2018-04-01

An Analysis of Using Error Metrics to Determine the Accuracy of Modeled Historical Streamflow on a Global Scale

Elise Katherine Jackson
Brigham Young University

An Analysis of Using Error Metrics to Determine the Accuracy of Modeled Historical Streamflow on a Global Scale

Elise Katherine Jackson

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

E. James Nelson, Chair
Daniel P. Ames
Norman L. Jones
Gustavious P. Williams

Department of Civil and Environmental Engineering

Brigham Young University

ABSTRACT

An Analysis of Using Error Metrics to Determine the Accuracy of Modeled Historical Streamflow on a Global Scale

Elise Katherine Jackson
Department of Civil and Environmental Engineering, BYU
Master of Science

Streamflow data is used throughout the world in applications such as flooding, agriculture, and urban planning. Understanding daily and seasonal patterns in streamflow is important for decision makers, so that they can accurately predict and react to seasonal changes in streamflow for the region. This understanding of daily and seasonal patterns has historically been achieved through interpretation of observed historical data at stream reaches throughout the individual regions. Developing countries have limited and sporadic observed stream and rain gage data, making it difficult for stakeholders to manage their water resources to their fullest potential.

In areas where observed historical data is not readily available, the European Reanalysis Interim (ERA-Interim) data provided by the European Center for Medium-Range Weather Forecasts (ECMWF) can be used as a surrogate. The ERA-Interim data can be compared to historic observed flow to determine the accuracy of the ERA-Interim data using statistical measures such as the correlation coefficient, the mean difference, the root mean square error, R^2 coefficients and spectral angle metrics. These different statistical measures determine different aspects of the predicted data's accuracy. These metrics measure correlation, errors in magnitude, errors in timing, and errors in shape.

This thesis presents a suite of tests that can be used to determine the accuracy and correlation of the ERA-Interim data compared to the observed data, the accuracy of the ERA-Interim data in capturing the overall events, and the accuracy of the data in capturing the magnitude of events. From these tests, and the cases presented in this thesis, we can conclude that the ERA-Interim is a sufficient model for simulating historic data on a global scale. It is able to capture the seasonality of the historical data, the magnitude of the events, and the overall timing of the events sufficiently to be used as a surrogate dataset. The suite of tests can also be applied to other applications, to make comparing two datasets of flow data a quicker and easier process.

Keywords: correlation, streamflow modelling, statistical analysis, ERA-Interim

ACKNOWLEDGEMENTS

I would like to thank my incredible parents for all their support throughout my education and as I have worked on this research. They have been an unwavering source of love and encouragement throughout my life. I would also like to thank the rest of my family for their love and support throughout this process. I would like to thank Dr. Jim Nelson for allowing me to pursue this degree and be a part of the BYU Hydroinformatics research group. I would like to thank Dr. Gus Williams for his guidance and assistance in this project. I would like to thank SERVIR, NASA Applied Sciences Program, Grant No. NNX16AN45G for funding and supporting this research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Observed data are a necessary component for planning and developing water resources projects. It is an essential element for any study such as flood forecasting and mapping, hydropower exploration, agriculture and municipal planning, and any other hydrologic project. Improvements in data availability and communication of information derived from these data to decision-makers and stakeholders has many benefits (Pappenberger, et al., 2015). Developing countries however, have limited and sporadic observed stream and rain gage data, making it difficult for stakeholders to manage their water resources to their fullest potential.

1.1 Past Research and Background

Where historical data does not exist, simulations from retrospective meteorological data can be done to provide estimates for historical data. This is done in areas where resources are not, or have not been sufficient to characterize historical precipitation or flow data. Generating artificial data via numerical and hydrologic modeling to fill in the historical gaps then allows these areas to better develop their water resources. The European Center for Medium-Range Weather Forecasts (ECMWF) is an independent, intergovernmental organization supported by many European countries, and is the world's largest archive of modeled weather prediction data (ECMWF, 2017). The ECMWF provides an ensemble of 52 medium-range weather forecasts out to 15 days, and seasonal forecasts out to 12 months, based on a gridded runoff prediction dataset (Basalmo et al., 2009). These forecasts are provided to national weather services and are used as

a complement to their respective short-range and climatological studies (Wetterdienst, 2014). The ECMWF forecasts are also used as an early warning system of severe weather, such as hurricanes and flooding (Roulstone & Norbury, 2013).

The ECMWF produces these forecasts by assimilating meteorological data from satellites and earth observations. This data assimilation serves as the initial state for a computer model, which uses the resulting atmospheric model to forecast the weather. These ensemble forecasts are computed by a medium-resolution forecasting system operating at a 16-kilometer resolution. The high-resolution forecasting system operates at an 8-kilometer resolution, giving the forecasts better range with high quality.

Using the same basic atmospheric model, the ECMWF also produces a historical meteorological dataset based on global atmospheric reanalysis. This historical meteorological data known as the European Reanalysis Interim (ERA-Interim) data can be used as a 35-year historical record for sites where observed data do not exist (Dee et al., 2011). The ECMWF then validates this 35-year simulated record against available observed data and further calibrates the ERA-Interim data as necessary.

1.1.1 ERA-Interim Data

The ECMWF's ERA-Interim forecast was created to address difficulties with data assimilation related to the representation of the hydrological cycle, the quality of the stratospheric circulation, and time consistency of reanalyzed geophysical fields. The ERA-Interim data is paired with the Hydrology-Tiled ECMWF Scheme for Surface Exchanges over Land (H-TESSEL), a land surface numerical scheme to create the ERA-Interim/Land, a global land-surface dataset. The ERA-Interim/Land dataset preserves the closure of the water balance,

and includes the defining equations of the land surface scheme with respect to the original ERA-Interim precipitation data (ECMWF, 2017). The ERA-Interim/Land dataset is compared with ground-based and remote sensing observations to assess the quality of the dataset. Effects of this calibration have been verified in monthly and seasonal forecasts used by the ECMWF (Balsamo, et al., 2012). The ECMWF uses the Root Mean Square Error (RMSE) metric for different vectors and variables, to further calibrate the ERA-Interim/Land dataset (Uppala, Dee, Kobayashi, & Simmons, 2008). Others have analyzed the effect of climate variables, and the ability of the ERA-Interim dataset for capturing monthly variability (Simmons, Willet, Jones, Thorne, & Dee, 2010).

This reanalyzed, calibrated data can be used to provide a multivariate, spatially complete, and coherent record of global atmospheric circulation (Dee, et al., 2011). The ERA-Interim/Land data provides a global estimate of weather patterns, and can be used as the initialization for numerical weather prediction and climate models.

1.1.2 Modeled Historical Streamflow Data

To use the ERA-Interim/Land data with respect to streamflow, the modeled historical precipitation data must be routed through stream networks to develop flow predictions. The Routing Application for Parallel computatIon of Discharge (RAPID), uses the ERA-Interim/Land data on the ECMWF grid to generate streamflow data for stream reaches on a global scale (Snow, 2015). The RAPID process creates a modeled historical streamflow (MHS) dataset, with a daily streamflow value for every date within a 35-year period. This data set can then be changed by calibrating the RAPID model and used as surrogate data where historical data do not exist.

The MHS data has been improved by performing RAPID stream routing on a higher density stream network. This increases the number of locations where simulated historical data can be accessed and analyzed. The Streamflow Prediction Tool (SPT), available on the Tethys platform makes these data available through a web interface (Snow, 2015). An example of using the SPT is shown in Figure 1-1.

Figure 1-1: Using the SPT to find data at Camu Bayacanes, Dominican Republic.

The SPT allows us to view the MHS data, as well as the monthly and daily averages over the 35-year period, for any specified stream reach. An example of the dialog box, showing the MHS data, is shown in Figure 1-2.

Figure 1-2: MHS data for Camu Bayacanes, Dominican Republic.

The data from the SPT can be merged with historical observed data and downloaded as a .csv file of the two different data series. Once the MHS data are made accessible, the data can be validated through the library of statistical analysis tools.

1.1.3 Statistical Analysis

Validating and calibrating modeled data is important to understand associated limitations and applications for the given model. Researchers have developed methods for defining hydrologic trends in hydrologic variables such as streamflow and climate (Burn & Hag Elnur, 2002). These same types of analysis can be used to characterize the correlation between observed streamflow, and simulated streamflow. The correlation between the observed and predicted stream flow is a measure of the model accuracy and can be used to determine the quality of the simulated data. Correlation measures can be affected by the resolution of the predicted data and the variability in the streamflow, so acceptable correlation metric values are site or region specific, rather than absolute. Some researchers have used regionalization to determine the accuracy of models, allowing predictions to be better calibrated for the specific region (Post & Jakeman, 1999).

There are a very large number of metrics or tests that can be used to determine the correlation between two time-series datasets, some more useful than others for streamflow data. Reich, et al. (2016) analyzed different error metrics such as the correlation coefficient R^2, the mean difference, and the variance of the difference to assess their applicability. Tornquist, Vartia and Vartia (1985) have investigated indicators of relative change between two datasets, and the effect of log transformations on error statistics. Also of interest are lag correlation metrics for the predicted model which can show timing or offset errors. These metrics are often calculated by

using the R^2 value and the spectral angle over a range of lagged values (i.e., data set offsets). Hyndman and Khandakar (2008) used these lag correlation measures to better understand how the dataset has captured specific events, even if the timing is slightly off.

1.2 Research Objectives

The purpose of this research is to develop a library of statistical analysis tools that allows investigators to quickly and easily conduct a statistical analysis of the modeled historical streamflow (MHS) data as compared to historical data and determine the appropriateness of using the MHS data for water resources planning and development. These statistical analysis tools specifically determine:

1) The accuracy and correlation of the MHS data compared to the observed data.
2) The accuracy of the MHS dataset in capturing the overall events shown by the observed data.
3) The accuracy of the MHS dataset in capturing the magnitude and timing of events shown by the observed data.

These analysis tools are available through a Python or MATLAB library, making them easily accessible to water resources agencies to aid in the development of water resources applications.

This thesis will also present a case study using the MHS data in a hydropower case study in the Dominican Republic. Using the MHS data as surrogates for observed data has the potential to expand the ability of countries to analyze potential hydropower sites and improve current estimates of hydropower capacity. The MHS data can also be used as a resource for agricultural and other water supply applications.

2 METHODS

Several different statistical tests were developed to determine the accuracy of the modeled historical streamflow (MHS) data. Specifically, the accuracy and correlation of the MHS data compared to the observed data overall, as well as with respect to timing of specific events and magnitude of specific events. Examples of the different methods using data from stations in Nepal are included in this section as examples.

To begin the analysis, the MHS data corresponding with an observed data station were obtained using the SPT. Both the MHS data and the observed data were then transformed using a natural log scale, and statistically analyzed. Because flow data in a time series are not normally distributed, the log scale is used to account for any outliers and skew in the data. The paired, logged dataset was used to determine the accuracy of the MHS data.

2.1 Visual Tests

The first and most basic method for gauging the accuracy of any simulated dataset is to visually inspect the data. Most simulated data show one or more of the following trends: over prediction, under prediction, or a time shift. Figure 2-1 shows both an over predicted and an under predicted series for a base series.

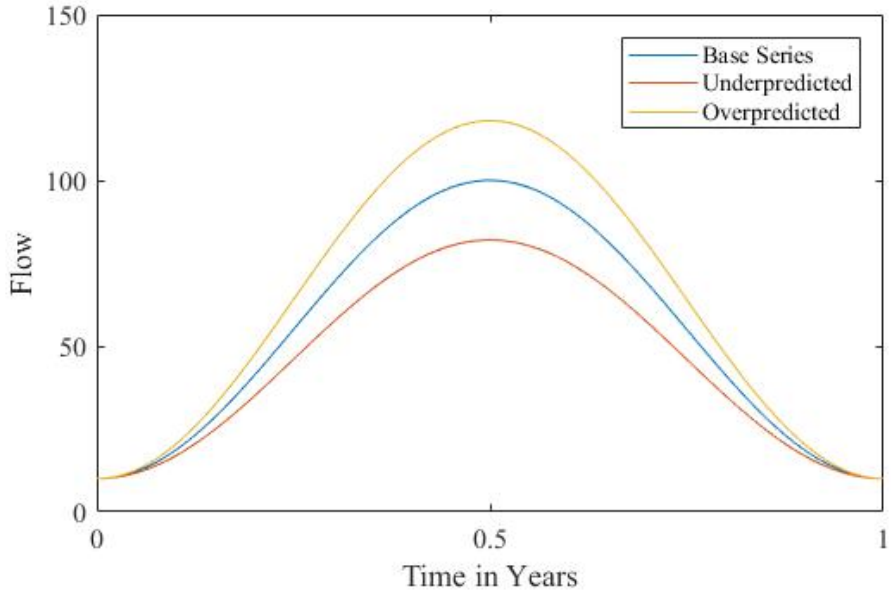

Figure 2-1: Example of over and under-prediction on a base series.

For these realizations, the timing of the flow event is correct for both the under predicted and over predicted series, but the magnitude of the event is incorrect. Figure 2-2 shows a realization with time shift from the same base series, without any under or over prediction.

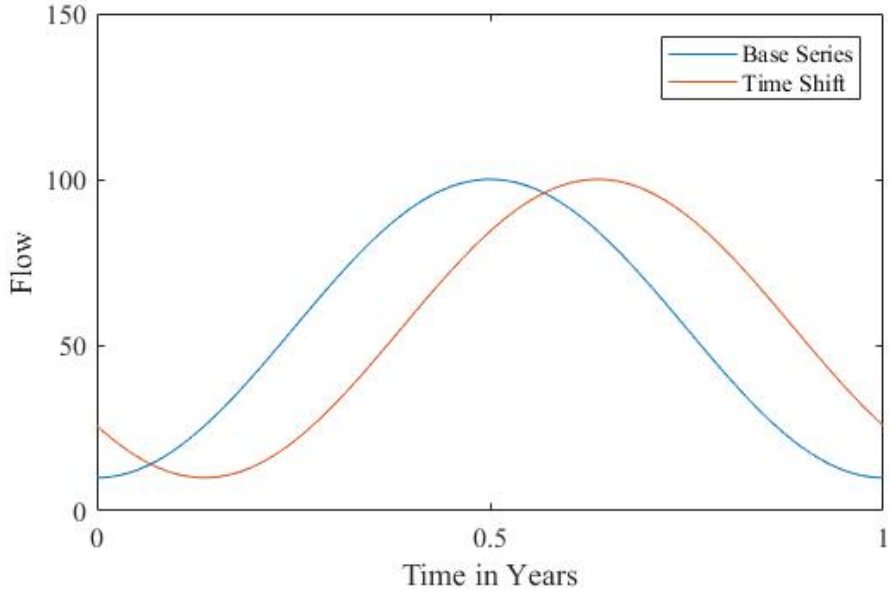

Figure 2-2: Example of time shift on a base series.

For this realization, the magnitude of the flow event is correct, but the timing of the event is incorrect. Figure 2-3 shows two time-shifted realizations with over prediction and under prediction compared to the base series.

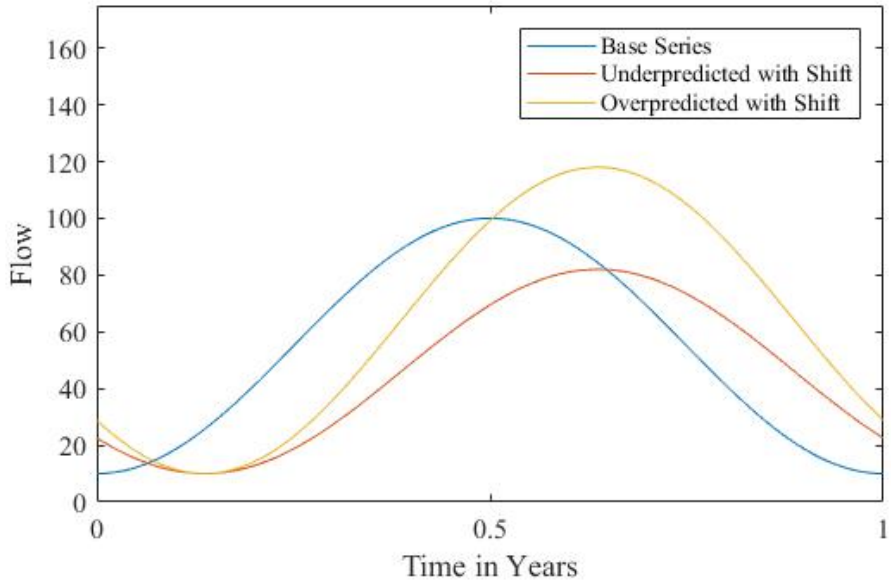

Figure 2-3: Example of time shift with over and under-prediction on a base series.

These types of errors are the most common for simulated hydrologic time series data. The simulated data usually model the overall trends relatively well with errors in the timing and magnitude of flow events.

Another visual analysis tool that can be used to determine the accuracy of the data is a quantile-quantile or Q-Q plot. The Q-Q plot is a graphical tool that helps researchers determine the distribution of the data, whether it is normally or otherwise distributed. There are two types of Q-Q plots, those that compare data to a specific distribution, or those that compare two datasets together to determine how the two datasets correlate.

The Q-Q plot of a distribution plots the data versus a normal distribution, and if the data falls along a 45° line, the data is normally distributed (University of Virginia Library, 2018). The quantiles can be thought of as percentiles, or the points in your data below which a certain proportion of the data falls. In a normal distribution, half of the data should fall below the 50th percentile, or 0. An example of the synthetic hydrograph without any changes is shown in Figure 2-4.

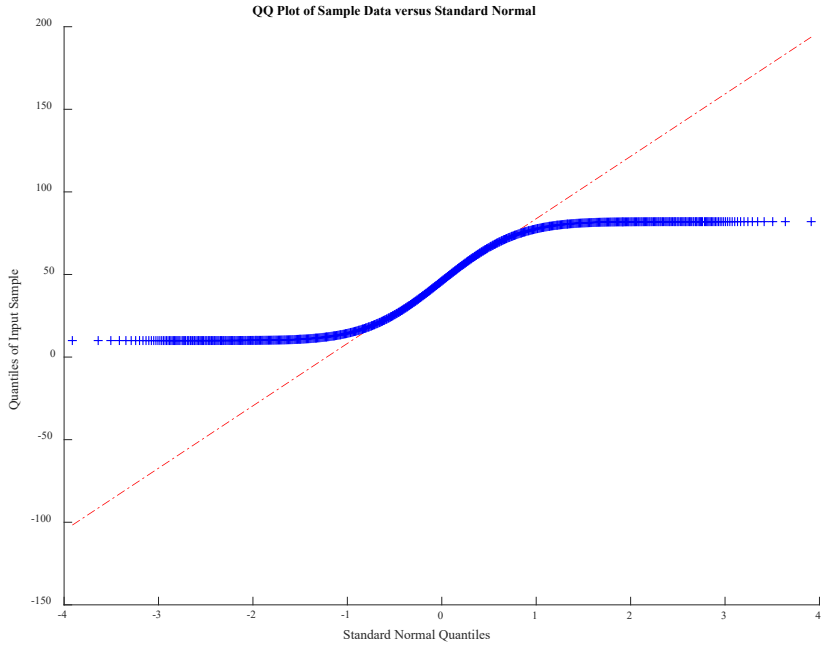

Figure 2-4: Q-Q plot of synthetic hydrograph.

