Developing a Relationship between Radar and Rain Gauge Datasets in Central and South Florida

Courtney Skinner[1], F. Bloetscher[2], C. Pathak[3] and P. Scarlatos[4]

[1]Department of Civil Engineering, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431-0991; PH (561) 951-3713; FAX (561) 297-0493; email: cskinne3@fau.edu
[2]Department of Civil Engineering, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431-0991; PH (239) 250-2423; FAX (561) 297-0493; email: fbloetsc@fau.edu
[3]Operations and Hydro Data Department, South Florida Water Management District, 3301 Gun Club Road, West Palm Beach, FL 33416-4680; PH (561) 682-2567; FAX (561) 681-6265; email: cpathak@sfwmd.gov
[4]Department of Civil Engineering, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431-0991; PH (561) 297-0466; FAX (561) 297-0493; email: scarlatos@civil.fau.edu

Abstract

The South Florida Water Management District (SFWMD) is an agency that relies on a network of nearly 300 rain gauges in order to provide rainfall data for use in operations, modeling, water supply planning, and environmental projects. However, the prevalence of convective and tropical rainfall events in South Florida during the wet season presents a challenge in that the current rain gauge network may not fully capture rain events which demonstrate high spatial variability. NEXRAD (Next Generation Radar) technology offers the advantage of providing a spatial account of rainfall, although the relative quality of radar rainfall measurements remains largely unknown. The intent of this paper is to examine the relationship between gauge-adjusted NEXRAD data and corresponding rain gauge measurements in order to assess the relative performance of radar and rain gauge technologies for different conditions.

Introduction

The South Florida Water Management District (SFWMD) is one of five government agencies responsible for the oversight and protection of water resources in the State of Florida. The SFWMD coverage area extends south from Orlando, along the boundaries of the Kissimmee River Basin to Lake Okeechobee, and from the Atlantic Ocean to the Gulf of Mexico in South Florida. This region includes an area of 17,930 square miles (46,439 square kilometers) and boasts a population of over seven million people. Key features of the South Florida hydrosystem include Everglades National Park and the Kissimmee River, both sites of major restoration efforts; Lake Okeechobee, the nation's second largest freshwater lake; water conservation areas; coastal and estuarine systems; as well as expansive agricultural areas and urban districts. The SFWMD manages this system through a complex network of water control structures, canals, and pump stations (Huebner et al., 2003; Pathak and Palermo, 2006; Sangoyomi et al., 2005; SFWMD[1], 2006).

The SFWMD maintains an extensive network of rain gauging stations in order to monitor rainfall and obtain precipitation data necessary for use in operations, planning, and regulatory aspects of water management. Several limitations are known to exist with the current dependence on rain gauge technology, including introduction of error through the spatial extrapolation of point measurements to surrounding areas (Bedient and Huber, 2002). The problem of accounting for spatial rainfall distributions is of particular concern in South Florida, where intense, highly variable convective and tropical rain events predominate in the wet season (Bras, 2004).

NEXRAD, or Next Generation Radar technology offers the advantage of providing water management officials with a spatial and temporal account of rainfall variability, although the quality of radar measurements remains largely unknown. The SFWMD presently acquires NEXRAD derived precipitation data from the OneRain Company in order to supplement data from the existing rain gauge network (Huebner et al., 2003). However, before NEXRAD data can be successfully extended to applications involving operations and hydrologic analysis, the relative quality of radar generated rainfall values must first be assessed.

The intent of the research was to determine the relationship between NEXRAD data received at the SFWMD and rain gauge measurements in order to assess the relative performance of NEXRAD technology for a variety of different conditions. The study concentrated on the Upper and Lower Kissimmee River Rain Areas, located in the northern portion of the SFWMD service area. The Kissimmee River Basin was selected as ongoing restoration projects require rainfall data for use in hydrologic models. The region also experienced the greatest influence, in terms of rainfall, of four tropical events which impacted the State of Florida in 2004 and 2005, which proved advantageous in investigating the effect of extreme tropical rainfall on the relationship between radar and gauge measurements (Pathak, 2006).

