



Western Michigan University  
ScholarWorks at WMU

---

Dissertations

Graduate College

---

8-2020

## Statistical Downscaling Techniques to Enhance the Spatial Resolution of the Grace Satellite Data and to Fill Temporal Gaps

Hossein Sahour

Western Michigan University, [hossein.sahour@gmail.com](mailto:hossein.sahour@gmail.com)

Follow this and additional works at: <https://scholarworks.wmich.edu/dissertations>



Part of the Geology Commons, and the Hydrology Commons

---

### Recommended Citation

Sahour, Hossein, "Statistical Downscaling Techniques to Enhance the Spatial Resolution of the Grace Satellite Data and to Fill Temporal Gaps" (2020). *Dissertations*. 3634.

<https://scholarworks.wmich.edu/dissertations/3634>

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact [wmu-scholarworks@wmich.edu](mailto:wmu-scholarworks@wmich.edu).



STATISTICAL DOWNSCALING TECHNIQUES TO ENHANCE THE SPATIAL RESOLUTION  
OF THE GRACE SATELLITE DATA AND TO FILL TEMPORAL GAPS

by

Hossein Sahour

A dissertation submitted to Graduate College  
in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy  
Geological and Environmental Sciences  
Western Michigan University  
August 2020

Doctoral Committee:

Mohamed Sultan, Ph.D., Chair  
Alan Kehew, Ph.D.  
Richard Becker, Ph.D.

Copyright by  
Hossein Sahour  
2020

# STATISTICAL DOWNSCALING TECHNIQUES TO ENHANCE THE SPATIAL RESOLUTION OF THE GRACE SATELLITE DATA AND TO FILL TEMPORAL GAPS

Hossein Sahour, Ph.D.

Western Michigan University, 2020

The Gravity Recovery and Climate Experiment (GRACE) has been successfully used to monitor variations in terrestrial water storage ( $GRACE_{TWS}$ ) and groundwater storage ( $GRACE_{GWS}$ ) across the globe, yet such applications are hindered on local scales by the limited spatial resolution of GRACE data. Using the Lower Peninsula of Michigan as a test site, I developed optimum procedures to downscale GRACE Release-06 monthly mascon solutions. A four-fold exercise is conducted. Cluster analysis is performed to identify the optimum number and distribution of clusters (areas) of contiguous pixels of similar geophysical signals ( $GRACE_{TWS}$  time series); three clusters are identified (cluster 1: 13,700 km<sup>2</sup>; cluster 2: 59,200 km<sup>2</sup>; cluster 3: 33,100 km<sup>2</sup>; Step I). Variables (total precipitation, normalized difference vegetation index [NDVI], snow cover, Lake Michigan level, Lake Huron level, land surface temperature, soil moisture, air temperature, and evapotranspiration [ET]) which could potentially contribute to, or correlate with,  $GRACE_{TWS}$  over the test site are identified, and the dataset are randomly partitioned into training (80%) and testing (20%) datasets (Step II). Multivariate regression, artificial neural network, and extreme gradient boosting techniques are applied on the training dataset for each of the identified clusters to extract relationships between the identified hydro-climatic variables and  $GRACE_{TWS}$  solutions on a coarser scale (13,700 – 33,100 km<sup>2</sup>) and are used to estimate  $GRACE_{TWS}$  at a spatial resolution matching that of the fine-scale ( $0.125^\circ \times 0.125^\circ$  or 120 km<sup>2</sup>) inputs. The statistical models were evaluated by comparing the observed and modeled  $GRACE_{TWS}$  values using the R-squared, the Nash-Sutcliffe model efficiency coefficient (NSE), and the normalized root-mean-square error (NRMSE; Step III). Lastly, temporal variations in  $GRACE_{GWS}$  are extracted

using outputs of land surface models and those of the optimum downscaling methodology (downscaled GRACE<sub>TWS</sub>) (Step IV). Findings demonstrate that (1) consideration should be given to the cluster-based extreme gradient boosting technique in downscaling GRACE<sub>TWS</sub> for local applications given their apparent enhanced performance (average value: R-squared: 0.81; NRMSE 0.46; NSE 0.77) over the multivariate regression (R-squared: 0.69; NRMSE 0.60; NSE 0.62) and artificial neural network (R-squared: 0.72; NRMSE 0.53; NSE 0.70) methods; and (2) identifying local hydrologic variables and the optimum downscaling approach for individual clusters is critical to implementing this method. The adopted method could potentially be used for groundwater management purposes on local scales in the study area and in similar settings elsewhere.

## ACKNOWLEDGMENTS

I would like to express my deepest appreciation and sincere gratitude to my advisor Dr. Mohamed Sultan, for his advice, encouragement, and his unconditional help and support during my academic journey as a Ph.D. student. I would also like to express my appreciation to my committee members, Dr. Alan Kehew, from the Geological and Environmental Science Department of Western Michigan University, and Dr. Richard Becker from the University of Toledo for their constructive comments and review.

Additionally, I would like to thank the institutions, administrations, and organizations that provided financial assistance for this research. In particular, I would like to thank NASA Earth Science Division (grants numbers NNX12AJ94G and 80NSSC18K1681), the Earth Remote Sensing Laboratory of the Western Michigan University, and the Michigan Geological Survey.

Finally, I would like to thank my parents for their unconditional love and support throughout my life and for their encouragement to pursue my academic and scientific ambitions.

Hossein Sahour

## TABLE OF CONTENTS

ACKNOWLEDGMENTS .....	ii
LIST OF TABLES .....	v
LIST OF FIGURES .....	vi
CHAPTER	
1. INTRODUCTION .....	1
1.1 Background .....	1
1.2 Objectives .....	5
2. OVERVIEW OF THE STUDY AREA .....	7
2.1 Lower Peninsula of Michigan .....	7
3. METHODOLOGY .....	11
3.1 Introduction .....	11
3.2 Cluster Analysis .....	13
3.3 Identification of Variables that Correlate with and/or Control GRACE <sub>TWS</sub> .....	14
3.4 Construction, Evaluation, and Selection of Optimum Model for Downscaling .....	21
3.5 Extraction of Temporal GW Storage Using Outputs of Land Surface Models .....	29
3.6 Sources and Propagation of Errors .....	31
4. RESULTS .....	33
4.1 Cluster Analysis .....	33
4.2 Evaluation and Comparison of the Models .....	35

## Table of Contents - Continued

4.3 Factors Controlling the TWS and GWS Variations over the Study Area .....	38
5. DISCUSSION AND IMPLICATIONS .....	42
5.1 Introduction .....	42
5.2 Inspection of Trends .....	42
6. SUMMARY AND CONCLUSION .....	50
REFERENCES .....	53
APPENDIX .....	65



## LIST OF TABLES

1.	Initial input variables for the statistical models.....	15
2.	Secular trends for GRACE <sub>TWS</sub> and GRACE <sub>GWS</sub> from 2002 to 2016.....	18
3.	Performance of the applied models.....	22
4.	Correlation matrix for GRACE <sub>TWS</sub> values over land (three clusters) and Lake Michigan water levels.....	34
5.	Statistical coefficients (R-squared, NRMSE, and NSE) for each of the examined models (Extreme Gradient Boosting, Multivariate Regression, and Artificial Neural Network) over clusters 1, 2, and 3 and calculated uncertainties.....	35
6.	Percent contribution of each variable in the outputs of the XGBoost models and their optimum lag times.....	39

## LIST OF FIGURES

1. Location map showing the distribution of the three clusters (1, 2, and 3), the Holland and Harbor Beach lake level measuring stations, stream gauges, and monitoring.....	9
2. Estimated yield in glacial deposits in gallons/minute (gpm) and hydrological provinces (1 through 9) of the LP.....	10
3. Flow chart showing the three main steps that were used to downscale $GRACE_{TWS}$ from $12,000\text{ km}^2$ to $120\text{ km}^2$ and to extract fine-resolution ( $120\text{ km}^2$ ) $GRACE_{GWS}$ .....	12
4. Configurations of the GRACE satellite mission.....	16
5. Schematic diagram of tree-based gradient boosting method.....	27
6. Secular $GRACE_{TWS}$ and $GRACE_{GWS}$ trend images (mm/year) over the LP for the period 2002 to 2016.....	30
7. Comparison of the time series of $GRACE_{TWS}$ over land (cluster 3) and Michigan lake level.....	34
8. Scatter plots of the $GRACE_{TWS}$ predicted values (using XGBoost models) versus observed values (from testing subsets) $GRACE_{TWS}$ values for the three clusters.....	37
9. The estimated total error rate for three arbitrary pixels (1, 2, and 3).....	38
10. $GRACE_{TWS}$ and $GRACE_{GWS}$ trend images and time series for downscaled data.....	43
11. Average annual rainfall and snow water equivalent for periods I (2002-2012) and II (2013-2016).....	46
12. Comparison between the downscaled $GRACE_{GWS}$ data for three pixels and groundwater levels from monitoring wells within each of the three GRACE pixels in Kalamazoo (well A) and Lansing (wells B and C).....	48
13. Comparison between the downscaled $GRACE_{TWS}$ data for two pixels and two inland lakes, namely Otsego Lake and Austin Lake .....	49

## CHAPTER 1

### INTRODUCTION

#### 1.1 Background

The Gravity Recovery and Climate Experiment (GRACE) is a satellite mission that was jointly implemented by the National Aeronautics and Space Administration (NASA) in the United States and the Deutschen Zentrum für Luftund Raumfahrt (DLR) in Germany to map the temporal variations in the global gravity field (Tapley et al, 2004). The GRACE satellites were launched in March 2002, and the GRACE Follow-On (GRACE-FO) mission was launched in May 2018. Since then, the application of GRACE has resulted in advances of hydrologic sciences in the assessment and monitoring of spatial and temporal variations in groundwater storage (GWS) in many parts of the world including Africa (Ahmed et al., 2014; Sultan et al., 2019); the Middle East (Othman et al., 2018; Sultan et al., 2019) China (Feng et al., 2013), India (Rodell et al., 2009), California (Scanlon et al., 2012), and Mexico (Castellazzi et al., 2018). However, such applications are hampered by the relatively low horizontal resolution of GRACE data and the fact that GRACE does not have vertical resolution (wahr et al, 2004). In other words, GRACE cannot determine in which compartment (e.g., surface water, groundwater, or soil moisture) the observed mass variations are occurring.

Many studies utilizing GRACE data for hydrological research and applications (e.g., Rodell et al., 2004; Chen et al., 2005) target large aquifers and watersheds (areas of  $450 \times 10^3$  to  $6 \times 10^6$  km<sup>2</sup>). However, the majority of the world's aquifers and watersheds are much smaller; even for the larger ones, one often needs to understand

the partitioning of water on the sub-basin level. A finer resolution of GRACE solutions would be a useful tool for tracking the changes in GWS ( $\text{GRACE}_{\text{GWS}}$ ) on local scales, especially for regions that do not have sufficient in-situ monitoring sites.

Downscaling techniques allow predictions to be made at a finer spatial resolution than that of the original dataset (Atkinson, 2013). Downscaling approaches, especially those developed for climate models and later applied to remotely acquired data, can be classified into two main groups: dynamic downscaling and statistical downscaling (Atkinson, 2013). The former approach has been successfully applied to downscale global climate models (GCMs) over regions of interest by integrating GCM outputs with the physical characteristics of Earth's surface in the area of interest. Monthly GRACE terrestrial water storage ( $\text{GRACE}_{\text{TWS}}$ ) solutions from Center for Space Research (CSR), Jet Propulsion Laboratory (JPL), and Deutsches GeoForschungsZentrum (GFZ) (spatial resolution: 150,000 km<sup>2</sup>) were assimilated into a land surface model to generate high-resolution water storage changes within the major watersheds of the Mississippi River (Zaitchik et al., 2008).  $\text{GRACE}_{\text{TWS}}$  derived from spherical harmonic (SH) solutions were assimilated into the Catchment Land Surface Model to extract GRACE-based drought indicators (spatial resolution: 1° × 1.25°) for North America (Houborg et al., 2012). Gridded (25 km<sup>2</sup>) Advanced Microwave Scanning Radiometer—Earth Observing System (AMSR-E) data were assimilated into a fine-scale (1 km<sup>2</sup>) NOAH land surface model using three-dimensional and one-dimensional Kalman filters (Sahoo et al., 2013). The JPL  $\text{GRACE}_{\text{TWS}}$  mascon solutions (3.0° × 3.0°) were assimilated into the fine resolution (0.05° × 0.05°) hydrologic models (Shokri et al., 2019). The scale factor was used to minimize the leakage errors and to improve the spatial resolution of the JPL spherical

harmonics solutions (Landerer and Swenson, 2012). Such applications often require extensive computing time and resources that are not available for many researchers (Schoof, 2013a). Also, many of the above-mentioned procedures depend on the selected hydrological model, some of which are lacking surface or groundwater components.

Statistical downscaling, on the other hand, does not require these resources. Statistical downscaling evaluates observed spatial and temporal relationships between inputs (independent variables) and outputs (dependent variables) using coarse-scale datasets (inputs and outputs) and applies the extracted relationships to produce the dependent variables at a spatial resolution matching that of the fine-scale inputs (Le Roux et al., 2018). A variety of statistical methods have been applied to downscale remote sensing or ground-based data, including Markov chains and support-vector machines (Hou et al., 2017), regression-kriging (Wang et al., 2018), neural networks (Miro and Famiglietti, 2018), and stochastic models (So et al., 2017). Stepwise regression was successfully applied to downscale satellite-based precipitation data (TRMM3B43 products) and average daily precipitation and air temperature data from weather stations (Joshi et al., 2015; Ezzine et al., 2017). Artificial neural networks (ANN) were used to downscale GCM outputs (Chadwick et al., 2011) and rainfall (Vu et al., 2016). The major limitation of the statistical approaches comes from the assumption of stationarity between the coarse- and fine-scale dynamics and from the uncertainty and probability associated with this assumption (Schoof, 2013b).

Statistical approaches were successfully used to downscale GRACE data to a high-resolution ( $\sim 16 \text{ km}^2$ ) dataset of groundwater storage changes over a portion of California's Central Valley using ANN techniques (Miro and Famiglietti, 2018). In that

study, temporal  $GRACE_{TWS}$  and a series of widely available hydrologic variables were used as model inputs, and target data were extracted from groundwater storage changes that were estimated from an extensive well network dataset (2189 wells). Similarly, variations in GWS were extracted and tested against temporal variations in groundwater levels in Texas, Nebraska, and Illinois (Sun, 2013) by using coarse-resolution ( $3^\circ \times 3^\circ$ )  $GRACE_{TWS}$  JPL Release-05 monthly mass concentration (mascons; JPL RL-05M) and high-resolution hydrologic variables and applying ANN techniques. Statistical downscaling was also used to successfully downscale GWS anomalies from 110 to 2 km in the North China Plain on both interannual and monthly scales in areas where a strong relationship between GWS and ET was detected and where the relationship was established under different spatial resolutions (Hu et al., 2018). The successful applications in the first two examples (in California's Central Valley, Texas, Nebraska, and Illinois) relies heavily on the availability of temporal head data from dense networks of wells, and for the latter (North China Plain) on the presence of a strong relationship between GWS and ET—conditions that are not necessarily available in many of the basins worldwide. In a fourth study, temporal  $GRACE_{TWS}$  solutions and land surface and hydro-climatic variables were used to predict groundwater level anomaly (GWLA). A network of 32 wells (21 wells for training and 11 wells for testing) was used to establish and test the relationship between  $GRACE_{TWS}$  and hydro-climatic variables as input and GWLA as the response variable using a downscaling algorithm based on machine learning (ML) (Seyoum et al., 2019). In many of the applied statistical downscaling approaches, including the latter study, a dense network of wells is required to establish a relationship between hydrological variables and groundwater anomalies extracted from the well data. Those methods cannot be

applied in many parts of the world with limited monitoring well sites.

In this study, I applied statistical techniques to extract relationships between coarse-resolution GRACE solutions (target data) and hydrologic variables (total precipitation, normalized difference vegetation index [NDVI], snow cover, Lake Michigan level, Lake Huron level, land surface temperature, soil moisture, air temperature, and evapotranspiration [ET]). These variables could potentially correlate with, or contribute to, the temporal variations of GRACE<sub>TWS</sub>. I used those relationships, and high-resolution hydrologic variables to generate high-resolution, modeled monthly GRACE solutions. The Lower Peninsula (LP) of Michigan was used as a test site. I applied, and compared the findings from three statistical methods—stepwise multivariate regression (MR) models, ANN, and extreme gradient boosting (XGBoost)—to downscale GRACE data and to fill the temporal gaps in the time series data over the LP throughout the investigated period (2002–2016).

## 1.2 Objectives

There are two major shortcomings associated with GRACE-derived products. First, the coarse spatial resolution of the GRACE-derived product that makes them difficult to use for local investigations, and second, the gaps in the time series resulted from battery failure during the first GRACE mission. The main objective of this study was to address the above-mentioned shortcomings using statistical downscaling techniques and extracting information from newly produced GRACE-derived data. The objectives of the study can be summarized as follows.

- To enhance the spatial resolution of the GRACE<sub>TWS</sub> from 12,000 km<sup>2</sup> to 120 km<sup>2</sup>

- To fill the temporal gaps in GRACE-derived TWS
- To extract the trend in TWS using downscaled  $\text{GRACE}_{\text{TWS}}$  in the study period
- To extract the trend in GWS using downscaled  $\text{GRACE}_{\text{TWS}}$  and outputs of land surface models.



## CHAPTER 2

### OVERVIEW OF THE STUDY AREA

#### 2.1 Lower Peninsula of Michigan

The LP (Fig. 1) depends heavily on its groundwater resources to support a population of 10 million citizens, its agricultural sector (35% of its area is agricultural land (Tayyebi et al., 2017), and its industry. The area in general, and the local communities, in particular, can benefit from reliable datasets that show spatial and temporal variations in GWS at the local scale. This is especially true for the southern counties of the LP where intensive agricultural activities are established, and groundwater withdrawal rates (30 to 90 million gallons/day (Grannemann et al., 2000) are the highest in the state.

Groundwater availability differs from one place to another in the LP; it is plentiful in some regions (e.g., the southwest) and less so in other areas (e.g., the southeast) (Rheaume, 1991). The major aquifers in Michigan are largely found in (1) glacial deposits, where yields are mostly from outwash and glacio-fluvial deposits, and (2) sedimentary bedrock units, where yields come largely from the Mississippian and Pennsylvanian rocks (Rheaume, 1991). The surface and groundwater in the area drain into the Great Lakes (Lake Michigan, Lake Huron, and Lake Erie) (Vugrinovich, 1986).

