

Article

# Comparison of Hydrologic Model Performance Statistics Using Rain Gauge and NEXRAD Precipitation Input at Different Watershed Spatial Scales and Rainfall Return Frequencies for the Upper St. Johns River, Florida USA

Amanda Bredesen, P.E. <sup>1</sup>, Christopher J. Brown, P.E., Ph.D. <sup>2,\*</sup>

<sup>1</sup> U.S. Army Corps of Engineers, Jacksonville, District; aetan625@aol.com

<sup>2</sup> School of Engineering, University of North Florida, USA; christopher.j.brown@unf.edu

\* Correspondence: christopher.j.brown@unf.edu; Tel.: +01-904-620-2811

Received: date; Accepted: date; Published: date

**Abstract:** Water resources numerical models are dependent upon various input hydrologic field data. As models become increasingly complex and model simulation times expand, it is critical to understand the inherent value in using different input datasets available. One important category of model input is precipitation data. For hydrologic models, the precipitation data inputs are perhaps the most critical. Common precipitation model input includes either rain gauge or remotely-sensed data such next-generation radar-based (NEXRAD) data. NEXRAD data provides a higher level of spatial resolution than point rain gauge coverage, but is subject to more extensive data pre and post processing along with additional computational requirements. This study first documents the development and initial calibration of a HEC-HMS model of a sub-tropical watershed in the Upper St. Johns River Basin in Florida, USA. Then, the study compares calibration performance of the same HEC-HMS model using either rain gauge or NEXRAD precipitation inputs. The results are further discretized by comparing key calibration statistics such as Nash-Sutcliffe Efficiency for different spatial scale and at different rainfall return frequencies. The study revealed that at larger spatial scale, the calibration performance of the model was about the same for the two different precipitation datasets while the study showed some benefit of NEXRAD for smaller watersheds. Similarly, the study showed that for smaller return frequency precipitation events, NEXRAD data was superior.

**Keywords:** Thiessen Polygon; NEXRAD; Rain Gauge Precipitation; HEC-HMS; Hydrologic Model; St. Johns River

---

## 1. Introduction

Computer models that simulate hydrologic runoff processes are essential tools for understanding and describing the overall hydrologic cycle. They are routinely used for important studies regarding water management, water quality issues, land use changes, flood inundation, and many other forecasting applications. The success of current model development and subsequent hydrologic prediction lies in the proper selection of model input parameters. Some researchers believe spatial and temporal variability of precipitation data are the main source of input data uncertainty when rainfall-runoff models are applied [1]. Since precipitation is the main driving mechanism for hydrologic models, selecting the suitable meteorological input dataset for precipitation is perhaps the most critical step in model development.

Traditionally, precipitation measurements from rain gauges or meteorological stations have been used as the only reliable source of precipitation in watershed modeling [2]. The benefit of rain gauges is their ability to obtain a precise point value for precipitation, with minimal data processing needed for use in hydrologic model applications. The main limitation in the reliance on rain gauge technology is that the rain gauge network cannot supply information about rainfall occurring between the gauges, and as a result, the network may not fully capture rainfall events demonstrating high spatial variability [3-4]. In an effort to use these point measurement locations, areal averaging of measured precipitation amounts is necessary. A commonly used method, the Thiessen Polygon method, calculates the weight of each rain gauge according to the rain gauge location to create a polygon network and applies the gauge rainfall quantity uniformly over the polygon area.

In recent years, technological advancements have provided rainfall estimates derived from remotely-sensed data. The use of remotely-sensed, radar-based data for rainfall estimation has evolved, presenting the opportunity of more accurate rainfall predictions. This has led to the use of satellite-based rainfall products, such as the Next-Generation Weather Radar (NEXRAD) data available in the USA and select overseas locations, in many hydrological modeling applications due to the availability of spatially continuous estimations at relatively fine spatial resolution. The NEXRAD data system was initiated in the early 1990s by the United States National Weather Service (NWS) and consists of a national network of radars known as the WSR-88D (Weather Surveillance Radar 1988) [5]. Due to the indirect nature of radar rainfall measurements, NEXRAD data may be subject to many sources of uncertainty such as radar-based factors (antenna, transmitter, and receiver), ground clutter, anomalous beam propagations, radar beam overshooting, and range effects caused by an increase in beam elevation and degradation of resolution due to beam spreading [2, 6-7]. Additionally, one of the largest sources of error can be the chosen reflectivity-to-rainfall (Z-R) relationship because it is directly related to the amount of precipitation estimated. Many empirical Z-R relationships have been developed because different climate conditions and rainfall characteristics can impact raindrop size distributions [8].

Since the acquisition of precipitation data and its application becomes more time consuming and expensive as the spatial resolution is increased (through the incorporation of NEXRAD data), the determination of the appropriate data source needed to provide satisfactory results in a hydrologic model is critical. Due to the limitations inherent with gauge-measured and radar-derived data, it is important to understand the quality of the data (in terms of precipitation quantity), possible bias, and systematic offsets. Many authors have found that NEXRAD underestimated rain when compared to rain gauge measurements [9,4], whereas Mazari [10] concluded that radar data was comparable to rain gauge measurements. Other researchers recommend understanding the potential differences in precipitation values, quality of radar data, and the overall bias of radar generated rainfall compared to rain gauge rainfall quantities, as they have a direct impact on hydrologic simulations [11].

The potential for improved accuracy of hydrologic model simulations and forecasts using radar data instead of point gauge data has been studied previously [12, 1, 13, 11]. Other research determined that NEXRAD data was generally less accurate in predicting the streamflow volumes using the hydrologic simulation code Hydrologic Engineering Center- Hydrologic Modeling System (HEC-HMS) [14] as compared to gauge-only simulations for two basins in central Tennessee, USA. Results of Kalin and Hantush [1] showed that during calibration of a Soil Water Assessment Tool (SWAT) model [15-16] in the Pocono Creek Watershed (~120 km<sup>2</sup>) the NEXRAD and rain gauge driven model performance statistics were comparable and the simulated hydrographs were similar to the observed flow hydrographs. Looper and Vieux [17] investigated streamflow prediction accuracy using radar-derived precipitation estimates and gauge observations on a physics-based distributed (PBD) hydrologic model, Vflo [18-19]. It was determined that the radar data showed more accurate hydrologic prediction for flash flood events. Results show that there are mixed conclusions in regards to whether radar-derived precipitation is a better alternative than traditional fixed instrument-based rain gauge data. Only limited inner comparison of watershed response improvement and its relationship at the spatial scale and rain intensity level has been completed.