Methodology

The analysis of NEXRAD and rain gauge performance at the SFWMD was realized through the statistical comparison of measurements from each; a comparison defined as the consideration of NEXRAD rainfall data produced for each 2-km by 2-km pixel containing a SFWMD rain gauge and precipitation data generated by the corresponding gauge. Daily time-interval rainfall data were selected for study as daily rainfall measurements are preferred at the SFWMD for many analyses, including hydrologic modeling, and because the daily time-interval makes use of the full compliment of precipitation data available from the rain gauge network. The time period of 2002-2005 was chosen for data comparison as the SFWMD began acquiring gauge-adjusted NEXRAD data from the OneRain Company in January 2002 (Huebner et. al., 2003). The period of record (POR) was further partitioned according to season in order to perform specific analyses, with the dry season identified as the months of November-May and the wet season defined as the months of June-October (Skinner, 2006).

Daily time-series precipitation data were obtained for active rain gauges located within the study area from the SFWMD DBHYDRO database. The rain gauges and corresponding reporting designations considered for the present research are shown in **Tables 1** and **2**. Gauge-adjusted NEXRAD radar rainfall data were retrieved for each pixel containing an active SFWMD rain gauge through the SFWMD GIS-based ArcHydro® interface.

Table 1: SFWMD Rain Gauge Reporter Types, Adapted From (Sangoyomi et al., 2005)

Reporting Designation	Data Transfer	Rain Gauge Type	Data Collection
LoggerNet	Telemetry	Tipping Bucket	Near Real Time
RACU	Telemetry	Tipping Bucket	Near Real Time
MOSCAD	Telemetry	Tipping Bucket	Near Real Time
ARDAMS	Phone Lines	Tipping Bucket	Daily
CR10	Manual	Tipping Bucket	Monthly
Graphic Chart	Manual	Float-Type	Monthly
Manual Log	Manual	Standard	Daily

Table 2: SFWMD Rain Gauge Attributes: Upper and Lower Kissimmee River Basins, Adapted From (SFWMD[2], 2006)

Reporter Type	Number of Gauges
LoggerNet	17
RACU	7
MOSCAD	0
ARDAMS	9
CR10	10
Graphic Chart	6
Manual Log	4
TOTAL	**53**

Data Preprocessing

The data-preprocessing component of the analysis was performed in order to remove select data points from NEXRAD and rain gauge datasets prior to statistical analysis. Preprocessing occurs in three steps, including the removal of tagged data, removal of zero-zero data points, and the removal of precipitation data that are 0.01 inches of rainfall or less. The first stage entails eliminating NEXRAD and corresponding rain gauge data when a SFWMD QA/QC (Quality Assurance/Quality Control) code accompanies the rain gauge measurement, signaling that the data quality may have been compromised. The second stage involves removing NEXRAD and rain gauge values when both measurements indicate zero rainfall volumes to avoid deceptive indications in statistical measures of correlation. Finally, NEXRAD and rain gauge data pairs are eliminated when either measurement is less than 0.01 inches as rain gauge instrumentation is not capable of discerning precipitation volumes less than this amount (Skinner, 2006).

Analysis Tools

Several statistical approaches were investigated for the determination of a relationship appropriate for describing NEXRAD measured rainfall quantities as a function of rain gauge measurements. The selection of rain gauge measurements as the independent variable and NEXRAD measurements as the dependant variable comes as an outcome of the NEXRAD gauge-adjustment process as well as the understanding that rain gauge technology is generally more accurate. The former suggests that NEXRAD and rain gauge data samples may not be independent of one another in a statistical sense. Several procedures were investigated to test this presumption including parametric tests such as the Statistical *t*-test and analysis of variance (ANOVA), and nonparametric approaches, namely the Mann-Whitney *U* test and the Kolmogorov-Smirnoff (K-S) test for independence. The K-S test for independence was ultimately selected for implementation, as this procedure does not maintain requirements for the shape of underlying population distributions (Sprinthall, 1997; Sheskin, 2000).

Rainfall frequency analyses were performed for rain gauge and corresponding NEXRAD precipitation data in order to study and visualize the two rainfall distributions. Linear regression was employed, accompanied by an analysis of correlation, in order to examine the suitability of a linear function in describing the relationship between the datasets. Significance of the correlation was also investigated. Bias frequency analyses were performed to identify systematic offsets among the datasets, where bias is considered the difference between rain gauge and NEXRAD rainfall values. The root mean square error (RMSE) was employed to determine goodness-of-fit for linear and nonlinear functions representing the data through computing:

$$RMSE = \left[\frac{\sum_{i=1}^{n} (y_i - y_i')^2}{n} \right]^{1/2}$$

where y_i are observed NEXRAD measurements and y_i' are NEXRAD values predicted by the devised relationship. RMSE communicates improved ability of the relationship to fit the data as the computed value is minimized. Finally, residuals, or the individual differences between actual and predicated NEXRAD values were examined. Residuals frequency analyses contributed additional information about the relative ability of each relationship to predict NEXRAD generated rainfall values with respect to rain gauge measurements (Sprinthall, 1997; Sheskin, 2000).