Nine hydrologic provinces with varying sedimentary deposits and aquifer thicknesses (Fig. 2) were identified in the LP (Rheaume, 1991). Province 1 has a relatively thin (1 to 60 m) glacial-lacustrine sand unit that overlies Silurian and Devonian limestone and dolomite. Aquifers in Silurian and Devonian rocks are the

main source of groundwater in the area. Province 2 is largely covered by thick (>300 m in some areas), coarse-grained, sandy glacial deposits, whereas Province 3 is characterized by variable thicknesses (15 to 75 m), low-yield lacustrine deposits that overlie the Mississippian bedrocks. Province 4 is in the southern sections of the LP and is characterized by thick (30 to 180 m), coarse-grained glacial deposits that overlie the low-yield Mississippian rocks; it is the main source for groundwater in the area. Province 5 glacial deposits vary in thickness (7 to 150 m) and overlie the high-yield Pennsylvanian aquifers. In this province, the thickness of the glacial aquifer thins to the south and groundwater is largely extracted from the Pennsylvanian sandstone. Province 6 is characterized by thin to moderate glacial drift that overlies high-yield Mississippian bedrock aquifers. The latter is largely composed of the Marshall Sandstone, whereas the glacial deposits are absent in some areas but thicken (up to 120 m) in others. Province 7 is characterized by thin glacial deposits (<10 m in most areas) that overlie moderate-yield Silurian and Devonian limestone and dolomite intercalated with sand and shale layers. Province 8 is characterized by low-yield, moderate to thick (15 to 120 m), lacustrine clay deposits that overlie low-yield Devonian and Mississippian sandstone. Lastly, Province 9 consists of featureless lacustrine and low-yield glacial sand and clay deposits of variable thickness (7 to 90 m) that overlie Pennsylvanian and Mississippian sandstone aquifers (Rheume, 1991). In general, the vertical hydraulic conductivity of the glacial aquifer is high ( $9.64 \times 10^{-7}$  to  $3.8 \times 10^{-5}$  m/day) in the southern sections compared to the northern sections ( $3.8 \times 10^{-5}$  to 0.45 m/day).

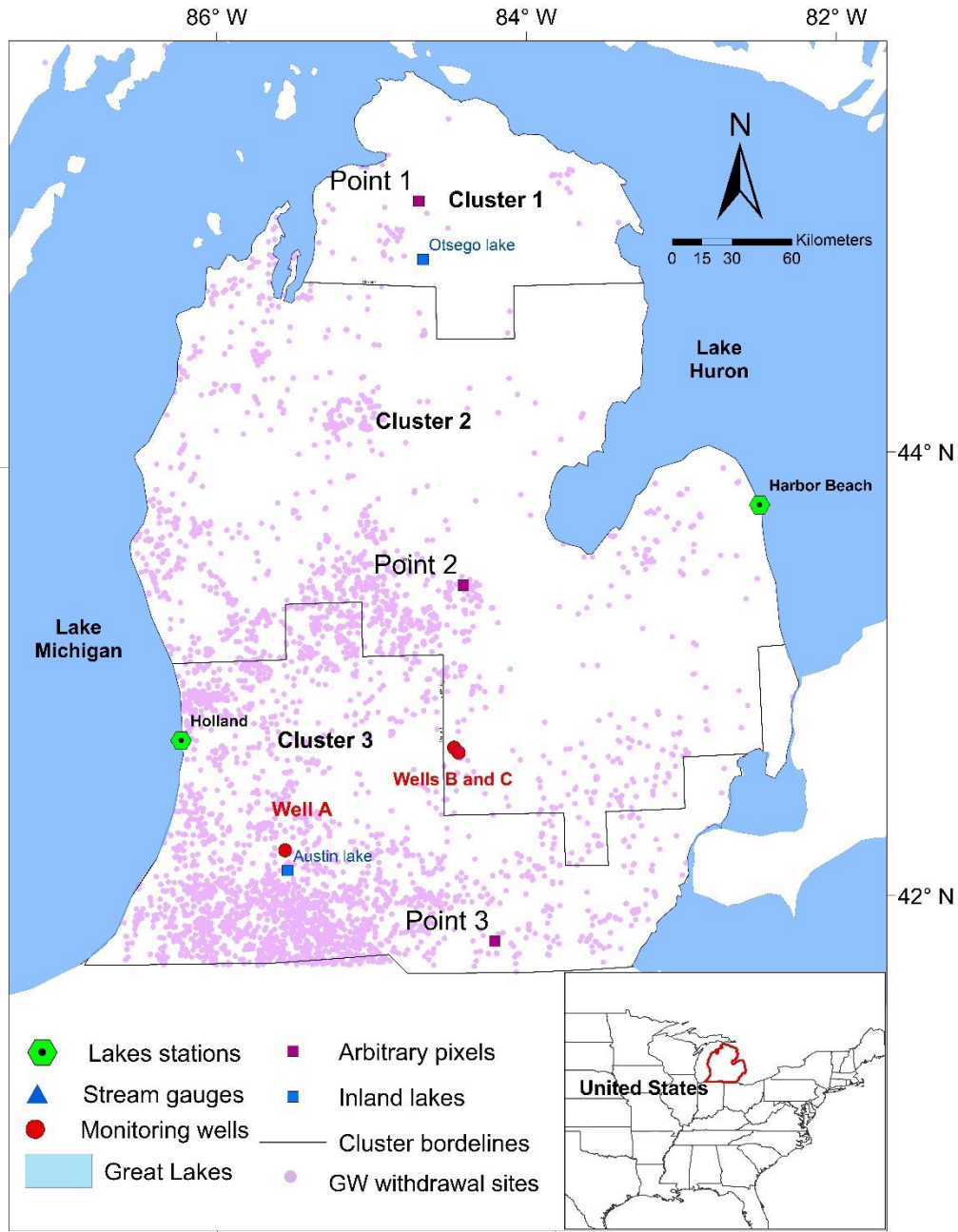


Figure 1. Location map showing the distribution of the three clusters (1, 2, and 3), the Holland and Harbor Beach lake level measuring stations, and monitoring wells in Kalamazoo (well A: site name 02S 11W 22CDBB 01) and Lansing (well B: site name 04N 02W 26BBDB 01 and well C: site name: 04N 02W 16DAAA 01). Inset shows location of study area in the USA.

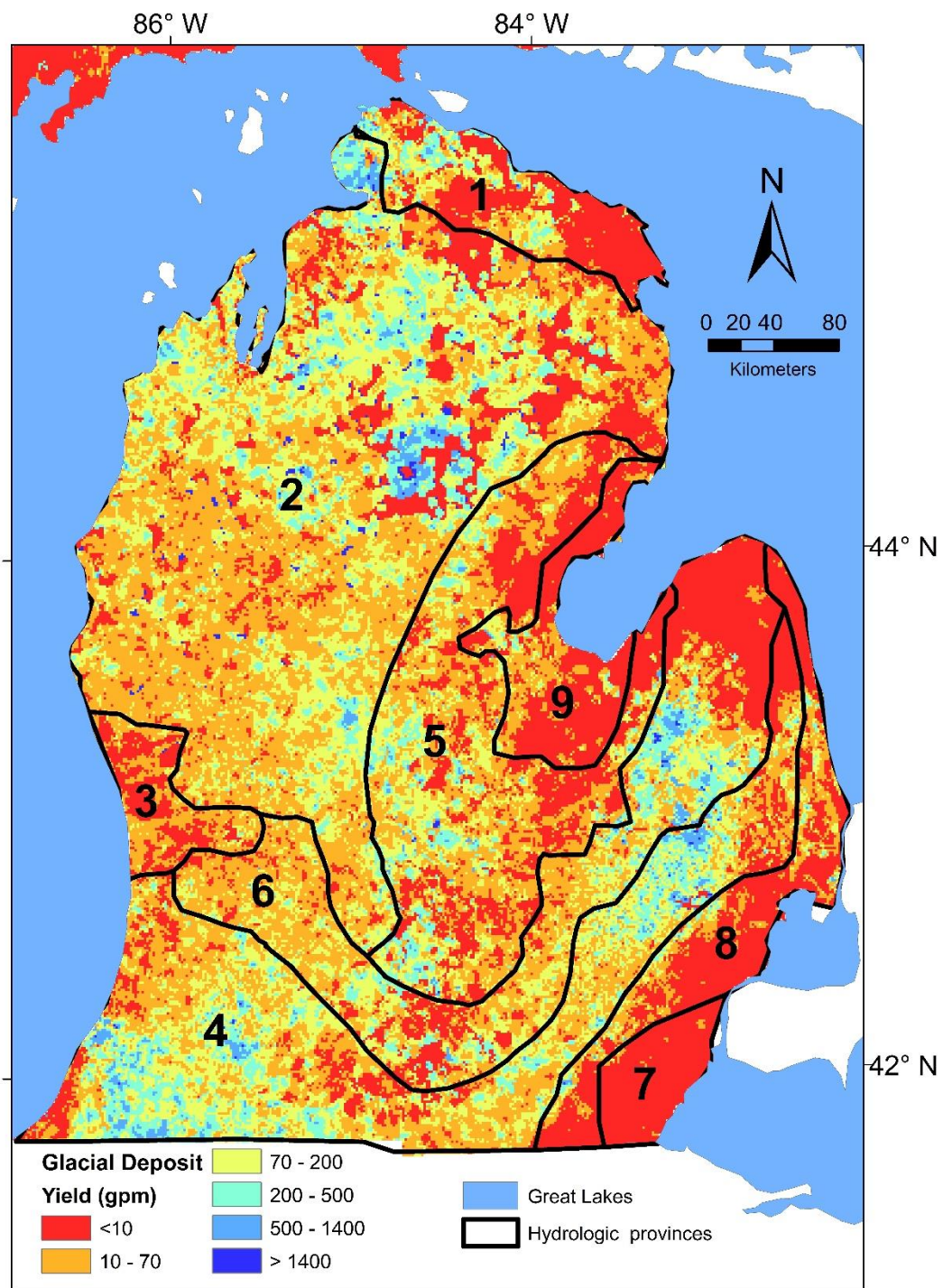


Figure 2. Estimated yield in glacial deposits in gallons/minute (gpm) and hydrological provinces (1 through 9) of the LP modified from (Vugrinovich, 1986).

## CHAPTER 3

### METHODOLOGY

#### 3.1 Introduction

A four-fold exercise was conducted to accomplish the statistical downscaling of  $GRACE_{TWS}$  (Fig. 3). First, clustering analysis was conducted on  $GRACE_{TWS}$  data over the LP to identify clusters of pixels, where pixels within each cluster have similar  $GRACE_{TWS}$  values yet are statistically different from those of neighboring clusters (Step I). Variables that correlate with and/or control  $GRACE_{TWS}$  were then identified (Step II). In Step III, for each cluster, relationships between coarse-scale inputs (independent variables) and outputs ( $GRACE_{TWS}$ ; dependent variables or target) were extracted using three statistical approaches—MR, ANN, and XGBoost—to produce the dependent variables at a spatial resolution matching that of the fine-scale inputs ( $0.125^\circ \times 0.125^\circ$  or  $120 \text{ km}^2$ ). In doing so, we assumed that the extracted statistical relationship between the input variables and the target applies at the finer scales as well. In this step, the adopted statistical models were compared and evaluated to select the optimum methodology. Finally, GWS variations were extracted using outputs of the applied land surface model and outputs of the optimum downscaling method (downscaled  $GRACE_{TWS}$ ) (Step IV).

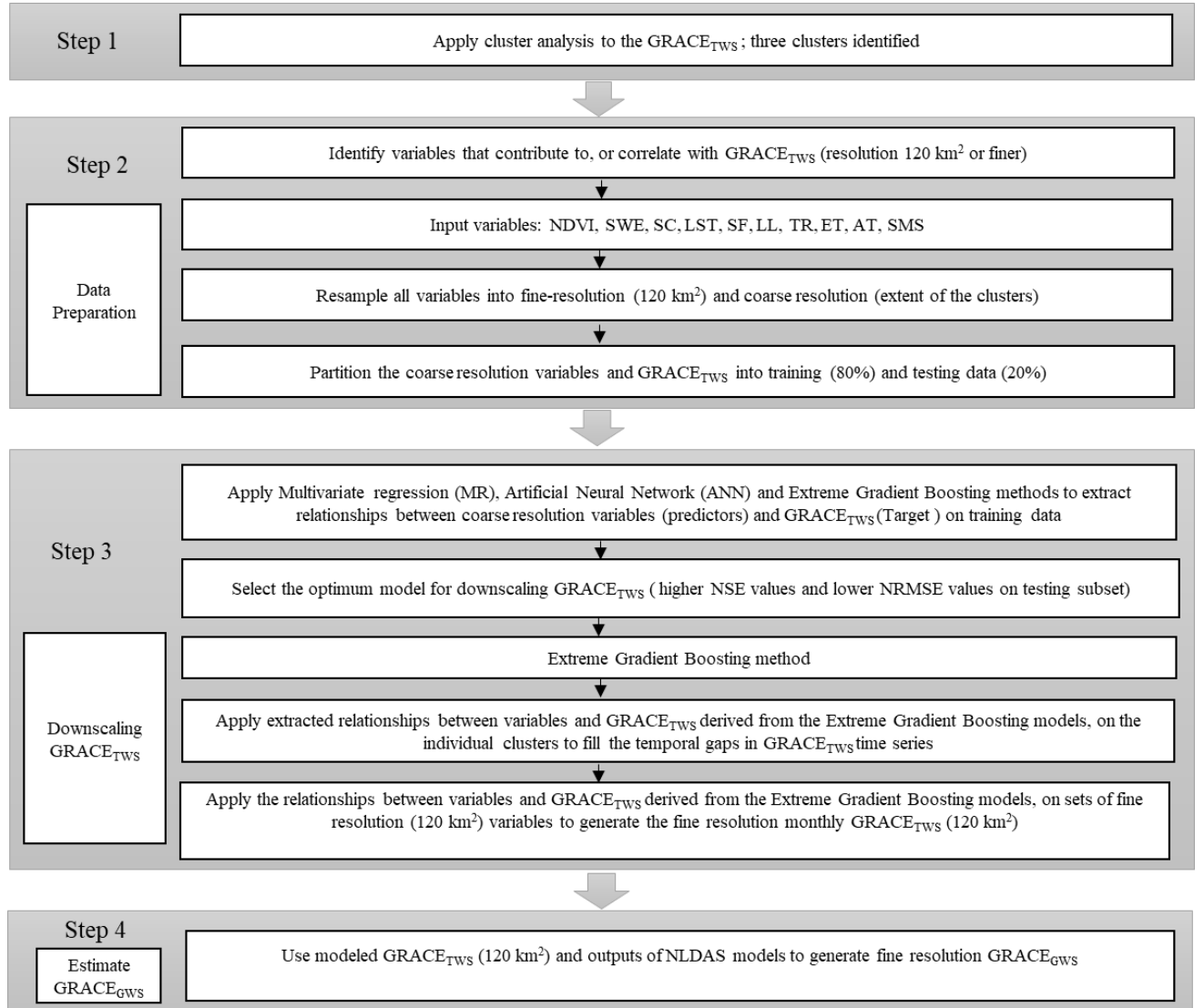


Figure 3. Flow chart showing the three main steps that were used to downscale  $GRACE_{TWS}$  from  $12,000\text{ km}^2$  to  $120\text{ km}^2$  and to extract fine-resolution ( $120\text{ km}^2$ )  $GRACE_{GWS}$ .

Initially, we selected 11 variables that could correlate with, or contribute to, the temporal variations of  $GRACE_{TWS}$ . All input variables listed below (section 3.2) are available at a spatial resolution ranging from  $0.05^\circ \times 0.05^\circ$  to  $0.125^\circ \times 0.125^\circ$ , and the output target values for each of the identified clusters were calculated from the gridded values within each of the clusters. Input variables were resampled to the size of their corresponding clusters for the step I and resampled to  $0.125^\circ \times 0.125^\circ$  ( $120\text{ km}^2$ ) for

the step III.

We adopted three statistical approaches to model the relationships between the 11 variables and  $GRACE_{TWS}$ , namely, the MR, ANN, and XGboost approaches for each cluster. For each of the three approaches, the data were randomly partitioned into two subsets, training and testing. The training subset comprised 80% of the data points (percentage of the months in the time series data) and was used to construct the model, whereas the remaining 20% were used to evaluate the performance of the model. This approach was applied to each of the identified clusters.

### 3.2 Cluster Analysis

Working with small areas (individual pixels) introduces significant leakages from neighboring pixels. I adopted cluster analysis to identify larger areas (contiguous pixels) with similar geophysical signatures ( $GRACE_{TWS}$  time series) and hence reduce leakage errors. Clustering is the partitioning of the set of objects into groups in such a way that the objects within a group are more similar to each other than those in other groups. I applied K-means, one of the most popular methods for clustering analysis to the monthly GRACE Release-06 (CSR RL06) solutions over the study area. In this method (K-means), the dataset is partitioned into K clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster (Dhanachandra et al., 2015) The optimum number of clusters was estimated, and the area was partitioned into the identified number of clusters. The monthly values of a variable for each cluster represents the average values of all pixels within that cluster in the investigated month.

The optimal number of clusters was determined using the gap statistic implementation (Tibshirani et al., 2001), where three steps are undertaken:

Step 1: Estimate the gap statistic using equation 1.

$$\text{Gap}(k) = \frac{1}{B} \sum_{b=1}^B (\log W_{kb} - \log W_k)$$

In which  $k$  is the number of clusters,  $B$  is the number of reference datasets generated using a uniform prescription,  $W_{kb}$  is the within-dispersion measures, and  $W_k$  is the within-cluster dispersion.

Step 2: Compute the standard deviation.

$$sd_k = \left[ \left( \frac{1}{B} \right) \sum_b \{ \log(W_{kb}) - \bar{l} \}^2 \right]^{\frac{1}{2}}$$

Where  $\bar{l} = \left( \frac{1}{B} \right) \sum_b \log(W_{kb})$ .

Step 3: Define  $s_k = sd_k \sqrt{1 + \frac{1}{B}}$  and choose the number of clusters via

$\hat{k} = \text{smallest } k \text{ such that } \text{Gap}(k) \geq \text{Gap}(k+1) - s_{k+1}$ .