This study was designed to assess the intrinsic value of the two precipitation input types (rain gauge versus NEXRAD) in the development of a rainfall-runoff hydrologic simulation model. Available stream gauges in the watershed were used to compare simulated discharges to actual discharges. These watersheds are typical for low-topographic drive systems situated around the globe. In order to assess and compare the value of the two input data types, a hydrologic simulation model was developed for sub-tropical watersheds located in central Florida, USA. The overall development of the model using HEC-HMS is briefly discussed while the focus of the research centers on the overall value of the two precipitation input data sets. In order to measure the value of each precipitation input dataset, a comparative assessment of the improvement or decline in calibration and validation performance statistics of the hydrological model results regarding:

(i) Watershed scale and, (ii) rainfall return frequency was completed. These proxies were then used to determine when NEXRAD data provides little to no advantage over rain gauge data, estimated using the Thiessen Polygon method, across a model domain. This work is very important as it clearly illustrates that automatically using the most complicated data input for hydrologic models may not be prudent if the results of the simulations deliver no advantage in model performance or accuracy. Ultimately, the research revealed that at larger spatial scale, the calibration performance of the model was about the same for the two different precipitation datasets while the study showed some benefit of NEXRAD for smaller watersheds. Similarly, the study showed that for smaller return frequency precipitation events, NEXRAD data was superior.

## 2. Materials and Methods

As was noted in the Introduction, this research compares the inherent value of two different precipitation input data sets for use in rainfall-runoff hydrologic modeling. The study approach uses a hydrologic model of a portion of the St. Johns River (SJR) watershed to do a comparative assessment. The study area chosen for simulation purposes was the SJR located in northeast, Florida USA. The general topography of the project area is flat, with an average slope of approximately 0.003 m/km, which results in a more lacustrine than riverine characteristic [20]. The climate of the central and eastern portion of Florida is mainly humid subtropical, similar to many other regions of the world at the same latitude. The average rainfall is between 1,100 mm to 1,500 mm per year [21]. The watershed is comprised of multiple sub-basins of various sizes and hydrologic properties that drain toward the flow-way of the SJR. Due to its size, the SJR has been divided into major "Basins" by the St. Johns River Water Management District (SJRWMD), a state agency whose work focuses on managing water supply, water quality and natural systems management, and flood protection (SJRWMD, 2012). The Upper St. Johns River Basin (USJRB) acts as the headwaters of the SJR and has an area of approximately 4530 km<sup>2</sup>. This Basin is mainly comprised of marsh and agricultural land with multiple storage areas used for flood control and environmental management. Additionally, there are multiple flood control projects within the area which include flood levees and water control structures. The Middle St. Johns River Basin (MSJRB) is downstream of the USJRB and encompasses approximately 3100 km<sup>2</sup>. It includes a variety of natural land types, but a majority of the basin is comprised of highly urbanized areas. For this study, a majority of the USJRB and portion of the MSJRB was simulated. The model domain covers roughly 5200 km<sup>2</sup> with a rain gauge density of approximately 217 km<sup>2</sup>/gauge. The watershed was divided into multiple sub-basins, delineated previously by the U.S. Army Corps of Engineers (USACE) and SJRWMD. Figure 1 shows the model domain with sub-basin delineation and available rain gauge locations. Gauges are maintained by the United States Geological Survey (USGS), USACE, South Florida Water Management District (SFWMD) and SJRWMD.



USDA Soil Conservation Services (SCS) Curve Number Method [22] was selected as the Loss Method for calculation of the infiltration of each sub-basin element. The SCS runoff equation is an empirical model that was developed for estimating runoff potential from the rainfall event based on the relationship between soil type, land use and antecedent soil conditions. The retention parameter, or Curve Number, was estimated using the Florida Land Use and Cover Classification System (FLUCCS) [23] and state soil geographic (STATSGO) database information. The SCS unit hydrograph Transform Method was used to compute the resulting surface runoff hydrograph based on calculated lag time. The Baseflow Method represents the runoff of prior precipitation and subsurface volumes for each sub-basin and the constant monthly baseflow method was selected.

The channel network is created through the use of reach elements, which are used to represent the main flow-way of the St. Johns River. The Muskingum-Cunge Routing Method, which is based on the conservation of mass and the diffusion representation of the conservation of momentum, was used to represent overland flow within the reach elements. In relation to the St. Johns River, one of the largest issues with routing is properly modeling the flood plain storage. According to HEC-HMS, flood flows through extremely flat and wide flood plains may not be modeled adequately as one-dimensional flow [24]. To overcome the potential overestimation in flow due to inadequate modeling of storage availability, a loss method was used in areas known to have low runoff potential, surface water withdrawals, and recharge rates.

Four simulation periods in the years between 2007 and 2011 were used for calibration and validation of the model, each representing a different rainfall return frequency. The calibration period occurred from August 2008 to October 2008 and represented a return frequency between a 5 to 10 year-24 hour duration. The validation periods occurred in October 2007, March 2010, and October 2011 with return frequencies of less than 1 year- 24 hour duration, approximately 1 year- 24 hour duration, and between a 10 to 25 year-24 hour duration, respectively. Two sets of daily precipitation data were used in the models: rain gauge observations and radar-derived estimates. USGS observed gauge measurements occur at multiple different locations within the model. Certain gauge locations are at the outlet of sub-basins, thereby representing the area of the upstream sub-basin(s). Any gauges located within the flow-way of the SJR would be representative of all sub-basins upstream, thereby allowing a large array of spatial scales to be analyzed. This paper will provide an abbreviated summary of the extensive model development and calibration effort herein, but the reader is referred to Tancreto [25] for complete details regarding the model. The complete thesis study can be downloaded at <http://digitalcommons.unf.edu/etd/584/>.

