

---


Electronic Theses and Dissertations, 2004-2019

---

2015

## Drinking Water Infrastructure Assessment with Teleconnection Signals, Satellite Data Fusion and Mining

Sanaz Imen  
*University of Central Florida*

 Part of the [Civil Engineering Commons](#)  
Find similar works at: <https://stars.library.ucf.edu/etd>  
University of Central Florida Libraries <http://library.ucf.edu>

This Doctoral Dissertation (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2004-2019 by an authorized administrator of STARS. For more information, please contact [STARS@ucf.edu](mailto:STARS@ucf.edu).

---

### STARS Citation

Imen, Sanaz, "Drinking Water Infrastructure Assessment with Teleconnection Signals, Satellite Data Fusion and Mining" (2015). *Electronic Theses and Dissertations, 2004-2019*. 1138.  
<https://stars.library.ucf.edu/etd/1138>

**DRINKING WATER INFRASTRUCTURE ASSESSMENT WITH  
TELECONNECTION SIGNALS, SATELLITE DATA FUSION AND  
MINING**

by

SANAZ IMEN

B.Sc. University of Science and Culture, 2005

M.S. University of Tehran, 2007

A dissertation submitted in partial fulfillment of the requirements  
of the degree of Doctor of Philosophy  
in the Department of Civil, Environmental, and Construction Engineering  
in the College of Engineering and Computer Science  
at the University of Central Florida  
Orlando, Florida

Spring Term  
2015

Major Professor: Ni-Bin Chang

© 2015 Sanaz Imen

## **ABSTRACT**

Adjustment of the drinking water treatment process as a simultaneous response to climate variations and water quality impact has been a grand challenge in water resource management in recent years. This desired and preferred capability depends on timely and quantitative knowledge to monitor the quality and availability of water. This issue is of great importance for the largest reservoir in the United States, Lake Mead, which is located in the proximity of a big metropolitan region - Las Vegas, Nevada. The water quality in Lake Mead is impaired by forest fires, soil erosion, and land use changes in nearby watersheds and wastewater effluents from the Las Vegas Wash. In addition, more than a decade of drought has caused a sharp drop by about 100 feet in the elevation of Lake Mead. These hydrological processes in the drought event led to the increased concentration of total organic carbon (TOC) and total suspended solids (TSS) in the lake. TOC in surface water is known as a precursor of disinfection byproducts in drinking water, and high TSS concentration in source water is a threat leading to possible clogging in the water treatment process. Since Lake Mead is a principal source of drinking water for over 25 million people, high concentrations of TOC and TSS may have a potential health impact. Therefore, it is crucial to develop an early warning system which is able to support rapid forecasting of water quality and availability.

In this study, the creation of the nowcasting water quality model with satellite remote sensing technologies lays down the foundation for monitoring TSS and TOC, on a near real-time basis. Yet the novelty of this study lies in the development of a forecasting model to predict TOC and TSS values with the aid of remote sensing technologies on a daily basis. The forecasting process is aided by an iterative scheme via updating the daily satellite imagery in concert with

retrieving the long-term memory from the past states with the aid of nonlinear autoregressive neural network with external input on a rolling basis onward. To account for the potential impact of long-term hydrological droughts, teleconnection signals were included on a seasonal basis in the Upper Colorado River basin which provides 97% of the inflow into Lake Mead. Identification of teleconnection patterns at a local scale is challenging, largely due to the coexistence of non-stationary and non-linear signals embedded within the ocean-atmosphere system. Empirical mode decomposition as well as wavelet analysis are utilized to extract the intrinsic trend and the dominant oscillation of the sea surface temperature (SST) and precipitation time series. After finding possible associations between the dominant oscillation of seasonal precipitation and global SST through lagged correlation analysis, the statistically significant index regions in the oceans are extracted. With these characterized associations, individual contribution of these SST forcing regions that are linked to the related precipitation responses are further quantified through the use of the extreme learning machine. Results indicate that the non-leading SST regions also contribute saliently to the terrestrial precipitation variability compared to some of the known leading SST regions and confirm the capability of predicting the hydrological drought events one season ahead of time. With such an integrated advancement, an early warning system can be constructed to bridge the current gap in source water monitoring for water supply.

## ACKNOWLEDGMENTS

I would like to express my outstanding gratitude to my advisor, Dr. Ni-Bin Chang for guiding and supporting me over the past three and a half years. I would not have been able to finish my PhD program without his help and guidance.

I also wish to thank the members of my dissertation committee, Dr. Martin Wanielista, Dr. Patrick Bohlen, and Dr. Dingbao Wang for their constructive criticism.

I would like to thank my research colleagues: Zhemin Xuan, Zachary Marimon, Riz Islam, Jessica Cormier, Cameron Houmann, Golam Mohiuddin, James Crawford, Nick Hartshorn, Subrina Tahsin, and Justin Joyce. In particular, I would like to thank Kaixu Bai, Carolina Doña Monzó, Jamie Jones, Ben Vannah, and Lee Mullon. They've become more than colleagues, they are my best friends and I consider them a part of my extended family.

Special thanks to my fiancé, who was always standing by me, with his love and support, in my hard time during this work.

Finally, I would like to thank my family in Iran for their love and support. I would like to dedicate this dissertation to my grandmother, my mom, my dad, and my sister for all the love, support, and encouragement that they have given me over the years. I undoubtedly could not accomplish this goal without you.

# TABLE OF CONTENTS

LIST OF FIGURES .....	viii
LIST OF TABLES .....	xii
CHAPTER 1 : INTRODUCTION .....	1
1.1. Data Fusion and Data Mining .....	2
1.2. Hydroclimatic Teleconnection .....	8
1.3. Research Objectives .....	11
1.4. Dissertation Organization .....	12
1.5. References .....	14
CHAPTER 2 : ATTRIBUTIONAL ASSESSMENT OF LOCAL TELECONNECTION SIGNALS DURING RECENT DROUGHTS IN THE UPPER COLORADO RIVER BASIN TOWARD LONG-TERM PRECIPITATION FORECASTING .....	18
2.1. Introduction .....	18
2.2. Study Area .....	24
2.2.1. Precipitation Regimes over the Upper Colorado River Basin .....	25
2.3. Data and Method .....	29
2.3.1. Data Collection .....	30
2.3.2. Phase I: Identification the Possible Index Regions .....	33
2.3.3. Phase II. Screening and Ranking the Identified Index Regions .....	37
2.3.4. Phase III. Seasonal Precipitation Forecasting .....	40
2.4. Results .....	44
2.4.1. Results of Phase I: Identification the Possible Index Regions .....	44
2.4.2. Results of Phase II: Screening and Ranking the Identified Oceanic Index Regions 49	
2.4.3. Results of Phase III: Seasonal Precipitation Forecasting .....	53
2.5. Application .....	57
2.6. Conclusion .....	59
2.7. Reference .....	61
CHAPTER 3 : DEVELOPING THE REMOTE SENSING-BASED EARLY WARNING SYSTEM FOR MONITORING TSS CONCENTRATIONS IN LAKE MEAD .....	67
3.1. Introduction .....	67
3.2. Study Area .....	70
3.3. Input Data .....	73
3.3.1. Ground-truth Data Collection and Analysis .....	73

3.3.2.	Remote Sensing Data Collection and Pre-processing.....	74
3.4.	Methodology .....	76
3.4.1.	Phase I: Nowcasting.....	77
3.4.2.	Phase II: Forecasting.....	78
3.5.	Results and Discussion.....	86
3.5.1.	Nowcasting Results.....	86
3.5.2.	TSS Concentrations at the Location of Inflows into the Lake.....	89
3.5.3.	Forecasting Results.....	100
3.6.	Final Remarks .....	104
3.7.	Conclusions .....	105
3.8.	References .....	106
CHAPTER 4 : MULTI-SENSOR FUSION FOR MONITORING TOTAL ORGANIC CARBON CONCENTRATIONS IN LAKE MEAD TOWARD WATER QUALITY FORECASTING.. 113		
4.1.	Introduction .....	113
4.2.	Problem Identification.....	115
4.3.	Data Collection and Analysis.....	116
4.3.1.	In-Situ Data.....	116
4.3.2.	Satellite Data.....	119
4.4.	Data Synthesis Based On Remote Sensing Imageries .....	119
4.4.1.	Nowcasting Method.....	119
4.4.2.	Forecasting Method .....	122
4.5.	Results and Discussion.....	124
4.5.1.	Nowcasting Results.....	124
4.5.2.	Forecasting Results.....	133
4.6.	Final Remarks .....	135
4.7.	Conclusions .....	136
4.8.	Reference.....	138
CHAPTER 5 : GENERAL CONCLUSIONS AND STUDY LIMITATIONS..... 144		
5.1.	Conclusions .....	144
5.2.	Study Limitations and Recommendations .....	146
5.3.	Reference.....	147
APPENDIX: MATERIALS UNDER REVIEW .....		
		148



## LIST OF FIGURES

Figure 1-1. The chronological trend of inversion methods ( <sup>1</sup> ANN: Artificial Neural Network, <sup>2</sup> GP: Genetic Programming, <sup>3</sup> RBFN: Radial Basis Function Neural Network Models, <sup>4</sup> GEGA: Grammatical Evolution in Genetic Algorithm, <sup>5</sup> GA-PLS: Genetic Algorithms and Partial Least Square( obtained from: Chang et al., 2015) .....	8
Figure 1-2. Organizational framework of this dissertation from the aspect of the research objectives.....	13
Figure 2-1. The location of Colorado River Basin in the U.S. The area marked by bold line is the present study area. The size of circles shows the magnitude of average annual flow at the location of different stations. A larger circle shows large magnitude of average annual flow. ....	24
Figure 2-2. Annual cycle of precipitation in the Upper Colorado River Basin (1997-2014). Bars indicate monthly standard deviations. ....	26
Figure 2-3. (a) winter season (December to February); (b) spring season (March to May); (c) summer season (June –August); (d) fall season (September to November); Dashed lines indicate $\pm 1$ stdv for the winter (8.58 mm), spring (8.18 mm), summer (8.03 mm), fall (9.40 mm); Middle dashed line indicates average for winter (28 mm), spring (32 mm), summer (29 mm), fall (29 mm). ....	27
Figure 2-4. Calculated SPI for monthly precipitation in the period of (1997-2014). Y-axis shows the defined category of SPI. The definition corresponding to each vale in y-axis are shown at the top of the diagram. ....	29
Figure 2-5. The study framework .....	31
Figure 2-6. Wavelet decomposition example of a single SST grid in fall season; (a) seasonal anomaly SST data with the X and Y axes representing time and SST in degrees, respectively. (b) SST wavelet power spectrum; (c) global wavelet spectrum; (d) SAWP time series which was calculated based on the 2-4 year period band. The dashed blue lines show the boundary of cone of influence. ....	36
Figure 2-7. Typical scheme of single layer feedforward neural networks (Chang et al., 2010)...	39
Figure 2-8. The framework of the applied SPI-based algorithm to determine the intensity of hydrological droughts a season ahead of time.....	42
Figure 2-9. Comparison between the raw and detrended time series for a single grid in fall season: (a) Precipitation data; (b) SST data .....	45
Figure 2-10. Spatial correlation between PC1 of seasonal precipitation and the SAWP of individual grid: (a) DJF, (b)MAM, (c) JJA, (d) SON. ....	46

Figure 2-11. Pixel-wise linear correlations between PC1 of precipitation SAWPs and SST SAWP of each grid in fall season. Shaded color represents significant correlation at 95% confidence intervals. The areas marked with rectangles show the identified significant regions. ....	48
Figure 2-12. Example of determining time lag for each oceanic index region. Right figure shows the correlation of the Index7 with PC1 of precipitation SAWPs in different time lags. ....	49
Figure 2-13. Comparison between the output and the target of the ELM algorithm for (a) spring; (b) summer, (c) fall, (d) winter. ....	50
Figure 2-14. Oceanic index regions which shows the highest contribution on precipitation variability. “S” stands for summer and spring, “F” and “W” stands for fall and winter, respectively. ....	51
Figure 2-15. Comparison between the SST of the identified oceanic index regions and the known teleconnection pattern in (a) fall; (b) summer; (c) spring. ....	52
Figure 2-16. Spatial comparison between predicted and observed SPI. ....	56
Figure 2-17. The current procedure of water supply forecasts on the Colorado River and the application of forecasts. ....	58
Figure 3-1. Location of Lake Mead and all rivers inflow into the lake. ....	70
Figure 3-2. Location of recorded forest fire events (right) and the percentage of forest fire events which occurred in each year (lower left) in the Virgin River Watershed – (Data source: USDA Forest Service, 2014). ....	71
Figure 3-3. TSS sampling stations by sampling agencies (Data source: Lower Colorado River regional water quality database). ....	73
Figure 3-4. Schematic diagram of the developed human-machine system in this study. ....	76
Figure 3-5. Methodological flowchart of the integrated data fusion and data mining method (obtained from Chang et al., 2014a). ....	78
Figure 3-6. The framework of the developed water quality forecasting method. ....	82
Figure 3-7. The diagram of NARXNET (Obtained from MATLAB R2013a). ....	84
Figure 3-8. The framework of the early warning system based on integrated nowcasting and forecasting analysis. ....	85
Figure 3-9. Vertical profile of TSS at the location of drinking water intake in Lake Mead. ....	86
Figure 3-10. Temporal distribution of TSS concentration at the location of inflows into the lake; white represents “no data” value as a result of inaccessibility of cloud-free Landsat images. ....	90
Figure 3-11. Temporal distribution of streamflow at the location of USGS stations with associated color legend for each arm. ....	92

Figure 3-12. Temporal distribution of the mean monthly TSS load (mg/s) at each arm; White color represents “no data” values as a result of inaccessibility of cloud-free Landsat images ..... 93

Figure 3-13. Comparison between temporal variations of TSS concentration and Lake Mead at Hoover dam elevation at (A) CRLM (B) VR25.1 (C) LVB2.7. Green line shows the trend of TSS concentrations, and red line indicates the variation of elevation during (2005-2009). . 94

Figure 3-14. Estimating TSS concentration maps using IDFM Method from 19 Sep 2008 to 25 Sep 2008. The top row in this figure shows the available MODIS images and the second row indicates the available Landsat images in the Lake Mead for the cloud-free days of the selected time period. Third and the last rows show fused images and daily TSS concentration maps, respectively. Since using the atmospheric correction algorithm changes the surface reflectance values in MOD09 to -100 –16,000 [27], pixels with negative surface reflectance values were categorized as white in this figure. .... 97

Figure 3-15. Estimating TSS concentration maps using IDFM Method from 12 May 2009 to 20 May 2009. The top row in this figure shows the available MODIS images and the second row indicates the available Landsat images in the Lake Mead for the cloud-free days of the selected time period. Third and the last rows show fused images and daily TSS concentration maps, respectively. Since using the atmospheric correction algorithm changes the surface reflectance values in MOD09 to -100 –16,000 [27], pixels with negative surface reflectance values were categorized as white in this figure ..... 98

Figure 3-16. Estimating TSS concentration maps using IDFM Method from 7 February 2006 to 13 February 2006. The top row in this figure shows the available MODIS images and the second row indicates the available Landsat images in the Lake Mead for the cloud-free days of the selected time period. Third and the last rows show fused images and daily TSS concentration maps, respectively..... 99

Figure 3-17. Comparing ground-truth data with NARXNET results in different scenarios. The first two dates belong to training and validation period and the last date is the prediction date. .... 102

Figure 4-1. TOC concentration compared to EPA advice note on DBPs in drinking water criteria. A) Temporal changes of TOC, white color show no data values; B) spatial variation of TOC, Boulder Basin (square), Overturn Arm (diamond), and Colorado River Arm (triangle).... 117

Figure 4-2. Vertical profile of TOC in Lake Mead at the location of the closest station to the drinking water intake ..... 118

Figure 4-3. IDFM Methodological Flowchart ..... 120

Figure 4-4. Water quality forecasting framework..... 123

Figure 4-5. The ability of IDFM in gap-filling of Landsat images. The fused image is shown in the dashed box, and all images are true color (RGB) images. .... 124

Figure 4-6. Comparison between the observed values and the estimated values of TOC concentration from 2000-to 2013 via two data mining models (A) ANN model, (B) GP model ..... 126

Figure 4-7. Spatial variations of (A) TOC concentration on June 20th, 2007, (B) TOC concentration on April 3rd, 2002. TOC concentration is categorized based on the EPA advice note on DBPs in drinking water. Pixels with negative surface reflectance values were shown in white..... 127

Figure 4-8. Estimating TOC concentration maps using IDFM Method from 6 May 2001 (DOY 126) to 14 May 2001(DOY 134). The top row in this figure shows the available MODIS images and the second row indicates the available Landsat images in the Lake Mead for the cloud-free days of the selected time period. Third and the last rows show fused images and daily TSS concentration maps, respectively (Obtained from: Imen et al., 2014) ..... 128

Figure 4-9. Temporal distribution of TOC concentration at the location of inflows into the lake; white represents “no data” value as a result of inaccessibility of cloud-free Landsat images ..... 130

Figure 4-10 Temporal distribution of streamflow at the location of USGS stations with associated color legend for each arm; white represents “no data” value ..... 131

Figure 4-11 Comparison between temporal variations of TOC concentration and Lake Mead elevation at Hoover dam at the location of stations: (A) CRLM (B) VR25.1 (C) LVB2.7. Green line shows the trend of TSS concentrations, and red line indicates the variation of elevation during (2000-2013)..... 132

## LIST OF TABLES

Table 1-1. Advantages and disadvantages of different data fusion algorithms .....	6
Table 2-1. Partial list of references focusing on the effects of regional SST's variability on hydroclimate variables in the Colorado River Basin.....	20
Table 2-2. Correlation between precipitation and inflow into the reservoirs within different time scales.....	28
Table 2-3. The SPI classification system (McKee te al., 1993).....	29
Table 2-4. Minimum and maximum precipitation (mm/month) of all precipitation grids within the study area. The selected year is shown in gray.....	55
Table 2-5. Comparison between the observed and predicted values over the study area in year 2012 (mm/month).....	56
Table 3-1. Ground-truth data within the Lake utilized by this study (Source: lower Colorado River regional water quality).....	74
Table 3-2. Number of satellite images for data fusion with selected bands utilized in this study	76
Table 3-3. Selected forecasting methods in water resources management and their pros and cons .....	80
Table 3-4. R_squared value between the spectral reflectance value of the fused image and the actual Landsat image .....	87
Table 3-5. Comparison between the observed TSS values and the estimated values from the trained ANN model .....	88
Table 3-6. Detected dates of hot spots ( $TSS \geq 25 \text{ mg.L}^{-1}$ ) at the location of each station- .....	91
Table 3-7. Summary of database used in different scenarios .....	101
Table 3-8. Summary of the results of error indices; Values of RMSE, PBIAS, and MAE equal to zero, and R-squared value equal to one show a perfect fit.....	104
Table 4-1. TOC Ground-truth data within the Lake utilized by this study.....	116
Table 4-2. Number of satellite images and selected bands utilized in this study .....	119
Table 4-3. Summary of database used in different scenarios .....	123
Table 4-4. R_squared value between the spectral reflectance value of the fused image and the actual Landsat image on 19-Oct-2006.....	125
Table 4-5. Comparison between observed values at the sampling location with the estimated values for TOC on April 3rd, 2002 and June 20th, 2007. ....	127

Table 4-6. Comparison between the estimated and observed values of TOC at LVB 6.7 during Aug14th –Oct 13th 2001 .....	134
Table 4-7. Summary of the results of error indices; Values of RMSE, PBIAS, and MAE equal to zero, and R-squared value equal to one show a perfect fit.....	134
Table 4-8. Comparison of predicted values in different scenarios .....	135

## **CHAPTER 1: INTRODUCTION**

Climate change will most likely increase the demand for water and decrease the supply. This shifting balance in lakes and reservoirs would challenge water managers to provide adequate water to the growing population, agricultural, ecological, and industrial sectors. Although much attention is focused on the impacts of climate change on the quantity of water supplies, the potential detrimental impacts of climate change on water quality also are of concern. Climate change can affect the quality and quantity of lakes and reservoirs by (1) changing the timing of snowmelt and inflow into the lake/reservoir, (2) changing the intensity and duration of precipitation events which cause washing more pollutants into water supply, (3) intensifying the concentration of contaminants during dry periods.

Climate variability and extreme events have led decision makers to pursue the source water assessment on lakes and reservoirs for adaptation. In modern smart cities, there is a need to utilize advanced data science, computational intelligence, and satellite remote sensing to collectively promote the sustainable operation of drinking water treatment plants in response to climate variability and water pollution. This study advances sustainability science by tackling the complexity embedded in the dynamics of the coupled natural systems and the built environment using data fusion and mining, and hydroclimatic teleconnection to forecast the quality and availability of water, respectively. With the aid of satellite data fusion and data mining, such a forecasting scheme can be made possible to address the aforementioned concerns.

## 1.1. Data Fusion and Data Mining

“More than 20,000 waterbodies across America have been identified as polluted waterbodies including more than 480,000 km (300,000 miles) of rivers and shoreline and 5 million acres of lakes” (Ritchie and Cooper, 2001). These statistics highlights the importance of water quality monitoring, and considering the effects of different factors such as ocean-atmosphere exchange, atmosphere decomposition, industrial outfalls, urban run-off, agricultural run-off, contaminant migration through groundwater, and sediment resuspension in water bodies. Among the aforementioned factors, ocean-atmosphere exchange, which includes the effects of global warming and consequences of extreme events such as droughts and floods, can increase different water quality parameters such as basic parameters (temperature, pH, dissolved oxygen,...), nutrients, dissolved organic matter (DOM), inorganic/organic micropollutants, pathogens, and cyanobacteria. For instance, water temperature can affect the internal lake process such as diffusion, mineralization, and vertical mixing, and lakes chemical process and increasing alkalinity and pH (Psenner and Schmidt, 1992; Malmaues et al., 2006). There is a positive correlation between temperature and nutrient concentrations in surface and groundwater (Delpa et al, 2009). Increase of both temperature and nutrient concentration leads to cyanobacteria blooms in water bodies (Hunter, 2003). In addition, higher precipitation intensity increases run-off and erosion which result in increase in external nutrient transports and TSS (Delpa et al., 2009). On the other hand, these hydrological parameters (i.e. precipitation and temperature) are the main climate drivers for changing the fate of organic micropollutants (Delpa et al., 2009).

The traditional techniques to assess water quality parameters are time consuming for a large area of water bodies and costly as a result of travel expenses, sampling equipment, and



sample analysis. Therefore, remote sensing has gained attention, since they are able to solve the problems of traditional techniques by providing ground and ocean monitoring with larger spatial and temporal resolution. The backscattering characteristics of surface water can be varied by the existence of substances in surface water. Remote sensing techniques became an applicable tool to measure the variation of spectral reflectance backscattered (Ritchie et al., 2003).

Since the 1970s, earth observations via remote sensing satellites have been providing enormous amount of information in the form of images for monitoring various environmental status and ecosystem state. Remote sensing can provide the images, such as (i) visible (i.e., in the range 0.4-0.7  $\mu\text{m}$ ); (ii) near infrared (NIR: 0.7-0.9  $\mu\text{m}$ ); (iii) shortwave infrared (SWIR: 1.0-2.5  $\mu\text{m}$ ); (iv) thermal infrared (TIR: 3-14  $\mu\text{m}$ ); and (v) microwave (i.e., 3 mm-3 m). These images can be categorized into two classes on the bases of their spectral characteristics, including multispectral and hyperspectral remote sensing. The relatively high spatial resolution images would normally have low temporal resolution and vice versa. Some environmental issues (e.g., forest fires, agricultural droughts, urban flooding, harmful algal bloom, and irrigation management) require high spatial (e.g., 30 m), high temporal (e.g., daily), and high spectral (e.g., hyperspectral) resolutions due to the needs for rapid changes detection in emergency response and mitigation actions. To achieve such goals, data fusion and mining becomes an indispensable tool providing rapid hindcasting, nowcasting, and forecasting information for risk-informed decision making.

The classical definition of data fusion produced by the Joint Directors of Laboratories is: “A process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete timely assessments of situations and threats as well as their significance” (White, 1991) . The

major application of data fusion is related to image processing. Multiple images of the same scene are integrated by different fusion methods to present a more complete image entitled fused images.

In the mid-1980s, image fusion has become a focal area in remote sensing community. These image fusion techniques were classified into two categories: (1) spatial domain fusion techniques, and (2) transform domain fusion techniques. The former deals with the image pixels whereas the latter transfers the image into frequency domain for feature extraction.

The spatial domain fusion techniques can be further divided into (1) color composition-based method (i.e., intensity-hue-saturation (IHS)), and (2) statistical/numerical method (i.e. Brovey transform, principle component analysis (PCA), high-pass filtering (HPF)). Gillespie et al. (1987) suggested the Brovey method which provides enhancement of color in the image. However, this method can cause major changes to the statistical parameters and colors of the original images. For example, IHS method was utilized in the field of remote sensing by Zhou et al. (1998) in the 1990s. They found out that although the fused images showed the ability of the IHS in providing an enhancement in spatial resolution, their original spectral information could be distorted. At the same time, the feature of dimensionality reduction of PCA was also employed for image fusion. The PCA is able to enhance image and benefit data compression. Given that information of the spectral bands cannot be conserved after applying PCA, there is a possibility of losing important information by using this method too. In 1991, Chavez et al. compared the fused images obtained from the PCA and the IHS methods with the HPF introduced by Schowengerdt (1980). Results showed that the HPF had the least spectral distortion.

Despite the spatial domain fusion methods present quite good visual effect, these methods resulted in spectral distortion in the fused images. As an alternative, the transform domain fusion technique was developed and applied due to their capability to decompose signals into frequency spaces and minimize the spectral distortion. This technique includes pyramid method, wavelet transform, and curvelet transform. Early work in image fusion using this method traces back to the 1980s. Burt and Kolczynski (1993) showed the application of pyramid method in image fusion can guarantee stability and noise immunity. In the late 1990s, the pyramid method was substituted with wavelet transform that can keep the spectral characteristics of multi-frequency images. However, among the transform domain fusion methods, curvelet method showed the best performance in keeping the spectral quality. Contrast to wavelet transform, curvelet is more powerful in delineating edges smoothly in two-dimensional platform and denoising in matching edges from images.

Overall, on one hand, the ability of spatial domain fusion methods may provide high spatial resolution images, and on the other hand, the ability of the transform domain fusion can minimize the distortion of spectral contents of images. These pros and cons led the community to develop hybrid methods incorporating both characteristics in the fused images. Even though these hybrid methods produce synthetic images with helpful spectral information and high spatial resolution by blending high spatial resolution of panchromatic bands, they are not able to enhance spatial and temporal resolutions simultaneously. To bridge the gap, in the early 2000s, the series of STAR-FM-based methods were developed to blend the coarse-resolution of MODIS with the high spatial resolution of Landsat and produce fused images with both high spatial and temporal resolution. The advantage and disadvantages of different STAR-FM-based methods are shown in Table 1-1.

**Table 1-1. Advantages and disadvantages of different data fusion algorithms**

<b>Method</b>	<b>Advantage</b>	<b>Disadvantage</b>	<b>Reference</b>
STAR-FM	Able to record the transition from one state to another.	Cannot predict disturbance events Any changes that are too subtle to be observed by the MODIS observations are not predictable with this algorithm. STARFM method does not explicitly handle the directional dependence of reflectance.	Gao et al., 2006
STAARCH	More capable in detecting the spatial and temporal changes in the landscape.	Prediction results degrade somewhat when used on heterogeneous fine-grained landscapes. Its ability in detecting landscape disturbance is limited to specific types of land-cover change.	Hilker et al., 2009
ESTARFM	ESTARM is more suitable for heterogeneous regions than STARFM owing to the use of “conversion coefficient” between the Landsat and MODIS data. Improve the accuracy of predicted fine-resolution reflectance	It is built under the assumption that land-cover types and the proportion of each land cover type do not change during the observation period. It cannot accurately predict objects for which shape changes with time. It cannot accurately predict short-term, transient changes Sensors with different spectral band passes may lead to nonlinear relationships. Compared with the original STARFM algorithm, the ESTARFM may be more computationally intensive and require at least two pairs of fine- and coarse-resolution images acquired at the same date Its automated processing is limited by necessity of setting the size of moving window and the number of classes	Zhu et al. 2010
SPSTFM	Accounts for all the reflectance changes during an observation period Capturing both the phenology and type change	Longer computation time than STAR-FM algorithm In many remote sensing applications, only one pair of prior images may be available.	Huang and Song, 2012

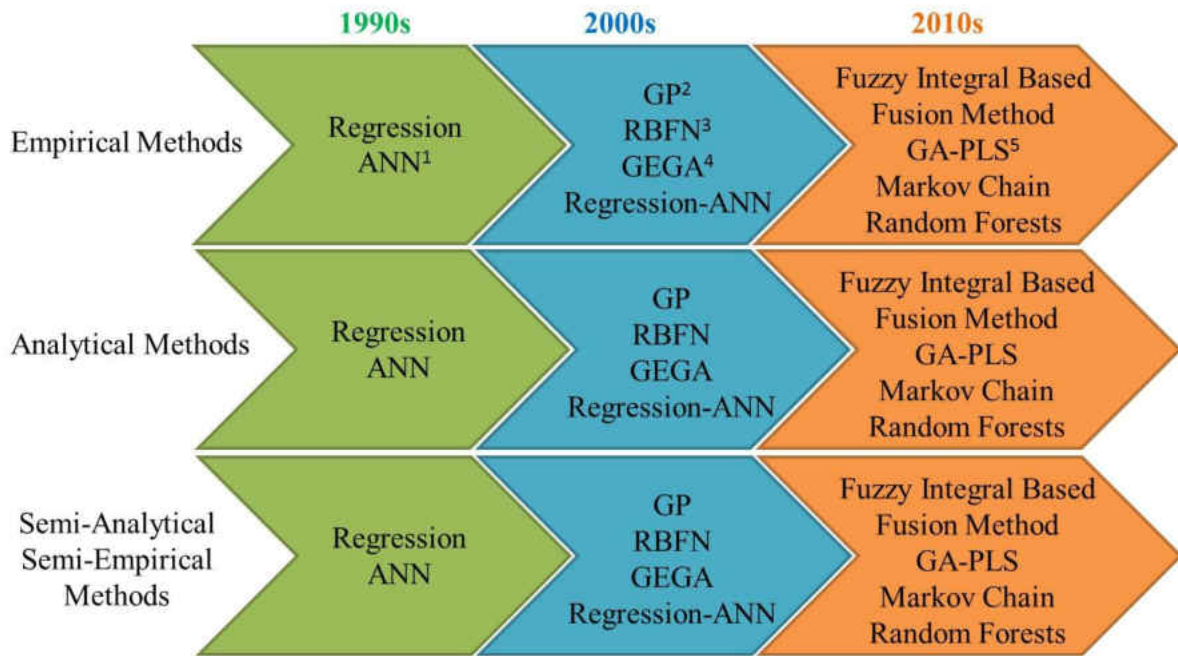
Data fusion methods create a new challenge of analyzing the massive data sets within fused images. For this purpose, data mining methods, as the automated, exploratory data analysis tool, were applied to identify the features of interest from fused images. The inversion models that are commonly used to find relationships between target environmental parameters and

remotely sensed data can be categorized into three types: empirical, analytical, and semi-empirical/semi-analytical methods. Appropriate selection of these inversion models based on their applications contributes to higher accuracy in evaluating the environmental parameters. These methods vary in complexity from simplified statistical based to physically based models (Lek et al., 1996).

Simplest methods started with empirical models which are based on the simple statistical relationship between the measured water quality parameters and spectral reflectance values. Since many parameters may change during remote sensing missions, these methods dependent on ground-truth data and their results does not have multi-temporal validity. On the other hand, analytical methods involve physical relationship between the target environmental parameters and the spectral reflectance values (Lek et al., 1996). Therefore, the hybrid methods entitled semi-empirical/semi-analytical methods were developed to improve the multi-temporal validity of relationships between the environmental parameters and spectral reflectance values. These hybrid models are based on the integration of empirical and analytical methods and can be expressed as empirical equations in which the optical theory in physics or radiative transfer theory can be employed (Chang et al., 2015).

The application of empirical methods is limited to a specific water body for which sampling data were collected and the formulas were derived. Therefore, its accuracy will be degraded in a case of using the same empirical model for a water body with different types of water quality parameters (Yu, 2005; Wang et al., 2003). However, the results of analytical and hybrid methods are more reliable for the transferable cases between different water bodies and

variety types of water quality parameters (Dekker et al., 1996). The chronological order of different inversion models is presented in Figure 1-1.



**Figure 1-1. The chronological trend of inversion methods ( <sup>1</sup>ANN: Artificial Neural Network, <sup>2</sup>GP: Genetic Programming, <sup>3</sup>RBFN: Radial Basis Function Neural Network Models, <sup>4</sup>GEGA: Grammatical Evolution in Genetic Algorithm, <sup>5</sup>GA-PLS: Genetic Algorithms and Partial Least Square( obtained from: Chang et al., 2015)**

## 1.2. Hydroclimatic Teleconnection

Climate change can intensify the global water cycle, which increases the intensity and frequency of extreme hydrological events and change the spatial and temporal availability of water resources (Huntington, 2006). The significant relationship between hydrological variables and the atmospheric circulation can be modeled through two different approaches: (1) simulation via General Circulation Models (GCMs), and (2) determining the statistical relationship between

hydrological variables and the atmospheric/oceanic variables separated by large distances, namely ‘hydroclimatic teleconnection’.

Although GCMs have been widely utilized to detect and attribute changing climate, to some extent, they are still limited by several factors: (1) relevant feedbacks between the land surface and the atmosphere are not included in the initial scenarios derived from GCMs (Arnell et al., 1996), and (2) the spatial resolution of GCMs, usually with grid-size over  $2.5^{\circ} \times 2.5^{\circ}$ , is too coarse to be applied in a regional study (Arnell et al., 1996). Due to the resolution limitation of GCMs, they cannot be used to model the terrain effects and mesoscale feedbacks in a case of complex terrain (Zhang et al., 2012). Within this context, various downscaling methods have been developed to improve the spatial/temporal resolution of GCMs output at the regional scale (Teutschbein et al., 2011). However, limitations and uncertainties of dynamical downscaling methods still exist, since they are not able to address the hydrological variability and extremes well (Chen et al., 2011). In addition, dynamical downscaling methods only can be applied for a short period of time (10-20 years) due to a large computational requirement, which are also limited to address a single or a few GCMs outputs. On the other hand, large bias has been reported between GCMs simulated scenario and the observations (Monckton et al., 2015).

In recent years, hydroclimatic teleconnection has received significant attention because of its potential application in analyzing hydrologic variability and addressing uncertainty associated with hydrological variables (Government of India Ministry of Water Resources, 2008). Moreover, it provides a basis for understanding the response of local-scale meteorology to changing climate patterns (Hodson et al., 2009). However, the non-stationary nature of the real world signals makes it difficult to identify the teleconnection if only relying on simple linear

analysis across a wide region at sub-continental scale, as weak teleconnection signals might be overwhelmed by anthropogenic effects. This effect can be intensified in areas over the Continental United States in which the impacts of climate change are not homogenous across the regions. This heterogeneity makes it even much more challenging and complicated.

Some statistical approaches such as multivariate Auto Regressive Moving Average (ARMA), Fourier, linear regression, and cross correlation are incapable of characterizing the non-linear and non-stationary nature of climatic signals. Such difficulties have triggered the use of models that can deal with non-stationarity such as maximum covariance analysis (MCA), wavelet analysis, wavelet empirical orthogonal function (WEOF), and Hilbert-Huang transform (HHT) (Mwale et al., 2004; Salim and Pawitan, 2007; Mullon et al., 2013; Shen et al., 2013). In addition, machine learning models were applied as powerful tools to overcome the problem of non-linearity in hydrology (Abbot and Marohasy, 2012). One of the most commonly used machine-learning models is artificial neural network (ANN) which has been in use in the field of water resources since the 1990s. Although ANN models are able to capture the complex non-linear nature of hydrological problems, the weaknesses of their input time series causes limited forecasting success (Mullon et al., 2015). Therefore, earlier studies (Wang and Ding, 2003; Mwale et al., 2004; Cannas et al., 2005; Kisi, 2008; Partal and Cigizoglu, 2009; Elsanabary and Gan, 2014; Mullon et al., 2015) applied different methods such as wavelet analysis, WEOF, HHT, and MCA to pre-treat the input time series into machine learning models in order to improve forecasting success.



### 1.3. Research Objectives

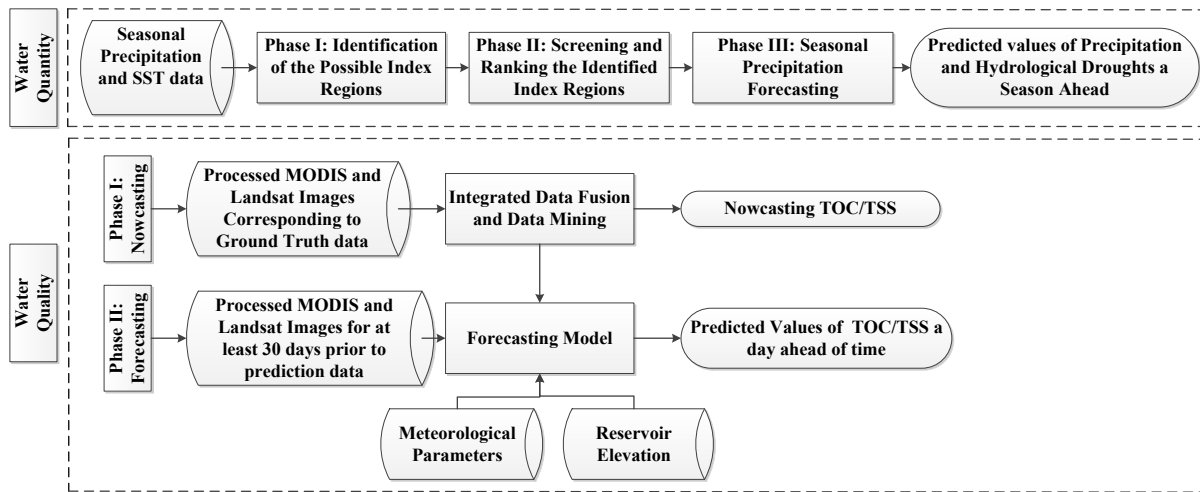
Creating secure, resilient, and sustainable infrastructure systems for fast growing urban regions is by no means an easy task. The novelty of this study lies at the intersection of the broadly-defined water resources engineering, environmental engineering, and earth system sciences, which assess a suite of diversified impacts on drinking water infrastructures. The major objectives of this dissertation are: (1) developing a remote sensing-based early warning system to predict TSS and TOC concentration at the drinking water intake in Lake Mead one day ahead, and (2) developing a diagnostic technique to assess hydrological extremes over the Upper Colorado River Basin one season ahead of time. The engineering community will benefit from having such early warning systems to manage the impact of a variety of natural hazards on water infrastructure systems at the same time. The developed algorithm would be beneficial for drinking water treatment plant operators to minimize the negative impacts on water quality before they occur, or otherwise achieves an actionable strategy either in plant or at the source water location due to unexpected events. The following specific objectives are formulated:

- (1) Developing a nowcasting model to monitor the spatiotemporal distribution of TOC and TSS concentration within Lake Mead on a daily basis.
- (2) Applying the nowcasting model to find the probable impacts of each pollutant source (i.e. transported sediments from Colorado River, forest fire events in the lower Virgin River Watershed, and transported urban stormwater runoff and treated wastewater effluent from Las Vegas Wash) on the increase of TSS and TOC concentration in the lake.
- (3) Developing a water-quality forecasting model to estimate water quality constituents (i.e. TSS and TOC) at the location of drinking water intake one day ahead of time.

- (4) Developing a diagnostic technique based on hydroclimatic teleconnection to estimate the spatial distribution of precipitation for a season ahead of time and assess the potential of hydrological extremes over the Upper Colorado River Basin three months ahead.

#### 1.4. Dissertation Organization

The organization of this dissertation from the aspect of research objectives is shown in Figure 1-2. As shown in this figure, this study focuses on two topics including water quality and water quantity. Chapter 2 focuses on the first topic (i.e. water quantity) through three phases. Phase 1 is based on assessing the use of hydroclimatic-teleconnection to find the oceanic index regions which show a statistically significant correlation with precipitation variability in different seasons over the Upper Colorado River Basin. Then, the identified oceanic index regions are ranked and screened in phase 2 in order to find the index region which has the highest contribution on precipitation variability over this region. The last phase is associated with the seasonal precipitation forecasting which leads into estimating the potential occurrence of hydrological extremes in this region for a season ahead. The output of the second chapter of this dissertation will be a standardized precipitation index (SPI) map which is generated to visually identify the spatial distribution of hydrological events over the Upper Colorado River Basin in each upcoming season. In addition, this map is compared with the observed value of SPI in year 2012 to validate the performance of the developed forecasting method.



**Figure 1-2. Organizational framework of this dissertation from the aspect of the research objectives.**

Chapters 3 and 4 of this dissertation are associated with monitoring and forecasting TSS and TOC concentration in Lake Mead, respectively. As shown in Figure 1-2, the first phase of the water quality section is associated with the nowcasting model. The nowcasting model is based on the integrated data fusion and data mining method developed by previous studies (Chang et al., 2014a; Chang et al., 2014b) to fuse MODIS and Landsat images and generate a synthetic image which has both high spatial and temporal resolutions. Results of phase 1 create daily concentration maps to depict the spatial-temporal distribution of TOC/TSS within the lake. These concentration maps were applied to provide information about the main pollutant sources of TSS/TOC and the relationship of TSS/TOC with increase in volumetric flow rate. The trained machine learning model in the phase 1 provides us with the daily TOC/TSS time series which was applied as inputs into the forecasting model in phase 2 as well as meteorological variables and reservoir elevation. The developed forecasting model provides the operator of the water treatment plant with the concentration value of TOC/TSS one day ahead of time which enables the operator to tackle unexpected events before they occur.