Extreme Event Analysis

An analysis of extreme events was performed in order to ascertain whether or not data associated with extreme tropical rain events in 2004 and 2005 should be removed as an additional preprocessing component. Rainfall data from Hurricanes Charley, Frances, Jeanne, and Wilma were examined to determine if these data points consistently corresponded to outliers from the selected prediction function. This was accomplished through the adoption of 95% confidence intervals, and the assumption of a normal residuals distribution. With these designations in place, the equation specifying the confidence limits becomes:

$$(\bar{x} - 1.96 s_x) < \mu < (\bar{x} + 1.96 s_x)$$

provided that n is large (greater then 100), where $\bar{x}$ represents the sample mean of residuals (in this case the predication function itself), s_x is the sample standard deviation of residuals, and μ is the population mean of residuals. Outliers are deemed those points falling outside of the established confidence intervals (Kachigan, 1986).

Results (Skinner, 2006)

The K-S test for independence was executed for rainfall datasets from each rain gauge within the study area. Results indicate that NEXRAD and corresponding rain gauge rainfall distributions are not derived from the same population when considering the period of record (POR). The assertion that NEXRAD and rain gauge precipitation data are significantly different is an important finding as this validates the direct comparison of the datasets in further statistical analyses. Similarly, the statistical dependence of NEXRAD and rain gauge datasets was examined for wet and dry season months, which revealed the tendency for the datasets to demonstrate less departure in the cumulative density functions (CDFs) in the dry season as opposed to the wet season.

Rainfall Frequency Analysis

Rainfall frequency analyses were conducted for each rain gauge within the study area in order to examine the relative distributions of rain gauge and corresponding NEXRAD precipitation measurements. Findings indicate rainfall distributions of a consistent shape as exemplified in **Figure 1** for the Poinciana rain gauge. The relative frequency distributions reveal that, for very low values of precipitation (less than 0.05 inches), NEXRAD tends to underestimate rainfall with respect to the rain gauge. This occurrence may be attributed to the fact that NEXRAD technology cannot always recognize precipitation at this level. More importantly, the graphic demonstrates the tendency for NEXRAD to overestimate precipitation relative to the rain gauge for precipitation measurements between 0.25 and 1.0 inch, and underestimate rainfall values for precipitation in the higher, 2.0 to 5.0 inch range. These results are consistent with observations made by Shaughnessy and Swartz (2006), that NEXRAD tends to over-represent rainfall for frontal, or stratiform disturbances, and under-represent rainfall for convective events.

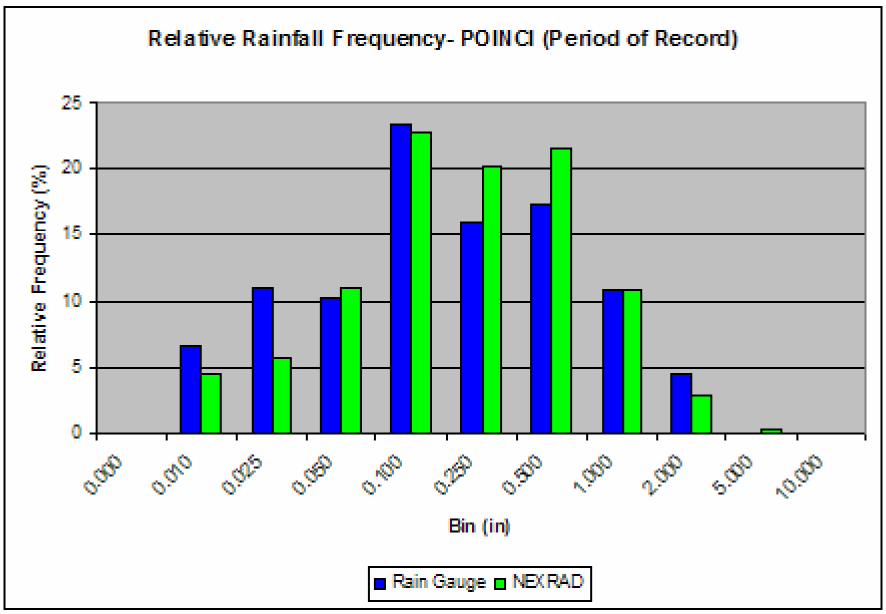

Figure 1: Relative Rainfall Frequency Distribution for POINCI Station (Typical)

