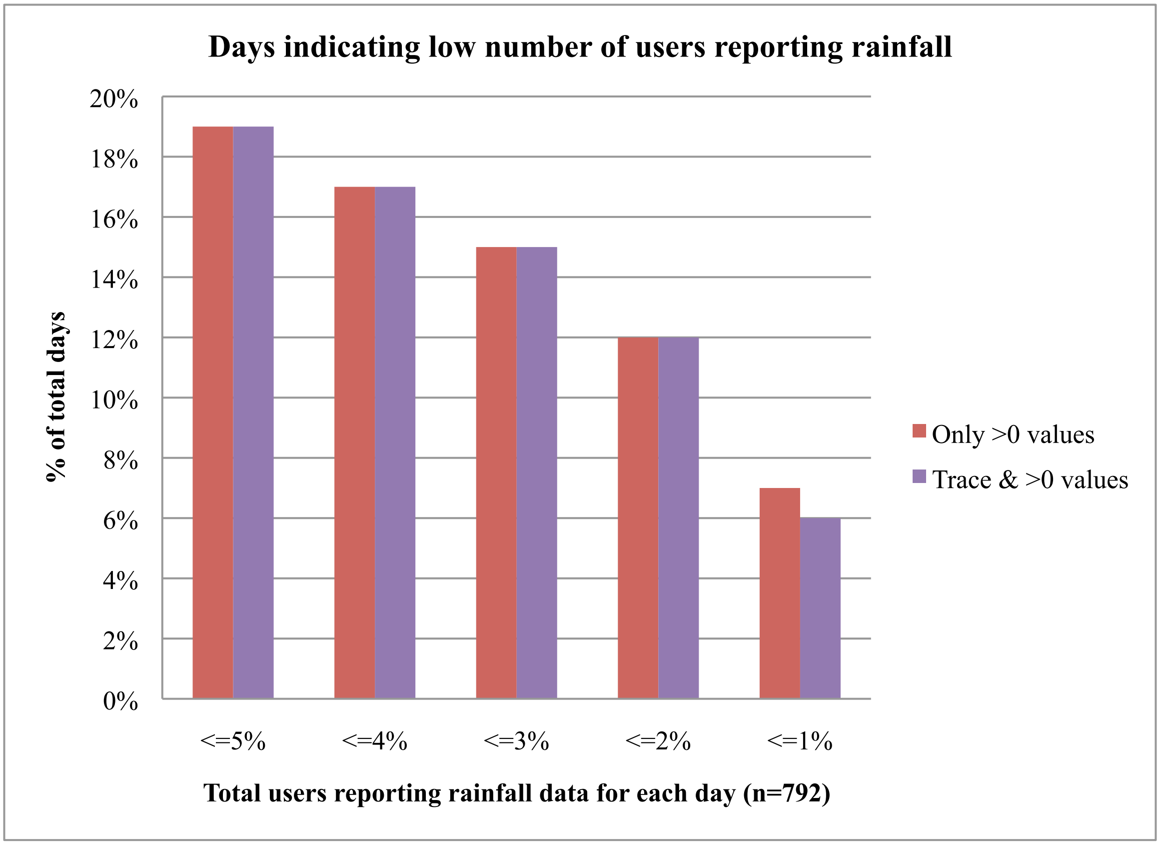

Figure 4.2- Graph illustrating the frequency of days with very low percentages of users reporting rainfall, suggesting wrong day reporting.

To further explore the issue of wrong day reporting, a second, more thorough analysis was performed to compare daily rainfall values from a given gauge to their 10 nearest neighbor gauges calculated from the IDW analysis. Although residuals can help identify users that are consistently reporting incorrectly or not at all, we found that high residual values indicating a user had reported rainfall on a day when none of the nearest neighbors did, were often offset by the following day's residual value that would be an approximately equivalent negative value. These two values would cancel each other out and therefore not provide an overall low or high residual trend that would indicate wrong day reporting. To address this issue we applied unique values for rain, non-rain and no reporting instances for each user and each nearest neighbor value for each date (see Table 4.1).

Type of report	Rain (>0 in.)	No Rain (0 in.)	No report
Unique value	1	-1	Blank

Table 4.1 - Values applied to each user and nearest neighbor data entry depending on whether rain, no rain or no data was reported.

The nearest neighbor values were summed by date and then multiplied by the user value, which was called the reported rainfall (RR) value. If the user and the nearest neighbors are all reporting rain or no rain on the same day, the RR value would be a positive integer (1-10) depending on how many nearest neighbors reports. If the user reports rainfall on a day no nearest neighbors do, or if the user reports no rain and one or more of the nearest neighbors report rainfall, the RR value will be a negative integer, indicating that the user is most likely reporting rainfall on the wrong day. If the user does

	USER	RR	NN1	RR1	NN2	RR2	NN3	RR3	NN4	RR4	NN5	RR5
Day 1	0.26	1	0	-1	0	-1			0	-1		
Day 2	0	-1	0.15	1			0.18	1	0.09	1	0.08	1
	NN6	RR6	NN7	RR7	NN8	RR8	NN9	RR9	NN10	RR10	Count NN RR	User RR
Day 1					0	-1	0	-1	0	-1	-6	-6
Day 2	0.1	1	0.2	1					0.24	1	7	-7

Table 4.2 - Example input table to assign a reported rainfall value to the user and each NN for two different dates. Notice that both examples result in a negative User RR because the user did not enter data consistent with the nearest neighbors.

not enter a value, but the nearest neighbors do, the result will be a '#VALUE' indicating missing data from the user, which will be discussed further in Section 4.1.2.

While negative integers suggest wrong day reporters, the RR value does not take into consideration how regularly users report data and how regularly nearest neighbors report data. These are important factors because there may be a highly negative RR value, but if the user entered data only a few times these are not systematic errors and instead of correcting the data the user's data should be removed from the database. Once the RR value has been calculated for every user, for every day, a reliability factor was multiplied by the RR value to more precisely identify potential wrong day reporters. This reliability factor is an average user reporting factor multiplied by an average nearest neighbor reporting factor (see Table 4.3). The user reporting factor is a percentage calculated by dividing how many days users reported data by how many total days in the dataset. Therefore the more consistently users report data, the higher their user reporting factor. The nearest neighbor reporting factor is a percentage calculated by finding the average number of nearest neighbors that report for each user, so again more nearest neighbors consistently reporting yields a higher NN reporting factor..

Reliability Factors	NN report less (25%)	NN report more (90%)
User reports less (10%)	0.1 x 0.25 = 0.025	0.1 x 0.9 = 0.09
User reports more (70%)	0.7 x 0.25 = 0.175	0.7 x 0.9 = 0.63

Table 4.3 - Reliability factors applied to every RR value to identify systematic wrong day reporters.

This reliability factor applies a weight to all the RR values and the resulting more negative numbers indicate users that may systematically report on the wrong day. Once the reliability factor has been applied to each RR value, a summary of users with the most negative reliability factors can be compiled and individual reports from these users can be explored to look at systematic reporting problems. Table 4.4.a illustrates the daily data from User 312, who has the highest reliability factor and Table 4.4.b lists the top ten users most likely to report on the wrong day based on their quantity of highly negative reliability factors. This analysis was only performed on data for the three winter months each year that receive precipitation from frontal storms: December, January and February. Wrong day reporting is easier to detect during frontal storms because they are larger than convective storms and the likelihood of a user and all nearest neighbors receiving precipitation is very likely, therefore removing bias from storms that are highly localized.

An example from Table 4.3.a of how these reliability factor calculations can be used to identify systematic wrong day reporters is given with user 312. The frequency of highly negative reliability factors suggests that this user may be systematically reporting incorrectly, but an exploration of the data associated with each potential flagged date proves that indeed, the user reports one day early, before the nearest neighbors report rainfall and then reports no rainfall on the following day when most or all nearest neighbors report rainfall. In fact, all of the top five potential wrong day users exhibit similar reporting patterns. Retroactively these data can be adjusted to appear on the correct day. To avoid future errors, these users should be contacted and made aware of their reporting problems.

USER ID 312	Sum of Reliability Factors
Date of reporting error	Daily Reliability Factors
12/29/2006	-1.64
01/20/2007	-2.18
01/30/2007	-1.09
01/31/2007	-3.28
02/02/2007	-0.55
02/20/2007	-1.64
12/03/2007	-4.91
12/08/2007	-3.28
12/12/2007	-4.91
12/22/2007	-0.55
01/04/2008	-3.28
01/07/2008	-3.28
01/08/2008	-4.91
01/27/2008	-2.18
02/05/2008	-3.82
02/06/2008	-4.37
Total	**-45.87**

USER ID	Sum of Reliability Factors
312	-45.87
129	-41.15
485	-32.84
201	-26.06
155	-22.75
15	-21.37
119	-20.01
80	-18.64
337	-17.97
193	-16.24

Table 4.4 (a & b) - Table 4.4.a (left) is an example of user 312 exhibiting wrong day reporting data. Table 4.4.b (right) lists the top ten users with highest negative reliability factors that are most likely to report data on the wrong day.

Systematic errors resulting from wrong day reporting are significant and must be corrected to ensure the accuracy of RainLog.org data; however users with highly negative reliability factors account for only 12% of the 425 reporting gauges in the Tucson area. This does not mean 12% of the data is skewed, instead 12% of users are frequently subject to wrong day reporting behaviors, which represent a small percentage of overall data. With the ability to identify these users, these reporting errors can be corrected in the database and the integrity of RainLog.org data will be preserved. To ensure that all wrong day reporters are identified, this analysis should be continued for all months.

4.1.2 Orders of Magnitude Analysis

To identify users that entered data of the wrong order of magnitude, the residual values calculated for the IDW interpolation were analyzed to identify extremely high (positive) and low (negative) values. Unfortunately, order of magnitude errors were not detected while discreetly analyzing over 100 extreme residual values. Interestingly, over 76% of the most extreme residuals were recorded during the summer monsoon season, 19% were during the frontal storm season (December-February) and 6% were during other months of the year. The high percentage of extreme residual values during the summer monsoon supports our accepted knowledge that monsoon storms are more spatially variable, which can result in nearest neighbors with significantly different rainfall values.

A second analysis to identify users reporting wrong orders of magnitude was done by calculating a ratio of each user value divided by each interpolated value. While this analysis is similar to the residual analysis in that the goal is to identify and explore extreme values, this ratio more easily identifies these order of magnitude errors. When a user enters a value greater than the nearest neighbors, the resulting ratio will be a large number (greater than 4, but up to several thousand) and if a user enters a value smaller than the nearest neighbors the ratio will be a small number, less than 1.0. To identify underreporting, ratios less than 0.2 were filtered, but represent only 1.3% of total entries. Conversely, ratios that are greater than 4 represent less than 3% of total entries, neither of which account for a large percentage of data. Ratios of 4 and 0.2 were chosen as the cutoffs due to the convective nature of storms and the reality that a gauge could receive three times more precipitation than a nearby gauge. These ratios were based on daily observations in Tucson using the RainLog.org graphic interface.