## 1.5. References

- Abbot J., Marohasy J., 2012, Application of artificial neural networks to rainfall forecasting in Queensland, Australia, *Advances in Atmospheric Sciences*, vol. 29, no. 4: 717-730. doi: 10.1007/s00376-012-1259-9.
- Arnell N. W., Bates B., Lang H., Manguson J.J., Mulholland P., Fisher S., Liu C., McKnight D., Strarosolszky O., Taylor M., Aquize E., Arnotte S., Brakke D., Braun L., Chalise S., Chen C., Folt C. L., Gafny S., Hanaki K., Hecky R., Leavesly G.H., Lins H., Nemeč J., Ramasastri K. S., Somlyódy L., Stakhiv E., 1996, Hydrology and freshwater ecology. In: Watson, R T, Zinyowera M C, Moss R H (eds) *Climate Change 1995- Impacts, Adaptions and Mitigation of Climate Change: Scientific-Technical Analysis. Contribution of working group II to the second assessment of the intergovernmental panel on climate change.* Cambridge University Press, Cambridge, pp. 325-363.
- Burt P. J., Kolczynski R. J., 1993, Enhanced image capture through fusion, in 1993 (4th) International Conference on Computer Vision, pp. 173–182.
- Cannas B., Fanni A., Sias G., Tronci S., Zedda M.K., 2005, Streamflow forecasting using neural networks and wavelet analysis, *Env. Geosci. Union*, vol. 7, pp. 45-51.
- Chang N.B., Imen S., Vannah B., 2015, Remote sensing for monitoring surface water quality status and ecosystem state in relation to the nutrient cycle: a 40-year perspective, *Critical Reviews of Environmental Science and Technology*, vol. 45(2), pp. 101-166.
- Chavez J., Sides S. C., Anderson J. A., 1991, Comparison of Three Different Methods to Merge Multiresolution and Multispectral Data: LANDSAT TM and SPOT Panchromatic, *Photogramm. Eng. Remote Sensing*, vol. 57, no. 3, pp. 295–303.
- Chen J., Brissette F. P., Leconte R., 2011, Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology*, vol. 401, no. 3-4, pp.190-202. doi:10.1016/j.jhydrol.2011.02.020
- Dekker A. G., Zamurović-nenad Ž., Hoogenboom H. J., Peters S. W. M., 1996, Remote sensing, ecological water quality modelling and in situ measurements: a case study in shallow lakes, *Hydrol. Sci. J.*, vol. 41, no. 4, pp. 531–547.
- Delpa I., Jung A.V., Baures E., Clement M., Thomas O., 2009, Impacts of climate change in surface water quality in relation to drinking water production, *Environment International*, vol. 35, 1225-1233.
- Elsanabary M.H., Gan T.Y., 2014, Wavelet analysis of seasonal rainfall variability of the Upper Blue Nile Basin, its teleconnection to global sea surface temperature, and its forecasting by an artificial neural network, *Mon. Wea. Rev.*, vol. 142, pp. 1771-1791.
- Gao F., Masek J., Schwaller M., Hall F., 2006, On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance, *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218.

- Gillespie A. R., Kahle A. B., and Walker R. E., 1987, Color enhancement of highly correlated images. II. Channel ratio and ‘chromaticity’ transformation techniques, *Remote Sens. Environ.*, vol. 22, no. 3, pp. 343–365.
- Government of India Ministry of Water Resources, 2008, Preliminary consolidated report on effect of climate change on water resources. Central Water Commission, National Institute of Hydrology
- Hilker T., Wulder M. A., Coops N. C., Linke J., McDermid G., Masek J. G., Gao F., White J. C., 2009, A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS, *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1613–1627.
- Hodson, D.L.R., Sutton R.T., Cassou C., Keenlyside N., Okumura Y., Zhou T., 2009, Climate impacts of recent multidecadal changes in Atlantic Ocean Sea Surface Temperature: a multimodel comparison. *Clim. Dyn.*, vol. 34, pp. 1041–1058. doi:10.1007/s00382-009-0571-2
- Huang B., Song H., 2012, Spatiotemporal Reflectance Fusion via Sparse Representation, *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716.
- Huntington T.G., 2006, Evidence for intensification of the global water cycle: review and synthesis, *Journal of Hydrology*, vol. 39, pp.83-95, doi:10.1016/j.jhydrol.2005.07.003
- Kisi O., 2008, Streamflow forecasting using neuro-wavelet technique, *Hydrol. Process.*, vol. 22, no. 20, pp. 4142-4152.
- Lek S., Delacoste M., Baran P., Dimopoulos I., Lauga J., Aulagnier S., 1996, Application of neural networks to modelling nonlinear relationships in ecology, *Ecol. Modell.*, vol. 90, no. 1, pp. 39–52.
- Malmaeus JM, Blenckner T, Markensten H, Persson I., 2006, Lake phosphorus dynamics and climate warming: a mechanistic model approach, *Ecol Model*, vol. 190, pp. 1–14.
- Monckton C., Soon W. W. H., Legates D.R., Briggs W.H., 2015, Why models run hot: results from an irreducibly simple, *climate model. Sci. Bull.*, vol. 60, no.1, pp. 122-135. doi: 10.1007/s11434-014-0699-2.
- Mullon L., Chang N. B., Yang Y. J., Weiss J., 2013, Integrated remote sensing and wavelet analysis for short-term teleconnection patterns in the Northeast America, *Journal of Hydrology*, vol. 499, pp. 247-264, doi:10.1016/j.jhydrol.2013.06.046
- Mullon L., Chang N.B., Imen S., Yang J., 2015, Forecasting regional terrestrial precipitation global nonlinear and non-stationary teleconnection signals in Northeast U.S., *Journal of Climate*, in review.
- Mwale D.M., Gan T. Y., Shen S.S. P., 2004, A new analysis of variability and predictability of seasonal rainfall of Central Southern Africa for 1950-94, *Int. J. Climatol.*, vol. 24, pp.1509-1530, doi: 10.1002/joc.1062
- Partal T., Cigizoglu H.K., 2009, Prediction of daily precipitation using wavelet-neural networks, *Hydrological Sciences*, vol. 54, no. 2, pp. 234-246.

- Psenner R, Schmidt R., 1992, Climate-driven pH control of remote Alpine lakes and effects of acid deposition. *Nature*, vol. 356, pp. 781–3
- Ritchie J.C., Cooper C.M., 2001, Remote sensing techniques for determining water quality: Applications to TMDLs. *In TMDL Science Issues Conference*, Water Environment Federation, Alexandria, VA, pp. 367–374.
- Ritchie J.C., Zimba P.V., Everitt J.H., 2003, Remote sensing techniques to assess water quality, *Photogrammetric Engineering and Remote Sensing*, pp. 695-704.
- Salim A., Pawitan Y., 2007, Model-Based maximum covariance analysis for irregularly observed climatological data, *Journal of Agricultural, Biological & Environmental Statistics*, vol. 12, no. 1, pp. 1-24.
- Schowengerdt R.A., 1980, Reconstruction of multispatial, multispectral image data using spatial frequency content., *Photogramm. Eng. Remote Sens.*, vol. 46, pp. 1325–1334.
- Shen S., New D., Smith T. M., Arkin P. A., 2013, Hibert-Huang transform analysis of the global average monthly precipitation. *Advances in Adaptive Data Analysis*, vol. 4, no. 3, 21p., doi: 10.1142/S1793536912500185.
- Smith R.B., 2012, Introduction to Remote Sensing of Environment (RSE), 31p, <http://www.microimages.com>.
- Teutschbein C., Wetterhal F., Seiber J., 2011, Evaluation of different downscaling techniques for hydrological climate change impact studies at the catchment scale, *Climate Dynamics*, vol. 37, no. 9-10, pp.2087-2105, doi 10.1007/s00382-010-0979-8
- Wang W., Ding J., 2003, Wavelet network model and its application to the prediction of hydrology, *Nature and Science*, vol. 1, no. 1, pp. 67-71.
- Wang W., Lu W., Wang X., Leung A. Y. T., 2003, Prediction of maximum daily ozone level using combined neural network and statistical characteristics., *Environ. Int.*, vol. 29, no. 5, pp. 555–62.
- Wang W. M., J., Niu Z., Zhao Y., Wang C., 2012, A model for spatial and temporal data fusion, *J. Infrared Millim. Waves*, vol. 31, no. 1, pp. 80–84.
- White F. E., 1991, Data Fusion Lexicon, in The Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C3, vol. 15, p. 15.
- Yu T., 2005, Utility of remote sensing data in retrieval of water quality constituents concentrations in coastal water of New Jersey, New Jersey Institute of Technology.
- Zhang W., Li A., Jin H., Bian J., Zhang Z., Lei G., Qin Z., Huang C., 2013, An Enhanced Spatial and Temporal Data Fusion Model for Fusing Landsat and MODIS Surface Reflectance to Generate High Temporal Landsat-Like Data, *Remote Sens.*, vol. 5, no. 10, pp. 5346–5368.
- Zhou J., Civco D. L., Silander J. A., 1998, A wavelet transform method to merge Landsat TM and SPOT panchromatic data, *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757.
- Zhu X., Chen J., Gao F., Chen X., Masek J. G., 2010, An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions, *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623.

Zhang Y., Qian Y., Duliere V., Salathe E., Leung L. R., 2012, ENSO anomalies over the Western United States: present and future patterns in regional climate simulations, *Climatic Change*, vol. 110, no. 1, pp. 315-346, doi:10.1007/s10584-011-0088-7

# **CHAPTER 2: ATTRIBUTIONAL ASSESSMENT OF LOCAL TELECONNECTION SIGNALS DURING RECENT DROUGHTS IN THE UPPER COLORADO RIVER BASIN TOWARD LONG-TERM PRECIPITATION FORECASTING**

## **2.1. Introduction**

Changes in climate patterns and extreme events have potential environmental and socioeconomic effects and may lead to disasters. As a result of the importance of these challenges, the Intergovernmental Panel on Climate Change (IPCC) has focused on “managing the risk of extreme events and disasters to advance climate change adaption” (IPCC, 2012). This issue is of great importance in the Western United States which experienced a decade-long drought.

The Bureau of Reclamation in the Phase 4 of the Colorado River Basin water supply and demand study stated that “based on preliminary assessments, large supply-demand imbalances greater than 3.5 million acre-feet are plausible over the next 50 years when considering a water supply scenario that incorporates changes in climate” (Bureau of Reclamation, 2011). This imbalance can be caused by a population growth of approximately 10 million from 1990 to 2008 especially in Arizona, Utah, Las Vegas, and Tijuana which leads to a rapid increase in water demand for municipal use, agriculture, and hydropower after signing the Colorado River Compact in 1922 (Cohen, 2011; Switanek et al., 2009). However, projections show a 10-30% decrease in runoff in the mid-latitude Western North America by the year 2050, and indicate



variation in the conditions of sub-basins in the Upper Colorado River Basin from semi-arid to arid conditions by the 2080s (Milly et al., 2005; Ficklin et al., 2013).

The impact of droughts in the Colorado River Basin is reflected in the storage of the two largest reservoirs, Lake Mead and Lake Powell. The unregulated inflow into Lake Powell for the past 15 years (2000-2014) has been the lowest of any 15-year period, since Glen Canyon Dam went into operation in 1963. In fact, the unregulated inflow into Lake Powell was above average in only 3 of those 15 years (Bureau of Reclamation, 2015). In addition, the reservoir elevation in Lake Mead has dropped by 100 feet during 2000-2005. It is evident that the drought which led to creation of the 2007 Guidelines did not end with their adoption and it is even projected that the shortage in the Lower Basin will reach 45% by the year 2026 (Bureau of Reclamation, n.d.). It is therefore necessary to develop a climate forecast model, which can be used as a warning mechanism, to support stakeholders in the decision making process for allocating this stressed resource in the Colorado River Basin.

An external forcing which plays a key role in controlling the weather and climate change is oceans (NOAA, 2012). Inclusion of the interaction between land, ocean, and atmosphere into the climate forecast model can capture the teleconnection signals associated with sea surface temperature (SST) and can provide us with predictive, local scale, hydroclimate information. A significant number of studies in the Colorado River Basin focused on the relationship between the regional SST variability (i.e. El Nino-Southern Oscillation (ENSO), Pacific Decadal Oscillation (PDO), and Atlantic Multi-decadal Oscillation (AMO)) and terrestrial responses (i.e. snowmelt, streamflow, and precipitation). Partial lists of these studies are shown in Table 2-1.

**Table 2-1. Partial list of references focusing on the effects of regional SST's variability on hydroclimate variables in the Colorado River Basin**

<b>Reference</b>	<b>Teleconnection Pattern</b>	<b>Hydroclimate variable</b>	<b>Forecasting</b>	<b>Method</b>
Hidalgo& Dracup, 2003	ENSO, PDO	Precipitation streamflow	×	Principle component analysis & linear correlation
Kim et al., 2006	SOI, PDO	Precipitation	×	Standardized precipitation index & conceptual influence index analysis
McCabe et al., 2007	AMO, PDO, ENSO, Indian Ocean	Streamflow	×	Monte Carlo, Linear Correlation, principle component analysis
Thomas, 2007	AMO, PDO, SOI	Streamflow	√	1. Linear least square multiple regression analysis 2. confidence intervals of the median
Kim et al., 2008	SOI, ESOI, WA, PDO, NAO, ATNH, Nino1+2, Nino3, Nino4, Nino3,4	Precipitation	×	Linear Correlation and conceptual influence index analysis
Hurkmans et al., 2009	ENSO, PDO, MEI, PNA	Precipitation soil moisture outflow	×	Linear correlation
Aziz et al., 2010	Hondo 34°N-24°S 150°E-160°W	Snowpack	×	Single value decomposition
Lamb et al., 2011	SOI, PDO, NAO, AMO	Streamflow	×	Rank sum analysis & single value decomposition
Karla et al., 2012	PDO, NAO, ENSO, Hondo	Streamflow	√	Support vector machine

\* ENSO- El-Nino Southern Oscillation; PDO-Pacific Decadal Oscillation,; SOI- Southern Oscillation Index; AMO-Atlantic meridional oscillation; ESOI- Equatorial Southern Oscillation Index; WA- West coast of America; NAO- North Atlantic Oscillation; ATNH-Average temperature at north hemisphere; MEI- multivariate ENSO index.

Although prior studies have been able to gain insight into the influence of the known teleconnection patterns on hydroclimate variables in this region, there have been few studies done from the perspective of detecting non-leading teleconnection signals which may have

influence on the hydroclimate variables of the Colorado River Basin (Baldwin et al., 2005; Switanek and Torch, 2009; Tootle and Piechota, 2005). Yet these studies tend to have large uncertainties in estimates by utilizing simple linear analyses to examine teleconnection relationships. The non-stationarity and non-linearity of the ocean-atmosphere system make the procedure of detecting teleconnection signals difficult at a local scale. One way to overcome the problem of non-stationarity in the ocean-atmosphere system is using wavelet analysis. Complementary to this, the application of a method for the removal of the long-term trend from non-linear and non-stationary time series will reduce the large uncertainty in linear correlation. To bridge the gap of previous studies, the aforementioned solutions will be applied in the current study.

The hydrologic regime of a river basin is dependent on the spatial and temporal variability of run-off and soil moisture which are both highly affected by the precipitation trend. Accurate estimation of precipitation will shed more light on understanding the hydrologic cycle and monitoring the variability in climate extremes (Karla and Ahmad, 2011). The hydrological systems are predicted by regressive analysis and stochastic theory. Although the stochastic models, such as ARIMA and SARIMA (which is able to control stochastic seasonality), have long been used for forecasting in the sphere of hydrology, the fundamental assumptions of these models are stationary over time (Delleur and Kavvas, 1978; Thapaliyal, 1981; Kaushik and Singh, (2008); Abdul-Aziz et al., 2013). The basic idea of these models based on the removal of trends from time series in order to transform them into a stationary time series, amplifies high frequency noise in the data (Coulibaly and Baldwin, 2005).

The non-linear characteristics of hydrological records have given rise to the use of the artificial intelligence (AI) type models, which is also known as machine-learning and data driven models (Karl et al., 2012). One of the most commonly used machine-learning models is artificial neural network (ANN) which has been in use in the field of water resources since the 1990s. ANN models are able to capture the complex non-linear nature of hydrological problems, consider the seasonal variability of data, and also show strong performance with limited data (Maier and Dandy, 1996). Although the application of ANN, fuzzy theory, and chaos theory has been broad in the field of hydrology and water resources, several studies, which applied these models, have pointed out a lack of precision with their forecasts (Wang and Ding, 2003).

To solve this problem, Wang and Ding (2003) proposed a hybrid ANN model called wavelet neural network (WNN). Therefore, the weakness in input time series can be solved by decomposing time series into their fundamental signals prior to apply into ANN model (Mullon et al., 2015). Anctil and Tail (2004) showed that the forecasting performance can be highly improved using the WNN model in case of having input time series from different sources (e.g. precipitation and evapotranspiration). In this case, the classic ANN model combine information at various frequency scales and producing different dynamic which cause the system switches to different hydrologic sources. This problem can be solved by the WNN model which includes multi-scale wavelet analysis into a set of neural networks.

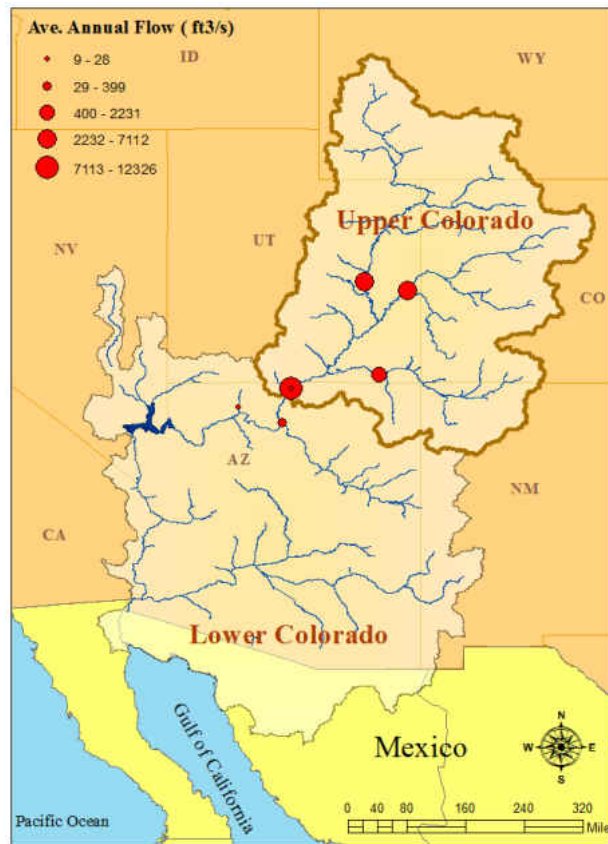
Several water resources disciplines, which require forecasts, have successfully implemented WNN modeling (Rajaei, 2011), municipal water demand forecasting (Mohammed and Ibrahim, 2012), shallow groundwater level forecasting (Wang and Ding, 2003), daily discharge forecasting (Wang and Ding, 2003), streamflow forecasting (Cannas et al., 2005;

Cannas et al., 2006; Kisi, 2008; Kisi, 2009; Pramanik et al., 2011), and precipitation forecasting (Nourani et al., 2008 Partal and Cigizoglu, 2009; Ramana et al., 2013; Solgi et al., 2014). Comparing the results of WNN model with neuro-fuzzy model showed the better forecasting performance of the WNN model in both short and long-term prediction (Solgi et al., 2014). Among previous studies, Mwale et al. (2004) and Elsanabary and Gan (2014) applied the WNN model to forecast seasonal precipitation based on global sea surface temperature as predictor data. They applied the wavelet principle component analysis to individual scale-average wavelet power to find the dominant oscillations of variability in time series. This study goes one step further by quantifying the individual contribution of the detected SST forcing regions through nonlinear modeling with the aid of extreme learning machine.

The main objective of this study is: (1) finding an oceanic index region which has statistically significant effect on precipitation over the Upper Colorado River Basin in each season, (2) develop a watershed-scale forecasting model to predict precipitation variability over the Upper Colorado River Basin one-season ahead using the historical precipitation data and the detected oceanic regions in step 1. (3) Estimate the hydrological events given the predicted seasonal precipitation, and precipitation range for each category of SPI. Key science questions to be answered include (1) Is there any non-leading teleconnection patterns affect terrestrial precipitation more than the known leading teleconnection patterns in this region? If so, to what extent? (2) Is the ability of the developed method in predicting the hydrological extremes over the upper Colorado River Basin the same in different seasons?

## 2.2. Study Area

The headwaters of Colorado River are located in Rocky Mountain National Park. The Colorado River flows toward west and it is joined by the Gunnison River at Grand Junction. As it reaches Utah, it is joined by the Green River. Then, right before it flows into Lake Powell it is joined by San Juan River, which emanates from San Juan Mountains. As shown in Figure 2-1, waters within or from parts of four U.S. states including Colorado, New-Mexico, Utah, and Wyoming are drained into Upper Colorado River Basin.



**Figure 2-1. The location of Colorado River Basin in the U.S. The area marked by bold line is the present study area. The size of circles shows the magnitude of average annual flow at the location of different stations. A larger circle shows large magnitude of average annual flow.**

The snowmelt from precipitation in three states: Colorado, Utah, and Wyoming provide about 90% of the inflow into the Colorado River Basin (Jacobs, 2011). Therefore, precipitation over the Lower Colorado River Basin does not affect water storage of Lake Mead and Lake Powell, which are both indicators of the water storage capacity in the Colorado River Basin (Jacobs, 2011). Comparison of the average annual flow of gauges located at the upstream and downstream of Lake Powell, shown in Figure 2-1, confirms this matter. Hence, this study is focused on precipitation variability of the Upper Basin.

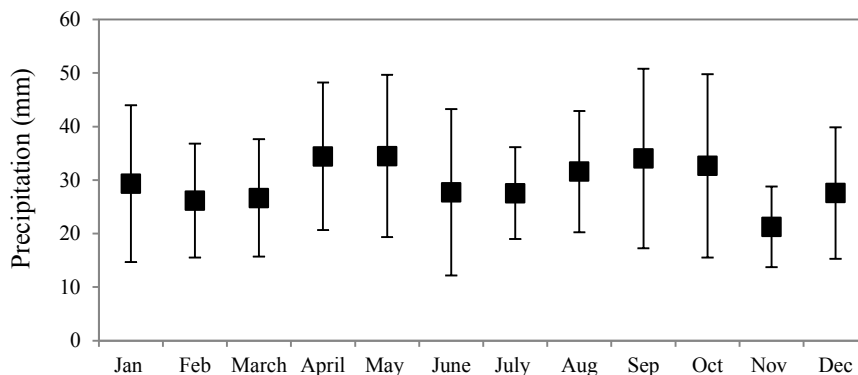
Variations in temperature and precipitation which are caused by climate change have significantly impacted the streamflow timing and availability (Muller and Piechota, 2008). In 2007, National Research Council (NRC) reported that “Temperature records across the Colorado River basin and the western United States document a warming trend over the past century. These temperature records, along with climate model projections, suggest that temperatures across the region will continue to rise in the foreseeable future”. However, there is less agreement on precipitation trend in future, and precipitation trend shows significant variability (NRC, 2007; U.S. Department of the Interior Bureau of Reclamation, 2012). As a result of this extreme variability in precipitation trend in the Colorado River Basin, estimating the amount of precipitation ahead of time would be beneficial for water managers in this basin. Prior to describing the developed forecasting method, the following section is focused on precipitation regimes over the Upper Colorado River Basin.

### *2.2.1. Precipitation Regimes over the Upper Colorado River Basin*

Precipitation in summer and winter seasons contributes significantly to the total annual in the Upper Colorado River Basin. Winter snow and summer monsoon are two main sources of

water in this region (Kim et al., 2006). Precipitation of these two seasons is affected by moisture content of frontal systems caused by atmospheric circulation patterns and sea surface temperature. In winter and spring, the frontal systems originating from the Pacific Ocean transfer a significant amount of moistures to the Colorado River Basin which increases precipitation over the elevated areas in this basin. In addition, cold front systems leads to snowpack at the elevation above 5,000 feet melted in the late spring or earlier by warm storm winters which causes high runoff on major rivers (USGS, 2004). In summer (especially during July-August), the North American Monsoon transfers moisture from east Pacific Ocean, Gulf of Mexico, and Gulf of California toward the Colorado River Basin which causes considerable precipitation in New Mexico, Utah, and Colorado during summer (USGS, 2004; U.S. Department of the Interior Bureau of Reclamation, 2012).

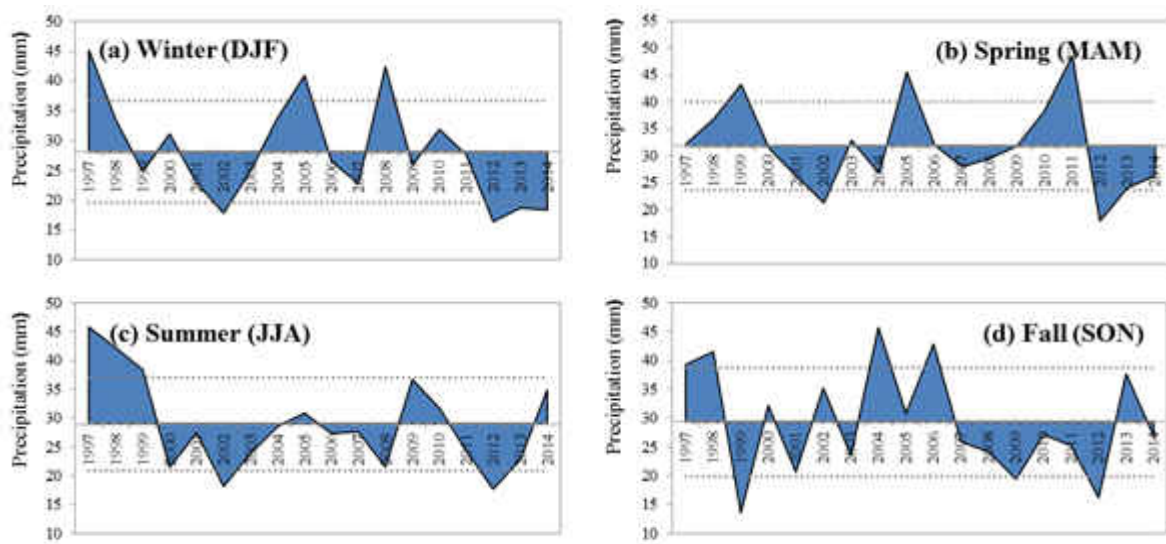
To find the best time scale for precipitation forecasting, it would be beneficial to find the answer to several questions. The first question is whether there is annual cycle of precipitation in this basin. To answer this question, it is important to have information about the variability of precipitation in the Upper Colorado River Basin.



**Figure 2-2. Annual cycle of precipitation in the Upper Colorado River Basin (1997-2014). Bars indicate monthly standard deviations.**



Variation of precipitation within the time period of this study (Jan 1997-Dec2014) is presented in Figure 2-2 and Figure 2-3. Based on Figure 2-2, there is no substantial difference in the average of precipitation in different months. In addition, as shown in Figure 2-3, the average annual precipitation of all seasons is almost the same, and there is no season in which precipitation dominates the annual total precipitation.



**Figure 2-3. (a) winter season (December to February); (b) spring season (March to May); (c) summer season (June –August); (d) fall season (September to November); Dashed lines indicate  $\pm 1$  stdv for the winter (8.58 mm), spring (8.18 mm), summer (8.03 mm), fall (9.40 mm); Middle dashed line indicates average for winter (28 mm), spring (32 mm), summer (29 mm), fall (29 mm).**

Since the main objective of this study is forecasting precipitation in order to help water managers, the second question is whether there is any relationship between precipitation and reservoir storage? If so, at which scale (monthly/seasonal) it is more significant? To answer this question, the association between precipitation and inflow into several reservoirs located at the Upper Colorado River Basin is assessed during 18-year (1997-2014) for two different time scales

(i.e. monthly and seasonal). As shown in Table 2-2, the correlations between precipitation and the inflow into the reservoirs are more significant at seasonal time scale.

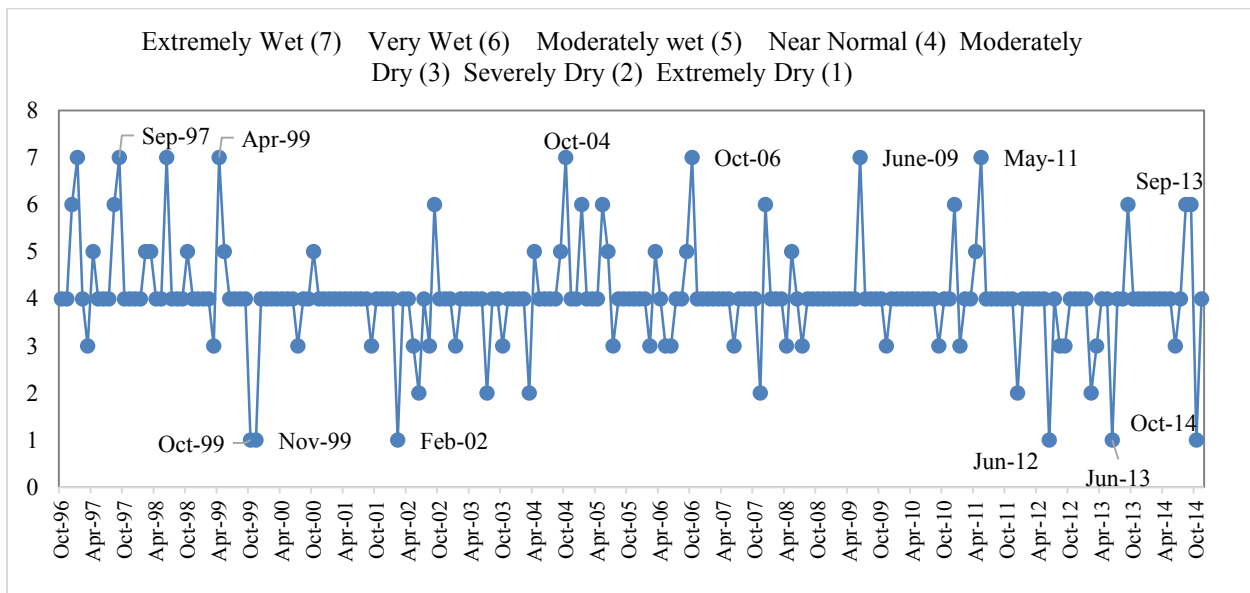
**Table 2-2. Correlation between precipitation and inflow into the reservoirs within different time scales**

	<b>Lake Powell</b>	<b>Fontenelle</b>	<b>Flaming Gorge</b>	<b>Taylor Park</b>	<b>Navajo</b>
Monthly	0.27	0.19	0.25	0.14	0.30
Winter (DJF)	0.45	0.43	0.28	0.48	0.56
Spring (MAM)	0.52	0.45	0.65	0.08	0.33
Summer (JJA)	0.49	0.66	0.66	0.45	0.69
Fall (SON)	0.48	0.38	0.25	0.16	0.53

One of the goals of this study is to find a linkage between the predicted precipitation and the severity of droughts in order to inform the water managers in advance. This poses a third question, whether or not there is a particular month or season associated with droughts? To answer this question droughts are quantified based on the precipitation deficit using the standardized precipitation index (SPI), which is described in the methodology section. McKee et al., (1993) defined the following classification system to quantify intensities of hydrological events based on the SPI value (Table 2-3). The calculated values of SPI for the period of (1997-2014) and their corresponding categories are shown in Figure 2-4. As shown in Figure 2-4, extremely dry conditions were detected in different months in fall, winter and summer. Therefore, there is not any particular month or season associated with droughts in this region.

**Table 2-3. The SPI classification system (McKee et al., 1993)**

<b>SPI</b>	<b>category</b>
2 and more	extremely wet
1.5 to 1.99	Very wet
1 to 1.49	Moderately wet
-0.99 to 0.99	Near normal
-1 to -1.49	Moderately dry
-1.5 to -1.99	Severely dry
-2 and less	Extremely dry



**Figure 2-4. Calculated SPI for monthly precipitation in the period of (1997-2014). Y-axis shows the defined category of SPI. The definition corresponding to each vale in y-axis are shown at the top of the diagram.**

Consequently, based on the results of precipitation analysis in this section, the seasonal scale was selected to be used in the developed precipitation forecasting model in this study.

### 2.3. Data and Method

In this study, empirical approaches are integrated to identify the dominant oceanic forcing regions and quantify the contribution of each oceanic index regions associated with precipitation responses over the Upper Colorado River Basin. Then, the extracted data are

applied to predict precipitation and estimate the intensity of the hydrological events over the study area for a season ahead of time. As shown in Figure 2-5, the proposed methodology includes three phases. The main objective of Phase I is based on identifying possible index regions by big data analytics which mainly consists of 6 steps. Firstly, SST and precipitation data are collected. Then, seasonal time series of each dataset are calculated. The long-term trends of each time series are removed, namely detrending, which is necessary prior to obtaining any meaningful results from the correlation analysis (Wu et al., 2007). After detrending, time series are decomposed at time-frequency domain to determine localized variation of wavelet power and dominant modes. Then, pixel-wise linear lagged correlation analysis is performed between the isolated dominant modes of SST and precipitation time series in order to identify the significant oceanic index regions. Next, in Phase II which is designed for screening and ranking the dominant index regions, extreme learning machine (ELM) is applied to simulate the possible nonlinear relationships between precipitation and the extracted SST index region. Finally, the contribution of each identified oceanic index region to precipitation responses is calculated through the sensitivity analysis. Lastly, in phase III the extracted data are applied as inputs into the ANN model to predict precipitation over the study area for a season ahead of time and to predict the hydrological extremes based on the SPI-based algorithm.

### *2.3.1. Data Collection*

Precipitation data is acquired from Global Precipitation Climatology Project (GPCP) dataset produced based on an optimal merge of estimates calculated from microwave, infrared and sounder data observed by the international constellation of precipitation-related satellites and precipitation gauge analysis.

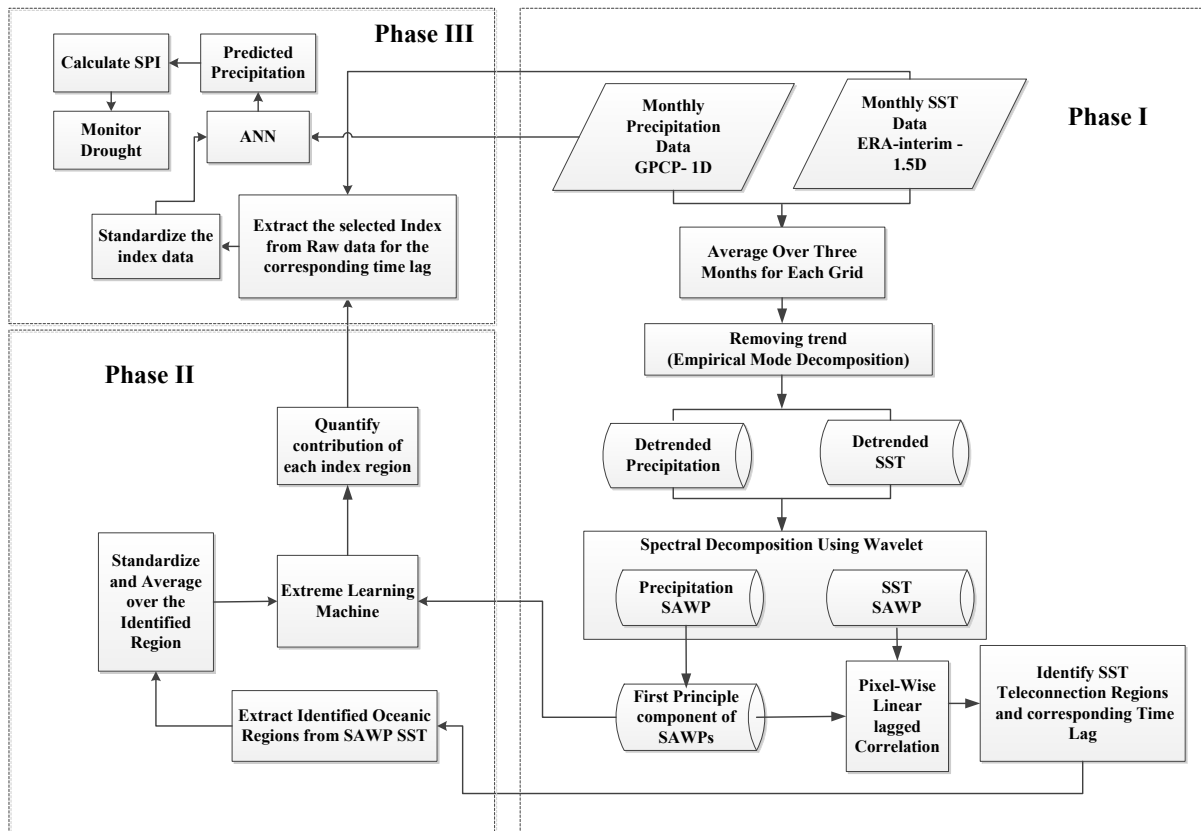


Figure 2-5. The study framework

Since 1996, GPCP data provides higher resolution estimates using the geo-synchronous-satellite operators around the world in order to collect histograms of infrared radiometers on geosynchronous satellites (geo-IR) brightness temperature ( $T_b$ ) on a  $1^\circ \times 1^\circ$  grid covering  $40^\circ\text{N}$ - $40^\circ\text{S}$  at 3-hourly intervals (Huffman et al., 2001). However, outside the geo-IR coverage, the sounding data is used. To estimate precipitation from the 3-hourly geo-IR histograms, the threshold-matched precipitation index (TMPI) is applied to compare the infrared brightness temperature (IR  $T_b$ ) with threshold  $T_b(\text{rain})$ . All pixels colder than  $T_b(\text{rain})$  are labeled as “precipitating”. It should be mentioned that both geo-IR and sounder data are responding to clouds instead of hydrometeors which cause some algorithmic uncertainties at the finest scales (Huffman et al., 2001). However, comparison between the GPCP 1DD data with rain gauge data

showed considerable agreement in both cold and warm seasons (Rubel and Rudolf, 1999). In addition, results of this comparison indicated excellent agreement at seasonal scale (Huffman et al., 2001). The daily GPCP data with spatial resolution of (1-degree) are employed in this study. This product estimates precipitation over the entire globe at 1-day for the period 1996 – 2014 and can be acquired via Global Precipitation Analysis home page (Global Precipitation Analysis, 2015). To meet the requirements of this study, monthly precipitation data is calculated based on summation of daily values in each month.

SST data is obtained from the latest ERA-Interim reanalysis products, which is generated by the European center for medium-range weather forecast (ECMWF). It includes hourly to monthly SST data which extends back to 1979 and continues to present time with spatial resolution of 1.5 degree latitude to 1.5 degree longitude. This product is produced by a 2006 version of the integrated forecast system (IFS). ERA- 40 was replaced with ERA-Interim, as a result of its improved atmospheric model and data assimilation system which can eliminate or significantly remove the several observed inaccuracies of ERA-40. The ERA-Interim SST data from Jan. 1996 through Dec. 2014 applied in this study is obtained from the ECMWF's homepage (ECMWF, 2015).

As a result of precipitation regimes in the Upper Colorado River Basin on one hand, and the significant seasonality revealed between terrestrial precipitation variability and SST forcing on the other hand, seasonal time series of each variable are thus calculated and used instead of monthly time series to assess the associations between the global SSTs and terrestrial precipitation. Seasonal time series are calculated by computing every 3 months averages, namely March-April-May (MAM) for spring, June-July-August (JJA) for summer, September-October-

November (SON) for fall, and December-January-February (DJF) for winter. In order to establish the possible associations between SSTs forcing and precipitation responses, lag correlation analysis is performed. Earlier studies in this region suggested selected lag time between 1-12 months as possible time for SSTs signal propagation to affect precipitation regime (Hidalgo and Dracup, 2003; Kim et al, 2006; Kim et al., 2008). Therefore, 13 different SSTs time series are computed with time lags from 0 to 12 months. For instance, to find the lagged correlation between precipitation of MAM, the associated SST time series are averaged for three month periods starting with MAM (0 lag), FMA (1 month lag), JFM (2 months lag), ..., MAM (12 months lag).

### *2.3.2. Phase I: Identification the Possible Index Regions*

As a result of complexity in climate systems, non-stationary and non-linearity characteristics of climate time series is expected (Wu et al., 2007). One of the important preprocessing steps prior to the further analysis is to remove long-term trend as it can lead to large uncertainty in linear correlation analyses (Pouzols and Lendasse, 2010). The most common approach to determine the trend is to fit a straight line to the time series. Since the aforementioned common approach is not suitable for non-linear signals, the empirical mode decomposition (EMD) was applied in this study. Earlier studies showed the ability of EMD approach in determining the global trend relative to wavelets, classical filtering average, and computational intelligence models (Pouzols and Lendasse, 2010; Amar and Guennoun, 2012). EMD approach was first suggested by Huang et al. (1998), and it has the ability of extracting the intrinsic and adaptive trends from non-linear and non-stationary time series. The procedure of EMD to extract adaptive trends is as follows (Huang et al., 1998):

- (1) Determine all the minima and maxima of input time series  $x(k)$ ,
- (2) Calculate the signal envelope passing through the minima  $e_{\min}(k)$  and maxima  $e_{\max}(k)$ ,
- (3) Compute the local mean time series,  $M(k)$ :

$$M(k) = \frac{e_{\min}(k) + e_{\max}(k)}{2} \quad (1)$$

- (4) Obtain the intrinsic mode function IMF candidate,  $S(k)$ :

$$S(k) = X(k) - M(k) \quad (2)$$

- (5) If the properties of  $S(k)$  is not the same as an IMF, replace  $X(k)$  with  $S(k)$  and repeat the procedure from the first step. Otherwise calculate the residual “ $r(k)$ ” as follows:

$$r(k) = X(k) - S(k) \quad (3)$$

- (6) Repeat step 1-5 till the residuals “ $r(k)$ ” is monotonic. The monotonic  $r(k)$  is defined as a trend.