Linear Regression/Correlation Analysis

Linear regression and correlation analyses were performed on rain gauge and NEXRAD data in order to determine the applicability of a linear function in describing the relationship between the two datasets. The linear correlation between NEXRAD and rain gauge measurements proved substantial for the majority of stations investigated, however results of the regression analysis exposed the tendency of the best-fit line to demonstrate a positive, non-zero intercept and a slope of less than unity. Further investigation revealed pronounced deterioration in correlation was found to be the consequence of forcing the regression line through the origin, as shown in **Figure 2** for the Alligator 2 rain gauge, which is typical for all gauges evaluated. Regression lines observed may attempt to correct for bias issues noted by Shaughnessy and Swartz (2006).

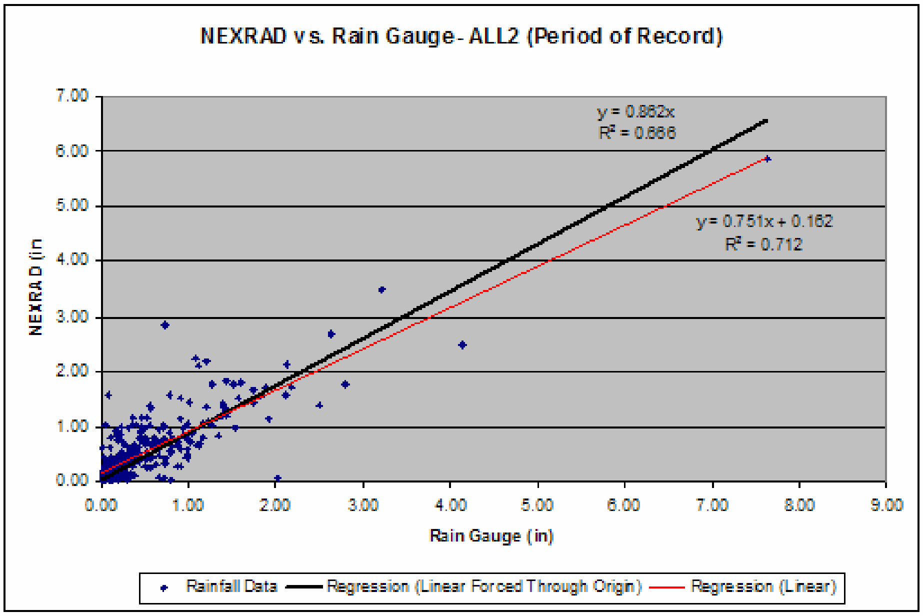

Figure 2: Regression Lines for ALL2 Station (Typical)

Four study gauges at Structure 61, Structure 65, Structure 65A, and Structure 65C were found to exhibit an exceptionally poor degree of correlation (r^2<0.40) between NEXRAD and rain gauge precipitation values. The feature common to these gauges is that they constitute the Standard rain gauges located within the study area, and consequently report rainfall as a 7 a.m. to 7 a.m. EST daily accumulation total, unlike NEXRAD and other gauge reporter types, which produce a midnight to midnight EST daily accumulation total. However, the poor correlation of data indicated at these stations may or may not be entirely due to the apparent discrepancy in reporting intervals. The relative inability of these gauges to provide information about the relationship between the two datasets, an issue which could not be resolved based on the data at hand, motivated the decision to disregard data associated with the four Standard rain gauges located within the study area for subsequent analyses.

Bias Analysis

Bias analyses were conducted to further assess the nature and severity of bias present in the rainfall data. Results consistently demonstrate a skewed bias distribution, as presented for the Avon Park rain gauge in **Figure 3**. The central portion of the distribution indicates that a greater probability exists for NEXRAD values to be slightly greater in magnitude than corresponding rain gauge measurements. The tails of the distribution suggest the opposite, that a greater chance exists for NEXRAD values to be understated greatly with respect to rain gauge measurements. Bias distributions observed are in agreement with the trend for NEXRAD to over-represent rainfall relative to the rain gauge for small precipitation amounts, and under-represent rainfall relative to the rain gauge for large precipitation amounts.