Order of Magnitude Example Errors		
	Greater than NN User 1280 on 7/11/2007	Less than NN User 1201 on 7/29/2007
User Value	0.75 in.	0.01 in.
Interpolated Value	0.002 in.	1.87 in.
Ratio	375	0.005

Table 4.5 - Order of magnitude example errors for users reporting greater than and less than nearest neighbor interpolations.

The difficulty of determining order of magnitude errors is exacerbated by many user entries having none or only a few nearest neighbor values to compare with, as well as nearest neighbors that are often too far from the user gauge to depend on the interpolated values. This causes interpolated values that are lower than expected (see Table 4.5). This analysis also determined a noticeable trend in gauges reporting 2 or more inches at the beginning of several months; this value most likely indicates the total rainfall for the previous month instead of a daily value and should not be included in the daily rainfall database. Additionally, edge effects can skew the data, especially when a user gauge is located along the mountain front and all of the nearest neighbor gauges are in one direction. Precipitation tends to increase along mountain fronts, especially as elevation increases, so it is not unlikely that a user accurately report a value that is an order of magnitude higher than its nearest neighbors.

To further address this issue, a denser gauge network including elevation information is needed to better model the horizontal and vertical relationships between gauges. Further exploration of residual values may reveal examples of order of magnitude errors, however automating a front-end check on data being entered into RainLog.org would help eliminate any question of whether extreme residual values are due to order of magnitude problems. This automated front-end check would prevent a user from entering a value less than 0.01 in. or greater than 1.0 in. without first confirming that the correct value is being entered.

4.1.3 Missing Data Analysis

Missing data can result from a variety of user behavior issues and examining missing data patterns helps identify solutions for the different reporting behaviors. Generally we see that the most users report data when it does rain and the fewest users report data when it has not rained for several weeks, which is probably attributed to users believing it is more important to report data for rain conditions than non-rain conditions and indicates that missing data are non-random. However, a lack of data for non-rain days can skew interpolations, so identifying cases where no data should actually be a no rain condition is critical. To identify users that have significant amounts of missing data, the same methodology was used as the second analysis for identifying wrong day reporters (Section 4.1.1). The reported rainfall (RR) values for both users and nearest neighbors (NN) were compared using IF statements to compare six different conditions that would analyze user behavior reporting patterns (see Table 4.6). Although each condition provides information about reporting behaviors, conditions 2 and 4 are the most important for this analysis because they indicate users that fail to report data when their nearest neighbors report precipitation.

Condition	User Behavior	Nearest Neighbor Behavior
1	User reporting no data	NN reporting no rain
2	User reporting no data	NN reporting rain
3	User reporting no rain	NN reporting no rain
4	User reporting no rain	NN reporting rain
5	User reporting rain	NN reporting rain
6	User reporting rain	NN reporting no rain

Table 4.6 - Description of the six different conditions that were compared to determine users with missing data.

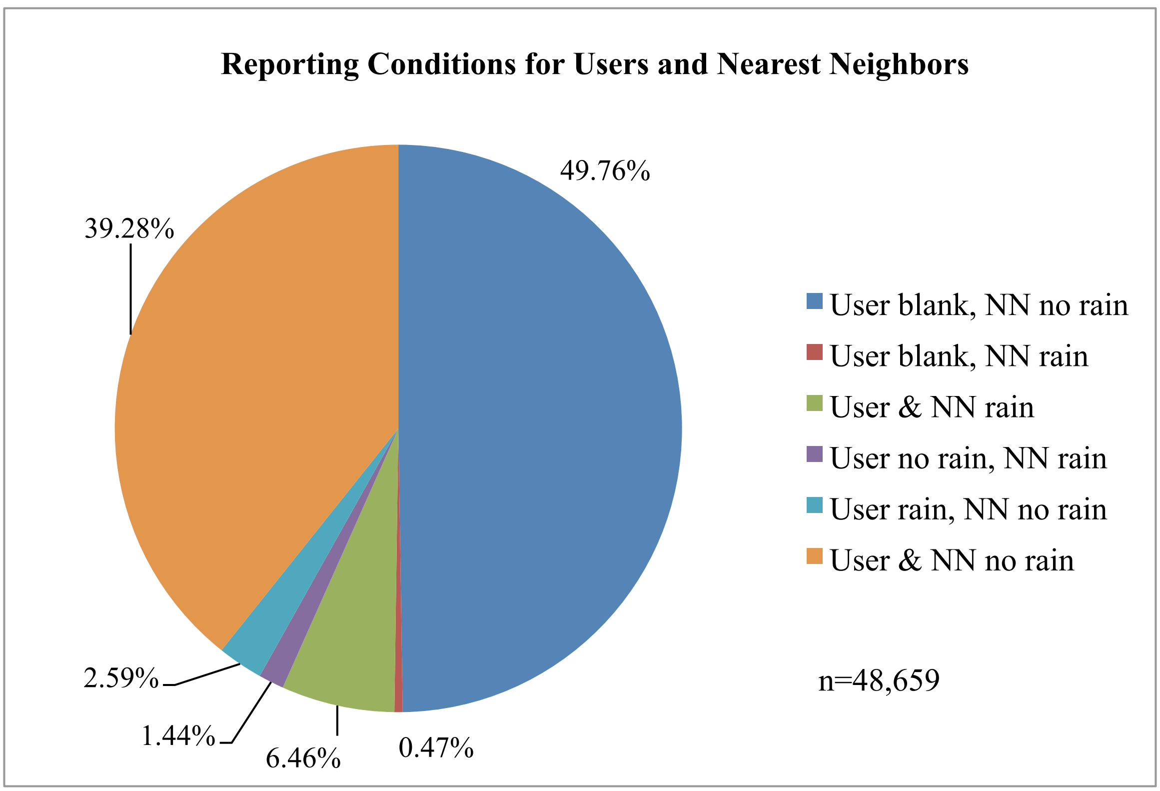

Figure 4.3 – Pie chart illustrating average user reporting for each of the 6 behavior conditions listed in Table 4.4.

Results from the missing data analysis indicate that the most common type of data condition (49.8%) is when a user does not report rainfall and no nearest neighbors report rainfall, followed by 39.3% of entries representing neither users or nearest neighbors reporting rainfall. Summary statistics for these two conditions indicate a significant amount of variance in user reporting behaviors (see Table 4.7), suggesting that some users rarely report data unless it rains, while others report consistently, regardless of precipitation. Analyzing users that do not enter data when their nearest neighbors do

User blank, NN rain		***User no rain, NN rain***		***User & NN reported rain***	
Mean	0.47%	Mean	1.43%	Mean	6.43%
Standard Error	0.04%	Standard Error	0.10%	Standard Error	0.27%
Median	0.00%	Median	1.10%	Median	5.52%
Mode	0.00%	Mode	0.00%	Mode	0.00%
Standard Deviation	0.73%	Standard Deviation	1.68%	Standard Deviation	4.46%
Sample Variance	0.01%	Sample Variance	0.03%	Sample Variance	0.20%
Kurtosis	-19.76%	Kurtosis	277.94%	Kurtosis	-119.55%
Skewness	118.36%	Skewness	165.45%	Skewness	15.46%
Range	2.31%	Range	7.73%	Range	16.02%
Minimum	0.00%	Minimum	0.00%	Minimum	0.00%
Maximum	2.31%	Maximum	7.73%	Maximum	16.02%
Count	271	Count	271	Count	271
User blank, NN no rain		***User no rain, NN no rain***		***User rain, NN no rain***	
Mean	50.03%	Mean	39.05%	Mean	2.58%
Standard Error	2.13%	Standard Error	1.84%	Standard Error	0.13%
Median	51.38%	Median	38.67%	Median	2.22%
Mode	49.72%	Mode	0.00%	Mode	0.00%
Standard Deviation	35.06%	Standard Deviation	30.33%	Standard Deviation	2.16%
Sample Variance	12.29%	Sample Variance	9.20%	Sample Variance	0.05%
Kurtosis	-140.60%	Kurtosis	-142.48%	Kurtosis	129.07%
Skewness	-3.78%	Skewness	3.20%	Skewness	98.61%
Range	98.88%	Range	82.87%	Range	12.15%
Minimum	0.00%	Minimum	0.00%	Minimum	0.00%
Maximum	98.88%	Maximum	82.87%	Maximum	12.15%
Count	271	Count	271	Count	271

Table 4.7 - Table of summary statistics for each of the 6 conditions analyzed for to identify users with missing data.

enter data indicates that very few user entries (0.47%) fail to report rainfall when one or more of their nearest neighbors do. This analysis suggests that missing data is not a significant data reporting problem, although identifying users with higher than average reporting in this category will help target and eliminate any missing data reporters.

The reporting average of users reporting rain, when nearest neighbors report no rain and users reporting no rain, when nearest neighbors report rain are also fairly low, at 2.58% and 1.43% respectively. Most importantly, based on the two conditions, user blank, while reported NN rain and user no rain, while NN reported rain, we are able to identify users who are subject to missing data and report in these categories more frequently than the average. These users are flagged in the database and listed below in Table 4.8.

User ID	User no rain, NN rain
312	7.73%
352	7.18%
193	7.18%
1242	6.74%
1599	6.63%
1433	6.63%
119	6.63%
288	6.63%
148	6.08%

User ID	User blank, NN rain
78	2.30%
1502	2.27%
3203	2.27%
3160	2.26%
2813	2.25%
3017	2.25%
3159	2.25%
1551	2.23%
3122	2.22%

Table 4.8 - Top 10 users that exceed average percentage reporting for each of the 2 conditions that indicate missing data.

Users that fail to report rainfall when rain is being reported by one or more nearest neighbors account for less than 2% of data entries, indicating that missing data is not a significant problem with RainLog.org data. However, due to the sample variance of these reported values, users that exceed the category mean should be flagged as users that may have missing data. Additionally, to create a more robust data set, zero values should be entered for all blank cells that correlate with non-rain days. This can be done either retroactively or a check can be built into the online reporting form that requires a zero or value greater than zero to be entered for every reporting day or a fill-in fuction at the end of each month for every day that was left blank. Assuming zeros for days that no rain was reported by the user or nearest neighbors provides a more complete and accurate data set that can be used more robustly.