- (7) Once the trend is extracted, it will be removed from the original time series, which can be shown as:

$$x'(k) = x(k) - r(k) \quad (4)$$

To overcome the non-stationarity problem of the ocean-atmosphere systems, wavelet analysis is applied to find the significant oscillations in time series of SST and precipitation at a seasonal time scale, and the range of dominant oscillations is then averaged into a scale-average wavelet power (SAWP) which depicts the modulation of many frequencies within a single time series. This analysis is performed through the following steps for all grids of precipitation within the Upper Colorado Basin and SST individually (Mwale et al., 2004):

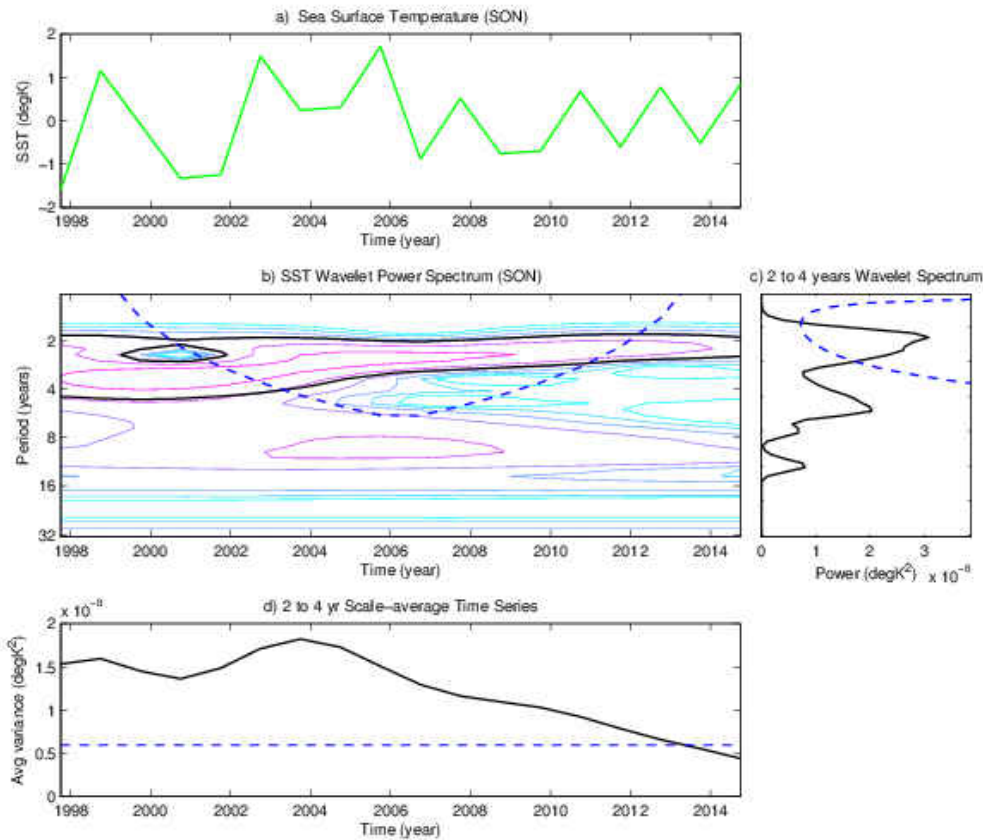
- (1) Wavelet Power Spectrum (Figure 2-6b): For each grid of precipitation and SST, time



series of these variables are decomposed into two-dimensional time-frequency domain using the continuous wavelet transform (CWT) process. Black solid line areas show statistical significant regions at 95% confidence level. As a result of applying finite length time series in this study, errors may occur at the beginning and the end of the wavelet spectrum which is padded with sufficient zeroes. To avoid these edge effects, all the regions inside the cone of influence (COI), which is shown by the inverted dashed blue arch shape line, are considered. This figure shows dominant oscillation in time series of SST in fall season as an example.

- (2) Global Wavelet Spectrum (Figure 2-6c): It represents the variation of dominant oscillations with time. Peaks of time series which is above the dashed curve areas show statistically significant region with the 95% confidence level,
- (3) Scale-average time series (Figure 2-6d): It depicts how the dominant oscillations defined in the previous steps in a certain band (2-4 years) vary with time. It is defined by calculating the weighted sum of the wavelet spectrum power over the defined band which is named as SAWP.

Singular value decomposition (SVD) is an applicable tool to analyze multivariate data. In this study, SVD is applied to the precipitation SAWPs in order to extract the common spatial and temporal variability of precipitation SAWPs and consolidate them into a single time series (i.e. principle component) without loss of information.



**Figure 2-6. Wavelet decomposition example of a single SST grid in fall season; (a) seasonal anomaly SST data with the X and Y axes representing time and SST in degrees, respectively. (b) SST wavelet power spectrum; (c) global wavelet spectrum; (d) SAWP time series which was calculated based on the 2-4 year period band. The dashed blue lines show the boundary of cone of influence.**

To find the association between precipitation over the Upper Colorado River Basin and global SST, linear lagged correlation between the SAWP of global SSTs and the PC1 of precipitation SAWP is calculated. The SAWP of SST and precipitation are applied to avoid using a wide range of unknown frequencies included in the raw data of SST and precipitation time series (Mwale et al., 2004). To compute the linear lagged correlation between terrestrial precipitation and global SST anomalies, the following Pearson correlation coefficient equation is applied:

$$r = \frac{\sum_i^n [(x_i - \bar{x}) \times (y_{i+d} - \bar{y})]}{\sqrt{\sum_i^n [(x_i - \bar{x})]^2} \times \sqrt{\sum_i^n [(y_{i+d} - \bar{y})]^2}} \quad (5)$$

where,  $r$  is the correlation coefficient,  $n$  is number of data,  $x$  and  $y$  are variables; in this study  $x$  represents the precipitation dataset and  $y$  represents the SST dataset;  $\bar{x}$  and  $\bar{y}$  are the mean;  $d$  is the time lag. Based on the calculated correlation coefficient of each pixel, the pixel-wise linear correlation maps are computed for the study area in all four seasons to identify the oceanic regions in which SSTs show significant correlation (at 95% confidence level) with terrestrial precipitation in different seasons.

From the pixel-wise correlation maps, regions with consistently statistical significance level in a long-term period (more than 3 month) are extracted. Besides extracting statistically significant regions, their corresponding time lags are determined. While the magnitude of correlation represents the intensity of SST forcing precipitation, the corresponding time lags indicate the possible time for the precipitation responding to SST forcing (Wu et al., 2008). Therefore, the lag corresponding to the maximum or minimum correlation coefficients between precipitation and each extracted SST region from 0-12 lags is determined. Therefore, each SST index region is extracted from the SAWP of global SST regarding to the determined time lag and then averaged to get individual SST indices.

### 2.3.3. Phase II. Screening and Ranking the Identified Index Regions

As several oceanic regions may show statistically significant relationship with precipitation in each season, it proposes a question of which one can be consider as a predictor for forecasting precipitation in the next step. To find the answer of this question, the advanced

machine learning tool, ELM, is applied to explore the nonlinear relationships between the extracted SST indices and precipitation. Although the backpropagation learning algorithm (BP) plays a significant role in training feedforward neural networks, it also comes with some challenges such as local minima, intensive human intervene, and slow learning speed (Huang, 2014). To overcome these challenging issues, ELM was proposed by Huang et al. in 2006. The structure of the single layer feedforward neural network is shown in Figure 2-7. For N arbitrary samples  $(x_i, y_i)$ , the mathematically modeled ELM is as follows (Huang et al., 2006):

$$\sum_{i=1}^N v_i g(a_i \cdot x_j + b_i) = y_j \quad j = 1, \dots, N \quad (6)$$

Equation (5) can be written as:

$$HV = Y \quad (7)$$

where

$$V = \begin{bmatrix} v_1^T \\ \vdots \\ v_L^T \end{bmatrix}_{L \times m} \quad Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m} \quad (8)$$

$$H = \begin{bmatrix} g(a_1, b_1, x_1) & g(a_2, b_2, x_1) & \dots & \dots & g(a_L, b_L, x_1) \\ \vdots & \vdots & & & \vdots \\ g(a_1, b_1, x_N) & g(a_2, b_2, x_N) & \dots & \dots & g(a_L, b_L, x_N) \end{bmatrix} \quad (9)$$

where  $g(\cdot)$  is an activation function which is sigmoid function  $g(x) = 1/(1 + e^{-x})$  in this study;  $a_i$  is the weight vector connecting the  $i^{\text{th}}$  hidden node and input nodes;  $v_i$  is the weight vector connecting the  $i^{\text{th}}$  hidden node and output nodes;  $b_i$  is the biases of the  $i^{\text{th}}$  hidden node;  $y_j$

is the expected value of  $j^{\text{th}}$  output node. The procedure of ELM algorithm is as follows (Huang et al., 2006):

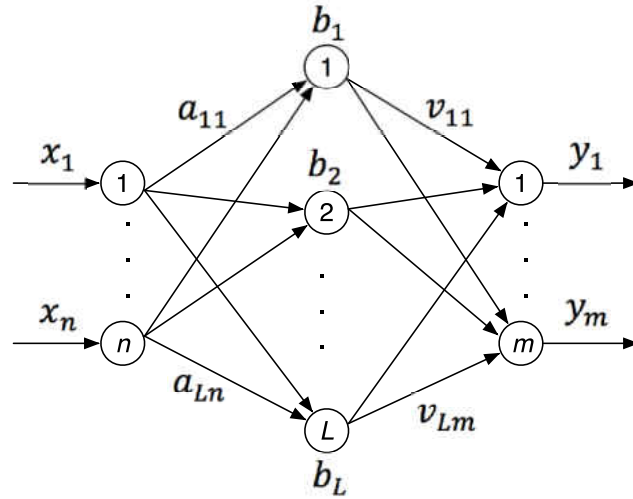


Figure 2-7. Typical scheme of single layer feedforward neural networks (Chang et al., 2010)

- (1) Given a training set, activation function  $g(\cdot)$ , and hidden node number ( $L$ ),
- (2) Randomly assign input weights  $a_i$  and bias  $b_i$ ,
- (3) Calculate the hidden layer output matrix  $H$ ,
- (4) Calculate the output weight  $V$  from Equation (9),

$$V = YH^+ \quad (10)$$

where  $H^+$  is the Moore-Penrose generalized inverse of matrix  $H$ . Therefore,  $V$  is the only parameter which needs to be solved. To establish a robust ELM algorithm in this study, all samples  $(x_i, y_i)$  are randomly separated into two datasets. First, 70% of samples are applied to train the model. Then, the rest 30% of samples (i.e. unseen data) are applied to validate the trained ELM model. In the training stage, the correlation coefficient between the model outputs

( $y_j$ ) and the observed data ( $x_i$ ) is calculated as stopping criterion. If the correlation coefficient reaches to 90%, the simulated ELM model will be considered as robust enough for predicting. Otherwise, training process will be repeated until it meets the stopping criterion. In this study, extracted SSTs and precipitation are considered as inputs and target of the ELM algorithm. The ELM was performed using MATLAB, with code developed by Nanyang Technological University in Singapore (Extreme Learning Machine, 2013).

Once robust ELM models are simulated, the contribution of each identified oceanic index region to precipitation at each site is calculated through the sensitivity analysis. For this purpose, two precipitation time series are simulated simultaneously based on the following scenarios:

- (1) Including all the identified oceanic index regions,
- (2) Excluding one of the identified oceanic index regions and including the rest indices,

Residuals between the two simulated precipitation time series are defined as the precipitation responses to the excluded index. To reduce the stochastic error, this procedure is repeated for 200 times for each index, and the average of these results is considered as the contribution of each index to precipitation at each site.

#### *2.3.4. Phase III. Seasonal Precipitation Forecasting*

A traditional feedforward ANN model is applied in this study to forecast precipitation one season ahead for all precipitation grids located in the study area. The structure of ANN consists of interconnected nodes arranged in three layers: input, hidden, and output layer. Input layer only distributes inputs to the hidden layer; after processing data in this layer, they are transferred to the output layer. The backpropagation algorithm which is the most common

algorithm to train the feed forward ANNs is applied in this study. The main objective in the backpropagation algorithm is modifying the weights in the network by minimizing the error between a target and computed output (Tokar and Johnson, 1999). The forward pass phase of training is started by connecting nodes within neighboring layers by links and assigning initial weights to each link. Then, a neuron  $j$ , sums the value of all inputs based on the following equation (Dawson and Wilby, 1998):

$$S_j = \sum_{i=1}^n W_{ij}U_i + W_{oj} \quad (11)$$

where,  $W_{oj}$  is the bias,  $i$  is number of inputs to a neuron  $U_i$  to  $U_j$ ,  $j$  is associated to neuron,  $W_{ij}$  is a weight associated to each input. The activation function is applied to the value  $S_j$  to produce the output of a neuron, and it is then transferred to the output layer. If the difference between the trained outputs of ANN and the desired ultimate outputs (error) is less than the selected value, the training procedure is stopped. Otherwise, the reverse pass phase is started by modifying the weights and biases of each neuron using a factor based on derivative of activation function, error, and actual neuron outputs (Dawson and Wilby, 1998).

The MATLAB ANN toolbox with feed forward neural network scheme trained by “trainsecg” activation function is applied in this study. This function updates weights and bias values based on the scale conjugate gradient method. The selected oceanic index region and seasonal precipitation data at each grid were applied as inputs and target, respectively. The MATLAB ANN toolbox was modified by adding a loop to check Pearson correlation coefficient and P-value for each iteration. This iterative process is repeated until the Pearson correlation coefficient value reaches the selected value (more than 80%) while the P-value satisfies a 95% confidence interval.

Once precipitation is predicted for all grids within the study area for a season ahead of time using the ANN model, the predicted values are applied to predict hydrological extremes based on SPI-based algorithm shown in Figure 2-8. To calculate an index based on precipitation deficit for assessing hydrological extremes, the SPI index is calculated based on the following procedure:

Let  $P_{ij}$  ( $i=1,2,\dots,N$ ,  $j=1,2,\dots,K$ ) be the observed time series of monthly precipitation over the Upper Colorado River Basin, where  $N$  is the number of precipitation grids and  $K$  is the number of precipitation observations. First, an individual Gamma distribution is fitted to precipitation data of each grid to determine the parameters of this distribution including shape parameter ( $\beta_i$ ) and scale parameter ( $\alpha_i$ ) via Equations (11) to (12) (Kumar et al., 2009).

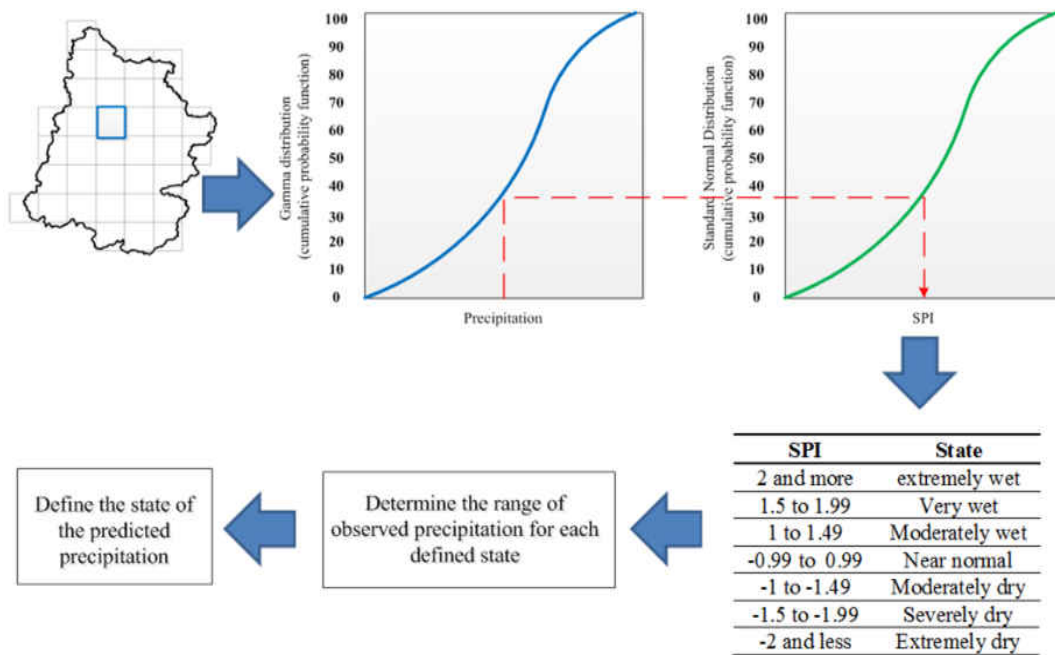


Figure 2-8. The framework of the applied SPI-based algorithm to determine the intensity of hydrological droughts a season ahead of time.



$$\beta_i = \frac{1 + \sqrt{1 + \frac{4}{3} \left( \ln(\bar{P}_{ij}) - \frac{\sum_{j=1}^K \ln(P_{ij})}{K} \right)}}{4 \left( \ln(\bar{P}_{ij}) - \frac{\sum_{j=1}^K \ln(P_{ij})}{K} \right)} \quad (12)$$

$$\alpha_i = \frac{\bar{P}_{ij}}{\beta_i} \quad (13)$$

where,  $\bar{P}_{ij}$  is the mean of precipitation in grid i.

Then, the mathematical cumulative probability function can be assigned to the observed precipitation event at each grid via Equation (13):

$$F(x) = \frac{\int_0^x x^{\alpha-1} e^{-\frac{x}{\beta}} dx}{\beta^\alpha \Gamma(\alpha)} \quad (14)$$

Since the observed precipitation (i.e. “x” in Equation (13)) can be zero, the gamma function is modified as Equation (14):

$$H(x) = q + (1 - q)F(x) \quad (15)$$

where, q is the probability of zero.

In order to take advantages of SPI in comparing precipitation of different grids which have different precipitation characteristics, the calculated cumulative probability functions of each grid is transformed into a standard normal random variable z with mean of zero and standard deviation of one. Transferring the calculated cumulative probability of all precipitation grids in this way puts the transformed variants in the same units, and makes the comparison of different grids possible.

Based on the aforementioned procedure, the series of  $SPI_{ij}$  is calculated and the time series of states is obtained based on the categories defined by McKee et al. (1993). Consequently, the range of precipitation value in different states is defined for each grid. By knowing the range of precipitation value at each grid corresponding to different states, the hydrological extremes at each pixel can be estimated given the predicted value of precipitation.

## 2.4. Results

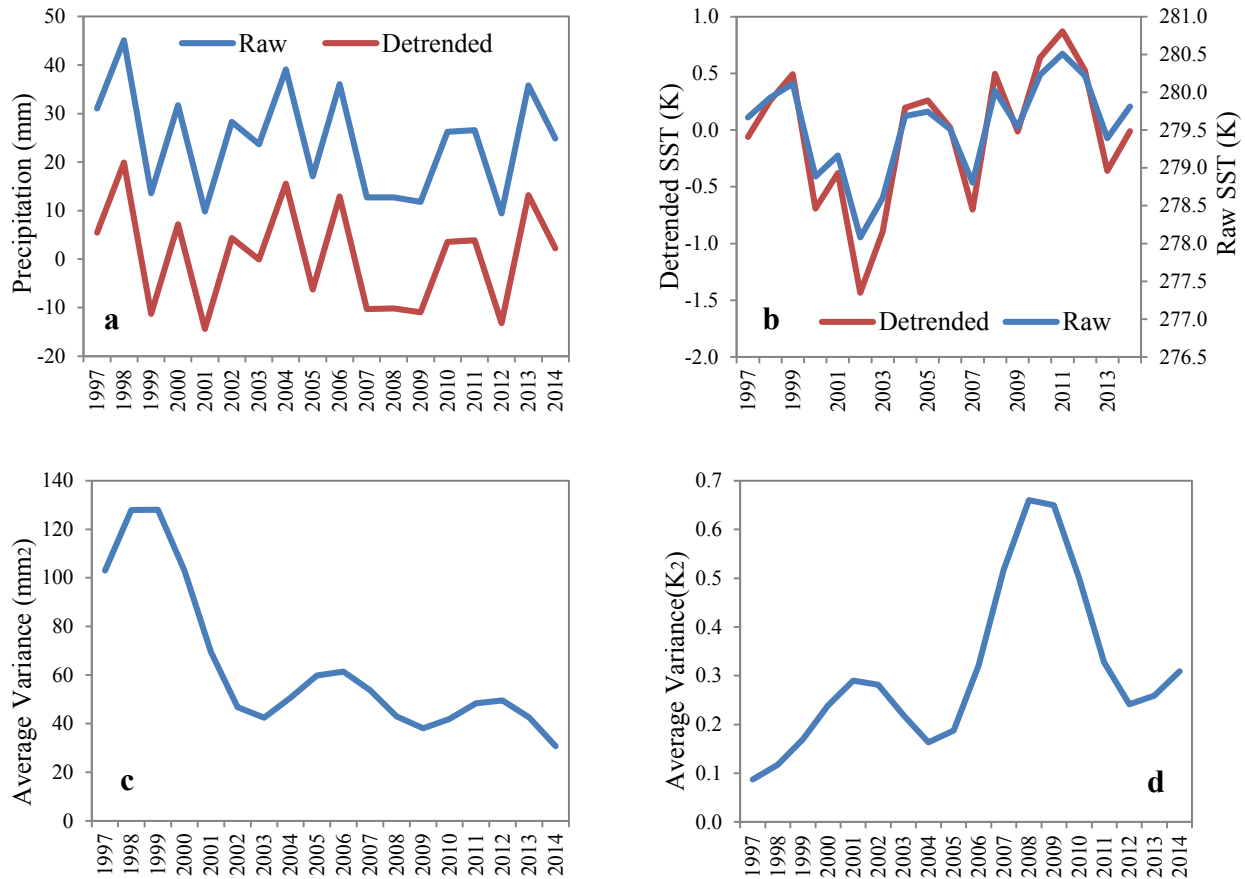
### *2.4.1. Results of Phase I: Identification the Possible Index Regions*

Prior to obtaining any meaningful results from the correlation analysis, the long-term trend of each time series needs to be removed (Wu et al., 2007). As an example, the detrended time series of SST and precipitation are compared with the raw time series for a single grid in Figure 2-9.

As shown in this figure, the detrending process has well removed the long-term trends while reserved the oscillation characteristics of original signals. Once the long-term trend is removed from the time series of all grids of SST and precipitation, the wavelet analysis is applied to capture the dominant oscillation of the detrended time series.

In addition, the differences between SAWP with the raw time series and detrended times are shown in this figure for a single grid of precipitation and SST. This figure shows that the SAWPs are totally different from the original and detrended time series, as they are reconstructed from the significant wavelet power at selected frequency band (i.e., 2-4 years), which is dedicated to maximize the dominant oscillation embedded in the raw time series. The reconstructed SAWP shown in Figure 2-9 fully demonstrate the significant long-term variability

of precipitation and SST, which is much intuitive and clear to capture the dominant oscillation of each index.



**Figure 2-9. Comparison between the raw and detrended time series for a single grid in fall season: (a) Precipitation data; (b) SST data**

As mentioned earlier, the SAWP was calculated for all 35 precipitation grids in the Upper Colorado River Basin. To consolidate all these precipitation grids into a single time series, SVD is applied to calculate the principle component. The first principle component (PC1) was able to describe 92% of precipitation variability in fall. In addition, PC1s of spring, summer, and winter were able to represent 87%, 82%, and 86% of the overall variability, respectively. The spatial

correlation patterns between PC1 of precipitation and the SAWP in all seasons is shown in Figure 2-10.

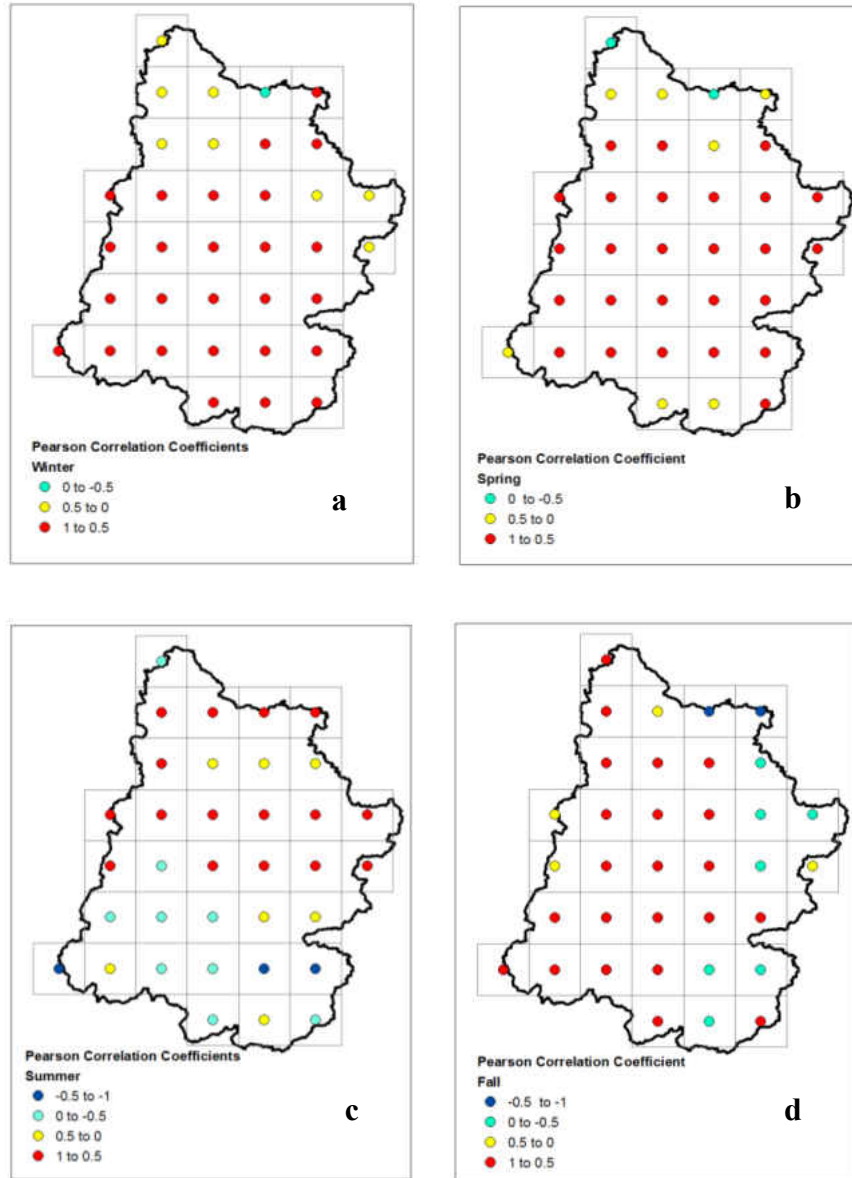
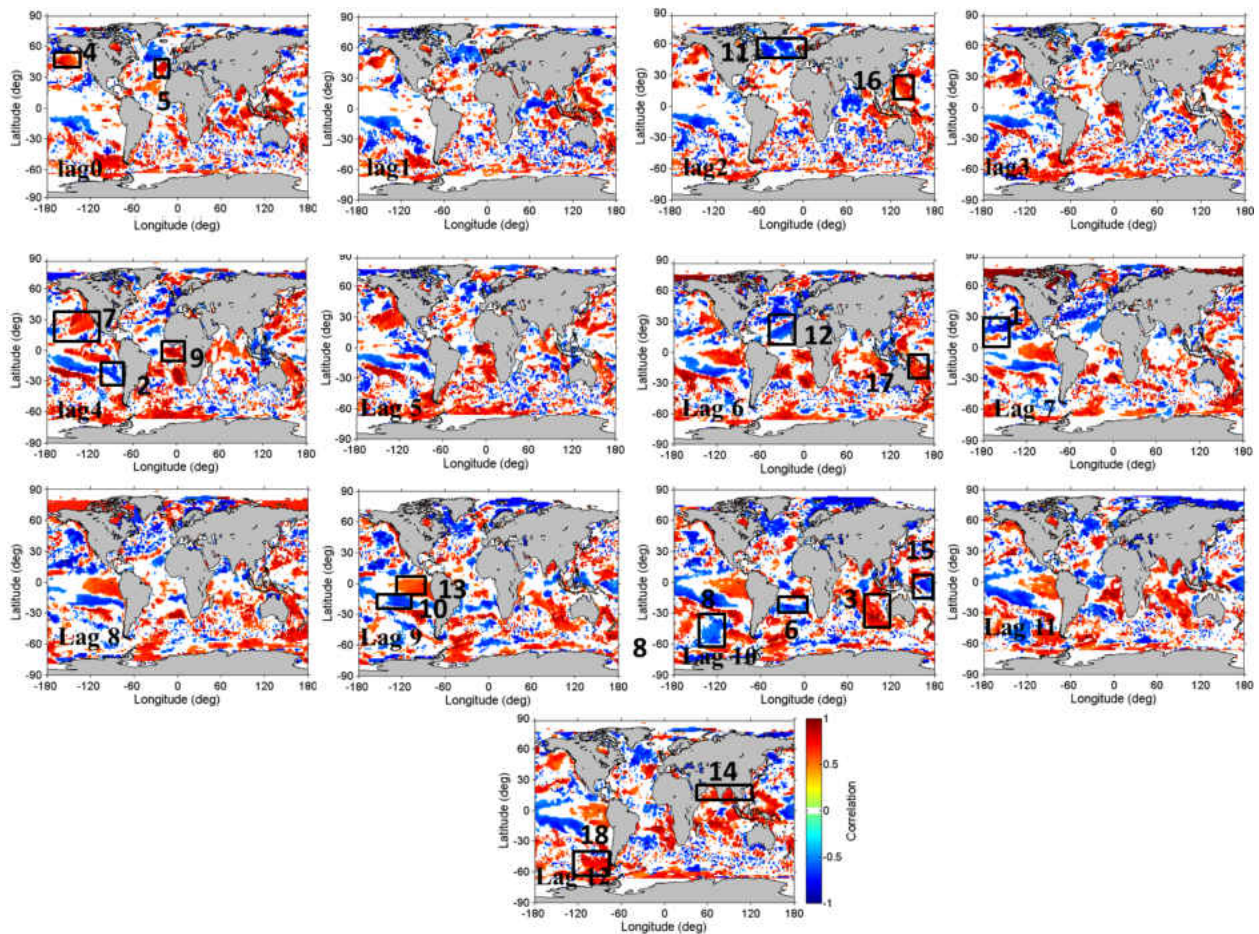


Figure 2-10. Spatial correlation between PC1 of seasonal precipitation and the SAWP of individual grid: (a) DJF, (b)MAM, (c) JJA, (d) SON.

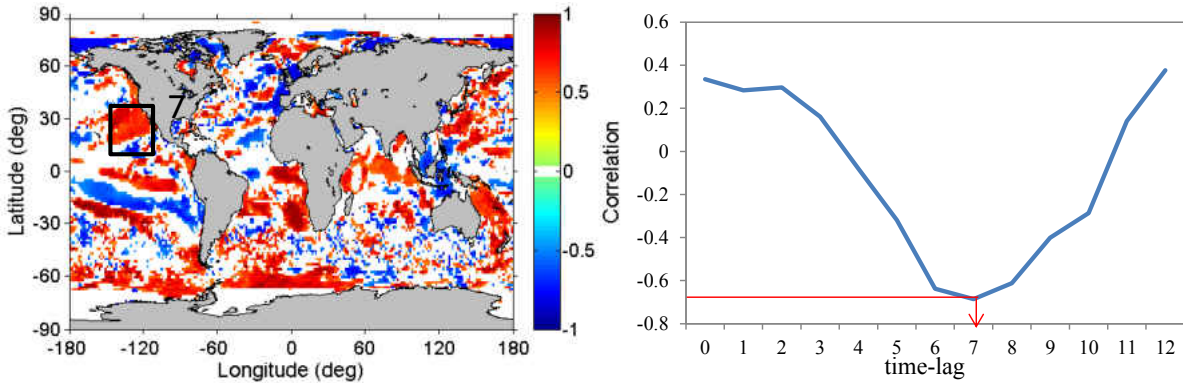
As shown in this figure, except summer, statistically high positive correlation (greater than 50%) between PC1 and the SAWP occurs across the entire region, except the pixels located at the eastern and northern part of the region in which topography affects precipitation variability. As shown in Figure 2-10, the spatial correlation patterns between PC1 and precipitation in summer is different from other seasons. This could be caused by the moist air known as “North American monsoon” that occur in July and August. It delivers high-intensity rainfall to the elevation below 7,000 feet and lower-intensity rainfall at higher elevations (USGS, 2004).

To find the association between global SST and precipitation variability in the Upper Colorado River Basin, the pixel-wise linear lag correlation was calculated between the each grid of SST SAWPs and PC1 of precipitation SAWP. As an example, spatial distributions of the correlation value in fall is shown in Figure 2-11, where the shaded colors show statistically significant correlation at the 95% confidence interval. Here, areas with a consistent significant correlation (lasting for more than 3 months) were extracted as possible forcing regions. Based on this criterion, 18 different areas (outlined regions in Figure 2-11) were extracted for winter season. These areas are termed as oceanic index regions hereafter. The same procedure is performed for all other seasons to extract the significant oceanic index regions. Different selected time lags shown in Figure 2-11 indicates how long it takes for precipitation in this region to respond to SST. Once the oceanic index regions are determined, associated time lags for each oceanic index regions are also identified by selecting the corresponding time lags with the maximum correlation coefficient between the oceanic index regions and precipitation. Figure 2-12 is presented as an example for a selected index in fall season. As shown in Figure 2-12, the most significant correlation between the selected oceanic index region and

precipitation occurred at lag 7, which suggests that the oceanic index of this region should be extracted with a 7 month time-lag compared with precipitation. Following this procedure, the associated oceanic indices were extracted for each index regions in different seasons. These oceanic indices will be further screened and ranked by simulating the possible contribution of each indices associated with the precipitation responses.



**Figure 2-11. Pixel-wise linear correlations between PC1 of precipitation SAWPs and SST SAWP of each grid in fall season. Shaded color represents significant correlation at 95% confidence intervals. The areas marked with rectangles show the identified significant regions.**

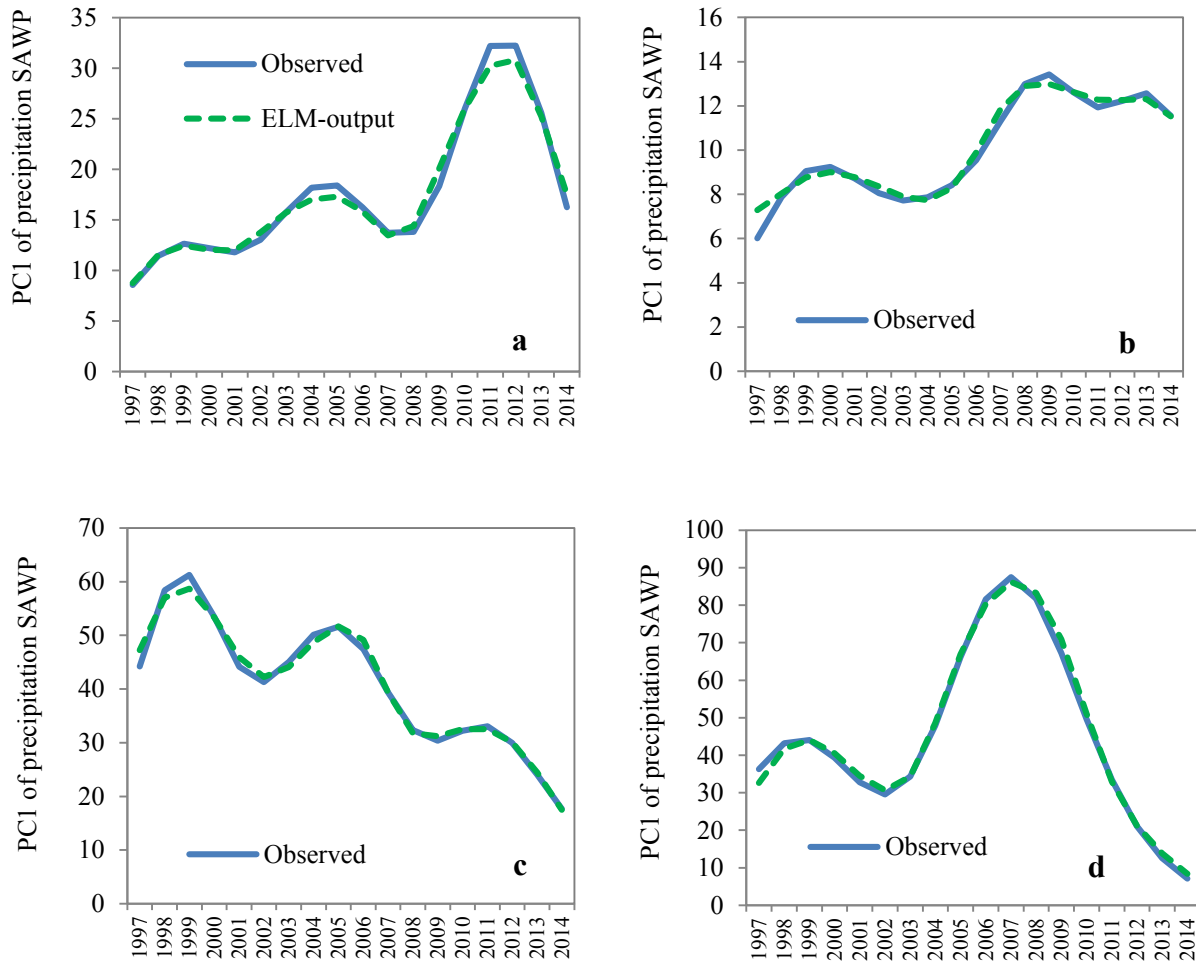


**Figure 2-12. Example of determining time lag for each oceanic index region. Right figure shows the correlation of the Index7 with PC1 of precipitation SAWPs in different time lags.**

#### *2.4.2. Results of Phase II: Screening and Ranking the Identified Oceanic Index Regions*

The main purpose of this phase of the study is finding the contribution of the identified oceanic index regions on precipitation variability over the study area in different seasons. For this purpose, the ELM algorithm is applied to assess the non-linear relationship between SST and precipitation. For each season, 200 trials were simulated, and the outputs of these 200 trials simulation were averaged to compare with the observed data as they were shown in Figure 2-13.

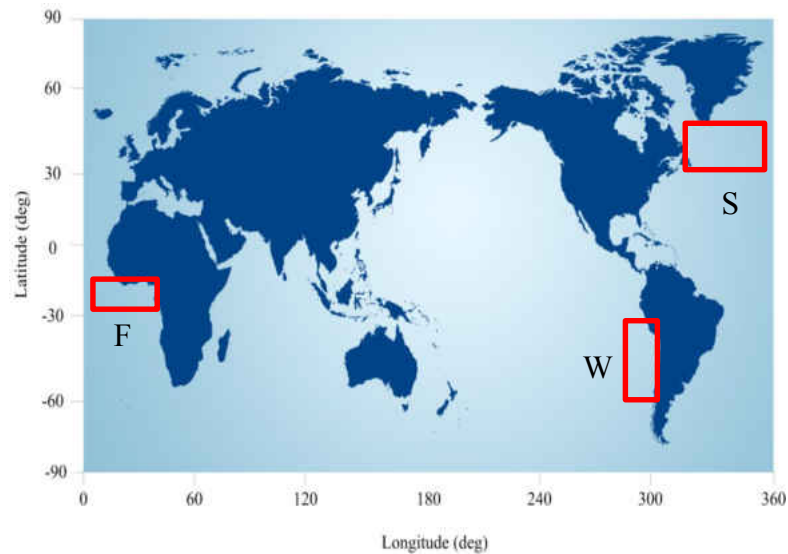
Fair agreements are observed between the simulated and observed precipitation time series over each site, which suggests the capability of ELM to characterize the complex relationships between the oceanic index regions and terrestrial precipitations. In addition, the simulated models can be used with confidence to simulate the precipitation responses to each oceanic index. As the robust ELM model is simulated, the contribution of each identified oceanic index region associated with precipitation variability over the study area can be quantified.



**Figure 2-13. Comparison between the output and the target of the ELM algorithm for (a) spring; (b) summer, (c) fall, (d) winter.**

For each season, the sensitivity analysis is repeated for at least 10 times in order to reduce the stochastic error. The oceanic index region with at least 70% contribution on precipitation variability in the most trials was selected. Based on this criterion, one index region was selected for each season. The locations of the selected index regions are shown in Figure 2-14. As shown in this figure, the same oceanic index region was detected for spring and summer seasons, which shows 84% and 75% contribution on precipitation variability over the study area, respectively.

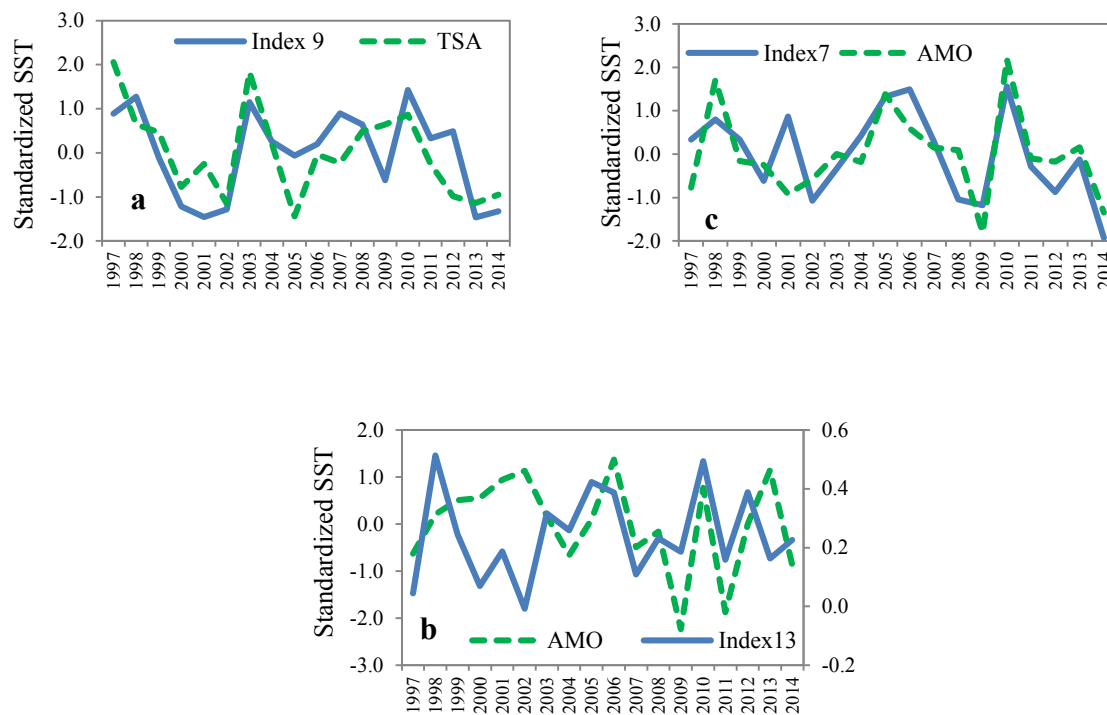




**Figure 2-14. Oceanic index regions which shows the highest contribution on precipitation variability. “S” stands for summer and spring, “F” and “W” stands for fall and winter, respectively.**

Comparing the location of the identified oceanic index region with the known teleconnection patterns indicates the compatibility of these indices with time series of Atlantic Multi-decadal Oscillation (AMO). The detected oceanic index region in summer and spring is consistent with the earlier studies by (McCabe et al., 2007; Enfield et al., 2001; Sutton and Hodson, 2005; Curtis, 2008; Mo et al., 2009; Feng et al., 2010) which show the impacts of AMO on rainfall and river flow variability and droughts in the Southwest United States and the Upper Colorado River Basin. On decadal to multidecadal time scales, AMO has the most significant association with the flow variability in the Upper Colorado River Basin compared to combination effects of Indian Ocean and Pacific Ocean variability (McCabe et al., 2007). In addition, during the warm phase of AMO, the anomalous cyclonic circulation located in the East Coast of U.S. reduces the amount of moistures transfers to the Southwest from Gulf of Mexico. Lower amount of moisture and higher temperature during boreal summer cause scarce rainfall

and intensify the potential of droughts over the Western United States (Curtis, 2008; Sutton and Hodson, 2005). The influence of AMO on rainfall over the Colorado River Basin is large when the phase of AMO is opposite to the phase of ENSO. Warm phase of AMO in a negative phase of ENSO leads to drought over the Southwest and Colorado River Basin (Mo et al., 2009). As shown in Figure 2-15, the identified oceanic index region in the both summer and spring seasons are consistent with the time series of the corresponding known teleconnection pattern (i.e. AMO). This finding suggests the efficiency of the proposed integrated approach to detect the possible linkage between global SSTs and precipitation responses.