The synthetic hydrograph shows a skew in the data, especially on the tails. This is typical of hydrographs as they naturally are rarely normally distributed due to outliers in the high flow seasons, as well as differing base flows in the stream reach throughout the year. The data must be transformed using a log transformation to better approximate a normal distribution in the data. The Q-Q plot is used to justify using the log transformation of the data. This same procedure can

be done by plotting two datasets against each other, as shown in Figure 2-5. In this type of Q-Q plot, quantiles are not used to describe the data.

2.2 Distribution Tests

Different types of statistical tests can be used to describe and quantify errors in the distribution of the simulated data compared to distribution of the observed data. The distribution of the data refers to how the different data points within the datasets are spread across the time period in question. The distribution of the data is dependent on the timing of events within the dataset. Correlation tests are most sensitive to the effects of timing errors, while error tests are sensitive to the effects of magnitude changes, caused by either timing or under- and over-prediction. Lag correlation tests can quantify timing errors and can be used in conjunction with the correlation and error tests to determine the accuracy of the simulated data.

2.2.1 Visual Tests

Visual tests can be used as a precursor to statistical metrics to determine the need of any error metrics. They also can give researchers an initial idea of the differences in the datasets, and their distributions.

Figure 2-5 shows an X-Y scatterplot of the observed flow compared to the predicted flow at Marsyangdi, Nepal. This differs from the Q-Q plot in Figure 2-4 in that we are comparing the predicted and observed flows, rather than the data to a statistical distribution.

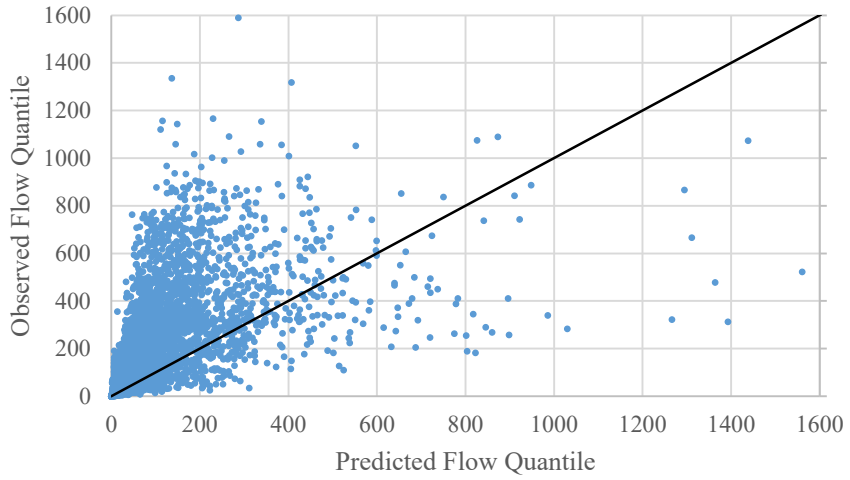

Figure 2-5: X-Y scatterplot of predicted flow vs. observed flow for Marsyangdi, Nepal.

The X-Y scatterplot shows that the distributions of the two data are not the same, and that there is significant difference between the two datasets. This conclusion can also be seen by creating an X-Y scatterplot by plotting the predicted flow versus the observed flow, as well as the log predicted flow versus the log observed flow. If the predicted flow were the same as the observed flow, the data would fall along the black line. The plot for the log transformed data is shown in Figure 2-6.

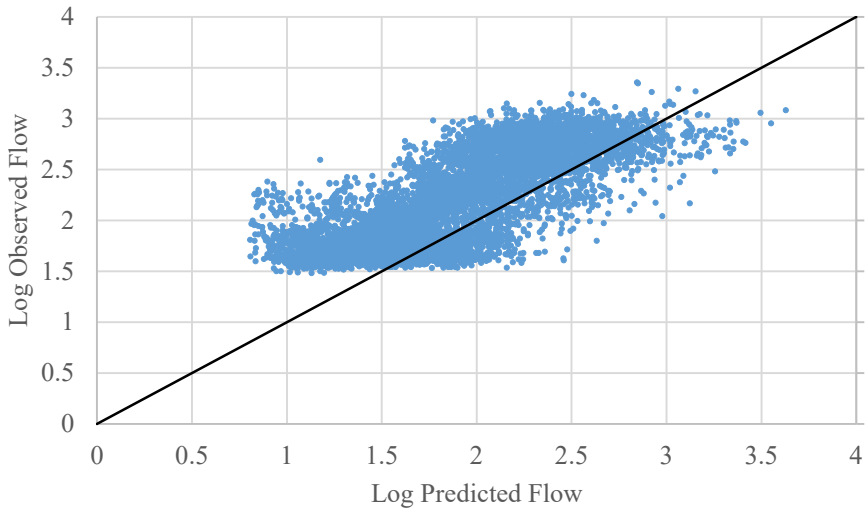

Figure 2-6: X-Y Scatterplot of log predicted flow vs. log observed flow.

From these two plots, the differences between the two flows can be seen. The comparison between the predicted and the observed flow is somewhat distributed along the match line for flows below 500 cubic meters per second. As the flow values increase, they fall farther from the match line. From Figure 2-6, the data are more similar. The log transformation removes any outliers from the datasets, removing the error shown in Figure 2-5. Both of these figures show the differences between the MHS data and the observed data, which must then be further explained using statistical metrics. A paired t-test was performed on the transformed data to quantify the correlation between the datasets by each month. An example result of a paired t-test is shown in Figure 2-7.

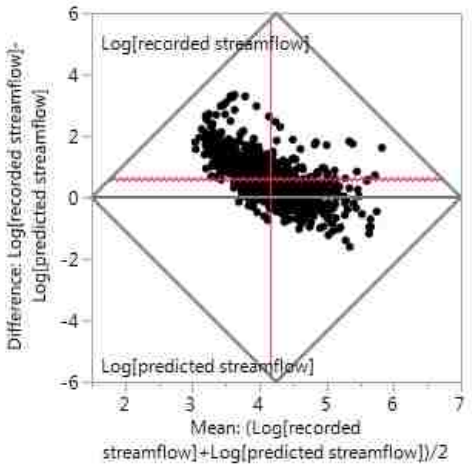

Figure 2-7: Paired t-test for Marsyangdi, Nepal for the month of May.

The paired t-test shows the overall trend between the observed and predicted data. The correlation value for the datasets can also be analyzed. The paired t-test graph consists of a diamond which encloses all data points when the range of the differences is greater than half the range of the data. The black, horizontal line defines the points where the observed streamflow is equal to the predicted streamflow. A perfectly correlated series would follow this line perfectly, while a negative correlation would be oriented vertically. From this graph, we can see the slight,

positive correlation between the predicted flow and the historical flow, as most of the points are plotted horizontally, with some outliers. This slight, positive correlation shows that the predicted data are similar to the observed data with respect to the actual values. The paired t-test also generates the mean difference value between the two datasets, which will be explained in the next section.

The paired t-test is more commonly used by statisticians, and requires previous knowledge of statistics to understand completely. As a result, hydrologists and researchers more commonly use the Q-Q plot, or an X-Y scatterplot as an initial test to understand the distribution of datasets.

2.2.2 Correlation Coefficient

The anomaly correlation coefficient is a measure for describing the linear association, or one-to-one comparison between any two variables. The correlation coefficient ignores any potential or biases, minimizing the seasonal effect (Stevenson, 2006). The equation for calculating the correlation coefficient is shown in Equation 2-1.

$$r_{XY} = \frac{\sum_{I=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) / (n-1)}{s_x s_y} \quad (2\text{-}1)$$

Where s_x and s_y are the sample standard deviations. The sample correlation coefficient is dimension free, ranging from -1 to +1, where the extremes correspond to cases of perfect, 1:1 correlation between the datasets. A positive correlation means that as the values in one dataset increase, the values in the other dataset also increase. A negative correlation means that as the values in one dataset increase, the values in the other dataset decrease. A correlation coefficient of zero corresponds to situations where there is no linear association (Ramsey & Schafer, 2013).

P-values can be calculated for each error metric, including the correlation coefficient. The p-value is a measure of the probability associated with finding an error metric larger than the one observed. In other words, a small p-value is associated with a high confidence in the test statistic, while a p-value greater than 0.10 gives no evidence of confidence. A p-value less than 0.001 indicates that the error metric is accurate. However, because the p-value is dependent on the number of samples, large datasets will generally have very small p-values for each error metric. As the datasets used in this paper contain many different data points, p-values are not used as measures of significance for each error metric, but can be calculated to show the relative importance of an error metric between different stations.

2.2.3 Example of Correlation Tests

The correlation coefficient, as well as the mean difference value and variance value are summarized below in Table 2-1 for the Marsyangdi station in Nepal.

Table 2-1: Correlation Test Summary by Month for Marsyangdi, Nepal

Month	Correlation Coefficient	Mean Difference	Variance
January	0.2400	2.2845	0.0944
February	-0.0516	1.7557	0.2172
March	0.1428	1.3343	0.5241
April	0.1882	1.6437	0.6049
May	0.3765	1.8576	0.6141
June	0.4993	1.9650	0.8413
July	0.2066	2.2853	0.5882
August	0.2661	2.4303	0.4160
September	0.4347	2.2842	0.3132
October	0.4511	2.5001	0.2307
November	0.3934	2.5797	0.0849
December	0.2810	2.5305	0.0887
Average	0.2857	2.1209	0.3848

The mean difference value is the geometric mean of the two datasets. The Marsyangdi mean difference value of 2.2845 for the month of January signifies that the observed value is 2.2845 times larger than the predicted value, with a confidence interval for the mean difference between 2.2362 and 2.3339. The variance value for January of 0.0944 signifies that the average data point for the month of January is 0.0944 greater or lesser than the average mean value for the station. The average correlation coefficient for Marsyangdi is 0.2857, signifying a slightly positive correlation between the predicted and observed flow data.

2.2.4 Error Tests

Calculating error can be done by considering the relative differences between observed and predicted data. Most flow data are transformed using a logarithmic function to account for the non-normal distribution, and any outliers that may occur. The most common terms used to define forecast error are the Root Mean Square Error (RMSE) and the Mean Squared Error (MSE). The equation for the MSE is shown below in Equation 2-2.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \quad (2\text{-}2)$$

Where y_i represents the observed value, and x_i represents the predicted value (Lehmann & Casella, 1998). The mean squared error measures the average of the squares of the standard deviations, or rather the variance in the differences between the observed values and the predicted values. An MSE of zero signifies a perfect predicted dataset.

The RMSE is a function of the MSE, and is the square root of the mean square error. It is simply the average distance of a data point from the observed data point measured along a vertical line. The equation for the RMSE is found in Equation 2-3.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2} \quad (2\text{-}3)$$

Figure 2-8 shows an example of the RMSE between an observed base series and an under predicted dataset.

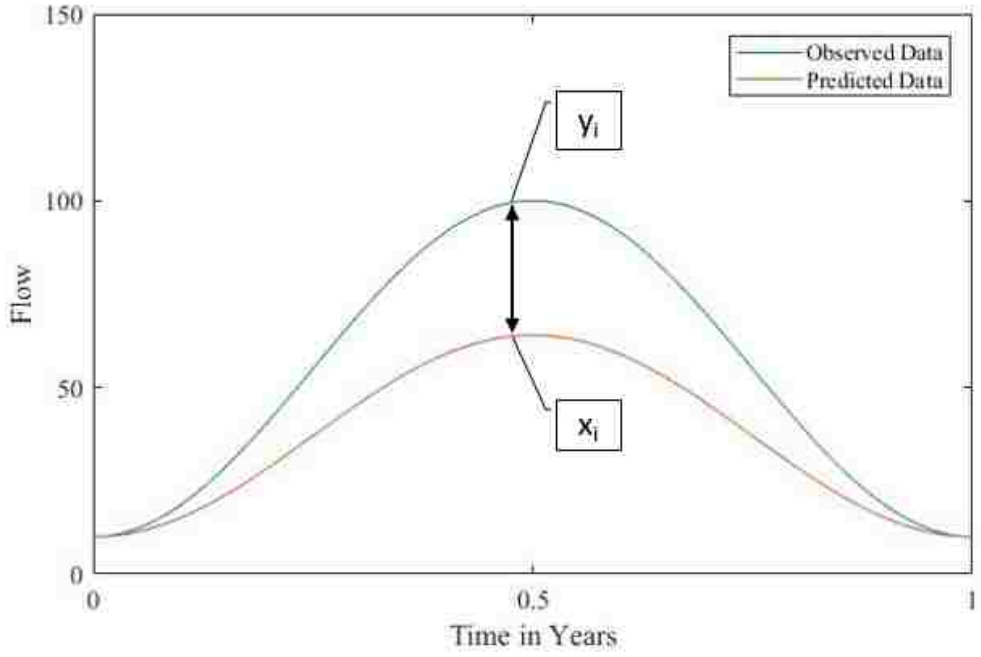

Figure 2-8: Visual depiction of RMSE for observed and under predicted data.

The RMSE is very sensitive to magnitude changes, as shown in Figure 2-8. Changes in timing have a very small effect on the RMSE, as shown in Figure 2-9.

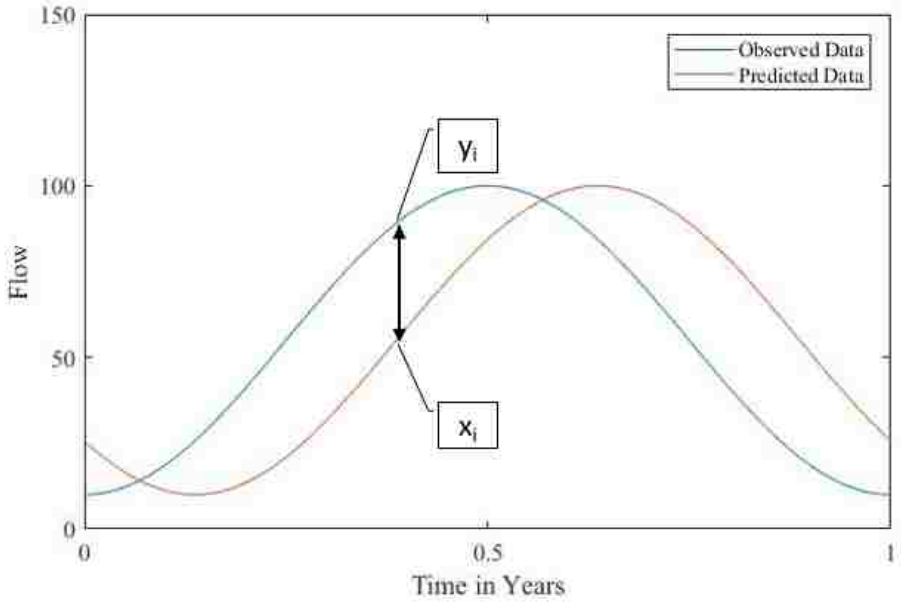

Figure 2-9: Visual depiction of RMSE for a time-shifted predicted dataset.

The RMSE is adaptable to situations when large errors are undesirable. The RMSE does not increase with the variance of the errors, but rather with the variance of the frequency of the error magnitude. This means that as the same error reappears in the dataset, the RMSE increases. The RMSE of the log error (RMSLE) can also be used to analyze the data where the variance in the data has been reduced by the log transformation (Tornquist, Vartia, & Vartia, 1985). These two measurements of error are the most commonly used in hydrological analysis (Boyle, Guta, & Sorooshian, 2000).

Table 2-2 shows an example of these error tests are shown for the Kankai station in Nepal. The standard error is the squared difference between the observed and predicted value divided by the number of data points. From these values, the MSE and RMSE can be calculated. These are tabulated in Table 2-3.

Table 2-2: Standard Error for Kankai, Nepal

Month	Observed Flow	Predicted Flow	Standard Error
January	11.73	10.48	0.129
February	9.98	8.33	0.227
March	9.24	6.16	0.790
April	10.01	6.65	0.945
May	15.12	15.88	0.048
June	38.21	51.69	15.140
July	122.78	121.88	0.067
August	124.71	124.72	0.000
September	103.85	119.75	21.056
October	48.03	65.70	26.044
November	23.86	25.71	0.287
December	14.60	14.31	0.007

Table 2-3: RMSE and MSE values for Kankai, Nepal

RMSE	8.046
MSE	64.741

These error tests constitute another method of determining the overall accuracy of the predicted dataset. The difference between these tests and the correlation tests is that the RMSE is a measure of unexplained variation, whereas correlation measures the positive association between the predicted and observed values.

A common measurement of error in hydrological models is the Nash-Sutcliffe model efficiency coefficient. The Nash-Sutcliffe efficiency (NSE) can be used to describe the accuracy of model outputs, and is shown in Equation 2-4.

$$E = 1 - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(X_i - \bar{Y}_i)^2} \quad (2\text{-}4)$$

Where X_i is the predicted flow, Y_i is the observed flow, and $\bar{Y}_i$ is the average of the observed flows (Nash & Sutcliffe, 1970). The NSE can range from negative infinity to +1, where +1

corresponds to a perfect match of predicted to observed flows. A NSE of zero shows that the predicted data is as accurate as the average of the observed data, while a negative NSE shows that the mean of the observed data is a better predictor than the modeled data. In other words, a NSE value less than zero means that a better prediction of the data would be found by simply taking the average of the observed data, rather than using the modeled data. Figure 2-10 shows a visual depiction of the NSE.

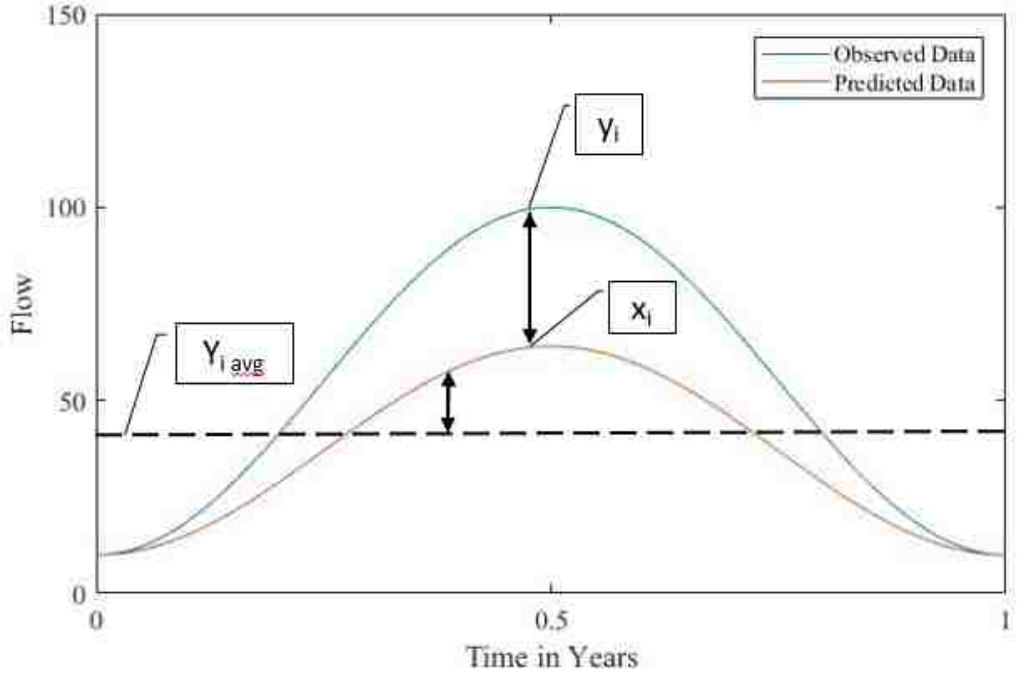

Figure 2-10: Visual depiction of NSE for observed and under predicted series.