4.1.4 Numeric Estimation Analysis

The 5/10 rounding bias was determined by filtering the database to find all reported rainfall values, excluding Trace and zero entries. Then, the COUNTIF function was used to find all entries that were divisible by 5 or 10. These entries ending with a 5 or 0 account for 68% of the total reported rainfall entries, which indicates a substantial rounding bias, since probability suggests that numbers that end with a 5 or 0 only account for 20% of the possible entries. While this is technically an inaccuracy, it does compare well with the results from the Daly et al. (2007) COOP study that found 75% of the stations failed the observer bias tests, meaning these users exhibited 5/10 biases. Rain gauges tend to encourage rounding because they are designed with larger ticks at the 5/10 intervals and rounding off is an acceptable, common practice in many disciplines that involve taking measurements. Assuming the volunteer enters the data on the correct date, these errors are acceptable because they are only off by two-hundredths at most, which is no more than a 10% error and within the range of other potential errors that exist within the RainLog.org network.

To more effectively analyze the 5/10 gauge bias, rainfall data were separated by gauge type to determine whether reporting biases exist due to different gauge type. To determine whether a 5/10 bias existed, the dataset was separated by all values divisible by five- and/or ten-hundredths of an inch and those values not divisible by five- or ten-hundredths of an inch. Then, based on the work of Daly et al. (2007), the mean of each population was calculated; a significant difference would indicate a 5/10 bias. In all gauge types, the mean of values ending in 5 and 0 is greater than the mean of all other values, which validates a 5/10 reporting bias (see Figure **4.9**). However, cylinder type

Gauge Type	**Mean of values excluding 5 & 0**	**Mean of values ending in 5**	**Mean of values ending in 0**	**% 5 and 0 values of total population**	**% 0.01 values of total population**
Cylinder (n=1,472)	0.317	0.406	0.510	**76.2%**	2.58%
Funnel Catch (n=214)	0.211	0.424	0.484	36.0%	24.3%
Funnel with Overflow (n=4,023)	0.254	0.344	0.482	21.5%	7.56%
Tipping Bucket (n=4,322)	0.267	0.403	0.513	25.8%	7.91%
Tru-Chek (12,225)	0.239	0.435	0.546	37.5%	6.63%
Unknown (n=2,030)	0.241	0.405	0.493	46.6%	6.56%

Table 4.9 - Results of the 5/10 bias analysis, indicating that all gauge types express some gauge bias with respect to values that are divisible by five and ten-hundredths, although more so for values divisible by ten-hundredths.

gauges also report an extreme amount of 5 and 0 values (76.2%) that account for all reported values, suggesting a significant 5/10 bias.

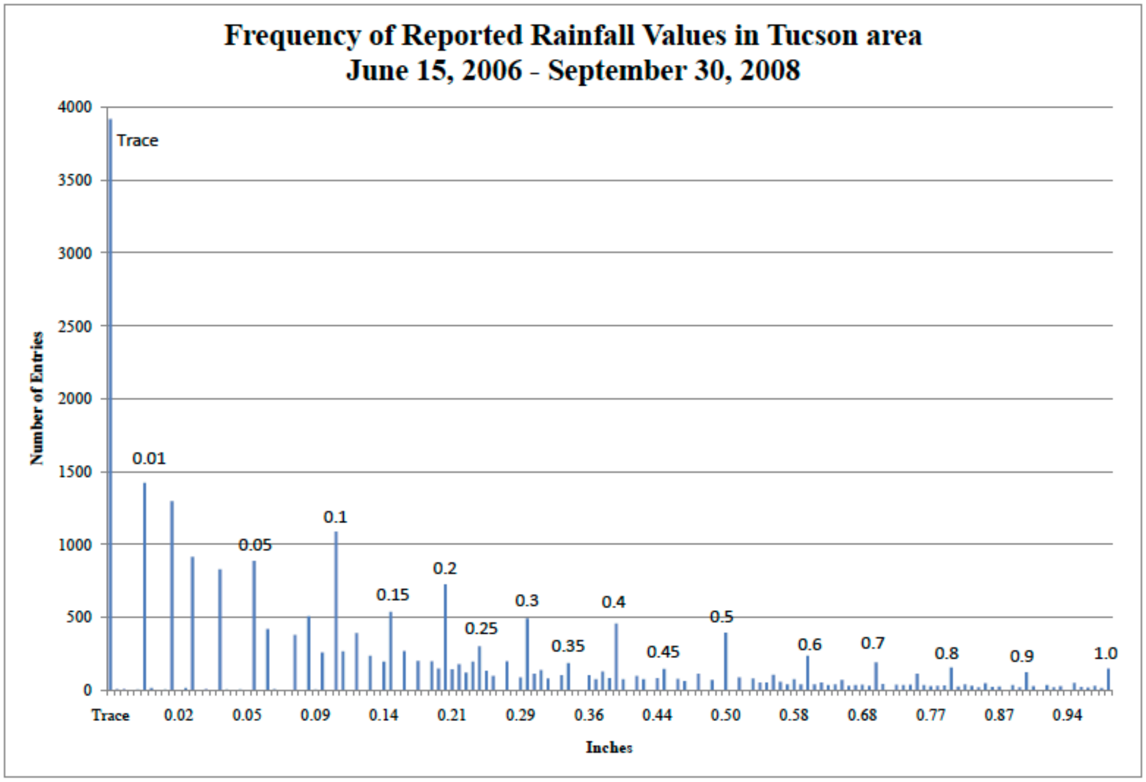

Figure 4.4 - Graph of the frequency of all reported rainfall values during the study period. The spikes indicate that a 5/10 bias does exist in our data because the spikes occur only in values divisible by 5 or 10.

To further explore the 5/10 bias, the frequency of rainfall entries that end with a 4 or 6 and a 9 or 1 were analyzed. Due to rounding that results in a 5/10 bias as well as people's instinct to be able to report the largest possible quantity of rainfall, we thought than the entries ending in 6 and 1 would have a higher frequency than the 4 and 9 values that would be rounded up. However, there is no evidence that this occurs with 4 and 6 values, as the total count of values ending in 4 is 732 and of those ending in 6, the frequency is 728. There is only a slight difference between 9 and 1 because the total count of values ending in 9 is 507 and values ending in 1 is 551, which results in a 4% difference.

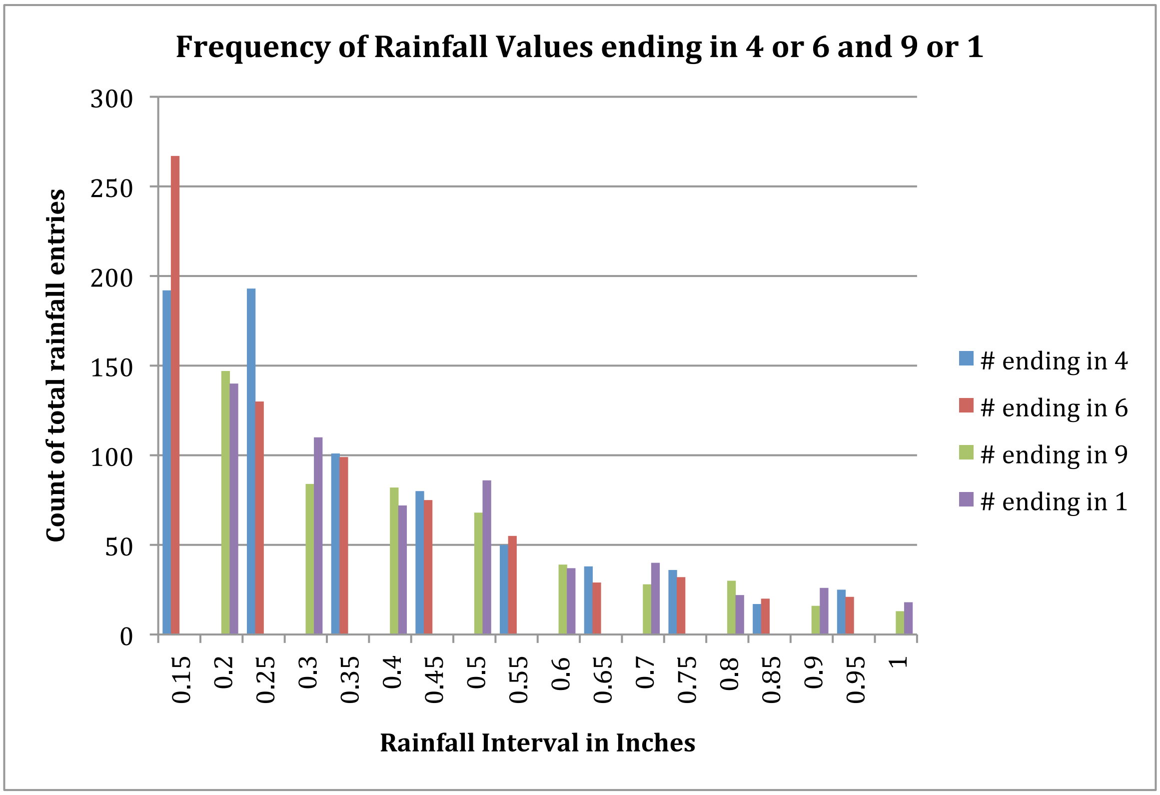

Figure 4.5 – Graph of the frequency of values only ending in 4 or 6 and 9 or 1 to determine whether there was a bias toward rounding upward by reporting larger values, like 6 and 1. There is a slight bias toward numbers ending in 1 over 9, but it is not statistically significant.

The 5/10 bias analysis confirms the results found in the Daly et al. (2007) study that systematic 5/10 bias does exist with RainLog.org data. However, considering the convective nature of storms in Arizona and the priority of collecting data from a dense gauge network, correcting the 5/10 bias is less important than other data reporting corrections because gauge design encourages this bias and the 5/10 bias does not significantly skew data. Educating participants about the influence of different rain gauge types may help reduce some of this bias, although it will probably always exist to some degree.

To determine whether false precision was a significant problem with RainLog.org data, the database was filtered to find data that had reported values extending to the thousandths. Of the 425 registered users in the Tucson area, only 28 users (6.6%) have ever entered data with three or more numbers to the right of the decimal point and of these 28 users, only 8 users did this more than five times. Most of these users had Tru-Chek type gauges, so they should not have been able to resolve their data to such a precise value. False precision accounts for only 0.12% of all data entries and of these, only 0.013% are less than 0.01 inches. These findings suggest that neither false precision or order of magnitude under reporting are significant sources of error for RainLog.org. False precision does not affect the accuracy of data when comparing to nearest neighbor gauge values and the extent of the problem is not very large, so no further analysis needs to be done to correct this type of data error. An automated front-end check could prevent false precision from occurring by preventing users from entering data with more than three decimals unless they confirm their input value.