### 3.3. Identification of Variables that Correlate with and/or Control GRACE<sub>TWS</sub>

In this section, I briefly describe the variables and target data for the MR, ANN, and XGboost models. I selected the input variables (total precipitation, normalized difference vegetation index [NDVI], snow cover, Lake Michigan level, Lake Huron level, land surface temperature, soil moisture, air temperature, and evapotranspiration



[ET]) based on their probable correlation and/or contribution to the target (GRACE<sub>TWS</sub>). The spatial resolution, format, and sources of these datasets are given in Table 1. A number of these variables (e.g., soil moisture, NDVI, and evapotranspiration) may be related to, or affected by, the soil characteristics and the underlying glacial aquifer parameters.

Table 1. Initial input variables for the statistical models

Variable Name	Format	Resolution	Source
NDVI	raster	(0.05° × 0.05°)	MODIS
Snow Cover	raster	(0.05° × 0.05°)	MODIS
Land surface Temperature	raster	(0.05° × 0.05°)	MODIS
Total precipitation	raster	(0.125° × 0.125°)	NLDAS
Air temperature	raster	(0.125° × 0.125°)	NLDAS
Soil moisture	raster	(0.125° × 0.125°)	NLDAS
Lakes Level	numerical	N/A	NOAA
Evapotranspiration	raster	(0.125° × 0.125°)	NLDAS

### GRACE-derived TWS

The Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) is a continuing joint satellite mission between the National Aeronautics and Space Administration (NASA) in the United States and the German Aerospace Center (DLR) in Germany. GRACE satellite comprises two parts (GRACE A and GRACE B) that rotate the Earth close to each other. The twin satellites were launched in March 2002 from Plesetsk Cosmodrome in Russia. The two satellites orbit the Earth, one in front of the other, rotating in a polar orbit with an approximate

separation of 200 km, and an inclination of  $89.5^\circ$  and an altitude of 500 km. (Tapley et al., 2004). The two satellites keep the same distance, speed, and altitude until a gravity anomaly is recognized. Mass anomalies are identified as disturbances in the distance between the two satellites that are used to measure temporal variations in the global earth gravity field (Fig. 4) (Tapley et al., 2004). The temporal variations in the gravity field within the investigated short time periods (months, years) are mostly related to mass variations in hydrologic systems ( e.g., groundwater, surface water, and oceans).

## GRACE Mission

### Science Goals

High resolution, mean & time variable gravity field mapping for Earth System Science

### Mission Systems

#### Instruments

- KBR (JPL/SSL)
- ACC (ONERA)
- SCA (DTU)
- GPS (JPL)

Satellite (JPL/DDS)

Launcher (DLR/Eurockot)

Science (CSR/JPL/GFZ)

### Orbit

Launch: March 2002

Altitude: 500 km

Inclination: 89 deg

Eccentricity: ~0.001

Lifetime: (Plan: 5 years)  
(Until 2016)

Non-Repeat Ground Track  
Earth Pointed, 3-Axis Stable

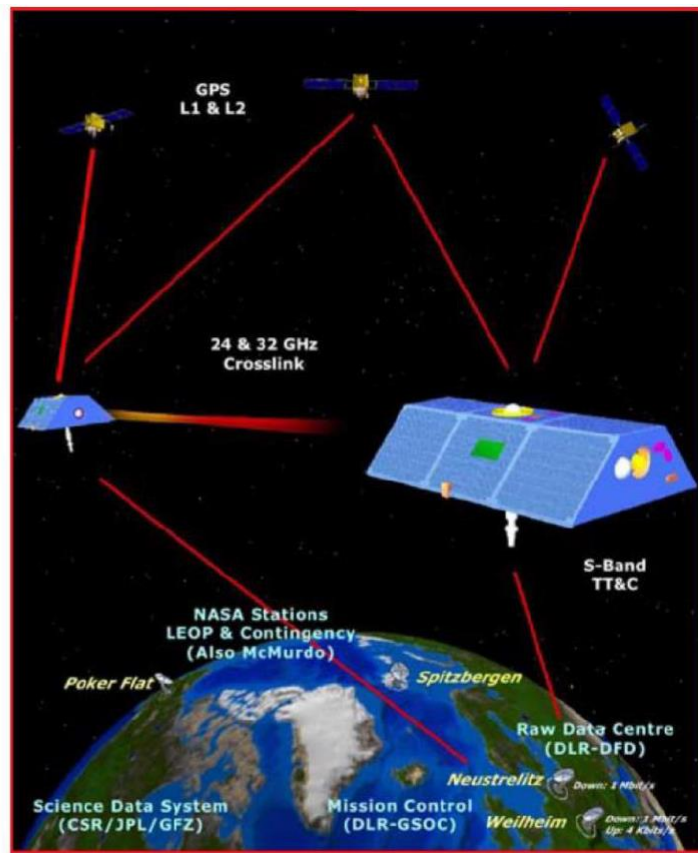


Figure 4. The configuration of the GRACE satellite mission. Source of image : <http://www.csr.utexas.edu/grace/>.

Three communal GRACE mascon solutions (from April 2002 through June

2016) were applied in this study and reported relative to a 2004–2009 mean baseline. The first is the GRACE CSR-RL06M solutions provided by the University of Texas Center for Space Research (UT-CSR); they were derived using Tikhonov regularization (Save et al, 2016), and resolved on a geodesic grid (grid size: 12,000 km<sup>2</sup>) (Save et al., 2012; Save et al., 2016). The second is the mascon solutions from the Jet Propulsion Laboratory (Watkins et al., 2015) (available at JPL website, JPL;L06M; <ftp://podaacftp.jpl.nasa.gov/allData/tellus/L3/mascon/RL06/JPL/CRI/netcdf/>) and the third is the mascon solution from NASA Goddard Space Flight Center (GSFC-Mascon; <https://earth.gsfc.nasa.gov/geo/data/grace-mascons>) (Luthcke et al., 2013).

The CSR-RL06M solutions were selected as the initial mascon solutions for extracting trends and time series over the investigated areas. The uncertainty associated with the calculated trend values were calculated from the differences in trend values extracted from the three solutions (CSR-RL06M, JPL-RL06M, and GSFC-M) ( Rodell et al., 2018; Scanlon et al., 2018) (Table 2). No post-processing and/or filtering or application of empirical scaling factors were applied (Luthcke et al., 2006; Save et al., 2016; Watkins et al., 2015). SH solutions of GRACE data have been successfully applied in many studies to monitor variations in TWS in large scales (Ahmed, Sultan, Wahr, & Yan, 2014; Ahmed et al., 2011); however, its application on local scales was hindered by its coarse spatial resolution (>125,000 km<sup>2</sup>), leakage problems from adjacent pixels, and the required complex post-processing steps (Scanlon et al., 2016). Compared to SH, the GRACE RL06 mascon solutions have a higher signal-to-noise ratio, higher spatial resolution, and reduced leakage from neighboring mascons that are in separate constraint regions (Scanlon et al., 2016). The extracted trends and associated uncertainties for each of the investigated clusters are given in Table 2.

Table 2. Secular trends for GRACE<sub>TWS</sub> and GRACE<sub>GWS</sub> from 2002 to 2016.

Clu ster	$\Delta TWS$ (mm/year)	$\Delta SMS$ (mm/year)	$\Delta SWE$ (mm/year)	$\Delta GWS$ (mm/year)
1	$16.2 \pm 5$	$0.3 \pm 0.0$	$0.1 \pm 0.0$	$15.8 \pm 5$
2	$14.4 \pm 5.2$	$0.0 \pm 0.0$	$0.7 \pm 0.0$	$13.7 \pm 5.2$
3	$8.8 \pm 3.4$	$-0.7 \pm 0.0$	$0.1 \pm 0.0$	$9.5 \pm 3.4$

Locations for clusters are shown in Figure 1.

$\Delta TWS$ : Change in terrestrial water storage. The values are based on original GRACE<sub>TWS</sub> before gap filling for missing months.

$\Delta SMS$ : Change in soil moisture storage.

$\Delta SWE$ : Change in snow water equivalent.

$\Delta GWS$ : Change in groundwater storage.

$\Delta CWS$ : Change in canopy water storage for each of the three clusters was found to be negligible (0.0) and was ignored in estimating  $\Delta GWS$ .

## NDVI

I used NDVI products derived from the Moderate-resolution Imaging Spectroradiometer (MODIS) as one of the variables. NDVI uses the red and near-infrared bands which are sensitive to healthy vegetation. The data consists of global monthly NDVI values reported at a  $0.05^\circ \times 0.05^\circ$  spatial resolution downloaded from Land Processes Distributed Active Archive Center's website ([https://lpdaac.usgs.gov/dataset\\_discovery/modis/modis\\_products\\_table](https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table)). The uncertainties associated with MODIS NDVI products was estimated by comparing the NDVI products with that extracted from the Advanced Very-High-Resolution Radiometer (AVHRR) and from the Visible Infrared Imaging Radiometer Suite (VIRS) (Van Leeuwen and Herrmann, 2006). The reported statistical coefficients between the NDVI products of MODIS and each of the AVHRR (R-square 0.99) and VIRS (R-square 0.99) indicate high consistency of the NDVI values extracted from different

sensors (Van Leeuwen et al., 2006).

#### Snow Cover (SC)

The monthly average snow cover (spatial resolution:  $0.05^{\circ} \times 0.05^{\circ}$ ) values were computed from daily snow cover observations extracted from the MODIS/Terra Snow Cover Daily L3 Global 0.05Deg Climate Modeling Grid dataset. The normalized difference snow index—an index that is sensitive to the high reflectance over snow covered lands in the visible wavelength region and low reflectance in the shortwave infrared regions—was used to identify snow-covered land. The monthly averages are calculated from the corresponding 28 to 31 days of observations in the daily maximum snow cover extent data. The MODIS monthly snow cover data were downloaded from NSIDC's website (<https://nsidc.org/data/MOD10CM/versions/6>).

#### Lake Levels (LL)

Average monthly water level data for Lake Michigan and Lake Huron were obtained from the NOAA's Center for Operational Oceanographic Products and Services. The Lake Michigan water levels were obtained from the Holland station and the Lake Huron levels from the Harbor Beach station. The former station is located along Lake Michigan's eastern shoreline and the latter along Lake Huron's southwestern shoreline (Fig. 1). At both stations, water levels are measured every six minutes. The average monthly water level data were downloaded from NOAA's website (<https://tidesandcurrents.noaa.gov/stations.html?type=Water+Levels>).

#### Land Surface Temperature (LST)

The monthly MODIS/Terra land surface temperature (LST) is derived by averaging the daily values of MOD11C3 products. These products have been validated

through a series of field studies by the MODIS land team [(“MODIS/Aqua Land-Surface Temperature/Emissivity Monthly Global 0.05Deg CMG - LAADS DAAC,” n.d.)]. The spatial resolution of the monthly LST is  $0.05^{\circ} \times 0.05^{\circ}$ , and the data were obtained from the NASA's Earth Science Data Systems available at <https://earthdata.nasa.gov/>. The performance of the MODIS LST product was evaluated in several locations in the USA; results showed a good correspondence (absolute biases  $< 0.8^{\circ}\text{C}$  and RMSEs  $< \text{than } 1.7^{\circ}\text{C}$ ) between the in-situ measurements and MODIS LST products (Wang and Meyers, 2008).

Rainfall, Snow Water Equivalent, Soil Moisture, Air Temperature, and Evapotranspiration

The NASA's North American Land Data Assimilation System (NLDAS) NOAH model outputs (total monthly rainfall, snow water equivalent [SWE], soil moisture [SMS], air temperature [AT], and evapotranspiration [ET]; spatial resolution:  $0.125^{\circ} \times 0.125^{\circ}$ ) were used as inputs to our models. These data are produced from a daily ground-based precipitation analysis, bias-corrected shortwave radiation, and surface meteorology re-analyses to drive land surface models (Mitchell et al., 2004). The average monthly data used in this research were downloaded from NASA Goddard Earth Sciences Data and Information Services Center's website at <https://disc.gsfc.nasa.gov/datasets?keywords=NLDAS>. The monthly total precipitation was computed from the sum of the monthly SWE and rainfall. The outputs of NLDAS products have been validated and enhanced through several research works (Xu et al., 2019; Xia, et al., 2012). The accuracy of the NLDAS products varies from one product to another and from one location to another. For instance, the uncertainty for ET

products was reported to be 4-8 mm per month over the contiguous United States (Xu et al., 2019). Comparing the NLDAS products with in-situ measurements showed that the performance of the soil moisture product over the contiguous United States was very high (RMSE= 0.02 to 0.11) (Xia et al., 2014). The uncertainty of the NLDAS snow water equivalent is less than 20% based on comparisons with IMS (MultiSensor Snow and Ice Mapping System) observations (Xia et al., 2012). The uncertainties in air temperature over Michigan were estimated by comparing the monthly mean values between NLDAS and the National Centers for Environmental Prediction [NCEP] products. The reported difference was  $< 0.4^{\circ}\text{C}$  (Mo et al., 2012). The NLDAS precipitation data were compared with five other available datasets over the western United States, and the mean relative difference between them ranged from 11 to 18% (Henn et al, 2018).

#### 3.4. Construction, Evaluation, and Selection of Optimum Model for Downscaling

For each cluster, all datasets (variables and targets) were randomly partitioned into two groups: training (80% of the time series for each cluster) and testing (20% of the time series for each cluster). I constructed and applied three statistical models (MR, ANN, and XGBoost) to establish the relationships between the variables (predictors) and GRACE (the target) for each of the investigated clusters. The performance of each model was compared and evaluated. The evaluation of the models was carried out by comparison between observed values (testing subset) and predicted values using the coefficient of determination (R-squared), the normalized root-mean-square error (NRMSE), and the Nash-Sutcliffe model efficiency coefficient (NSE) (Table 3) (Nash and Sutcliffe, 1970). The R-squared values range from 0 to 1; those for the NRMSE

and NSE indices range from 0 to 1. The predictive power of the models increases with increasing R-squared and NSE values and with decreasing NRMSE values. The rate of the performance for each approach was based on classifications adopted by (Moriassi et al., 2007). Following the identification of the statistical model that yielded the highest performance, I used the selected model to downscale the GRACE solutions for each of the clusters to  $0.125^\circ \times 0.125^\circ$  throughout the investigated period.

Table 3. Performance of the applied models modified from ( Moriassi et al., 2007).

Performance Rating	NSE	NRMSE
Very Good	$NSE \geq 0.75$	$NRMSE \leq 0.5$
Good	$NSE \geq 0.65$ and $< 0.75$	$NRMSE > 0.50$ and $\leq 0.60$
Satisfactory	$NSE \geq 0.50$ and $< 0.65$	$NRMSE > 0.60$ and $\leq 0.70$
Unsatisfactory	$NSE < 0.5$	$NRMSE > 0.70$

### MR Models

The MR, or multiple linear regression method, derives patterns in the data and establishes the best fitting multivariate linear relationships between two or more independent variables and the target ( $GRACE_{TWS}$ ). As described earlier, I applied a stepwise MR method in which the selection of variables is carried out by addition to, or subtraction from, a set of dependent variables using some pre-specified coefficients such as the F-test, the t-test, and the coefficient of partial determination. In an MR model, every value of the independent variable  $X$  is associated with a value of the target variable  $Y$ . The regression line for  $n$  independent variables  $X_1, X_2, \dots, X_n$  can be explained as follows.



$$Y = B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_n \quad ,$$

Where  $Y$  is the predicted value of the target variable,  $B_0$  is the value of  $Y$  when all of the independent variables are equal to zero,  $X_1$  through  $X_n$  are independent variables, and  $B_1$  through  $B_n$  are the estimated regression coefficients. In multivariate linear regression, the response variable  $Y$  (GRACE<sub>TWS</sub> in our work) is assumed to be linearly related to a set of  $n$  explanatory independent variables  $X_1, X_2, \dots, X_n$ , and the independent variables are not highly correlated with each other. Observations are selected independently and randomly from the population. Also, residuals assumed to be normally distributed with a mean value of 0.

The parameters are trained in such a way to achieve the highest similarity between the modeled and observed values in the training data set. One optimization model is employed to minimize the sum of the squares of the vertical deviations from each data point to the regression equation. The ideal case is a model in which a data point lies completely on the fitted line (i.e., vertical deviation = zero).

I applied a stepwise multivariate regression approach. Stepwise regression fits the multivariate regression several times, each time removing the least correlated variable until the statistically significant variables are left. For a full description of variable selection in the stepwise method, see (Hocking, 1976). MR models were first developed for each of the identified clusters to establish linear relationships between coarse-resolution inputs (variables) and target (GRACE<sub>TWS</sub>) values for each of the identified clusters.

All input variables were available in both coarse and fine resolutions, whereas GRACE<sub>TWS</sub> values were available in coarse resolution only. The variables were resampled to the size of each cluster using bilinear resampling techniques.

## ANN

The ANN method establishes empirical, possibly non-linear, relationships between a set of “input” variables and corresponding “target” variables. An ANN is based on a series of connected units or neurons which are intended to replicate the functions of neurons in animal or human brains; they pass information between one another, a structure that enables ANNs to be trained and learn. The ANN method used in this study is known as a multilayer perceptron (MLP). An MLP consists of units called perceptrons. Perceptrons have one or more inputs, an activation function, and an output. An MLP model is built up by combining perceptrons in structured layers. The perceptrons in a given layer are independent of each other, but each connects to all of the perceptrons in the following layer. Each layer is composed of a set of neurons and is trained with a backpropagation algorithm.

Backpropagation is one of the most extensively used algorithms for supervised training of multilayered neural networks (Mohaghegi et al, 2005; Sahour et al, 2020; Alshehri et al, 2020). Backpropagation works by approximating the non-linear relationship between the input and the target by altering the weight values internally. The processes of the Backpropagation can be divided into two stages: feedforward and Backpropagation. In the feedforward step, a pattern is applied to the input layer, and its effect propagates, layer by layer, through the network until the output is generated. The network's sample output value is then compared to the anticipated value for a given

input, and an error signal is estimated for each of the output neurons. Since all neurons within the hidden layer contributed to the signal errors in the output layer, the output errors are transmitted backward from the output layer to each neuron within the hidden layer that contributed to the output layer. This process is then reiterated, layer by layer, until each neuron in the network has received an error signal that represents its relative contribution to the total error. Once the error signal for each neuron has been computed, the errors are then applied by the neuron to adjust the values for each connection weights. The goal is to minimize the value of the error function in weight space. The weights with minimum error functions are then considered to be a solution to the learning problem. In an ANN, hyperparameter is a parameter whose value is set before the learning process begins, and it controls the model structure (e.g., number of layers, number of hidden neurons, number of epochs). Additional information about the theory behind ANN applications can be found in (Rumelhart et al, 1986).