This statistic works best when the coefficient of variation is large, and model bias produced by calibration is low (McCuen, Knight, & Cutter, 2006). It is possible to calculate a high efficiency value when the variance in the data is very high, even when the fit of the model is relatively poor (Jain & Sudheer, 2007). The efficiency coefficient is sensitive to outliers, so data

must be transformed to a normal distribution before testing. An example of the Nash-Sutcliffe Efficiency, as well as the observed and predicted flow, is shown in Table 2-4.

Table 2-4: Nash-Sutcliffe Efficiency for Kankai, Nepal

Month	Observed Flow	Predicted Flow	Variance	Nash-Sutcliffe Efficiency
January	11.73	10.48	0.120	-1.277
February	9.98	8.33	0.187	-2.071
March	9.24	6.16	0.238	-4.263
April	10.01	6.65	0.507	-3.202
May	15.12	15.88	0.895	-1.472
June	38.21	51.69	0.894	-0.195
July	122.78	121.88	0.821	-0.087
August	124.71	124.72	0.628	0.075
September	103.85	119.75	0.635	-0.465
October	48.03	65.70	0.500	-0.684
November	23.86	25.71	0.198	-0.517
December	14.60	14.31	0.115	-0.283
Year Total	27.37	26.82	0.529	0.590

From the Nash-Sutcliffe efficiency values, it can be determined that the predicted data is a satisfactory predictor of flow on a yearly basis. The station had a yearly Nash-Sutcliffe efficiency of 0.59, and a variance of 0.529. In other words, the MHS data is a satisfactory predictor on a yearly basis, however the monthly predictions are not as accurate as the observed data. The negative values for the monthly values are due to the subdivision of the data into months, changing the values of data that were being compared. The low values of variance for the station may also explain how the months are less accurate than the overall year values, as the NSE works best with datasets with large variances.

2.3 Population Metrics

Population metrics determine if the magnitude of data values are similar or different. The mean difference is a population metric that measures the difference in magnitude relative to the mean of the data. The mean difference is generated by the paired t-test as the geometric mean difference between the transformed predicted data and transformed observed data. The equation for the mean difference is shown below in Equation 2-5.

$$\bar{X}_{geom} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 * x_2 * \ldots * x_n} \quad (2\text{-}5)$$

Where x is the difference between the observed streamflow and the predicted streamflow value. Each mean difference metric must then by transformed from the natural log value to the actual value by using the exponential transform. Figure 2-11 shows a visual depiction of the mean difference for observed and under predicted datasets.

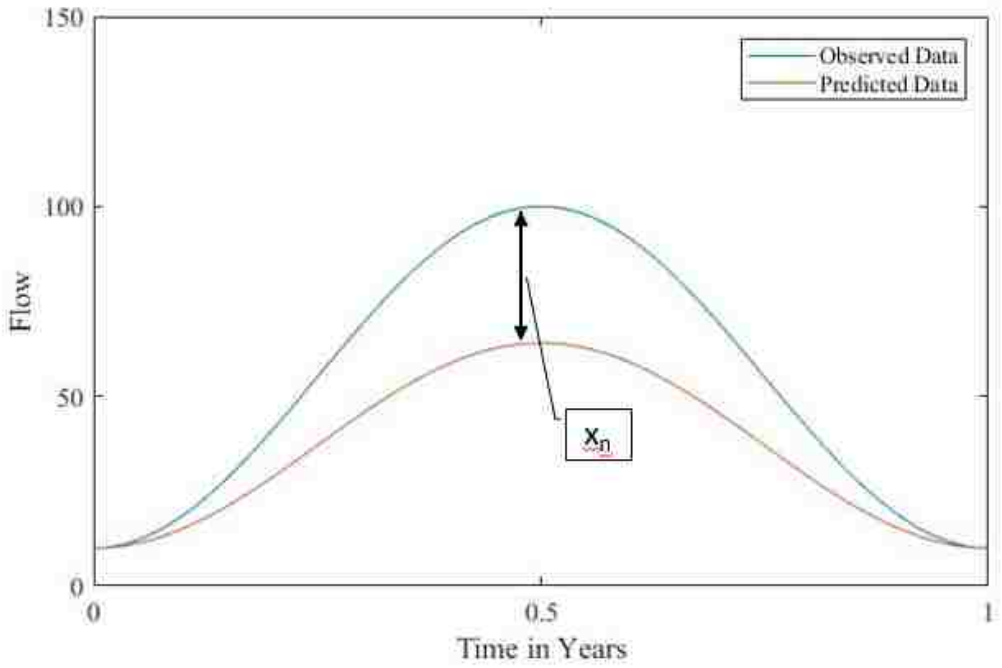

Figure 2-11: Visual depiction of mean difference.

The mean difference of the geometric mean emphasizes any large values within the datasets. For hydrological data, this is optimal as researchers generally are more interested in large events and magnitudes rather than base flow values. The mean difference is not affected by any timing shifts within the data, but is affected by any subdivision of the data into months or seasons, as this changes the populations being compared.

The variance on the mean difference can also be calculated, and used as a standard of comparison. The variance is an estimate of the deviation of a variable from its mean. In other words, the variance measures the spread of the variable from the average value.

2.4 Timing Tests

Another aspect of analyzing the accuracy of the MHS data is determining how well the MHS data captures time events, such as high flow months and days. This analysis was conducted in two parts, by calculating the R^2 coefficient and the spectral angle while changing the lag time for the predicted dataset in case there were timing errors associated with the input or routing of the flows through the watershed. This analysis can be done with any error metric, however the R^2 and spectral angle coefficient were chosen to specifically analyze the shape of the hydrograph with respect to timing. To generate the lagged time series, additional points were interpolated for the datasets between the initial daily values using a cubic spline. These interpolated points created 6-hour time steps for the MHS and observed datasets, which can then be used as lag times to determine the best lag time for the predicted data.

2.4.1 R^2 Analysis

The R^2 coefficient, or the coefficient of determination, is a measure of the correlation between the original or observed data and the modeled values (Everitt, 2002). The R^2 value

ranges between 0 and +1, where +1 signifies a perfect 1:1 correlation between the datasets, and zero signifies absolutely no correlation. By lagging the predicted dataset by a time step of six hours, and then calculating the R^2 value, the optimal lag time can be predicted.

The optimal lag time helps determine the accuracy of the predicted dataset in simulating events such as storms and dry seasons. The R^2 analysis determines the absolute difference between each data point. This helps researchers understand the overall accuracy of the predicted dataset. Figure 2-12 shows the trend in R^2 values as the time series is lagged both forward and backwards 15 time steps.

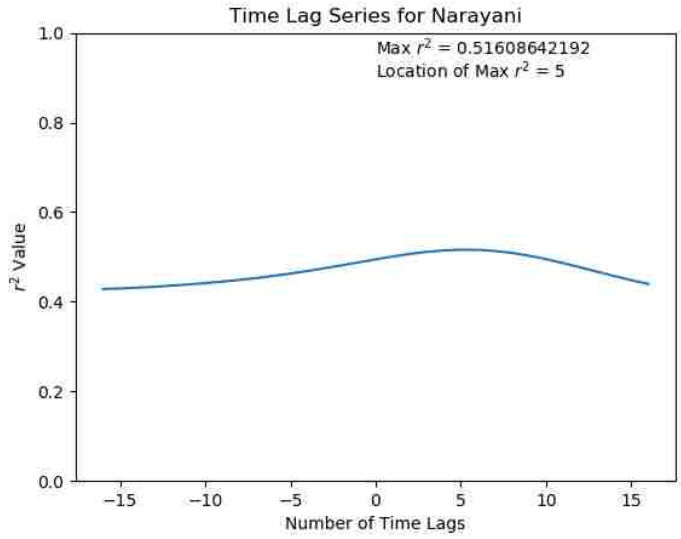

Figure 2-12: R^2 comparison with time lag for Narayani, Nepal.

This graph shows the trend in the R^2 values as the time series is lagged. The maximum R^2 value occurs when the time series is lagged forward five time steps, or 30 hours.

The number of time lags indicates the original accuracy of the predicted data set, as well as the necessary time shift for a maximum R^2 value.

2.4.2 Spectral Angle Metric

The spectral angle metric is most often used in applications that deal with wave functions, as the analysis determines the accuracy of the shape, not the magnitude of the waves. For hydrographs and extreme events such as floods, this statistical analysis is important to gauge if the simulated data is able to catch the event itself. If a hydrograph has a similar shape, even if the magnitude is not captured perfectly, we can say that the event is being modeled. For example, if a modeled hydrograph can consistently predict the shape of events, resources can then be allocated to prepare for large events. The process of calculating the spectral angle relies on measuring the similarities between the different series, and producing the spectral angle coefficient, which measures the actual angle between the two vectors. The equation for this the coefficient is shown in Equation 2-6.

$$Spectral\ Angle(x,y) = cos^{-1}\left(\frac{(x \cdot y)}{\|x\|_2\|y\|_2}\right) \quad (2\text{-}6)$$

Or the arccosine of the dot product of x and y (Robila & Gershman, 2005). The spectral angle coefficient ranges from zero to +1, where +1 corresponds to a perfect match in shape of the datasets. Figure 2-13 shows an example of the spectral angle coefficient for the Narayani station in Nepal corresponding to different numbers of time lags.

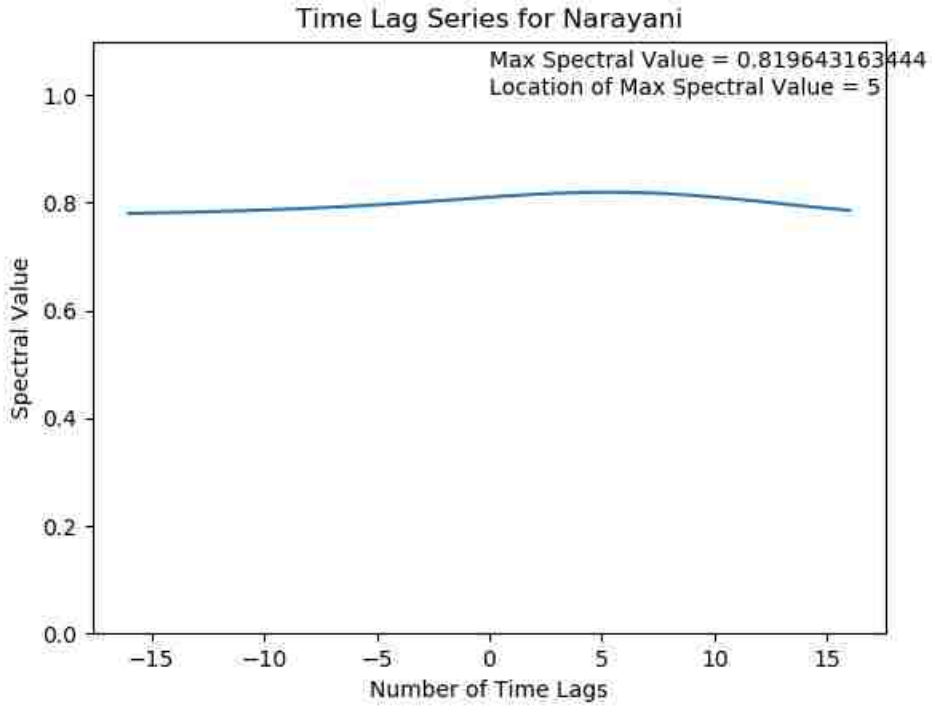

Figure 2-13: Spectral Angle Coefficient comparison with time lag for Narayani, Nepal.

In this case, the spectral angle coefficient increased with a time lag of 30 hours. This indicates that the correlation between the observed data and the MHS data could be improved by lagging the data by 30 hours.

2.5 Generalized Application

To define the tests best suited for analyzing different types of events and error, the suite of statistical tests was first tested on a series of synthetic hydrographs describing trends like under and over-prediction, timing shifts, outliers, and combinations of each. An example of the pertinent results from two of these tests, as well as the optimum results, is summarized in Table 2-5.

Table 2-5: Select Metrics for Synthetic Hydrograph

Metric	Under predict by 5%	Over predict by 5%	Perfect Match
Mean Error	-2.2498	2.2498	0
Mean Absolute Error	2.2498	2.2498	0
Mean Squared Error	7.5931	7.5931	0
Root Mean Square Error	2.7556	2.7556	0
Root Mean Squared Log Error	0.0369	0.0354	0
Mean Absolute Scaled Error	45.6208	45.6208	0
R^2	1.0000	1.0000	1
Anomaly Correlation Coefficient	1.0000	1.0000	1
Mean Absolute Percentage Error	3.4185	3.4185	0
Nash-Sutcliffe Efficiency	0.9925	0.9925	1
Modified Nash-Sutcliffe Efficiency	0.9215	0.9215	1
Relative Nash-Sutcliffe Efficiency	0.9960	0.9960	1
Spectral Angle	0.9999	0.9999	1

From this example, researchers can compare the calculated metric with the optimal value, to determine the accuracy of the synthetic data based on the metric. For the synthetic hydrograph shown above (under predicted by a specific scale) the R^2 value, as well as the correlation coefficient for both the under and over predicted hydrograph were determined as 1.0, showing that the shape of the hydrograph was identical to the original hydrograph. The error tests for the hydrograph showed variations in error, with an RMSE of 2.756, and a RMSLE of 0.0369 for the under predicted hydrograph and an RMSE of 2.756 and a RMSLE of 0.0354 for the over predicted hydrograph.

Once the tests had been run on each series, the best statistical tests for each trend was summarized in a table. The results for the different metrics varied with the different synthetic hydrographs, which made it possible to determine the best metrics for different hydrograph scenarios. This table can be found in the results chapter of this thesis.

2.6 Global Application

As part of this analysis, both a MATLAB and Python library were created that could perform the statistical analysis for any given station, provided that sufficient observed data could be obtained at the station. The MHS data was obtained through the SPT, and then merged with the observed data in a .csv file. The Python library was then used to conduct the paired t-test and correlation tests, the error tests, the R^2 and spectral angle coefficient analysis, and generated graphs of the average predicted and observed stream flows by month.

The MHS data was compared to observed historical data at sites within the Dominican Republic, Nepal and East Africa to give both a nation-wide and world-wide comparison between the datasets. The statistical analysis was conducted monthly to determine the accuracy of the MHS data. The correlation values between the predicted and observed flows, as well as the mean difference were then tabulated and analyzed to determine the overall correlation between the two datasets. The timing tests and error tests were also conducted on each nations' dataset to determine the overall accuracy of the MHS data compared with observed data.

3 RESULTS

The analysis was conducted in the Dominican Republic, Nepal, and Tanzania. The results are summarized by type as well as by country. This allowed the results to be analyzed at different resolutions; both nation-wide and world-wide level. Figure 3-1, Figure 3-2, and Figure 3-3 show the spread of stations across the different countries.

Figure 3-1: Observed streamflow stations in Nepal.

Figure 3-2: Observed streamflow stations in Tanzania.

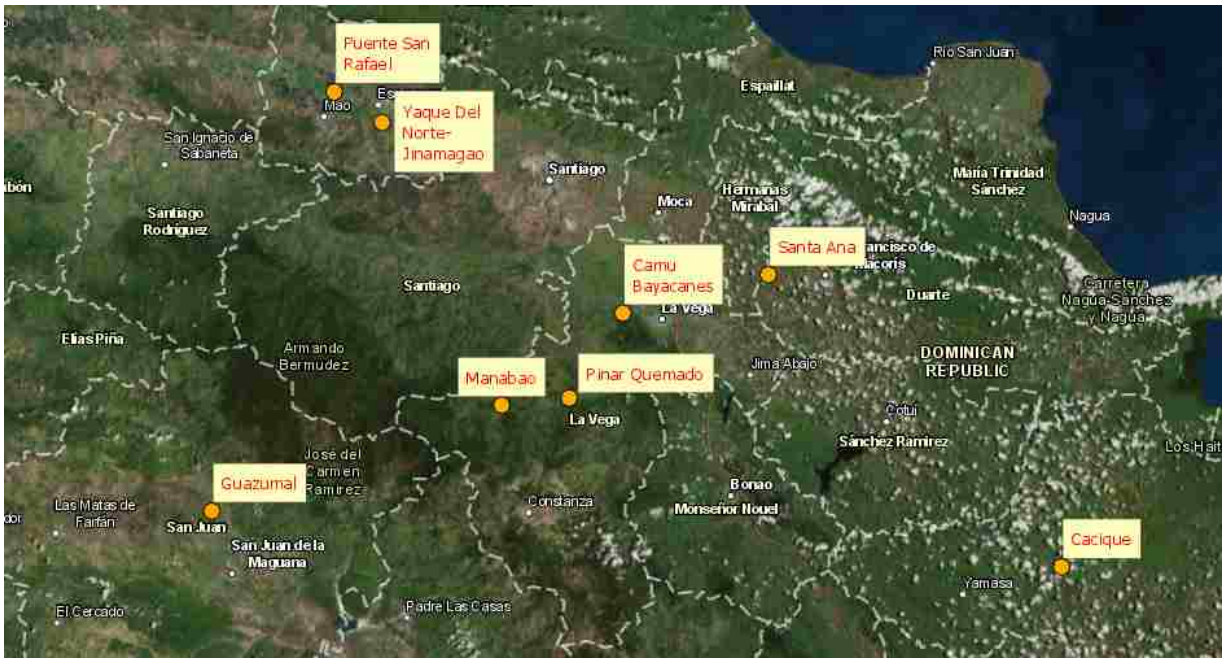

Figure 3-3: Observed streamflow stations in the Dominican Republic.

The stations were chosen by the respective water management agencies within the countries. As the stations are spread across the countries, the results of the statistical analysis can be applied throughout the country. This spread of observed data also validates the findings of the statistical analysis by determining any trends that exist across the country, rather than trends that are centralized at specific locations.

3.1 Distribution and Population Metrics

The correlation coefficient values and mean difference values for each country can be found in Appendix A. Correlation coefficient values for each country are shown in Table 0-1, Table 0-2, and Table 0-3. The mean difference values are shown in Table 0-4, Table 0-5, and Table 0-6. The correlation coefficients and mean difference values are summarized by station and by month.