4.1.5 Trace Report Analysis

System wide, Trace reports account for 3.6% of all data entries. As a comparison, all data entry values greater than 0.00 in. account for 18.1% of all data entries, so Trace reports actually account for 16.5% of all non-zero entries in the Tucson area (see further discussion in Section 4.2.2). While the quantity of this rainfall remains miniscule, even compared to Tucson's low annual average rainfall, it remains significant in ascertaining the spatial variability of rainfall and ecosystem responses from these small quantity events. Unlike other behavior-related errors, trace rainfall reports do not inherently result in errors that skew total reported rainfall values. The first analysis in Section 4.1.1 offers evidence that Trace rainfall is not reported disproportionately. Applying a correction factor is not necessary, however researchers interested in storm patterns and rainfall distribution may want to perform additional spatial analyses comparing the proximity of Trace data reports to each other, to greater than 0.00 in. values, and to radar data during single storm events and for long-term records.

4.2 Volunteer Analysis

Of all the RainLog.org network participants, there are 425 participants located within the Tucson metropolitan area, which includes Tucson, Marana, Oro Valley, Catalina, Oracle, Corona de Tucson, Vail, Saddlebrooke and Sahuarita. Of these 425 participants, 9% or 36 individuals have never actually participated in the RainLog.org system.

When compared with volunteer organizations across the United States, RainLog.org compares very well with the national average volunteer retention rate of nearly 70% for all age groups over 35 (Corporation for National & Community Service, 2007). Participation efforts within RainLog.org vary widely, but the fact that over half of the registered users are still active and that the average number of participation days is over one year suggest that recruitment efforts are successful in getting participants committed to our first expectation, data reporting. These findings are comparable with CoCoRAHS's experience that ~50% of participants remain with the project after initial training (Reges et al., 2006). If CoCoRAHS is not experiencing any greater retention than RainLog.org, additional time and money invested in hosting training sessions may not increase participation. RainLog.org's online reporting and data visualization fosters a virtual community that appeals to volunteers who want to participate because it is easy to do at home and does not involve others, but offers a sense of community involvement. For these reasons, hosting training sessions and continuing education opportunities would not necessarily attract existing or new participants. Additional participants should be recruited through media outlets that emphasize the need for dense gauge networks, as well as the ease of use and accessibility of RainLog.org.

4.2.1 User Reliability

To take a closer look at user reporting patterns and determine whether participants are meeting our RainLog.org volunteer expectations, 2007 data from all users in the Tucson area who were registered as of January 1, 2007 was analyzed. To characterize the volunteer reporting patterns in the Tucson area, initial registration dates were compared to last readings entered. Of the 425 participants that have registered in the Tucson area, 26 users never entered any data and 10 users entered data only once. Of the 425 registered users, 301 users, or 71%, are considered active, meaning they have entered data within the last two months. The average number of days users have participated in RainLog.org is 490 days or about 1.3 years; 55% (277) of users have participated for at least one year, 24% (121) for two years and only six users for three or more years.

To address the question of whether RainLog.org volunteers are reliable, which means they follow our expectations for an ideal RainLog.org volunteer (see Section 1.5), a reliability percentage was applied to each user, based on how many data entries they made over an entire year. To analyze user reliability in the Tucson area for 2007, only those 232 users that had registered by January 1, 2007 were included; once users who registered, but never entered data or have since stopped reporting were sorted out, 150 users who had entered data at least once in 2007 remained. Results indicate that the reliability mean is 62.2%, meaning that on average these 150 individuals reported data on 226 of the 365 days in 2007. This statistic is not impressively high, although the mode is 100%, indicating that the largest group in this population is users that report data every day (See Figure 4.5).

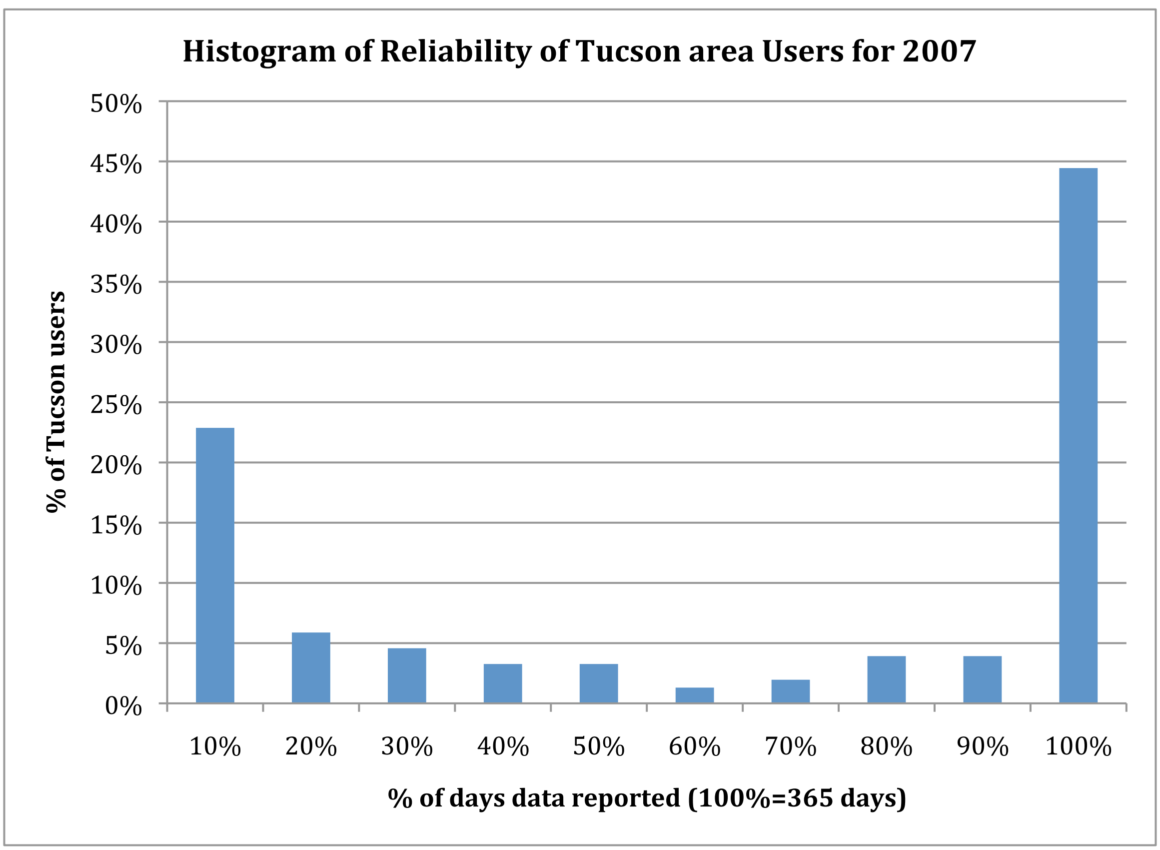

Figure 4.6 - This graph shows that although the user reliability mean is only 62%, nearly 45% of users always, or almost always report data. Unfortunately, nearly 25% of users report data on less than 10% of days.

Comparing the number of days rainfall is reported is important due to the discrepancy in averages and Tucson's variable precipitation patterns. Users reported rainfall on 182 days, or 50% of all days in 2007. Our own observations tell us it does not rain 50% of the time in Tucson and the Tucson NWS station reports measurable rainfall on an average of 53 days, or 15% of days per year (WRCC, 2008). The fact that rainfall is reported on so many days is a red flag that users are reporting data on the wrong date and analysis of this issue is discussed in Section 4.1.1.

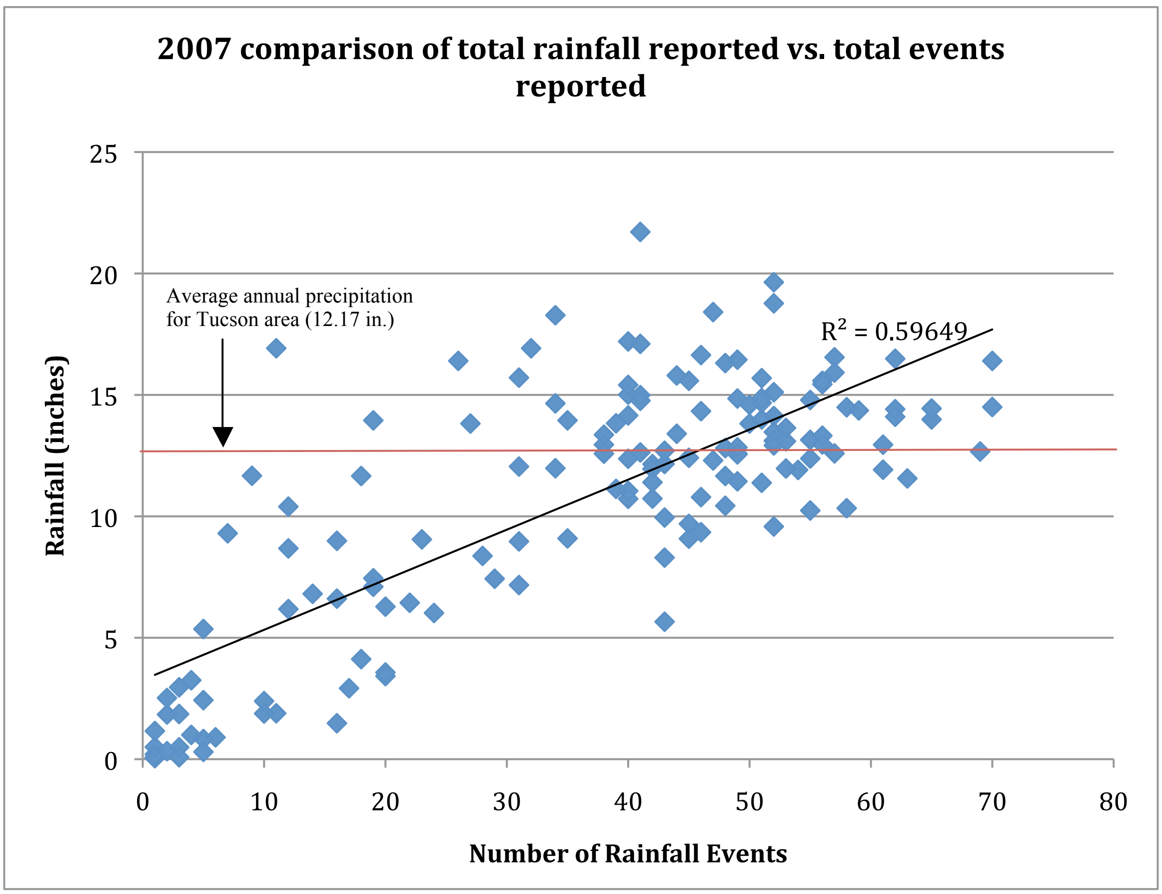

Figure 4.7 - XY-Graph of individual users reporting behaviors, comparing the number of rainfall events entered in 2007 (x-axis) to the total annual rainfall value entered (y-axis). These plots indicate a logical trend that annual rainfall increases as the number of storm events increases. Also note the cluster near zero, indicated users never or rarely reported and the cluster ranging from 40-60 events (x-axis) indicating the range about 60% of users fall into.