In our study, I applied a trial and error technique to determine the optimal number of hyperparameters where the numbers were added gradually until the predicted, and observed values start to match by evaluating the model performance using the Mean squared normalized error (MSE) performance function. In our study, individual ANNs were constructed for each cluster. In all three clusters, the ANNs consist of one input layer, one hidden layer, and one output layer. The number of hidden neurons in our study ranged from 12 to 18. The number of epochs is the number of times the entire training data are used to update the weights. In other words, it is the number of times that the backpropagation algorithm works through the entire training dataset. The number of epochs ranged from 350 to 500.

The final model evaluation was carried out by the comparison between observed and predicted values on testing data (out of sample data set) using the above-mentioned statistical coefficients (NRMSE, and NSE).

### XGBoost

Gradient boosting was used with decision trees. Decision tree learning is a predictive modeling approach in machine learning that uses a tree-like model to go from observations of predictors (branches of the tree) to the prediction of the target value (leaves of the tree). The goal of our study was to create a model to predict the GRACE<sub>TWS</sub> values from sets of input variables for each month. Using trees has several advantages, including, but not limited to, the ability to handle various types of target variables (e.g., numerical, categorical, and multivariate), modeling complex interactions, and managing missing values with minimal loss of information (De'ath, 2007). However, there are two main limitations with trees: weakness of prediction and difficulty in the interpretation of large trees (De'ath, 2007). To overcome these limitations gradient boosting algorithm was introduced by (Friedman, 2001) and developed by many others (Mason et al., 2000; Chen et al., 2016)

In gradient boosting, the goal is to use a set of predictors  $(X_1, \dots, X_n)$  to predict a set of target data  $(Y_1, \dots, Y_n)$  by fitting a model  $F(X) \rightarrow Y$  and minimize the sum of the loss function  $J = \sum_{i=1}^n L(Y_i, F(X_i))$  by improving the model  $F(X)$  (in our work the loss function  $L(x, y) = (x - y)^2$ ). Then the following iteration is performed:

1. Calculate the negative gradients of  $J$  with respect to  $F(X_i)$ , which is  $-\frac{\partial J}{\partial F(X_i)}$ .
2. Fit a regression tree  $h$  to negative gradients  $-\frac{\partial J}{\partial F(X_i)}$ .
3. Let our new  $F(X_i)$  to be  $F(X_i) + \gamma h$ , where  $\gamma$  is the step size in our algorithm to reach

the estimated minimum of  $J$ .

As a significant improvement over gradient boosting, in XGBoost I start with a loss function  $L(Y_i, F(X_i) + h)$  and minimize  $J = \sum_{i=1}^n L(Y_i, F(X_i) + h) + \Omega(h)$ , where  $\Omega(h) = \gamma T + \frac{1}{2} \lambda ||w||^2$ . Here  $T$  is the number of leaves in the tree and  $w$  is the leaf weights. Fig. 5 shows a schematic diagram of gradient boosting method.

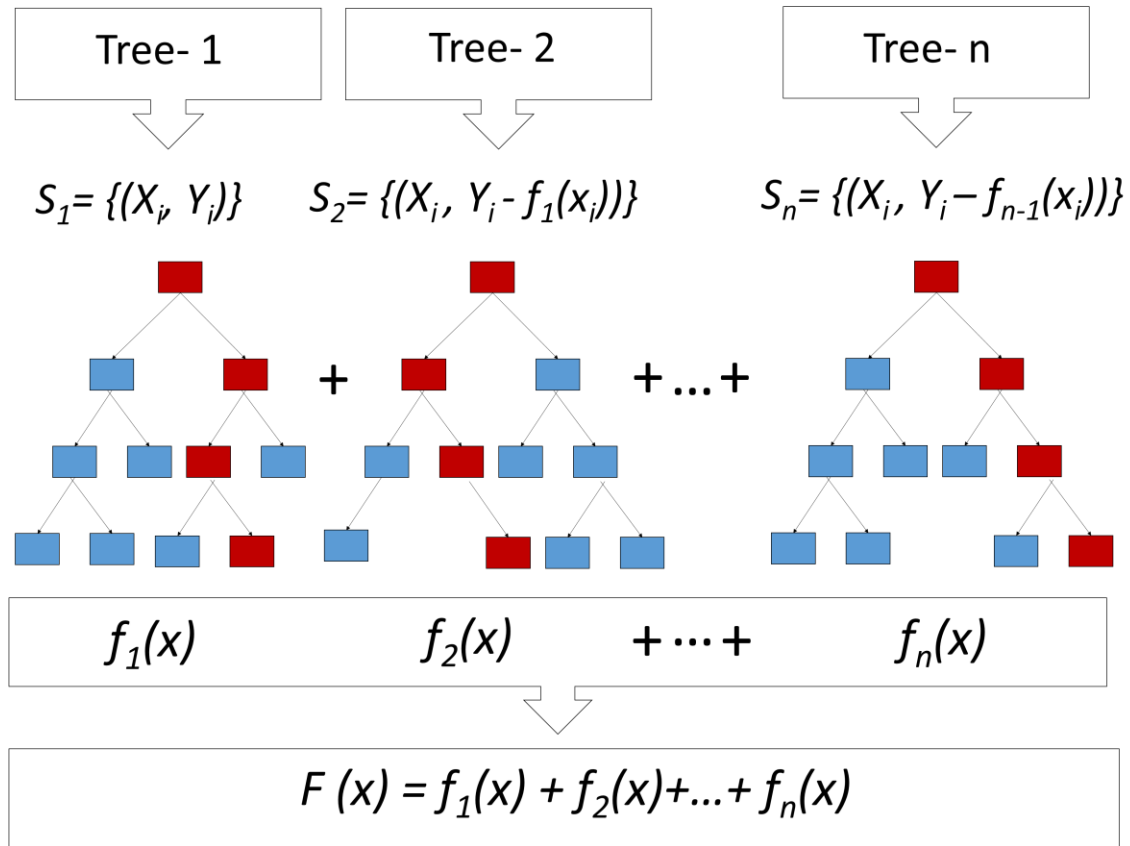


Figure 5. Schematic diagram of a tree-based gradient boosting method.

### Selection of optimum statistical model and Gap Filling

The performance of each of the three models was evaluated by comparing the predicted values with observed values on the testing subset using R-squared, Normalized Root Mean Squared Error (NRMSE) ( Eq.3, Eq.4) and Nash-Sutcliffe efficiency (NSE) (Eq.5) as follows.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(Y_o - \hat{Y}_p)^2}{n}} \quad (3)$$

$$NRMSE = \frac{RMSE}{\bar{Y}_{oi}} \quad (4)$$

$$NSE = 1 - \frac{\sum_{i=1}^n (\hat{Y}_p - Y_o)^2}{\sum_{i=1}^n (Y_o - \bar{Y}_{oi})^2} \quad (5)$$

where  $Y_o$  is the observed value,  $\hat{Y}_p$  is the predicted value, n is the number of observations, and  $\bar{Y}_{oi}$  is the mean of observed data.

The model that produced the highest R-squared and NSE and the lowest NRMSE value was selected. Using the optimum model, the relationships were established between input variables (total precipitation, NDVI, snow cover, Lake Michigan level, Lake Huron level, soil moisture, air temperature, LST, and ET) and GRACE<sub>TWS</sub> as the target variable. These relationships were used to estimate the

missing GRACE<sub>TWS</sub> months.

### 3.5. Extraction of Temporal Groundwater Storage Using Outputs of Land Surface Models

I used the downscaled GRACE data to extract for each of the  $0.125^\circ \times 0.125^\circ$  pixels the changes in GRACE<sub>TWS</sub> ( $\Delta\text{GRACE}_{\text{TWS}}$ ) and the secular trend for each of these pixels (Fig. 6). Then, changes in groundwater storage ( $\Delta\text{GWS}$ ) were calculated using the downscaled  $\Delta\text{GRACE}_{\text{TWS}}$  and outputs of land surface models (NLDAS NOAH) and applying the following equation (Eq. 6):

$$\Delta\text{GWS} = \Delta\text{GRACE}_{\text{TWS}} - (\Delta\text{SMS}_{\text{NLDAS}} + \Delta\text{CWS}_{\text{NLDAS}} + \Delta\text{SWE}_{\text{NLDAS}}),$$

Where  $\Delta\text{SMS}_{\text{NLDAS}}$ ,  $\Delta\text{CWS}_{\text{NLDAS}}$ , and  $\Delta\text{SWE}_{\text{NLDAS}}$  are the changes in soil moisture, canopy water storage, and snow water equivalent, respectively, as extracted from the NLDAS model. All these data are provided in a spatial resolution of  $0.125^\circ \times 0.125^\circ$  ( $\sim 120 \text{ km}^2$ ).

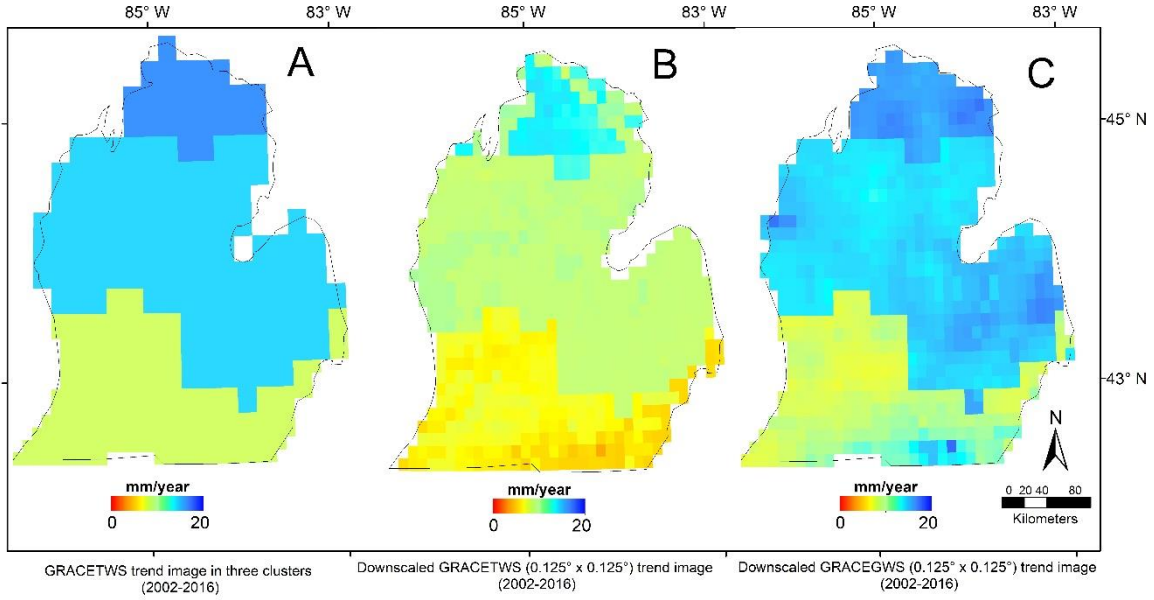


Figure 6. Secular  $GRACE_{TWS}$  and  $GRACE_{GWS}$  trend images (mm/year) over the LP for the period 2002 to 2016. The trend from the  $GRACE_{TWS}$  data prior to gap-filling and associated uncertainty for each cluster is presented in Table 2. (A) Secular  $GRACE_{TWS}$  trend extracted from GRACE mascon data after filling the gaps for missing months. (B) Secular  $GRACE_{TWS}$  trend image extracted from the downscaled GRACE solutions ( $120 \text{ km}^2$ ) after filling the gaps for the missing months. (C) Secular  $GRACE_{GWS}$  trend image extracted from downscaled  $GRACE_{TWS}$  and NLDAS NOAH land surface model outputs. The uncertainties for the fine-resolution  $GRACE_{TWS}$  ( $120 \text{ km}^2$ ) pixels ranged from 0.6 mm/year to 1.68 mm/year.

### 3.6 Sources and Propagation of Errors

The uncertainties in the  $GRACE_{TWS}$  trends reflect the variations between trend values that were extracted from three GRACE solutions (CSR-RL06M, JPL-RL06M, and GSFC-M) (Table 2; (Scanlon et al., 2018)).

The uncertainties in the downscaled  $GRACE_{TWS}$  are related to (1) uncertainties in the variables (remote sensing-based and land surface model-based) that were used as inputs to the statistical models, and (2) errors introduced by the applied statistical models. The statistical coefficients (R-square, NSE, and NRMSE) describe the



accuracy of the extracted model, namely the degree to which the extracted statistical model can, or cannot predict the target ( $GRACE_{TWS}$  in our case). Statistical models that have R-square, NSE, and NRMSE values of 0.9, 0.9, and 0.1, respectively can predict the target with an accuracy of 90%. Since the reported accuracy of the models was estimated by comparing the modeled and observed  $GRACE_{TWS}$  values in the test dataset, the reported model errors incorporate the errors associated with the individual variables as well. In this respect, the coefficients could be used to estimate the errors introduced by both the variables and the statistical models (Seyoum et al., 2019).

I assumed that the model-based accuracy of  $GRACE_{TWS}$  in area A applies to all downscaled pixels within this area. Similarly, the  $GRACE_{TWS}$  of the downscaled pixels within areas B and C will inherit the estimated accuracy for areas B and C, respectively. These are reasonable assumptions given that all of the CSR06M pixels within each of areas A, B, and C have similar geophysical signals.

The errors (uncertainties) associated with the estimated downscaled  $GRACE_{GWS}$  values were propagated from the estimated errors in the  $GRACE_{TWS}$  and in the land surface model outputs ( $SMS_{NLDAS}$ ,  $CWS_{NLDAS}$ , and  $SWE_{NLDAS}$ ) were used in calculating changes in groundwater storage (eq. 7). The errors in each of these land surface model outputs were calculated as the standard deviation of the values extracted from the three NLDAS simulations (NOAH, VIC, and Mosaic (Castle et al., 2014) (Voss et al., 2013). The errors in the estimated GWS ( $\sigma_{GWS}$ ) was calculated by adding, in quadrature, the uncertainties related to  $GRACE_{TWS}$ ,  $SMS_{NLDAS}$ ,  $CWS_{NLDAS}$ , and  $SWE_{NLDAS}$  values (Eq. 7).

$$\sigma_{GWS} = \sqrt{(\sigma_{TWS})^2 + (\sigma_{SMS})^2 + (\sigma_{SWE})^2 + (\sigma_{CWS})^2} \quad (7)$$

Where  $\sigma_{TWS}$  ,  $\sigma_{SMS}$ ,  $\sigma_{SWE}$ , and  $\sigma_{CWS}$  are the uncertainties related to terrestrial water storage, soil moisture, snow water equivalent, and canopy water storage, respectively.

## CHAPTER 4

### RESULTS

#### 4.1. Cluster Analysis

The optimum number of clusters was estimated at 3. Three clusters were identified (cluster 1 area: 13,700 km<sup>2</sup>, cluster 2 area: 59,200 km<sup>2</sup>, cluster 3 area: 33,100 km<sup>2</sup>; Fig.1). The correlation coefficients of the GRACE time series between clusters were evaluated through the construction of a correlation matrix (Table 4). The correlation coefficients of the GRACE time series between clusters varies from 0.41 to 0.66, and those between clusters and the Michigan lake level varies from 0.43 to 0.74. In the generation of the correlation matrix, the secular trends and seasonal cycles were removed from the time series. Although I cannot rule out leakage from the adjacent water bodies and areas, I suggest that there are significant geophysical signals in each of the investigated areas, as evidenced by the following observations. First, higher correlation coefficients than those observed (0.41–0.66) are to be expected between GRACE<sub>TWS</sub> over areas 1, 2, and 3 if the leakage was significant. Second, lake levels lag behind GRACE<sub>TWS</sub> by 1 to 2 months for areas 1, 2, and 3 (Fig. 7).

Table 4. Correlation matrix for GRACE<sub>TWS</sub> values over land (three clusters) and Lake Michigan water levels. The values are presented after the removal of the seasonal cycle and secular trends.

	Cluster 1	Cluster 2	Cluster 3	Lake Level
Cluster1	1			
Cluster2	0.41	1		
Cluster3	0.66	0.56	1	
Lake Level	0.74	0.43	0.58	1

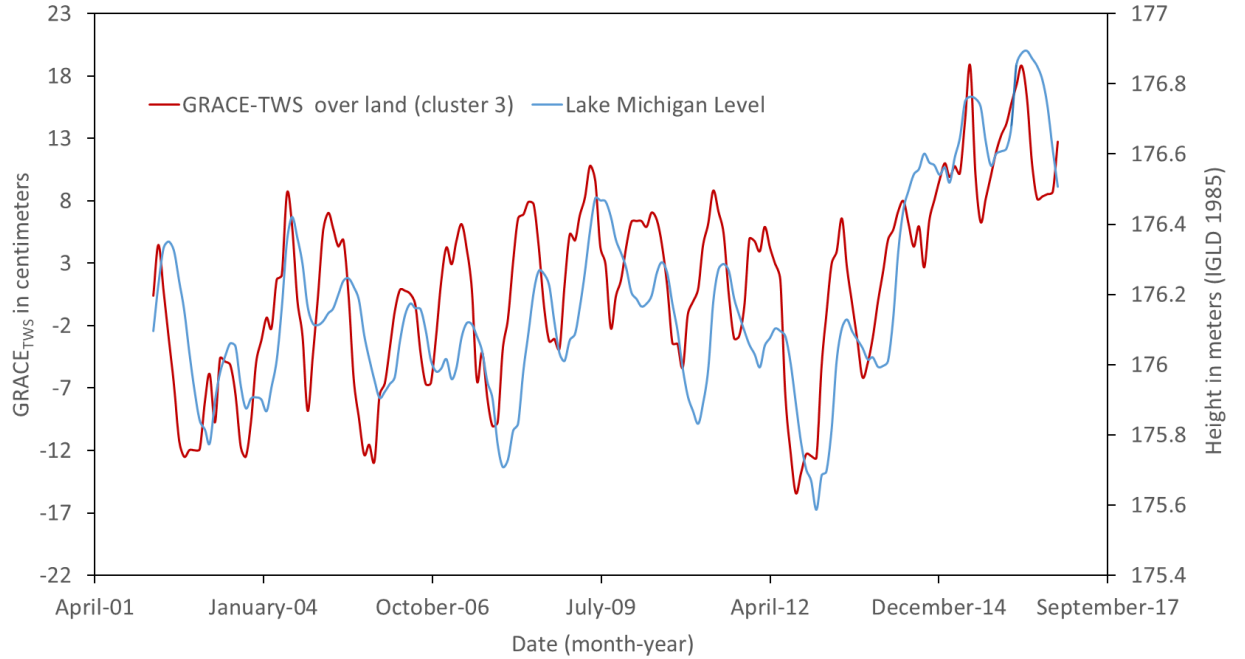


Figure 7. Comparison of the time series of GRACE<sub>TWS</sub> over land (cluster 3) and Michigan lake level. The time series for the lake levels lagged by two months.