From the correlation and mean difference tables, both the correlation coefficients and mean differences for each country can be determined at a monthly and overall level. The different countries had different trends in average correlation coefficient and mean difference corresponding to high flow seasons and low flow seasons. Using a statistical method known as a Student's t-test, the correlation coefficients and mean difference values were compared by month to see any trends. The Student's t-test compares the differences in means to determine where significant differences occur. The correlation coefficients for all three countries increased during high flow seasons, and decreased during low flow seasons. The error in mean difference values for the three countries followed similar trends. For Tanzania, the mean difference value decreased during the high flow seasons, and increased during low flow seasons, but were closest to one during the high flow seasons. Nepal did not show an appreciable difference in the mean

difference between the different months. Table 3-1 summarizes the correlation statistics for the three countries by month.

Table 3-1: Statistical Summary by Country

	Correlation Coefficient	Mean Difference	Observed Flow	Predicted Flow	Variance
Dominican Republic					
January	0.4749	5.0277	9.5761	2.2071	0.6095
February	0.3162	4.6258	8.9512	2.2047	0.6893
March	0.2526	4.0193	10.0038	2.3280	0.7961
April	0.4394	2.7016	11.5027	4.1789	0.9772
May	0.6120	2.1101	21.0807	11.2307	1.1555
June	0.4088	3.2992	25.0217	6.2043	0.7772
July	0.4583	3.5605	12.2170	3.2653	0.5314
August	0.4826	2.9119	10.1645	3.4701	0.5494
September	0.4328	2.8910	12.1085	4.2783	0.6447
October	0.4434	3.4898	15.3880	4.1328	0.7910
November	0.4387	4.3554	14.6662	3.5718	0.7646
December	0.5765	5.8011	10.7269	2.2306	0.6657
Average	0.4447	3.7328	13.4506	4.1085	0.7460
Nepal					
January	0.3318	1.6518	127.3038	83.2350	0.4692
February	0.3612	1.5274	110.9548	78.0137	0.5028
March	0.2896	1.5757	108.4400	85.2476	0.5538
April	0.3278	1.6316	132.8855	130.8335	0.6776
May	0.3591	1.9959	226.5781	220.3405	0.8475
June	0.6048	2.2649	489.7523	284.2585	1.0511
July	0.4732	1.9014	1239.9085	740.7930	0.6755
August	0.3388	1.8960	1541.8047	941.7744	0.5533
September	0.5160	1.6160	1050.8243	728.8233	0.4591
October	0.5392	1.7700	475.1222	316.7528	0.4524
November	0.4558	1.6578	246.2587	155.9685	0.4983
December	0.3603	1.5666	163.0488	103.1103	0.8351
Average	0.4131	1.7517	492.7401	322.4292	0.6313
Tanzania					
January	0.4537	0.9391	9.4184	17.3918	1.6794
February	0.4394	0.9193	13.1232	24.5919	1.6435
March	0.3487	0.8617	17.2920	30.3960	0.9963
April	0.3196	1.1688	19.6769	31.5737	0.8837
May	0.3896	2.1778	11.8760	15.1235	0.7152

Table 3-1: Statistical Summary by Country (cont.)

	Correlation Coefficient	Mean Difference	Observed Flow	Predicted Flow	Variance
June	0.4099	2.7092	6.8937	8.1914	0.5179
July	0.3452	3.0951	4.7592	4.9940	0.8430
August	0.3923	3.5907	3.6117	3.1539	0.7695
September	0.3357	4.2366	2.6998	2.0466	0.6262
October	0.2637	3.8218	2.0630	1.4828	0.9618
November	0.3534	2.7423	1.9614	1.9125	2.0161
December	0.5556	1.3339	4.3820	8.4403	2.1012
Average	0.3839	2.2997	8.1465	12.4415	1.1461
Grand Total	**0.4101**	**2.4561**	**191.1951**	**126.6037**	**0.8530**

Nepal showed an overall correlation coefficient of 0.4131, Tanzania showed an overall correlation coefficient of 0.3839, and the Dominican Republic showed an overall correlation coefficient of 0.4447.

The mean difference values varied between months and between countries. In Nepal, the months with the mean difference closest to one were the months of neither high nor low traditional flows. In Tanzania, the months with the mean difference closest to one were the months of April, January and February. These months are near the end of the high flow season. In the Dominican Republic, the smallest mean difference values occurred during the months of April and May, which mark the beginnings of the high flow season.

From this analysis, we can see that worldwide the modeled historical streamflow (MHS) data is positively correlated with the observed values with an average correlation coefficient of 0.41. The mean difference values for all three countries show that, on average, the MHS is under predicting the actual flow by a factor of 2.46. This means that the MHS data is modeling around half of the actual flow on a global scale. Some stations, such as Chimala in Tanzania show the opposite result, that the MHS data is over predicting, sometimes twice as much as the observed

flow. This over-prediction by the MHS, while not standard throughout the world, could be a result of dams on the river reaches, something that the RAPID routing procedure does not account for. In the case of Chimala, this discrepancy is a result of the MHS data not being available at the exact site of the observed data, but rather further downstream. These over-predicted streams must then be calibrated in the RAPID model, in order to correctly estimate the statistical tests for the stream reaches.

3.2 Error Tests

The error analysis for the different stations show the relationship between the different error metrics and the variance of the mean difference. The RMSE and the RMSLE were both low for the three countries, while the Nash-Sutcliffe efficiency was found to be negative for all three countries on average. These negative values for the Nash-Sutcliffe efficiency could be due to the small variance in the data for all three countries, but also indicate that there are errors in the model. Table 3-2 summarizes the error analysis for Nepal.

Table 3-2: Error Summary for Nepal

Station	Variance	RMSE	RMSLE	Nash-Sutcliffe Efficiency
Asaraghat	0.2817	7.2069	0.0143	-5.0404
Babai	0.3800	2.6620	0.0589	-3.5777
Bheri	0.2636	11.2160	0.0450	-31.1788
Kaligandaki	0.2996	8.7285	0.0274	-9.8843
Kamali	0.2433	18.4823	0.0152	-2.9693
Kankai	0.4782	1.4437	0.0179	-1.2034
Marsyangdi	0.3848	4.5470	0.0298	-18.6078
Narayani	0.2311	24.2055	0.0209	-7.4208
Rapti	0.3445	3.5240	0.0300	-1.4779
Saptakosi	0.1435	17.3572	0.0107	-1.6789
Seti	0.3652	13.8719	0.1050	-18.8570
Tinaukhola	4.1601	7.3230	0.0634	-0.2429
Total Average	**0.6313**	**10.0473**	**0.0365**	**-8.5116**

The average RMSE for Nepal was 10.0473, with a RMSLE of 0.0365. The Nash-Sutcliffe efficiency for Nepal was -8.5116; implying that variance in the mean difference is smaller than the variance of the observed data. In other words, when the Nash-Sutcliffe efficiency is negative, we can conclude that there are large differences between the predicted and observed flows. Table 3-3 summarizes the error statistics for the Dominican Republic.

Table 3-3: Error Analysis for the Dominican Republic

Station	Variance	RMSE	RMSLE	Nash-Sutcliffe Efficiency
Cacique	0.5191	0.0869	0.0283	-0.0792
Camu Bayacanes	1.2695	0.1682	0.0608	-1.9039
Cenovi	0.5384	1.4096	0.4298	-15.7397
Guazumal	1.1852	1.1853	0.0618	-1.9821
Jinamagao	0.2970	1.3047	0.4147	-28.1913
Manabao	0.2713	1.9932	0.4286	-22.1422
Pinar Quemado	0.4900	19.2785	0.4149	-3.6382
Puente San Rafael	1.3974	0.0538	0.0577	-0.1401
Santa Ana	0.5191	0.0869	0.0283	-0.0792
Cacique	1.2695	0.1682	0.0608	-1.9039
Camu Bayacanes	0.5384	1.4096	0.4298	-15.7397
Cenovi	1.1852	1.1853	0.0618	-1.9821
Total Average	**0.7460**	**3.1850**	**0.2371**	**-9.2271**

This table shows the same trends as in Nepal for the different error statistics. The RMSE and RMSLE were both quite low, with a negative efficiency coefficient. The difference between the RMSE and RMSLE between Nepal and the Dominican Republic is due to the difference in flow between the two regions. To compare these metrics between the different regions, they must first be normalized by maximum flow. Table 3-4 summarizes the error statistics for Tanzania.

Table 3-4: Error Analysis for Tanzania

Station	Variance	RMSE	RMSLE	Nash-Sutcliffe Efficiency
Chimala at Chitakelo	1.2683	0.2013	0.0825	-3.9336
Igawa	0.2660	0.4686	0.0277	-2.8317
Ihimbu	0.7005	0.6483	0.0569	-8.0842
Ilongo	0.9786	0.2263	0.0490	-2.1382
Ipatagwa	1.5920	0.1303	0.0882	-3.7825
Kimani	0.3849	0.1993	0.0275	-1.1565
Mawande	0.7004	0.5236	0.0387	-1.1329
Msembe	2.6170	3.7412	0.1614	-3.7779
Mswisi	2.5617	0.1422	0.1024	-3.8806
Mtandika	1.0073	0.8285	0.0982	-61.5474
Mtitu	1.0860	0.1370	0.0871	-51.3613
Ndiuka	0.5912	0.5697	0.0511	-9.8578
Total Average	**1.1461**	**0.6513**	**0.0726**	**-12.7904**

The Tanzania stations showed similar values for the error statistics as those in the Dominican Republic and Nepal. The RMSE and RMSLE were the lowest in Tanzania, while the Nash-Sutcliffe efficiency was the most negative. The overall negative values of the Nash-Sutcliffe efficiency indicate that there are differences between the predicted and observed values. However, the efficiency coefficient is sensitive to extreme values, and works best if the variance in the data is large. The low variance for the stations may explain the negative efficiency coefficients.

From these statistical analysis, the RMSE and RMSLE shown to be the most accurate tests for this data. The Nash-Sutcliffe efficiency does not perform as well for this dataset, due to the small variance in the data. Both the RMSE and RMSLE can be used to determine the overall error between the MHS data and the observed data.

3.3 Timing Tests

The R^2 analysis and spectral angle coefficient analysis for Nepal and Tanzania showed no overall trends for the countries with respect to time lag. Table 3-5 summarizes the maximum R^2 coefficient, the maximum spectral angle coefficient and the respective time lags for the maximum coefficients for each station in Nepal. Each time lag is equal to a period of six hours.

Table 3-5: R^2 and Spectral Angle Analysis for Nepal

Station	R^2 Coefficient	Number of Time Lags (R^2)	Spectral Angle Coefficient	Number of Time Lags (Spectral Angle)
Asaraghat	0.4710	3	0.8215	3
Babai	0.1387	5	0.4888	5
Bheri	0.5422	10	0.8301	10
Kaligandaki	0.5134	6	0.8114	6
Kamali	0.5703	10	0.8534	10
Kankai	0.2153	2	0.5483	2
Marsyangdi	0.3638	1	0.7308	1
Narayani	0.5161	5	0.8196	5
Rapti	0.2261	7	0.5686	7
Saptakosi	0.6340	9	0.8845	9
Seti	0.4935	2	0.7892	2
Tinaukhola	0.0199	16	0.2418	16
Total Average	**0.3920**	**6.3333**	**0.6990**	**6.3333**

The R^2 coefficient for the Nepal stations ranged from 0.0199 to 0.6340, with an average value of 0.3920. Each station's R^2 increased as the time series was shifted, with an average time shift of 6.333 lag steps, or 38 hours. The spectral angle coefficient for Nepal stations ranged from 0.2418 to 0.8845, with an average value of 0.6990. The time steps for the R^2 values match those for the spectral angle coefficients.

Table 3-6 summarizes the same analysis for the stations in the Dominican Republic where sufficient data was provided.

Table 3-6: R^2 and Spectral Angle Coefficients for the Dominican Republic

Station	R^2 Coefficient	Number of Time Lags (R^2)	Spectral Angle Coefficient	Number of Time Lags (Spectral Angle)
Cacique	0.0882	4	0.4774	4
Camu Bayacanes	0.1452	-3	0.4737	-3
Cenovi	0.0910	-2	0.5885	-2
Guazumal	0.1487	-1	0.6627	-1
Jinamagao	0.1200	2	0.5017	2
Manabao	0.2485	-1	0.7574	-1
Pinar Quemado	0.4384	0	0.8274	0
Puente San Rafael	0.2776	-1	0.7308	0
Santa Ana	0.0156	0	0.2262	0
Cacique	0.0882	4	0.4774	4
Camu Bayacanes	0.1452	-3	0.4737	-3
Cenovi	0.0910	-2	0.5885	-2
Total Average	**0.1748**	**-0.2222**	**0.5829**	**-0.1111**

For the stations in the Dominican Republic, the R^2 values were significantly lower than for Nepal. The R^2 values ranged from 0.0156 to 0.43842, with an average value of 0.1748. Five of the nine stations were shifted, one being back-shifted 3 time steps, or 18 hours. This backwards shift implies that the data for Camu Bayacanes is predicting flow events after they actually happen. Positive lag times signify that the MHS data is predicting events before they actually happen. The maximum spectral angle coefficient for the Dominican Republic ranged from 0.2262 to 0.8274, with an average value of 0.5829. This is lower than the maximum values for Nepal.

Table 3-7 summarizes the same analysis for Tanzania, and shows similar results to the earlier tables for Nepal and the Dominican Republic.

Table 3-7: R^2 and Spectral Angle Coefficients for Tanzania

Station	R^2 Coefficient	Number of Time Lags (R^2)	Spectral Angle Coefficient	Number of Time Lags (Spectral Angle)
Chimala at Chitakelo	0.0692	0	0.5440	0
Igawa	0.1826	0	0.5448	6
Ihimbu	0.1546	0	0.5735	0
Ilongo	0.1961	0	0.5880	11
Ipatagwa	0.2278	11	0.5914	0
Kimani	0.1160	0	0.6184	0
Mawande	0.2579	0	0.6537	0
Msembe	0.2472	0	0.6589	0
Mswisi	0.3376	0	0.6794	0
Mtandika	0.3179	0	0.6900	0
Mtitu	0.3925	0	0.7365	0
Ndiuka	0.3853	6	0.7735	0
Total Average	**0.2404**	**1.4167**	**0.6377**	**1.4167**

The R^2 values for the Tanzania stations were slightly lower than those from Nepal, but higher than those from the Dominican Republic. The R^2 values ranged from 0.0692 to 0.3925, with an average value of 0.2404. Only two of the twelve stations required any shift for a maximum R^2 coefficient value or maximum spectral angle coefficient, with the maximum of 11 time steps, or 66 hours. The spectral angle coefficients for the Tanzania stations were slightly higher than those from Nepal, and significantly higher than those from the Dominican Republic. The spectral angle coefficient values ranged from 0.5440 to 0.7735, with an average value of 0.6377. The comparison between the Tanzania, Dominican Republic, and Nepal R^2 values is shown in Figure 3-4.

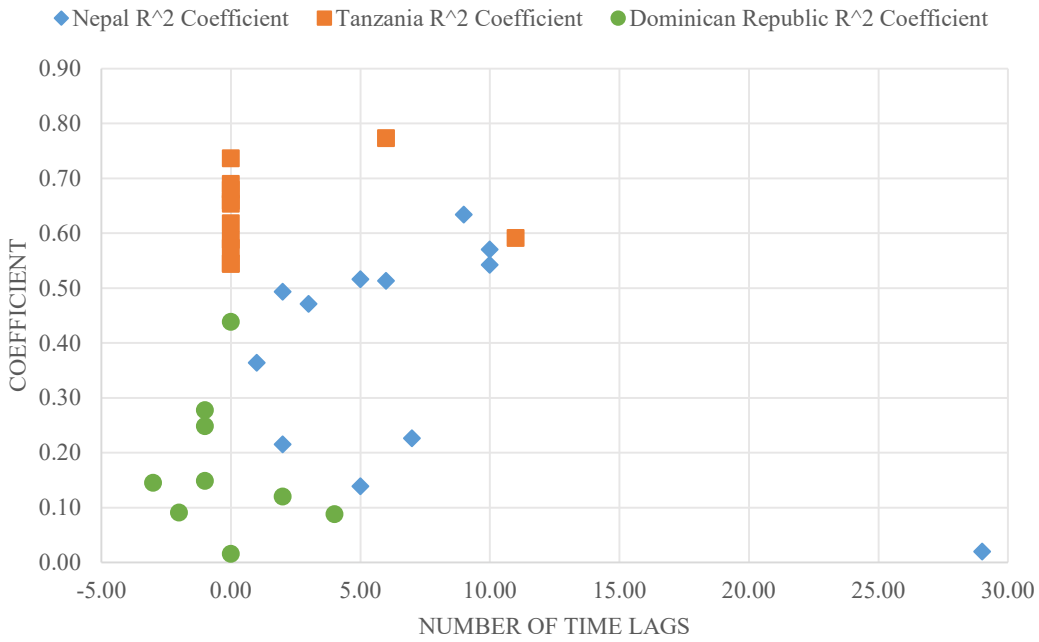

Figure 3-4: Comparison of R^2 values for the Dominican Republic, Nepal and Tanzania.

This graph better illustrates the general spread of R^2 values for Nepal, and the close clumping of R^2 values for the Tanzania stations. The low values for the Dominican Republic, and the spread of the time lags for the three regions are also shown here. Nepal showed the largest range of time lags, ranging from one time lag to 30, or a time shift of six hours to several days, while the maximum values of R^2 values for Tanzania mainly occurred at zero time lag, indicating no time shift was necessary in order to calculate the maximum R^2 value. Figure 3-5 below summarizes the spread and similarity between the spectral angle coefficients for the Dominican Republic, Nepal and Tanzania.

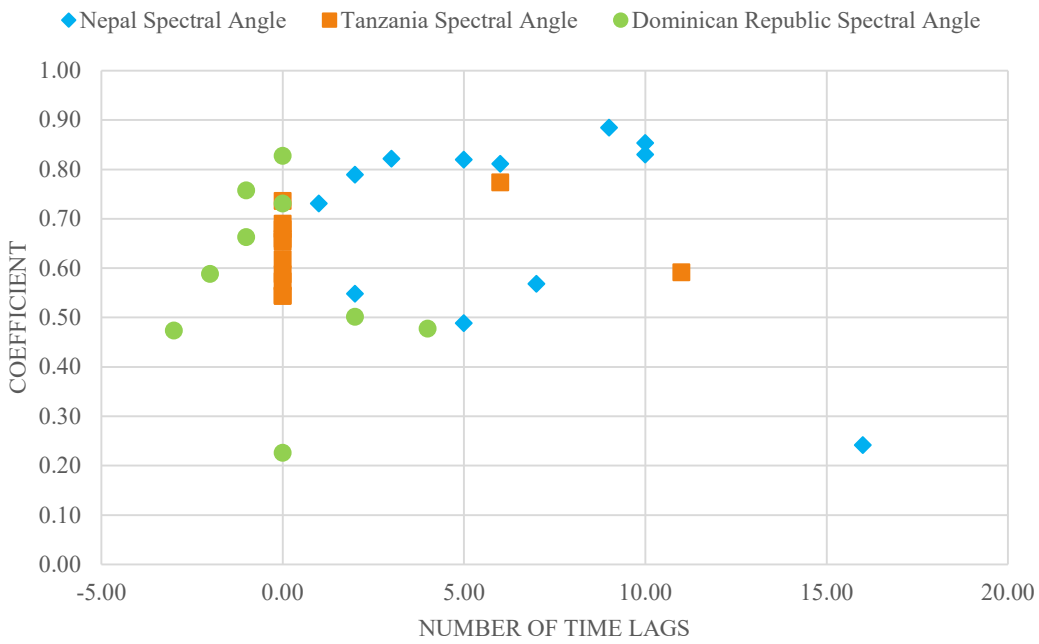

Figure 3-5: Spectral angle coefficients for the Dominican Republic, Nepal and Tanzania.