**Figure 2-15. Comparison between the SST of the identified oceanic index regions and the known teleconnection pattern in (a) fall; (b) summer; (c) spring**

As shown in Figure 2-15, the oceanic index region which shows significant compatibility with the time series of the tropical South Atlantic Ocean (TSA) indicates the highest contribution

(79%) on precipitation variability in fall. SST warm in this region of the ocean during spring makes the Atlantic inter-tropical convergence zone (ITCZ) sensitive to any changes in interhemispheric SST gradient. ITCZ moves toward north of the equator as a result of the development of the cold tongue during June–September. As a result of the large heat capacity of the ocean mixed layer, the oceanic ITCZ lags its movement over the continent and reaching its northernmost position in September (Xie and Carton, 2004).

The non-leading teleconnection patterns located in the west coast of the South America shows the highest contribution (86%) on precipitation patterns over the Upper Colorado River Basin in winter. This region can be affected by “sea-saw” effect in sea level pressure between the Pacific Ocean and Indian Ocean (USGS, 2004). Concurrent with the low sea level pressure in the equatorial Pacific ocean and high pressure in the Indian Ocean, the trade winds over the Pacific Ocean are weakened which lead to spread the warm water from equatorial Pacific region to the west coast of South America, producing El-Nino. On the other hand, the inverse sea level pressure between two oceans intensifies the trade winds over the Pacific Ocean which leads to move the cold water along the west coast toward the eastern and central part of the Pacific Ocean, producing La-Nina (USGS, 2004).

#### *2.4.3. Results of Phase III: Seasonal Precipitation Forecasting*

The existence of 2-4 year period band of dominant oscillations in SST and precipitation time series, and the significant seasonal association between the PC1 of precipitation SAWPs and SST SAWPs suggests that precipitation over the Upper Colorado River Basin is predictable at intra-annual scale.

For each season, the significant oceanic index regions detected in the previous section are extracted from the raw SST data for the determined time lag. For instance, four-month time lag was selected for the detected index region in fall and spring, 3-month lag for winter index, and 10-month lag for summer index. The extracted SST data are averaged and normalized by scaling between 0 and 1 to use as inputs into the ANN model. In addition, the averaged raw precipitation data is considered as target of the ANN model. To determine the spatiotemporal ability of the developed method in predicting hydrological droughts over the Upper Colorado River Basin, the ANN model was simulated for each single precipitation grid. Therefore, 35 ANN models were simulated. For each grid, the ANN model was run 30 times and the average of 30 predicted values was reported as the predicted value of each grid in order to reduce the stochastic errors of the model. The time period of the seasonal forecasting model starts from 1997 to 2014 (18 years). All ANN models were trained based on 70% of inputs (12 years), and tested based on 30% of input (6years). Since the main objective of this study is developing a forecasting model to predict the hydrological droughts ahead of time, the ability of model in detecting precipitation deficits is assessed by testing the capability of the model for the year in which the lowest precipitation is recorded in all seasons. The maximum and minimum amount of precipitation of all grids within the Upper Colorado River Basin during (1997 – 2014) is shown in Table 2-4.

**Table 2-4. Minimum and maximum precipitation (mm/month) of all precipitation grids within the study area. The selected year is shown in gray.**

Season/ Year	Winter		Summer		Spring		Fall	
	Min	Max	Min	Max	Min	Max	Min	Max
1997	24	115	26.8	65	16.8	49	28	56.6
1998	17.7	57	13.8	60	23	57	29	49
1999	10	65.7	21	55	20	57.7	3.9	21
2000	16	62.6	11.7	31	17.7	45.5	16.5	44.5
2001	12.5	33.5	14	37.6	15	36	9	36
2002	7.8	49	5.7	37.7	6	43	20	60
2003	16	43.5	13.8	33	17	47	12.8	33.8
2004	19.5	64.7	10	53	16	39	30.5	66.7
2005	13.6	68	15.7	55	23	68	14.9	67.6
2006	9	85.6	12.9	37.6	15.6	57	24.8	65
2007	11.7	52	13.5	43.6	13.7	40	11	53.5
2008	22	70	10.4	29	13	52	14.6	38.6
2009	12	50.7	11	66.9	17.9	46	11	31.8
2010	9.8	53	16.9	48.6	22	55.9	14.5	50
2011	12	53	12	35.4	18	81	15	37.5
2012	5.9	32	9	27.6	7.5	36	9	31
2013	6	36	11	30.5	14.6	32.4	21	54
2014	10	41	21	47	15.7	39	17.8	38

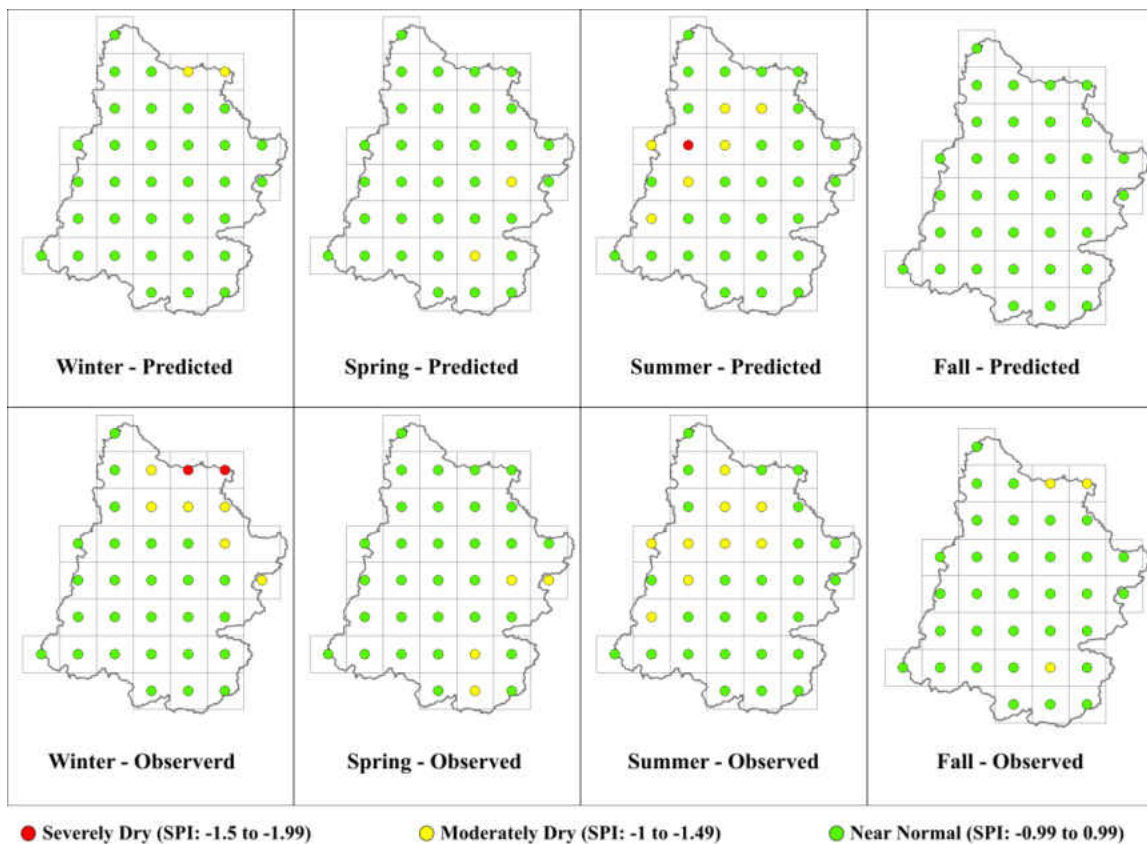
As shown in Table 2-4, although the lowest precipitation amount was recorded in one or two seasons of year 1999 and 2002, the year 2012 was the only year in which the low precipitation value was recorded in all seasons. Therefore, this year was selected to assess the ability of model in predicting the possible hydrological extremes.

As shown in Table 2-4, precipitation has significant variability within the study area in all seasons. However, comparison between the averaged, minimum, and maximum values of predicted and observed precipitation for all grids within the study area indicates the ability of the model in predicting the variability of precipitation (Table 2-5).

**Table 2-5. Comparison between the observed and predicted values over the study area in year 2012 (mm/month)**

Season	Average		Min		Max	
	Observed	Predicted	Observed	Predicted	Observed	Predicted
Winter	17	22.8	5.9	7	31	31
Spring	16.8	21	7.5	9	36	38
Summer	15	19.8	9	8.8	27.6	33
Fall	18	22	5.9	7	32	35.7

Once, the predicted values of precipitation are calculated, their corresponding SPI values are determined based on the method described in the methodology section. The results of the predicted and observed SPI values are shown for all seasons in Figure 2-16. Results show the ability of the model in detecting the states including near normal, and moderately dry.



**Figure 2-16. Spatial comparison between predicted and observed SPI**

As shown in Figure 2-16, the model was not able to detect the states of moderately dry for a few numbers of grids in all seasons, since the model overestimates the precipitation values of those grids. Also, in winter, for two pixels located on the boarder of the study area, the state of “severely dry” was predicted as “moderately dry”.

The percent error formula as given in Equation (15) was applied to assess the precision of the prediction method.

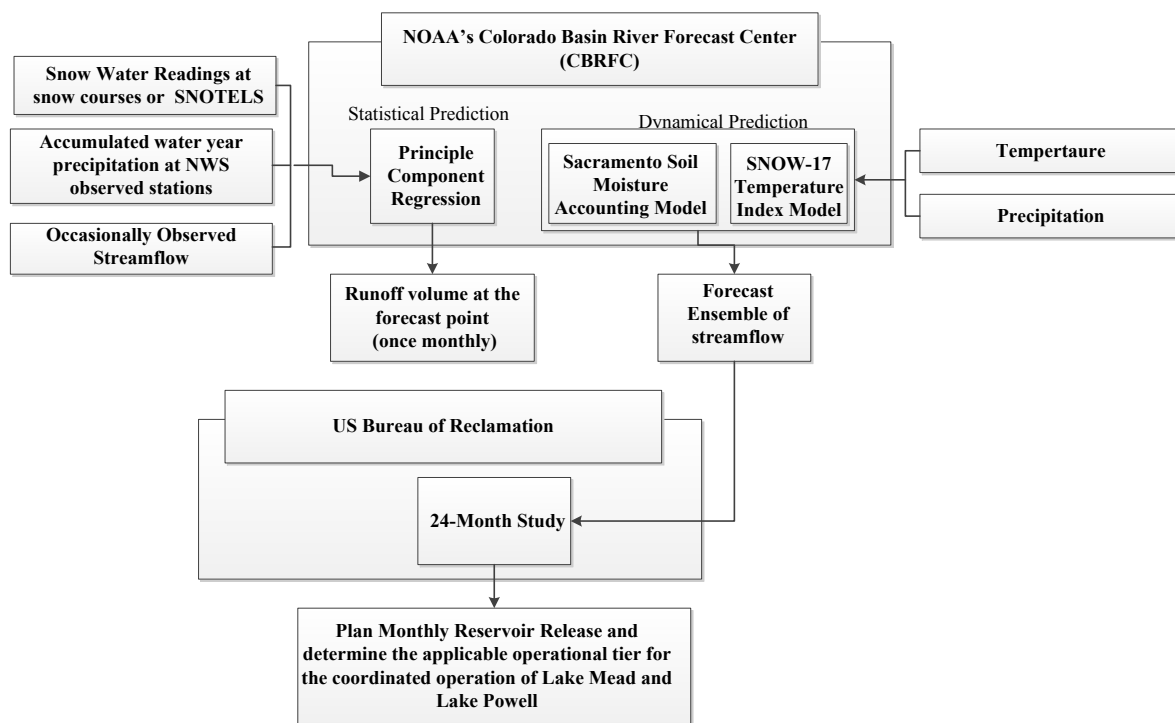
$$\%_{error} = \frac{n(S_p \neq S_o)}{N} \times 100 \quad (16)$$

where  $n(S_p \neq S_o)$  is the number of pixels in which the predicted state is not the same as the observed; N is the total number of pixels within the study area. A percentage closer to zero indicates the higher precision of the results. In this study, the percentage errors obtained for winter, spring, summer, and fall are 22%, 5%, 8.6%, and 8.6%, which confirm the reliability of the predicted method.

## 2.5. Application

Based on the latest report from the U.S. Bureau of Reclamation in 2012, use of Colorado River was exceeded the amount of water supply in multiple years as a result of increase in water demand and reduction in water supply related to the drought in the 2000’s (USRB, 2012). This issue leads to the creation of an interim operating agreement in 2007 to govern shortages and surpluses on the system by the year 2026 (USBR, 2007). The current procedure of forecasting water supply for Colorado Basin and reservoir operation management and planning models is described in Figure 2-17. As shown in this figure, the “24-month study” model is used in the U.S. Bureau of Reclamation to plan monthly reservoir release for two years ahead of time using

the predicted streamflow ensembles provided by NOAA’s Colorado Basin River Forecast Center (CBRFC). However, there is an information gap between the current results of projections and stakeholders’ requirement (Werner, 2011). Some concern has been voiced by stakeholders over the projections of the second year of the 24-month study model having low skill, and the 10th and 90th percentile forecasted monthly runoff releases of this model are not able to provide adequate information for the risk-based decision making (Warner, 2011). These concerns have directed CBRFC to improve forecasts, and Bureau of Reclamation to apply the mid-term ensemble-based operation model (MTOM) parallel to the “24-month study” model since 2010.



**Figure 2-17. The current procedure of water supply forecasts on the Colorado River and the application of forecasts**

The developed methodology for forecasting precipitation in this study can be applied to predict precipitation as one of the inputs fed into the streamflow forecasting model by NOAA.



On the other hand, having knowledge about the condition of hydrological extremes in at least next three months can provide valuable information to achieve the goal of low impact development. Currently, at the beginning of each month, the “24-month study” model is updated based on the monthly inflow forecast from CBRFC. Therefore, the results of this study can be considered as another input along with the predicted monthly inflow to update the model at the beginning of each month.

## 2.6. Conclusion

To overcome the problem of non-stationarity of the ocean-atmosphere system, wavelet analysis and empirical mode decomposition are applied in this study to exclude a wide range of unknown frequencies and intrinsic trend embedded in raw data, respectively. Wavelet analysis decomposes a time-series at time-frequency domain to determine localized variation of wavelet power and dominant modes in time series of precipitation and SSTs. In this study, the dominant wavelet power period for both precipitation and SST mainly concentrates within the inter-annual frequency band (2-4 years). To find the significant oceanic index regions, pixel-wise linear correlation maps are calculated based on the linear correlation between the averaged SAWP of precipitation and the SAWP of each SST grid. While the magnitude of correlation represents the intensity of SST forcing on precipitation, it is important to indicate how quickly the precipitation responds to SST. For this purpose, the corresponding time lag of each identified oceanic index region is calculated and the identified oceanic index regions are extracted for the corresponding lag. Since the linear-lagged correlation can only be able to show the statistically significant regions, there is still a need to develop a method which considers the non-linear relationships between oceanic index regions and precipitation. For this purpose, the integrated approach of the

advanced machine learning tool (ELM) and sensitivity analysis is applied to screen the identified oceanic indices and to calculate their contributions.

Assessing the impacts of different oceans on seasonal precipitation variability over the Upper Colorado River Basin showed the significant effects of AMO in both spring and summer. The impacts of the aforementioned teleconnection pattern on rainfall, streamflow, and moisture over the Upper Colorado River Basin was detected in earlier studies (McCable et al., 2007; Enfield et al., 2001; Sutton and Hodson, 2005; Curtis, 2008; Mo et al., 2009; Feng et al., 2010). In fall, the oceanic index region located in the tropical South Atlantic Ocean shows the highest contribution to precipitation variability. Comparing the detected oceanic index regions with time series of AMO and TSA confirms the consistency between the identified oceanic index regions and the known teleconnection patterns in most of the years within the selected time period. In addition, a non-leading teleconnection signals was detected in winter which points out the importance of considering the non-leading teleconnection signals as well as the known teleconnection patterns. Therefore, these oceanic index regions were extracted from the raw SST data to use as inputs into the ANN model.

The capability of the ANN model in predicting the hydrological drought was assessed for the year 2012 in which the recorded precipitation in all seasons were low compare to the other years in 1997-2014. Although the model was successful in capturing the wide range of precipitation variability in all seasons, the model overestimated the precipitation values for some grids located in the eastern and northern boundaries of the basin adjacent to Rocky Mountains and San Juan Mountains. The highest prediction error occurred in winter which means sea surface temperature for this season may not suffice for good prediction of precipitation and other

variables such as temperature may improve forecasts. However, results show the capability of the model in predicting the hydrological states for a season ahead of time.

## 2.7. Reference

- Abdul-Aziz A.R., Anokye M., Kwame A., Munyakazi L., Nsowah-Nuamah N.N.N., 2013, Modeling and forecasting rainfall pattern in Ghana as a seasonal ARIMA process: the case of Ashanti region, *International Journal of Humanities and Social Science*, vol. 31, no. 3, pp. 224-233.
- Aziz O.A., Tootle G.A., Gray S.T., Piechota T.C., 2010, Identification of Pacific Ocean sea surface temperature influences of Upper Colorado River Basin Snowpack, *Water Resour. Res.*, vol. 46, W07536, doi: 10.1029/2009WR008053.
- Baldwin, C., Waage, M., Steger, R., 2005, Acclimatizing Water Managers to Climate Forecasts through Decision Experiments. Climate Variations, *Climate Change, and Water Resources Engineering*, pp. 115-131. doi: 10.1061/9780784408247.ch08
- Bureau of Reclamation, 2011, Colorado River Basin Water Supply and Demand Study Seeks Input to Help Resolve Projected Future Supply and Demand Imbalances, <http://www.usbr.gov/newsroom/newsrelease/detail.cfm?RecordID=38644>.
- Bureau of Reclamation, 2015, Upper Colorado Region-Glen Canyon Dam/Lake Powell, <http://www.usbr.gov/uc/water/crsp/cs/gcd.html>.
- Cannas B., Fanni A., Sias G., Tronci S., Zedda M.K., 2005, Streamflow forecasting using neural networks and wavelet analysis, *Env. Geosci. Union*, vol. 7, pp. 45-51.
- Cannas B., Fanni A., See L., Sias G., 2006, Forecasting data pre-processing for river flow forecasting data pre-processing for river flow forecasting using neural networks: wavelet transforms and data partitioning, *Phys. Chem. Earth*, vol. 31, no. 18, pp. 1164-1171.
- Cohen M.J., 2011, Municipal deliveries of Colorado River Basin water, Pacific Institute, ISBN: 1-893790-34-7.
- Curtis S., 2008, The Atlantic multi-decadal oscillation and extreme daily precipitation over the U.S. and Mexico during the hurricane season, *Clim. Dyn.*, vol. 30, pp. 343-351, DOI:10.1007/s0038-007-0295-0.
- Dawson C.W., Wilby R., 1998, An artificial neural network approach to rainfall-runoff modeling, *Hydrological Sciences*, vol. 43, no. 1, pp. 47-66.

- Duller J.W., Kawas M.L. 1978, Stochastic models for monthly rainfall forecasting and synthetic generation, *J. Appl. Meteor.*, vol. 17, pp. 1528-1536.
- ECMWF, retrieved March 8, 2015, from <http://apps.ecmwf.int/datasets/>.
- Enfield D.B., Metas-Nunez A.M., Trimble P.J., 2001, The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental U.S., *Geophysical Research Letters*, vol. 28, no. 10, pp. 2077-2080.
- Global Precipitation Analysis, retrieved March 8, 2015, from <http://precip.gsfc.nasa.gov/>.
- Huang N. E., Shen Z., Long S. R., Wu M. C., Shih H. H., Zheng Q., Yen N.C., Tung C. C., Liu H.H., 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond. A*, vol. 454, pp. 903-995. DOI: 10.1098/rspa.
- Huang G.B., Zhu Q.Y., Siew C.K., 2006, Extreme learning machine: theory and applications. *Neurocomputing*, vol. 70, pp. 489–50. doi:10.1016/j.neucom.2005.12.126
- Huffman G.J. Adler R.F., Morrissey M.M., Bolvin D.T., Curtis S., Loyce R., McGavock B., Susskind J., 2001., Global precipitation at one degree daily resolution from multi-satellite observations, *J. Hydrometeor*, vol. 2, pp. 36-50.
- IPCC, 2012, Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.
- Elsanabary M.H., Gan T.Y., 2014, Wavelet analysis of seasonal rainfall variability of the Upper Blue Nile Basin, its teleconnection to global sea surface temperature, and its forecasting by an artificial neural network, *Mon. Wea. Rev.*, vol. 142, pp. 1771-1791.
- Feng S., Hu Q., Oglesby R.J., 2010, Influence of Atlantic sea surface temperatures on persistent drought in North America, *Clim. Dyn.*, DOI: 10.1007/s00382-010-0835-x.
- Hidalgo H.G., Dracup J.A., 2003, ENSO and PDO Effects on Hydroclimatic Variations of the Upper Colorado River Basin, *J. Hydrometeor*, vol. 4, pp. 5-23.
- Hurkmans R., Troch P.A., Uijlenhoet R., Torfs R., Durcik M., 2009, Effects of climate variability of water storage of the Colorado River Basin, *Journal of Hydrometeorology*, vol. 10, pp. 1257-1270.
- Ficklin D.L., Stewart I.T., Maurer E.P., 2013, Climate Change Impacts on Streamflow and Subbasin-Scale Hydrology in the Upper Colorado River Basin, *PLoS ONE*, vol. 8, no. 8, e71297. doi:10.1371/journal.pone.0071297

- Jacobs J., 2011, The sustainability of water resources in the Colorado River Basin, *The bridge – national academy of engineering*, vol. 41, no. 4, ISSN: 0737-6278.
- Karla A., Ahmad S., 2011, Evaluating changes and estimating seasonal precipitation for the Colorado River Basin using a stochastic non-parametric disaggregation technique, *Water Resource Research*, vol. 4, w05555, doi:10.1029/2010WR009118.
- Karla A., Miller W.P., Lamb K.W., Ahmad S., Piechota T., 2012, Using long-scale climate patterns for improving long-lead time streamflow forecasts for Gunnison and San Juan River Basins, *Hydrologic Process*, DOI:10.1002/hyp.
- Kaushik I., Singh S.M., 2008, Seasonal ARIMA model for forecasting of monthly rainfall and temperature, *Journal of Environmental Research and Development*, vol. 3, no. 2, pp. 506-514.
- Kim, T.W., Valdes J.B., Nijssen B., Roncoyolo D., 2006, Quantification of linkage between large-scale climate patterns and precipitation in Colorado River Basin, *Journal of Hydrology*, vol. 321, pp. 173-186.
- Kim T., Yoo C., Ahn J., 2008, Influence of climate variation on seasonal precipitation in the Colorado River Basin, *Stoch. Environ. Res. Risk Assess.*, vol. 22, no. 3, pp. 4111-420.
- Kisi O., 2009, Neural networks and wavelet conjunction model for intermittent streamflow forecasting, *J. Hydrol. Eng.*, vol. 14, no. 8, pp. 773-782.
- Kisi O., 2008, Streamflow forecasting using neuro-wavelet technique, *Hydrol. Process.*, vol. 22, no. 20, pp. 4142-4152.
- Kumar M.N., Murthy C.S., Sesh Sai M.V.R., Roy P.S., 2009, On the use of standardized precipitation index (SPI) for drought intensity assessment, *Meteorol. Appl.*, vol. 16, pp. 381-389, DOI:10.1002/met.136.
- Lamb K.W., Piechoto T.C., Aziz O.A., Tootle G.A., 2011, A basis for extending long-term streamflow forecasts in the Colorado River Basin, *Journal of Hydrologic Engineering*, DOI: 10.1001/(ASCE)HE.1943-5584.0000153.
- McKee, T. B., Doesken N. J., Kleist J., 1993, The relationship of drought frequency and duration of time scales, Eighth Conference on Applied Climatology, American Meteorological Society, Jan17-23, 1993, Anaheim CA, pp.179-186.
- Muller W.P., Piechota T.C., 2008, Regional analysis of trend and step changes observed in hydroclimatic variables around the Colorado River Basin, *Journal of Hydrometeorology*, vol. 9, pp.1020-1034.
- Mullon L., Chang, N.B., Imen S., Yang J., 2015, Forecasting Regional Terrestrial Precipitation with Global Nonlinear and Nonstationary Teleconnection Signals in Northeast U.S, *Journal of Climate*, in review.

- Mwale D.M., Gan T. Y., Shen S. S. P., 2004, A new analysis of variability and predictability of seasonal rainfall of Central Southern Africa for 1950-94, *Int. J. Climatol.*, vol. 24, pp. 1509-1530. doi: 10.1002/joc.1062
- NOAA, Climate Monitoring, February Release: 6 March 2015, <http://www.ncdc.noaa.gov/monitoring-references/dyk/colorado-basin-drought>
- NOAA Ocean Service Education, "The Ocean's Role in Weather and Climate". [http://oceanservice.noaa.gov/education/pd/oceans\\_weather\\_climate/](http://oceanservice.noaa.gov/education/pd/oceans_weather_climate/). Revised July 18, 2012
- Nowak K., 2011, stochastic streamflow simulation at inter-decadal time scale and implications to water resources management in the Colorado River Basin, Dissertation, University of Colorado.
- National Research Council (NRC), 2007, Colorado River Basin water management: evaluating and adjusting to hydroclimate variability, Washington, DC: National academics press, ISBN: 978-0-309-10524-8.
- Maier H.R. and Dandy G.C., 1996, Neural network models for forecasting univariate time series, *Neural Network World*, vol. 6, no. 5, pp. 747-771.
- McCabe G.J., Dettinger M.D., 2002, Primary modes and predictability of year-to-year snowpack variations in the Western United States from teleconnections with Pacific Ocean climate, *Journal of Hydrometeorology*, vol. 3, no. 1, pp.13-25.
- McCabe G.J., Betancourt J.L., Hidalgo H.G., 2007, Associations of Decadal to Multi-Decadal Sea Surface Temperature Variability with Upper Colorado River Flow, *Journal of the American Water Resources Association*, vol. 43, no. 1, pp. 183-192.
- Milly P.C.D., Dunne K.A., Vecchia A.V., 2005, Global pattern of trends in streamflow and water availability in a changing climate, *Nature*, vol. 438, pp. 347-350, doi:10.1038/nature04312.
- Mo K.C., Schemm J.K.E., Yoo S.H., 2009, Influence of ENSO and the Atlantic multi-decadal oscillation on drought over the United States, *Journal of Climate*, vol. 22, pp. 5962-5982.
- Mohammed J.R., Ibrahim H.K., 2012, Hybrid wavelet artificial neural network model for municipal water demand forecasting, *ARPN Journal of Engineering and applied science*, vol. 7, no. 8, pp. 1047-1065.
- Nourani Y., Alami M.T., Aminifar M.H., 2008, A combined neural-wavelet model for prediction of Ligvanchai Watershed precipitation, *Eng. Appl. Artif. Intell.*, vol. 22, no. 3, pp. 466-472.
- Partal T., Cigizoglu H.K., 2009, Prediction of daily precipitation using wavelet-neural networks, *Hydrological Sciences*, vol. 54, no. 2, pp. 234-246.
- Pramanik N., Panda R.K., Singh A., 2011, Daily river flow forecasting using wavelet ANN hybrid models, *Journal of Hydroinformatics*, vol. 13.1, pp. 49-63.

- Rajae T., 2011, Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers, *Science of the Total Environment*, vol. 409, no. 15, pp. 2917-2928.
- Ramana R.V., Krishna B., Kumar S.R., Pandey N.G., 2013, Monthly rainfall prediction using wavelet neural network analysis, *Water Resources Management*, vol. 27, no. 10, pp. 3697-3711.
- Rubel F., Rudolf B., 1999, Verification of GPCPs satellite-based daily precipitation analysis using corrected rain gauge data of BALTEX for the Baltic Sea catchment are. Working Group Biometeor., Institute of medical physics and biostatistics, Veterinarmedizinische Universitat, Vienna, Austria. [Available online at <http://www-med-physik.vu-wien.ac.at/staff/rub/pro/GPCC/applications.htm>]
- Solgi A., Nourani V., Pourhangi A., 2014, Forecasting daily precipitation using model of wavelet-artificial neural network and comparison with adaptive neuro-fuzzy inference system (case study: Verayneh Station, Nahavand), *Advance in Civil Engineering*, 279368, 12p.
- Sutton R.T., Hodson D.L.R., 2005, Atlantic Ocean forcing of North American and European summer climate, *Science*, vol. 309, pp. 115-118.
- Switanek M.B., Troch P.A., 2009, Improving seasonal prediction of climate variability and water availability at the catchment scale, *Journal of Hydrometeorology*, vol. 10, pp. 1521-1533.
- Thapaliyal V., 1981, ARIMA model for long-range prediction of monsoon rainfall in Peninsular India, India Meteor. Monograph, Climatology, 12, India Meteorological Department, Poona, India, 22 pp.
- Thomas B.F., 2007, Climate fluctuations and forecasting of streamflow in the lower Colorado River Basin, *JAWRA Journal of American Water Resource Association*, vol. 43, no. 6, pp. 1550-1569.
- Tokar, A., Johnson, P., 1999, Rainfall-Runoff Modeling Using Artificial Neural Networks, *J. Hydrol. Eng.*, vol. 4, no. 3, pp. 232-239.
- Tootle, G., Piechota, T., 2005, Climate Variability, Water Supply, and Drought in the Upper Colorado River Basin, *Climate Variations, Climate Change, and Water Resources Engineering*: pp. 132-142. doi: 10.1061/9780784408247.ch09
- U.S. Geological Survey (USGS), 2004, Climate fluctuations, drought, and flow in the Colorado River Basin, USGS-fact sheet 2004-3062 version2.
- USBR, 2007, Record of Decision: Colorado River Interim Guidelines for Lower Basin Shortages and the Coordinated Operations for Lake Powell and Lake Mead, U.S. Bureau of Reclamation, Department of Interior.

- USBR, 2012, Colorado River Basin Water Supply and Demand Study-Study report, U.S. Department of the Interior, Bureau of Reclamation.
- Wang W., Ding J., 2003, Wavelet network model and its application to the prediction of hydrology, *Nature and Science*, vol. 1, no. 1, pp. 67-71.
- Werner K., 2011, Climate services on the Colorado River: capabilities, gaps, and Chasms, US national oceanic and atmospheric administration, *Climate Test Bed Joint Seminar Series*, NCEP, Camp Springs, Maryland.
- Wolock D.M., McCabe G.J., 1999, Estimates of Runoff Using Water-Balance and Atmospheric General Circulation Models, *J. Amer. Water Resour. Assoc.*, Vol. 35, pp. 1341-1350.
- Wu Z., Huang N.E., Long S.R., Peng C.K., 2007, On the trend, detrending and variability of nonlinear and nonstationary time series, *PNAS*, vol. 104, no. 38, pp. 14889-14894. doi: 10.1073/pnas.0701020104
- Wu R., Kirtman B. P., Pegion K., 2008, Local rainfall-SST relationship on sub-seasonal time scales in satellite observations and CFS, COLA Technical Report 266, 17p.
- Xie S.P., Carton J.A., 2004, Tropical Atlantic variability patterns, mechanisms, and impacts, Earth's climate: the ocean-atmosphere interaction, *Geophys. Monogr.*, vol. 147, Amer. Geophys. Union, pp. 121-142.



# **CHAPTER 3: DEVELOPING THE REMOTE SENSING-BASED EARLY WARNING SYSTEM FOR MONITORING TSS CONCENTRATIONS IN LAKE MEAD**

## **3.1. Introduction**

Generally, sustainable water resources management is often built on continuous water quality monitoring efforts and predictive analyses at regional and local scales (Kundzewicz et al., 2007). Manual sampling or automatic sampling in a monitoring network of point measurements is costly, laborious, and time consuming (Chang and Xuan, 2013). In addition, it is not able to characterize dynamics of water quality parameters such as total suspended solids which has a wide range of spatiotemporal variability (Chen et al., 2015). Satellite remote sensing through rigorous ground-truthing in calibration and validation makes it possible to analyze spatiotemporal variability of TSS concentration in surface water bodies. Having information about the spatiotemporal variation of TSS is of great importance, since high TSS concentration increase water temperature, and decrease levels of dissolved oxygen (DO) in water (EPA, 2012). In addition, high level of turbidity scatters light in different directions and decreases the light penetration into the depth of water which inhibits photosynthesis (Chesapeake Bay Program, 2012). Furthermore, pollutants can attach suspended particles and enter to water (Perlman, 2014). In drinking water treatment process, treating turbid water is costly and can result in odor and taste problem (FONDRIEST, 2015). In addition, turbidity can cause clogging tanks and pipes (WHO, n.d.).

Prior studies show the potential of high temporal and spatial resolutions of satellite sensors for TSS retrieval. Although several satellite sensors such as Landsat Thematic Mapper

(TM) (Khorram, 1985; Lavery et al. 1993; Mertes et al., 1993; Keiner and Yan, 1998; Panda et al. 2004; Wang et al. 2008; Montanher et al., 2014), MODIS (Chang et al., 2009; Zhang et al., 2010; Petus et al., 2010; Katlane et al., 2012; Villar et al., 2013; Son et al., 2014; Chen et al., 2015), and SeaWiFS (Tassan, 1994; Myint and Walker, 2002; Fettweis et al., 2007) can be used to retrieve TSS from remote sensed images, the retrieval of water quality is limited by the spatial, temporal, and spectral resolutions of the satellite sensors. In other words, the water quality information and its usefulness depend on the fly-over time interval and types of the sensors used.

For this purpose, the concept of data fusion, which was firstly introduced in the 1990s, was applied to monitor inland water bodies taking advantage of high spatial and temporal resolution available from different satellite sensors (Zhang et al. 2002; Koponen et al. 2002; Zhang et al., 2003). Although the recently developed integrated data fusion and mining (IDFM) technique provides a reliable near real-time monitoring system (Chang et al. 2014a; Chang et al. 2014b), there is still an acute need to develop a forecasting capacity to assess the changing conditions of water quality for both watershed management and drinking water supplies. Developing such new tools will require innovative creation of predictive methods leading to continuous evaluation of management outcomes.

There are a wide range of quantitative forecasting methods to handle both linearity and non-linearity of time series. However, the developed forecasting methods can cause drawbacks in real-world applications. For instance, traditional statistical models are not able to handle highly non-linear patterns of time series in a powerful way. Therefore, evolutionary algorithms were developed to bridge the gap of traditional statistical models. Although the competence of evolutionary algorithms in forecasting the chaotic and nonlinear time series were established by several authors (Lion et al. 1995; Mahfoud and Mani 1996; Szpiro 1997; López et al. 2000),

these models cannot simultaneously capture both linear and non-linear patterns. For this purpose, the soft computing and hybrid methods were developed. Whereas soft computing is applicable to investigate the vague and unclear relationship between variable and results, the hybrid methods combine statistical and evolutionary algorithms to capture both linearity and non-linearity embedded in time series data.

In this paper, the integrated data fusion and mining (IDFM) technique was applied to develop a near real-time monitoring system for daily nowcasting of the TSS concentrations. Then a nonlinear autoregressive neural network with external input (NARXNET) model was selected and applied for forecasting analysis of the changes of TSS concentrations over time on a rolling basis onward using the IDFM technique. Forecasted information was utilized to develop an early warning system which provides the spatiotemporal variation of TSS in source water around the plant intake and leads to the assistance in the operations of water treatment plants during episodes of high TSS concentration. Implementation of this remote sensing-based early warning system was assessed by a case study at Lake Mead.

The following scientific questions are explored: (1) Is the developed nowcasting model able to estimate TSS concentration at the inflows into the lake for the dates in which sampling was not performed? (2) Is there a linkage between the occurrences of high TSS concentration within the lake and the increase in the volumetric flow rate of inflows into the lake? 3) Does the applied nowcasting method have the capability of finding the probable impact of each source (i.e. transported sediments, forest fire events, and transported urban stormwater runoff and treated wastewater effluent) on the increase of TSS concentration in the lake on a daily basis? 4) What is the reliability of the developed forecasting method in estimating TSS concentration at the location of the drinking water intake?

### 3.2. Study Area

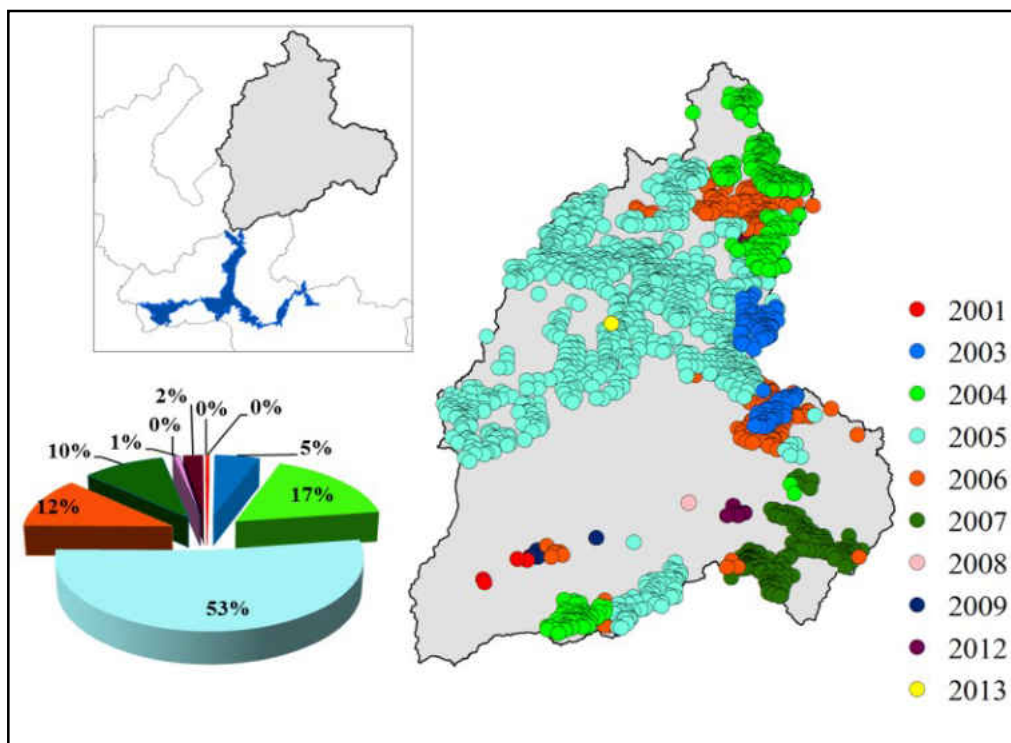
Lake Mead, formed by the Hoover Dam along the Colorado River, is located on the border of Nevada and Arizona (Figure 3-1). The lake is the largest reservoir in the United States and the 16<sup>th</sup> largest reservoir in the world. There are three inflows into Lake Mead: the Colorado River in the east, the Muddy and Virgin Rivers in the north, and the Las Vegas Wash in the west.



**Figure 3-1. Location of Lake Mead and all rivers inflow into the lake**

Water quality in Lake Mead, Nevada in the arid western United States varies in response to areal forest fires on the northern side, the treated wastewater effluent discharge into the Las Vegas Wash on the western side, and sediments transported from the upper Colorado River on the eastern side. 98 percent of the inflow into Lake Mead is originated from the upper Colorado River which transports discharged sediments from the upper Colorado River Basin (Edward, 1991). Although the amount of sediment transported into Lake Mead from the upper Colorado

River has declined by 90% since the completion of the Glen Canyon Dam in 1963, it is still necessary to assess the effects of the imported sediments into the lake on the water quality of the lake (National Park Service, 2014). On the other hand, there are a number of forest fire events in the recent decade in the upper part of the lower Virgin River Watershed, having caused drastic land-use and land-cover changes (Figure 3-2). Higher rainfall and runoff caused by climate change may lead to considerable erosion and may affect the TSS concentration in Lake Mead (Megal et al., 2009).



**Figure 3-2. Location of recorded forest fire events (right) and the percentage of forest fire events which occurred in each year (lower left) in the Virgin River Watershed – (Data source: USDA Forest Service, 2014)**

As shown in Figure 3-2, the majority of forest fire events occurred in 2005. Even though forest fires have not been documented as a cause for increases in TSS, the Lower Virgin River was identified in the 2006/2008 impaired waters list from the Arizona Department of

Environmental Quality due to the detection of contaminants which included TSS (Arizona Department of Environmental Quality, 2006).

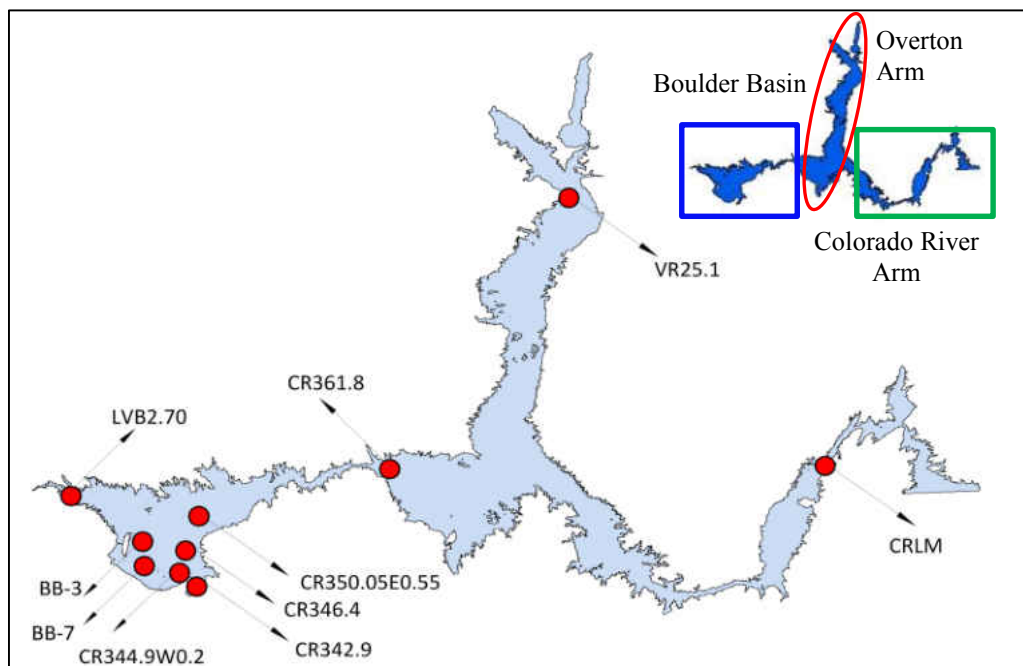
Another tributary, Las Vegas Wash, in the west (See Figure 3-1) transports urban stormwater runoff and treated wastewater effluent into Lake Mead and deteriorates the water quality in the lower Las Vegas Wash. In 2002, the lower Las Vegas Wash was included in the 303(d) list of impaired waters in Nevada due to high TSS impact. “The term "303(d) list" is short for the list of impaired and threatened waters (stream/river segments, lakes) that the Clean Water Act requires all states to submit for EPA approval every two years on even-numbered years” (EPA, 2012b). Even though the lower reach was removed from the aforementioned list in 2004, TSS and other pollutants remain a concern for water quality of the Lake.