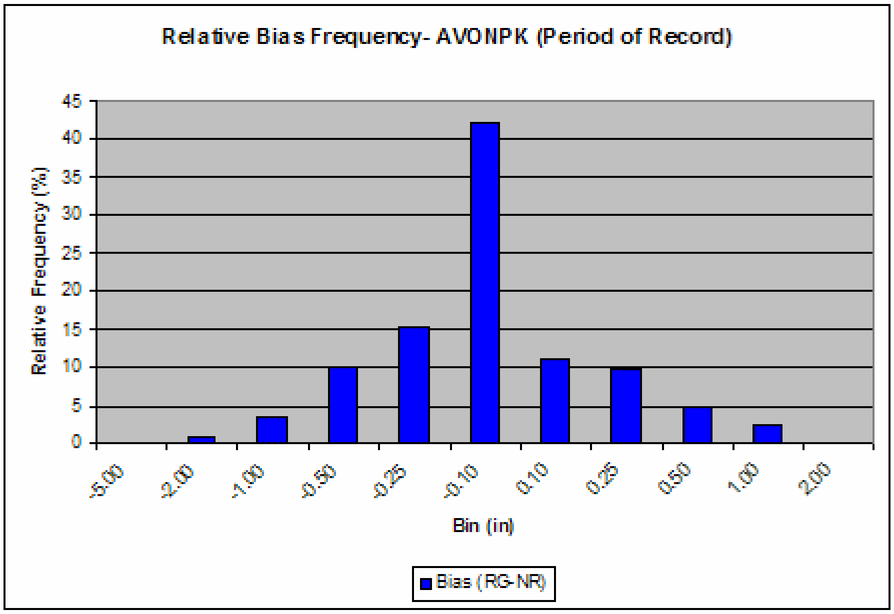

Figure 3: Relative Bias Frequency Distribution for AVONPK Station (Typical)

Relationship Formulation

The rainfall data indicate the relative tendency for NEXRAD to overestimate rainfall with respect to the rain gauge for low-end data, and underestimate rainfall with respect to the rain gauge for high-end data. The disparity found with regard to bias tends to suggest that a power relationship is appropriate for representing the data. Moreover, a power function resolves the issue that the linear relationship does not pass through (0,0) and is supported in the literature by Austin (1987) and Ciach et al. (2000). Consequently, the investigation turned to the formulation of a power function to improve upon the linear relationship.

Development of the prediction function focused on comparing low (less than 1.0 inch) and high-end (greater than 1.0 inch) rainfall data for the Alligator 2 rain gauge separately. Linear regression performed for each subset resulted in two distinct lines. Since large precipitation amounts are of great importance in the operational control and allocation of water by the SFWMD, it was essential to achieve high correlation for extreme rain events. The radar-rainfall relationship was arrived at through an initial, visual comparison of slightly varying curves to the lines generated, and a supplemental analysis of RMSE. Results of this analysis are visualized in **Figures 4** and **5** where it is observed that RMSE is minimized for the curve,

$$y = 0.9x^{0.9}$$

where *y* is the NEXRAD measurement and *x* is the corresponding rain gauge measurement, indicating the improved ability of this function to describe the rainfall data.