This graph illustrates the similarity of spectral angle coefficients for the stations in Nepal and Tanzania, as well as the low values for the Dominican Republic. Both Nepal and Tanzania showed few outliers, with the majority of the maximums occurring with minimal time lag correction. However, the maximum values for the spectral angle coefficients occurred with some time lag in Nepal, with a few outliers around 10 or 16 time steps. The negative values of time steps in the Dominican Republic indicate that the model is too slow, and must be sped up for optimal spectral angle coefficients. Figure 3-6 shows the relationship between the R^2 coefficient and the spectral angle coefficient for both countries.

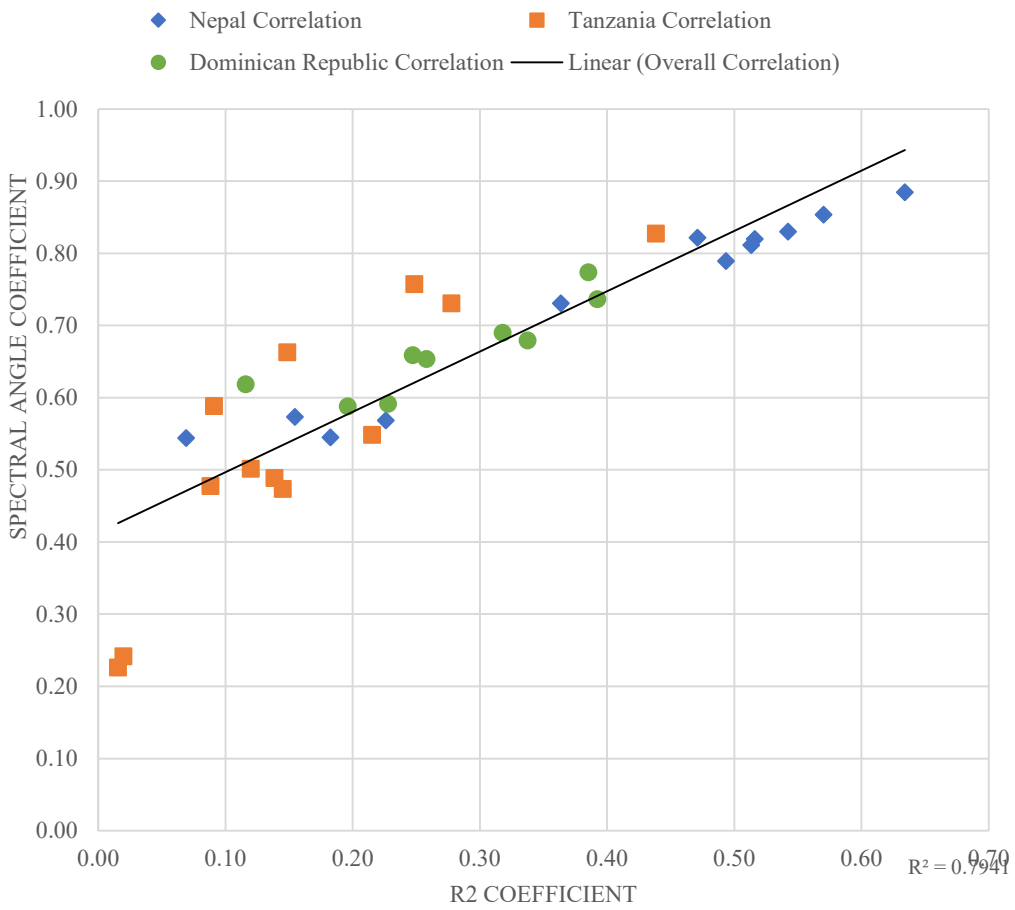

Figure 3-6: R^2 and spectral angle coefficient comparison.

Figure 3-6 defines the positive correlation between the R^2 values and the spectral angle coefficients. Generally, as the R^2 values increase, so do the spectral angle coefficients. The R^2 value for the correlation is 0.7941, or a very high correlation between the two coefficients. The correlation between the spectral angle and R^2 values for Tanzania is a little higher than average, however the overall values of both the R^2 values and the spectral angle coefficients imply that the MHS data is predicting the shape of flow events well over the course of the 35-year period, if not predicting the exact magnitude. The higher spectral angle coefficients support this conclusion as the spectral angle measures the overall direction of the data, rather than the absolute difference between data points.

3.4 MHS Analysis Results

Table 3-8 summarizes the error metrics for each of the three countries, as well as the perfect measure of error. The RMSE and RMSLE were both normalized by dividing by the observed flow to allow the metrics to be compared between countries.

Table 3-8: Overall Summary of Error

	Nepal	Dominican Republic	Tanzania	Average Results	Perfect
Observed Flow	492.74	13.45	191.20	232.46	-
Predicted Flow	322.43	4.11	126.60	151.05	-
RMSE	10.05	3.19	0.65	4.63	0.00
RMSLE	0.04	0.24	0.07	0.12	0.00
Normalized RMSE	2%	24%	0.3%	9%	0%
Normalized RMSLE	0.2%	9.4%	0.6%	3%	0%
NSE	-8.51	-9.23	-12.79	-10.18	1.00
Correlation Coefficient	0.41	0.45	0.41	0.42	1.00
Mean Difference	1.75	3.73	2.46	2.65	1.00
R2	0.39	0.17	0.24	0.27	1.00
Time Lag Difference	37.98	-1.33	8.50	15.05	0.00
Spectral Angle	0.70	0.58	0.64	0.64	1.00
Time Lag Difference	38.00	-0.67	8.50	15.28	0.00

From these overall values, the errors were normalized and then color-coded to determine the overall performance of the MHS data in the three countries. The red cells indicate the metrics with the highest normalized error, while the green cells indicate the metrics with low normalized error. These values are summarized in Table 3-9.

Table 3-9: Color-Coded Error Summary

	Nepal	Dominican Republic	Tanzania
Observed Flow	492.74	13.45	191.20
Predicted Flow	322.43	4.11	126.60
Normalized RMSE	0.09	1.00	0.01
Normalized RMSLE	0.02	1.00	0.06
NSE	0.69	0.74	1.00
Correlation Coefficient	0.99	0.94	1.00
Mean Difference	0.27	1.00	0.53
R^2	0.74	1.00	0.92
Time Lag Difference	1.00	0.04	0.22
Spectral Angle	0.72	1.00	0.87
Time Lag Difference	1.00	0.02	0.22

From Table 3-9, we can see that the Dominican Republic had the most error, while Tanzania had very low error. This error summary can be used to determine the overall accuracy of the MHS data, as well as help researchers determine the type of errors present in the data for the specific regions. The high error for the time lag difference, as well as the high error for correlation coefficient indicate that timing errors exist for the Nepal stations. The high errors for RMSE and RMSLE for the Dominican Republic, paired with high errors in the R^2 and spectral angle coefficients indicate that the magnitude and timing of events need to be calibrated for those stations. The overall low error for Tanzania indicate that the MHS data accurately predicted both overall events, as well as correct magnitude and timing of individual events.

3.5 Generalized Application Results

By calculating the metrics for each of the synthetic hydrographs, trends in the results and the metrics themselves can be determined. Table 0-7 in Appendix A. shows the summary for the different hydrographs, as well as the resulting error for the metric normalized by the maximum

error for the metric. A value of 1.0 signifies that the hydrograph had the largest error for that metric, while a value of zero signifies that there was no error for that metric. The results are also color-coded from low error (red) to high error (green) to better identify trends in both the metrics and hydrographs. The green cells indicate which metrics calculated significant error for each hydrograph, while the red cells indicate the metrics that calculated very little error for the different hydrographs. From this table, it can be determined that a hydrograph that is over predicted, with shift, noise and outliers has the largest error metrics, while hydrographs that are simply over or under predicted have the smallest error metrics. The R^2 metric and the correlation coefficient had the largest variance in errors, ranging from no error in the hydrographs that were only scaled to an error of 0.032 in the hydrograph with over prediction, shift, outliers, and noise.

The hydrographs without timing shifts had lower total errors than the hydrographs with timing shifts. Even the hydrographs with over or under prediction, noise, and outliers had less total error than the hydrographs with only a timing shift. The exception to this was the mean error metric, which showed the smallest error for the hydrograph with over prediction, time shift, noise and outliers, and the largest error for the hydrographs that were under predicted with noise and outliers.

From these results, specific metrics can be seen to be the most useful for different types of hydrographs. For hydrographs without timing shifts, all the metrics showed low error. Specifically, the R^2 metric and the correlation coefficient showed very low error. For hydrographs with timing shifts, the spectral angle metric showed the lowest error. The Nash-Sutcliffe efficiency also showed low error for hydrographs with timing shifts.

In conclusion, hydrographs without timing errors can be described using any of the metrics discussed here. The spectral angle metric is best used with hydrographs with timing

errors, but overall performance is similar for all hydrographs. The Nash-Sutcliffe Efficiency and correlation coefficient are also appropriate for all hydrographs. The R^2 metric is best used with hydrographs without any timing shifts but can be used with hydrographs with outliers and noise as well.

3.6 MHS Analysis Package

The Python library created for this analysis performs the statistical analysis for any given station, based on the SPT data and the observed data at the station. The merged .csv file is used as an input for the Python library, and a .csv file with the results for the station is given as an output. The Python library conducts the paired t-test and correlation tests, error tests, lag analysis, as well as generates a graph of the average predicted and observed stream flows by month.

The graphical representation of the observed and predicted flows by month, such as the comparison shown in Figure 3-7 gives a first impression of the accuracy of the data on a monthly basis.

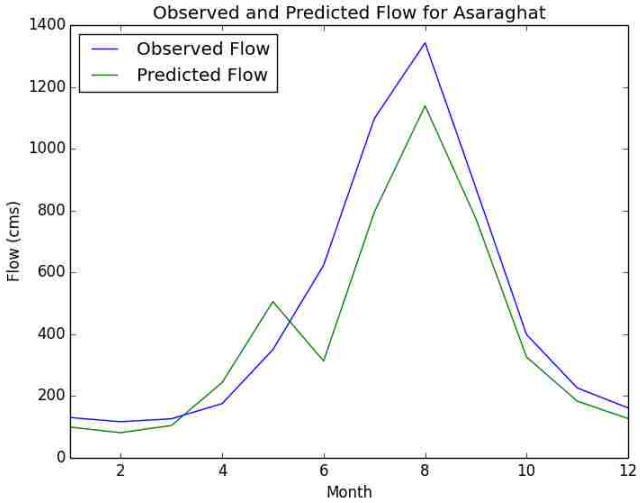

Figure 3-7: Example of visual representation of flows generated by Python script.

The statistical analysis library also generates a summary file for all stations analyzed, where researchers can inspect specific statistics and stations to better understand the accuracy of the MHS data.

4 STATISTICAL ANALYSIS CASE STUDY

Appropriately using statistical analysis to compare two datasets is a challenging topic for those who have not exclusively studied statistics. In this case study, the modeled historic streamflow (MHS) data for three stations chosen at random in Nepal was analyzed to determine the validity of using the MHS data as a surrogate for observed data for the region. The stations chosen were Saptakosi, Bheri, and Marsyangdi. The three stations are shown in Figure 4-1.

Figure 4-1: Station location in Nepal.

The data was analyzed first using visual analysis, then using error metrics associated with distribution and population differences. The results were then summarized to determine if the MHS data could be used as a surrogate for observed data where observed data does not exist in Nepal.

4.1 Visual Analysis

The first step in statistical analysis is always to plot the data and analyze it visually. Figure 4-2 shows the predicted and observed flows for Saptakosi for the entire 35-year historical period.

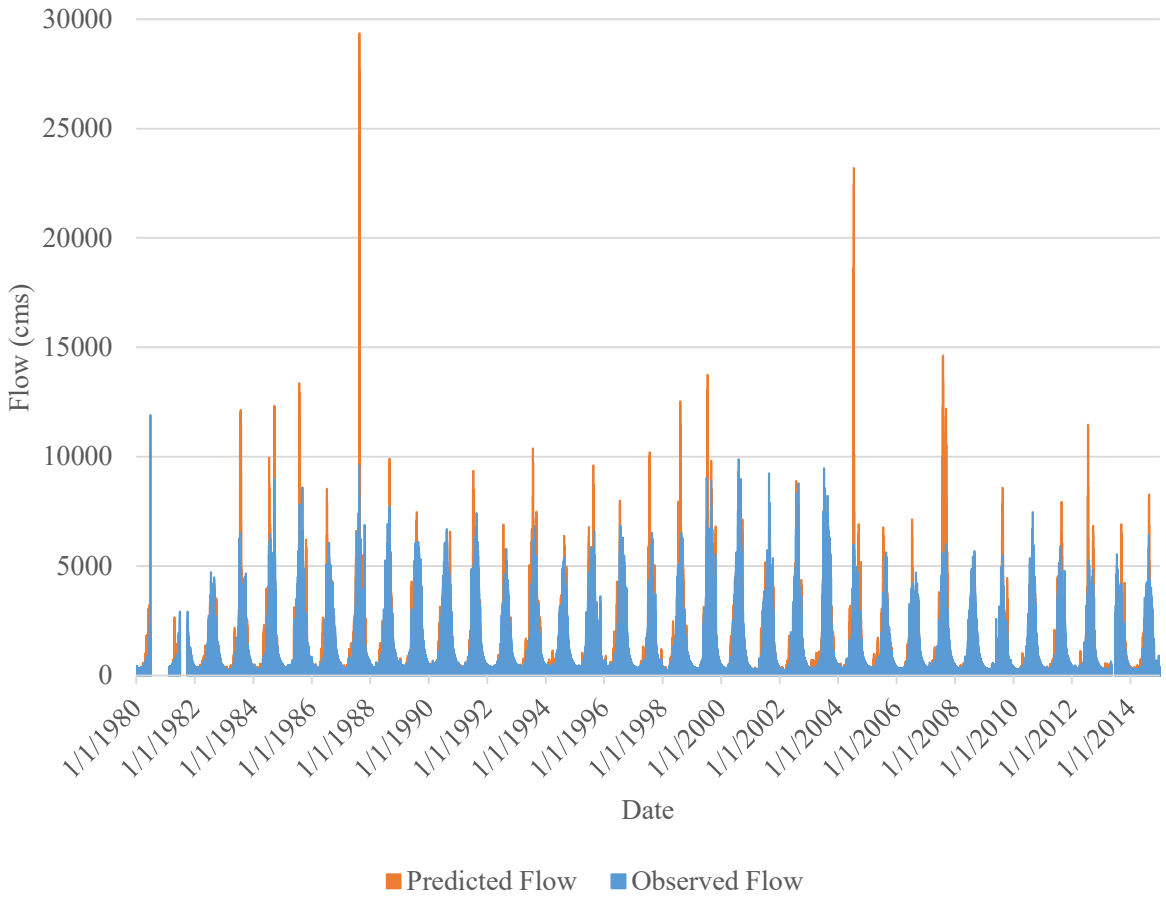

Figure 4-2: Predicted and observed flows for Saptakosi for the 35-year historical period.

Figure 4-2 shows the daily streamflow value for both the predicted and observed datasets. From this figure, some outliers and over prediction for the MHS dataset can be seen, however it is difficult to view any specific trends in the data. To better understand the trends in the data, the 35-year dataset is then summarized by month to generate an average yearly hydrograph for the station. Figure 4-3, Figure 4-4, and Figure 4-5 show the difference between the observed flow and the flow from the MHS, or predicted data on a monthly basis.

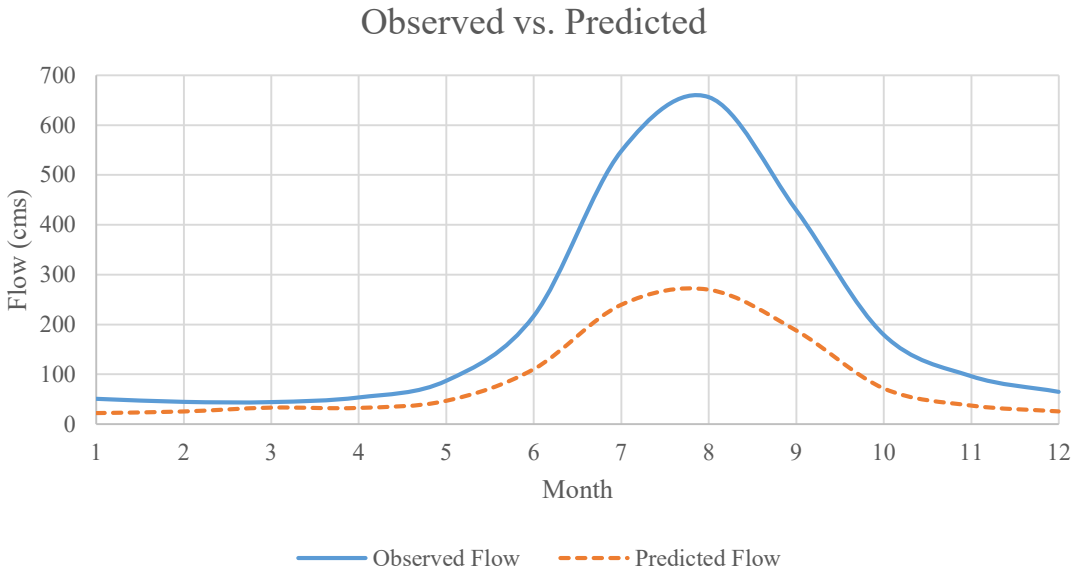

Figure 4-3: Observed vs. predicted flow for Station Marsyangdi, Nepal

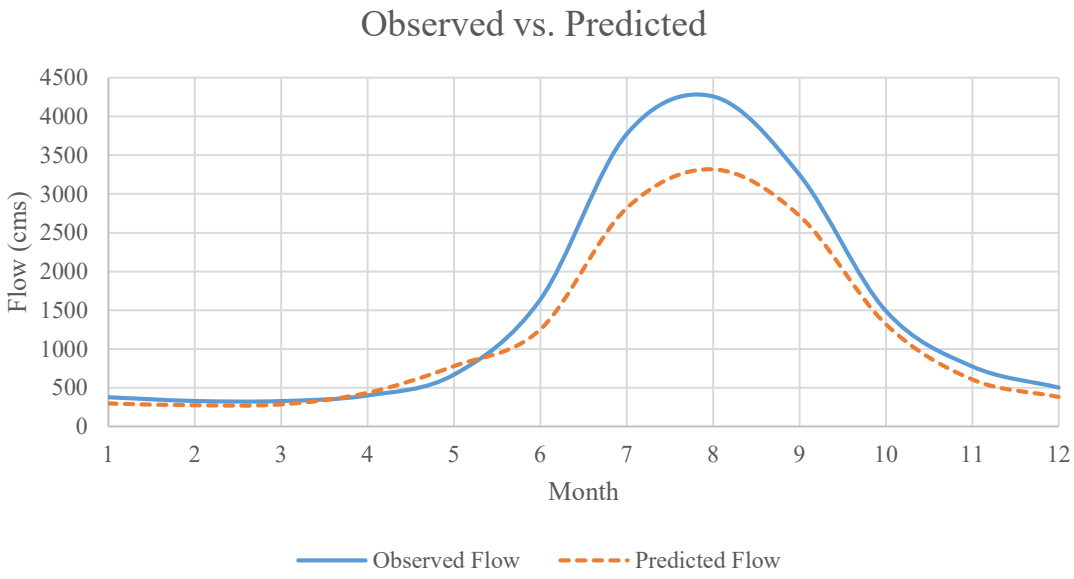

Figure 4-4: Observed vs. predicted flow for Station Saptakosi, Nepal.