In addition to the aforementioned factors, the water level in Lake Mead has declined by about 33 m (100 feet) from 2000 to 2008 as a result of a decade of drought. The hydrological processes and the aforementioned factors have led to the increased concentration of total suspended solid (TSS) in the lake. Lake Mead is the principal source of water in the Las Vegas metropolitan region. The surface water is treated at two treatment facilities using the conventional flocculation, coagulation, and filtration processes. Produced water is disinfected by chlorination or ozonation before being distributed to the region’s consumers. For any water treatment operations, high TSS and algal blooms in source water are always a threat leading to poor water treatment efficiency. At this moment, the development of early warning system for at-risk water quality conditions in surface water bodies as the drinking water source is not fully available by the use of an integrated satellite remote sensing and sensor networks.

### 3.3. Input Data

#### 3.3.1. Ground-truth Data Collection and Analysis

Historical water quality data from collected samples at 10 local monitoring stations within the lake was obtained from the lower Colorado River Regional Water Quality Database. The locations of these stations are shown in Figure 3-3. In addition, the names of these stations, number of samples, and title of performing agencies are shown in Table 3-1. The mean, maximum, and minimum TSS concentrations at each station are also presented in this table.



**Figure 3-3. TSS sampling stations by sampling agencies (Data source: Lower Colorado River regional water quality database)**

Based on the location of these stations, they can be geographically classified into three categories, Boulder Basin, Overturn Arm, and Colorado River Arm. The regions associated with these categories are shown in Figure 3-1. As shown in Table 3-1, the average TSS concentration

was calculated based on a limited number of samples (i.e. at most 17 samples) and in different time periods. According to this table, TSS concentration at Boulder Basin was relatively constant. However, the highest variation in TSS concentration was detected at the location of the inflows into the lake (i.e. LVB2.7, CRLM, and VR25.1). Therefore, it is necessary to develop a method to obtain a holistic overview of spatiotemporal variation of TSS within the lake.

**Table 3-1. Ground-truth data within the Lake utilized by this study (Source: lower Colorado River regional water quality)**

Location	Station	No. of Samples	Time Period	Sampling Agency	value (mg·L <sup>-1</sup> )		
					Min	Max	Mean
Boulder Basin	BB-3	12	2005-2007	City of Las Vegas	5	5	5
	BB-7	6	2007	City of Las Vegas	5	5	5
	CR342.9	7	2007	City of Las Vegas	5	5	5
	CR344.9W0.2	7	2007	City of Las Vegas	5	5	5
	CR346.4	12	2005-2007	City of Las Vegas	5	5	5
	CR350.0SE0.55	12	2005-2007	City of Las Vegas	5	5	5
	CR361.8	7	2007	City of Las Vegas	5	6	5.3
	LVB2.70	2	2013	U.S. Bureau of Reclamation (Boulder City)	6.7	7.2	6.95
Colorado River Arm	CRLM	3	2003-2004	Southern Nevada Water System	6	12	9.3
Overton Arm	VR25.1	17	2007-2009	U.S. Bureau of Reclamation (Denver)	3	44	8.3

### 3.3.2. Remote Sensing Data Collection and Pre-processing

Two types of satellite sensors, including MODIS and Landsat TM/ETM+, were applied in this study. Although the daily revisit time of MODIS provides high temporal resolution, land bands of MODIS have poor spatial resolution. On the other hand, Landsat TM/ETM+, with its



16-day revisit cycle, has a high spatial resolution of 30 m. Therefore, we applied the IDFM method developed by previous studies (Chang et al. (2014a); (2014b)) to fuse MODIS and Landsat images and generate a synthetic image which has both high spatial and temporal resolutions for nowcasting analysis. The number of images utilized in this study and their selected bands are shown in Table 3-2. Only cloud-free Landsat images close to the sampling dates were selected. In addition, cloud-free MODIS images corresponding to ground-truth dates and selected Landsat images were gathered.

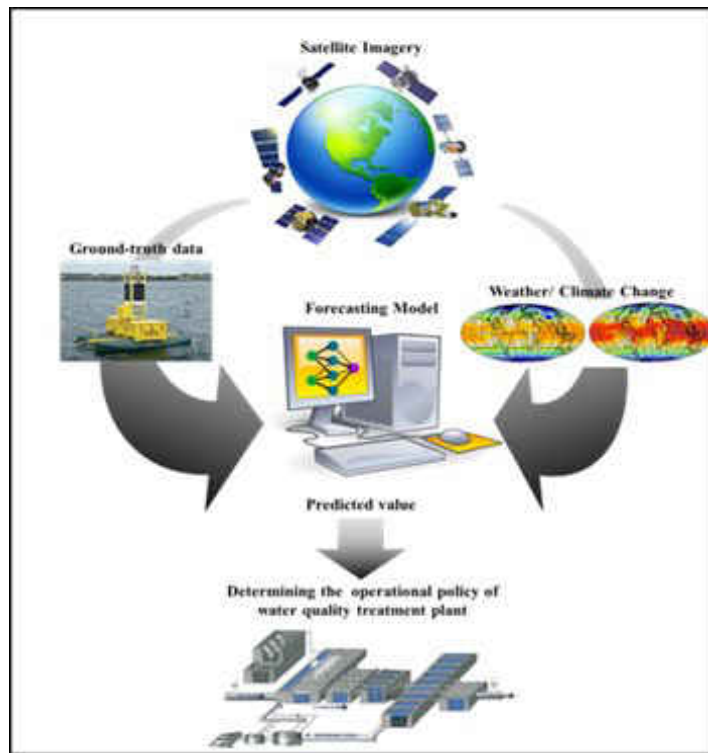
As the first step in the pre-processing satellite imagery, atmospheric correction should be done to remove the effects of atmospheric gases and aerosols on the reflectance values of the images. MODIS land surface reflectance (MOD09GA) and surface reflectance for climate data record (CDR) for Landsat 5 (TM) and Landsat 7 (ETM+) were obtained from the USGS MODIS reprojection tool web interface (MRTWeb, 2010) and the United States Geological Survey (USGS, 2014), respectively. MOD09 product has been radiometrically and atmospherically calibrated (Vermote and Kotchenova, 2011). In addition, the surface reflectance CDR applied in this study for Landsat is generated from software in which the MODIS atmospheric correction routines are applied to Level-1 Landsat 4-5 (TM) or Landsat 7 (ETM+) data. Next, all satellite images were reprojected to UTM zone 11N and cropped to include only the study area. All images were visually compared in order to make sure they were co-registered. Lastly, the spatial resolution of the applied images needs to be the same to produce fused images. For this purpose, MODIS images were resampled to 30 m to be the same pixel size as Landsat images.

**Table 3-2. Number of satellite images for data fusion with selected bands utilized in this study**

Satellite	Spatial Resolution (m)	Band	Number of Satellite Imagery
MODIS	500	1-4, 6-7	139
Landsat 5 TM Landsat 7 ETM+	30	1-5,7	27

### 3.4. Methodology

The proposed methodology includes two phases. The main objective of phase one is based on finding the relationship between TSS concentration data available in the lake and surface reflectance data embedded in satellite imagery from the years 2005 to 2009. For this purpose the IDFM method described by Chang et al., (2014a); (2014b) was applied to develop a near real-time monitoring system for daily nowcasting TSS concentrations.



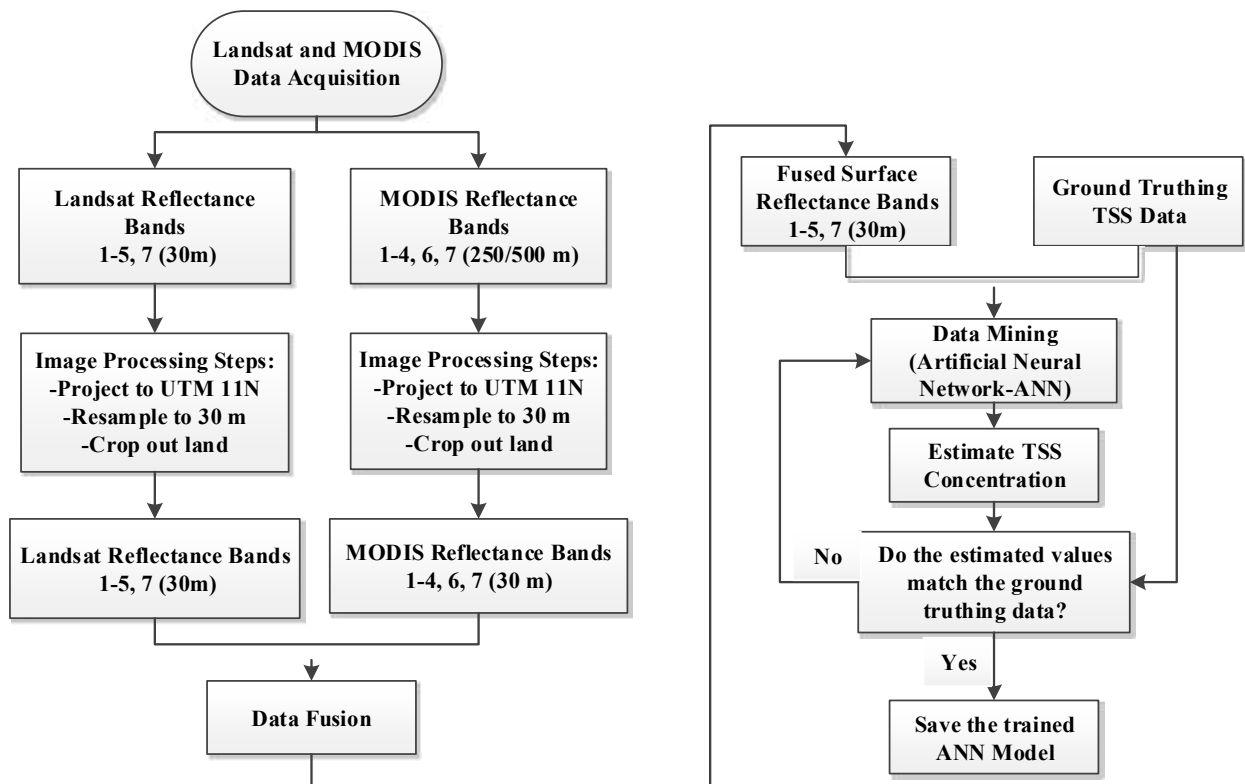
**Figure 3-4. Schematic diagram of the developed human-machine system in this study**

Next, in phase II, which is designed for predicting TSS concentration for a day ahead of time, the NARXNET model was applied to forecast TSS concentration on a rolling basis onward using the meteorological data and the trained ANN model from the phase I. As shown in Figure 3-4, the forecasting model provides information for the water treatment plant operator in order to tackle unexpected events before they occur.

#### *3.4.1. Phase I: Nowcasting*

The framework of the IDFM method developed by Chang et al., (2014a); (2014b) is shown in Figure 3-5. First, all MODIS and Landsat images were collected and processed for the dates closest to those of ground-truth data. In the data fusion step in Figure 3-5, the processed pairs were defined as inputs to the spatial and temporal adaptive reflectance fusion model (STAR-FM) algorithm developed by the National Aeronautics and Space Administration (NASA) (Chang et al., 2014a; Chang et al., 2014b; Gao et al., 2006). Using two pairs of images for both before and after the desired date may increase the accuracy of the produced fused images (Gao et al., 2006). The STAR-FM algorithm produces fused images corresponding to ground-truth dates. Since ground-truth data were collected during 2005-2009, the fused images were produced for the same time period. Then, the spectral reflectance values of all bands were extracted at the location of the ground-truth data throughout the lake. The extracted spectral reflectance values of all bands corresponding to ground-truth dates and the ground-truth data were applied as the inputs and target of the artificial neural network model, respectively. The MATLAB ANN toolbox with feed-forward neural network trained by the Levenberge-Marquardt back propagation algorithm with 30 neurons in the hidden layer was applied in this study. In order to overcome the problem of overtraining which is caused by the ANN model's tendency to

adopt to any data, even noise, the model data was split into two datasets entitled training and testing (Lohninger, 1999). In this study, 60% of all input data were used for training, and 40% for testing the model. The net with the best performance which is called the trained ANN model from now on was saved to be applied in the forecasting procedure.



**Figure 3-5. Methodological flowchart of the integrated data fusion and data mining method (obtained from Chang et al., 2014a)**

### 3.4.2. Phase II: Forecasting

#### 3.4.2.1. Model Selection

Brief description about the previous forecasting methods applied in the field of water resource management and their pros and cons are presented in Table 3-3. In this study, a forecasting method is needed to predict the TSS concentrations based on the long term memory

of the spatiotemporal relationship between TSS concentration and extracted surface reflectance values within the lake. For this purpose, artificial neural network (ANN), which has powerful pattern classification and recognition capabilities by learning and generalizing from experience with memory effect, was applied in our forecasting scheme. Although the forecasting ability of ANNs can be improved by proposing a conjunction models via the hybrid approach, the complexity of hybrid model which may affect the prediction accuracy leads us to apply recurrent dynamic ANNs instead. The selected ANN is the NARXNET, which is able to learn the dynamic relationship of time series. The output of NARXNET not only depends on the current input, but also depends on the current and previous inputs and outputs of the network. The results of prior studies showed the higher forecasting accuracy of NARXNET, and its promising qualities for dynamic system applications (Arbain and Wibowo, 2012; Menezes and Barreto, 2006; Diaconescu, 2008).

**Table 3-3. Selected forecasting methods in water resources management and their pros and cons**

Model	Reference	Advantage	Disadvantage
Traditional Statistical Model (AR*, MA*, ARMA*, ARIMA*)	<ul style="list-style-type: none"> <li>• Islam and Sivakumal, 2002</li> <li>• Girmaldi et al., 2005</li> <li>• Soltani et al., 2007</li> <li>• Chattopadhyay and Chattopadhyay, 2010</li> </ul>	<ul style="list-style-type: none"> <li>• Simple and easy to use                             <ul style="list-style-type: none"> <li>• Flexible</li> </ul> </li> <li>• Multiple variables</li> </ul>	<ul style="list-style-type: none"> <li>• Restrictive assumptions</li> <li>• Less accurate</li> <li>• Inappropriately for nonlinear cases</li> <li>• Sensitivity to outliers</li> <li>• They are only good for short-term prediction</li> </ul>
Evolutionary Computing (ANN*GP*, GA*)	<ul style="list-style-type: none"> <li>• Principe et al., 1992</li> <li>• Kuligowski and Barros, 1998</li> <li>• Oliveira et al., 2000</li> <li>• Silvermand and Dracup, 2000</li> </ul>	<ul style="list-style-type: none"> <li>• Potential to solve pattern-recognition problems</li> <li>• Competence in forecasting chaotic time series                             <ul style="list-style-type: none"> <li>• Robust with respect to noisy evaluation functions</li> <li>• Easy adoption</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Inability to handle both linear and non-linear patterns equally well.</li> <li>• Premature convergence to local extreme</li> <li>• Time consuming</li> <li>• Configuration dependent</li> <li>• Over-fitting</li> <li>• Does not yield to a mathematical equation</li> </ul>
Soft Computing (Fuzzy theory, Grey system theory)	<ul style="list-style-type: none"> <li>• Yu et al., 2001</li> <li>• Mahabin et al., 2003</li> <li>• Trivedi and Singh, 2005</li> <li>• Altunkaynak et al., 2005</li> </ul>	<ul style="list-style-type: none"> <li>• Fuzzy logic is applicable when the relationship between variable and results are vague.</li> <li>• Gray system theory investigate the relationship of variable and results with unclear linear relationship, uncertain mechanisms, and insufficient information                             <ul style="list-style-type: none"> <li>• The ability to implement expert knowledge into the fuzzy rules</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• The lack of self-learning capability of fuzzy logic                             <ul style="list-style-type: none"> <li>• Use of large number of data in the fuzzy logic</li> <li>• Gray forecast is not suitable for dealing with fluctuating data.</li> <li>• In long term forecast, gray theory does not show its advantage.</li> </ul> </li> </ul>

\*AR: autoregressive- MA: moving average- ARMA: autoregressive-moving-average; ARIMA: autoregressive integrated moving-average ANN: artificial neural network- GP: genetic programming, GA: genetic algorithm

Model	Reference	Advantage	Disadvantage
Hybrid Method (ARIMA -ANN, ARIMA-GP, Wavelet Transform and ANN, ANN and Fuzzy Logic, GP and Fuzzy Logic, and Grey system-Fuzzy logic, and dynamic programming.	<ul style="list-style-type: none"> <li>• Chang et al., 2002</li> <li>• Cheng et al., 2002</li> <li>• Nayak et al., 2004</li> <li>• Dixon, 2005</li> <li>• Díaz-Robles et al., 2008</li> <li>• Faruk, 2010</li> <li>• Lee and Tong, 2011</li> <li>• Adamowski and Chan, 2011</li> <li>• Kim and Valdés 2003</li> <li>• Zhang 2003</li> <li>• Aladag et al. 2009</li> <li>Egrioglu et al. 2009</li> </ul>	<ul style="list-style-type: none"> <li>• ARIMA-ANN can handle both linear and nonlinear patterns</li> <li>• ANN-Fuzzy eases the model building processes.</li> <li>• In contrast to most time series modeling techniques, ANN-Fuzzy does not require the model structure to be known priori.</li> <li>• Hybrid methods outperform evolutionary computing models in terms of computational speed, forecast errors, efficiency, and peak estimation.</li> </ul>	<ul style="list-style-type: none"> <li>• Explaining the hidden layer of ANN is difficult in the ARIMA-ANN model,</li> <li>• ARIMA-ANN does not yield to a mathematical equation.</li> <li>• In wavelet transform-ANN, the accurate wavelet decomposition of series is still a problem and it depends heavily on the choice of wavelet basis function.</li> <li>• In wavelet transform-ANN, disregarding denoising may influence the prediction accuracy.</li> </ul>

\*AR: autoregressive- MA: moving average- ARMA: autoregressive-moving-average; ARIMA: autoregressive integrated moving-average ANN: artificial neural network- GP: genetic programming, GA: genetic algorithm

#### 3.4.2.2. Forecasting Method

Framework of the developed forecasting model in this study is shown in Figure 3-6. The essential steps include:

**Step (1) data acquisition:** MODIS and Landsat images for a minimum of 30 days before the first forecasting date are collected and pre-processed according to the IDFM procedure described in the previous section.

**Step (2) data fusion:** STAR-FM was applied to fuse high-frequency reflectance data from MODIS with fine spatial resolution data from Landsat for all cloud-free days for at least 30 days prior to the forecasting date.

**Step (3) extract reflectance values:** To retrieve TSS values from satellite imagery, the surface reflectance values of all 6 bands of fused images were extracted at the location of station (BB-7) which is the closest station to the drinking water intake.

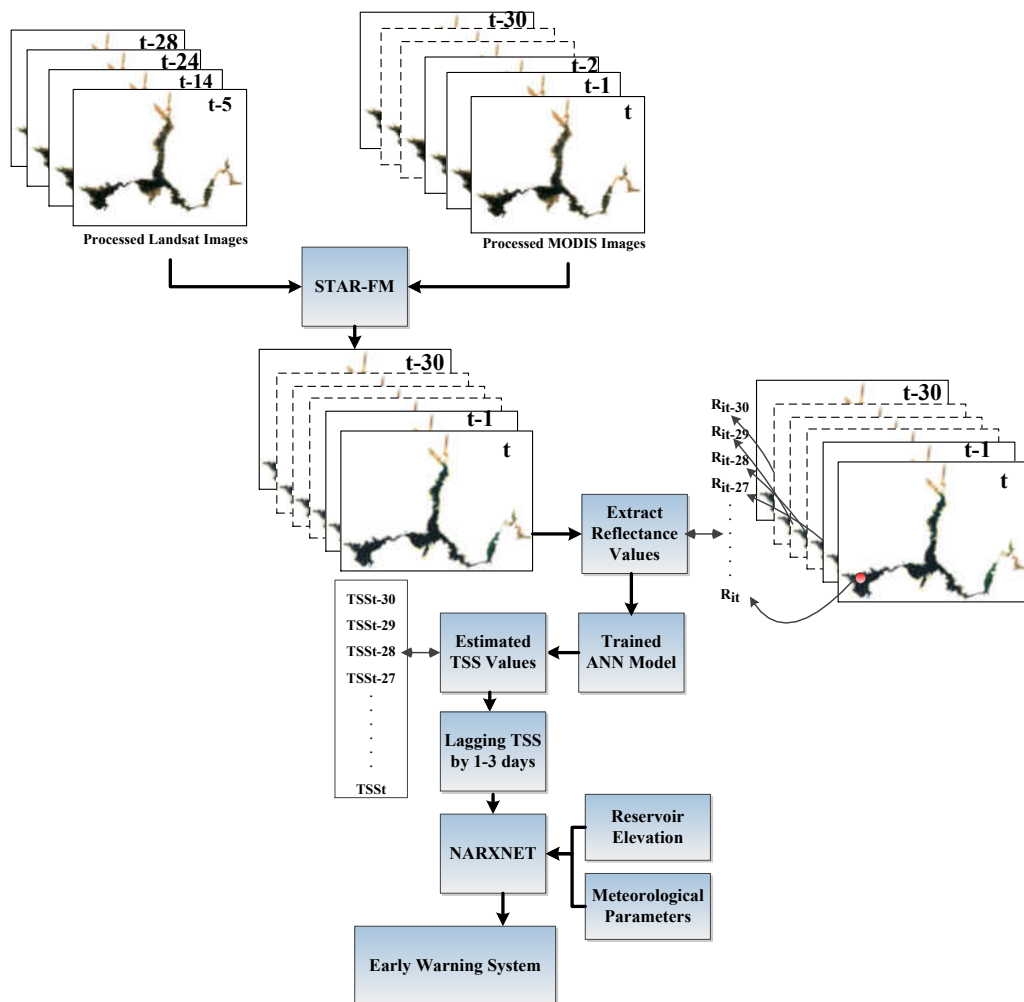


Figure 3-6. The framework of the developed water quality forecasting method



**Step (4) train the ANN model:** The trained ANN model, which was described in the IDFM section, was applied to estimate daily TSS values in the developed forecasting method using the extracted spectral reflectance values from Step (3). The main reason for using the trained ANN model in the developed forecasting model is to keep the long term memory (i.e. 2005-2009) of the relationship between TSS concentration and spectral reflectance values in Lake Mead

**Step (5) apply the ANN model:** The MATLAB ANN toolbox entitled NARXNET model was applied for forecasting TSS values based on different combinations of three types of input variables, surface reflectance values, local meteorological parameters, and water level in the reservoir. The impact of climate change due to the long-term drought, which started in the year 2000, on water quality in Lake Mead is therefore considered in the last two aforementioned input variables. The time series of all input variables include a minimum of 30 days before the first forecasting date. NARXNET can be used to predict the TSS values for a day ahead of time. The output of NARXNET is dependent on not only time series of input variables but also the outcome of the model that may become the input of the model iteratively (Figure 3-7). This relationship can be written as follows (Cai et al., 2012):

$$y(t) = f(y(t - 1), \dots, y(t - d), x(t - 1), \dots, x(t - d)) \quad (1)$$

where  $y(t)$  is the predicted time series (TSS time series),  $x(t)$  is the time series for each of the input variables (i.e. spectral reflectance values, meteorological parameters, and reservoir elevation),  $d$  is the input and feedback delay node.

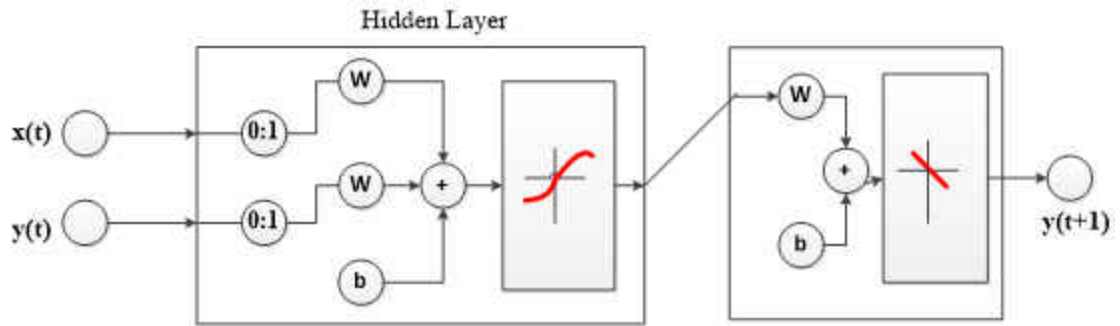
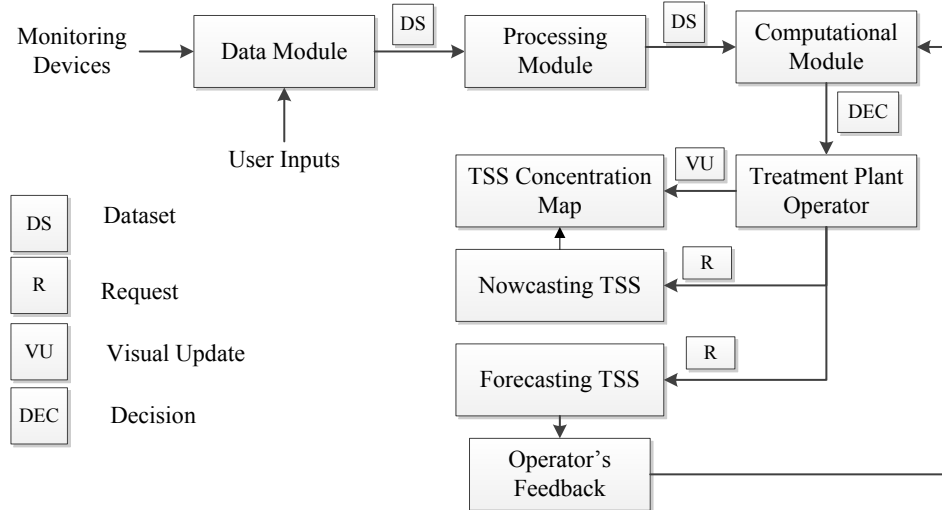


Figure 3-7. The diagram of NARXNET (Obtained from MATLAB R2013a)

As shown in Figure 3-7, the NARXNET model consists of two layers. The number of neurons in the hidden layer was set to 30. Also, the “trainlm” training function was selected to update bias and weight values based on the Levenberg-Marquardt optimization. The model was trained using 70% of inputs. Also, 20% and 10% of inputs were selected for validation and testing the model, respectively.

**Step (6) Early warning system:** The vast amount of information produced by the developed forecasting and nowcasting methods does not guarantee better decisions, since machines do not have the human ability of considering the implications of unexpected events and situations. On the other hand, human decision making is usually inconsistent and contains errors in judgment. Therefore, the early warning system in the current study is able to benefit from the results of the developed forecasting model to continuously generate forward-looking and risk-informed data to make a reliable decision. The framework of the early warning system is shown in Figure 3-8.



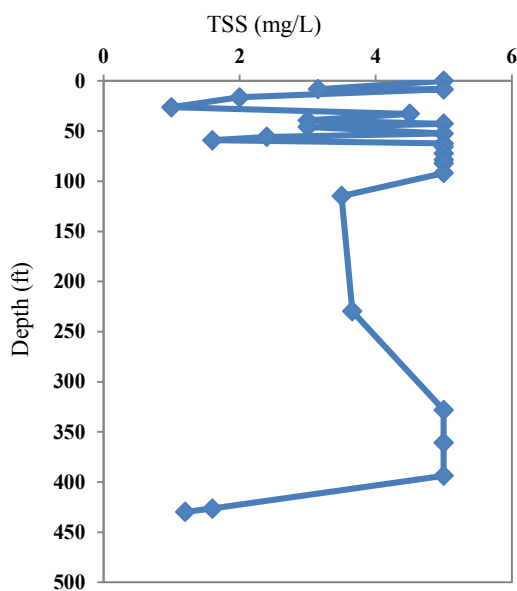
**Figure 3-8. The framework of the early warning system based on integrated nowcasting and forecasting analysis**

As shown in Figure 3-8, the early warning system consists of data module, processing module, and computational module. The data module is a database which contains current and historical ground-truth data and satellite imagery. The processing module is responsible for preparing and pre-processing satellite images to make them applicable for data fusion. The forecasting and nowcasting methods are embedded in these computational modules to form an integrated capacity. Based on the decision of the treatment plant operator, the results of nowcasting and forecasting of TSS values around the whole lake and nearby the water intake may be grouped together to realize the global trend for the requested date. Therefore, the water treatment plant operators will be benefit from such evaluation of water quality in the lake based on a suite of TSS concentration maps to minimize the water quality impact before the negative impact occurs, or otherwise achieves an actionable in-plant strategy to avoid cascading negative impacts.

### 3.5. Results and Discussion

#### 3.5.1. Nowcasting Results

Surface reflectance values detected by satellite sensors can only estimate the concentration of water quality parameters at water surface. Since the main objective of this study is predicting TSS concentration at the location of drinking water intake, it is important to assess the vertical profile of TSS to assure that estimated values of TSS at the water surface are representative of TSS values at the intake's depth. The lower Colorado River regional water quality database provided us with the TSS concentration at the location of the aforementioned stations in Table 3-1 at different depths from 0 to 130 m (0-430 ft). Based on this data, the vertical profile of TSS is prepared in Figure 3-9.



**Figure 3-9. Vertical profile of TSS at the location of drinking water intake in Lake Mead**

This figure shows that TSS from the surface to a depth of 130 m (430 ft) is either the same as the surface or less than the surface. Therefore, we can assume that the prediction of TSS

concentration at the water surface can be applied to the location of intake at 300 m (1000 feet) below the lake surface.

As mentioned in the previous section, the output of the nowcasting will be the net with the best performance which is saved to be applied in the forecasting procedure. Therefore, it would be beneficial to assess the accuracy of the nowcasting method. First, the reliability of the resulted fused images was assessed by calculating the correlation between the spectral reflectance values of each pixel in a fused image with its corresponding pixel in an actual Landsat image (Table 3-4). According to Table 3-4, the calculated R-Squared values vary from 0.64 in band 1 to 0.85 in band 4, which indicates a good fit. Therefore, the reliability of the resulting fused images is sufficient to support subsequent machine learning techniques for TSS estimation.

**Table 3-4. R\_squared value between the spectral reflectance value of the fused image and the actual Landsat image**

<b>Band</b>	<b>Band1</b>	<b>Band2</b>	<b>Band3</b>	<b>Band4</b>	<b>Band5</b>	<b>Band7</b>
<b>R<sup>2</sup></b>	0.64	0.76	0.80	0.85	0.83	0.81

After assuring the reliability of the resulting fused images to be applied in machine learning techniques, the accuracy of the trained ANN model in estimating TSS concentration values were assessed. For this purpose, all fused images were applied and the spectral reflectance values of all 6 bands were extracted at the location of station BB-7 to use as inputs into the trained ANN model. First, 60% of data was applied to train the ANN model through an iterative process to find the best performance by comparing the outputs (i.e. estimated TSS values) against the desired outputs (observed TSS values). The ANN model with the best performance was saved and the testing dataset (i.e. unseen data) was used to validate the model. This

procedure is helpful in overcoming the overtraining issues with the ANN model. The results showed  $R^2=0.98$ ,  $RMSE=0.489$  for training and  $R^2=0.99$ ,  $RMSE=1.066$  for testing the model. In addition, to assess the reliability of the saved ANN model, the surface reflectance values at the location of all stations corresponding to ground-truth dates were used as inputs to the ANN model to estimate TSS values for those dates. Table 3-5 shows the comparison between the estimated TSS values with the observed data. As shown in this table, the model was capable of estimating TSS concentrations with high precision.

**Table 3-5. Comparison between the observed TSS values and the estimated values from the trained ANN model**

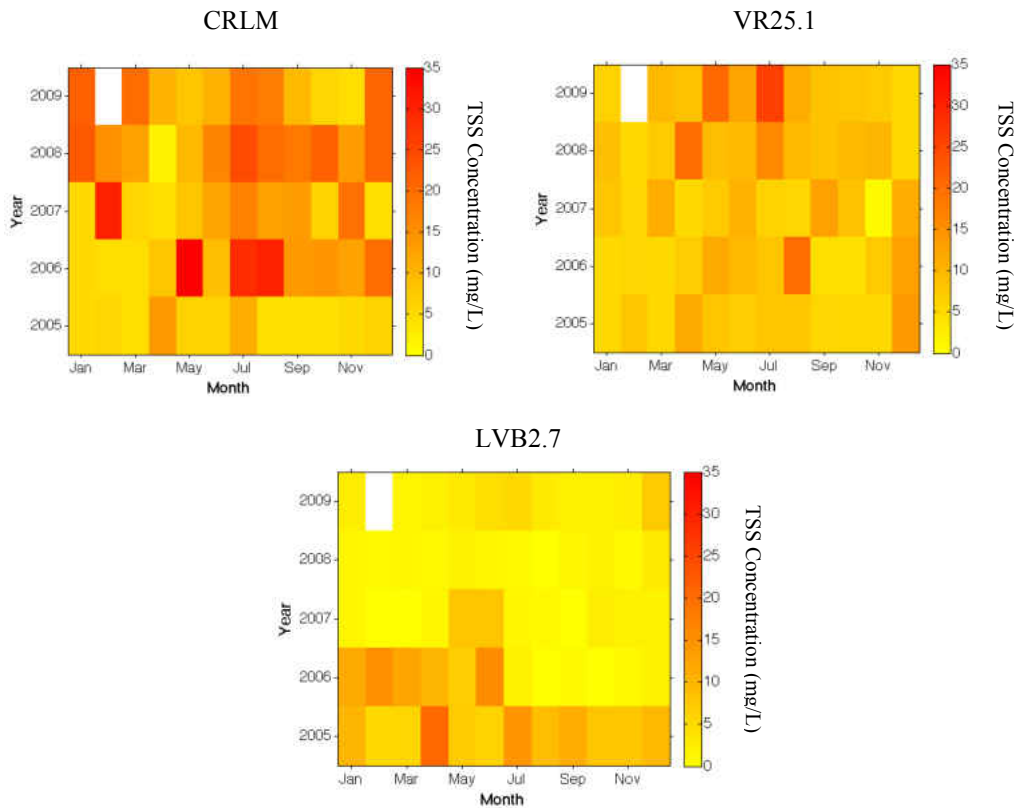
Sampling Station	Date	Observed	Estimated	Sampling Station	Date	Observed	Estimated
BB-3	4/5/2005	5	5.00	CR346.4	7/3/2007	5	5.18
	7/12/2005	5	4.98		10/3/2007	5	4.99
	4/6/2006	5	4.99	CR350.0SE0.55	4/5/2005	5	4.99
	1/23/2007	5	5.00		1/23/2007	5	5.00
	4/10/2007	5	4.99		4/3/2007	5	5.017
	7/3/2007	5	5.11		10/17/2007	5	5.00
	10/9/2007	5	4.99		10/17/2007	5	5.00
BB_7	10/3/2007	5	5.00	CR342.9	10/31/2007	5	5.00
	10/17/2007	5	5.00		11/7/2007	5	5.00
	10/31/2007	5	5.00	CR361.8	9/25/2007	5	5.00
	11/7/2007	5	5.00		10/9/2007	5	4.99
CR344.9W0.2	10/3/2007	5	5.00	VR25.1	10/23/2007	5	4.99
	10/17/2007	5	5.00		11/7/2007	5	5.00
	10/31/2007	5	5.00		9/11/2007	25	25.00
	11/7/2007	5	5.00		10/17/2007	13	13.00
CR346.4	7/11/2006	5	4.98	1/17/2008	9.2	9.203	
	1/23/2007	5	5.00	6/19/2008	9	6.55	
	4/10/2007	5	5.00	10/14/2008	7.4	10.93	

### 3.5.2. TSS Concentrations at the Location of Inflows into the Lake

To find the time of abrupt increase in TSS concentration in the lake, it is necessary to assess the monthly variation of TSS values at three different arms of the lake (i.e. Colorado River Arm, Overton Arm, and Boulder Basin). Sampling data within Lake Mead is available for sporadic time periods and in limited numbers to support the local assessment (Table 3-1). For instance, in the Colorado River Arm only three ground-truth values were recorded (Table 3-1). Therefore, the trained ANN model was applied to estimate TSS concentration for all days for which Landsat cloud-free images were available, and to make the temporal distribution of TSS at three different arms (Figure 3-10). This figure shows the temporal distribution of TSS within the lake on a monthly basis at the location of stations CRLM, VR25.1, and LVB2.7. These stations are representatives of water quality at the location of inflows from Colorado River, Lower Virgin River Watershed, and Las Vegas Wash, respectively. Figure 3-10 was calculated based on the averaged TSS concentrations for each month at the location of each selected station using all available cloud free Landsat images.

The U.S. EPA does not define any quantitative criteria for TSS. However, the water quality standards for Nevada are controlled by the Nevada Administrative Code (NAC). Based on the NAC 445A.194, the water quality standard for TSS for beneficial uses is limited to less than or equal to  $25 \text{ mg.L}^{-1}$  (Bureau of Reclamation, 2006). As shown in Figure 3-10, months in which the highest TSS concentration was detected are different for each station. Therefore, it is important to find the probable factors which may increase TSS concentration at the three different arms. As shown in Figure 3-10, TSS concentration values from the highest to lowest were detected at CRLM, VR25.1, and LVB2.7, respectively. In addition, the dates of detected hot spots (i.e. TSS concentration  $\geq 25 \text{ mg.L}^{-1}$ ) are shown in Table 3-6. As shown in this table,

there is not any evidence of high TSS concentration at any station in the years 2005 and 2008. In addition, during 2005-2009, there is no evidence of high TSS concentration (i.e more than 25 mg.L-1) at LVB2.7 located at Boulder Basin. Therefore, results show the probable impacts of the inflows from Colorado River and Lower Virgin River Watershed on increasing TSS concentration within the lake.



**Figure 3-10. Temporal distribution of TSS concentration at the location of inflows into the lake; white represents “no data” value as a result of inaccessibility of cloud-free Landsat images**

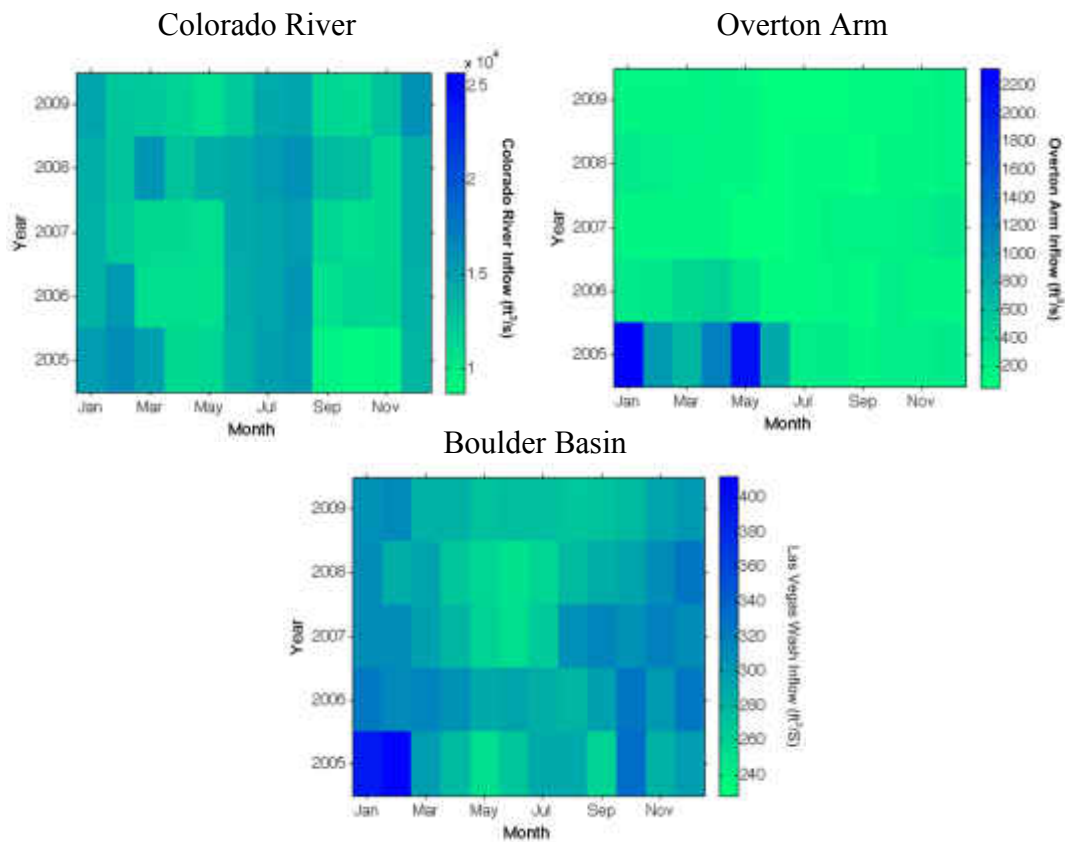


**Table 3-6. Detected dates of hot spots (TSS  $\geq$  25 mg.L<sup>-1</sup>) at the location of each station-**

Station	Location	Year				
		2005	2006	2007	2008	2009
CRLM	Colorado River Arm	-	May, Jun-Jul	Feb	-	-
VR25.1	Overton Arm	-	-	-	-	Jul
LVB2.7	Boulder Basin	-	-	-	-	-

Higher inflow rates may accelerate the rise in TSS concentration as a result of its impact on higher sediment transport and erosion rate. Therefore, it would be beneficial to assess the temporal variation of inflows into the lake. Figure 3-11 was calculated based on the monthly streamflow at the location of four USGS streamflow stations: 09419800, 09404200, 09419507, and 09415000. The first two stations are representatives of inflows into the Boulder Basin and the Colorado River, respectively. In addition, the summation of streamflow of the last two stations will demonstrate the inflow from Overton Arm.