Once overall reliability was determined, the reporting frequency of storm events was explored since our hypothesis is that users are more likely to report data on days that it rains. Analysis of the data reported in 2007 by these 150 users suggests that individuals reporting data with less than 62% reliability are skewing total annual rainfall values. The mean number of rainfall events is 36.5 (S.D. 19.1), while the mode remains at 52, suggesting that many people are not reporting data for every storm event. In fact, the range of storm events was from 1 to 70 (see Figure 4.8). The mean annual rainfall was 10.78 inches (S.D. 5.11) which is nearly 1.5 inches, or 11% lower, than the historical annual average of 12.17 inches (NWS, 2009).

A mode of 52 events per year compares very well with the NWS long-term average of 53 rainfall days per year and the median rainfall of 12.3 inches suggests that user average would be closer to the historical average if the users reporting low annual rainfall totals and low storm event counts were removed. To calculate total rainfall and storm event means that are closer to their respective median and mode values, these unreliable users should be excluded.

Summary statistics were recalculated after excluding all users that reported 30 or fewer events per year (less than 62% reliability). The storm event mean increased to 48.01 (S.D. 9.05), while the median was 48 and the mode remained 52. Although there was still some variability in annual rainfall, the mean increased to 13.41 inches (S.D. 2.62) while the median increased to 13.32 inches and the mode was 12.58 inches. While removing participants with less than 62% reliability improved the precision of storm event reporting, total annual rainfall for these 103 users remains variable.

To ensure that only data from reliable users is included in the RainLog.org database, inactive user data should be filtered out regularly and not assumed to be a zero value. In addition to monthly emails, follow-up emails may help entice some volunteers back into reporting. The variability of total rainfall reported for 2007, even after less reliable users were eliminated, suggests that the spatial variability of convective storms results in significantly different daily rainfall reports across a region. This variability is of great interest to many researchers, as well as emergency managers, and should be explored with more detailed modeling.

4.2.2 Seasonal Reporting Characteristics

Since the characteristics of frontal and convective storm events are different in duration, intensity and spatial coverage, it may be informative to compare the reporting characteristics and trends of data entered by participants during these different periods. On average, 32.62% of participants report rainfall on any given day during the monsoon season, whereas only 11.55% of participants report rainfall on any given day during the rest of the year.

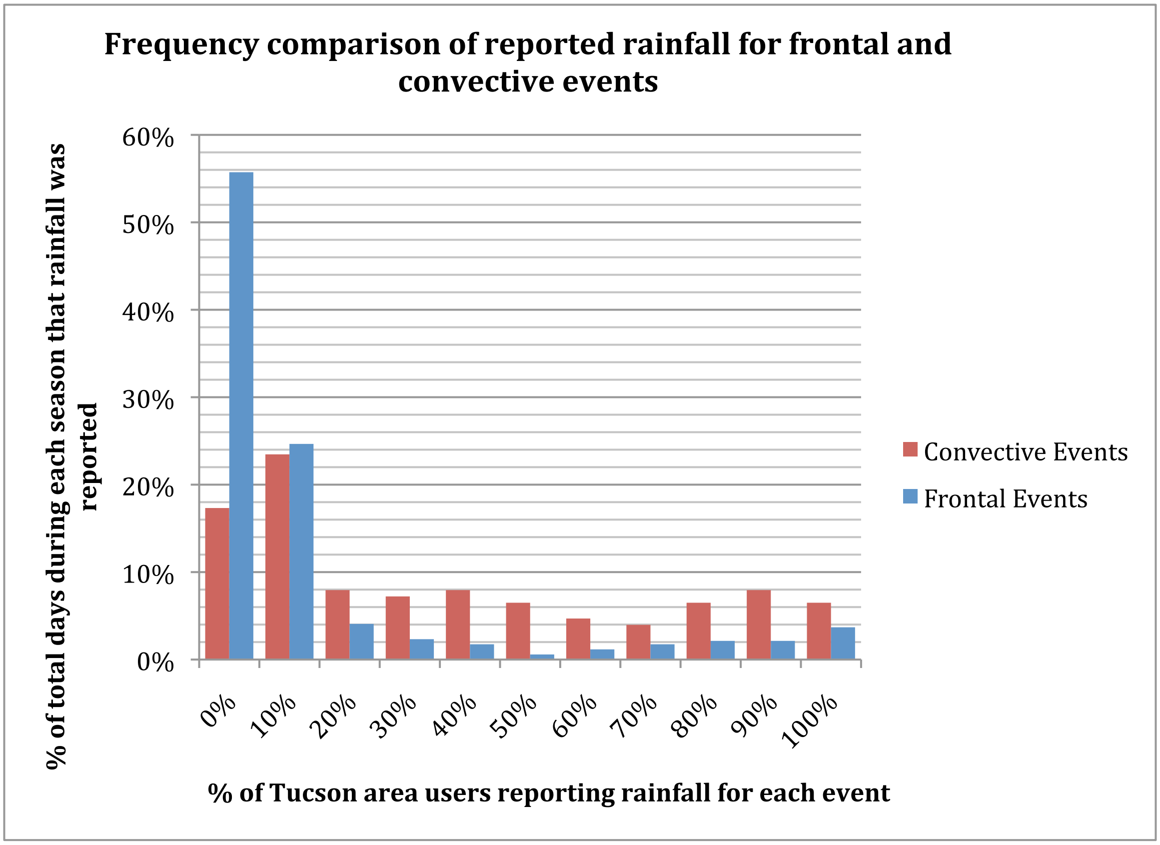

Figure 4.8 - Graph illustrating the difference between the percentages of users on a daily basis that receive rainfall in frontal and convective events.

The graph in Figure 4.8 was created using a histogram to bin the percent of users reporting rainfall from zero to 100%. Unsurprisingly, for frontal storm types, most daily reports indicate no rainfall. For convective storm types, the highest category of users reporting rainfall is 10% to 20%, meaning that there are more days during the monsoon season than the rest of the year when some people throughout Tucson report precipitation in their gauges. Additionally, convective storms have more users reporting rainfall when between 20% to 90% of all users report rainfall. However, the frequency of frontal storms when 90% to 100% of all users report rainfall is higher, which supports the fact that frontal storms tend to cover entire regions and therefore more users would receive precipitation at their house.

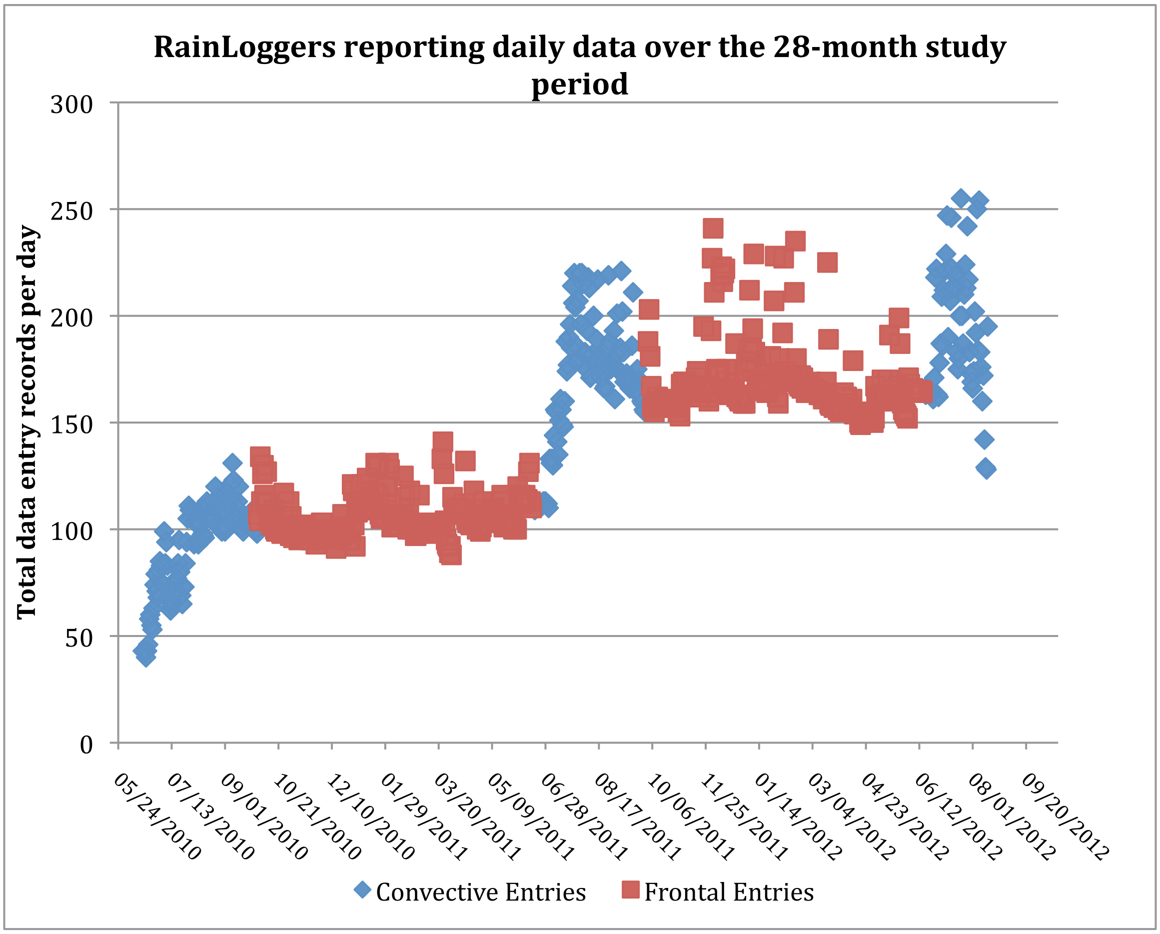

Figure 4.9 - Graph comparing the percentage of users that report rainfall on a daily basis and total daily entries including 0, Trace and reported rainfall.

Figure 4.8 shows that more, new participants enter data during the convective season than the frontal season, which is nearly three times as long. Reasons for this trend include more excitement about precipitation during the convective season due to the intensity and frequency of storm events. This suggests that the nature of less frequent storms during the frontal season does not continuously engage participants to enter daily data. Although end-of-month emails are sent out asking people to enter uncompleted data, this pattern suggests that an automated system that enters a zero reading until otherwise changed to a 'Trace' or '>0' reading may be more accurate.