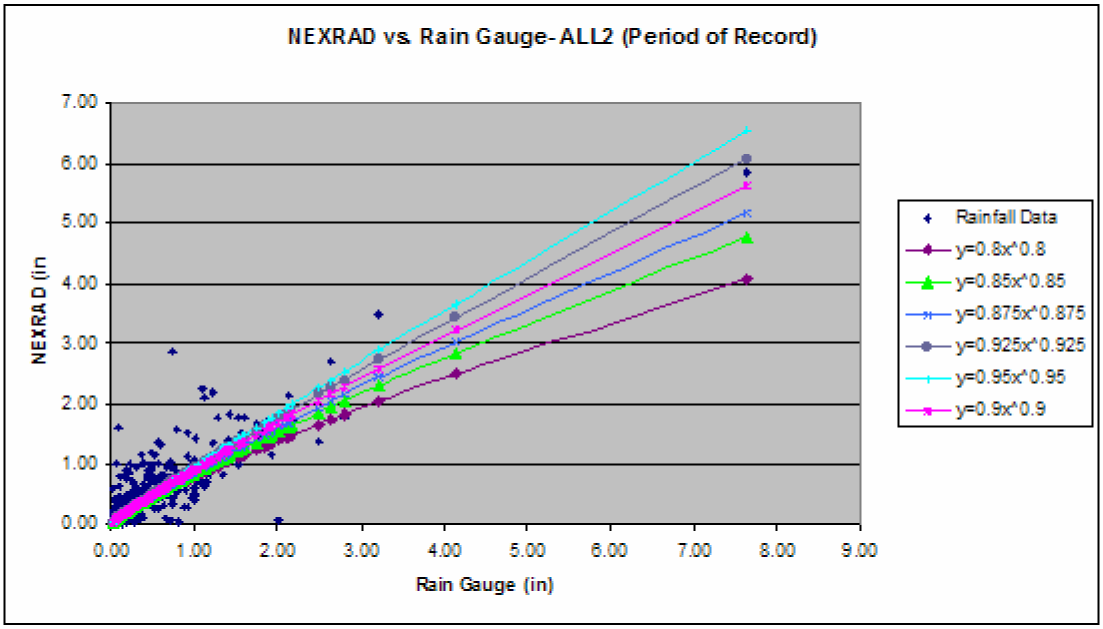

Figure 4: Select Power Functions Analyzed for ALL2 Station

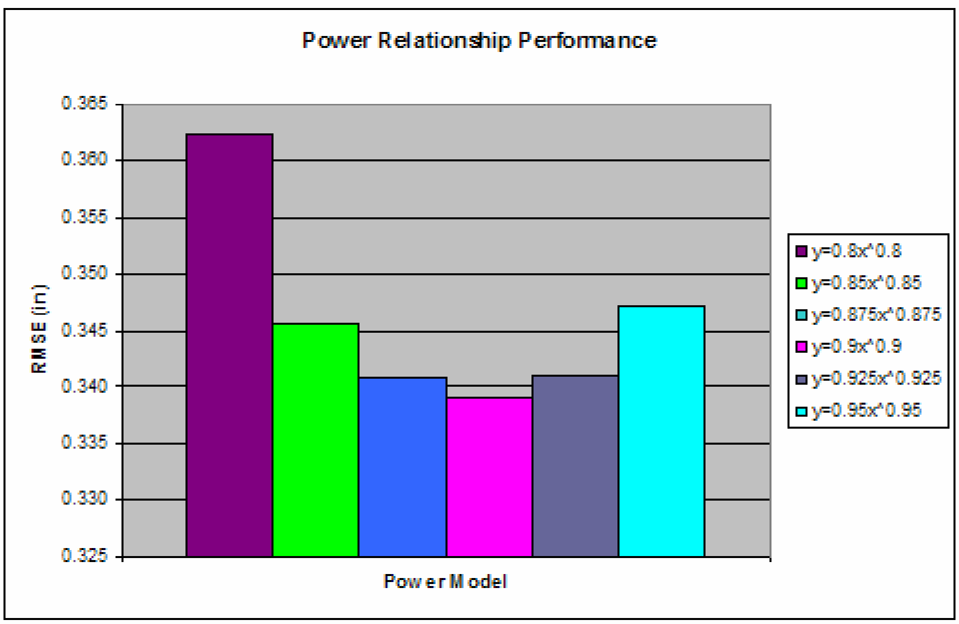

Figure 5: Select Power Functions and Associated RMSE Values for ALL2 Station

The derived prediction function was extended to other SFWMD gauges, as well as all study gauges in order to establish applicability for other gauge stations. RMSE was employed to evaluate goodness-of-fit where it was found that, overall, performance of the power relationship is comparable to that of the best-fit linear relationship established for each gauge. Rain gauges associated with upper quartile (75^{th} percentile) RMSE values were selected to further study the relative inability of the radar-rainfall relationship to describe NEXRAD measurements at certain gauges.

Results offered in **Table 3** indicate that Float-Type rain gauges tend to exhibit a lower degree of fit to the prediction function, with RMSE values greater than 0.35 inch, the performance measure imposed by the upper quartile. Furthermore, it was discovered that the Float-Type gauges that fail the performance threshold (EL MAXIMO, MAXYN, and MAXYS) also exhibit the greatest RMSE values. The trend that the derived radar-rainfall relationship, which seems well-suited for the majority of the gauge sites surveyed, does not fit rainfall data from Float-Type gauges to the same extent as Tipping Bucket gauges suggests that the relative quality of rain gauge measurements at these stations may be lacking. This may be attributed to several factors specific to Float-Type gauges, such as the facts that they are not capable of transmitting near real time (NRT) rainfall data via telemetry and that they rely on the mechanics of a relatively outdated float and pen-driven chart system.