As shown in Figure 3-11, the associated color legend for each arm is different, and the majority of inflow into Lake Mead is coming from Colorado River. The attempt to find a linkage between increases in TSS concentration and the potential factors (transported sediments from the Colorado River, forest fire events, and transported urban stormwater runoff and treated wastewater effluent from the Las Vegas Wash) motivated the further observations of average mean monthly TSS load, which show the rate of discharge of the sediment from each inflow into the lake (Figure 3-12). These values were calculated by taking the average of TSS loads from the selected days in each month.

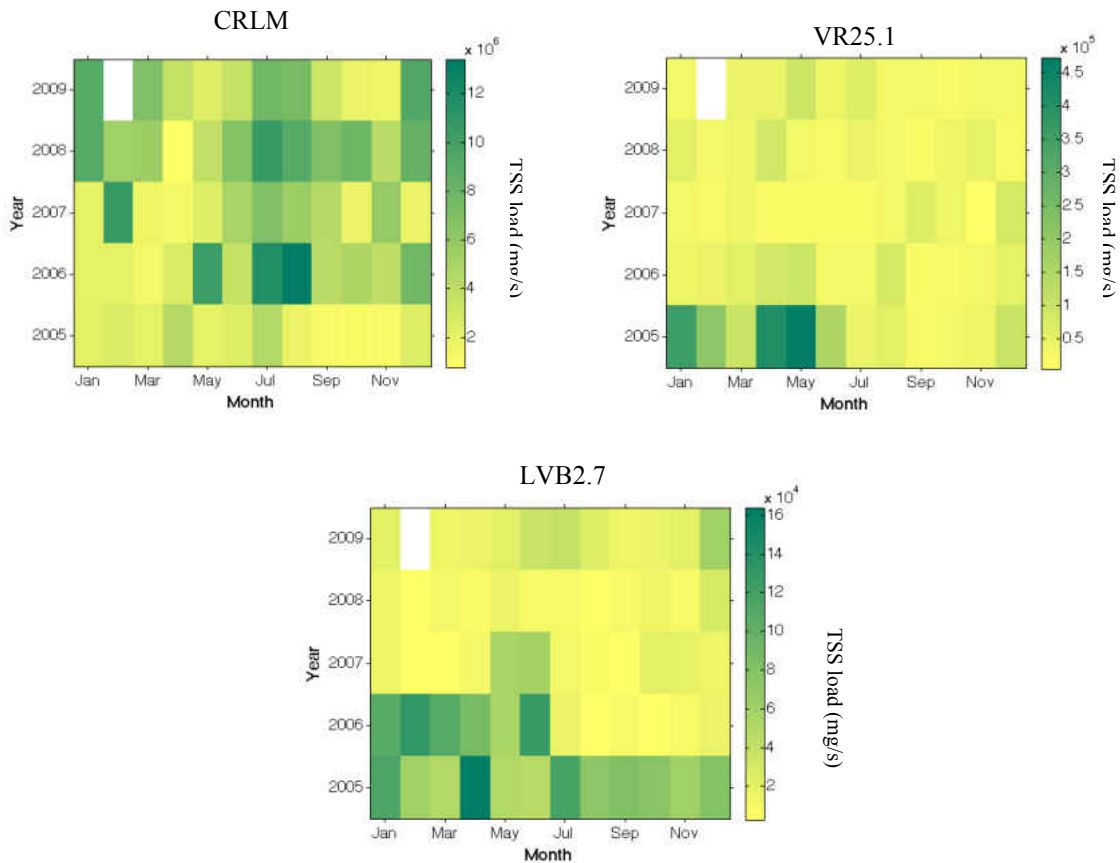


**Figure 3-11. Temporal distribution of streamflow at the location of USGS stations with associated color legend for each arm**

As shown in Figure 3-12, TSS loads in Colorado River Arm and Boulder Basin are more correlated to TSS concentrations than inflows. However, in the Overton Arm, the TSS load is highly dependent to the inflows than TSS concentrations.

Although consistency between high TSS concentration and high inflow rates were detected in some months in the Colorado River Arm (i.e July-Aug 2009), there are still some months in this part of the lake in which the highest TSS concentration occurred at the relatively low inflow rates (i.e. May 2006). In addition, the lower Virgin River Watershed and Las Vegas Wash does not reveal any evidence of a correlation between high inflow rates and increases in TSS concentration. In addition, increase in TSS concentration at the Overton Arm was not

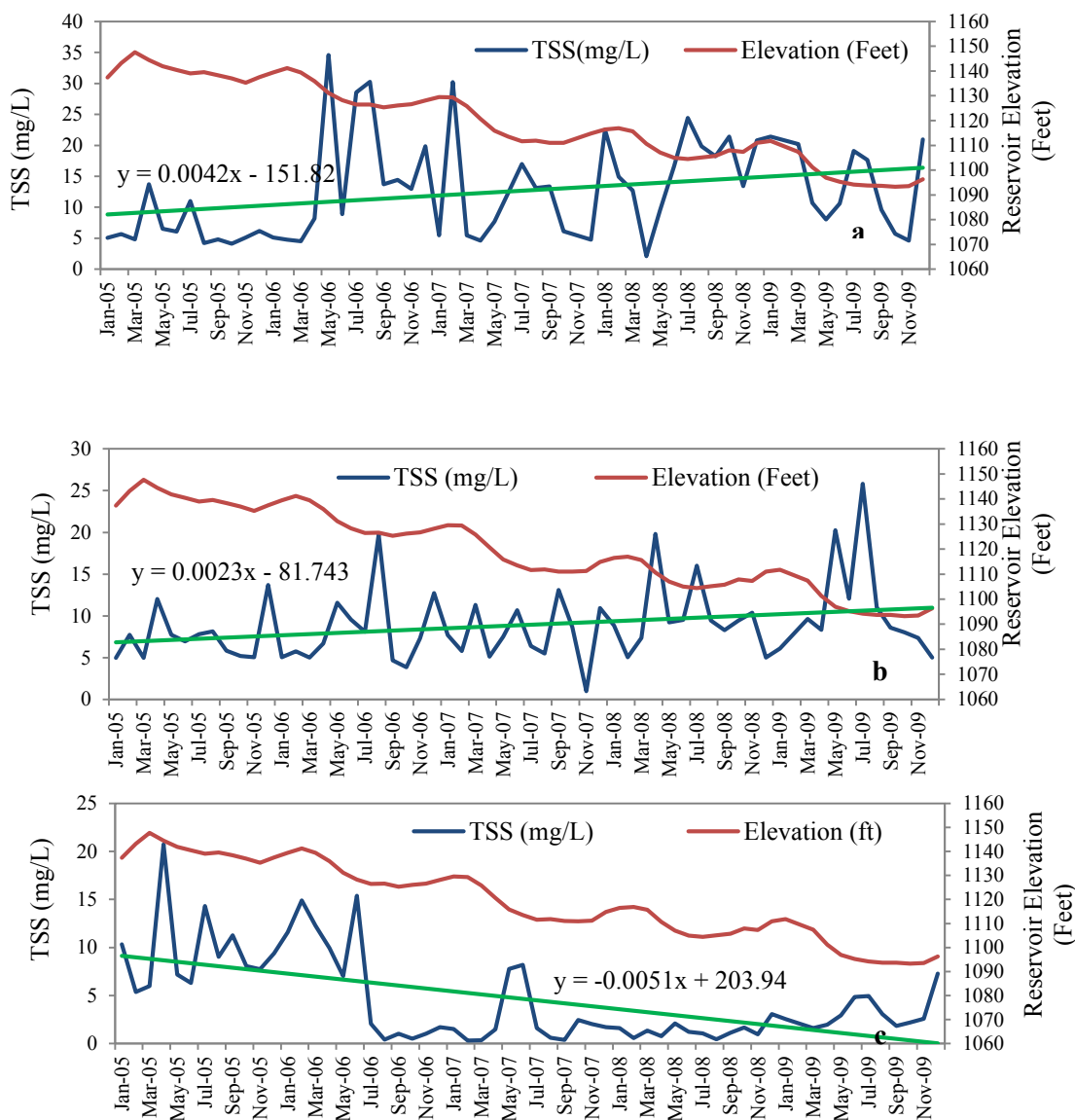
detected during the year 2005 in which the largest forest fire event occurred. Although there is not any simultaneous impact from forest fire events on increasing TSS concentration in Lake Mead, the negative impacts of forest fire events on soil erosion may have lag time to show up.



**Figure 3-12. Temporal distribution of the mean monthly TSS load (mg/s) at each arm; White color represents “no data” values as a result of inaccessibility of cloud-free Landsat images**

A decade of drought in the Southwest will also modify water quality in Lake Mead by direct effects on concentration of water quality parameters. Elevation of Lake Mead has been dramatically reduced by more than a decade of drought in the Southwest. Lower reservoir levels may concentrate all the constituents in the water. For this purpose, the comparison between TSS concentration and reservoir elevation is shown in Figure 3-13. TSS concentrations at the Colorado River Arm and Overton Arm (Figure 3-13 a, b) have increasing trend while reservoir

level is decreasing. On the other hand, TSS concentration at the Boulder Basin (Figure 3-13c) indicates a decreasing trend as a result of construction of erosion control structures, restoration of wetland areas, and removal of invasive vegetation upstream of the Las Vegas Wash inflow into the lake.



**Figure 3-13. Comparison between temporal variations of TSS concentration and Lake Mead at Hoover dam elevation at (A) CRLM (B) VR25.1 (C) LVB2.7. Green line shows the trend of TSS concentrations, and red line indicates the variation of elevation during (2005-2009).**

To assess the reliability of the IDFM algorithm for early warning system, this algorithm was applied to create synthetic fused images and daily TSS concentration maps for the months with the high TSS concentration in each arm within the time period of 2005-2009. As an example, this algorithm was applied for three different dates in each arm (i.e. Figure 3-14 to Figure 3-16). As shown in these figures, this algorithm can provide daily TSS concentrations for at least a week reflecting the high TSS concentration in Figure 3-10.

As shown in Figure 3-10, during June-Oct 2008, high TSS concentration was continuously detected at CRLM. Therefore, as an example, one of the weeks in September 2008 was selected to prepare daily TSS concentration maps (Figure 3-14). As shown in this figure, high TSS concentration (greater than  $30 \text{ mg}\cdot\text{L}^{-1}$ ) was detected at the Colorado River Arm in the first day of the selected week. Also, the TSS concentration values for the rest of the days of the same week were estimated in the range of  $10\text{-}20 \text{ mg}\cdot\text{L}^{-1}$  at the Colorado River Arm. In addition, TSS values at Overton Arm and Boulder Basin were estimated to be less than  $10 \text{ mg}\cdot\text{L}^{-1}$  which is in consistent with Figure 3-10.

Although the highest TSS concentration at Overton Arm was detected on July 9 of 2009, we were not able to demonstrate a continuous daily concentration maps for a week of the aforementioned date as a result of the existence of cloudy days. The Star-FM algorithms require cloud-free images in TSS concentration retrieval. Therefore, a week in May 2009 was selected. High TSS concentration (greater than  $30 \text{ mg}\cdot\text{L}^{-1}$ ) was detected at Overton Arm in two days of the selected week (Figure 3-15). High TSS concentration was also detected at the Colorado River Arm and Boulder Basin on 19 May 2009 (Figure 3-15), which is not consistent with Figure 3-10.

This inconsistency was caused by calculating Figure 3-10 based on only available cloud-free Landsat images.

As shown in Figure 3-10, the highest TSS concentration at the Boulder Basin was detected in April 2005. However, as a result of the existence of cloudy days in this month, the continuous daily concentration maps were created for a week of February 2006 (Figure 3-16). As shown in Figure 3-16, the first day of the selected week showed the highest TSS concentration in Boulder Basin. Therefore, these figures can provide water treatment plant operators with information about not only the specific days of detected hot spots, but also the locations of high TSS concentration.

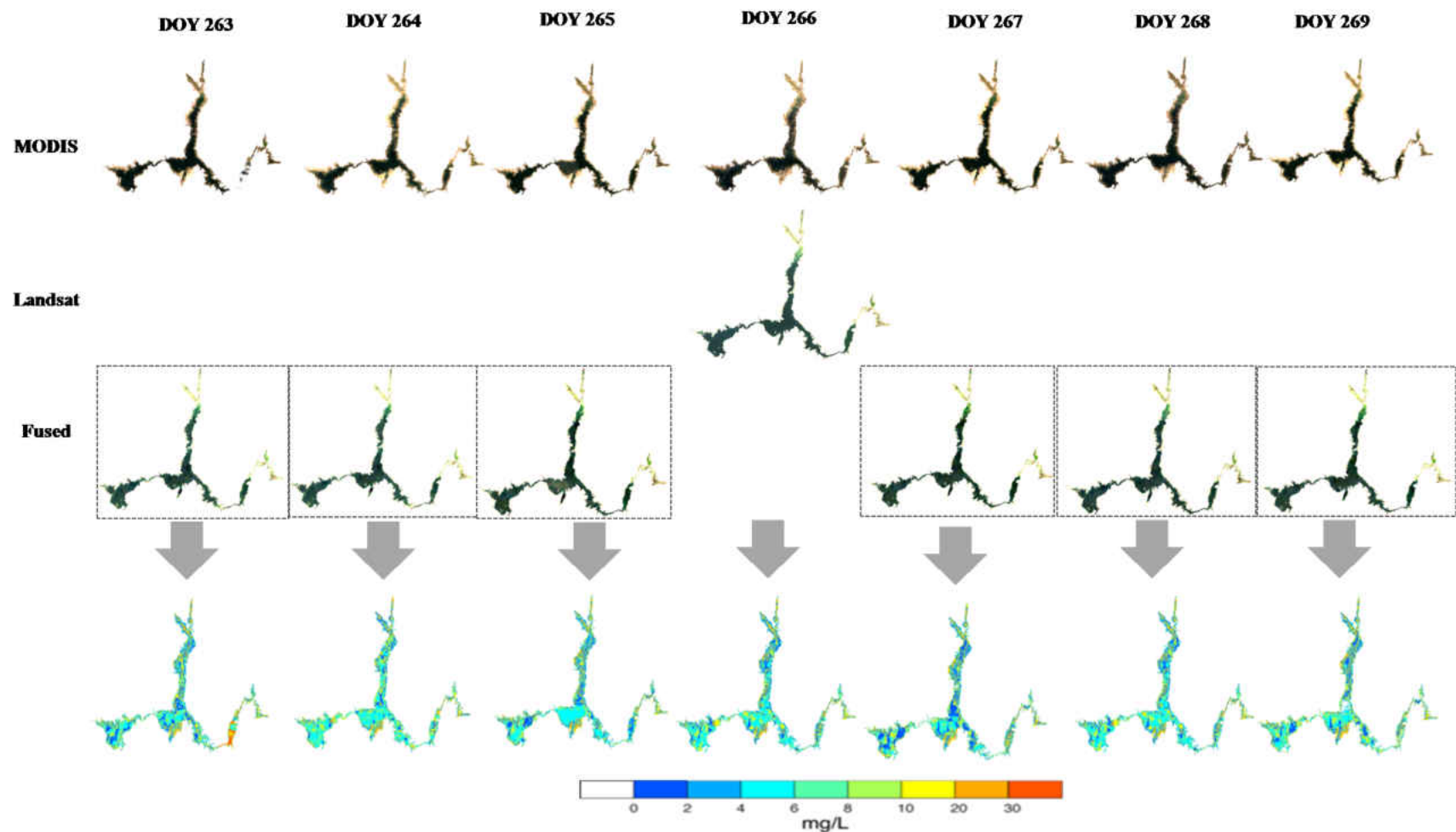


Figure 3-14. Estimating TSS concentration maps using IDFM Method from 19 Sep 2008 to 25 Sep 2008. The top row in this figure shows the available MODIS images and the second row indicates the available Landsat images in the Lake Mead for the cloud-free days of the selected time period. Third and the last rows show fused images and daily TSS concentration maps, respectively. Since using the atmospheric correction algorithm changes the surface reflectance values in MOD09 to -100 –16,000 [27], pixels with negative surface reflectance values were categorized as white in this figure.

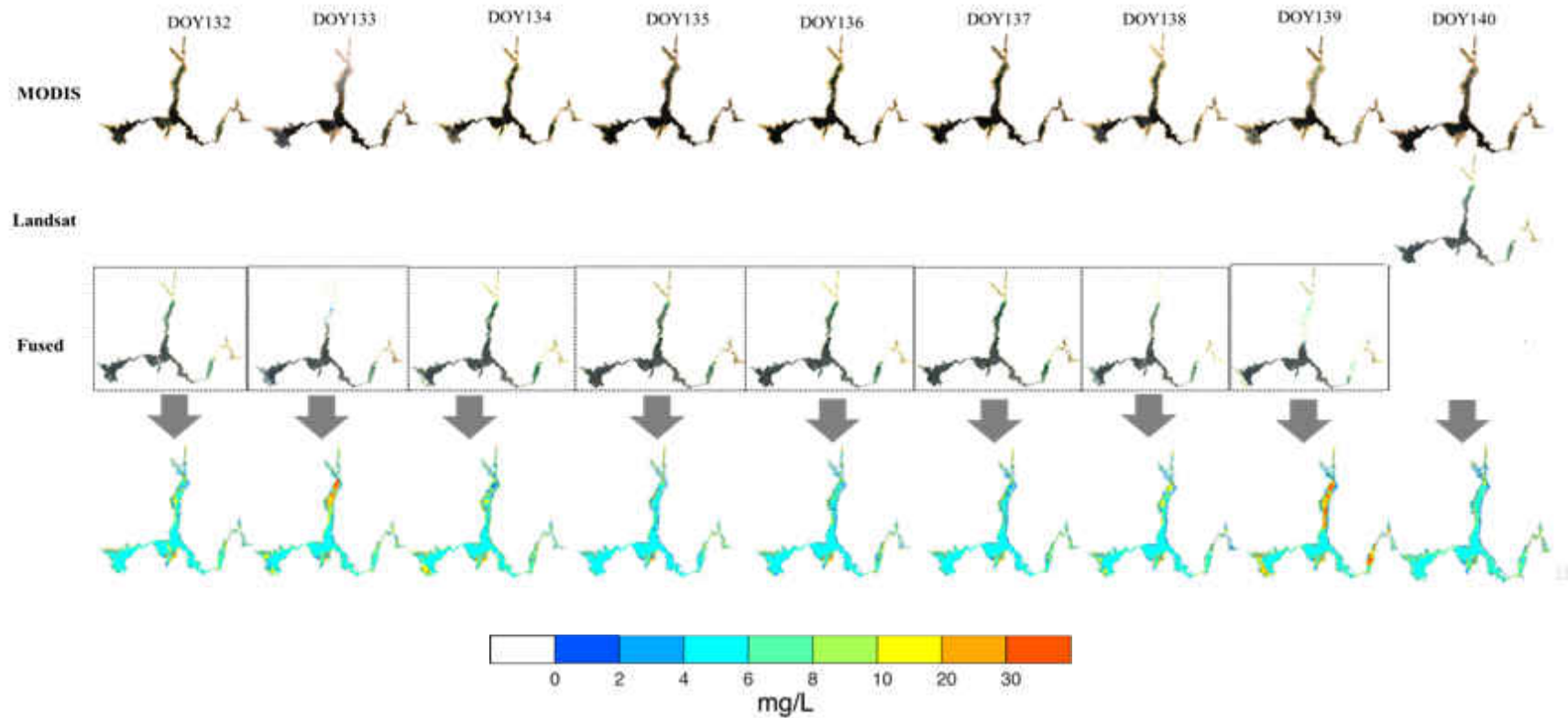


Figure 3-15. Estimating TSS concentration maps using IDFM Method from 12 May 2009 to 20 May 2009. The top row in this figure shows the available MODIS images and the second row indicates the available Landsat images in the Lake Mead for the cloud-free days of the selected time period. Third and the last rows show fused images and daily TSS concentration maps, respectively. Since using the atmospheric correction algorithm changes the surface reflectance values in MOD09 to -100 –16,000 [27], pixels with negative surface reflectance values were categorized as white in this figure



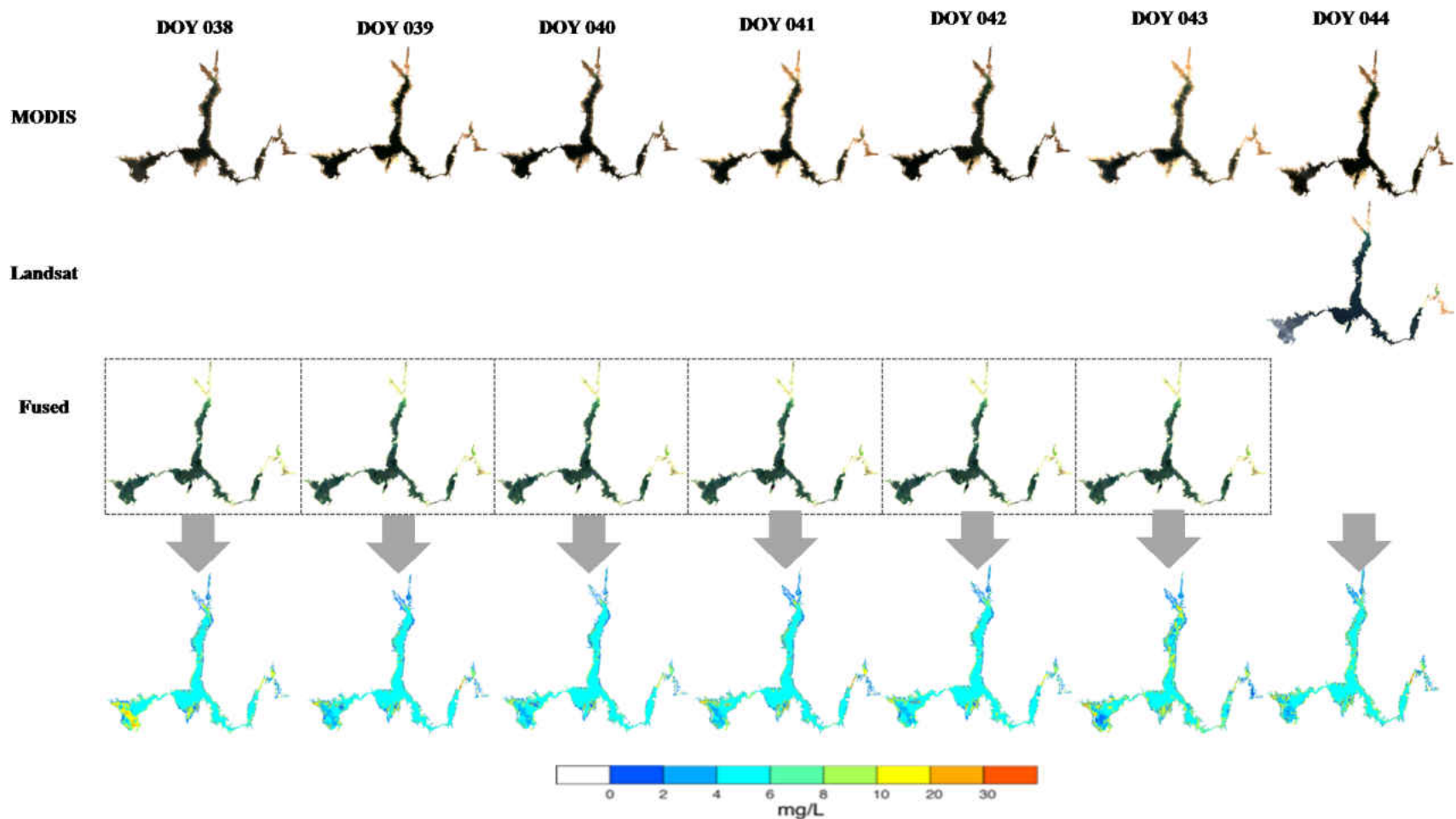


Figure 3-16. Estimating TSS concentration maps using IDFM Method from 7 February 2006 to 13 February 2006. The top row in this figure shows the available MODIS images and the second row indicates the available Landsat images in the Lake Mead for the cloud-free days of the selected time period. Third and the last rows show fused images and daily TSS concentration maps, respectively.

### 3.5.3. Forecasting Results

The reliability of the developed forecasting model to foresee the impacts of high TSS concentration at the drinking water intake is investigated for the time period in which the highest concentration of TSS detected in the Boulder Basin. As shown in Table 3-1, the concentration of TSS in Boulder Basin for all the available ground-truth data (2005-2007) at the stations close to the drinking water intake (BB-3, BB-7) was constant ( $5 \text{ mg}\cdot\text{L}^{-1}$ ), and high TSS concentration was not detected during this period. Since the ground-truth data at the location of these stations were mostly collected in the year 2007, the performance of the developed forecasting model was assessed for the months of 2007 in which the TSS concentration was detected in the lake at the location of BB-7. Therefore, as an example, a time period of 11-Sep-2007 to 30-Oct-2007 was selected to train and validate the model and to investigate the ability of the model in forecasting TSS concentration for a day ahead of time (i.e. 31-Oct-2007). To develop the forecasting model, all cloud-free Landsat and MODIS images for more than 30-days (from 11-Sep-2007 to 30-Oct-2007) before the forecasting date (31-Oct-2007) were downloaded and processed. Then, the STAR-FM algorithm was applied to produce fused images for those days during the aforementioned time period in which Landsat images were not available. The extracted reflectance values of all bands of the fused/Landsat images at the location of BB-7 were used as inputs into the trained ANN model to estimate the daily TSS concentration at the location of BB-7 for the selected time period. Then, the sensitivity analysis was performed by comparing multiple scenarios in the context of big data analytics in order to make the forecasting outputs easier for implementation in the water treatment plants based on varying availability of different data. Three different scenarios were defined in this study (Table 3-7). The first scenario includes

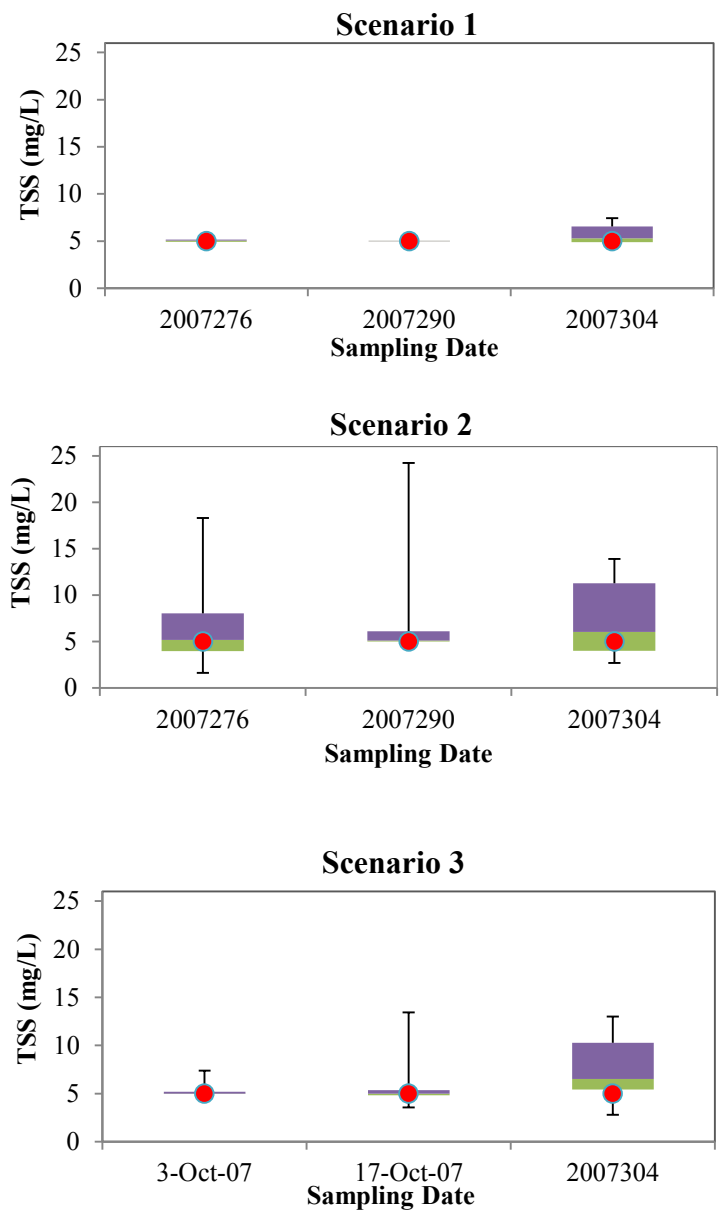
both remote sensing and water level data; Scenario 2 includes all meteorological parameters and remote sensing data; Scenario 3 contains remote sensing, water level data, and meteorological data. To reduce the uncertainty, multiple runs (10 runs) for cross-validation were performed for each scenario, and the box-plot of each scenario was presented in (Figure 3-17). During the selected time period, three ground-truth data for days 2007276 (3-Oct-2007), 2007290 (17-Oct-2007), and 2007304 (31-Oct-2007) were available at the location of BB-7. The first two dates were applied to test the performance of the developed forecasting model in training and testing periods, and the last date was applied to assess the precision of the model in forecasting TSS concentration for a day ahead of time.

**Table 3-7. Summary of database used in different scenarios**

<b>Database</b>	<b>Scenario 1</b>	<b>Scenario 2</b>	<b>Scenario 3</b>
Spectral reflectance Value	*	*	*
Meteorological Parameters <sup>1</sup>		*	*
Water Level in the Reservoir	*		*

<sup>1</sup>Average wind speed, average air temperature, and average humidity were considered as meteorological parameters.

In Figure 3-17, the vertical distance between the smallest value to the largest one shows the spread of all data. As shown in this figure, scenario 1 has the least spread of data. In addition, the middle half of the data set falls within the interquartile range shown with the width of the box. Therefore, for both scenario 1 and 2, the predicted value (date 2007304) is within the interquartile range. In addition to comparing scenarios using box plots, different statistical indices such as root mean squared error (RMSE), mean absolute error (MAE), percent bias (PBIAS), and R-Squared ( $R^2$ ) were used in this study.



**Figure 3-17. Comparing ground-truth data with NARXNET results in different scenarios. The first two dates belong to training and validation period and the last date is the prediction date.**

RMSE is the most suitable measurement tool amongst the four aforementioned error indices for calculating goodness of fitness. This error index is calculated based on the following equation (Willmott and Matsuura 2005):

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_{Pi} - Y_{Oi})^2}{N}} \quad (2)$$

where  $Y_{Pi}$  is the  $i^{\text{th}}$  predicted value,  $Y_{Oi}$  is the  $i^{\text{th}}$  observed value, and  $N$  is the total number of ground truth data.

MAE is the average of absolute errors calculated based on equation (3) (Willmott and Matsuura 2005). Values of RMSE and MAE equal to zero would indicate a perfect fit.

$$MAE = \frac{\sum_{i=1}^N |Y_{Pi} - Y_{Oi}|}{N} \quad (3)$$

PBIAS indicates the average tendency of predicted data to be smaller or larger than the observed value. Its optimal value is zero and its positive/negative value indicates model underestimation/overestimation bias (Gupta et al. 2004).

$$PBIAS = \left[ \frac{\sum_{i=1}^N (Y_{Oi} - Y_{Pi}) * 100}{\sum_{i=1}^N Y_{Oi}} \right] \quad (4)$$

R-squared value should be as close to one as possible which would indicate the best match between the model output and the target value (Legates and McCabe 1999).

$$R^2 = \left[ \frac{\sum_{i=1}^N (Y_{Oi} - \bar{Y}_{Oi})(Y_{Pi} - \bar{Y}_{Pi})}{\sqrt{\sum_{i=1}^N (Y_{Pi} - \bar{Y}_{Pi})^2 \sum_{i=1}^N (Y_{Oi} - \bar{Y}_{Oi})^2}} \right]^2 \quad (5)$$

The average of the calculated error indices over the multiple running of the model were reported in Table 3-8. As shown in this table, scenario 1 showed the highest R-squared value,

and the lowest estimation error. Adding additional inputs (i.e. meteorological parameters) decreased R-squared value and increased errors. Therefore, among the defined scenarios, scenario 1 shows the best result; this finding highlights the reservoir’s water level as a significant parameter in predicting TSS concentration in Lake Mead.

**Table 3-8. Summary of the results of error indices; Values of RMSE, PBIAS, and MAE equal to zero, and R-squared value equal to one show a perfect fit.**

<b>Scenario</b>	<b>R<sup>2</sup></b>	<b>RMSE</b>	<b>MAE</b>	<b>PBIAS (%)</b>
Scenario1	0.98	0.5	0.35	1.6
Scenario2	0.76	2.13	1.59	6
Scenario 3	0.94	1.16	0.77	6.3

### 3.6. Final Remarks

Taking advantage of the wide range of information provided by the aforementioned nowcasting and forecasting methods, an early warning system can be built including context aware pervasive sensor system, algorithms, and networks. As shown in Figure 3-8, treatment plant operators will be able to access the visual preview of TSS concentration throughout the lake for specific dates using the historical or current database. Besides visualizing spatial distribution, the early warning system will be capable of providing the plant operators with temporal variations of TSS concentration at the specific location (i.e., drinking water intake). The last but most important result of the early warning system is related to the forecasted TSS values at the drinking water intake or anywhere as needed to achieve an adaptation strategy for the water treatment plant or even the lake watershed management. It enables the operator of the water treatment plant to tackle unexpected events before they occur.

### 3.7. Conclusions

With the aid of remote sensing imagery, local sensor networks, and machine learning techniques, the developed forecasting and nowcasting models are able to capture the spatial and temporal distributions of TSS concentrations throughout a lake and estimate TSS concentration at a designated location one day ahead of time. This capability is demonstrated in Lake Mead serving as the major water source for the Las Vegas metropolitan region.

Comparing TSS concentration at the location of inflows into Lake Mead showed the importance of the Colorado River in the TSS brought into the lake. Although TSS concentration at the Colorado River and Overton Arm had an increasing trend, Las Vegas Wash indicates a decreasing trend as a result of the construction of erosion control structures, restoration of wetland areas, and removal of invasive vegetation upstream of the Las Vegas Wash inflow into the lake.

Assessing the relationship between the higher inflow rates with TSS concentration does not show any impact from forest fire events on accelerating the rise in TSS concentration. However, comparing the trend of TSS concentration with reservoir level showed the probable impacts of a decade of drought on increasing the TSS concentration in the lake.

As shown in the previous section, except for cloudy days, the applied nowcasting method is able to show the daily spatial distribution of TSS for at least a week in which the highest TSS concentration was detected at the location of the inflows into the lake. These concentration maps provide the reservoir management and monitoring system with TSS values for the dates in which sampling was not performed and for those locations which lack of sampling stations. This information can be used by the monitoring system to pin point hot spots within the lake which

are currently lacking sampling locations. Results showed the necessity of establishing more sampling locations at the Colorado River Arm, Overton Arm, and the southern section of the Overton Arm within Lake Mead.

Comparing the results of the forecasting model with ground truth data showed the reliability of the developed forecasting model. Evaluating the obtained results of the developed forecasting model to build up the early warning system enables the water treatment plant operator to achieve an actionable strategy either in plant or at the source water location due to unexpected events. However, cloud contamination is a big hindrance in exploitation of satellite images retrieved from the visible and infrared spectral ranges, and the developed forecasting model is not able to forecast TSS concentration for cloudy days. Therefore, the developed early warning system can be further improved in future studies by including the cloud removal algorithms developed by Chang et al. (2015).

### 3.8. References

- Adamowski J., Chan H.F., 2011, A wavelet neural network conjunction model for groundwater level forecasting, *Journal of Hydrology*, vol. 407, no. 1-4, pp. 28-40.
- Aladag, C.H., Egrioglu E., Kadilar C., 2009, Forecasting nonlinear time series with a hybrid methodology, *Appl. Math. Lett.*, vol. 22, no. 9, pp. 1467–1470, doi:10.1016/j.aml.2009.02.006
- Altunkaynak, A., Oezger, M., Cakmakci, M., 2005, Water consumption prediction of Istanbul City by using fuzzy logic approach, *Water Resources Management*, vol. 19, pp. 641-654.
- Arbain, S., Wibowo A., 2012, Neural Networks Based Nonlinear Time Series Regression for Water Level Forecasting of Dungun River, *J. Comput. Sci.*, vol. 8, no. 9, pp. 1506–1513.
- Arizona Department of Environmental Quality (ADEQ). Draft 2006 status of ambient surface water quality in Arizona- Arizona's integrated 305 (b) assessment and 303(d) listing report. Arizona Department of Environmental Quality, Phoenix, Arizona.



- Bureau of Reclamation, 2006, Final Environmental Impact statement for clean water coalition systems conveyance and operations programs Lake Mead national recreation area, Clark County, Nevada, [www.nps.gov/lake/leam/management/upload/scope\\_feis\\_chp3.pdf](http://www.nps.gov/lake/leam/management/upload/scope_feis_chp3.pdf).
- Cai S., Kang X., Zhang L., 2012, Time series prediction based on artificial neural network for estimation of forest biomass, *International Journal of Advancements in Computing Technology*, vol. 4, no. 12, pp. 318-326.
- Chang F.J., Hui S.C., Chen Y.C., 2002, Reservoir operation using grey fuzzy dynamic programming, *Hydrological Processes*, vol. 16, no. 12, pp. 2395-2408.
- Chang N.B., Daranpob A., Yang J., Jin K.R., 2009, A Comparative Data Mining Analysis for Information Retrieval of MODIS Images: Monitoring Lake Turbidity Changes at Lake Okeechobee, Florida, *Journal of Applied Remote Sensing*, vol. 3, no. 1, 033549.
- Chang N.B., Xuan Z., 2013, Monitoring Nutrient-Concentration in Tampa Bay with MODIS Images and Machine Learning Model, IEEE Conference, pp. 702-707, 978-1-4673-5200-0/13/\$31.00©2013IEEE.
- Chang N.B., Vannah B.W., Yang Y.J., Elovitz M., 2014a, Integrated Data Fusion and Mining Techniques for Monitoring Total Organic Carbon Concentrations in Lake, *International Journal of Remote Sensing*, vol. 35, no. 3, pp. 1064-1093.
- Chang N.B., Vannah B.W., Yang Y.J., Elovitz M., 2014b, Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in Lake Erie, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2426-2442. DOI: 10.1109/JSTARS.2014.2329913.
- Chang N.B., Bai K., Chen C.F., 2015. Smart information reconstruction via time-space-spectrum continuum for cloud removal in satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. vol. PP, no. 99, pp. 1-15.
- Chattopadhyay S., Chattopadhyay G., 2010, Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ARNN, *Comptes Rendus Geoscience*, vol. 342, no. 2, pp. 100-107.
- Chen L., Tan C.H., Kao S.J., Wang T.S., 2008, Improvement of Remote Sensing on Water Quality in a Subtropical Reservoir by Incorporating Grammatical Evolution with Parallel Genetic Algorithms into Satellite Imagery, *Water Research*, vol. 42, pp. 296-306.
- Chen L., 2003, A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data, *International Journal of Remote Sensing*, vol. 24, no. 11, pp. 2265-2275.
- Chen S., Han L., Chen X., Li D., Sun L., Li Y., 2015. Estimating wide range total suspended solids concentrations from MODIS 250-m imagery: an improved method. *ISPRS Journal of Photogrammetry and Remote Sensing*. vol. 99, pp. 58-69.

- Cheng C.T., Ou C.P., Chau K.W., 2002, Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration, *Journal of Hydrology*, vol. 268, no. 1-4, pp. 72-86.
- The Clean Water Coalition Systems Conveyance and Operations Program (SCOP), Final EIS, October 2006, [http://www.nps.gov/lake/learn/management/upload/scope\\_feis\\_chp3.pdf](http://www.nps.gov/lake/learn/management/upload/scope_feis_chp3.pdf).
- Diaconescu, E., 2008, The use of NARX neural networks to predict chaotic time series, *WSEAS Transactions on computer research*, vol. 3, no. 3, pp. 182-192, ISSN: 1991-8755
- Díaz-Robles L.A., Ortega J.C., Fu J.S., Reed G.D., Chow J.C., Watson J.G., Moncada-Herrera J.A. 2008, A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment*, vol. 42, no. 35, pp. 8331-8340.
- Edward D.A., 1991, Colorado river Ecology and Dam Management: Proceeding of a Symposium, May 24-25 1990, Santafe, New Mexico, Committee to review the Glen Canyon Environmental Studies, Water Science and Technology Board, National Research Council, ISBN:0-309-59772-2,288 p.
- Egrioglu, E., Aladag C.H., Yolcu U., Uslu V.R., Basaran M.A., 2009, A new approach based on artificial neural networks for high order multivariate fuzzy time series, *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10589–10594, doi:10.1016/j.eswa.2009.02.057.
- Faruk D.Ö., 2010, A hybrid neural network and ARIMA model for water quality time series prediction, *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 586-594.
- Ferrari. R.L., 2008, 2001 Lake Mead Sedimentation Survey. U.S. department of the interior bureau of reclamation technical service center water resources service division. sedimentation and river hydraulic group, Denver, Colorado.
- Fettweis M., Nechad B., Eynde D.V., 2007. An estimation of the suspended particular matter (SPM) transport in the southern North Sea using SeaWiFS images, in situ measurement and numerical method results. *Cont. Shelf Res.* vol. 27, pp. 1568-1583.
- Gao F., Masek J., Schwaller M., Hall F., 2006, On the blending of Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp.2207-2218.
- Grimaldi S., Serinaldi F., Tallorini C., 2005, Multivariate linear parametric models applied to daily rainfall time series, *Advances in Geosciences*, vol. 2, pp. 87-92.
- Islam N. N., Sivakumar B., 2002, Characterization and prediction of runoff dynamics: a nonlinear dynamical view, *Advances in Water Resources*, vol. 25, pp. 179-190.

- Jayawardena A.W., Lai F., 1994, Analysis and prediction chaos in rainfall and streamflow time series, *Journal of Hydrology*, pp. 153, 23-52.
- Katlane R., Dupouy C., Zargouni F., 2012, Chlorophyll and Turbidity Concentration as an Index of Water Quality of the Gulf of Gabes from MODIS in 2009, *Revue Teledetection*, vol. 11, no. 1, pp. 263-271.
- Keiner L.E., Yan X.H., 1998. A Neural Network Model for Estimating Sea Surface Chlorophyll and Sediments from Thematic Mapper Imagery, *Remote Sensing of Environment*, vol. 66, pp. 63-165.
- Khorram S., 1985. Development of water quality models applicable throughout the entire San Francisco Bay and Delta, *Photogrammetric Engineering and Remote Sensing*. vol. 51, pp. 53-62.
- Kim, T., Valdés J., 2003, Nonlinear Model for Drought Forecasting Based on a Conjunction of Wavelet Transforms and Neural Networks, *Journal of Hydrologic Engineering*, vol. 8, no. 6, pp. 319–328.
- Koponen S., Pulliainen J., Kallio K., Hallikaine M., 2002, Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data, *Remote Sensing of Environment*, vol. 79, no. 1, pp. 51-59.
- Kuligowski R.J., Barros A.P., 1998, Experiments in short-term precipitation forecasting using artificial neural networks, *Monthly Weather Review*, vol. 126, pp. 470-482.
- Kundzewicz Z.W., Mata L.J., Arnell N.W., Döll P., Kabat P., Jiménez B., Miller K.A., Oki T., Sen Z., Shiklomanov I.A., 2007. Freshwater resources and their management. Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 173-210.
- Lavery P., Pattiaratchi C., Wyllie A., Hick P., 1993, Water quality monitoring in estuarine water using the Landsat thematic mapper. *Remote Sensing of Environment*, vol. 46, no. 3, pp. 268-280.
- Lee Y.S., Tong L.L., 2011, Forecasting time series using a methodology based on autoregressive integrated moving averaged and Genetic programming, *Knowledge-Based Systems*, vol. 24, no. 1, pp. 66-72.
- Liong, S., Chan W., ShreeRam J., 1995, Peak-flow forecasting with genetic algorithm and SWMM, *Journal of Hydraulic Engineering*, vol. 121, no. 8, pp. 613-617.
- López, C., Álvarez A., Hernández-García E., 2000, Forecasting confined spatiotemporal chaos with genetic algorithms, *Phys. Rev. Lett.*, vol. 85, pp. 2300–2303, doi:10.1103/PhysRevLett.85.2300.