The distribution of 'Trace' and '>0' entries is nearly constant when compared between the frontal and convective seasons (see Figure 4.10). On average, the percent 'Trace' entries compared to daily convective entries that suggest rain fell (i.e. data are classified as either 'Trace' or '>0') is 16.3% and >0 entries are 83.7%. During the frontal season, Trace entries account for 16.7% and >0 entries account for 83.3%. These results support the previous discussions that additional spatial analysis on trace events may help determine the significance of these events in the total water budget.

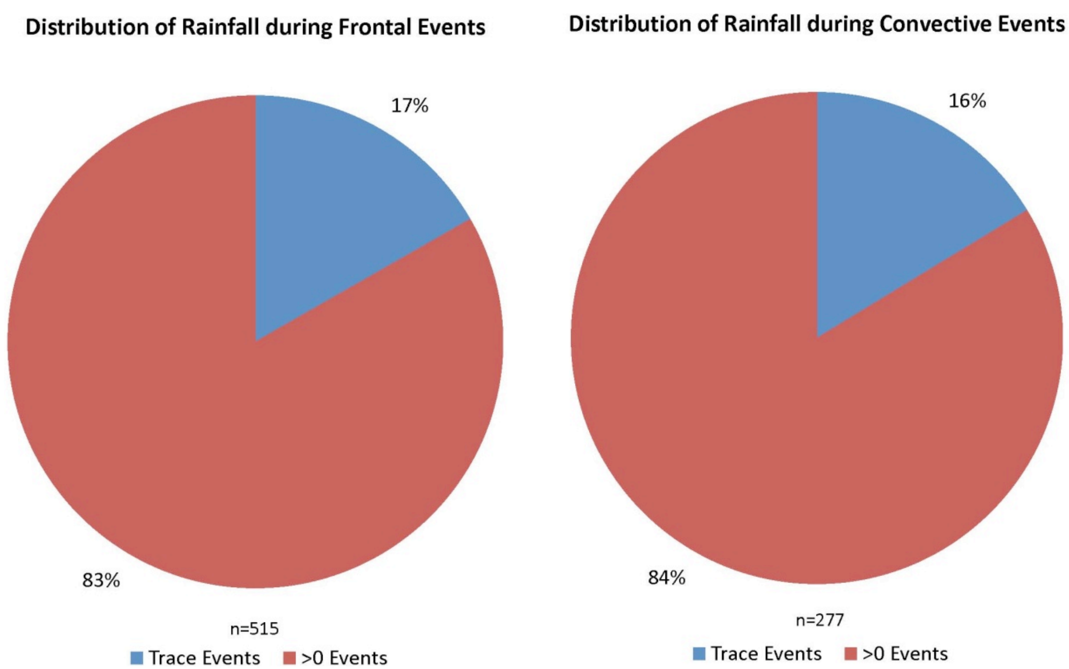

Figure 4.10 - Pie charts for frontal (left) and convective (right) events indicating that on average, Trace events are reported as frequently during frontal storms as during convective storms.

5. DATA USEFULNESS DISCUSSION

5.1 Uses of Data

Data entered into RainLog.org are targeted for use by two primary stakeholder groups that will use citizen-collected precipitation data, both directly and indirectly. In general, the volunteers collecting and reporting precipitation data are one stakeholder group that supports the direct use of data by the second stakeholder group, which includes agencies, municipalities and researchers/scientists. The main difference among the latter stakeholder groups is the type and timeliness of data they need; community-based networks such as RainLog.org must determine whether they can meet the needs of these different stakeholder groups.

In addition to supporting the RainLog.org goal of better characterizing large-scale rainfall patterns, volunteers may utilize their data and data from nearby participants to make household decisions, such as when to skip irrigating landscape plants or when to add a supplemental watering if rainfall at their house was less than recorded at other places in the area. To meet the need of interested citizens, RainLog.org has created RainMapper, which is an automated subscription service that emails people rainfall information from nearby RainLog.org gauges. RainMappers do not have to be RainLog.org volunteers, so it provides a community-wide service that can help improve water management and may increase water conservation.

Although data accuracy is always a best practice, the value of RainLog.org maps and the RainMapper service for the general public is primarily educational at this point.

RainLog.org provides a user-friendly map interface that allows people to view precipitation patterns, both spatially and historically. While data can be copied from an individual user's chart, the website does not have the functionality to download datasets, so it functions mostly to inform visitors about where rainfall is being reported and basic precipitation trends. A future direction for RainMapper may be to couple reported rainfall with evapotranspiration-based estimated watering requirements for different plant types; this comparison will help homeowners determine whether and how much they need to water.

In contrast, water managers modeling drought trends and estimating water availability and storage for a watershed, need to know with as much certainty as possible, how precipitation varies throughout the entire watershed area. This requires that not only is actual RainLog.org data accurately reported, but also that estimates between actual reported data locations are interpolated with certainty. Within the scope of using RainLog.org data to make water management decisions, the type of data needed may vary; a dense network may be needed to minimize estimation errors, while it may be equally as helpful to have long-term data sets to include climate change trends in water resources planning.

RainLog.org is a tool to support various stakeholder groups by providing accurate and precise precipitation data that will be used to answer questions at all scales that could not be addressed without large-scale rainfall data. Table 5.1 provides an overview of the different stakeholders served by RainLog.org and their precipitation data needs; this table was compiled based on observed general stakeholder needs, as well as discussions between RainLog.org developers and stakeholders. These data needs have been categorized into accuracy, precision, length of data record and network size to help discern the type of data needed by each stakeholder group. The remainder of this section explores the research applications of RainLog.org data for stakeholder needs ranging from science, to health to policy, by supporting decision-making and outreach from the individual level to the federal level.

Characteristics of Precipitation Data needed by Stakeholders						
Stakeholder	Accurate	Precise	Timely (robust, short-term data)	Historic (robust, long-term data)	Small Dense Network	Large Areal Network
Drought Managers		x		x		x
Educators		x	x	x	x	x
Flood Control Districts	x		x	x	x	x
Health Officials		x	x			x
Homeowners	x		x		x	
Invasive Species Managers		x	x			x
Researchers	x	x	x	x	x	x
Utility Managers	x		x		x	
Water Managers		x		x		x
Weather Forecasters		x		x		x
Weather Reporters	x	x	x		x	x

Table 5.1- This table lists the different stakeholders that are interested in using RainLog.org data and illustrates the needs of each specific stakeholder group. Notice that precision, timely data and a large areal network are the most desired features of a volunteer precipitation network, which validates the need to do interpolations that ensure that precise data can be reported in near real-time, covering as large of an area as possible. Accuracy is defined as ensuring the exactness of measurements from each individual gauge and precision is defined as ensuring the repeatability or similarity of measurements from multiple, nearby gauges (Wikipedia, 2009).

5.1.1 In Public Health

In Tucson, scientists and health department officials are using RainLog.org data to help determine which neighborhoods may be at highest risk for West Nile Virus due to mosquito breeding from heavy rains and standing water. Precipitation patterns throughout Tucson can identify neighborhoods that received significant amounts of precipitation and may have standing water, either in natural depressions or manmade collection devices like pools, flower pots, old tires and rainwater catchment cisterns. Mosquitoes larvae laid in standing water can hatch within a couple days, growing to adult mosquitoes within a week (Miller, 2004), so localized treatment and neighborhood education must begin immediately after large storm events. RainLog.org data helps health department officials target and prioritize locations susceptible to standing water, which can help reduce the incidence of mosquito outbreaks and ultimately the spread of West Nile Virus.

5.1.2 In Watershed Management and Water Quality

Research currently being conducted by University of Arizona scientists utilizes RainLog.org data to correlate precipitation patterns with urban storm runoff throughout neighborhoods in Tucson. This research is being conducted to determine the potential of urban storm runoff as groundwater recharge, the quality of this recharged water, the impacts of different wash substrates and how residential development age and density alter the quantity and quality of storm water runoff. Collocating rainfall gauges with water quality sampling stations allows researchers to monitor how runoff and water quality constituents move throughout a watershed during precipitation events. Ultimately, this type of research will be used to develop best management practices for promoting nonpoint source pollution reduction and aquifer water quality by evaluating the tradeoffs between enhanced groundwater recharge and the increased possibility of pollutants in surface and groundwater. As water resources become more limited and more expensive to attain, ensuring the quality of reusable and reclaimed resources like urban runoff recharge will become necessary. These and similar research findings play a significant role in protecting water supplies in already water-limited regions like Arizona and the southwestern U.S.

5.1.3 In Decision-Making and Land Management

RainLog.org's statewide reach provides decision makers at regional and state levels with precipitation data that will improve their ability to predict short and long-term needs and changes within their geographic boundaries. Both drought monitoring and water management efforts can be aided by better gauging rainfall over large areas to determine precipitation inputs in water budgets, ranging from plant water demands to reservoir storage. Using citizen-based science to support management decisions can improve the accuracy and efficiency of the decision-making process, freeing up more resources to help meet budget shortfalls elsewhere. In addition, decisions made with citizen science data can improve the transparency of these processes and will hopefully lead to increased trust and acceptance of natural resource-based management decisions.

RainLog.org data is now being used in planning buffelgrass eradication efforts in the Tucson area. For several years local officials and many volunteers have teamed-up to remove buffelgrass from around Pima County, while scientists have continued studying its ecosystem effects, including its increased wildfire risk that poses a serious threat to native plant and animal species. Combined with ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) satellite data, scientists at the University of Arizona are using RainLog.org data to develop a relationship between the amount of precipitation that falls in an area and the growth of buffelgrass in that same area (Olsson, 2009). This research will improve future removal efforts and will help prevent the spread of buffelgrass by targeting locations (with herbicide and volunteer labor) that are positioned to spread buffelgrass if they receive rainfall in a storm event. Herbicide must be applied to buffelgrass soon after it receives precipitation to work effectively, so utilizing RainLog.org data increases project effectiveness. If this effort is successful, this method of using satellite imagery and real-time precipitation data to combat quickly spreading invasive species may be used by municipalities and agencies throughout the country to monitor and manage various species that depend on localized rainfall.

5.1.4 In Risk Management and Forecasting

Many smaller networks desire online reporting with real-time or near real-time results and RainLog.org is able to provide this service, which is especially favorable with county flood control districts that need precipitation data in a timely matter. Both the Maricopa County and Pima County Flood Control Districts are working with RainLog.org to share data and increase their volunteer observer base, which Maricopa County says will "…help the District improve the safety and welfare of County residents through sophisticated flood control and floodplain management programs" (2007).