## 2.2. Precipitation Input to Models

The SJRWMD and SFWMD each operate a large network of rain gauges in or near the project area. A majority of the data has undergone quality assurance procedures, which aids in improving the accuracy and reliability of the data. Spatial estimates of precipitation across the entire model domain were developed using the Thiessen Polygon method, which defines an individual area of influence surrounding each gauge using the application of a Thiessen Polygon network. Each polygon is formed by the perpendicular bisectors of the lines joining adjacent gauges and represents areas of effective uniform depth. It is assumed that the gauge data, which was collected at a single point, is representative of the entire Thiessen Polygon. A common problem with the Thiessen Polygon method is having an insufficient number of rain gauges in a network, thereby assuming the areal rainfall is spatially homogeneous over large areas.

Hourly WSR-88D NEXRAD radar data, that has been gauge-adjusted from the network of rain gauges within the project area, was obtained from the SJRWMD for this study. The gauge and radar data are combined to calculate a gauge-radar ratio, which is applied in a radar calibration algorithm to derive a gauge-adjusted rainfall dataset. Additionally, proprietary geographic information system (GIS) algorithms are used to help reduce or eliminate any discontinuities and ground clutter from the radar station data [26]. The SJRWMD radar data product is generated at a 2 km pixel resolution grid format, which covers the entire project area. The data was aggregated to a daily time step to ensure consistent temporal scaling across precipitation input types.

Both precipitation input types required preprocessing in order to be used within HEC-HMS due to the various models selected to compute runoff processes being lumped in nature. The precipitation method selected in HEC-HMS was the Gauge Weights method, which uses separate parameter data for each gauge and for each sub-basin in the model. The ArcGIS 10.1 software [27] was used to compute the Thiessen polygons and the relative area of each sub-basin (from the sub-basin shape file) within each polygon.

### 2.3. Comparison of precipitation measurements

The quality of the radar-rainfall measurements, although continuously advancing, remains largely unknown. It is important to distinguish the relationship between the precipitation datasets, as they have a direct impact on hydrologic modeling results. Because the comparison of precipitation input in hydrologic simulation is the main objective of the research, the degree of similarity between the rainfall datasets must be analyzed. Comparative statistics will aid in understanding the potential differences in precipitation values, quality of the radar data, and the overall bias of radar generated rainfall compared to rain gauge rainfall quantities. The two statistical measures employed are the bias (B) and root mean square difference (RMSD), as utilized in several studies [11, 9, 4]. The estimation bias (B) is the ratio of the total difference in precipitation between the radar total and rain gauge total to the rain gauge total, as seen in Equation 1.

$$\text{Estimation Bias (B) (\%)} = \frac{\text{Radar}_{\text{Total}} - \text{Gauge}_{\text{Total}}}{\text{Gauge}_{\text{Total}}} * 100 \quad (1)$$

The RMSD was calculated to determine the degree of deviation or difference between the radar data predicted to the gauge value, which was actually observed. The RMSD represents the standard deviation of the differences between the predicted radar data and the observed gauge data. It is used to evaluate the goodness of fit for the study rain gauge and NEXRAD data. The RMSD equation can be seen below in Equation 2.

$$\text{RMSD} = \sqrt{\frac{1}{n} \sum_{i=1}^n (R_i - G_i)^2} \quad (2)$$

Where:

$G_i$  and  $R_i$  represent the  $i$ th day precipitation rate of gauge and radar, respectively  
 $n$  is the sample size of radar and gauge pairs

In previous studies, radar estimates are compared with the corresponding gauge observations [11, 6, 12]. Comparative statistics are normally performed for the radar pixels (one grid) that contain gauges. As discussed earlier, the NEXRAD data processing uses the gauge observations to adjust, correct, and sometimes replace radar estimates. This causes a lack of independence between the two sets of estimates at these locations [11]. Therefore, a direct comparison of the 2 by 2 km grid at the corresponding rain gauge data (point data) location may not provide an adequate representation of bias present within the dataset. To quantify the amount of total bias present between the datasets, precipitation inputs of particular sub-basins were chosen as the comparative sampling locations. Additionally, distance measurements were computed from the rain gauge location to the centroid of the sub-basin. This was important because as stated above, bias between the gauge and radar data may be reduced (through radar-gauge correction) at locations close to the gauges. For this reason, the distance was used to determine if the calculated biases were influenced by gauge location through the use of the coefficient of determination. It was not the purpose of this comparison to examine, in full detail, the accuracy of the NEXRAD data provided, but to provide insight on the bias that may be present when performing simulations. The sub-basin rainfall data chosen for comparison are believed to be representative of the entire basin as they are located throughout the modeled area.

Table 1 illustrates that for seven representative sub-basin rainfall measurements, the total percent bias can range from -36.1% to 53.3%. The average overall bias for 2007 and 2008 were the

greatest at 16% and 9%, respectively. The average bias of 2010 and 2011 was close to zero. This may be due to the fact that radar estimation accuracy has been continuously increasing over the past few years, resulting in higher quality radar data. The higher values of bias indicate that significant random differences may exist between the radar estimates and the corresponding gauge measurements for many of the sub-basins within the model. Additionally, a majority of the sub-basins show a positive bias which indicates that the NEXRAD data may be overestimating rainfall compared to the corresponding gauge. The total bias was plotted against the distance from the rain gauge to the centroid of the sub-basin. Linear regression was performed and the associated coefficient of determination ranged from 0.13 to 0.45. Therefore, no strong correlation can be detected regarding total bias in terms of distance from the rain gauge.