- Mahabir C., Hicks F.E., Robinson Fayak A., 2003, Application of fuzzy logic to forecast seasonal runoff, *Hydrological Processes*, vol. 17, no. 18, pp. 3749-3762
- Mahfoud, S., Mani G., 1996, Financial forecasting using genetic algorithms. *Appl. Artif. Intell.* Vol. 10, pp. 543–566, doi:10.1080/088395196118425, 1996.
- Maritorena S., Siegel D., Peterson A., 2002, Optimization of a Semianalytical Ocean Color Model for Global-Scale Applications. *Appl. Optics.*, vol. 41, pp. 2705-2714.
- Matsuoka A., Hooker S.B., Bricaud A., Gentili B., Babin M. 2012, Estimating Absorption Coefficients of Colored Dissolved Organic Matter (CDOM) using a Semi-analytical algorithm for Southern Beaufort Sea (Canadian Arctic) Waters: Application to Deriving Concentrations of Dissolved Organic Carbon from Space, *Biogeosciences Discuss*, vol. 9, pp. 13743-13771.
- Megal S., McKay D.L., Peterson B., Gillen S., 2009, Lower Virgin River Watershed: Rapid Watershed Assessment Report, USDA Natural Resource Conservation Service and University of Arizona. 58 p.
- Menezes, J. M. P., Barreto G.A., 2006, On recurrent neural networks for auto-similar traffic prediction: A performance evaluation, in 2006 International Telecommunications Symposium, pp. 534–539, E-ISBN:978-85-89748-04-9.
- Mertes L., Smith M., Adams J., 1993. Estimating suspended sediment concentrations in surface waters of the Amazn River wetlands from Landsat images. *Remote Sens. Environ.* vol. 43, pp. 281-301.
- Mindell D.A., 2000, Cybernetics- Knowledge domains in Engineering systems: <http://21stcenturywiener.org/wp-content/uploads/2013/11/Cybernetics-by-D.A.-Mindell.pdf>.
- MODIS Reprojection Tool Web Interface(MRTWeb), 2010. <https://mrtweb.cr.usgs.gov/>, Page Last Modified: April 15, 2010
- Montanher O.C., Novo E. M.L.M., Barbosa C.C.F., Renno C.D., Silva T.S.F., 2014. Empirical models for estimating the suspended sediment concentration in Amazonian white water rivers using Landsat 5/TM. *International Journal of Applied Earth Observation and Geoinformation.* vol. 29, pp. 67-77.
- Myint S.W., Walker N.D., 2002. Quantification of surface suspended sediments along a river dominated coast with NOAA and SeaWiFS measurements: Louisiana, USA. *Int. J. Remote Sensing.* vol. 23, no. 16, pp. 3229-3249.
- National Park Service, Sedimentation Lake Mead, <http://www.nps.gov/lake/naturescience/sedimentation-lake-mead.htm>, Last Updated: 06/29/2014.
- National Interagency Coordination Center, Statistics and Summary, 2005, [http://www.predictiveservices.nifc.gov/intelligence/2005\\_statsumm/intro\\_summary.pdf](http://www.predictiveservices.nifc.gov/intelligence/2005_statsumm/intro_summary.pdf).

- Nayak P.C., Sudheer K.P., Rangan D.M., Ramasastri K.S., 2004, A neuro-fuzzy computing technique for modeling hydrological time series, *Journal of Hydrology*, vol. 291, no. 1-2, pp. 52-66.
- Oliveira K.A.D., Vannucci A., Silva D. E.C., 2000, Using artificial neural networks to forecast chaotic time series. *Physica A: Statistical Mechanics and its Applications*, vol. 284, no. 1-4, pp. 393-404.
- Panda S.S., V. Garg, Chaubey I., 2004, Artificial Neural Networks Application in Lake Water Quality Estimation using Satellite Imagery, *Journal of Environmental Informatics*, vol. 4, no. 2, pp. 65-74.
- Petus C., Chust G., Gohin F., Doxaran D., Froidefond J.M., Sagarminaga Y., 2010. Estimating turbidity and total suspended matter in the Adour River plume (South bay of Biscay) using MODIS 250-m imagery. *Cont. Shelf. Res.* vol. 30, pp. 379-392.
- Principe J.C., Rathie A., Kuo J.M., 1992, Prediction of chaotic time series with neural networks and the issue of dynamic modeling, *International Journal of Bifurcation and Chaos*, vol. 20, pp. 989-996, DOI: 10.1142/S0218127492000598.
- Riha S., and H. Krawczyk, 2011, Development of a Remote Sensing Algorithm for Cyanobacterial Phycocyanin Pigment in the Baltic Sea using Neural Network Approach, Proc. SPIE 8175, Remote Sensing of the Ocean, Sea Ice, Coastal Waters, and Large Water Regions, 817504 , DOI:10.1117/12.898081.
- Silverman D., Dracup J.A., 2000, Artificial neural networks and long range precipitation prediction in California, *Journal of Applied Methodology*, vol. 39, pp. 57-66.
- Smyth T.J., Moore G.F., Hirata T., Aiken J., 2006, Semi-analytical Model for the Derivation of Ocean Color Inherent Optical Properties: Description, Implementation, and Performance Assessment, *Applied Optics*, vol. 45, no. 31, pp. 8116-8131.
- Soltani S., Modarres R. , Eslamian S.S., 2007, The use of time series modeling for the determination of rainfall climates of Iran, *Int. J. Climatol.*, vol. 27, no. 6, pp. 819-829.
- Son S.H., Kim Y.H., Kwon J.I., Kim H.C., Park K.S., 2014. Characterization of spatial and temporal variation of suspended sediments in the Yellow and East China Ses using satellite ocean color data, *GIScience and Remote Sensing*, vol. 51, no. 2, pp. 212-226.
- Szpiro, G. G., 1997, Forecasting chaotic time series with genetic algorithms, *Phys. Rev. E*, vol. 55, pp. 2557–2568, doi:10.1103/PhysRevE.55.2557, 1997.
- Tassan S., 1994. Local algorithms using SeaWiFS data for the retrieval of ohytoplankton, pigments, suspended sediment, and yellow substancein coastal waters. *Appl. Opt.* vol. 33, no. 12,pp. 2369-2378.
- Trivedi H.V., Singh J.K., 2005, Application of grey system theory in the development of a runoff prediction model, *Biosystems Engineering*, vol. 92, no. 4, pp. 521-526.

- USDA Forest Service, <http://activefiremaps.fs.fed.us/>, Last Updated: 07/08/2014.
- USGS, Earth Explorer, <http://earthexplorer.usgs.gov/>, Page Last Modified: 11/03/2014.
- Vermote E., Kotchenova S., Ray J., 2011, MODIS surface reflectance user's guide, MODIS Land Surface Reflectance Science Computing Facility. 1.3, 40p.
- Villar R.E., Martinez J.M., Le Texier M., Guyot J.L., Fraizy P., Meneses P.R., et al., 2013. A study of sediment transport in the Madeira River, Brazil, using MODIS remote sensing images. *J.S. Am. Earth Sci.* vol. 44, pp. 45-54.
- Wang J.P., Cheng S.T., Jia H.F., 2008, Application of artificial neural network technology in water color remote sensing inversion of inland waterbody using TM data, In the International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences. Istanbul: ISPRS Congress, Proceedings of Commission IV.
- Yu P.S., Chen C.J., Chen S.J., Lin S.C., 2001, Application of grey model toward runoff forecasting, *Journal of American water resources association*, vol. 37, no. 1, pp. 151-166.
- Zhang Y., Pulliainen J., Koponen S., Hallikainen M., 2002, Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data, *Remote Sensing of Environment*, vol. 81, no. 2-3, pp. 327-336.
- Zhang Y., Pulliainen J.T., Koponen S.S., Hallikainen M.T., 2003, Water quality retrievals from combined Landsat TM data and ERS-2 SAR data in the Gulf of Finland, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 622-629.
- Zhang, G., 2003, Time series forecasting using a hybrid ARIMA and neural network model, *Neurocomputing*, vol. 50, pp. 159–175, doi:10.1016/S0925-2312(01)00702-0.
- Zhang M., Tang J., Song Q., Ding J., 2010. Retrieval of total suspended matter concentration in the Yellow and East China Seas from MODIS Imagery. *Remote Sensing of Environment*. vol. 114, pp. 392-403, doi:10.1016/j.rse.2009.09.016.

# **CHAPTER 4: MULTI-SENSOR FUSION FOR MONITORING TOTAL ORGANIC CARBON CONCENTRATIONS IN LAKE MEAD TOWARD WATER QUALITY FORECASTING**

## **4.1. Introduction**

Only a small fraction of the Earth's total surface area is made up of reservoirs and lakes, yet they are considered important sources of water for the maintenance of ecosystems and human use. As the condition of reservoirs deteriorates throughout the world, it necessitates the requirement of forecasting water quality status to characterize the risk of water use in reservoirs. This is especially evident in Lake Mead, one of the most important water resources in the Western United States, as provides drinking water for 25 million people, irrigation water for 10,000 square kilometers of land, and electricity for major cities. It is also considered a recreational area used by 8 million people and an important habitat for federally endangered fish and wildlife species. Consequently, monitoring and forecasting water quality in Lake Mead can be of great use for helping water quality management in the municipal water supply system and assuring the sustainability of this important natural resource.

Reservoirs, including Lake Mead, cause increasing in hydraulic residence time which leads to more opportunities for elevated total organic carbon (TOC) concentrations and the possible formation of disinfection by products (DBPs) in the water treatment processes (AWWA 2002). One of the most important types of DBPs detected in drinking water distribution networks is trihalomethanes (THMs) which have dangerous side effects on the human body when ingested through drinking water, inhaled while bathing, and from both inhalation and osmosis through the skin while swimming in waters that have been disinfected using chlorine (Nikolaou et al. 1999;

Mohamadshafiee and Taghavi, 2012). DBPs are formed when Natural Organic Materials (NOMs) in surface water come into contact with chlorine which is often used to disinfect the drinking water supply. Since TOC includes the majority of organic matter, and is easier to measure than NOM, it is applied as the indicator for controlling the formation of DBPs (USEPA 1998).

The goal of this chapter is to extend the current IDFM method developed by Chang et al. (2014a); (2014b) and elevate it to be a cost-effective, forward-looking, and risk-informed forecasting scheme for water quality monitoring in Lake Mead with respect to TOC concentration at the location nearby the drinking water intake. Since the IDFM technique is a case-based engineering approach, the current site-specific study was exclusively developed for Lake Mead via a two-stage analysis. The objective in the first stage is to train the machine learning model to interpret fused satellite imageries based on its ground truth data in the Lake Mead environment to verify the effective nowcasting via IDFM, and then the second stage advances IDFM to forecast TOC concentration onward with updated water quality parameters and satellite imagery on a daily basis. The following scientific questions are explored in this study: 1) How the water levels drop affect the TOC concentration in the lake? 2) Is there any linkages between forest fires and high TOC concentrations? (3) Is there any difference between the results of applying different machine learning techniques (i.e. artificial neural network and genetic programming)? (3) What is the reliability of the developed forecasting method to predict TOC a day ahead of time in Lake Mead?



## 4.2. Problem Identification

Majority of the inflow into Lake Mead is from the Colorado River, a small percentage of the inflow that is also deemed important in terms of water quality management is provided by the Muddy River and the Virgin River on its northern side and from the Las Vegas Wash on its western side. Las Vegas Wash transports stormwater, urban runoff, and treated wastewater effluent into Boulder Basin in Lake Mead and causes water quality impairment to the lower Las Vegas Wash. In addition, forest fires which occur on the northern side of the lake increase stormwater runoff and sediment load. The completion of the third drinking water intake at Saddle Island for the city of Las Vegas will make these potential factors even more sensitive (YSI 2006).

In 2013, the maximum level of total THMs was reported greater than the official limit (i.e., 4 ppb) of United States Environmental Protection Agency (US EPA) due to the presence of high TOC concentration levels in the lake (Las Vegas Valley Water District 2013). Although this violation is allowable as long as either a running annual average or a locational running annual average does not exceed the US EPA permit, it highlights the importance of monitoring TOC in the lake as its concentration can be used as an indicator of DBP production. The US EPA advice note on DBPs in drinking water states that in the case of TOC concentrations greater than 4 mg·L<sup>-1</sup>, THMs will likely be high as well and remedial work will be necessary. In addition, in case of TOC concentrations between 2 and 4 mg·L<sup>-1</sup>, more assessment is required to check if it results in unexpected increases in THMs levels. Finally, it is suggested to keep the TOC concentration in treated water at a level less than 2 mg·L<sup>-1</sup> prior to disinfection. To address the dynamic conditions of the target water body, this chapter focuses on both nowcasting and

forecasting of TOC concentrations in Lake Mead using both remote sensing and machine learning techniques.

### 4.3. Data Collection and Analysis

#### 4.3.1. In-Situ Data

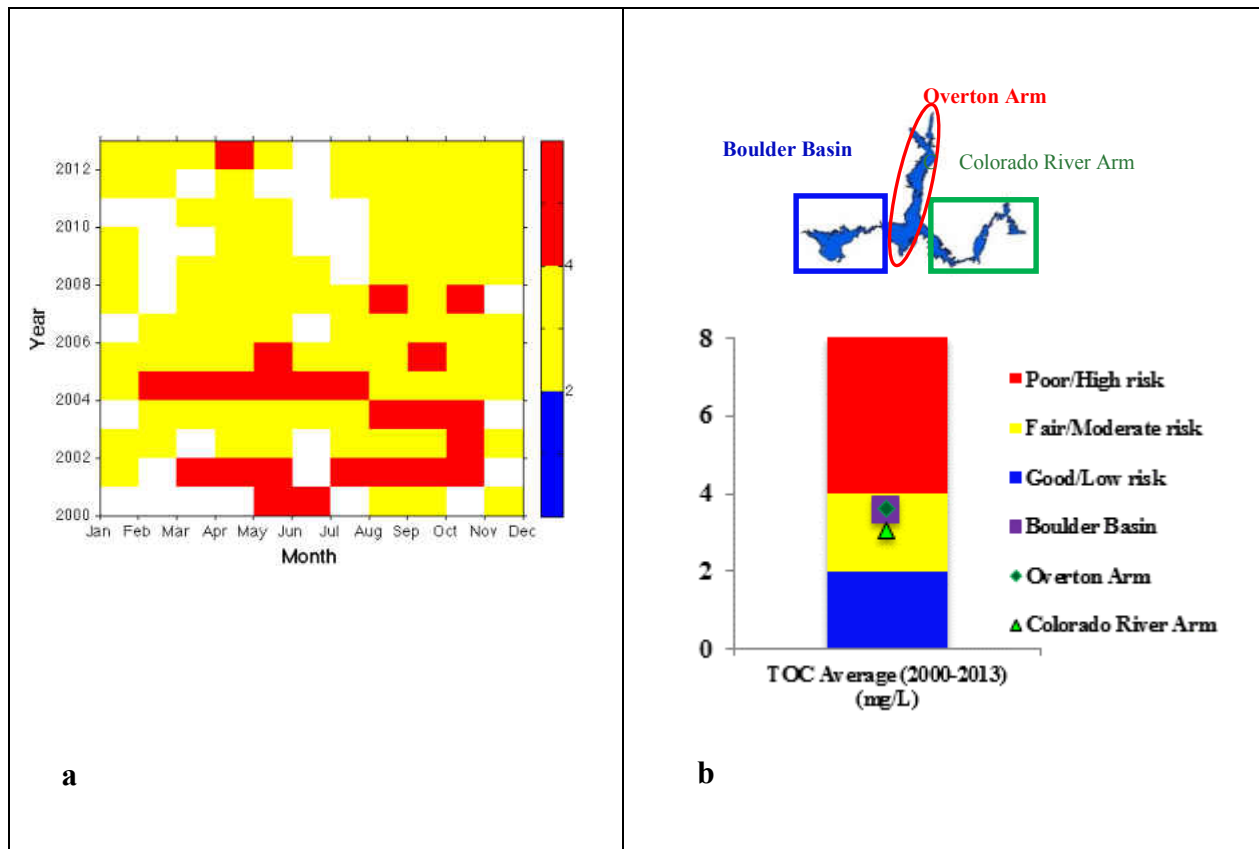
For the purpose of assuring the availability and quality of water in Lake Mead, it has become one of the most intensely sampled water bodies in the United States (Tietjen et al. 2012). All the data sets used in this study were obtained from the lower Colorado River regional water quality database which has more than 2 million records of about 1,000 different water quality parameters provided by different sampling agencies. The titles of these performing agencies that have contributed to the data collection of TOC concentration are shown in Table 4-1. In addition, the number of stations, number of samples, and the maximum, minimum, and mean TOC concentration are listed in this table as well.

**Table 4-1. TOC Ground-truth data within the Lake utilized by this study**

No. of Samples	No. of Stations	Time Period	Sampling Agency	TOC Concentration (mg·L <sup>-1</sup> )		
				Min	Max	Mean
1252	26	2000-2013	1. City of Las Vegas 2. Southern Nevada Water Authority(SNWA) 3. U.S. Bureau of Reclamation (Denver) 4. Southern Nevada Water System	1.6	31.9	3.59

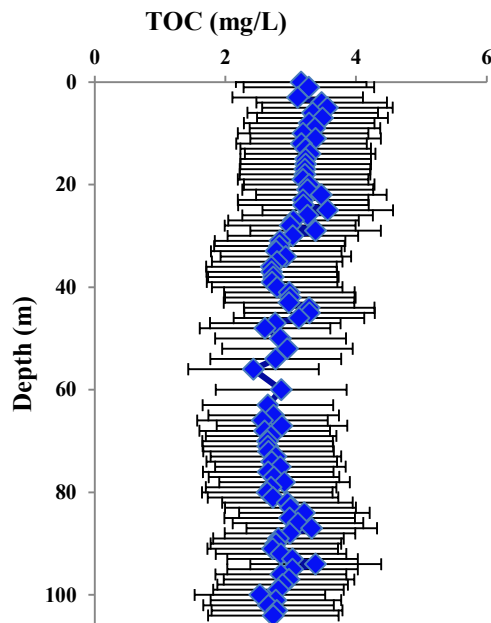
The temporal and spatial changes for TOC in the Lake are shown in Figure 4-1 a and Figure 4-1b, respectively. It should be mentioned that Figure 4-1 was calculated based on the average TOC concentration within the lake. TOC values were compared to the thresholds of the

aforementioned EPA criteria to portrait a risk-based assessment. In Figure 4-1a, the highest concentration of TOC ( $>4 \text{ mg}\cdot\text{L}^{-1}$ ) occurred during the 2000 to 2006 period, and during some isolated months in 2007 and 2012. In addition, the spatial variation of TOC shows that the concentration in Overturn Arm, Colorado River Arm, and Boulder Basin was in the range of 2-4  $\text{mg}\cdot\text{L}^{-1}$  (Figure 4-1b).



**Figure 4-1. TOC concentration compared to EPA advice note on DBPs in drinking water criteria. A) Temporal changes of TOC, white color show no data values; B) spatial variation of TOC, Boulder Basin (square), Overturn Arm (diamond), and Colorado River Arm (triangle).**

Satellite sensors can only detect surface reflectance emissions which can be applied to determine the concentration level of water quality parameters present in a water column. To assure that the water quality monitoring at the surface is representative of water quality at the drinking water intake's depth at about 300 m (1000 feet) below the lake surface, the vertical profile of TOC was assessed according to Figure 4-2. This figure shows that the concentration of TOC is relatively stable from the surface to the depth of 100 m (i.e. the depth in which the TOC data were available) (Figure 4-2). Therefore, we can assume that the prediction of TOC concentration at the water surface can be applied to the location of drinking water intake.



**Figure 4-2. Vertical profile of TOC in Lake Mead at the location of the closest station to the drinking water intake**

### 4.3.2. Satellite Data

MODIS Terra (MOD09GA) surface reflectance images at 500 m resolution were obtained from the United States Geological Survey (USGS) MODIS reprojection tool web interface (MRTWeb, 2010). In addition, the surface reflectance climate data record (CDR) for Landsat 5 (TM) and Landsat 7 (ETM+) were obtained from the USGS site (USGS, 2014). All images were downloaded for cloud-free days consistent with the collection dates of ground-truth data. In Table 4-2, number of downloaded images and applied land surface reflectance bands are presented.

**Table 4-2. Number of satellite images and selected bands utilized in this study**

<b>Satellite</b>	<b>Spatial Resolution (m)</b>	<b>Band</b>	<b>No. of Satellite imagery</b>
MODIS	500	1-4, 6-7	1384
Landsat 5 TM Landsat 7 ETM+	30	1-5,7	381

## 4.4. Data Synthesis Based On Remote Sensing Imageries

### 4.4.1. Nowcasting Method

The nowcasting method applied in this chapter is based on the IDFM technique developed by Chang et al. (2014 a); (2014b), and it is consistent with the same conceptual framework described in the previous chapter (Figure 4-3). However, in this chapter, the ability of two different machine-learning techniques, including artificial neural network (ANN) and genetic programming (GP) were assessed to estimate surface water quality parameter (i.e. TOC) using extracted spectral reflectance values corresponding to ground-truth dates and the ground-truth data. In both models, 70% of all samples were used to train the model, 15% were applied to

validate the model, and the remaining 15% of the samples were used for testing. The model which showed the best performance is applied to produce TOC concentration maps.

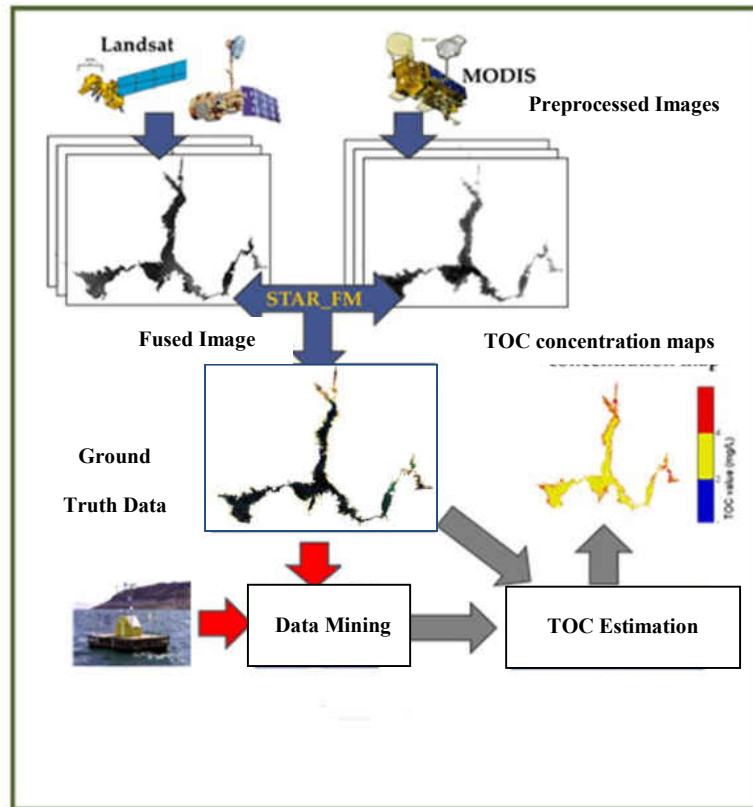


Figure 4-3. IDFM Methodological Flowchart

#### 4.4.1.1. Artificial Neural Network

The MATLAB ANN toolbox with feed forward neural network scheme trained by the Levenberge- Marquardt back propagation algorithm was applied in this study. The MATLAB ANN toolbox was modified by adding a loop to check R-Squared and P-value for each iteration. All inputs (spectral reflectance values from 6 bands of the fused images) in the first layer are distributed to the hidden layer. Each input parameter is multiplied by its connection weight which is randomly determined to make prediction happen with respect to each water quality

parameter. Suitable weights for each connection are found by processing the inputs and comparing the trained outputs of the ANN against the desired ultimate outputs. This comparison is performed through an iterative process, which assigns random connection weights for training and learning and calculates the corresponding P-value and R-squared value. This process is repeated until the highest R-Squared value is found while the P-value satisfies a 95% confidence interval. The ANN model is completed once the validation and testing steps can be successfully completed following a satisfactory performance from the training step.

#### 4.4.1.2. Genetic Programming

The Discipulus™ software applied in this study consists of hundreds of genetic programming (GP) algorithms which are run sequentially for each project. Genetic programming is based on evolutionary computation techniques similar to those used by the genetic algorithm developed by Holland, (1975) with the difference of increasing the complexity of the structures undergoing adaption (Koza, 1992). Different parameters such as crossover rate, mutation rate, population size, instruction set, distribution of initial program size, termination criteria, and parsimony pressure can be adjusted by users in Discipulus (Foster, 2001). The procedure of a single run by Discipulus GP software package is based on the following steps (Francone, 1998):

**Step 1.** Initialize the population size: The initial population size of this study was 80 with a maximum population size of 256. It indicates the size of the population of programs in each project.

**Step 2.** Discipulus selects four programs from the population of programs, compares them based on fitness, and selects two winners and two losers. Fitness is calculated based on comparing the outputs of the evolved program with the target output in the training data.

**Step 3.** Discipulus applies search operators on the winners of the previous section to produce “Children” or “Offspring” as follows:

- a. Replace the losers of section 2 by copying the two winners
- b. Crossover the copies of the winners by the assumed crossover frequency of 50%
- c. Mutate one of the programs by the assumed mutation frequency of 90%
- d. Mutate the other programs resulting from step 3a by the assumed mutation frequency.

**Step 4.** Repeat the last two steps until the run is terminated.

#### *4.4.2. Forecasting Method*

The framework of the applied forecasting model, as shown in (Figure 4-4), is the same as the one described in the previous section for TSS. However, the sensitivity analysis is performed based on four different scenarios which are defined in Table 4-3. The first scenario only includes remote sensing data; Scenario 2 includes all meteorological parameters and remote sensing data; Scenario 3 contains both remote sensing and water level data; Scenario 4 contains all three types of input variables. Scenarios 2 and 3 help us in assessing the importance of meteorological parameters and reservoir water level in regard to the performance of the water quality forecasting model.



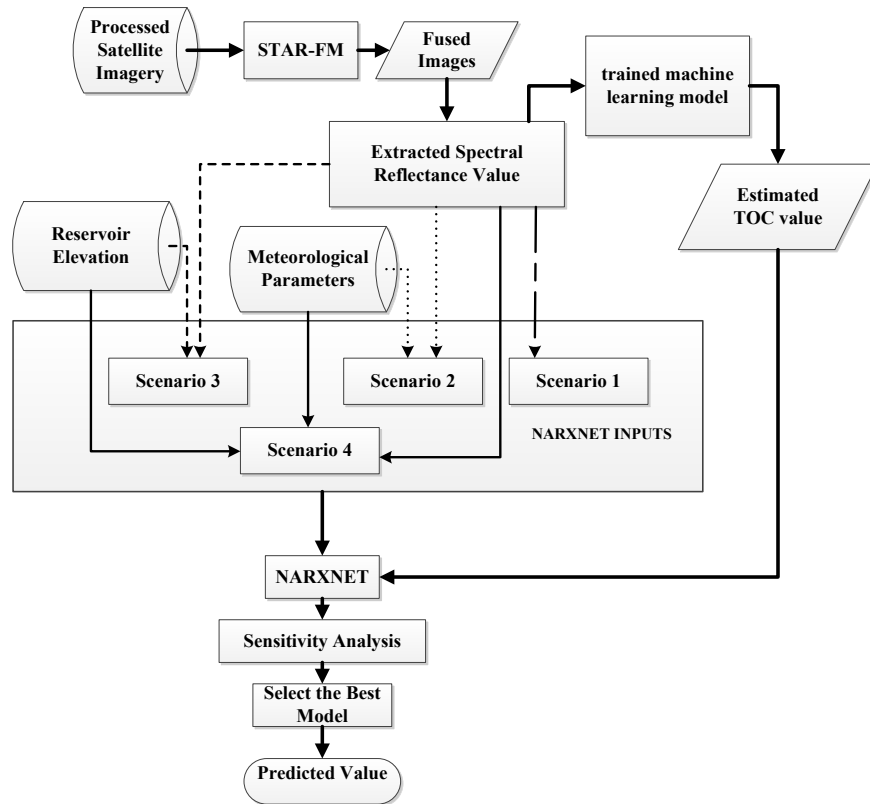


Figure 4-4. Water quality forecasting framework

Table 4-3. Summary of database used in different scenarios

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Spectral reflectance Value	*	*	*	*
Meteorological Parameters <sup>1</sup>		*		*
Water Level in the Reservoir			*	*

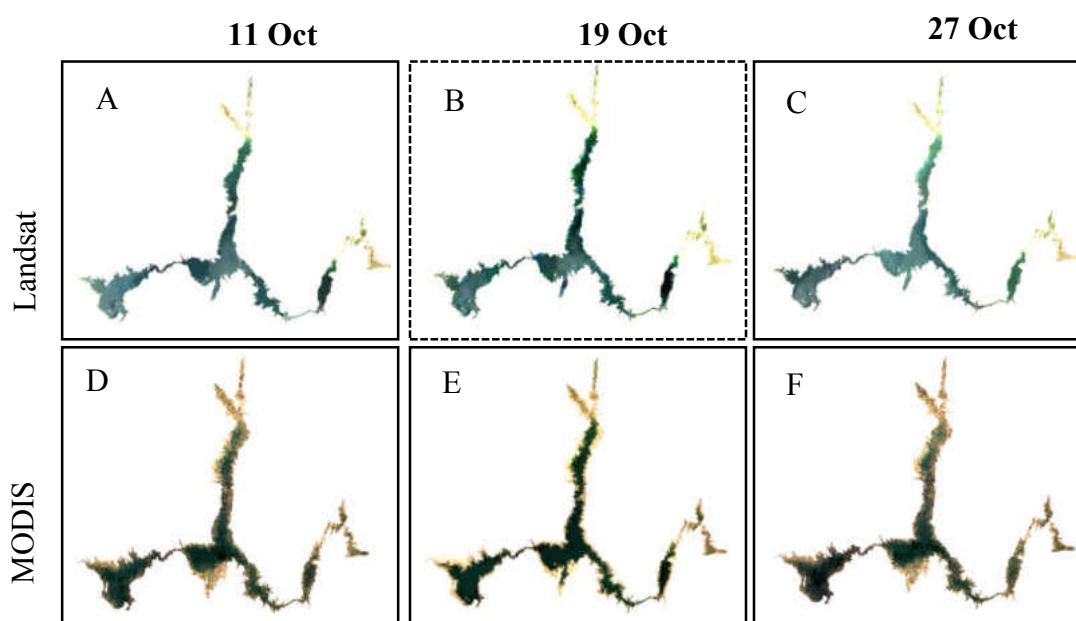
<sup>1</sup>Average wind speed, average air temperature, and average humidity were considered as meteorological parameters.

To compare multiple scenarios, several indices such as root mean squared error (RMSE), mean absolute error (MAE), percent bias (PBIAS), and R-Squared ( $R^2$ ), which were described in the previous chapter, were applied.

## 4.5. Results and Discussion

### 4.5.1. Nowcasting Results

The procedure of generating fused images in Lake Mead is described in Figure 4-5 by using two pairs of images taken before (11-Oct-2006) and after (27-Oct-2006) the ground-truth date (19-Oct-2006). All images in this figure are true-color RGB images. To generate fused image (B), two synthetic images for the pre- and post-condition were created. The pre-condition and post-condition synthetic images were generated using the group of images of A, D, and E as well as C, F, and E, separately. Then, these two synthetic images were used to create the fused image (B).



**Figure 4-5. The ability of IDFM in gap-filling of Landsat images. The fused image is shown in the dashed box, and all images are true color (RGB) images.**

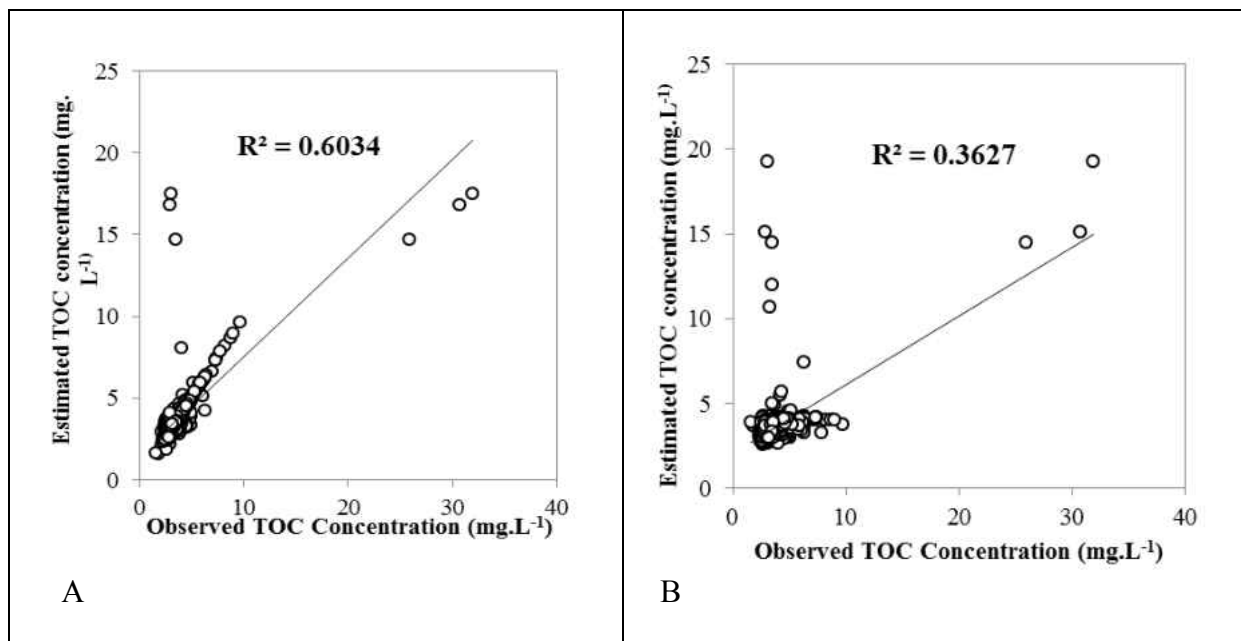
To check the accuracy of the STAR-FM algorithm, R-Squared value between the spectral reflectance values of fused and actual images was calculated for each band (Table 4-4). As

shown in Table 4-4, the calculated R-Squared values vary from 0.95 for band 1 to 0.68 for band 7, which indicates a sufficient reliability of the fused images for machine learning applications.

**Table 4-4. R\_squared value between the spectral reflectance value of the fused image and the actual Landsat image on 19-Oct-2006**

<b>Band</b>	<b>Band 1</b>	<b>Band 2</b>	<b>Band 3</b>	<b>Band 4</b>	<b>Band 5</b>	<b>Band 7</b>
<b>R<sup>2</sup></b>	0.95	0.94	0.94	0.89	0.71	0.68

These fused images and ground truth data were integrated as inputs to feed in the GP and ANN models. Statistical plots between observed and estimated values (Figure 4-6) exhibit comparative advantages of each model. Based on this figure, the ANN model is more suitable than the GP model for estimating the global trend of TOC concentration, although the ANN model is not capable of estimating 3 peak values of TOC accurately. The calculated R-squared values are 0.6 for ANN model and 0.36 for GP model. In addition, the calculated RMSE for ANN and GP are 0.98 and 1.24, respectively. Therefore, results show a stronger correlation between the observed and estimated values in the ANN model as compared to the counterpart. Therefore, the trained ANN model may be fair enough for producing TOC concentration maps which present the spatiotemporal variability of this parameter within the lake. For instance, the resulting concentration maps for TOC reconstructed for two different days are shown in Figure 4-7. It should be mentioned that the surface reflectance values in MOD09 are between 0-10,000. However, using the atmospheric correction algorithm changes this range to -100 to 16,000 (Vermonte et al. 2011). As a consequence, white color in the legend of Figure 4-7 represents pixels with negative surface reflectance values. Pixels with negative surface reflectance values were excluded from the modeling process.

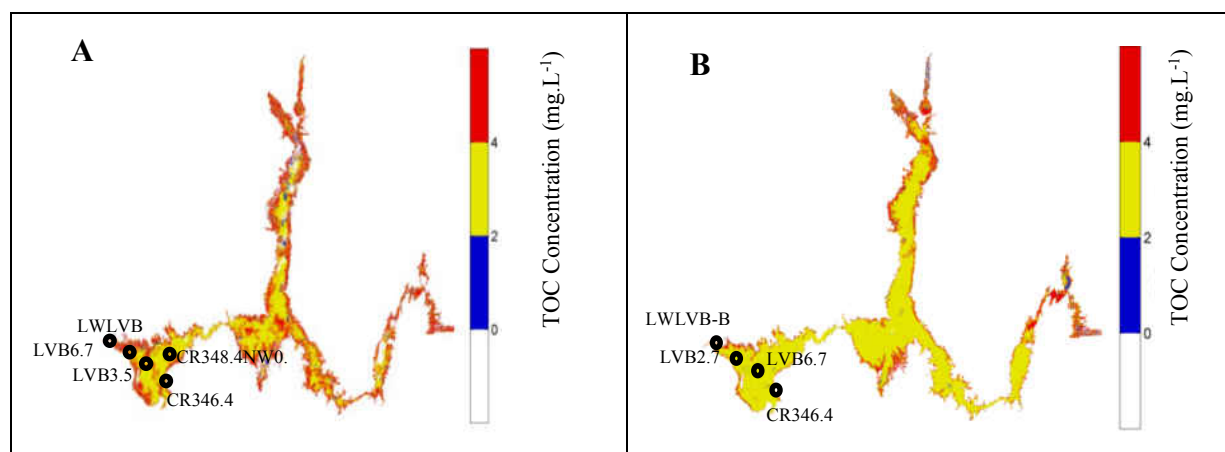


**Figure 4-6. Comparison between the observed values and the estimated values of TOC concentration from 2000-to 2013 via two data mining models (A) ANN model, (B) GP model**

To compare the estimated values with the observed values at the location of the stations for the aforementioned dates, the TOC concentration was categorized into three ranges based on the EPA advice note on DBPs in drinking water which was mentioned earlier. The results of this comparison are shown in Table 4-5. Figure 4-7 and Table 4-5 shows the consistency between the observed TOC values collected at the sampling locations and the categorized ranges estimated based on the fused images developed by the IDFM model. However, adding more ground truth data which are distributed all over the lake may further improve the reliability of the IDFM model developed for water quality estimation.

**Table 4-5. Comparison between observed values at the sampling location with the estimated values for TOC on April 3rd, 2002 and June 20th, 2007.**

Date	TOC Measurement - (mg·L <sup>-1</sup> )					
	Station	LWLVB-B	LVB6.7	LVB2.7	CR346.4	
April 3, 2002	Observed Value	4.598	2.75	3.99	2.61	
	Estimated Value	TOC>4	2< TOC <4			
June 20, 2007	Station	LWLVB	LVB6.7	LVB3.5	CR346.4	CR348.4NW0.8
	Observed Value	4.5	3.9	3.1	3.2	3.1
	Estimated Value	TOC>4		2< TOC <4		



**Figure 4-7. Spatial variations of (A) TOC concentration on June 20th, 2007, (B) TOC concentration on April 3rd, 2002. TOC concentration is categorized based on the EPA advice note on DBPs in drinking water. Pixels with negative surface reflectance values were shown in white.**

In addition to calculating TOC concentration maps for the ground truth dates, the applied nowcasting method is able to provide daily TOC concentration maps for at least a week in which the high TOC concentration was detected. As an example, a week in May 2001 in which the average TOC concentration within the lake was more than 4 mg·L<sup>-1</sup> (Figure 4-1) was selected and the results were shown in Figure 4-8. Therefore, these daily concentration maps provide decision makers and water treatment plant operators with spatiotemporal distribution of TOC throughout the lake.