Table 3: Rain Gauges Which Fail RMSE Performance Measure

Rain Gauge Type	Number of Gauges in Violation	Total Number of Gauges	Percentage of Gauges in Violation
Tipping Bucket	8	43	19%
Float-Type	3	6	50%
TOTAL	**11**	**49**	**22%**

Residuals Analysis

Performance of the radar-rainfall relationship was also assessed through conducting a residuals frequency analysis for each gauge within the study area. Typical findings are presented in **Figure 6** for the Peavine Trail rain gauge. Overall, it is observed that residuals associated with the application of the devised power relationship demonstrate a highly symmetric normal distribution, while the residuals distribution resulting from the use of the linear relationship is highly skewed in shape. Results illustrate that the developed prediction function is superior in describing the relationship between rain gauge and NEXRAD rainfall data at the SFWMD as much of the systematic component of error between the two datasets has been removed, leaving primarily random error contributions as designated by the normal distribution.

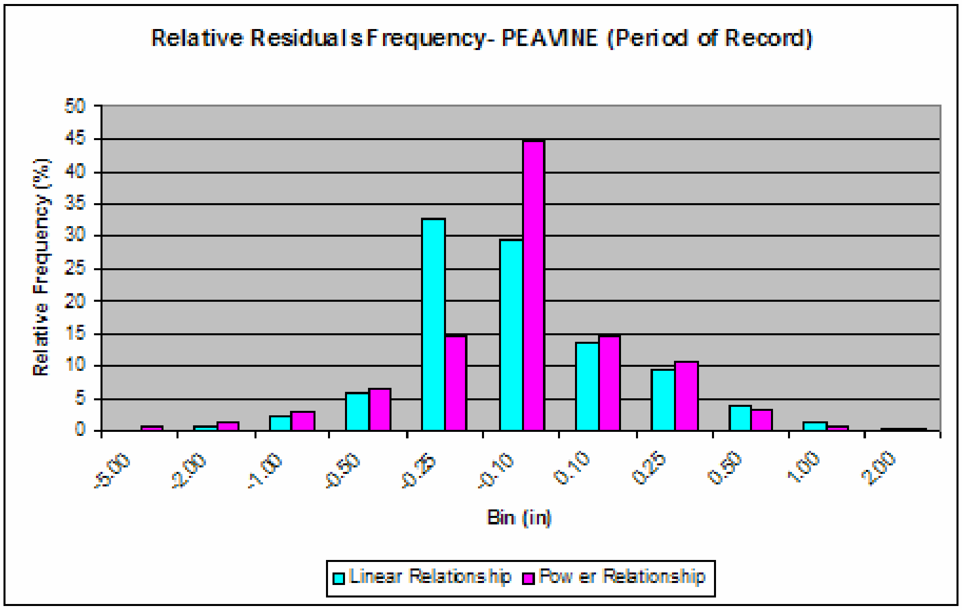

Figure 6: Relative Residuals Frequency Distribution for PEAVINE Station (Typical)

Extreme Event Analysis

Noteworthy results were obtained from the comparison of RMSE values on an annual basis, where it was found that a pronounced decrease in correlation is evident in 2004. Curtis (2006) notes that poor data relative data quality observed in 2004 may be the result of an active hurricane season, among other factors, as three tropical events of magnitude affected Central and South Florida in 2004; Hurricane Charley, which made landfall August 13-14; Hurricane Frances, which made landfall September 5-6; and Hurricane Jeanne, which made landfall on September 21. Hurricane Wilma impacted the study area on October 24-25, 2005 (National Hurricane Center, 2006).

To assess the validity of these statements, RMSE for the developed prediction equation was examined on a monthly basis. Overall, it was found that greater RMSE values tended to be associated with the months of August and September. In order to determine if increased deviation was solely the result of tropical precipitation, or if rain gauges damaged during the storms could have influenced these outcomes, daily precipitation data from all study gauges were plotted and examined for August and September (2004). Confidence intervals (95%) were employed as bounds for the radar-rainfall relationship to identify outlier points.

This analysis revealed that tropical events do not produce a greater number of severe outliers when compared to data not associated with the three hurricanes. Furthermore, it was observed that several of the decided outliers occur on the same day, September 26, 2004, which may indicate post-storm damage to rain gauges. A parallel investigation of extreme rainfall events was performed for each rain gauge to establish the influence of tropical events on the radar-rainfall relationship. Results from this analysis illustrate that data points associated with hurricanes are not consistently outliers and that some of the most obvious outliers produced for the POR are not linked to tropical events.

Conclusions (Skinner, 2006)

The K-S Test for independence was employed to determine whether NEXRAD and corresponding rain gauge rainfall distributions are significantly different statistically. This procedure revealed that, for the POR data, the precipitation distributions are not derived from the same population. The conclusion that the two datasets are deemed appreciably different is important as this validates the direct comparison of NEXRAD and rain gauge data for the study period.