**Table 1.** Total Bias and RMSD between rain gauge and NEXRAD precipitation values.

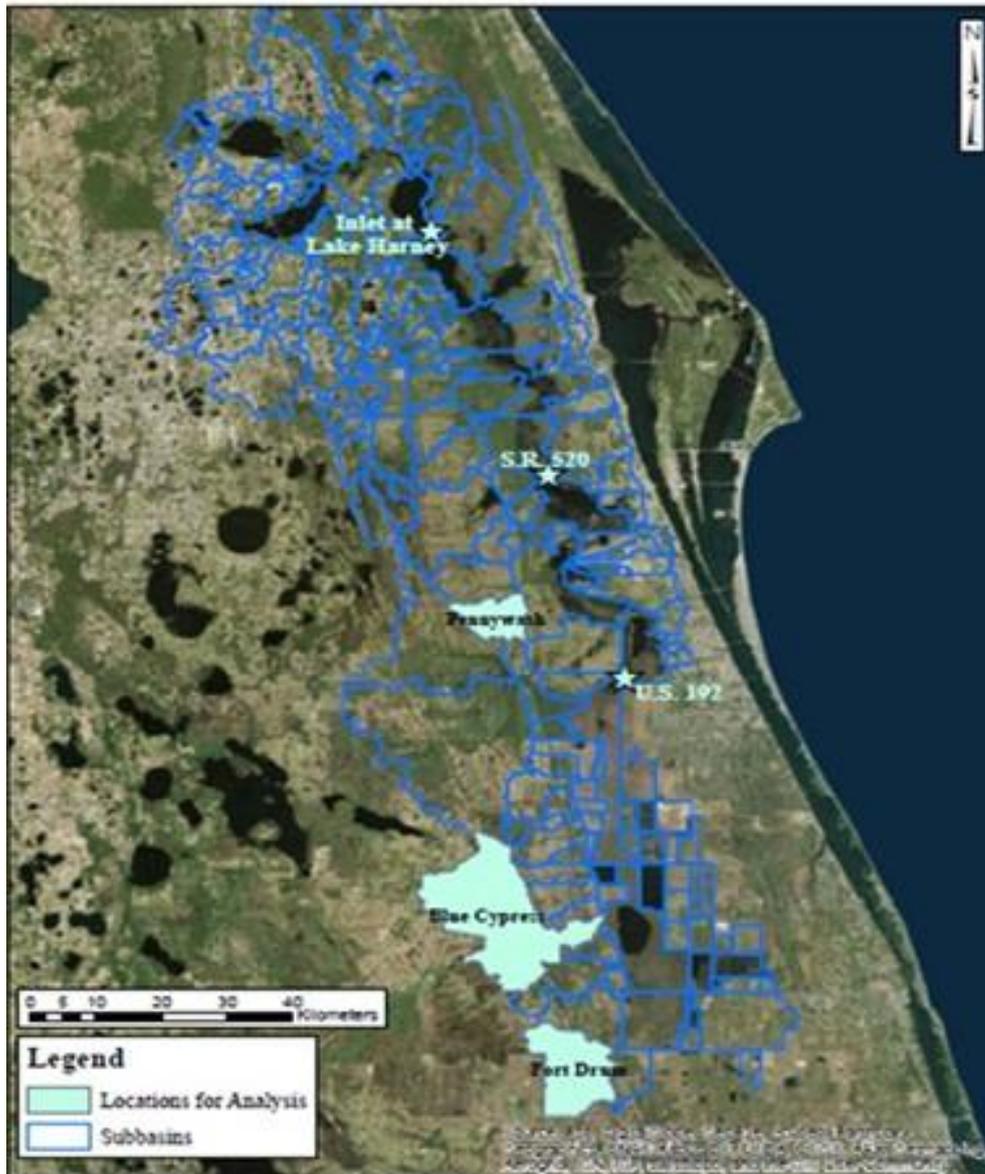
Sub-basin I.D.	Rain Gauge I.D.	Distance (m)	Total Bias (%)				Total RMSD (mm.)			
			Simulation Event				Simulation Event			
			2007	2008	2010	2011	2007	2008	2010	2011
Evans Grove	1880101	12,250	29.9%	-5.5%	-5.0%	-24.4%	7.37	8.38	4.32	12.70
C-54 Retention	510206	2,850	-3.5%	5.7%	1.4%	10.6%	3.81	7.11	2.29	6.60
Bull Creek	590263	8,300	53.3%	-1.1%	-21.1%	-2.9%	9.14	6.60	7.11	4.32
St. Johns Trib. 3	100109	3,500	28.0%	19.9%	6.0%	19.6%	2.54	7.11	2.29	7.62
Cox Creek Upper	530225	4,200	18.9%	19.9%	14.8%	-0.6%	3.05	5.08	2.29	2.03
Green Branch	11523739	1,500	24.1%	21.1%	4.6%	-0.8%	3.56	8.13	3.56	2.03
S.J.R. Puzzle	1140474	16,500	-36.1%	5.1%	1.2%	-1.3%	5.59	7.87	2.79	2.54

The RMSD values help quantify the random differences between the data sets to show the magnitude of error. As can be seen from Table 1, error up to 12.7 mm was calculated. This error can cause significant implications in hydrologic runoff calculations due to the undesired over- or under-estimation of rainfall amounts.

The comparative statistical tests show that bias and error can occur between the radar and rain gauge data sets. The model calibration process largely acts to correct this error or bias in precipitation data via parameter adjustment [13]. Many studies discuss the importance of model recalibration when switching precipitation products [28, 11, 2, 13]. Therefore, minor recalibration of the model was performed to ensure the model parameters are acceptable in order to produce the best possible hydrologic simulation results for each product.

#### 2.4. Model calibration and validation

The HEC-HMS model was calibrated and validated using existing observed storm events in the study area. The initial calibration of the model was completed using the rain gauge data. Once the model was deemed calibrated, the NEXRAD-derived data were input and the model was recalibrated. The primary goal of the calibration process is to match the simulation results to the observed data as closely as possible. The calibration locations were determined by the available discharge data locations available during the calibration timeframe and are shown in Figure 2. Table 2 provides a summary of the initial calibration performance of the model using rain gauge precipitation input, which was considered highly satisfactory. Each of the performance goodness-of-fit statistics are discussed in more detail below.