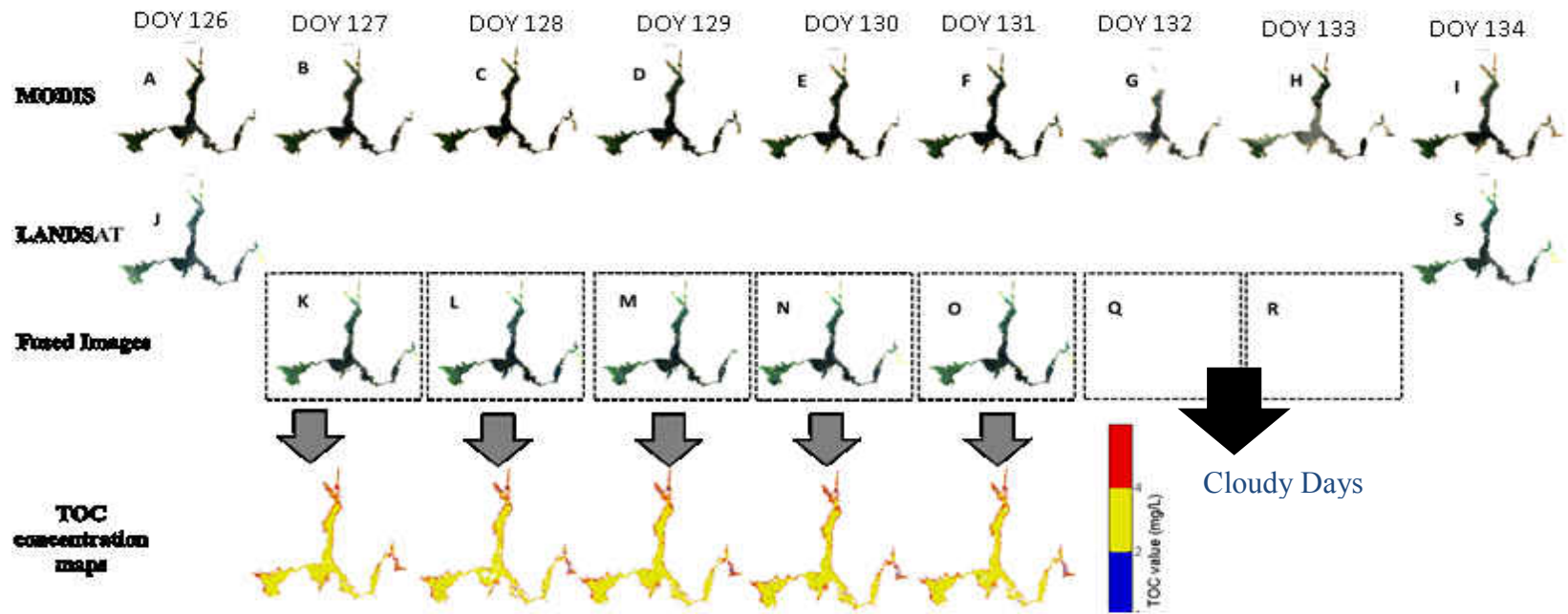
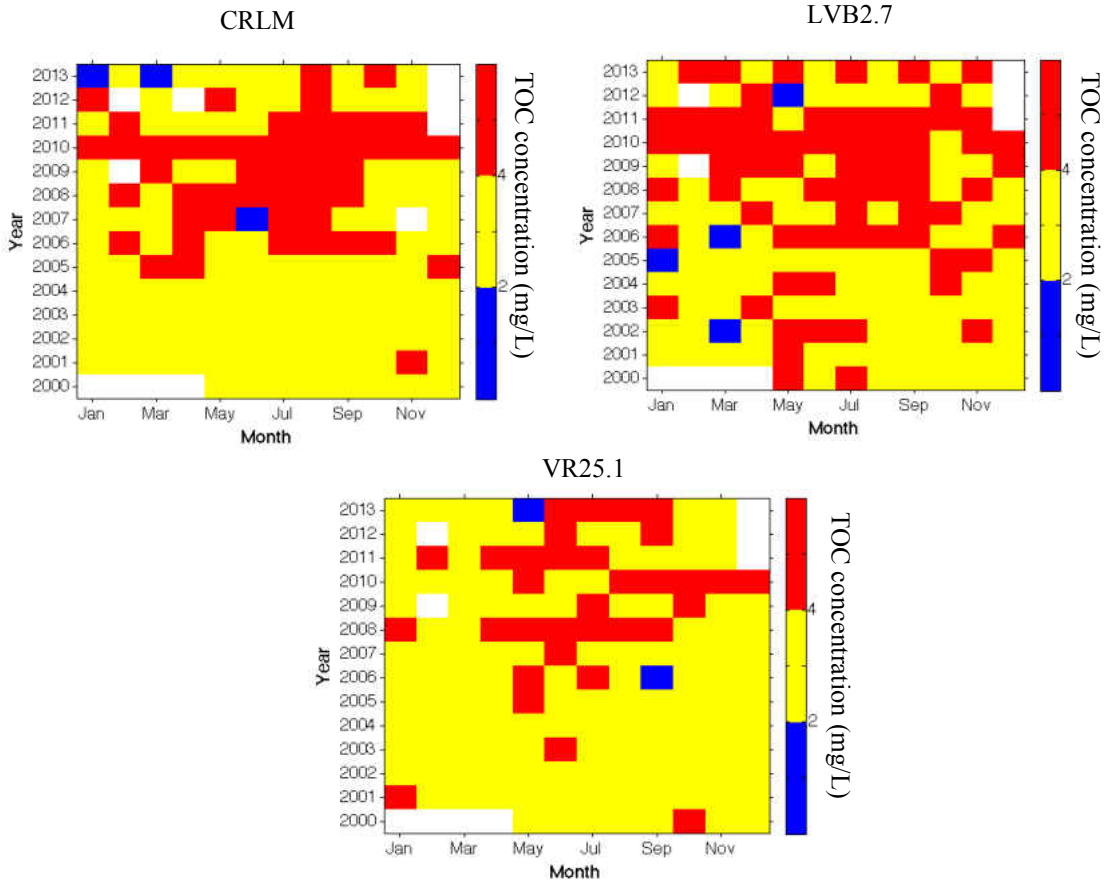


Figure 4-8. Estimating TOC concentration maps using IDFM Method from 6 May 2001 (DOY 126) to 14 May 2001 (DOY 134). The top row in this figure shows the available MODIS images and the second row indicates the available Landsat images in the Lake Mead for the cloud-free days of the selected time period. Third and the last rows show fused images and daily TSS concentration maps, respectively (Obtained from: Imen et al., 2014)

It should be mentioned that only cloud free images can be applied in STAR-FM algorithm. Therefore, STAR-FM algorithm was not able to generate fused images for cloudy days (i.e. DOY 132, 133). Therefore, cloud contamination is a big hindrance in the developed forecasting model.

In addition to having information about spatiotemporal distribution of TOC within the lake, it is important to have an overall view of the impacts of each inflow into the lake on increasing TOC concentration. For this purpose, we need to access the continuous time series of TOC concentration at the location of each arm. However, sampling data within Lake Mead is available for sporadic time periods and in limited numbers to support the local assessment. Therefore, the trained ANN model was applied to estimate TOC concentration using the extracted reflectance values from Landsat or fused images at the location of stations CRLM, VR25.1, and LVB2.7, as shown in Figure 4-9. These stations are representatives of water quality of inflows from Colorado River, Lower Virgin Watershed, and Las Vegas Wash, respectively. As shown in Figure 4-9, the concentration of TOC is categorized into three ranges which were defined by EPA advice note on disinfection by-products in drinking water. Since more than one Landsat image may be available in a specific month, the average of all estimated TOC values for each month was reported in Figure 4-9. As shown in this figure, the frequency of detecting high TOC concentration values has increased since 2006 at all inflow locations, especially at the Colorado River Arm and Las Vegas Wash. As discussed in the previous chapter, the majority of forest fire events occurred in 2005 and only a small percentage occurred outside of the 2003-2007 period (Figure 3-2). However, concentration of TOC within the aforementioned time period is less than the recent years in Overton Arm (i.e. station VR25.1).

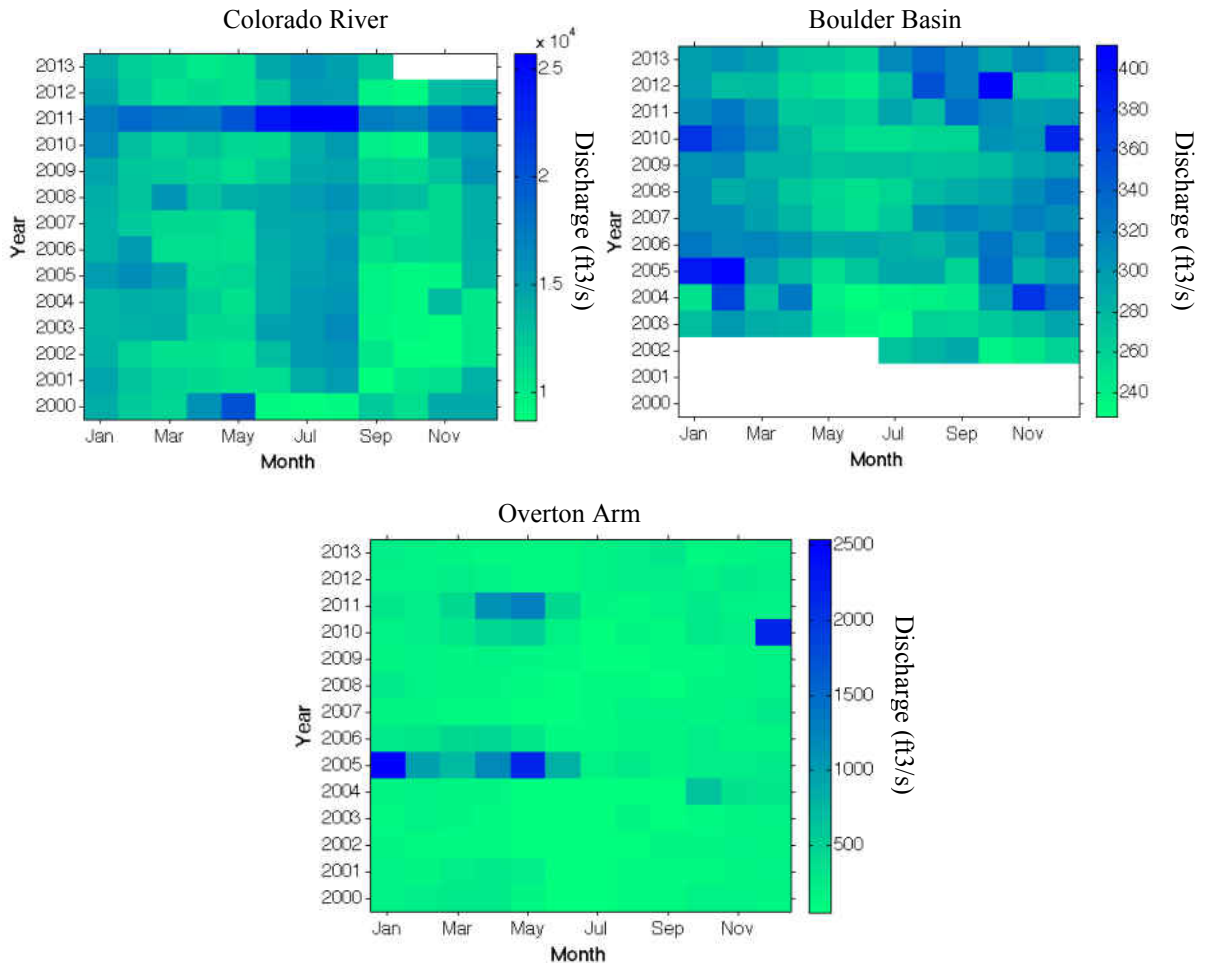


**Figure 4-9. Temporal distribution of TOC concentration at the location of inflows into the lake; white represents “no data” value as a result of inaccessibility of cloud-free Landsat images**

It would be beneficial to assess the temporal variation of inflows into the lake to investigate the probable linkage between the high rates of inflows with higher TOC concentration throughout the lake. For this purpose, the temporal distribution of streamflow is shown in Figure 4-10. Comparing Figure 4-9 to Figure 4-10 does not reveal any evidence of a linkage between high inflow rates and increases in TOC concentration. For instance, the high inflow rates detected during Jan-Apr 2005, which coincides with the largest forest fire events in the Virgin River Watershed, do not impact TOC concentration at Overton Arm. Therefore, there



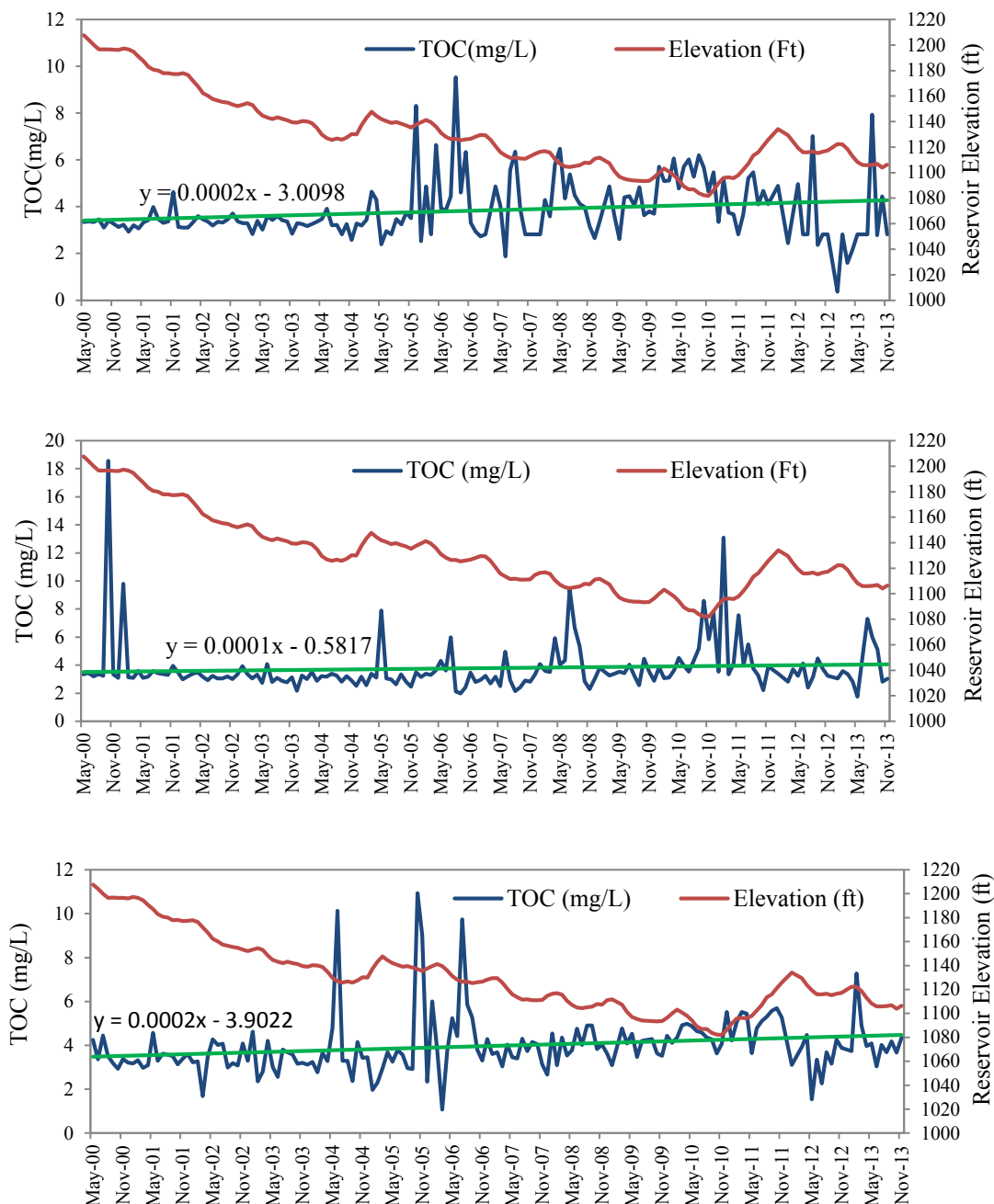
is not any evidence which reveals the effects of forest fire events on increasing TOC concentration within the lake.



**Figure 4-10 Temporal distribution of streamflow at the location of USGS stations with associated color legend for each arm; white represents “no data” value**

One of the possible factors causing an increase in TOC concentration since 2005 in all arms of the lake could be the decade-long drought. The important parameter which can be used as a representative of the impacts of drought on Lake Mead is the lake’s elevation. For this purpose, the comparison between TOC concentration and reservoir elevation is shown in Figure 4-11. While the lake’s elevation has a drastic drop, TOC concentration has a gradually

increasing trend. Except some months which show abrupt increase in TOC concentration, the TOC concentration is almost constant within 2000-2013.



**Figure 4-11 Comparison between temporal variations of TOC concentration and Lake Mead elevation at Hoover dam at the location of stations: (A) CRLM (B) VR25.1 (C) LVB2.7. Green line shows the trend of TSS concentrations, and red line indicates the variation of elevation during (2000-2013).**

#### 4.5.2. Forecasting Results

The successful fitting process resulted from the ANN nowcasting analysis is extended in this section to shape the forecasting model for TOC concentration on a daily basis onwards. This ANN nowcasting model is called the trained-ANN model hereafter. As the first step, fused images were produced for a minimum of 30 days before the first forecasting date using the STAR-FM algorithm as long as the available MODIS and Landsat images within the selected time period have no severe cloud coverage. These 30 days were selected from the years in which the higher concentration of TOC was detected (Figure 4-1). As an example, the time period from August 14, 2001 to October 14, 2001 was selected to assess the reliability of the forecasting model when tackling the high episodes of TOC concentration values. To apply the results of the forecasting model in the drinking water infrastructure assessment, the closest station to the drinking water intake was selected (i.e., station LVB6.7 in Figure 4-7). It should be mentioned that the last date of the selected time period was applied to assess the reliability of the forecasting model by comparing the forecasted value with the ground truth data on October 14, 2001. First, spectral reflectance values of all 6 bands were extracted from the Landsat and produced fused images in the aforementioned time period at the location the LVB6.7. These extracted values were applied as inputs into the trained-ANN model to estimate daily time series of TOC for selected time period. Comparison between the estimated and observed TOC values at the location of LVB6.7 during the selected time period is shown in Table 4-6. As shown in the table, although the trained-ANN model mostly underestimated TOC concentration, the difference between the estimated and observed values is not significant.

**Table 4-6. Comparison between the estimated and observed values of TOC at LVB 6.7 during Aug14th –Oct 13th 2001**

<b>Date</b>	<b>Observed TOC (mg·L-1)</b>	<b>Estimated TOC (mg·L-1)</b>
28-Aug-2001	4.69	3.9
4-Sep-2001	3.98	3.86
25-Sep-2001	3.83	3.44
7-Oct-2001	3.74	4.13

The subsequent step was to construct the NARXNET model to forecast TOC at the location of drinking water intake (i.e. station LVB6.7) for a day ahead of time (i.e. Oct 14, 2001). To reduce the uncertainty, multiple runs (10 runs) for cross-validation were performed for each scenario, and the average of calculated error indices over the rounds were reported in Table 4-7. As shown in this table, scenario 1 with having only spectral reflectance value of remote sensing showed the lowest R-squared value and the highest errors. Although adding the additional inputs such as meteorological parameters or reservoir level improved the results, scenario 4 which includes both of these environmental parameters cannot exhibit the best outcome. Among the whole group of 4 scenarios, scenario 3 shows the highest R-squared value and the lowest estimation errors. This finding indicates the importance of considering the water level in the TOC forecasting model which is consistent with the result obtained in the previous chapter for TSS.

**Table 4-7. Summary of the results of error indices; Values of RMSE, PBIAS, and MAE equal to zero, and R-squared value equal to one show a perfect fit.**

<b>Scenario</b>	<b>R<sup>2</sup></b>	<b>RMSE</b>	<b>MAE</b>	<b>PBIAS (%)</b>
Scenario1	0.78	0.24	0.13	1.38
Scenario2	0.81	0.26	0.12	-0.45
Scenario 3	0.85	0.23	0.10	-0.91
Scenario 4	0.83	0.35	0.13	-1.06

In addition to comparing scenarios based on statistical indices, the accuracy of predicted TOC values for each scenario was assessed by comparing the forecasting results with synchronous ground truth data at the same location. Since NARXNET was run 10 times for each scenario, the minimum, maximum, mean, and standard deviation of predicted values for each scenario are calculated (Table 4-8). As shown in Table 4-8, scenarios 1 and 4 show the highest standard deviation and the lowest minimum TOC concentration. Although the amount of the standard deviation in scenario 2 is lower than scenario 3, the range of predicted values in scenario 3 was mostly closer to the recorded ground truth data (4.74 mg·L<sup>-1</sup>). Therefore, the results in Table 4-7 and Table 4-8 collectively confirm that scenario 3 is the most reliable scenario out of the 4 scenarios considered as inputs into the NARXNET model.

**Table 4-8. Comparison of predicted values in different scenarios**

<b>Scenario</b>	<b>Mean</b>	<b>Min</b>	<b>Max</b>	<b>Standard Deviation</b>
1	3.89	2.76	4.68	0.71
2	3.49	3.02	4.18	0.43
3	3.67	2.91	4.79	0.50
4	2.98	2.00	4.26	0.77

#### 4.6. Final Remarks

The treatment technique for removal of TOC lessens the concentration of disinfection by-product precursors (DBP-Ps) available to form DBPs during disinfection. Therefore, having information of probable increase of TOC a day ahead of time can help the water treatment plant operator in preventing the formation of THMS which may pose a risk to human health. Therefore, the main purpose of this study is to develop a water quality forecasting model which can be applied in an early warning system to help the water treatment plant operator achieve this

goal. This early warning system turns a broad range of information from multiple sources into a reliable decision.

#### 4.7. Conclusions

The accuracy of the IDFM method for nowcasting of TOC concentrations was assessed with three steps. First, the reliability of the fused images was assessed and confirmed by examining the correlation between the spectral reflectance values of each pixel in a fused image with its corresponding pixel in an actual Landsat image. The resulting R-squared values varied between 0.68 and 0.95. After assuring the reliability of the resulting fused images for machine learning applications to retrieve the water quality information, the capability of the trained ANN and GP models for estimating TOC concentrations were assessed in comparison with the ground truth data. Results show that the trained ANN model has the higher correlation between the estimated and observed values for the ANN model. In the last stage, the accuracy of the developed concentration maps were evaluated by comparing some of the ground truth data with the estimated range of TOC concentration values to prove the fidelity of the trained ANN model for nowcasting. Besides the ability of the nowcasting method in providing information about the spatiotemporal variability of TOC concentration, this method can provide us with daily TOC concentration maps for at least a week in which high concentration of TOC was detected. The calculated monthly temporal distribution of TOC concentration in each arm showed increasing in TOC concentration since 2005. In addition, comparing these figures with the temporal distribution of inflows into the lake did not show any linkage between high inflow rates with increasing in TOC concentration. Furthermore, there is not any evidence of impacts of forest fire event on increasing TOC concentration at Overton Arm.

Forecasting TOC concentration is of great importance in drinking water treatment. It provides information for water treatment plant operators and makes them aware of the occurrence of DBP's in drinking water ahead of time which is considered as a threat to human health. The forecasting model developed in this study also showed the application potential by having a model-based iterative process which may be constantly updated by newly acquired satellite imagery, and reservoir elevation. Forecasting can be made possible periodically by retrieving the memory of past states of water quality to be in concert with the most updated information of the current state of water quality. Such a numerical scheme is formulated by the NARXNET model, which is able to predict TOC concentration one day ahead of the time using updated input and output values repetitively on a rolling basis onwards. The reliability of the forecasting model developed herein was evaluated by comparing its forecasted results with the sampling data for those years in which the highest concentration of TOC was detected. The results confirmed the potential application associated with scenario 3 in forecasting TOC concentration. Such potential application is limited for the areas which are rarely overcast. This is caused by the STAR-FM algorithms requirement of cloud-free images. However, this problem can be remedied through the integration of this model with some cloud removal algorithms.

Overall, this study develops a decision making framework for nowcasting and forecasting TOC concentrations at the location close to the drinking water intake in Lake Mead. With the aid of IDFM algorithm, the developed framework will be able to capture the spatial and temporal distributions of TOC concentrations at the finer scales throughout the lake. In addition, the iterative process for updating the data set in the model supports its forecasting processes and makes the model a cost-effective, forward-looking, and risk-informed decision support component for drinking water treatment processes.

#### 4.8. Reference

- Aguilera P.A., Frenich A.G., Torres J.A., Castro H., Vidal J.L.M., Canton M. 2001, Application of the Kohonen neural network in coastal water management: methodological development for the assessment and prediction of water quality, *Water Research*, vol. 35, no. 17, pp. 4053-4062.
- Ahn, Y. 1999, Development of inverse model from ocean reflectance, *Marine Technology Society Journal*, vol. 33, no. 1, pp. 69-80.
- Brivio P.A., Giardino C., Zilioli E. 2001, Determining of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images, *International Journal of Remote Sensing*, vol. 22, pp. 487-502.
- AWWA with assistance from economic an engineering servise, Inc. 2002. Finished water storage facilities, U.S. Environmental Protection Agency office of ground water and drinking water standards and risk division, pp.22.
- Brezonik P., Menken K.D., Bauer M., 2005, Landsat-based remote sensing of lake water quality characteristics, including chlorophyll and colored dissolved organic matter (CDOM), *Lake and Reservoir Management*, vol. 21, no. 4, pp. 373-382.
- Cai S., Kang X., Zhang L. 2012, Time series prediction based on artificial neural network for estimation of forest biomass, *International Journal of Advancements in Computing Technology*, vol. 4, no. 12, pp. 318-326.
- Chen C., Shi P., Mao Q. 2003, Application of Remote Sensing Techniques for Monitoring the Thermal Pollution of Cooling-Water Discharge from Nuclear Power Plant, *Journal of Environmental Science and Health, Part A- toxic/Hazardous Substances and Environmental Engineering*, A38(8), 1659-1668.
- Chang, N. B., Daranpob, A., Yang, J., Jin, K. R., 2009, A Comparative Data Mining Analysis for Information Retrieval of MODIS Images: Monitoring Lake Turbidity Changes at Lake Okeechobee, Florida, *Journal of Applied Remote Sensing*, vol. 3, 033549.
- Chang N.B., Xuan Z., 2013, Monitoring Nutrient-Concentration in Tampa Bay with MODIS Images and Machine Learning Models, IEEE Conference. pp. 702-707, 978-1-4673-5200-0/13/\$31.00©2013IEEE
- Chang N.B., Vannah B.W., Yang Y.J., Elovitz M., 2014a, Integrated Data Fusion and Mining Techniques for Monitoring Total Organic Carbon Concentrations in Lake, *International Journal of Remote Sensing*, vol. 35, no. 3, pp. 1064-1093.
- Chang N.B., Vannah B.W., Yang Y.J., Elovitz M. 2014b, Comparative sensor fusion between hyperspectral and multispectral satellite sensors for monitoring microcystin distribution in



- Lake Erie, *IEEE Journal of Selected Topics on Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2426-2441.
- Chen L. 2003, A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data, *International Journal of Remote Sensing*, vol. 24, no. 11, pp. 2265-2275.
- Chen L., Tan C.H., Kao S.J., Wang T.S., 2008, Improvement of remote monitoring on water quality in a subtropical reservoir by incorporating grammatical evaluation with parallel genetic algorithms into satellite imagery, *Water Research*, vol. 42, no. 1-2, pp. 296-306.
- Covay, K.J., and Beck, D.A., 2001, Sediment-deposition rates and organic compounds in bottom sediment at four sites in Lake Mead, Nevada, May 1998: U.S. Geological Survey Open-File Report 2001-282, 34 p. (Also available at <http://pubs.er.usgs.gov/publication/ofr01282>.)
- Dalponte M., Bruzzone L., Gianelle D., 2008, Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416-1427.
- Dekker A.G., 1993, Detection of optical water parameters for eutrophic lakes by high resolution remote sensing, PhD dissertation, Free University, Amsterdam
- Dekker, A.G., Zamurovic-Nenad, Z., Hoogenboom, H.J., Peters, S.W.M., 1996, Remote Sensing, Ecological Water Quality Modelling and In situ Measurements: a Case study in Shallow Lakes, *Hydrological Sciences*, vol. 41, no. 4, pp. 531-547.
- Dekker, A.G., Vos, R.J., Peters, S.W.M., 2001, Comparison of Remote Sensing Data, Model Results and In Situ for Total Suspended Matter (TSM) in the Southern Frisian Lakes, *The Science of the Total Environment*, vol. 268, pp. 197-214
- Dekkere, A.G., Vos, R.J., Peters, S.W.M., 2002, Analytical algorithms of lake water TSM estimation for retrospective analyses of TM and SPOT sensor data, *International Journal of Remote Sensing*, vol. 23, no. 1, pp. 15-35.
- Ehler M. 1991, Multisensor image fusion techniques in remote sensing, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 46, no. 1, pp. 19-30.
- Extreme Learning Machine, 2013, [http://www.ntu.edu.sg/home/egbhuang/elm\\_codes.html](http://www.ntu.edu.sg/home/egbhuang/elm_codes.html).
- Foster J.A., 2001, Review: Discipulus: A commercial genetic programming system, *Genetic Programming and Evolvable Machines*, vol. 2, no. 2, pp. 201-203.
- Francone, D.F., 1998, Discipulus Software Owner's Manual, Version 3.0 DRAFT. Colorado: Machine Learning Technologies, Inc.
- USEPA, 1998. National primary drinking water regulations: disinfectants and disinfection byproducts: final rule, Final Rul. Fed. Reg. 63/24/69478.

- Gao F., Masek J., Schwaller M., Hall F., 2006, On the blending of Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2207-2218.
- Ge Z.F., Frick W.E., 2009, Time-frequency analysis of beach bacteria variations and implication for recreational water quality modeling, *Environmental Science and Technology*, vol. 43, no. 4, pp. 1128-1133.
- Giardino C., Brrando V.E., Dekker A.G., Strombeck N., Candiani G., 2007, Assessment of water quality in Lake Garda (Italy) using Hyperion, *Remote Sensing of Environment*, vol. 109, no. 2, pp. 183-195.
- Gitelson A., Garbuzov G., Szilagyi F., Mittenzwe K.H., Karnieli A., Kaiser A., 1993, Quantitative Remote Sensing Methods for Real-Time Monitoring of Inland Water Quality, *International Journal of Remote Sensing*, vol.14, no. 7, pp. 1269-1295.
- Gupta H.V., Sorooshian S., Yapo P.O., 2004, Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration, *J. Hydrologic Eng.*, vol. 4, no. 2, pp. 135-143
- Hamilton H.G., Schladow S.G., 1997, Prediction of water quality in lakes and reservoirs. Part I-model description, *Ecological Modelling*. Vol. 96, no. 1-3, pp. 91-110.
- Haw H.G., Chan Q.I., Qiao J.F., 2011, An efficient self-organizing RBF network for water quality prediction, *Neural Networks*, vol. 24, no. 7, pp. 717-725.
- He L.M., He Z.L., 2008, Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in Southern California, USA. *Water Research*. Vol. 42, pp. 2563-2573.
- Holland J.M., 1975, Adaption in natural and artificial systems, Ann Arbor, MI: University of Michigan Press.
- Imen S., Chang N.B., Yang Y.J., 2014, Monitoring spatiotemporal total organic carbon concentrations in Lake Mead with integrated data fusion and mining (IDFM) technique, *11<sup>th</sup> international conference in hydroinformatics, HIC 2014*, New York, USA.
- Keiner L.E., Yan X.H. 1998, A neural network for estimating sea surface chlorophyll and sediments from Thematic Mapper imagery, *Remote Sensing of Environment*, vol. 66, no. 2, pp. 153-165.
- Kishino, M., Ishimaru, T., Furuya, K., Oishi, T., Kawasaki, K., 1998, In-water Algorithms for ADEOS/OCTS, *Journal of Oceanography*, vol. 54, no. 5, pp. 431-436.
- Koponen S., Pulliainen J., Kallio K., Hallikainen M., 2002, Lake water quality classification with airborne hyperspectral spectrometer and simulated MERIS data, *Remote Sensing of Environment*, vol. 79, no. 1, pp. 51-59.

- Koza J.R., 1992, Genetic programming: on the programming of computers by means of natural selection, Cambridge, MA: MIT Press.
- Las Vegas Valley Water District, 2013, Water quality report, website UPL: [www.lvvwd.com/assets/pdf/wqreport.pdf](http://www.lvvwd.com/assets/pdf/wqreport.pdf)
- Legates D.R., McCabe G.J., 1999, Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Res*, vol. 35, no. 1, pp. 233-241
- Ge Z.F., Frick W.E., 2009, Tim-frequency analysis of beach bacteria variations and implication for recreational water quality modeling, *Environmental Science and Technology*. Vol. 43, no. 4, pp. 1128-1133.
- Maier H.R., Dandy G.C., 1996, The use of artificial neural networks for prediction of water quality parameters, *Water Resources Research*, vol. 32, no. 4, pp. 1013-1022.
- Makkeasorn A., Chang N.B., Zhou X., 2008, Short-term streamflow forecasting with global climate change implications: A comparative study between genetic programming and neural network models, *Journal of Hydrology*, vol. 352, pp. 336-354.
- MATLAB R 2013a, <http://www.mathworks.com/>
- Methews M.W., Bernard S., Winter K., 2010, Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoelei, a small hypertrophic lake, using MERIS, *Remote Sensing of Environment*, vol. 114, no. 9, pp. 2070-2087.
- Minghelli-Roman A., Polidori L., Mathieu-Blanc S., Laubersac L., Cauneau F., 2006, Spatial Resolution Improvement by merging MERIS-ETM images for coastal water monitoring, *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 2, pp. 227-231.
- MODIS Reprojection Tool Web Interface(MRTWeb), 2010. <https://mrtweb.cr.usgs.gov/>, Page Last Modified: April 15, 2010
- Mohamadshafiee M.R., Taghavi L., 2012, Health effects of Trihalomethanes as chlorinated disinfection by products: A review article, *World Academy of Science, Engineering, and Technology*, vol. 6, pp. 1773-1779.
- Nikolaou A.D., Kostopoulou M.N., Lekkas T.D., 1999, Organic by-products of drinking water chlorination, *Glob Nest: the Int J.*, vol. 1, no. 3, pp. 143-156.
- Panda, S.S., Garg, V., Chaubey, I., 2004, Artificial Neural Networks Application in Lake Water Quality Estimation using Satellite Imagery, *Journal of Environmental Informatics*, vol. 4, no. 2, pp. 65-74.
- Phinn, S.R., Dekker, A.G., Brando, V.E., Roelfsema, C.M., 2005, Mapping Water Quality and Substrate Cover in Optically Complex Coastal and Reef Waters: an Integrated Approach, *Marine Pollution Bulletin*, vol. 51, pp. 459-469.

- Pohl, C., Van Genderen, J.L., 1998, Multisensor image fusion in remote sensing: concepts, methods and applications, *International Journal of Remote Sensing*, vol. 19, no. 5, pp. 823-854.
- Pullianen J., Kallio K., Elohemio K., Koponen S., Servomaa H., Hannonen T., Tauriainen S., Hallikainen H., 2001, A semi-operative approach to lake water quality retrieval from remote sensing data, *Science of the Total Environment*, vol. 268, no. 1-3, pp. 79-93.
- Randolph, K., Wilson, J., Tedesco, L., Li, L., Pascual, D.L., Soyeux, E., 2008, Hyperspectral Remote Sensing of Cyanobacteria in Turbid Productive Water using Optically Active Pigments, Chlorophyll-a and Phycocyanin, *Remote Sensing of Environment*, vol. 112, pp. 4009-4019.
- Seebach L., Strobl P., San-Miguel-Ayanz J., 2011, Data fusion of different spatial resolution remote sensing images applied to forest-type mapping, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 4977-4986.
- Solberg A.H., Schistad, Jain A.K., Taxt T., 1994, Multisource classification of remotely sensed data: fusion of Landsat TM and SAR images, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 768-778.
- Song, K., Lu, D., Li, L., Li, S., Wang, Z., Du, J., 2012, Remote Sensing of Chlorophyll-a Concentration for Drinking Water Source Using Genetic Algorithms (GA)-Partial Least Square (PLS) Modeling, *Ecological Informatics*, vol. 10, pp. 25-36.
- Sudheer, K.P. Chaubey, I., Garg, V., 2006, Lake water quality assessment from Landsat TM Data using Neural Network: An approach to optimal band combination selection, *Journal of American Water Resources Society*, vol. 42, no. 6, pp. 1683–1695.
- Tietjen T., Holdren G.C., Rosen M.R., Veley R.J., Moran M.J., Vanderford B., Wong W.H., Drury D.D., 2012, Lake water quality: chapter 4 in a synthesis of aquatic science for management of Lake Mead and Mohave, USGS Circular, vol. 1381-4, pp. 35-68.
- Tyler A.N., Svab E., Preston T., Presing M., Kovacs W.A., 2006, Remote sensing of the water quality of shallow Lakes: a mixture modelling approach to quantifying phytoplankton in water characterized by high-suspended sediment, *International Journal of Remote Sensing*, vol. 27, no. 8, pp. 1521-1537.
- USGS, Earth Explorer, <http://earthexplorer.usgs.gov/>, Page Last Modified: 11/03/2014.
- Vermote E., Kotchenova S., Ray J., 2011, MODIS surface reflectance user's guide, MODIS Land Surface Reflectance Science Computing Facility. 1.3, 40p
- YSI, 2006, Monitoring and Protecting Water Quality in Lake Mead, Nevada. Application Note, <http://www.ysisystems.com/media/pdfs/A539-Monitoring-and-Protecting-Water-Quality-in-Lake-Mead-Nevada.pdf>

Zhang Y., Pulliainen J., Koponen S., Hallikainen M., 2002, Application of an empirical neural network to surface water quality estimation in the Gulf of Finland using combined optical data and microwave data, *Remote Sensing of Environment*, vol. 81, no. 2-3, pp. 327-336.

Zhang Y., Pulliainen J.T., Koponen S.S., Hallikainen M.T., 2003, Water quality retrievals from combined Landsat TM data and ERS-2 SAR data in the Gulf of Finland, *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 622-629.

## CHAPTER 5: GENERAL CONCLUSIONS AND STUDY LIMITATIONS

### 5.1. Conclusions

Using empirical mode decomposition as well as wavelet analysis in this dissertation overcomes the problem of non-stationarity of the ocean-atmosphere system by excluding a wide range of unknown frequencies and intrinsic trends embedded in the raw data. The dominant oscillation of both SST and precipitation was detected within the inter-annual frequency band (2-4 years). The process of screening the oceanic index regions led to the detection of two regions in the Atlantic Ocean which contribute more than 75% of the precipitation variability over the Upper Colorado River Basin. One of these regions is located adjacent to the known teleconnection pattern AMO, and another one is found in the tropical Atlantic Ocean. In addition, a non-leading teleconnection pattern located on the west coast of South America was detected which has a substantial contribution (86%) on precipitation variability in winter compared to all identified oceanic index regions. This finding emphasizes the importance of considering non-leading teleconnection patterns as well as the leading teleconnection patterns. Results of the developed forecasting model demonstrate its capability in capturing the precipitation variability over the Upper Colorado River Basin. This model also showed negligible error in estimating the potential of hydrological droughts in fall, spring, and summer. The calculated error in winter (about 20%) implies that sea surface temperatures alone may not suffice for predicting precipitation in this season and other meteorological variables may need to be included to improve forecasts. Results of the developed precipitation forecasting model can be applied as inputs into the CBRFC streamflow forecasting model in order to improve the forecasted streamflow ensembles. In addition, the predictions produced by the developed model

for the intensity of hydrological droughts over the Upper Colorado River a season ahead, alongside the updated monthly inflow forecast, could be used at the beginning of each month to update the U.S. Bureau of Reclamation's "24-Month Study" model.

The creation of the nowcasting water quality model lays down a foundation for monitoring TSS and TOC, on a near real-time basis, using the integrated remote sensing data fusion and mining technique. Results demonstrated that the use of the nowcasting method was suitable for monitoring spatial and temporal variability of TSS and TOC in Lake Mead. The nowcasting results for TSS showed the importance of the Colorado River in bringing TSS into the Lake. Also, the decreasing trend of TSS at the Las Vegas Wash indicated the effects of the construction of erosion control structures upstream of the Las Vegas Wash. Results didn't show any association between the increase in TSS/TOC concentration and the occurrence of forest fire events. However, the probable impacts of a decade of drought was obvious in the increase of TSS concentration compare to TOC in the Lake. Comparing the TOC concentration maps with the dates of recorded high TSS/TOC concentration showed the capability of the nowcasting model in detecting hot spots.

The novelty of this study lies in the development of a forecasting model to predict TOC and TSS values with the aid of remote sensing technologies on a daily basis. Results showed a strong correlation of 94% and 85% between the observed and predicted values for TSS and TOC, respectively. In addition, the calculated error indices showed the capability of the forecasting model in capturing the range of both TSS and TOC. Among the defined scenarios, the scenario which includes both satellite spectral reflectance values and reservoir elevation was selected as

the best scenario for both TSS and TOC. This finding highlights the reservoir water level as a significant parameter in predicting the concentration of TSS and TOC a day ahead.

## 5.2. Study Limitations and Recommendations

Since the dominant oceanic forcing regions associated with precipitation responses are identified based on the integrated empirical approaches, the developed methodology is a case-based approach. Therefore, the identified oceanic index regions cannot be applied for other regions in the continental United States; even for a region located within the Upper Colorado River Basin (i.e. Weminuche Wilderness area), the oceanic index regions which show the highest contribution to precipitation variability can be different (Chang et al., 2015).

Furthermore, the detected dominant oscillation for SST and precipitation (i.e. 2-4 years) is highly dependent on the selected time period (1997-2014). Therefore, for time periods which are longer or shorter than the selected time period of this study, the wavelet analysis needs to be performed again.

Data mining methods applied in this study to find the relationship between target environmental parameters and remotely sensed data are dependent on the range of ground-truth data, and their results do not have multi-temporal variability. Therefore, these methods are not able to predict the amount of TOC/TSS which is outside of the recorded ground-truth data range. This issue can be remedied with time by calibrating the data mining methods using the future ground-truth data. Hence, periodic sampling can calibrate the model for the gradual changes of TOC/TSS over long timespans.



Predictions of ELM and ANN models include uncertainties. To reduce the effects of these uncertainties and improve results, each model was run several times (e.g. 200 trials) and the average of the results of all trials is reported.

Cloud contamination is a big hindrance in exploitation of satellite images retrieved from the visible and infrared spectral ranges, and the developed forecasting model is not able to forecast TSS/TOC concentration for cloudy days. The developed early warning system can be further improved by including a cloud removal algorithm developed by Chang et al. (2015) to be able to forecast TOC/TSS at the location of drinking water intake during cloudy days.

### 5.3. Reference

- Chang N.B., Bai K., Chen C.F., 2015. Smart information reconstruction via time-space-spectrum continuum for cloud removal in satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. vol. PP, no. 99, pp. 1-15.
- Chang N.B., Imen S., Bai K., Yang J., 2015, The impacts of global non-leading teleconnection signals on terrestrial precipitation across North and Central America, *International Journal of Climatology*, In Review.

## **APPENDIX: MATERIALS UNDER REVIEW**

The contents of chapters 1, 2, 3, and 4 are submitted for review as follows:

- The content in chapter 1 has been submitted for publication as: Chang, N. B., Imen, S., Bai, K. X., and Yang, J., “The Impact of Global Non-leading Teleconnection Signals on Terrestrial Precipitation across North and Central America”, *International Journal of Climatology*, in review, 2015
- The content in chapter 2 will be submitted for publication as: Imen, S., Chang, N. B., and Yang, Y.J., “Attributional Assessment of Local Teleconnection Signals during Recent Droughts in the Upper Colorado River Basin for Long-term Precipitation Forecasting”, *Journal of Hydrology*, in preparation, 2015.
- The content in chapter 3 has been submitted for publication as: Imen S., Chang N.B. , Yang Y.J., “Developing the remote sensing-based early warning system for monitoring TSS concentrations in Lake Mead”, *Journal of Environmental Management*, in review, 2015.
- The content in chapter 4 will be submitted for publication as: Imen, S., Chang, N. B., Yang, Y.J., “ Remote Sensing for Monitoring Total Organic Carbon Concentrations in Lake Mead under Drought and Forest Fire Impact”, *Journal of Hydrology*, in preparation, 2015.