Weather forecasting also requires time-sensitive data that are reported and analyzed within a single day. Technically, weather forecasting cannot be done with RainLog.org data because this data is only available to report rainfall. However, forecasters can use RainLog.org data as an on-the-ground check to compare against radar and satellite predictions if enough RainLog.org data is available over a large geographic area. Forecasters that appear on television use RainLog.org graphics to illustrate precipitation patterns to viewers and may be able to discuss the progression of storm events through a region.

Television station 13 in Tucson (KOLD) collaborates with RainLog.org to display daily precipitation maps during the weather segments. Often, weather reports include only one or two values that represent rainfall from an entire region, but displaying a RainLog.org map provides viewers with a more accurate estimate of the rainfall amount received near their homes. Additionally, rainfall maps produced from radar and satellite can be difficult for viewers to interpret because these maps often illustrate storm intensity or cloud cover instead of actual rainfall values. Encouraging television stations and newspapers to utilize RainLog.org data provides weather reporters with spatially-distributed data that viewers and readers throughout the region can relate to, while publicizing this citizen-science opportunity to recruit new participants, both for RainLog.org and RainMapper.

5.1.5 In Education and Outreach

Because of its user-friendly, map interface and online accessibility, RainLog.org is well positioned to provide meteorologists with near real-time, spatially dispersed weather information. As mentioned earlier, 44% of individuals still get their information about science and technology from television, so it is critical that news channel meteorologists provide viewers with timely and engaging information about current weather, as well as historic climate trends and local area impacts. When interested in a specific scientific issue, 44% of individuals use the Internet to research information and weather news is the most popular news category sought online, so RainLog.org is the perfect companion for meteorologists wanting to convey locally relevant information to viewers, especially during monsoon season when people have a heightened sense of awareness about storm impacts and dangers (NSB, 2004). Viewers are able to relate to RainLog.org maps because they can see where they live compared to reported rainfall, which may interest them in visiting the dynamic website. Website visitors may or may not become interested in participating in RainLog.org, but once they visit the website they have increased their awareness that other citizens are collecting rainfall data and that it is spatially variable. During monsoon season people often talk about "how much rain they got at their house;" RainLog.org validates the fact that people in different parts of a city do receive different amounts of rainfall and in fact, this is the norm.

The next question asked by individuals is why this variability occurs and at what scale. While scientists have intensely studied orographic effects on precipitation patterns and it is widely known that mountain ranges like the Santa Catalina Mountains northeast of Tucson intensify the convective action of thunderstorms, many microclimate features related to elevation, wind direction, urban density and urban heating are not yet understood. Although answering these questions may require more sophisticated equipment and computer modeling, the RainLog.org system is certainly a positive step toward addressing these unknown variables and helping illustrate regional variability to scientists and citizens alike.

5.2 The Future of RainLog.org

RainLog.org continues to remain on the forefront of citizen science and precipitation monitoring throughout Arizona. Success of this network is largely due to its excellent user interface and graphic capabilities that have linked valuable scientific data with user-friendly and interactive Google Earth™ technology. RainLog.org has the opportunity to expand into additional states and internationally. In fact, the international market may be better served with RainLog.org technology, especially in countries where no rain gauge networks currently exist. RainLog.org has been developed using Google Maps™ technology, which has a global reach, so the only physical limitations to expanding RainLog.org would be developing training materials in other languages and managing an increased volume of participant emails.

Once RainLog.org implements the recommendations discussed in Chapter 4 it will have all the tools necessary to function not only as a data repository, but more importantly, as a data screening system. RainLog.org's continuous growth is evidence that the volunteer precipitation monitoring market is not saturated and that targeted campaigns in certain locations will play an important role in maintaining this growth. Although similar volunteer monitoring networks exist nationally, RainLog.org has features that no other volunteering precipitation monitoring system has, providing an opportunity to collaborate and provide these other systems will some of RainLog.org's valuable, user-friendly tools.

As RainLog.org continues to develop more technical services, it will have the opportunity to act as a consultant for other parties (ranging from teachers, to flood control districts to countries) interested in starting volunteer precipitation monitoring programs. RainLog.org is in a position to offer advice and technical services, which should be utilized by developing programs; RainLog.org can play an important role in developing seamless volunteer monitoring programs with a user-friendly interface and QA/QC checks instead of disjointed volunteer precipitation networks that lack technical services. Additionally, RainLog.org should prepare to help scientists compile precipitation data needed for research purposes. For example, these queries may include certain dates, certain regions or certain quantities, as well as interpolated values. To address these, RainLog.org should continue developing online interpolation methods that allow users to perform their own analyses to answer additional research questions. An online data analysis tool that includes spatial interpolation will certainly set RainLog.org apart from any other volunteer precipitation monitoring network and will make it the preeminent system for analyzing rainfall data.

6. CONCLUSIONS

As researchers look for inexpensive and non-burdensome ways to collect spatially distributed data ranging from plants budding to invasive species entering a region to daily precipitation, volunteer networks are becoming increasingly popular. This popularity presents a unique opportunity to engage a large population of citizen scientists who are motivated and willing to contribute their time and energy to furthering science and decision-making and may become more informed and educated. But this also presents unique challenge of training and managing these dispersed volunteers to collect accurate and useful data that can further research goals and our societal understanding of the earth's physical processes and our changing environment. RainLog.org has taken great strides to ensure that the website is user-friendly and that the training materials are clear and easily accessible; however, to this point, there has been no analysis of whether the data being collected are accurate. To extend RainLog.org into other regions, both nationally and internationally, it is crucial that we know what the biggest sources of data errors are and whether these errors can be resolved.

Unlike scientists and professionals that are trained to collect data following specific QA/QC procedures, relying on volunteers with different levels of training and experience may result in unreliable data. Through this project, various QA/AC issues related to volunteer-collected and reported data were analyzed to address these questions:

1. What are the most likely causes of error?
2. How can we identify systematic and random errors?
3. How can the identified errors be reduced or eliminated?

4. How “good” are the RainLog.org data?
5. To what degree can we trust these data?

A series of analyses was performed, indicating that the most likely sources of error (in order of greater to least potential source) are: user reliability; wrong day reporting; missing data; order of magnitude errors; false precision; and 5/10 rounding bias. In doing these analyses we determined that detecting random errors was difficult, and not significant, because they happened so infrequently and therefore did not skew overall data accuracy. However, systematic errors were detected for both missing data and wrong day reporting for certain users that frequently exhibit these reporting errors. Since these users can be identified, we can flag and exclude their results in the database; correcting these issues involves contacting the users to communicate reporting problems, as well as developing automated functions to ensure and improve data accuracy when users enter data.

These automated front-end checks include: a fill-in function for zero values; an automatic interpolation to warn users if they are entering data different from their nearest neighbors or outside of normal ranges; a warning if users are entering data with decimal values in the thousandths; and a warning if users are entering data greater than 1.0 in. or less than 0.1 in. because most 24-hr rainfall amounts are within this range. These checks will help correct behaviors that contribute to systematic errors and will eliminate this potentially incorrect data from being entered into the database.

The issue of reporting bias toward values than end in 5 or 0 can be addressed with additional training materials that help users learn to read gauge values properly.

However, to some extent this problem may never be corrected because rounding is human nature; if most users are rounding equally up and down, as the analysis suggests, then overall these data are considered accurate. Because all gauge types exhibit the 5/10 reporting bias, RainLog.org is correct in assuming that a variety of rain gauge types (excluding cylinder-type gauges) equally report accurate and precise rainfall data. Inherent spatial variability of rainfall, especially in the Southwest, outweighs the need to require gauges that are precise to more than hundredths of an inch.

Finally, any network can be improved by increasing the density of the network. Although this is not possible in many rural areas, more dense networks will increase the accuracy of interpolation results, which is especially appealing in areas like Tucson that face increasing urbanization abutting mountains that significantly affect rainfall patterns. This unique geography converging with growth is seen throughout the West and understanding storm patterns and their relation to elevation will be increasingly important as flooding risk increases due to increased urbanization and climate change predictions of more extreme events. Additionally, more dense networks provide data users with more comprehensive datasets.

Overall, volunteer data from the RainLog.org system are reliable data that can be used by a multitude of stakeholders. However, the equally if not more important outcome, is that this project developed an overall process or model for addressing QA/QC issues within a volunteer monitoring network of any kind. This model can be organized into four steps, starting with identifying and describing types of data problems and errors that may exist in a specific citizen science project and understanding what these errors would look like in a database. Next, manually find errors by comparing reported values with expected values, which are often drawn from historical data averages, nearby neighbor reports and interpolation analyses, will help assess the extent of problems with the data. Once these errors are identified, developing approaches to automatically screen all reported data for the various problems and more easily compile error frequencies and patterns. Finally, this information will allow project managers to determine whether data should be flagged (notation made to indicate error), corrected or removed and whether users making errors should be contacted for remedial instruction.

The results of this project support the emerging field of Citizen Science, offering evidence that by following this four-step process and incorporating additional automated data checks, volunteer data are good and can be trusted. RainLog.org and similar projects should continue to partner with large-scale research projects and funding to develop and expand new citizen science efforts should be made available because of the mutual benefits of cost-efficient data collection and citizen involvement in science questions that lead to greater engagement and understanding of current research and environmental conditions.

7. REFERENCES

Adams, D.K., and A.C. Comrie (1997). The North American Monsoon. *Bulletin of the American Meteorlogical Society,* 78: 2197–2213.

The Arizona Meterological Network (2008). <http://cals.arizona.edu/azmet/azdata.htm> Accessed July 15, 2008.

Allerup, P., Madsen, H. (1979). Accuracy of point precipitation measurements. Climatological Papers, Danish Meteorological Institute, No. 5.

Ally, M. (2004). *Theory and Practice of Online Learning*. Athabasca, Canada: Athabasca University.

Blöschl, G., and M. Sivapalan (1995). Scale issues in hydrological modelling—A review. *Hydrologic Processes*, 9: 251–290.

Bradford, B. and Israel, G. (2004). Evaluating Volunteer Motivation for Sea Turtle Conservation in Florida. University of Florida – IFAS Extension, AEC 372.

Branom, M. (2008, March 10). Weather Service defines 'Arizona Monsoon Season. *East Valley Tribune*.

Brossard, D., Lewenstein, B., and Bonney, R. (2005). Scientific knowledge and attitude change: The impact of a citizen science project. *International Journal of Science Education*, 27(9): 1099-1121.

Chaubey, I., C.T. Haan, S. Grunwald and Salisbury, J.M. (1999). Uncertainty in the model parameters due to spatial variability of rainfall. *Journal of Hydrology*. 220(1-2): 48-61.