**Figure 2.** Location of representative calibration points where stream flow data exists.

**Table 2.** Initial model calibration performance statistics for representative basins.

Gauge Location	$r^2$	NSE
Fort Drum	0.89	0.85
Blue Cypress	0.84	0.84
Penneywash	0.86	0.79
US 192	0.88	0.86
FL 520	0.98	0.98
Inlet Lake Harney	0.93	0.90

After the initial calibration effort was completed, a comparison of calibration and validation performance was completed comparing both precipitation input datasets. Having known discharge gauge and simulation data allows a direct comparison to be made to evaluate the performance of the model using the two different precipitation input datasets. This comparison was completed using two different statistical measures: the coefficient of determination ( $r^2$ ) and the Nash-Sutcliffe efficiency coefficient (NSE). These statistics were used to quantitatively compare the hydrologic simulation results to determine which precipitation input method yielded the most accurate results for the various simulation events.

The  $r^2$  gives the variance of the data and assesses a goodness of fit at each calibration point for the model. The equation for  $r^2$  can be seen below in Equation 3. It can help explain the variability of the model and how well the model may produce results for future predictions. The coefficient of determination is between 0 and 1, with 1 indicating a perfect fit with all variation explained.

$$r^2 = \left( \frac{\sum_{i=1}^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}} \right)^2 \quad (3)$$

Where:  $O_i$  is the observed data on the  $i$ th day

$S_i$  is the simulated data on the  $i$ th day

$\bar{O}$  and  $\bar{S}$  is the observed and simulated mean values, respectively

$n$  is the number of observations

The NSE [29] is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance and indicates how well the plot of the observed data versus the simulated data fits the 1:1 line [30]. The NSE equation can be seen below in Equation 4.

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - S_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

Where:  $O_i$  is the observed data on the  $i$ th day

$S_i$  is the simulated data on the  $i$ th day

$\bar{O}$  is the observed mean value

$n$  is the number of observations

The ranges for NSE can vary between  $-\infty$  to 1, where:  $NSE=1$  corresponds to a perfect match between discharge data and observed data;  $NSE=0$  shows that the model predictions are as accurate as the mean of the observed data; and  $-\infty < NSE < 0$  occurs when the observed mean is a better predictor than the model, which indicates unacceptable performance [30]. The St. Johns River Water Supply Impact Study [20] used the Nash-Sutcliffe statistic to explain the calibration performance for their hydraulic model. Following similar methodology, the NSE values will be divided into intervals which explain performance rating. The intervals are as follows:  $0.75 < NSE < 1$  is a "very good" performance rating,  $0.65 < NSE < 0.75$  is a "good" performance rating,  $0.50 < NSE < 0.65$  is a "satisfactory" performance rating, and  $NSE < 0.50$  is an "unsatisfactory" performance rating.

The Wilcoxon signed-rank test is a nonparametric test of hypothesis used to determine whether there is a significant difference between the medians of two related groups. It is an analysis that is useful to determine if the population median-ranks differ between measurements that are repeated, also known as a paired difference test. The calculation of the Wilcoxon signed-rank test involves a  $W$  test statistic, whose distribution under the null hypothesis (distributions between the pairs are equal) is known. Using the test statistic  $W$ , a  $z$ -score and  $p$ -value can be calculated. If the  $p$ -value calculated is less than the significance level ( $\alpha$ ), then the groups are statistically significant and different with the null hypothesis rejected. The selected significance level for all tests was chosen as 0.05. This test was used to determine if the resulting NSE and  $r^2$  between the groups of data were statistically significant, that is, there is a distinct difference between the NEXRAD-generated and rain gauge-generated simulation results. If the null hypothesis is rejected then the simulation results are unlikely to have occurred by chance and the simulation results from each precipitation input type do indeed differ.

### 3. Results

#### 3.1. Simulation results at different rainfall return frequencies

The four simulation runs, representing the different rainfall return frequencies, were completed using two different rainfall input types. The final statistical results are presented on a per simulation basis, with a comparative framework designed to determine which precipitation input yielded more accurate results. The  $r^2$  and NSE values for each precipitation input type for the model calibration and validation periods are shown below in Table 3 and Table 4, respectively.

The model performance measures during the 2008 calibration period show a strong agreement between the simulated and observed streamflow data, with a high level of accuracy, as can be inferred from the high  $r^2$  and NSE values. Outcomes of the two statistical tests were compared using the Wilcoxon signed-rank test for the two precipitation input types. The Wilcoxon signed-rank test determined that the distributions between the two groups did not differ significantly ( $W=34.5$ ,  $z=0.13$ ,  $p=0.917$  for  $r^2$  and  $W=325$ ,  $z=0.297$ ,  $p=0.797$  for NSE). Both Wilcoxon sign-rank test p values suggest that there is no significant improvement of simulated streamflow values when using the NEXRAD data as compared to the rain gauge data.

The validation performance results for the October 2007 simulation event differ significantly depending on the discharge location, with a majority of locations showing relatively low performance values. The NEXRAD  $r^2$  values were compared against the rain gauge  $r^2$  values to determine improvement or decline. The median  $r^2$  values were calculated as 0.53 and 0.7 for the rain gauge and NEXRAD data, respectively, which is a favorable increase in performance. The Wilcoxon signed-rank test calculated W value was 55 and a p value of 0.233 was determined, thus the null hypothesis can only be rejected at an alpha value of 0.23, which is higher than the specific alpha value of 0.05. Many of the locations had little to no correlation (negative values), as determined by the NSE results, when using rain gauge data as the precipitation input method. The use of the NEXRAD data allowed significant improvement at ten USGS discharge gage locations. The median values for the NSE values were -0.06 and 0.53 for the rain gauge and NEXRAD data, respectively. For the NSE results, the calculated W value was 81 with a p-value of 0.01. This shows that a statistically significant difference of median is present at the alpha level of 0.05. Therefore, from the results it can be inferred that the NEXRAD calibrated data simulations estimate flow more accurately than the rain gauge simulations for rainfall return frequencies less than 1 year at a duration of 24 hours.

The 2010 and 2011 validation performance results show a strong relationship between the simulated and observed streamflow data for both precipitation input types. The  $r^2$  and NSE values are relatively close to one for a majority of the locations. An interesting observation made during the 2010 and 2011 validation effort was that minimal recalibration of model was needed between precipitation input types. The Wilcoxon signed-rank test determined that there is no significant improvement of streamflow values using the NEXRAD data over the rain gauge data for both validation events.