#### 4.2. Evaluation and Comparison of the Models

Comparison of the performance of the three models revealed that, in general, the XGBoost models perform better than the other two models (Table 5) as indicated by their lower NRMSE values and their higher NSE and R-squared values and ranking. For example, the R-squared, NRMSE and NSE values for the

XGBoost models ranged from 0.45 to 0.47 (average: 0.46), 0.76 to 0.78 (average: 0.46), and 0.76 to 0.78 (average: 0.77), respectively; those for the ANN models ranged from 0.66 to 0.75 (average: 0.72), 0.51 to 0.53 (average: 0.52) and from 0.68 to 0.72 (average: 0.70), respectively, and those for the MR models from 0.63 to 0.74 (average: 0.69), 0.53 to 0.71 (average: 0.62) and from 0.53 to 0.67 (average: 0.60), respectively. The performance of the XGBoost is high (clusters 1, 2, and 3: Very Good), compared to that for MR models (clusters 2 and 3: Very Good; cluster 1: Good) and for the unified ANN model (clusters 2 and 3: Very Good; cluster 1: Unsatisfactory). One plausible explanation for the enhanced performance of the XGBoost models over the ANN and MR models is that it can better account for the specific characteristics or significant variables that control, or relate to, the observed temporal GRACE solutions in each cluster. It is flexible and performs well with categorical and numerical values [(T. Chen & Guestrin, 2016)], as is the case with our datasets (Fig. 8).

Table 5. Statistical coefficients (R-squared, NRMSE, and NSE) for each of the examined models (Extreme Gradient Boosting, Multivariate Regression, and Artificial Neural Network) over clusters 1, 2, and 3 and calculated uncertainties.

Method		Cluster1	Cluster2	Cluster3
Extreme Gradient Boosting	Rsquared	0.80	0.81	0.81
	NSE	0.75	0.76	0.78
	NRMSE	0.45	0.47	0.46
	Average	30	30	29
	Uncertainty (%)			
	Ranking*	VG	VG	VG
Artificial Neural Networks	Rsquared	0.66	0.74	0.75
	NSE	0.68	0.71	0.72
	NRMSE	0.53	0.53	0.51
	Average	42.3	36	34.7
	Uncertainty (%)			

Table 5- continued

	Ranking	G	G	G
	Rsquared	0.63	0.72	0.74
	NSE	0.53	0.63	0.71
Multivariate	NRMSE	0.67	0.59	0.53
Regression	Average	50	48	36
	Uncertainty (%)			
	Ranking*	S	S	G

\*VG: Very Good; G: Good; S: Satisfactory; US: Unsatisfactory.

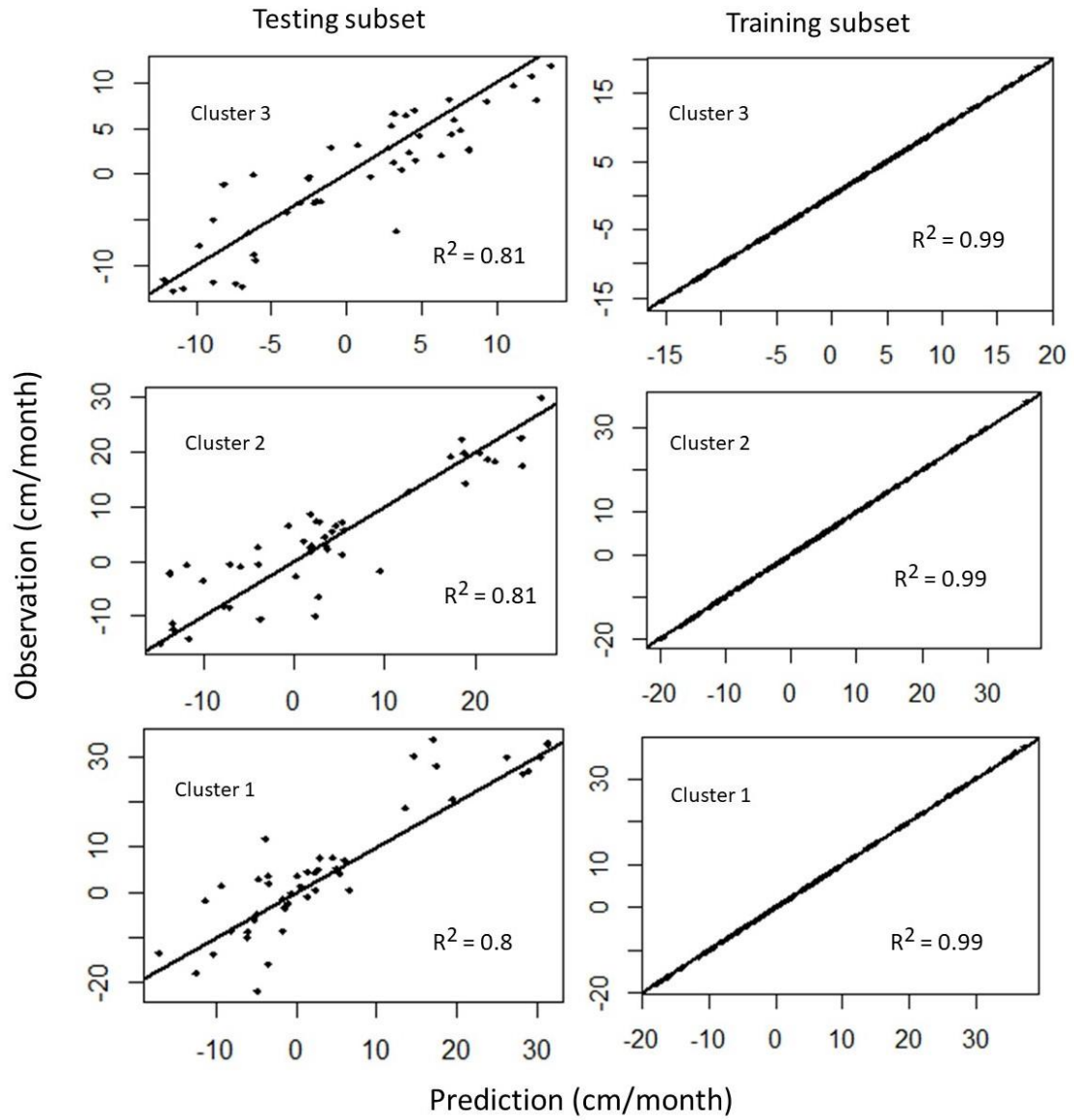


Figure 8. Scatter plots of the GRACE<sub>TWS</sub> predicted values (using XGBoost models) versus observed (from testing subsets) GRACE<sub>TWS</sub> values for the three clusters.

The combined errors were estimated by averaging the three coefficients and were presented in table 4 as percent uncertainty of the output (GRACE<sub>TWS</sub>). The estimated total error rate for fine-scale GWS in three arbitrary pixels (location given in Fig. 1) is presented in figure 9.

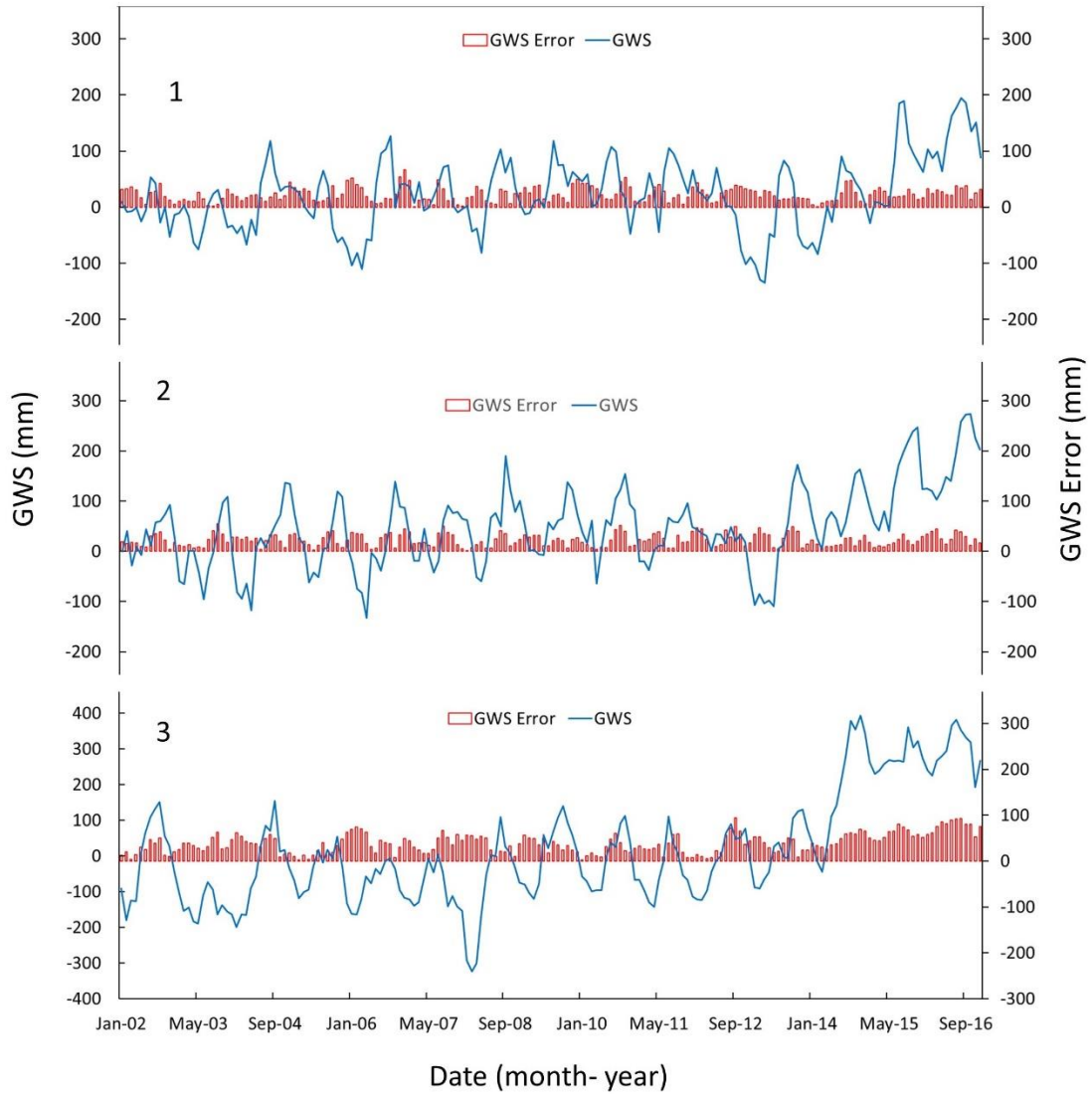


Figure 9. The estimated total error rate for three arbitrary pixels (1,2, and 3). The locations of the pixels are shown in figure 1.

#### 4.3. Factors Controlling the TWS and GWS Variations over the Study Area

Six out of 11 variables showed statistical significance with the  $GRACE_{TWS}$  values in XGboost models (Table 6). They were used for the downscaling process based on cluster-based XGBoost models. Those variables



are ET, air temperature, NDVI, total precipitation, soil moisture, and Lake Michigan water level. The significance of the variables is determined by their p-values. The insignificant variables were omitted due to their high p-value (probability value). The p-value represents the probability of the occurrence of a given event and helps determine the significance of the results. The higher p-values (typically  $>0.05$ ) indicate weak evidence against the null hypothesis (i.e., there is no significant relationship between the independent variable and the target, and therefore the variable is insignificant (Fisher, 1992)). The smaller p-values (typically  $\leq 0.05$ ) indicate the opposite: there is strong evidence in favor of the alternative hypothesis, and there is a significant relationship between the independent variable and the target. The percent contribution of variables was determined using the variable importance (VI) technique in the XGBoost model (Table 6). The average values of the variables were presented in appendix 1.

Table 6. Percent contribution of each variable in the outputs of the XGBoost models and their optimum lag times.

Variables						
Clusters	Total Precipitation	Temperature	NDVI	Soil Moisture	Lake Michigan Level	Evapotranspiration
1	6.1 (1)*	1. 2	4 .1 (1)	18. 3 (2)	69.1 (1)	1.2 (2)
2	1.3 (3)	16 .2 (3)	1 3.5	8.4 (1)	34.1 (2)	26.5 (2)
3	4.6 (1)	2. 9	2 .4 (2)	51. 1 (1)	37.6 (2)	1.4

\*The number in parentheses shows the optimum lag time (in months) for each variable.

Multicollinearity, a condition in which two or more predictors are highly correlated with one another in linear regression models, was addressed using

variance inflation factor (VIF). Multicollinearity makes it difficult to determine the effect of the individual predictors on the response and to identify the variables to be included in the model. VIF is one of the most widely used diagnostic indices for multicollinearity (Alin, 2010). It estimates how much the variance of a coefficient is “inflated” because of linear dependences with other predictors (Alin, 2010). Using a VIF value of 11 in this study, one of two variables (Lake Michigan and Lake Huron water levels) that show multicollinearity was omitted. Lake Huron lake level was automatically removed from the set of individual variables in the stepwise procedure and was not used in our models. Five lag times (months 1 through 5) were assigned to each of the investigated variables to identify the optimum lag time for the individual variables. Four of the examined variables (total precipitation, temperature, ET, and NDVI) were found to have optimum lag times ranging from 1 to 3 months; none exceeded 3 months, and the remaining variables had no lag times (Table 6). The optimal lag time was found to vary from one cluster to another; for example, the lag in total precipitation varied from 1 month in cluster 1 to 3 months in cluster 2. Again, less significant lag times for an individual variable were automatically removed throughout the application of the stepwise regression.

The significant variables are the ones that correlate well with, respond fairly quickly to, and either drive or are driven by the variations in  $GRACE_{TWS}$ . An increase in soil moisture over a cluster will increase its  $GRACE_{TWS}$  values, whereas an increase in land surface temperature or ET will probably decrease its TWS values. Interestingly, lake levels correlated well with  $GRACE_{TWS}$ , which is to be expected given that both the land (clusters 1–3) and surrounding water

bodies (Lakes Michigan and Huron) receive added water contributions (precipitation and SWE) which will increase the water levels in the lakes and increase the  $GRACE_{TWS}$  over the land. However, the lake water levels lagged behind  $GRACE_{TWS}$  by 1 to 2 months. I suspect that this lag time is related to the time period it takes for runoff to reach the lake. Starting in 2013, there has been an increase in water levels in both Lake Huron and Lake Michigan. A thorough investigation revealed that the recent rise in water level in Lake Michigan-Huron is due to above-average spring runoff, which drains into the lakes and excess precipitation over the lake as well (Gronewold et al., 2016).

## CHAPTER 5

### DISCUSSION AND IMPLICATIONS

#### 5.1 Introduction

The original size of the pixels over the LP (irregular grid, pixel size  $\sim 12000 \text{ km}^2$ ) is coarse for monitoring TWS and GWS on the county scales (size range:  $500 \text{ km}^2$  [e.g., St. Joseph County] to  $850 \text{ km}^2$  [e.g., Kent County]). The adopted downscaling technique addresses this issue through the generation of downscaled  $\text{GRACE}_{\text{TWS}}$  and  $\text{GRACE}_{\text{GWS}}$  (spatial resolution:  $0.125^\circ \times 0.125^\circ$ ;  $10 \times 14 \text{ km} = 140 \text{ km}^2$ ).

#### 5.2 Inspection of Trends

Inspection of the secular trends in  $\text{GRACE}_{\text{TWS}}$  and  $\text{GRACE}_{\text{GWS}}$  revealed two general patterns, a near-steady-state in  $\text{GRACE}_{\text{TWS}}$  and  $\text{GRACE}_{\text{GWS}}$  ( $-1$  to  $+1 \text{ mm/year}$ ) for an earlier period (2002 to 2012), hereafter referred to as period I, followed by an increase in  $\text{GRACE}_{\text{TWS}}$  ( $28$  to  $120 \text{ mm/year}$ ) and  $\text{GRACE}_{\text{GWS}}$  ( $10$  to  $130 \text{ mm/year}$ ) for a later period (2013 to 2016), hereafter referred to as period II (Fig. 10).

The breakpoints were identified using the regime shift detection method with a 95% confidence (Andersen et al., 2009). The two above-mentioned general patterns were observed throughout the entire investigated area. For all of the downscaled pixels, no major differences in  $\text{GRACE}_{\text{TWS}}$  and  $\text{GRACE}_{\text{GWS}}$  trends were observed during Period I, all of which show near-steady trends. However, distinct variations in  $\text{GRACE}_{\text{TWS}}$  and  $\text{GRACE}_{\text{GWS}}$  trend values are observed in Period II between the three clusters (Fig. 10).