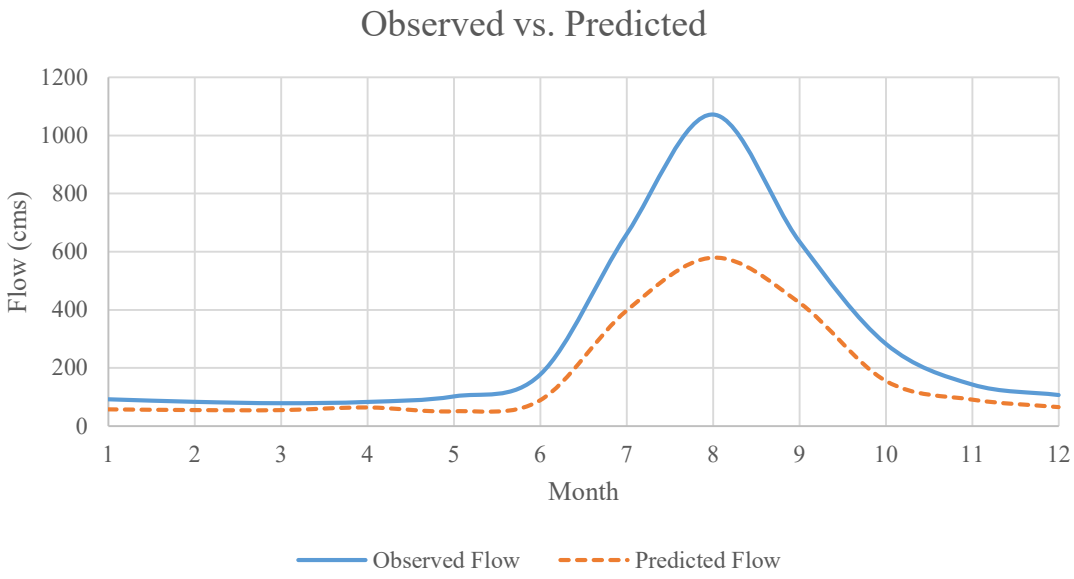

Figure 4-5: Observed vs. predicted flow for Station Bheri, Nepal.

From these figures, the difference between the observed data, and the MHS predicted flow can be seen. For all three stations, the MHS data under predicted the flow. Saptakosi had the best correlation between the two datasets, while Bheri and Marsyangdi both were under predicted, but both showed similar timing trends. To better understand the relationship and error between the MHS and observed flow, error metrics must be calculated.

4.2 Distribution Tests

To determine the accuracy of the MHS data, the error metrics related to the distribution of the two datasets must be calculated. The different error metrics vary as the magnitude of the predicted flow changes, or if there exists any timing error. These effects are summarized in the following sections.

4.2.1 Effects of Magnitude and Timing Change

To first quantify the effect of magnitude change on distribution metrics, synthetic hydrographs were used to simulate different levels of magnitude change. These synthetic hydrographs were generated as variations of a basic sine wave to simulate the basic trends found in hydrographs such as noise, weighted noise, outliers, over and under prediction, time shifts, and combinations of these trends. To determine the effect of magnitude shifts, the magnitude of over and under prediction was varied at ±50% to simulate different levels of under and over prediction. Figure 4-6, Figure 4-7, Figure 4-8, Figure 4-9, Figure 4-10, and Figure 4-11 show the effects of magnitude on specific error metrics.

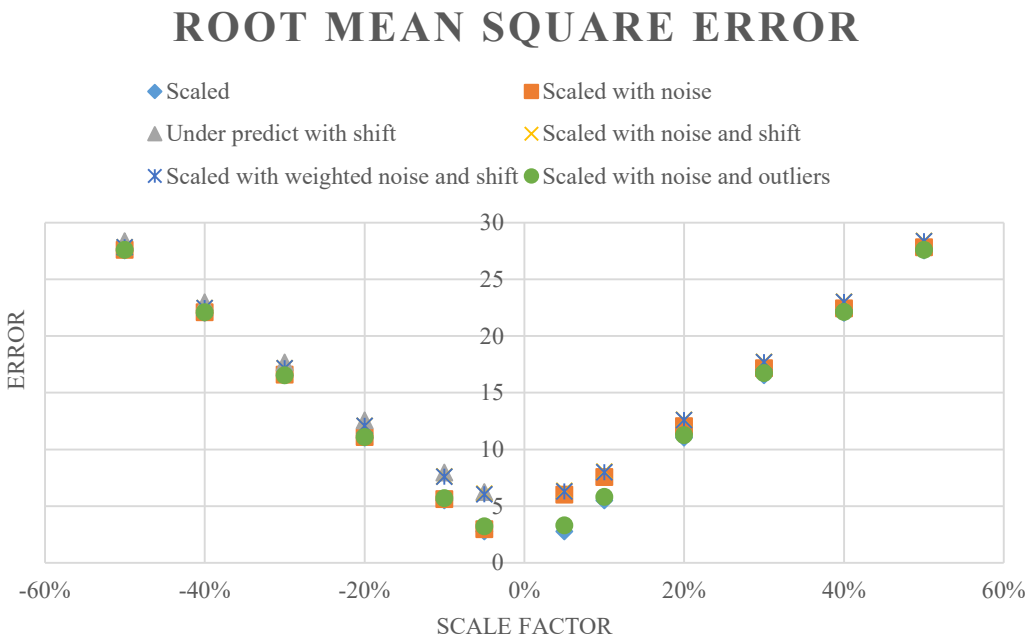

Figure 4-6: Change in RMSE with magnitude change.

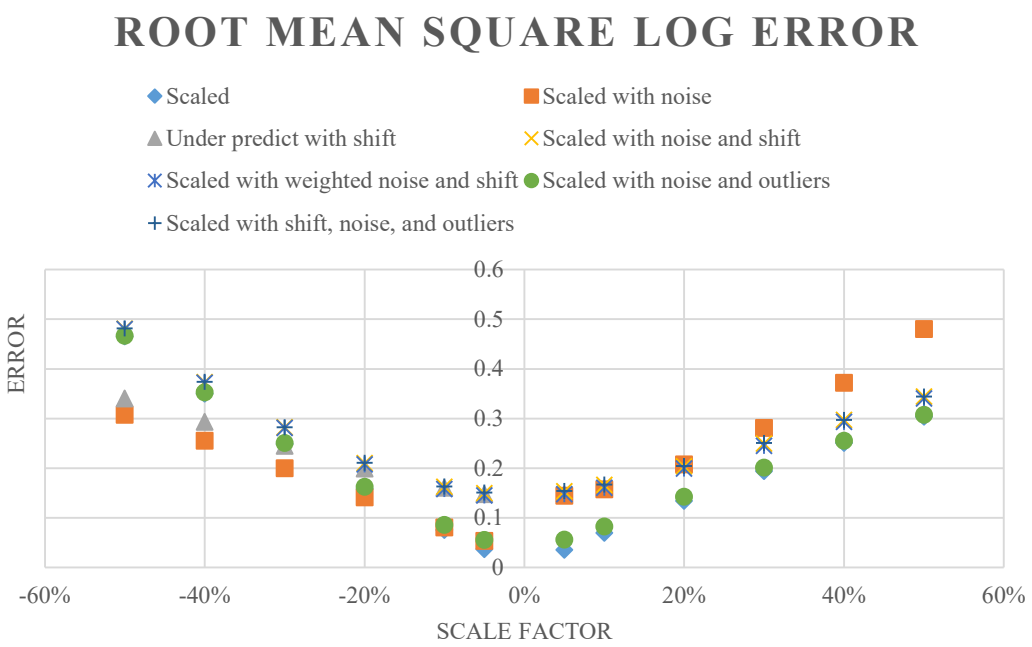

Figure 4-7: Change in RMSLE with magnitude change.

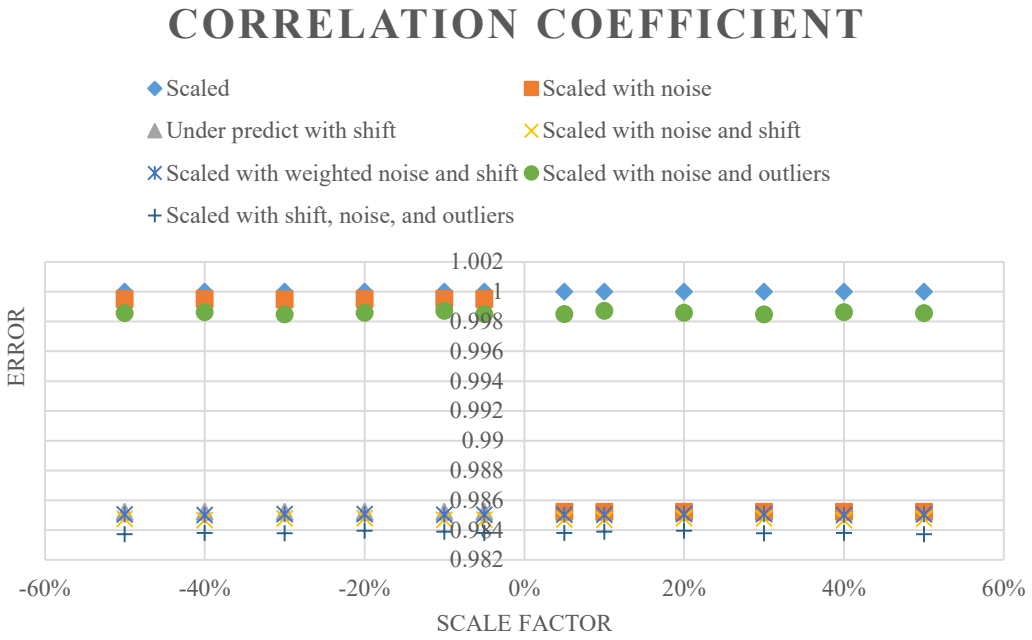

Figure 4-8: Change in correlation coefficient with magnitude change.

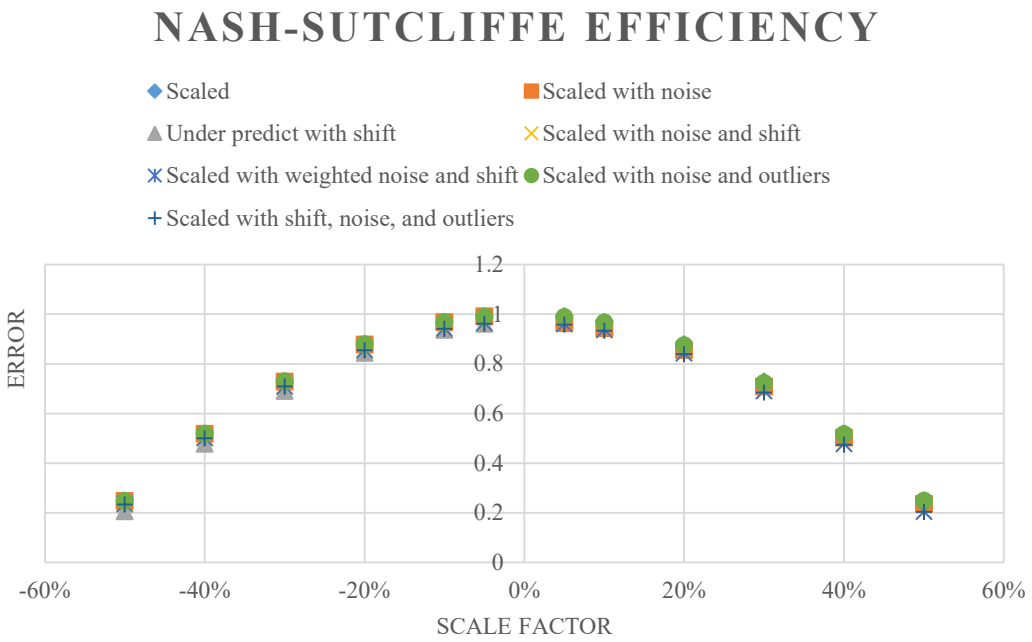

Figure 4-9: Change in Nash-Sutcliffe Efficiency with magnitude change.

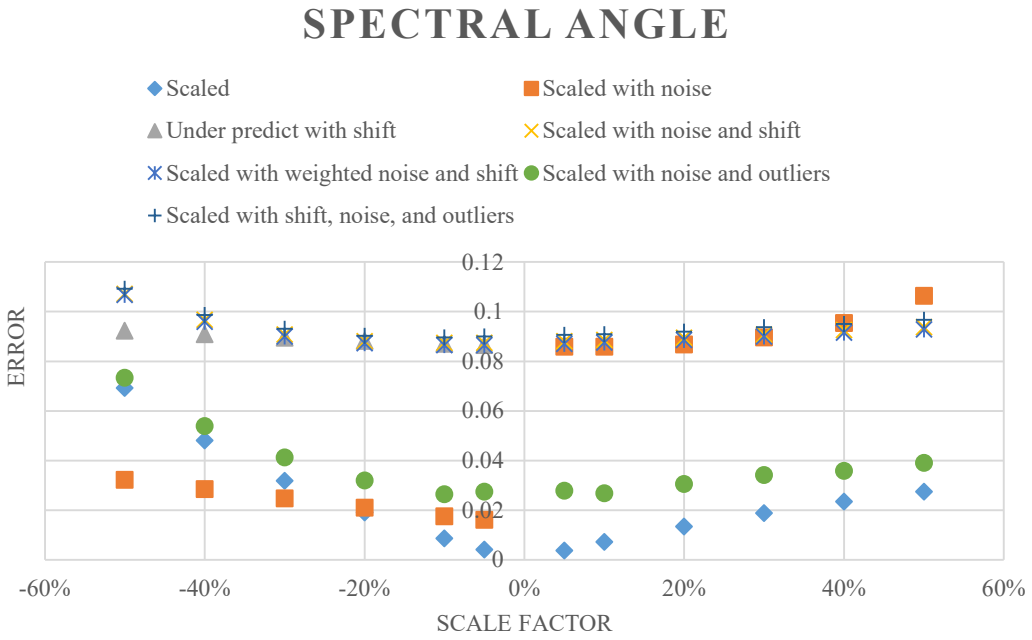

Figure 4-10: Change in spectral angle coefficient with magnitude change.

These figures show that for different levels of magnitude shift, such as those shown in the visual analysis, the RMSE, RMSLE, and Nash-Sutcliffe Efficiency (NSE) increase as the over

and under prediction factors increase, with the lowest errors occurring when the over and under prediction factor is at a minimum. The correlation coefficient and spectral angle coefficients showed little to no variation in error, even as the over and under prediction factor changed from -50% to 50%. From this analysis, researchers can see that the effects of magnitude shifts are best shown by the RMSE, RMSLE, and NSE.

To quantify the effect of timing change on distribution metrics, the same synthetic hydrographs were used to simulate different levels of timing shift. Figure 4-11, Figure 4-12, and Figure 4-13 show the changes in the different error metrics as the timing shift magnitude was changed between -10 days and +10 days while holding the other factors constant.

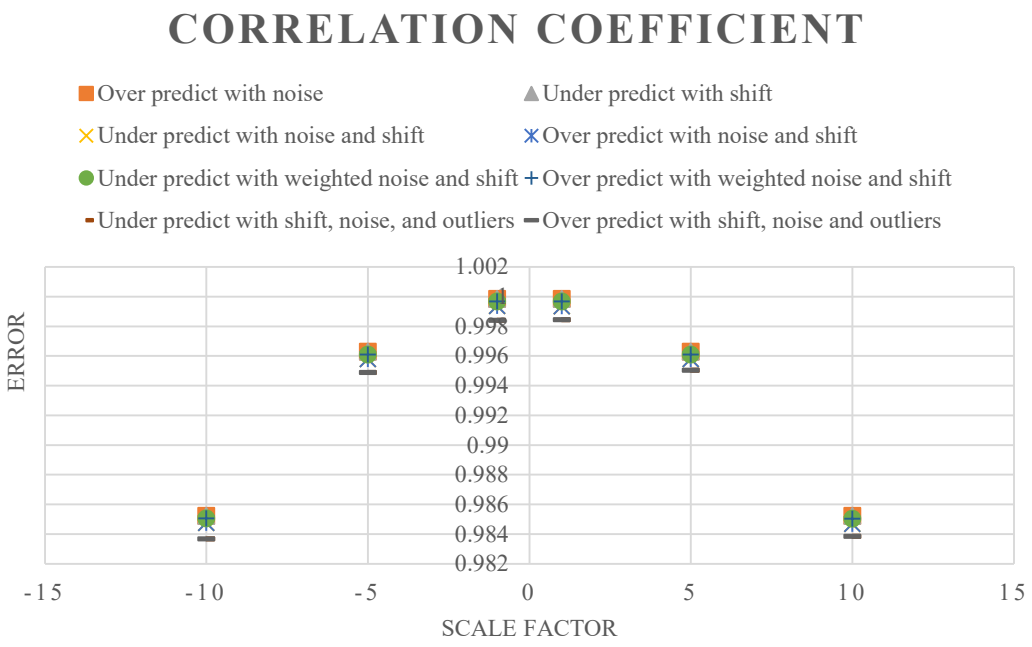

Figure 4-11: Change in correlation coefficient with timing shift.