**Table 3.** Coefficient of determination simulation results.

USGS Gauge Identification Number	Coefficient of Determination ( $r^2$ )							
	2008		2007		2010		2011	
	Rain Gauge	Radar	Rain Gauge	Radar	Rain Gauge	Radar	Rain Gauge	Radar
02231342	0.89	0.93	0.49	0.70	0.89	0.91	0.83	0.77
02231396	0.89	0.90	0.52	0.63	0.91	0.90	0.80	0.80
02231600	0.91	0.90	0.64	0.87	0.94	0.95	0.76	0.78
02232000	0.88	0.89	0.90	0.88	0.96	0.93	0.98	0.97
02232155	0.86	0.87	0.11	0.55	0.85	0.85	0.74	0.76
02232200	0.87	0.87	0.29	0.86	0.86	0.86	0.82	0.84
02232400	0.98	0.98	0.48	0.30	0.97	0.96	0.99	0.98
02232500	0.95	0.93	0.95	0.89	0.95	0.92	0.94	0.93
02234000	0.93	0.91	0.72	0.60	0.95	0.93	0.95	0.95
02233460	0.85	0.91	0.16	0.23	N/A	N/A	N/A	N/A
02233473	0.89	0.92	0.53	0.56	N/A	N/A	N/A	N/A
02233484	0.90	0.80	0.85	0.81	0.90	0.89	0.94	0.95
02233500	0.87	0.78	0.82	0.82	0.65	0.67	0.96	0.95

\*N/A denotes USGS gauge data that was not available

**Table 4.** Nash-Sutcliffe efficiency coefficient simulation results.

USGS Gauge Identification Number	Nash-Sutcliffe Coefficient (NSE)							
	2008		2007		2010		2011	
	Rain Gauge	Radar	Rain Gauge	Radar	Rain Gauge	Radar	Rain Gauge	Radar
02231342	0.85	0.93	0.33	0.65	0.85	0.86	0.67	0.72
02231396	0.84	0.84	-0.06	0.11	0.86	0.86	0.72	0.74
02231600	0.90	0.89	0.41	0.81	0.91	0.94	0.59	0.64
02232000	0.86	0.87	0.74	0.72	0.98	0.97	0.98	0.97
02232155	0.79	0.79	-1.27	0.14	0.85	0.84	0.64	0.69
02232200	0.87	0.87	-0.11	0.77	0.83	0.83	0.80	0.83
02232400	0.98	0.98	-1.00	-1.25	0.96	0.94	0.99	0.98
02232500	0.94	0.92	0.67	0.53	0.93	0.91	0.94	0.89
02234000	0.90	0.88	-0.47	0.21	0.87	0.88	0.93	0.94
02233460	0.83	0.89	-2.44	-0.72	N/A	N/A	N/A	N/A
02233473	0.86	0.88	-2.86	-0.30	N/A	N/A	N/A	N/A
02233484	0.89	0.77	0.57	0.58	0.79	0.84	0.94	0.94
02233500	0.85	0.77	0.19	0.74	0.24	0.44	0.95	0.94

\*N/A denotes USGS gauge data that was not available

### 3.2. Simulation results at different spatial scales

The model performance measures were compared across simulation events for both individual sub-basins and at locations downstream of multiple sub-basins. The purpose of this analysis was to determine if, at certain sub-basin sizes, the NSE or  $r^2$  values would show improvement during the various simulation runs. To determine if NEXRAD data input produced more accurate results at

particular sub-basin scales, the  $r^2$  and NSE performance evaluations for all simulations were compared using the Wilcoxon signed-rank test. The  $r^2$  and NSE values were compared across all simulations events at the three sub-basin size groupings; small (less than 250 km<sup>2</sup>), medium (250 km<sup>2</sup> <  $x$  < 1000 km<sup>2</sup>), and large (over 1000 km<sup>2</sup>). The p value for the  $r^2$  and NSE comparison for the small sub-basins were 0.002 and 0.0002, respectively. Due to the p values, difference in computed median values, and previously computed improvement of  $r^2$  and NSE results, it was inferred that at smaller sub-basin sizes the NEXRAD data performs better. For the medium sized sub-basins a p value of 0.818 for  $r^2$  and 0.719 for NSE were computed. Since the p values are much higher than the significance level of 0.05 or 5%, the null hypothesis cannot be rejected. There is no statistical significance suggesting that the distributions between the datasets are different, and thus the NEXRAD data input does not improve  $r^2$  and NSE values at the medium sub-basin size. For the large sized sub-basins, the p value for the comparison of  $r^2$  and NSE values were 0.003 and 0.05, respectively. Since the p value for NSE comparison is equal to the significance level, the null hypothesis should not be rejected. This p value is very close to rejecting the null hypothesis thus suggesting that the rain gauge data may produce more accurate results than the NEXRAD data. The p value, higher median value, and previously determined decline in results when using NEXRAD data would suggest that rain gauge data produces more accurate results at the large sub-basin scale.

#### 4. Discussion

In this paper, a comparison of HEC-HMS hydrologic simulation performance using rain gauge and NEXRAD precipitation input at varying spatial scales and rainfall return frequencies for a large sub-tropical watershed with minimal topographic drive. In addition to the comparative model performance analysis, total bias between NEXRAD and rain gauge data within the study location was investigated. Precipitation measurements arguably have the most critical influence on the model performance, thus the need for quality data input is apparent. Comparing hydrologic simulation results using radar and rain gauge input aids in identifying the thresholds for maximum gain when using the more cumbersome, but finer-resolution radar data. This research provided guidance for both spatial scale and rainfall return frequency scenarios for which the use of radar data would yield more accurate hydrologic results.

Calibration and validation of a HEC-HMS hydrologic model of the Upper and Middle St. Johns River Basins was completed for four storm simulation periods, each representing a different rainfall return frequency. The HEC-HMS model was first calibrated using precipitation data from rain gauges located within or near the watershed boundary. As an alternative precipitation input source, NEXRAD data was obtained. Rain gauge and NEXRAD precipitation estimates were compared at seven locations within the model domain. The evaluation showed that NEXRAD total precipitation was greater than gauge total precipitation for a majority of the sub-basins, and recalibration between precipitation inputs was necessary.