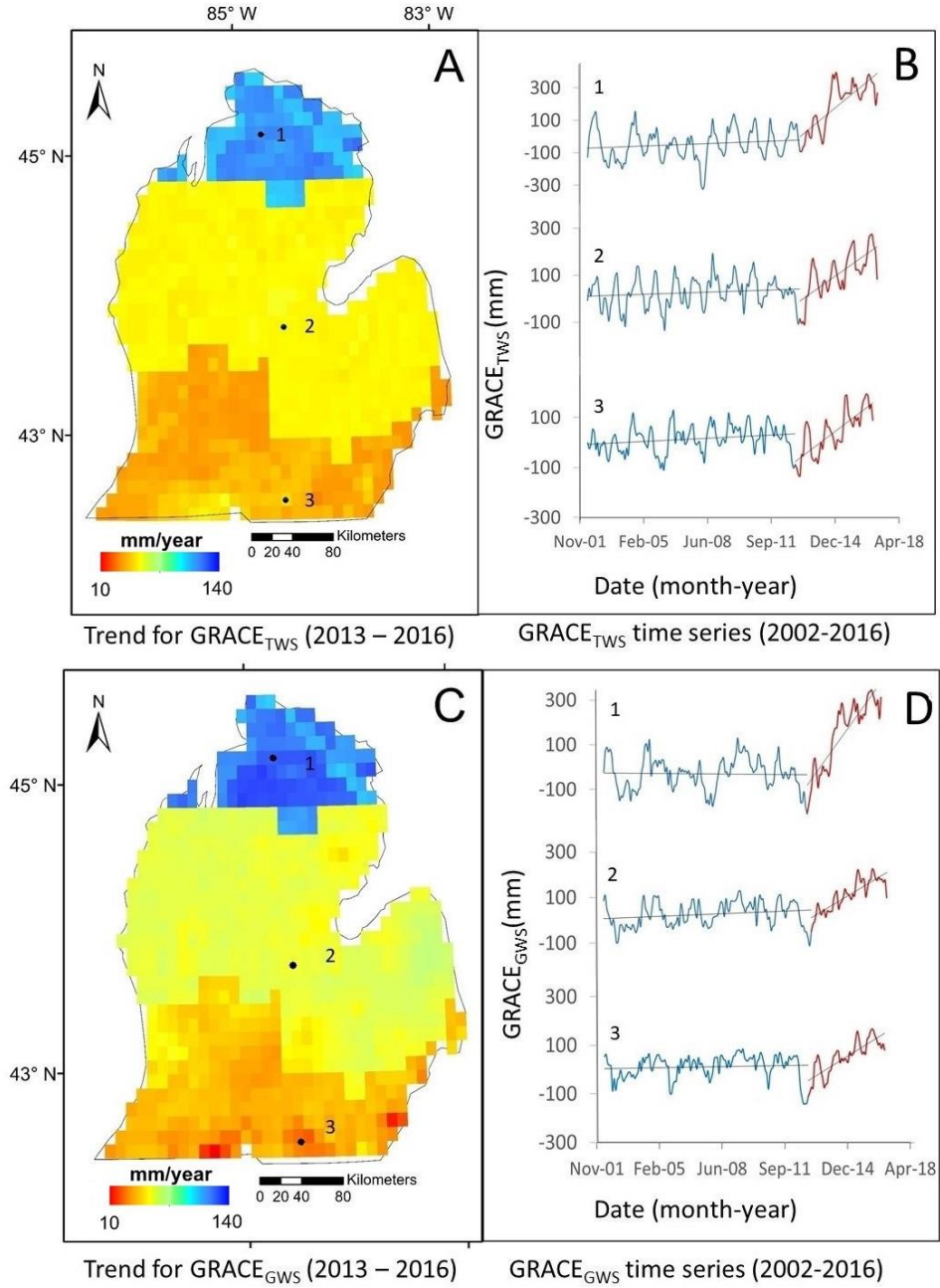


Figure 10. GRACE<sub>TWS</sub> and GRACE<sub>GWS</sub> trend images and time series for downscaled data. (A) GRACE<sub>TWS</sub> trend image for period II (2013 to 2016). (B) GRACE<sub>TWS</sub> time series (2012–2016) over four locations. (C) GRACE<sub>GWS</sub> trend image for period II. (D) GRACE<sub>GWS</sub> time series (2012–2016) over four locations.

Cluster 1 (represented by point 1; Fig. 10) has the highest GRACE<sub>TWS</sub> (103 to 122

mm/year) and  $GRACE_{GWS}$  (100 to 130 mm/year) trends, followed by cluster 2 (represented by point 2) with a TWS trend of 50 to 57 mm/year and GWS trend of 45 to 70 mm/year. Groups 1 and 2 are located within areas characterized by the highest average SWE (60 to 190 mm/year) and the highest average annual rainfall (800 to 1043 mm/year), respectively. Cluster 3 is located in the southern and southwestern parts of the LP, areas that are characterized by high groundwater extraction. This cluster (represented by point 3) shows a TWS trend of 28 to 55 mm/year and a GWS trend of 10 to 55 mm/year.

The glacial aquifers are widely distributed, they overlies all other aquifers, and crop out across large sectors in the State, and thus one would expect that the observed variations in  $GRACE_{GWS}$  are largely controlled by variations in glacial aquifer storage. Clusters 1 and 2 are located in the northern and central sections of the province 2 (Figs 1, 2) where the glacial deposit is relatively thick, whereas cluster 3 is located in the southwestern section of the LP, an area characterized by high groundwater withdrawal for agricultural activities (Fig. 1). The eastern part of cluster 3 is located in an area characterized by low yield (Fig. 2) and thin to moderate glacial deposits (refer to section 2). Also, in general, the northern, but not the southern sections, of the LP have high vertical conductivity and low groundwater extraction rates. The average annual rainfall over the entire LP increased from 774 mm/year (period I) to 783 mm/year (period II), and the average annual SWE increased from 50 mm/year (period I) to 55 mm/year (period II) (Fig. 11). For cluster 1, the average annual rainfall increased from 689 mm/year (period I) to 723 mm/year (period II), and the average annual SWE increased from 44 mm/year to 75 mm/year. Similarly, for cluster 2, the average annual rainfall increased from 761 mm/year (period I) to 785 mm/year (period II), and the average annual SWE increased from 52 mm/year to 56 mm/year in periods I and II, respectively. For cluster 3, the average annual SWE increased from 36 mm/year (period I) to 44 mm/year (period II), but the average

annual rainfall decreased from 834 mm/year (period I) to 823 mm/year (period II). These collective observations suggest that the observed steep  $GRACE_{TWS}$  and  $GRACE_{GWS}$  trends over the northern sections of the LP during period II could be related to one or more of these factors: (1) thickened glacial deposit, (2) high precipitation and/or snowfall rates, (3) high vertical conductivity and (4) low extraction rates.

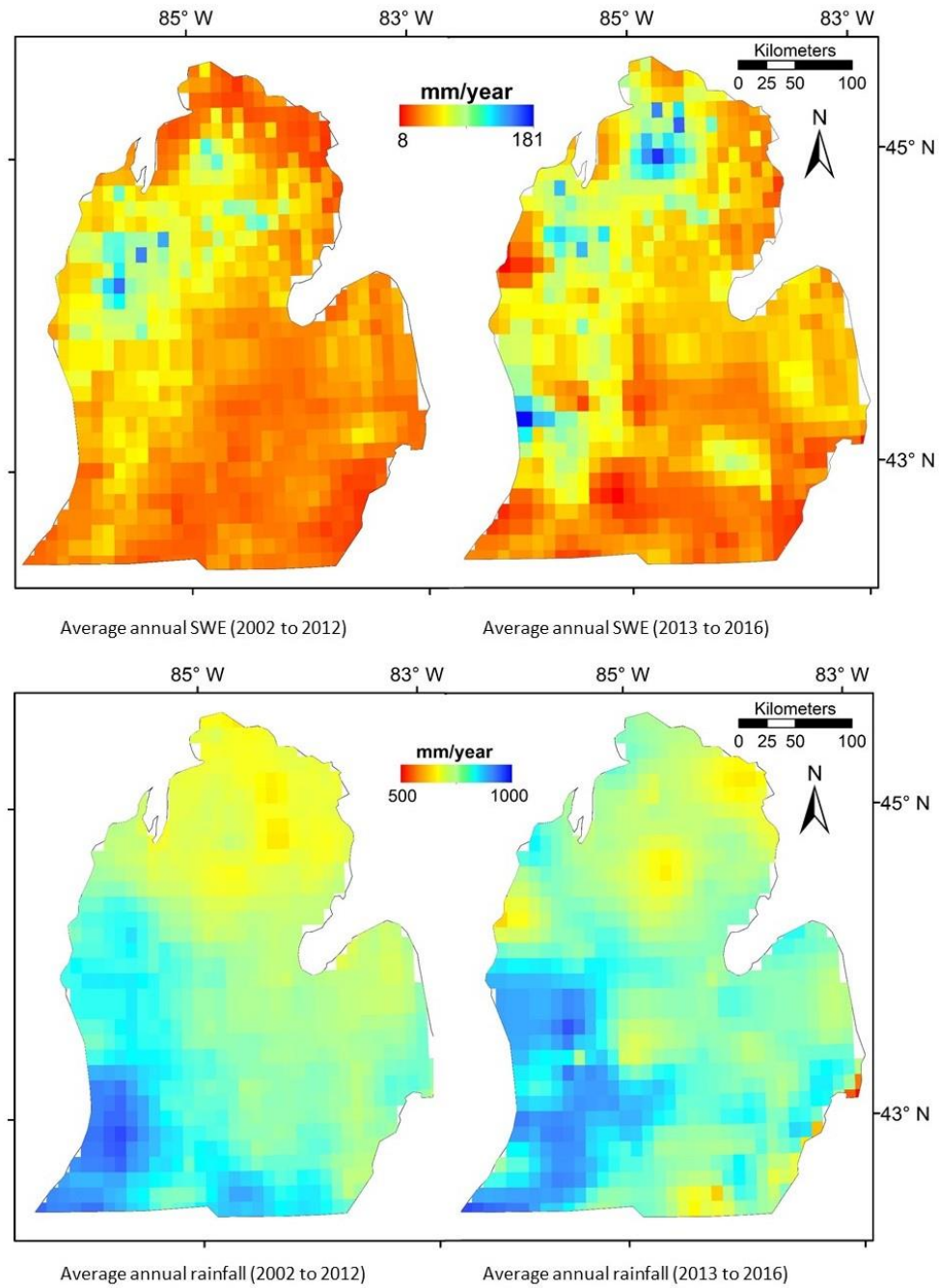


Figure 11. Average annual rainfall and snow water equivalent for periods I (2002–2012 and II (2013–2016). (A) Average annual rainfall (mm/year) for period I. (B) Average annual rainfall for period II. (C) Average annual snow water equivalent for period I. (D) Average annual snow water equivalent for period II.

One would expect the above-mentioned temporal variations in precipitation and SWE



to be reflected in the downscaled  $GRACE_{GWS}$  and groundwater levels. Figure 12 shows an overall correspondence between the downscaled  $GRACE_{GWS}$  data and groundwater levels from three monitoring wells (well A: site name 02S 11W 22CDBB 01, location Kalamazoo; well B: site name 04N 02W 26BBDB 01, location Lansing; and well C: site name 04N 02W 16DAAA 01, location Lansing; Fig. 1) within each of the downscaled  $GRACE_{GWS}$  pixels. One should not expect a one to one correspondence between the two datasets. The groundwater levels, but not the GWS, is affected by groundwater withdrawal and by the lag time (between precipitation and recharge). Unfortunately, only a few of the monitoring wells, all located in Kalamazoo and Lansing, have continuous measurements throughout the investigated period and across the study area, and none of which are located in the central or northern LP (Fig. 1). The correlation coefficients between the downscaled GWS ( $0.125^\circ \times 0.125^\circ$ ) and the observed groundwater level in wells A, B, and C were calculated at 0.4, 0.55, and 0.32, respectively, higher than those between the original  $GRACE_{GWS}$  and the wells (A: 0.14; B: 0.36; and C: 0.05).

I also compared the time series for surface water levels from two inland lakes (Otsego lake in northern LP and Austin lake in southern LP; Fig. 1) to downscaled  $GRACE_{TWS}$  time series in areas (pixels) proximal to these lakes (Fig. 13). The surface water lake levels approximate the groundwater table in the surrounding areas, and thus, the changes in lake level should be indicative of the temporal variations in  $GRACE_{GWS}$  as it was reported in previous studies (Krabbenhoft et al., 1990; Krabbenhoft and Webster, 1995). Fig. 13, shows a good correspondence between the downscaled  $GRACE_{GWS}$  and Otsego lake level (correlation coefficient: 0.73) and Austin lake level (correlation coefficient: 0.75), an observation that further validates the adopted downscaling procedures.

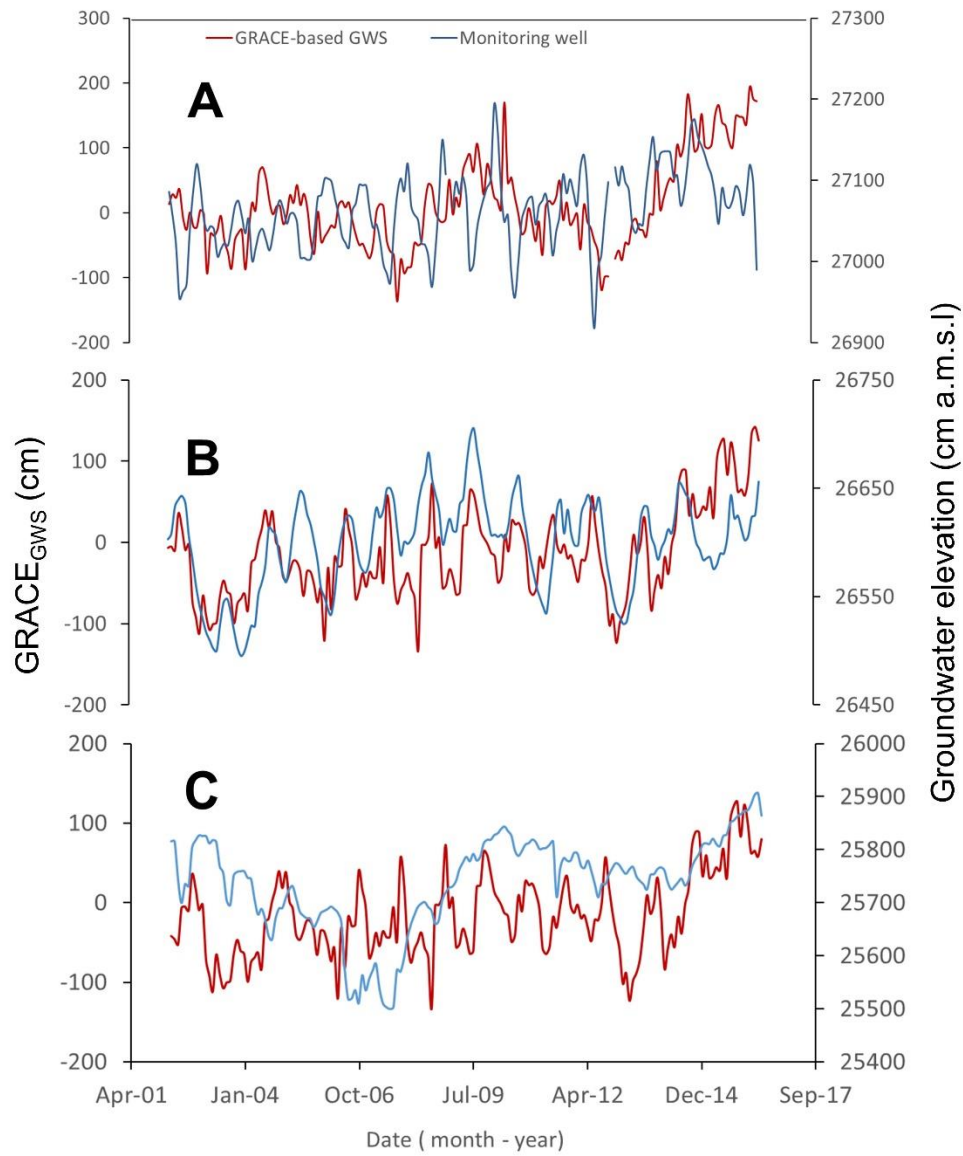


Figure 12. Comparison between the downscaled  $GRACE_{GWS}$  data for three pixels and groundwater levels from monitoring wells within each of the three GRACE pixels in Kalamazoo (well A) and Lansing (wells B and C) (see locations in Fig. 1). Groundwater level elevations are given in elevation above mean sea level (cm).

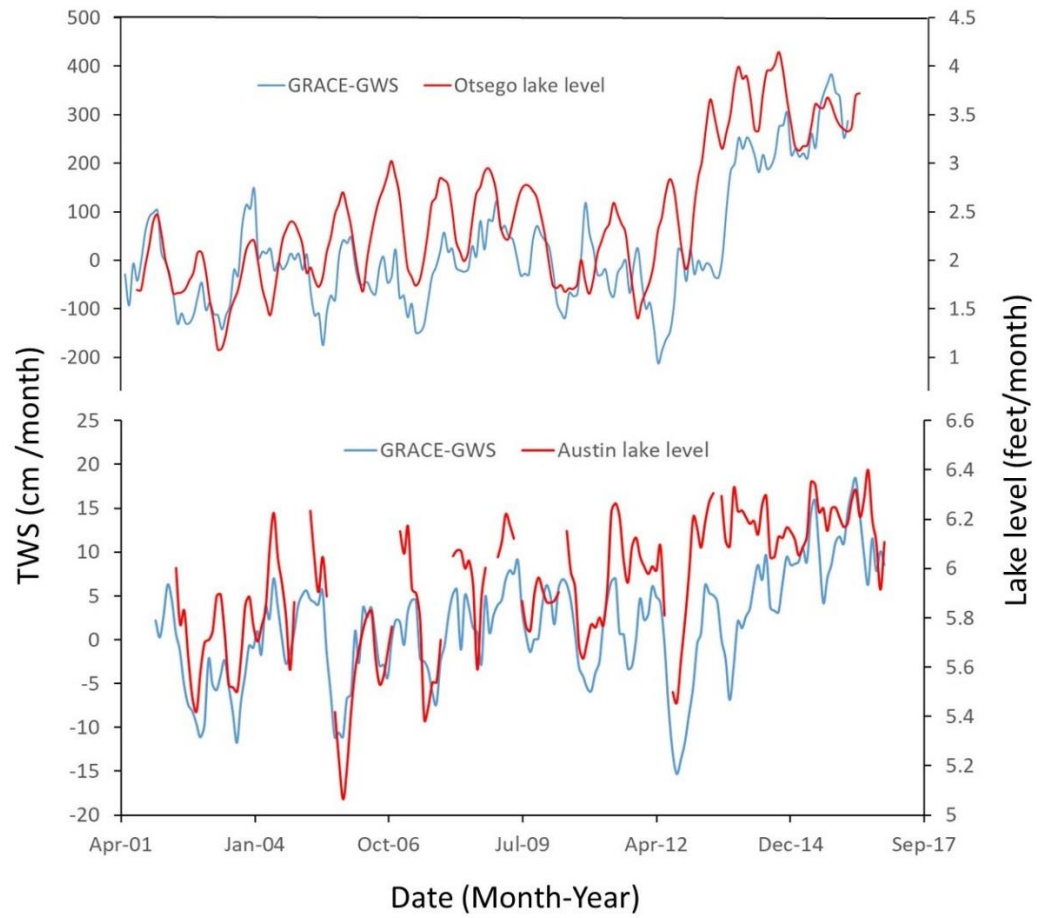


Figure 13. Comparison between the downscaled  $GRACE_{TWS}$  data for two pixels and two inland lakes, namely Otsego lake and Austin lake (see locations in Fig. 1). The discontinuity in the figure is due to temporal gaps in the original data.