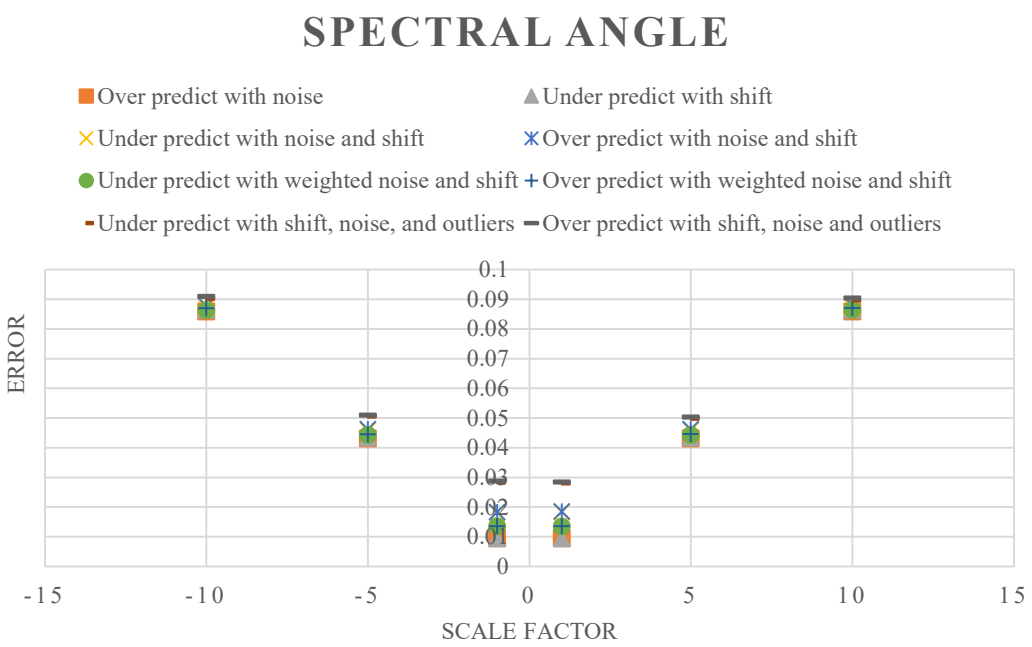

Figure 4-12: Change in spectral angle coefficient with timing shift.

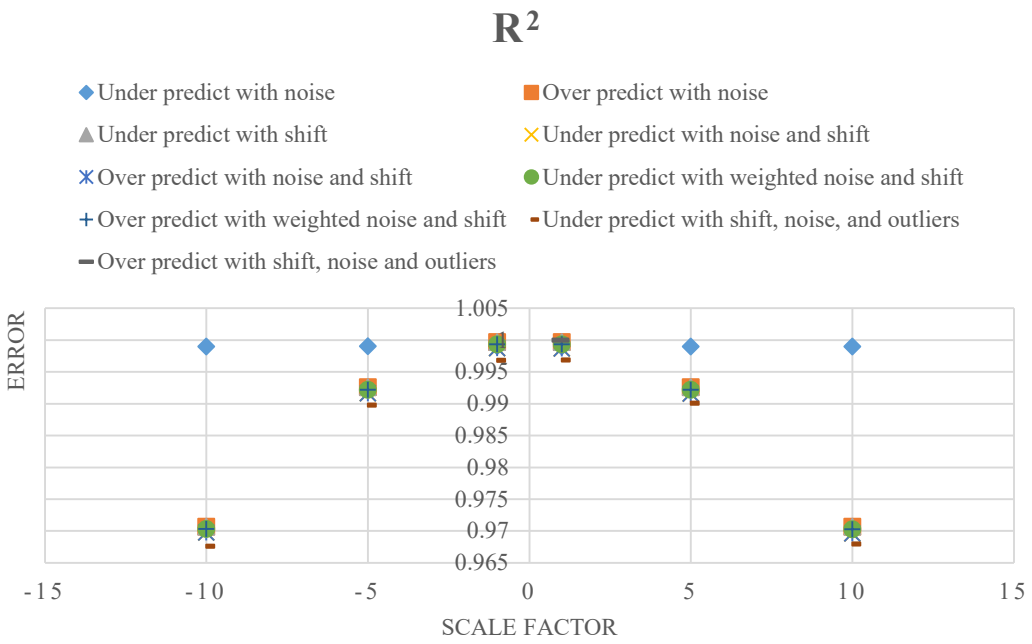

Figure 4-13: Change in R^2 with timing shift.

These figures show that for different levels of timing shift, such as those shown in the visual analysis, the correlation coefficient, R^2 coefficient, and the spectral angle coefficient react

to the change in timing shift, however the magnitude of these changes is very small. The RMSE, RMSLE, and NSE showed no change in error as the timing shift changed.

4.2.2 Results

Based on the results from changing the magnitude and timing, specific metrics were shown to be more effective than others when it comes to timing and magnitude changes. The anomaly correlation coefficient, RMSE, RMSLE, and Nash-Sutcliffe Efficiency were calculated for the three stations on both a monthly and yearly basis. The monthly values were calculated by subdividing the data into monthly increments, while the yearly value was calculated using the full datasets. Table 4-1,

Table 4-2, and Table 4-3 summarize the error metrics based on distribution for the three stations.

Table 4-1: Metrics Summary for Marsyangdi, Nepal

Month	Observed Flow	Predicted Flow	Correlation	RMSE	RMSLE	Nash-Sutcliffe Efficiency
January	50.989	22.319	0.250	1.010	0.030	-43.049
February	44.876	25.560	-0.052	0.776	0.023	-23.188
March	44.420	33.292	0.143	0.919	0.023	-16.439
April	53.754	32.704	0.188	1.098	0.026	-13.458
May	87.215	46.952	0.377	1.778	0.027	-4.830
June	217.098	110.481	0.499	5.244	0.031	-2.856
July	547.681	239.650	0.207	12.490	0.033	-8.085
August	655.973	269.912	0.266	14.443	0.033	-16.245
September	429.553	188.057	0.435	9.253	0.031	-7.566
October	179.540	71.813	0.451	4.076	0.033	-10.150
November	96.412	37.374	0.393	2.101	0.033	-26.366
December	64.955	25.669	0.281	1.376	0.032	-51.062
Yearly	129.439	61.910	0.812	1.343	0.009	0.052

Table 4-2: Metrics Summary for Saptakosi, Nepal

Month	Observed Flow	Predicted Flow	Correlation	RMSE	RMSLE	Nash-Sutcliffe Efficiency
January	377.255	299.771	0.433	2.622	0.008	-3.195
February	329.465	276.956	0.390	2.373	0.008	-2.358
March	328.017	287.843	0.435	2.840	0.009	-2.527
April	400.235	436.944	0.563	5.791	0.011	-2.534
May	668.020	780.399	0.478	11.646	0.014	-0.855
June	1636.301	1253.126	0.525	25.078	0.016	-0.447
July	3770.640	2816.260	0.424	55.166	0.016	-3.185
August	4257.869	3317.010	0.398	48.739	0.012	-1.970
September	3248.639	2719.062	0.644	30.440	0.009	-0.358
October	1491.013	1321.225	0.751	13.799	0.008	0.047
November	774.756	609.332	0.799	5.850	0.008	-0.692
December	502.272	384.242	0.633	3.942	0.009	-2.073
Yearly	941.774	804.642	0.917	4.955	0.003	0.805

Table 4-3: Metrics summary for Bheri, Nepal

Month	Observed Flow	Predicted Flow	Correlation	RMSE	RMSLE	Nash-Sutcliffe Efficiency
January	92.436	57.943	0.454	3.270	0.043	-72.083
February	83.719	55.459	0.489	2.428	0.037	-71.900
March	78.933	55.319	0.384	2.157	0.034	-64.673
April	83.362	64.471	0.259	4.241	0.054	-81.034
May	102.265	51.283	0.402	3.797	0.056	-15.578
June	177.610	89.422	0.750	12.500	0.066	-2.038
July	660.108	397.785	0.586	25.658	0.045	-0.794
August	1071.834	579.767	0.183	36.900	0.047	-5.841
September	634.971	424.802	0.586	25.649	0.040	-1.047
October	283.168	155.720	0.732	11.250	0.049	-2.640
November	143.241	91.415	0.302	3.859	0.034	-11.939
December	107.073	65.648	0.146	2.884	0.034	-44.579
Yearly	196.819	119.480	0.867	3.576	0.013	0.433

From these tables, the accuracy of the MHS data can be determined based on distribution metrics. All three stations had typical values of error for the metrics calculated. Bheri had the highest correlation coefficient, while Saptakosi had the lowest RMSE error. All three stations had very low RMSLE and Nash-Sutcliffe Efficiency errors on a yearly basis. For the individual months, the Nash-Sutcliffe efficiency error was very large, especially for the months outside of the high flow season. However, the error for all three metrics was less for the yearly value. This trend occurs because the data was not subdivided for the yearly calculation, allowing the metric to account for falling and rising limbs of the hydrograph that would otherwise be cut out of the monthly increment. This trend between the monthly and the yearly values justify calculating the metrics on both a monthly and yearly basis, to better understand any trends or errors in the data.

4.3 Population Metrics

To determine the accuracy of the MHS data based on population metrics, the mean difference and mean variance were calculated both per month, and for the total dataset. These population metrics were used to discount the effects of timing changes from the MHS data compared to the observed data points. The population metrics also help researchers to understand how the population of the predicted flow compares to the population of the observed flow. Table 4-4, Table 4-5, and Table 4-6 summarize the results from the population metric tests for the three stations.

Table 4-4: Population Metrics Summary for Marsyangdi, Nepal

Month	Observed Flow	Predicted Flow	Mean Difference	Mean Variance
January	50.989	22.319	1.617	0.110
February	44.876	25.560	1.527	0.083
March	44.420	33.292	1.461	0.180
April	53.754	32.704	1.387	0.573
May	87.215	46.952	2.079	0.317
June	217.098	110.481	2.297	0.860
July	547.681	239.650	1.685	0.352
August	655.973	269.912	1.871	0.285
September	429.553	188.057	1.511	0.227
October	179.540	71.813	1.835	0.097
November	96.412	37.374	1.569	0.035
December	64.955	25.669	1.638	0.044
Yearly	129.439	61.910	1.705	0.292

Table 4-5: Population Metrics Summary for Saptakosi, Nepal

Month	Observed Flow	Predicted Flow	Mean Difference	Mean Variance
January	377.255	299.771	1.267	0.062
February	329.465	276.956	1.194	0.059
March	328.017	287.843	1.150	0.112
April	400.235	436.944	0.934	0.211
May	668.020	780.399	0.868	0.283
June	1636.301	1253.126	1.318	0.309
July	3770.640	2816.260	1.360	0.273
August	4257.869	3317.010	1.291	0.157
September	3248.639	2719.062	1.200	0.100
October	1491.013	1321.225	1.136	0.093
November	774.756	609.332	1.275	0.032
December	502.272	384.242	1.309	0.030
Yearly	941.774	804.642	1.176	0.163

Table 4-6: Population Metrics Summary for Bheri, Nepal

Month	Observed Flow	Predicted Flow	Mean Difference	Mean Variance
January	92.436	57.943	1.617	0.110
February	83.719	55.459	1.527	0.083
March	78.933	55.319	1.461	0.180
April	83.362	64.471	1.387	0.573
May	102.265	51.283	2.079	0.317
June	177.610	89.422	2.297	0.860
July	660.108	397.785	1.685	0.352
August	1071.834	579.767	1.871	0.285
September	634.971	424.802	1.511	0.227
October	283.168	155.720	1.835	0.097
November	143.241	91.415	1.569	0.035
December	107.073	65.648	1.638	0.044
Yearly	196.819	119.480	1.705	0.292

The mean difference and mean variance for the three stations were very small. Each station had a different value for mean difference. Overall, the mean difference was greater than one, signifying that the MHS data consistently under predicted the flow for the station. However, the error for the mean difference was the smallest during the high flow months. The small values for mean variance signify that there is very little variation between the values of mean difference for the two datasets. The same trend in the yearly values of the distribution metrics occurred for the population metrics. Taking the metric of the full dataset avoids cutting off the rising and falling limbs of the hydrograph as they fall on monthly division lines.

4.4 Timing Tests

To account for the change in shape of the MHS data, and to determine the effect of any timing changes, the R^2 and spectral angle coefficient were calculated for the data. The results for the three stations are summarized in Table 4-7.

Table 4-7: Timing Metrics Summary for Nepal Stations

Error Metric	Marsyangdi	Saptakosi	Bheri
R^2	0.364	0.634	0.542
Correlation Coefficient	0.812	0.917	0.867
Spectral Angle Coefficient	0.731	0.884	0.830

This table shows that although the three stations had low R^2 coefficients, the correlation coefficients and spectral angle coefficients were very high. This shows that the MHS dataset had a very similar shape to the observed data, even if the timing of the events was inaccurate, or if the base flow for the stream reach was inaccurate.

4.5 Summary

The MHS data is an appropriate model for streamflow data in the Nepal watersheds analyzed. While the ultimate decision on how to use the model lies with the individual agencies, the values of the error metrics can be used to determine the next steps in model calibration. Understanding the values of the different error metrics allows users to determine the next step in using the MHS data.

The high values of spectral angle coefficient prove that the MHS data retains a similar shape to the observed data for the entire dataset. The distribution metrics show that the MHS data is accurate overall, but especially during the high flow months. The error is larger during the months of little flow. The mean difference for the three stations also support this conclusion, as the mean difference measures the error between the two time-series on a population basis, not on a timing basis. The mean difference also shows that the MHS data is under predicted for each

of the stations, which could be used to better calibrate the ERA-Interim data to improve the base flow in each stream reach.

To fully interpret the results of the statistical analysis, all the metrics must be used. Using one metric does not fully explain the trends and errors in the data. For example, a high spectral angle may indicate a good shape match, however if the R^2 coefficient is very low, that indicates a low overall match in the data. By using a combination of distribution, population and timing metrics, researchers can have a complete understanding of any errors in the data, and how to better calibrate the data.

5 CONCLUSION

The modeled historical streamflow (MHS) data model has the potential to be used worldwide as a surrogate for historical data where no observations exist. The data can be analyzed using correlation, error, and timing tests to define the accuracy and appropriateness of the MHS data when compared to observed data. By comparing the MHS and the observed data at different stations worldwide, researchers can better understand the overall accuracy and any bias of the MHS data. These tests can be performed quickly and easily at distinct stations, and researchers can then identify the specific tests that will help define the appropriateness of using the MHS data as a surrogate historical dataset.

The different metrics were calculated on both a monthly and yearly basis. The monthly values were calculated by subdividing the overall dataset into specific months, and then calculating the metric on the data values. This was done to identify any seasonal trends within the separate months, as well as allow researchers to identify the different high and low flows for each country. The yearly metrics were calculated on the entire dataset, which resulted in lower error for the different metrics. The yearly metrics resulted in lower error, as the data was not subdivided, meaning that the different metrics were able to capture the entire hydrographs of events.

Certain metrics are sensitive to different trends in the data. The RMSE, RMSLE, and NSE are all sensitive to changes in over and under prediction. The correlation coefficient, R^2

coefficient and spectral angle coefficients are relatively immune to changes in magnitude, while being slightly sensitive to shifts in timing. Understanding these trends, and other trends allows researchers to understand when to use specific metrics, as well as interpret the results from the different metrics.

The ultimate decision of whether to use the MHS data as a surrogate or not lies with the individual regions and agencies, however these error metrics can be used to determine whether the MHS data is accurate enough for the specific regions. For example, the data had correlation coefficients ranging from -0.6594 to 0.5751 worldwide, with mean differences ranging from 0.00115 to 17.8446. The MHS data was most accurate in Tanzania, with the lowest overall values of error. Overall, the MHS data had a RMSE error from 0.0071 to 74.8256. Different agencies may have distinct cutoffs and thresholds for these common types of metrics, which determine whether the model has performed well or poorly, however multiple metrics should be used to give a complete picture of the correlation in the datasets.

The statistical methods discussed in this thesis, along with other metrics, will be made available as a web application hosted on the Tethys development platform. This web application will allow countries and regions to upload their own observational data, access the MHS data from the SPT, and conduct the statistical analysis necessary to determine the overall accuracy of the MHS data for the region. This application will include the four analysis methods discussed in this thesis; visual, distribution, population, and timing tests. The metrics available through the application will include the metrics discussed in this thesis, as well as other metrics found in research and industry.

Using this statistical analysis method, the next step in the research would be change the MHS data by calibrating the RAPID model. This calibration of the RAPID model would correct the

errors defined by the error metrics, such as minimize the mean difference, maximize the spectral angle coefficient, and minimize the RMSE value. The calibration of the RAPID model would also correct for any bias in the MHS data due to timing errors or shifts. Further research could be done to analyze the 15-day forecasts produced by the MHS, and determine the forecast skill, as well as the significance in terms of flood events.

In conclusion, statistical analysis can be used to determine the accuracy of the MHS data as compared to observed data worldwide. The different metrics describe different trends and errors between the two datasets and using multiple metrics can give a better description of the data rather than using a single metric. Some metrics, such as the RMSE, RMSLE and NSE describe changes in magnitude in the distribution of the datasets, while others such as the correlation coefficient, R^2 coefficient and spectral angle coefficient describe any timing errors within the dataset. Researchers and users of the statistical analysis methods described here should calculate multiple metrics to best understand the data. The statistical analysis should also be done for each individual region, before agencies determine how to use the MHS data in the region. Once these metrics have been calculated, the MHS data can then be used to better understand streamflow trends throughout the individual regions.

REFERENCES

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., . . . Vitart, F. (2012). *ERA-Interim/Land: A global land-surface reanalysis based on ERA-Interim meteorological forcing.* Berkshire: European Centre for Medium-Range Weather Forecasts.

Basalmo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., & Betts, A. K. (2009). A Revised Hydrology for the ECMWF Model: Verification from Field Site to Terrestrial Water Storage and Impact in the Integrated Forecast System. *Journal of Hydrometeorolgy*.

Booker, D., & Snelder, T. (2012). Comparing methods for estimating flow duration curves at ungauged sites. *Journal of Hydrology*, 434,78-94.

Boyle, D. P., Guta, H. V., & Sorooshian, S. (2000). Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, 3663-3674.

Buehler, B. D. (2011). *Analyzing the Potential for Small Hydroelectric Power Installment in the Dominican Republic.* Provo: Brigham Young University.

Burn, D. H., & Hag Elnur, M. A. (2002). Detection of Hydrologic Trends and Variability. *Journal of Hydrology*, 107-122.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., & Bechtold, P. (2011). The ERA-Interim Analysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553-597.

ECMWF. (2017, 12 20). *ERA-Interim/Land*. Retrieved from European Center for Medium-Range Weather Forecasts: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim-land

ECMWF. (2017, 12 20). *Who we are*. Retrieved from European Center for Medium-range Weather Forecast: https://www.ecmwf.int/en/about/who-we-are

Everitt, B. S. (2002). *Cambridge Dictionary of Statistics.*

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: the forecast package for R. *Journal of Stastistical Software*.

Jain, S. K., & Sudheer, K. P. (2007). Fitting of Hydrologic Models: A Close Look at the Nash-Sutcliffe Index. *Journal of Hydrologic Engineering*, 981-986.

Lehmann, E. L., & Casella, G. (1998). *Springer Texts in Statistics.* New York: Springer-Verlag.

McCuen, R. H., Knight, Z., & Cutter, A. G. (2006). Evaluation of the Nash-Sutcliffe Efficiency Index. *Journal of Hydraulic Engineering*.

Ministerio de Economía, Planificación y Desarrollo. (2015). Estrategia Nacional de Desarrollo 2030. Santo Domingo, Distrito Naional, Republica Dominicana.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 282-290.

Office of Energy Efficiency and Renewable Energy. (2017, October 12). *Hydropower Basics*. Retrieved from Water Power Technologies: https://energy.gov/eere/water/hydropower-basics

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S., & Thielen, J. (2015). The monetary benefit of early flood warnings in Europe. *Environmental Science & Policy*, 278-291.

Post, D. A., & Jakeman, A. J. (1999). Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. *Ecological Modellingq*, 91-104.

Ramsey, F. L., & Schafer, D. W. (2013). *The Statistical Sleuth.* Boston: Cengage Learning.