Model performance was evaluated both visually and statistically against observed hydrograph data from USGS. The model performance measures, Nash–Sutcliffe efficiency (NSE) and coefficient of determination ( $r^2$ ), were used to quantitatively compare the NEXRAD and rain gauge hydrologic simulations to the observed USGS discharge data. Additionally, a statistical hypothesis test, the Wilcoxon Signed-Rank Test, was used to evaluate the difference in model performance results for the two precipitation input types. Overall, the calculated NSE and  $r^2$  values for the 2008, 2010, and 2011 simulations were similar and very promising (majority were > 0.75), indicating the model predicts streamflow values with a high level of accuracy for both NEXRAD and rain gauge data input. The Wilcoxon Signed-Rank Test results confirm that no significant improvement or decline in model streamflow accuracy is present when using NEXRAD data input for rainfall return frequencies of approximately 1 year-24 hour duration and greater. For the 2007 event, the NEXRAD precipitation data performed better than the rain gauge data at predicting the magnitude and timing of the peak, as reflected in the higher  $r^2$  and NSE values. A statistically significant difference of median is present (at the alpha level of 0.05). Thus, the NEXRAD data were shown to produce more accurate streamflow simulation results for rainfall return frequencies less than 1 year-24 hour duration. The model

performance measures were compared across simulation events at multiple spatial scales. NEXRAD data produced more accurate simulated streamflow values for the small sub-basin or watershed areas (less than 250 km<sup>2</sup>); neither NEXRAD nor rain gauge results show consistent improvement or decline in accuracy of the streamflow values for the medium sized sub-basin or watershed areas (250 km<sup>2</sup> < x < 1000 km<sup>2</sup>); and rain gauge data produced more accurate simulated streamflow values for the large sub-basin or watershed areas (over 1000 km<sup>2</sup>). The difference in median for the small and large sub-basin performance statistics were statistically significant.

The results of this study suggest that at small spatial scales and low return frequencies, NEXRAD data may produce more accurate streamflow estimations. Given that the performance of radar data in this study, it may be inferred that the spatial averaging of rain gauge Thiessen polygon data provides similar or more accurate rainfall estimations for large spatial scales and higher rainfall frequencies. This could be an indication of the spatial resolution of the rain gauges not capturing the spatial variability of smaller storm systems, which may be more convective in nature. Additionally, smaller spatial scales show a high level of sensitivity to rainfall input, thus the need for the higher spatial resolution of the NEXRAD data. It is important to note that the results of this study are conditioned on the modeling platform and precipitation data used. Many important factors should be analyzed before a precipitation input is selected such as; the spatial and temporal scale of the model, rain gauge data availability (coverage), quality of radar data, rainfall event, and model structure and spatial discretization. The conclusions of this study are not comprehensive of all watersheds, and thus care should be taken when assessing the results of this study for use in future modeling efforts. Further research should concentrate on identifying the rainfall return frequency threshold for which NEXRAD data may provide more accurate results for varying temporal scales. All simulation periods analyzed in this research were for a daily time period, over a relatively short duration. Introducing radar data at the hourly time scale may further improve model performance statistics, thus suggesting improvement of NEXRAD simulation results at certain return frequencies. Additionally, further research on model spatial discretization and its relation to streamflow accuracy at different spatial scales is needed. Due to the relative size of the model domain, the spatial discretization was relatively coarse for this project. Therefore, a full range of model benefits, including improvements in model accuracy, may not have been realized.

In summary, this study has clearly demonstrated that automatically using a larger and more complex precipitation input dataset is not justified in flatter, sub-tropical watersheds. In the future, it would be advisable for model developers to carefully consider the benefits of using more complicated and costly model inputs if the resulting benefits are negligible.

**Author Contributions:** Amanda Bredesen (formerly Amanda Tancreto) led the research efforts for this article as part of her thesis work. The conceptualization of the work was performed by Amanda Bredesen and Dr. Chris Brown. The model development was prepared by Amanda Bredesen with review by Dr. Chris Brown. The formal analysis was completed by Amanda Bredesen and reviewed by Dr. Chris Brown. This article was written by Amanda Bredesen and Dr. Chris Brown based upon work completed in the original thesis. Dr. Brown provided funding, project administration, and final article editing.

**Funding:** This study was partially funded with support from grant # 27884 from the St. Johns River Water Management District (SJRWMD) for the University of North Florida as a subcontractor to the University of Central Florida. It was part of a larger study for the economic evaluation of wetlands for the St. Johns River.