## CHAPTER 6

### SUMMARY AND CONCLUSION

The GRACE data has been widely used to monitor the temporal and spatial variations in TWS and GWS on large scales. However, such applications remained limited on the local scales due to the poor spatial resolution (irregular grid of 12000 km<sup>2</sup>) of the GRACE data. The objective of this study was to address this shortcoming by downscaling the CSR mascon solutions to a finer resolution ( $0.125^{\circ} \times 0.125^{\circ}$ ) to enable the monitoring of GRACE<sub>TWS</sub> and GRACE<sub>GWS</sub> on county scales and fill the gaps for missing months in GRACE time series over the study area. Using cluster analysis, areas of similar GRACE<sub>TWS</sub> patterns within the study area were first identified. For each of the identified clusters, variables (total precipitation, NDVI, snow cover, Lake Michigan water level, Lake Huron water level, soil moisture, land surface temperature, and ET) that presumably contributed to, or were correlated with, the GRACE data were identified and collected on a monthly basis over the investigated period (2002 to 2016). The data sets were randomly partitioned into two groups: training data (80%) and testing data (20%). XGBoost, MR, and ANN methods were applied to extract statistical relationships between the independent variables and the GRACE<sub>TWS</sub> (dependent variable).

The comparisons of the observed GRACE<sub>TWS</sub> (training dataset) versus the modeled GRACE<sub>TWS</sub> values showed that the XGBoost method outperformed the other two methods as indicated by their lower NRMSE and higher NSE values compared to those obtained from the MR and ANN models. The unified approaches have the advantage of providing adequate overall downscaling results over large areas, yet one would expect the performance of the model to vary from one area to another given that the selected variables and/or their significance is likely to vary across the investigated area. I suggest that if statistical

downscaling methods were selected for downscaling  $GRACE_{TWS}$  values on local scales, preference should be given to cluster-based approaches over the commonly used unified approaches.

The XGBoost model was used to downscale ( $12,000 \text{ km}^2$  to  $120 \text{ km}^2$ )  $GRACE_{TWS}$ , given the high performance of this model over all other models (ANN and MR) and its ability to estimate the contributions of the independent variables towards the response variable and to forecast missing months within the GRACE's time series data. Although the XGBoost model outperformed the ANN method in all three clusters in our study area, that might not necessarily be the case over other locations. I suggest that one should explore the use of multiple statistical approaches and select the one that performs the best over each of the investigated areas (clusters).

Since the individual variables and the degree to which they correlate with  $GRACE_{TWS}$  vary from one cluster to another, it is recommended to identify the local hydrologic components that are specific to the investigated area and to select the optimum cluster-based model to improve the accuracy of the downscaled  $GRACE_{TWS}$  values. The accuracy of the derived downscaled  $GRACE_{GWS}$  will largely depend on the accuracy of the land surface model outputs that were used in calculating  $GRACE_{GWS}$ , namely the  $SMS_{NLDAS}$  and  $SWE_{NLDAS}$  in our case. Unfortunately, the State of Michigan lacks a comprehensive groundwater monitoring program to adequately validate the downscaled  $GRACE_{GWS}$  data.

As discussed earlier, I cannot rule out leakage from the adjacent water bodies and/or areas, but I suggest that there are significant geophysical signals from each of the investigated areas (clusters) as evidenced by the modest correlation coefficients between the time series of areas 1, 2, and 3 and by the lag of lake levels by 1-2 months behind the  $GRACE_{TWS}$  over the land areas. Currently, efforts are underway to generate  $GRACE_{TWS}$  of higher spatial resolution

(1° x 1°) by NASA JPL, by combining satellite gravimetry and in-situ GNSS measurements [89]. If and when such data become available, I can apply the proposed methodologies on the individual pixels without worrying about the leakage from their surroundings.

developed a straightforward methodology that could be used to monitor temporal variations in  $GRACE_{TWS}$  and  $GRACE_{GWS}$  on local scales (county levels). The methodology takes advantage of readily available remote sensing datasets and outputs of land surface models that are of global nature, both of which come at no cost to users. These methodologies could be used by local communities and decision-makers for water management purposes in the State of Michigan. They can also provide a replicable model for local applications across the continental USA and possibly in similar settings worldwide as well. The performance of the statistical models can be enhanced by identifying and including local variables that control, or correlate with, the  $GRACE_{TWS}$  solutions over the investigated areas.

## REFERENCES

- Ahmed, M., Sultan, M., Wahr, J., Yan, E., Milewski, A., Sauck, W., ... Welton, B. (2011). Integration of GRACE (Gravity Recovery and Climate Experiment) data with traditional data sets for a better understanding of the time-dependent water partitioning in African watersheds. *Geology*, 39(5), 479–482. <https://doi.org/10.1130/G31812.1>
- Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3), 370–374. <https://doi.org/10.1002/wics.84>
- Alshehri, F., Sultan, M., Karki, S., Alwagdani, E., Alsefry, S., Alharbi, H., Sahour, H., & Sturchio, N. (2020). Mapping the Distribution of Shallow Groundwater Occurrences Using Remote Sensing-Based Statistical Modeling over Southwest Saudi Arabia. *Remote Sensing*, 12(9), 1361.
- Andersen, T., Carstensen, J., Hernández-García, E., & Duarte, C. M. (2009). Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evolution*, 24(1), 49–57. <https://doi.org/10.1016/J.TREE.2008.07.014>
- Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 22, 106–114. <https://doi.org/10.1016/J.JAG.2012.04.012>
- Castellazzi, P., Longuevergne, L., Martel, R., Rivera, A., Brouard, C., & Chaussard, E. (2018). Quantitative mapping of groundwater depletion at the water management scale using a combined GRACE/InSAR approach. *Remote Sensing of Environment*, 205, 408–418. <https://doi.org/10.1016/J.RSE.2017.11.025>

- Castle, S. L., Thomas, B. F., Reager, J. T., Rodell, M., Swenson, S. C., & Famiglietti, J. S. (2014). Groundwater depletion during drought threatens future water security of the Colorado River Basin. *Geophysical Research Letters*, 41(16), 5904–5911. <https://doi.org/10.1002/2014GL061055>
- Chadwick, R., Coppola, E., & Giorgi, F. (2011). An artificial neural network technique for downscaling GCM outputs to RCM spatial scale. *Nonlin. Processes Geophys*, 18, 1013–1028. <https://doi.org/10.5194/npg-18-1013-2011>
- Chen, J. L., Rodell, M., Wilson, C. R., & Famiglietti, J. S. (2005). Low degree spherical harmonic influences on Gravity Recovery and Climate Experiment (GRACE) water storage estimates. *Geophysical Research Letters*, 32(14), n/a-n/a. <https://doi.org/10.1029/2005GL022964>
- Chen, T., & Guestrin, C. (n.d.). *XGBoost: A Scalable Tree Boosting System*. <https://doi.org/10.1145/2939672.2939785>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August-2016*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- D. N. Moriasi, D. N., J. G. Arnold, J. G., M. W. Van Liew, M. W. Van, R. L. Bingner, R. L., R. D. Harmel, R. D., & T. L. Veith, T. L. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>



- Dhanachandra, N., Manglem, K., & Chanu, Y. J. (2015). Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Computer Science*, 54, 764-771.
- De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology*, 88(1), 243–251. [https://doi.org/10.1890/0012-9658\(2007\)88\[243:BTFFEMA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2007)88[243:BTFFEMA]2.0.CO;2)
- Ezzine, H., Bouziane, A., Ouazar, D., & Hasnaoui, M. D. (2017). Downscaling of TRMM3B43 Product Through Spatial and Statistical Analysis Based on Normalized Difference Water Index, Elevation, and Distance From Sea. *IEEE Geoscience and Remote Sensing Letters*, 14(9), 1449–1453. <https://doi.org/10.1109/LGRS.2017.2705430>
- Feng, W., Zhong, M., Lemoine, J.-M., Biancale, R., Hsu, H.-T., & Xia, J. (2013). Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements. *Water Resources Research*, 49(4), 2110–2118. <https://doi.org/10.1002/wrcr.20192>
- Fisher, R. A. (1992). *Statistical Methods for Research Workers*. [https://doi.org/10.1007/978-1-4612-4380-9\\_6](https://doi.org/10.1007/978-1-4612-4380-9_6)
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.2307/2699986>
- Grannemann, N. G., Hunt, R. J., Nicholas, J. R., Reilly, T. E., & Winter, T. C. (2000). The importance of ground water in the Great Lakes Region. In *Water-Resources Investigations Report*. <https://doi.org/10.3133/wri004008>

- Gronewold, A. D., Bruxer, J., Durnford, D., Smith, J. P., Clites, A. H., Seglenieks, F., ... Fortin, V. (2016). Hydrological drivers of record-setting water level rise on Earth's largest lake system. *Water Resources Research*, 52(5), 4026–4042.  
<https://doi.org/10.1002/2015WR018209>
- Henn, B., Newman, A. J., Livneh, B., Daly, C., & Lundquist, J. D. (2018). An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology*, 556, 1205–1219. <https://doi.org/10.1016/j.jhydrol.2017.03.008>
- Hocking, R. R. (1976). A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics*, 32(1), 1. <https://doi.org/10.2307/2529336>
- Hou, Y.-K., Chen, H., Xu, C.-Y., Chen, J., Guo, S.-L., Hou, Y.-K., ... Guo, S.-L. (2017). Coupling a Markov Chain and Support Vector Machine for At-Site Downscaling of Daily Precipitation. *Journal of Hydrometeorology*, 18(9), 2385–2406.  
<https://doi.org/10.1175/JHM-D-16-0130.1>
- Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research*, 48(7).  
<https://doi.org/10.1029/2011WR011291>
- Joshi, D., St-Hilaire, A., Ouara, T., & Daigle, A. (2015). Statistical downscaling of precipitation and temperature using sparse Bayesian learning, multiple linear regression and genetic programming frameworks. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, 40(4), 392–408.  
<https://doi.org/10.1080/07011784.2015.1089191>

- Krabbenhoft, D. P., Bowser, C. J., Anderson, M. P., & Valley, J. W. (1990). Estimating groundwater exchange with lakes: 1. The stable isotope mass balance method. *Water Resources Research*, 26(10), 2445–2453. <https://doi.org/10.1029/WR026i010p02445>
- Krabbenhoft, D. P., & Webster, K. E. (1995). Transient Hydrogeological Controls on the Chemistry of a Seepage Lake. *Water Resources Research*, 31(9), 2295–2305. <https://doi.org/10.1029/95WR01582>
- Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 48(4). <https://doi.org/10.1029/2011WR011453>
- Le Roux, R., Katurji, M., Zawar-Reza, P., Quénot, H., & Sturman, A. (2018). Comparison of statistical and dynamical downscaling results from the WRF model. *Environmental Modelling & Software*, 100, 67–73. <https://doi.org/10.1016/J.ENVSOFT.2017.11.002>
- Luthcke, S. B., Rowlands, D. D., Lemoine, F. G., Klosko, S. M., Chinn, D., & McCarthy, J. J. (2006). Monthly spherical harmonic gravity field solutions determined from GRACE inter-satellite range-rate data alone. *Geophysical Research Letters*, 33(2), L02402. <https://doi.org/10.1029/2005GL024846>
- Mason, L., Bartlett, P., Baxter, J., & Frean, M. (n.d.). *Boosting Algorithms as Gradient Descent*.
- Miro, M., & Famiglietti, J. (2018). Downscaling GRACE Remote Sensing Datasets to High-Resolution Groundwater Storage Change Maps of California's Central Valley. *Remote Sensing*, 10(1), 143. <https://doi.org/10.3390/rs10010143>

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., ...

Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, 109(D7). <https://doi.org/10.1029/2003JD003823>@10.1002/(ISSN)2169-8996.GCIP3

Mo, K. C., Chen, L.-C., Shukla, S., Bohn, T. J., & Lettenmaier, D. P. (2012). Uncertainties in North American Land Data Assimilation Systems over the Contiguous United States. *Journal of Hydrometeorology*, 13(3), 996–1009. <https://doi.org/10.1175/JHM-D-11-0132.1>

MODIS/Aqua Land-Surface Temperature/Emissivity Monthly Global 0.05Deg CMG - LAADS DAAC. (n.d.). Retrieved November 15, 2019, from <https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MYD11C3/>

Mohaghegi, S., Del Valle, Y., Venayagamoorthy, G. K., & Harley, R. G. (2005). A comparison of PSO and backpropagation for training RBF neural networks for identification of a power system with statcom. *Proceedings - 2005 IEEE Swarm Intelligence Symposium, SIS 2005*, 2005, 391–394. <https://doi.org/10.1109/SIS.2005.1501646>

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)

- Othman, A., Sultan, M., Becker, R., Alsefry, S., Alharbi, T., Gebremichael, E., ...
- Abdelmohsen, K. (2018). Use of Geophysical and Remote Sensing Data for Assessment of Aquifer Depletion and Related Land Deformation. *Surveys in Geophysics*, 39(3), 543–566. <https://doi.org/10.1007/s10712-017-9458-7>
- Rheaume, S. J. (1991). Hydrologic provinces of Michigan. In *Water-Resources Investigations Report*. <https://doi.org/10.3133/wri914120>
- Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. *Nature*, 557(7707), 651–659. <https://doi.org/10.1038/s41586-018-0123-1>
- Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., ... Toll, D. (2004). The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, 85(3), 381–394. <https://doi.org/10.1175/BAMS-85-3-381>
- Rodell, Matthew, Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, 460(7258), 999–1002. <https://doi.org/10.1038/nature08238>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Sahoo, A. K., De Lannoy, G. J. M., Reichle, R. H., & Houser, P. R. (2013). Assimilation and downscaling of satellite observed soil moisture over the Little River Experimental Watershed in Georgia, USA. *Advances in Water Resources*, 52, 19–33. <https://doi.org/10.1016/J.ADVWATRES.2012.08.007>

- Sahour, H., Sultan, M., Vazifedan, M., Abdelmohsen, K., Karki, S., Yellich, J. A., Gebremichael, E., Alshehri, F. and Elbayoumi, T.M. (2020). Statistical applications to downscale GRACE-derived terrestrial water storage data and to fill temporal gaps. *Remote Sensing*, 12(3), 533.
- Save, H., Bettadpur, S., & Tapley, B. D. (2012). Reducing errors in the GRACE gravity solutions using regularization. *Journal of Geodesy*, 86(9), 695–711.  
<https://doi.org/10.1007/s00190-012-0548-5>
- Save, H., Bettadpur, S., & Tapley, B. D. (2016). High-resolution CSR GRACE RL05 mascons. *Journal of Geophysical Research: Solid Earth*, 121(10), 7547–7569.  
<https://doi.org/10.1002/2016JB013007>
- Scanlon, B. R., Longuevergne, L., & Long, D. (2012). Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. *Water Resources Research*, 48(4). <https://doi.org/10.1029/2011WR011312>
- Scanlon, Bridget R., Zhang, Z., Save, H., Sun, A. Y., Schmied, H. M., Van Beek, L. P. H., ... Bierkens, M. F. P. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences of the United States of America*, 115(6), E1080–E1089.  
<https://doi.org/10.1073/pnas.1704665115>
- Scanlon, Bridget R., Zhang, Z., Save, H., Wiese, D. N., Landerer, F. W., Long, D., ... Chen, J. (2016). Global evaluation of new GRACE mascon products for hydrologic applications. *Water Resources Research*, 52(12), 9412–9429.  
<https://doi.org/10.1002/2016WR019494>
- Schoof, J. T. (2013a). Statistical Downscaling in Climatology. *Geography Compass*, 7(4), 249–265. <https://doi.org/10.1111/gec3.12036>

- Schoof, J. T. (2013b). Statistical Downscaling in Climatology. *Geography Compass*, 7(4), 249–265. <https://doi.org/10.1111/gec3.12036>
- Seyoum, W., Kwon, D., & Milewski, A. (2019). Downscaling GRACE TWSA Data into High-Resolution Groundwater Level Anomaly Using Machine Learning-Based Models in a Glacial Aquifer System. *Remote Sensing*, 11(7), 824. <https://doi.org/10.3390/rs11070824>
- Shokri, A., Walker, J. P., Dijk, A. I. J. M., & Pauwels, V. R. N. (2019). On the Use of Adaptive Ensemble Kalman Filtering to Mitigate Error Misspecifications in GRACE Data Assimilation. *Water Resources Research*, 55(9), 7622–7637. <https://doi.org/10.1029/2018WR024670>
- So, B.-J., Kim, J.-Y., Kwon, H.-H., & Lima, C. H. R. (2017). Stochastic extreme downscaling model for an assessment of changes in rainfall intensity-duration-frequency curves over South Korea using multiple regional climate models. *Journal of Hydrology*, 553, 321–337. <https://doi.org/10.1016/J.JHYDROL.2017.07.061>
- Sultan, M., Sturchio, N. C., Alsefry, S., Emil, M. K., Ahmed, M., Abdelmohsen, K., ... Chouinard, K. (2019). Assessment of age, origin, and sustainability of fossil aquifers: A geochemical and remote sensing-based approach. *Journal of Hydrology*, 576, 325–341. <https://doi.org/10.1016/J.JHYDROL.2019.06.017>
- Sun, A. Y. (2013). Predicting groundwater level changes using GRACE data. *Water Resources Research*, 49(9), 5900–5912. <https://doi.org/10.1002/wrcr.20421>

- Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. In *Science* (Vol. 305).  
<https://doi.org/10.1126/science.1099192>
- Tayyebi, A., Smidt, S., & Pijanowski, B. (2017). Long-Term Land Cover Data for the Lower Peninsula of Michigan, 2010–2050. *Data*, 2(2), 16. <https://doi.org/10.3390/data2020016>
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411–423.
- Van Leeuwen, W. J. D., Orr, B. J., Marsh, S. E., & Herrmann, S. M. (2006). Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications. *Remote Sensing of Environment*, 100(1), 67–81.  
<https://doi.org/10.1016/j.rse.2005.10.002>
- Voss, K. A., Famiglietti, J. S., Lo, M., de Linage, C., Rodell, M., & Swenson, S. C. (2013). Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water Resources Research*, 49(2), 904–914. <https://doi.org/10.1002/wrcr.20078>
- Vu, M. T., Aribarg, T., Supratid, S., Raghavan, S. V., & Liong, S.-Y. (2016). Statistical downscaling rainfall using artificial neural network: significantly wetter Bangkok? *Theoretical and Applied Climatology*, 126(3–4), 453–467.  
<https://doi.org/10.1007/s00704-015-1580-1>



- Vugrinovich, R. (n.d.). *PATTERNS OF REGIONAL SUBSURFACE FLUID MOVEMENT IN THE MICHIGAN BASIN*. Retrieved from [https://www.michigan.gov/documents/deq/GIMDL-OFR866\\_302614\\_7.pdf](https://www.michigan.gov/documents/deq/GIMDL-OFR866_302614_7.pdf)
- Wahr, J., Swenson, S., Zlotnicki, V., & Velicogna, I. (2004). Time-variable gravity from GRACE: First results. *Geophysical Research Letters*, *31*(11), n/a-n/a. <https://doi.org/10.1029/2004GL019779>
- Wang, W., Liang, S., & Meyers, T. (2008). Validating MODIS land surface temperature products using long-term nighttime ground measurements. *Remote Sensing of Environment*, *112*(3), 623–635. <https://doi.org/10.1016/j.rse.2007.05.024>
- Watkins, M. M., Wiese, D. N., Yuan, D. N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research B: Solid Earth*, *120*(4), 2648–2671. <https://doi.org/10.1002/2014JB011547>
- Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., ... Lohmann, D. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research: Atmospheres*, *117*(D3), n/a-n/a. <https://doi.org/10.1029/2011JD016051>
- Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., ... Mocko, D. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and

application of model products. *Journal of Geophysical Research: Atmospheres*, 117(D3), n/a-n/a. <https://doi.org/10.1029/2011JD016048>

Xia, Y., Sheffield, J., Ek, M. B., Dong, J., Chaney, N., Wei, H., ... Wood, E. F. (2014).