Reich, N. G., Lessler, J., Sakrejda, K., Lauer, S. A., Iamsirithaworn, S., & Cummings, D. A. (2016). *Case study in Evaluating Time Series Prediction Models using the Relative Mean Absolute Error.*

Robila, S. A., & Gershman, A. (2005). Spectral Matching Accuracy in Processing Hyperspectral Data. *Signals, Circuits and Systems, 2005. ISSCS 2005. International Symposium on*, (pp. 163-166).

Roulstone, I., & Norbury, J. (2013). How Math Helped Forecast Hurricane Sandy. *Scientific American*.

Simmons, A. J., Willet, K. M., Jones, P. D., Thorne, P. W., & Dee, D. P. (2010). Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *Journal of Geophysical Research: Atmospheres*.

Singh, V. P. (1987). On application of the Weibull distribution in hydrology. *Water Resources Management*, 33-43.

Snow, A. D. (2015). *A New Global Forecasting Model to Produce High-Resolution Stream Forecasts.* Provo: Brigham Young University.

Souffront, M. (2016). *Mekong Storage Capacity*. Retrieved from Tethys: http://tethys.byu.edu/apps/storagecapacitymekong/

Souffront, M., & Jackson, E. (2016, April 7). *Hydro Power*. Retrieved from BYU Hydroinformatics Lab Apps Portal: http://tethys.byu.edu/apps/hydropower/

Stevenson, M. (2006). Forecast verification: a practitioner's guide in atmospheric science. *International Journal of Forecasting*, 403-405.

Tomczak, M. (1998). Saptial Interpolation and its Uncertainty Using Automated Anisotropic Inverse Distance Weighting (IDW) - Cross-Validation/Jackknife Approach. *Journal of Geographic Informationa nd Decision Analysis*, 18-30.

Tornquist, L., Vartia, P., & Vartia, Y. O. (1985). How Should Relative Changes be Measured? *The American Statistician*, 43-46.

University of Virginia Library. (2018, January 9). *Understanding Q-Q Plots*. Retrieved from Research Data Services & Sciences: http://data.library.virginia.edu/understanding-q-q-plots/

Uppala, S. M., Dee, D. P., Kobayashi, S., & Simmons, A. J. (2008). Evolution of Reanalysis at ECMWF. *Proceedings of Third WCRP International Conference on Reanalysis.* Tokyo, Japan.

Vogel, R., & Fennessey, N. (1994). Flow-Duration Curves I: New interpretation and confidence intervals. *Journal of Water Resources Planning and Management*, 120,485-504.

Wetterdienst, D. (2014). *ECMWF - European Centre for Medium-Range Weather Forecasts.* Berlin: Federal Ministry of Transport and Digital Infrasctructure.

Young, D. F., Munson, B. R., Okiishi, T. H., & Huebsch, W. W. (2011). *A Brief Introduction to Fluid Mechanics.* Hoboken: John Wiley & Sons, Inc.

Zhang, Y., Vaze, J., Chiew, F. H., & Li, M. (2015). Comparing flow duration curve and rainfall-runoff modelling for predicting daily runoff in ungauged catchments. *Journal of Hydrology*, 72-86.

APPENDIX A: ERROR METRIC TABLES

Table 0-1: Correlation Coefficient Summary for Nepal

Station	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Yearly Correlation
Asaraghat	0.5199	0.4444	0.4016	0.6880	0.3931	0.4828	0.4934	0.2669	0.5427	0.6040	0.5474	0.6347	0.8240
Babai	0.1933	0.5686	0.2295	0.3292	0.6055	0.7635	0.4160	0.2837	0.3736	0.5123	0.3770	0.4307	0.8406
Bheri	0.4536	0.4892	0.3837	0.2586	0.4017	0.7501	0.5859	0.1833	0.5861	0.7316	0.3022	0.1463	0.8675
Kaligandaki	0.5483	0.4396	0.1879	0.3424	0.4724	0.7089	0.4931	0.5085	0.5716	0.6662	0.5516	0.5414	0.8674
Kamali	0.4343	0.5502	0.5016	0.6399	0.4733	0.6137	0.5645	0.1849	0.5243	0.6606	0.5639	0.5661	0.8622
Kankai	0.2174	0.1556	0.1862	0.4224	0.4344	0.4655	0.4192	0.4420	0.1651	0.2960	0.2972	0.2829	0.8289
Marsyangdi	0.2497	-0.0516	0.1428	0.1882	0.3765	0.4993	0.2066	0.2661	0.4347	0.4510	0.3934	0.2810	0.8122
Narayani	0.4377	0.3039	0.3188	0.4216	0.4799	0.6410	0.4274	0.3992	0.6340	0.7562	0.7671	0.6538	0.8775
Rapti	0.3682	0.5163	0.5256	0.4117	0.6604	0.7535	0.5870	0.3664	0.5633	0.4901	0.4867	0.4036	0.8869
Saptakosi	0.4330	0.3904	0.4353	0.5634	0.4780	0.5250	0.4240	0.3975	0.6444	0.7512	0.7993	0.6327	0.9168
Seti	0.5517	0.6578	0.2569	-0.2317	-0.3445	0.8343	0.6940	0.3660	0.7439	0.4820	0.5178	0.0325	0.7557
Tinaukhola	-0.4260	-0.1302	-0.0945	-0.1004	-0.1213	0.2205	0.3678	0.4005	0.4080	0.0691	-0.1345	-0.2815	0.3427
Average of Correlation	**0.3318**	**0.3612**	**0.2896**	**0.3278**	**0.3591**	**0.6048**	**0.4732**	**0.3388**	**0.5160**	**0.5392**	**0.4558**	**0.3603**	**0.8068**

Table 0-2: Correlation Coefficient Summary for Tanzania

Station	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Yearly Correlation
Chimala at Chitakelo	-0.0274	0.3397	0.2634	0.1679	0.0731	0.0354	-0.1101	-0.1053	-0.1776	0.0950	0.0016	0.1722	0.3058
Igawa	0.6951	0.5595	0.5179	0.3881	0.4438	0.3198	0.2072	0.2278	0.2642	0.2013	0.4261	0.7360	0.8338
Ihimbu	0.4536	0.4636	0.3015	0.3243	0.4339	0.4469	0.4291	0.5226	0.3942	0.0961	0.4409	0.5691	0.7799
Ilongo	0.4756	0.4728	0.3460	0.2493	0.3892	0.4198	0.5666	0.5876	0.6022	0.5478	0.3876	0.5642	0.7946
Ipatagwa	-0.0428	0.2776	0.1775	0.0946	0.0217	0.0322	-0.0943	-0.0797	-0.2185	0.1370	-0.0288	0.1547	0.2684
Kimani	0.6121	0.6309	0.4065	0.3991	0.4275	0.4861	0.6101	0.5555	0.5071	0.2872	0.5338	0.7608	0.8624
Mawande	0.6593	0.5112	0.4498	0.3963	0.6963	0.5404	0.5405	0.3695	0.3091	0.2988	0.2919	0.7055	0.7832
Msembe	0.6018	0.4335	0.5248	0.3462	0.6297	0.5596	0.6110	0.6727	0.3363	0.2344	0.6023	0.6070	0.7441
Mswisi	0.3466	0.0191	0.0349	0.2922	0.2673	0.4785	0.0348	0.1668	0.5634	0.2640	-0.1231	0.3245	0.5311
Mtandika	0.6373	0.6234	0.4437	0.5486	0.4822	0.6642	0.4587	0.7259	0.5406	0.3727	0.6887	0.7155	0.6715
Mtitu	0.5935	0.4504	0.3688	0.3509	0.4360	0.4870	0.4507	0.6034	0.4728	0.4125	0.5879	0.6812	0.7642
Ndiuka	0.4398	0.4908	0.3489	0.2783	0.3746	0.4486	0.4382	0.4608	0.4350	0.2175	0.4322	0.6766	0.8151
Average of Correlation	**0.4537**	**0.4394**	**0.3487**	**0.3196**	**0.3896**	**0.4099**	**0.3452**	**0.3923**	**0.3357**	**0.2637**	**0.3534**	**0.5556**	**0.6795**

Table 0-3: Correlation Coefficient Summary for the Dominican Republic

Station	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Yearly Correlation
Cacique	0.4749	0.4318	0.5083	0.6064	0.6432	0.6244	0.4555	0.4652	0.4799	0.4664	0.4401	0.6074	0.5973
Camu Bayacanes	0.4558	0.5179	0.3931	0.3970	0.4707	0.5060	0.3085	0.3092	0.2754	0.4411	0.4372	0.3818	0.4053
Guazumal	0.1776	-0.6594	-0.4691	0.3060	0.8295	0.2080	0.5112	0.3633	0.4361	0.2968	0.1998	0.4349	0.3210
Jinamagao	0.2855	0.3849	0.4718	0.3053	0.5118	0.3624	0.2159	0.3239	0.4083	0.4762	0.2979	0.4496	0.4067
Manabao	0.4707	0.5581	0.2329	0.4720	0.4897	0.3926	0.6553	0.6779	0.5701	0.4227	0.6704	0.8449	0.5726
Pinar Quemado	0.7167	0.3114	0.3496	0.4191	0.6947	0.4229	0.2812	0.5547	0.5897	0.5893	0.4161	0.7073	0.6334
Puente San Rafael	0.7478	0.4745	0.0530	0.5804	0.7270	0.3472	0.6487	0.5481	0.1484	0.3920	0.5252	0.6449	0.5676
Santa Ana	0.4705	0.5104	0.4813	0.4289	0.5293	0.4067	0.5901	0.6185	0.5544	0.4623	0.5230	0.5411	0.4727
Average of Correlation	**0.4749**	**0.3162**	**0.2526**	**0.4394**	**0.6120**	**0.4088**	**0.4583**	**0.4826**	**0.4328**	**0.4434**	**0.4387**	**0.5765**	**0.4971**

Table 0-4: Mean Difference Summary for Nepal

Station	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Yearly Mean Difference
Asaraghat	1.3185	1.4514	1.2403	0.7458	0.7255	2.1489	1.4245	1.2014	1.1463	1.2342	1.2421	1.2773	1.2195
Babai	1.6093	1.5567	2.0568	2.4772	2.9060	2.9377	1.8693	2.5009	1.9176	2.2617	1.9485	2.0834	2.1958
Bheri	1.6168	1.5270	1.4615	1.3865	2.0790	2.2973	1.6855	1.8715	1.5110	1.8348	1.5694	1.6378	1.7055
Kaligandaki	1.8278	1.7572	1.9273	1.4742	1.4287	1.9294	2.1183	2.1279	1.6565	1.6889	1.7361	1.7740	1.7768
Kamali	1.4681	1.3675	1.2093	0.9914	1.0005	2.0930	1.5918	1.5285	1.3341	1.5013	1.5527	1.5701	1.4011
Kankai	1.1270	1.2180	1.5502	1.6857	1.1207	0.7508	1.0021	0.9792	0.8551	0.7292	0.9286	1.0160	1.1226
Marsyangdi	2.3134	1.8011	1.4340	1.7874	2.0113	2.1844	2.3775	2.5114	2.3458	2.5574	2.6023	2.5587	2.1800
Narayani	2.0258	1.7358	1.3145	1.1623	1.4139	1.9876	2.2081	2.1932	1.9205	2.0911	2.2449	2.1692	1.8159
Rapti	1.4822	1.3550	1.6857	1.7946	1.6967	1.4946	1.6312	1.9316	1.6178	1.5212	1.3804	1.3658	1.6031
Saptakosi	1.2672	1.1937	1.1495	0.9342	0.8682	1.3184	1.3595	1.2914	1.2002	1.1363	1.2749	1.3087	1.1759
Seti	1.9162	1.4943	1.6543	1.7198	5.4631	6.4516	3.3776	3.1841	2.9306	3.9255	2.8186	1.7024	2.6916
Tinaukhola	1.6399	1.8325	2.2156	3.3939	3.1231	2.1827	2.1319	1.4222	0.9425	0.7325	0.5924	0.3368	1.4279
Average of Mean Difference	**1.6344**	**1.5242**	**1.5749**	**1.6294**	**1.9864**	**2.3147**	**1.8981**	**1.8953**	**1.6148**	**1.7678**	**1.6576**	**1.5667**	**1.6930**

Table 0-5: Mean Difference Summary for Tanzania

Station	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Yearly Mean Difference
Chimala at Chitakelo	0.0971	0.0921	0.1119	0.2026	0.2953	0.5433	0.5602	0.7020	0.9053	1.1597	1.1111	0.2284	0.3541
Igawa	1.2731	1.0283	0.7743	0.4481	0.3514	0.4290	0.5448	0.6469	0.7412	0.8074	0.9367	1.1247	0.7215
Ihimbu	1.2208	1.1704	1.1426	1.3458	1.8619	2.3680	3.0123	3.7172	4.3145	3.9490	3.7036	2.5406	2.4562
Ilongo	0.9524	1.2623	1.3439	1.4817	1.0452	0.6514	0.4870	0.4692	0.4154	0.2441	0.1856	0.3617	0.5300
Ipatagwa	0.1328	0.1394	0.1758	0.4639	0.8899	1.9532	2.1439	2.8419	3.7939	4.4864	3.0810	0.3545	0.9584
Kimani	1.5024	1.7514	1.5204	1.2244	0.9492	0.8306	0.7922	0.8571	0.9822	1.0793	0.9724	0.8778	1.0306
Mawande	1.0308	0.8909	1.0345	1.2020	1.4364	1.6945	1.9097	2.2235	2.4916	2.1052	1.3732	1.1986	1.5586
Msembe	0.0263	0.0841	0.1471	0.2409	0.3636	0.2548	0.1564	0.0784	0.0256	0.0042	0.0011	0.0077	0.0503
Mswisi	0.2510	0.3741	0.6627	2.3284	4.5190	5.3837	1.0871	1.0468	4.7650	4.4008	2.9898	0.4244	1.5297
Mtandika	1.2723	1.1447	0.8518	1.6744	7.6648	10.0309	15.4805	17.1574	17.8446	13.6583	8.3998	2.8335	5.7706
Mtitu	1.9833	1.8280	1.2432	1.9025	4.6254	5.7359	7.8732	9.5683	10.4764	10.4537	7.6232	3.7960	5.0499
Ndiuka	1.5271	1.2663	1.3323	1.5107	2.1316	2.6351	3.0945	3.7796	4.0834	3.5132	2.5301	2.2592	2.4841
Average of Mean Difference	**0.9391**	**0.9193**	**0.8617**	**1.1688**	**2.1778**	**2.7092**	**3.0951**	**3.5907**	**4.2366**	**3.8218**	**2.7423**	**1.3339**	**1.8745**

Table 0-6: Mean Difference Summary for the Dominican Republic

Station	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec	Yearly Mean Difference
Cacique	0.7601	0.8872	0.9314	0.8928	0.7518	0.8700	1.1392	1.3504	1.2644	1.0027	0.6957	0.6854	0.9146
Camu Bayacanes	4.6330	5.2912	4.7648	3.5205	2.3841	2.2537	1.9484	1.9924	1.6385	2.3353	3.2237	4.3726	3.0423
Guazumal	7.1677	5.6683	3.5060	1.1387	1.0852	1.8252	4.7228	4.0301	4.5186	3.8564	5.2730	10.1906	3.7965
Jinamagao	4.6960	4.0890	3.9705	2.2389	1.9589	3.4916	2.4957	1.7106	1.5743	2.7128	4.0500	4.7264	2.9603
Manabao	9.5865	8.9122	7.2032	5.0262	4.0114	6.6732	7.4182	6.0697	5.9565	7.3191	8.6992	11.0656	7.2123
Pinar Quemado	6.2033	6.2774	4.8351	3.8244	3.4663	4.3127	5.1898	4.0602	4.0860	5.0235	5.4378	7.4441	5.0118
Puente San Rafael	4.0845	3.6683	4.5524	3.1977	1.9617	4.9350	4.2778	3.3152	3.2816	4.5924	4.7536	5.2511	3.9940
Santa Ana	3.0904	2.2124	2.3910	1.7739	1.2614	2.0320	1.2921	0.7662	0.8080	1.0762	2.7101	2.6730	1.6840
Average of Mean Difference	**5.0277**	**4.6258**	**4.0193**	**2.7016**	**2.1101**	**3.2992**	**3.5605**	**2.9119**	**2.8910**	**3.4898**	**4.3554**	**5.8011**	**3.5770**

Table 0-7: Color-Coded Synthetic Metrics Summary

	Over predict with shift, noise and outliers	Over predict with noise and shift	Under predict with shift, noise, and outliers	Over predict with weighted noise and shift	Under predict with shift	Under predict with noise and shift	Under predict with weighted noise and shift	Over predict with noise	Over predict with noise and outliers	Under predict with noise and outliers	Under predict with noise	Under predict by scale	Over predict by Scale	Error Range
Mean Error	0.9859	0.9966	1.0000	0.9923	0.9933	0.9903	0.9942	0.9933	0.9859	1.0000	0.9965	0.9933	0.9933	0.0141
Mean Absolute Error	1.0000	0.9876	0.9595	0.9810	0.9773	0.9472	0.9416	0.9365	0.4563	0.4514	0.4384	0.4040	0.4040	0.5960
Mean Squared Error	1.0000	0.9561	0.9251	0.9373	0.9263	0.8810	0.8657	0.8552	0.2507	0.2473	0.2051	0.1803	0.1803	0.8197
Root Mean Square Error	1.0000	0.9778	0.9618	0.9681	0.9624	0.9386	0.9304	0.9248	0.5007	0.4973	0.4528	0.4246	0.4246	0.5754
Root Mean Squared Log Error	1.0000	0.9928	0.9800	0.9561	0.9541	0.9715	0.9390	0.9369	0.3655	0.3633	0.3463	0.2389	0.2294	0.7706
Mean Absolute Scaled Error	1.0000	0.9876	0.9595	0.9810	0.9773	0.9472	0.9416	0.9365	0.4563	0.4514	0.4384	0.4040	0.4040	0.5960
R^2	1.0000	0.9441	1.0000	0.9225	0.9088	0.9441	0.9225	0.9088	0.0886	0.0886	0.0298	0.0000	0.0000	1.0000
Anomaly Correlation Coefficient	1.0000	0.9437	1.0000	0.9219	0.9081	0.9437	0.9219	0.9081	0.0879	0.0879	0.0296	0.0000	0.0000	1.0000
Mean Absolute Percentage Error	1.0000	0.9937	0.9269	0.9739	0.9719	0.9205	0.9032	0.9006	0.3543	0.3407	0.3437	0.2450	0.2450	0.7550
Nash-Sutcliffe Efficiency	1.0000	0.9561	0.9251	0.9373	0.9263	0.8810	0.8657	0.8552	0.2507	0.2473	0.2051	0.1803	0.1803	0.8197
Spectral Angle	1.0000	0.9876	0.9595	0.9810	0.9773	0.9472	0.9416	0.9365	0.4563	0.4514	0.4384	0.4040	0.4040	0.9584
Error Total	21.815	21.451	20.937	20.721	20.634	20.527	19.958	19.857	7.876	7.776	7.252	5.605	5.508	