**Acknowledgments:** The authors would like to acknowledge support and data provided by the St. Johns River Water Management District, University of North Florida, University of Central Florida, and the U.S. Army Corps of Engineers, Jacksonville District.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Diaz-Ramirez, J. N.; McAnally, W. H.; Martin, J. L. Sensitivity of Simulating Hydrologic Processes to Gauge and Radar Rainfall Data in Subtropical Coastal Catchment. *Water Res Man* **2012**, *26*, 3515-3538. <http://dx.doi.org/10.1007/s11269-012-0088-z>.
2. Kalin, L.; Hantush, M. M. Hydrologic Modeling of an Eastern Pennsylvania Watershed. *J. of Hydrol Eng* **2006**, *11*(6), 555-569. <http://dx.doi.org/10.1061/ASCE1084-0699200611:6555>.
3. Huebner, S.R.; Pathak, C.; Hoblit, B.C. Development and Use of a NEXRAD Database for Water Management in South Florida. In ASCE EWRI World Water and Environmental Resources Congress Proceedings, Reston, Virginia USA, 2003.
4. Skinner, C.; Bloetscher, F.; Pathak, C. Comparison of NEXRAD and Rain Gauge Precipitation Measurements in South Florida. *J. Hydrol. Eng* **2009**, *14*(3), 248–260.
5. Winston, R. C.; SMITH, P. L.; Bigler, S. G.; Wilk, K. E.; Harbuck, A. C. History of Operational Use of Weather Radar by U.S. Weather Services. Part I: The Pre-NEXRAD Era. *Weather & Forecasting* **1998**, *13*(2), 219. Retrieved from <https://login.dax.lib.unf.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eih&AN=1011985&site=eds-live&scope=site>
6. Habib E.; Larson B.; Grascel J. Validation of nexrad multisensor precipitation estimates using an experimental dense raingauge network in south Louisiana. *J. of Hydrol* **2009**, *173*: 463–478.
7. Rendon, S. H.; Vieux B. E.; Pathak, C. S. Continuous Forecasting and Evaluation of Derived Z-R Relationships in a Sparse Rain Gauge Network Using NEXRAD. *J. of Hydrol Eng* **2013**, *18*(2), 175-182.
8. Pathak, C.; Curtis, D.; Kitzmiller, D.; Vieux, B. Identifying and Resolving the Barriers and Issues in Using Radar-Derived Rainfall Estimating Technology. *J. of Hydrol Eng* **2013**, *18*(10), 1193-1199.
9. Jayakrishnan, R.; Srinivasan, R.; Arnold, J. G. Comparison of raingage and WSR-88D Stage III precipitation data over the Texas-Gulf basin. *J. of Hydrol* **2004**, *292*(1), 135-152.
10. Mazari, N.; Xie, H.; Zeitler, J.; Sharif, H. O. Validation of the NEXRAD DSP product with a dense rain gauge network. *J. of Hydrol Eng* **2013**, *18*(2), 156-167.
11. Neary, V. S.; Habib, E.; Fleming, M.; Hydrologic modeling with NEXRAD precipitation in middle Tennessee. *J. Hydrol. Eng* **2004**, *95*, 339–349.
12. Sexton, A. M.; Sadeghi, A. M.; Zhang, X.; Srinivasan, R.; Shirmohammadi, A. Using NEXRAD and Rain Gauge Precipitation Data for Hydrologic Calibration of SWAT in a Northeastern Watershed. *American Society of Agricultural and Biological Engineers Journal* **2010**, *5*(3), 1501-1510.
13. Price, K.; Purucker, S. T.; Kraemer, S. R.; Babendreier, J. E.; Knightes, C. D. Comparison of radar and gauge precipitation data in watershed models across varying spatial and temporal scales. *Hyd Proc* **2014**, *28*(9), 3505-3520.
14. Hydrologic Engineering Center (HEC). (2010, August). Hydrologic Modeling System HEC-HMS User's Manual. U.S. Army Corps of Engineers, Davis, Calif.
15. Neitsch, S. L.; Arnold, J. G.; Kiniry, J. R.; Williams, J. R.; King, K. W. TWRI Report TR-191, Soil and Water Assesment Tool Theoretical Documentation Version 2000; Texas Water Resource Institute, College Station, Texas USA, 2002.
16. Neitsch, S. L.; Arnold, J. G.; Kiniry, J. R.; Williams, J. R.; King, K. W. TWRI Report TR-192, Soil and Water Assesment Tool User's Manual Version 2000; Texas Water Resource Institute, College Station, Texas USA, 2002.
17. Looper, J. P.; Vieux, B. E. An assessment of distributed flash flood forecasting accuracy using radar and rain gauge input for a physics-based distributed hydrologic model. *J. of Hydrol* **2012**, *412-413*(0), 114-132.
18. Vieux, B.E. Distributed hydrologic modelling for flood forecasting. In GIS and Remote Sensing in Hydrology, Water Resources and Environment. Post published Proceedings of the Three Gorges Dam Conference. IAHS Red Book Publication No. 289. ISBN:1-901502-72-4 (16–19 September 2003).
19. Vieux, B.E. Distributed Hydrologic Modeling Using GIS, 2nd ed; Water Science Technology Series, vol. 48; Kluwer Academic Publishers, Norwell, Massachusetts USA 2004, pp. 289, ISBN:1-4020-2459-2.
20. St. Johns River Water Management District. St. Johns River Water Supply Impact Study (Publication No. SJ2012-1), St. Johns River Water Management District 2012, Palatka, Florida USA.
21. National Academies Press (U.S.) Review of the St. Johns River Water Supply Impact Study : Final Report. Washington, D.C. USA 2012, National Academies Press. Retrieved from

- <https://login.dax.lib.unf.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=433014&site=eds-live&scope=site>
22. U.S. Department of Agriculture Soil Conservation Service, National Engineering Handbook, Section 4, Hydrology, available from U.S. Government Printing Office, Washington, DC USA, 1972.
  23. Department of State, State of Florida. The Florida land use and cover classification system : a technical report The Bureau of State, Tallahassee, Florida USA 1976. Retrieved from <https://login.dax.lib.unf.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat00488a&AN=unf.020459243&site=eds-live&scope=site>
  24. U.S. Army Corps of Engineers, Hydrologic Modeling System HEC-HMS, Rep. No. CPD-74B, at 145, Davis, CA USA 2000.
  25. Tancreto, A.E. "Comparison of Hydrologic Model Performance Statistics Using Thiessen Polygon Rain Gauge and NEXRAD Precipitation Input Methods at Different Water-shed Spatial Scales and Rainfall Return Frequencies". University of North Florida, Jacksonville, Florida USA, Theses and Dissertations, Master's Thesis 2015. Paper 584. <http://digitalcommons.unf.edu/etd/584>
  26. St. Johns Water Management District (SJRWMD). (2015). Radar Rainfall. Retrieved February 5, 2015, from <http://webapub.sjrwmd.com/agws10/radrain/index.htm>.
  27. Esri. ArcGIS 10.1 User Guide 2012, Redwood City, CA USA
  28. Bradley, A. A.; Kruger, A. Recalibration of hydrologic models for use with WSR-88D precipitation estimates, Preprints, Special Symposium on Hydrology; 78th Annual AMS Conf., American Meteorological Society, Boston, MA USA 1998.
  29. Nash J.E.; Sutcliffe, J.V. River Flow forecasting through conceptual models, 1, A discussion of principles. *J. of Hydrol* **1970**. 10: 282- 290.
  30. Wang, W. C.; Chau, K. W.; Cheng, C. T.; Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J. of Hydrol* **2009**, 374(3), 294-306.
  - 31.



© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).