Linear regression and a corresponding analysis of correlation were performed on the NEXRAD and corresponding rain gauge data pairs over the POR in order to determine if a linear correlation exists between the two datasets. Overall, it was found that while a significant linear correlation exists for the study data, the linear relationship does not pass through the origin and cannot explain observed low and high-end bias issues. Subsequent rainfall and bias distribution analyses confirmed the presence of bias and exposed the relative tendency for NEXRAD to overestimate low values of rainfall and underestimate high values of rainfall relative to individual gauge stations. Results regarding the nature of bias prompted the development of the radar-rainfall relationship $y = 0.9x^{0.9}$ to describe NEXRAD measurements as a function of rain gauge observations for the POR.

Performance of the devised relationship in representing the data was assessed through computation of RMSE and through an analysis of residuals. Results reveal that the developed power function is comparable in performance to the linear model but demonstrates the added benefits of passing through the origin and addressing bias. Subsequent analyses identified that precipitation data generated by Tipping Bucket rain gauges adhere to the developed relationship to a greater extent than data from Float-Type gauges. Further assessments exposed the finding that hurricane data do not consistently correspond to outliers from the radar-rainfall relationship, and ultimately provided the ground to retain tropical data points in the preprocessing scheme.

References

Austin, P.M., 1987. Relation between Measured Radar Reflectivity and Surface Rainfall. *Monthly Weather Review*. 115(5), 1053-1071.

Bedient, P.B. and Huber, W.C., 2002. *Hydrology and Floodplain Analysis, 3rd Edition*. Prentice Hall, Upper Saddle River, New Jersey.

Bras, R.L., 2004. *Thoughts on Precipitation Monitoring Network Design*. Personal Communication to C.S. Pathak. Obtained 1/26/2006.

Ciach, G.J., Morrissey, M.L., Krajewski, W.F., 2000. Conditional Bias in Radar Rainfall Estimation. *Journal of Applied Meteorology*. 39(11), 1941-1946.

Curtis, D.C., 2006. Personal Communication to C.S. Pathak. Dated 9/26/2006. RE: Thoughts on Calibration Datasets used for Radar Processing. Obtained 9/27/2006.

Huebner, S.R., Pathak, C. and Hoblit, B.C., 2003. Development and Use of a NEXRAD Database for Water Management in South Florida. *ASCE EWRI World Water and Environmental Resources Congress Proceedings*. ASCE, Reston, Virginia.

Kachigan, S.K., 1986. *Statistical Analysis*. Radius Press, New York, New York.

National Hurricane Center, 2006. Website: http://www.nhc.noaa.gov/pastall.shtml. Accessed 6/6/2006.

Pathak, C.S., 2006. Personal Communication. Rationale for Selection of Kissimmee River Basin as Study Area. 9/27/2006. SFWMD Lead Engineer, West Palm Beach, Florida.

Pathak, C.S. and Palermo, S., 2006. Rainfall-Based Management Plan for Water Conservation Area 3A in the Florida Everglades. *ASCE EWRI World Water and Environmental Resources Congress Proceedings*. ASCE, Reston, Virginia.

Sangoyomi, T. and Dawkins, D., 2005. *QA/QC of Rainfall Data Procedures (Rainfall Gauges and NEXRAD). Procedure Q202*. SFWMD, West Palm Beach, Florida.

SFWMD[1], 2006. Website: http://www.sfwmd.gov. Accessed 1/31/2006.

SFWMD[2], 2006. Internal Communication. Spreadsheet of SFWMD Rain Gauge Attributes. SFWMD, West Palm Beach, Florida. Obtained 2/8/2006.

Shaughnessy, G. and Swartz, E., 2006. Personal Communication. NEXRAD Consistently Overestimates Frontal and Underestimates Convective Rain Events. 4/11/2006. SFWMD Meteorologists, West Palm Beach, Florida.

Sheskin, D.J., 2000. *Handbook of Parametric and Nonparametric Statistical Procedures, 2nd Edition*. CRC Press, Boca Raton, Florida.

Skinner, C.L., 2006. *Developing a Relationship between NEXRAD Generated Rainfall Values and Rain Gauge Measurements in South Florida*. Thesis Completed to Fulfill Requirements of Master of Science Degree, Florida Atlantic University, Boca Raton, Florida.

Sprinthall, R.C., 1997. *Basic Statistical Analysis, 5th Edition*. Allyn and Bacon, Needham Heights, Massachusetts.