Chvila, B., Sevruk, B. and Ondras, M. (2005). Intercomparison Measurements of Recording Precipitation Gauges in Slovakia. World Meteorological Organization Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (TECO-2005, Report 82). Bucharest, Romania. <www.wmo.int/pages/prog/www/IMOP/publications/IOM-82-TECO_2005/Posters/P3(29)_Slovakia_Chvila.pdf>

Constantinescu, G. S., Krajewski, W. F., Ozdemir, C. E. and Tokyay, T. (2007). Simulation of airflow around rain gauges: Comparison of LES with RANS models. *Advances in Water Resources,* 30:43–58.

Corporation for National & Community Service (2007). Issue Brief: Volunteer Retention. <www.nationalservice.gov>

Daly, C., Gibson, W. Taylor, G., Doggett, M. and Smith, J. (2007). Observer Bias in Daily Precipitation Measurements at United States Cooperative Network Stations. *Bulletin of the American Meteorological Society,* 88(6): 899–912.

Doesken, N. J. (2005). Ten-Year Comparison of Daily Precipitation from the 4 Inch Diameter Clear Plastic Rain Gauge versus the 8 Inch Diameter Metal Standard Rain Gauge. 13th American Meteorological Society Symposium on Instrumentation and Meteorological Observations, Savannah, Georgia.

Doesken, N. J., and Judson, A. (1996). The Snow Booklet: A Guide to the Science, Climatology, and Measurement of Snow in the United States. Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado.

Doesken, N. J. and Weaver, J. F. (2000). Micro-scale rainfall variations measured by a local volunteer network. *American Meteorological Society 12th Conference on Applied Climatology*.

Ellis, R. and Waterton, C. (2004). Environmental citizenship in the making: the participation of volunteer naturalists in UK biological recording and biodiversity policy. *Science and Public Policy,* 31(2): 95-205.

The Flood Control District of Maricopa County. (2007, July 19). Become a Weather Watcher for the Flood Control District of Maricopa County through RainLog.org. <www.fcd.maricopa.gov/PIO/News/FCDNews003.aspx> Accessed November 13, 2008.

Germann, U., and Joss, J. (2001). Variograms of radar reflectivity to describe the spatial continuity of Alpine precipitation. *Journal of Applied Meteorology*, 40: 1042-1059.

Goodrich, D. C., Keefer, T. O., Unkrich, C. L., Nichols, M. H., Osborn, H. B., Stone, J. J. and Smith, J. R. (2008). Long-term precipitation database, Walnut Gulch Experimental Watershed, Arizona, United States. *Water Resources Research,* 44: W05S05.

Goodrich, D. C., Unkrich, C. L., Keefer, T. O., Nichols, M. H., Stone, J. J., Levick, L. R., and Scott, R. L. (2008). Event to multidecadal persistence in rainfall and runoff in southeast Arizona. *Water Resources Research*, 44: W05S14.

Lawrence, A. (2009). The first cuckoo in winter: Phenology, recording, credibility and meaning in Britain. *Global Environmental Change*, 19: 173-179.

Lawrence, A. (2006). 'No Personal Motive?' Volunteers, Biodiversity, and the False Dichotomies of Participation. *Ethics, Place and Environment*, 9(3): 279-298.
Lohse, K., Snyder, M., Gallo, E., Dejwakh, N., Brooks, P., McIntosh, J. and Meixner, T., (2008). Management practices for urban storm runoff: Investigating roles of channel substrates in modifying water quality. Soil and Water Conservation Society Annual Meeting, Tucson, Arizona. <http://www.allacademic.com/meta/p235537_index.html>

Malek, E., McCurdy, G. and Giles, B. (1999). Dew contribution to the annual water balances in semi-arid desert valleys. *Journal of Arid Environments*, 42: 71–80.

Mayer, P. and DeOreo, W. (1999). *Residential End Uses of Water*. AWWA Research Foundation; American Water Works Association.

Miller, M. (2004). Are Mosquitoes Bugging You? Pima County Consumer Health and Food Safety Division Brochure. <http://www.tucsonaz.gov/water/docs/problem.pdf>

Milton, K. (2002). Loving Nature: Towards An Ecology of Emotion. Routledge, London.

National Association of Counties (2005). NACo - Find a county. <http://www.naco.org/Template.cfm?Section=Find_a_County&Template=/cffiles/counties/state.cfm&state.cfm&statecode=AZ> Accessed February 25, 2009.

National Science Board. (2004). Science and technology: Public attitudes and public understanding. *Science & Engineering Indicators—2004*. Washington, D.C.; U.S. Government Printing Office.

Nespor, V. and Sevruk, B. (1999). Estimation of Wind-Induced Error of Rainfall Gauge Measurements Using a Numerical Simulation. *Journal of Atmospheric and Oceanic Technology*, 16: 450-463.

Nijssen, B., and Lettenmaier, D. (2004). Effect of precipitation sampling error on simulated hydrological fluxes and states: Anticipating the Global Precipitation Measurement satellites. *Journal of Geophysical Research*, 109: D02103.

Nystuen, J, Proni, J., Black, P. and Wilkerson, J. (1996). A Comparison of Automatic Rain Gauges. *Journal of Atmospheric and Ocean Technology*, 13: 62-73.

O'Connell, P.E. and Todini, E. (1996). Modeling of rainfall, flow and mass transport in hydrological systems: an overview. *Journal of Hydrology*, 175: 3-16.

Olsson, A., Marsh, S., Daniels, Swann, D. and Backer, D. (2009). Using Remote Sensing to Map and Manage Buffelgrass Infestations in the Tucson and Santa Catalina Mountains. <http://www.buffelgrass.org/ongoing.php> Accessed March 13, 2009.

Pima Association of Governments. (2008). Population Primer. <http://www.pagnet.org/RegionalData/Population/PopulationPrimer/tabid/407/Default.aspx> Accessed October 10, 2008.

Poole, B. (2008, June 13). Monsoon season starts Sunday. *Tucson Citizen.*

Reges, H., Cifelli, R.C. and Doesken, N.J. (2006). The Community Collaborative Rain, Hail and Snow Network (CoCoRaHS)— "Hands-On" Science for Communities Right in their Own Backyards. *15th American Meteorlogical Society Education Symposium*, Atlanta, GA.

Robinson, A. C., and Rodda, J. C. (1969). Rain, wind and the aerodynamic characteristics of rain-gauges, *The Journal of Meteorology Magazine*, 98: 113-120.

Roebber, P., Bruening, S., Schultz, D. and Cortinas, J. (2002). Improving Snowfall Forecasting by Diagnosing Snow Density. *Weather and Forecasting,* 18: 264-287.

Sampson, G. and Pytlak, E. (2009). Educating the Public on the North American Monsoon System. *18th American Meteorological Society Symposium on Education*, Phoenix, Arizona.

Serreze, M., Clark, M., Armstrong R., McGinnis D. and Pulwarty R. (1999). Characteristics of the western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research,* 35(7): 2145–2160.

Sevruk, B. (2008). Chapter 35: Rainfall Measurement: Gauges. In Anderson, M., (Ed.), *Encyclopedia of Hydrological Sciences*. New York, NY; .John Wiley & Sons, Inc.

Sevruk, B., and Nespor, V. (1998). Empirical and theoretical assessment of the wind induced error of rain measurement. *Water Science* and *Technology*, 37: 171-178.

Sevruk, B. (1996). Adjustment of tipping-bucket precipitation gauge measurements. *Atmospheric Research*, 42: 137-246.

Sevruk, B. (1989). Wind-induced measurement error for high-intensity rains. In Sevruk, B. (Ed.), *Precipitation Measurement*. World Meteorological Organization. Geneva, Switzerland, 199-204. Proceedings of International Workshop on Precipitation Measurement, St. Moritz, Switzerland,

Sheppard, P., Comrie, A., Packin, G., Angersbach, K. and Hughes, M. (2002). The climate of the Southwest. *Climate Research*, 21: 219–238.

Sieck, L. C., Burges, S. J., and Steiner, M. (2007). Challenges in obtaining reliable measurements of point rainfall. *Water Resources Research*, 43: W01420.

Sinclair, S. and Pegram, G., (2005). Combining radar and rain gauge rainfall estimates using conditional merging. *Atmospheric Science Letters,* 6: 19–22.

Southwest Hydrology. (2005). Product Review: Digital Rain Gauges Compared. *Southwest Hydrology*. The University of Arizona SAHRA Center, Tucson, AZ; 6(4): 31.

U.S. Census Bureau, (2007). Annual Estimates of the Population for Counties of Arizona: April 1, 2000 to July 1, 2007 <www.census.gov/popest/counties/tables/CO-EST2006-01-04.xls>. Accessed: March 10, 2009.

U.S. National Weather Service. (2008). National Weather Service Instruction 10-1307. <http://www.nws.noaa.gov/directives/>. Accessed: January 14, 2009.

U.S. National Weather Service. (2007). Cooperative Station Observations Manual 10-1315. <http://www.nws.noaa.gov/directives/>. Accessed: January 14, 2009.

U.S. National Weather Service. (1989). Cooperative Station Observations, NWS Observing Handbook No. 2. Silver Spring, MD; Observing Systems Branch, Office of Systems Operations.

U.S. National Weather Service Forecast Office. (2009). Monthly and Daily Normals (1971-2000) plus Daily Extremes (1895-2008) for Tucson, Arizona. <http://www.wrh.noaa.gov/twc/climate/tus.php>. Accessed: March 3, 2009.

U.S. National Weather Service. (2008). Tucson WFO Monthly/Daily Climate Data. <http://www.weather.gov/climate/index.php?wfo=twc.> Accessed March 3, 2009.

Vitaterna, M., Takahashi, J. and Turek, F. (2001). Overview of Circadian Rhythms. National Institute on Alcohol Abuse and Alcoholism. <http://pubs.niaaa.nih.gov/publications/arh25-2/85-93.htm>. Accessed: February 24, 2009.

Western Regional Climate Center. (2007). Arizona Climate Summaries. <http://www.wrcc.dri.edu/summary/climsmaz.html>. Accessed: July 9, 2008.

Wikipedia: The Free Encyclopedia. (2009) "Accuracy and precision." < http://en.wikipedia.org/wiki/Accuracy_and_precision>. Accessed: September 1, 2009.

World Meteorological Organization. (2006). *WMO Guide to Meteorological Instruments and Methods of Observation, 7th Edition*. Geneva, Switzerland; World Meteorological Organization.

World Meteorological Organization. (1983). *WMO Guide to Meteorological Instruments and Methods of Observation, 5th Edition*. Geneva, Switzerland; World Meteorological Organization.