Evaluation of multi-model simulated soil moisture in NLDAS-2. *Journal of Hydrology*, 512, 107–125. <https://doi.org/10.1016/j.jhydrol.2014.02.027>

Xu, T., Guo, Z., Xia, Y., Ferreira, V. G., Liu, S., Wang, K., ... Zhao, C. (2019). Evaluation of twelve evapotranspiration products from machine learning, remote sensing and land surface models over conterminous United States. *Journal of Hydrology*, 578.

<https://doi.org/10.1016/j.jhydrol.2019.124105>

Wiese D, Argus D, Yuan D, Landerer F. (2019). Combining satellite gravimetry and in-situ

GNSS measurements to improve spatial resolution of mass flux estimates. GRACE

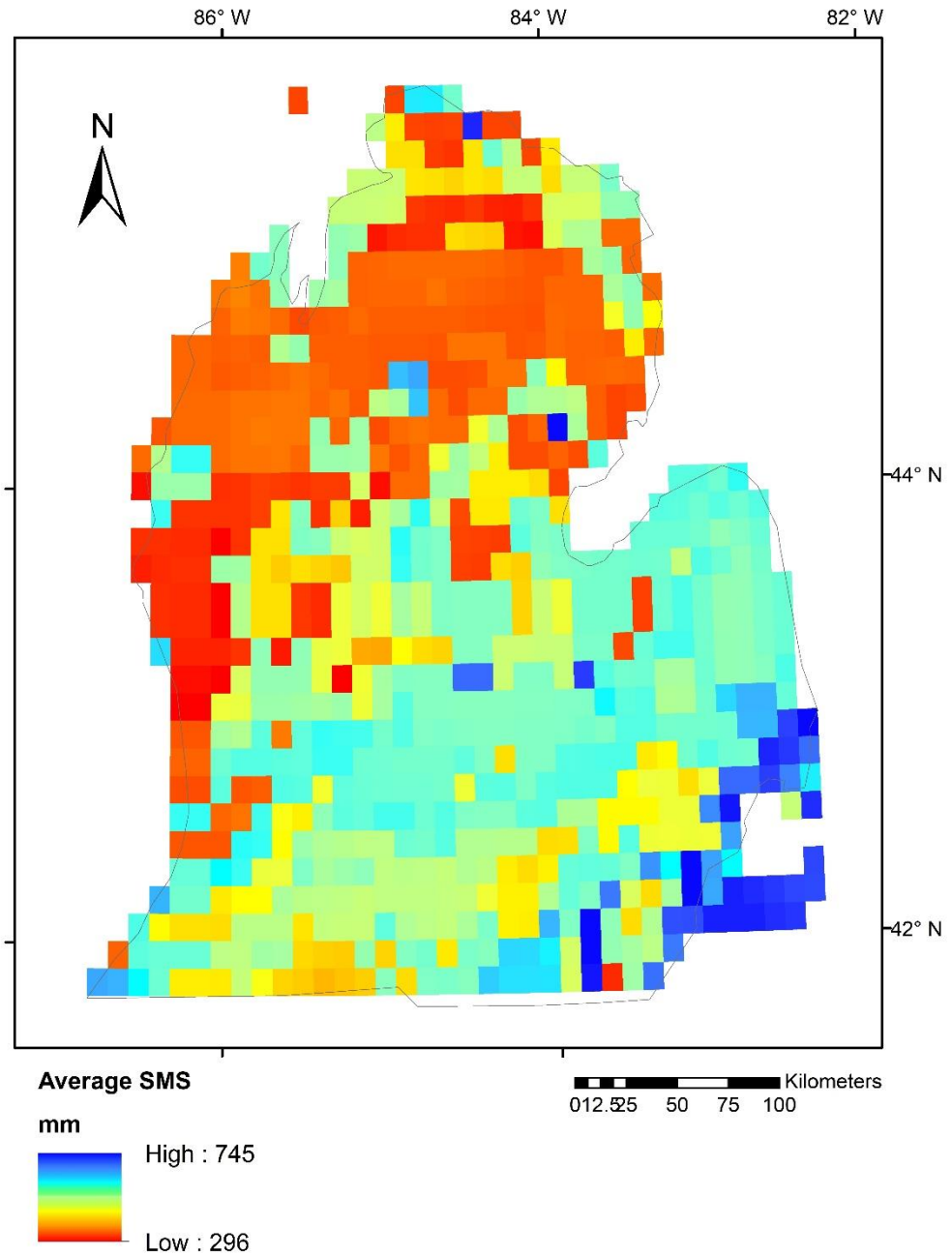
Science Team Meeting (GSTM), California Institute of Technology, Pasadena, CA, Oct. 2019

Zaitchik, B. F., Rodell, M., Reichle, R. H., Zaitchik, B. F., Rodell, M., & Reichle, R. H.

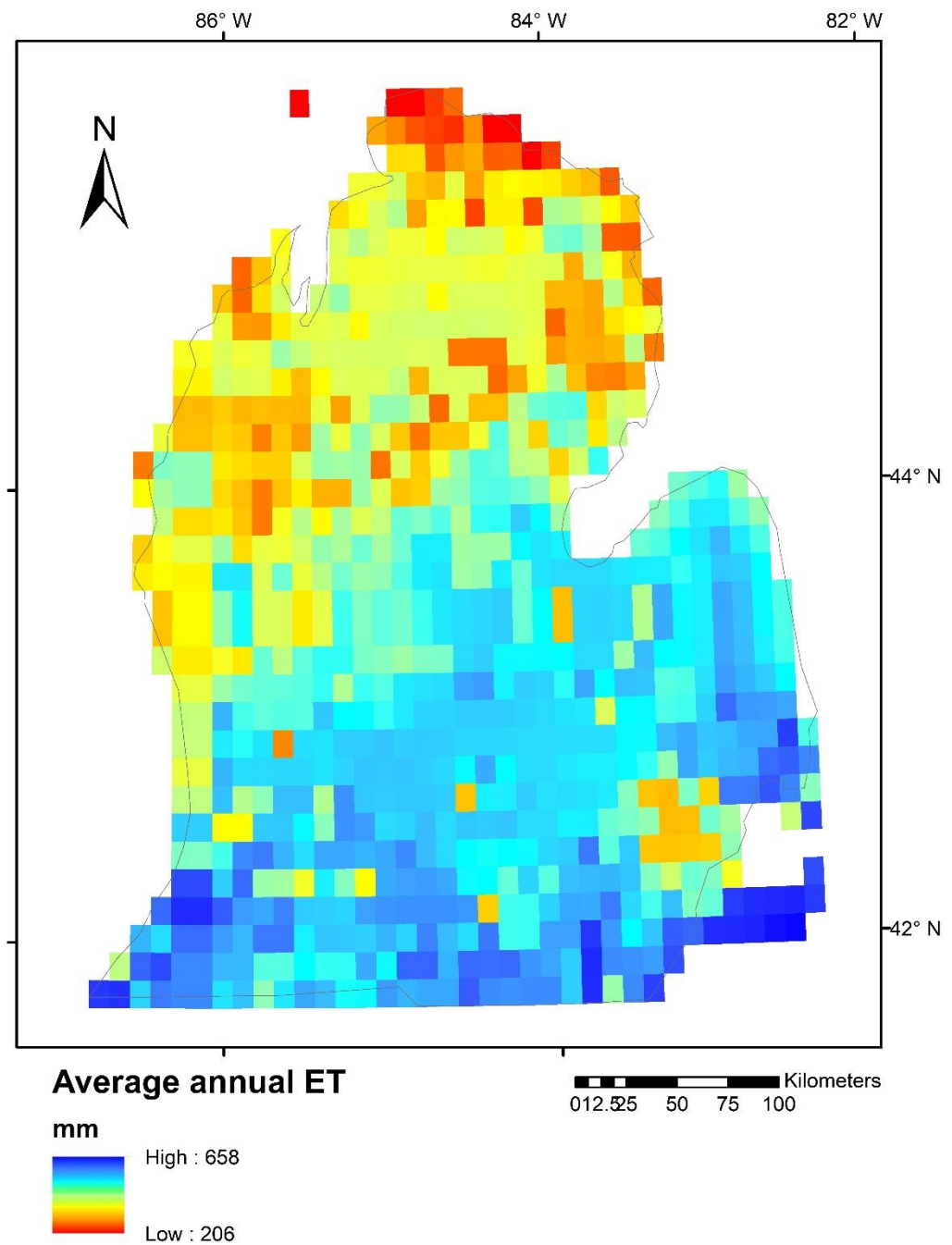
(2008). Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin. *Journal of Hydrometeorology*, 9(3), 535–548. <https://doi.org/10.1175/2007JHM951.1>

## APPENDIX

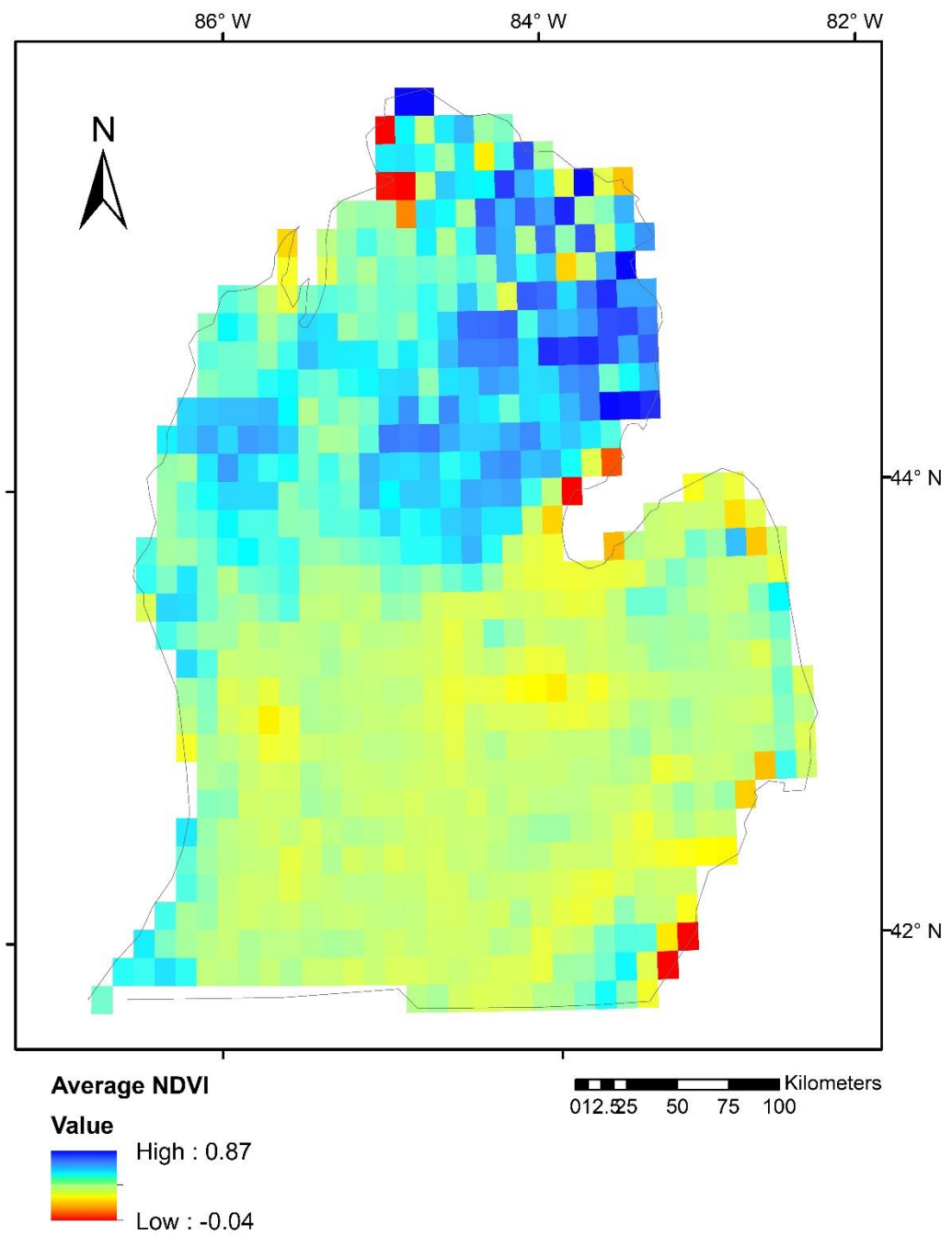
Average values of the variables



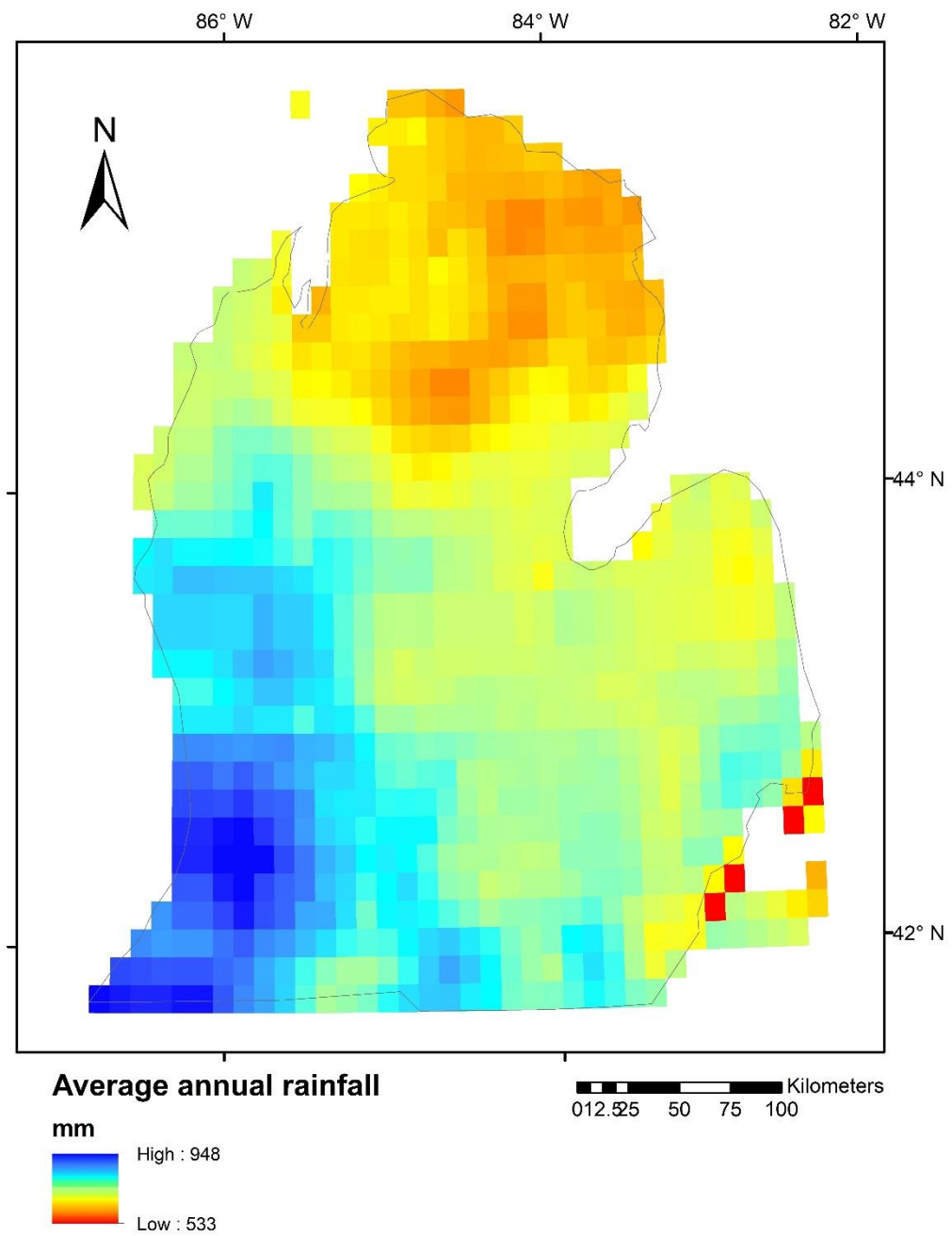
### 1. Average soil moisture



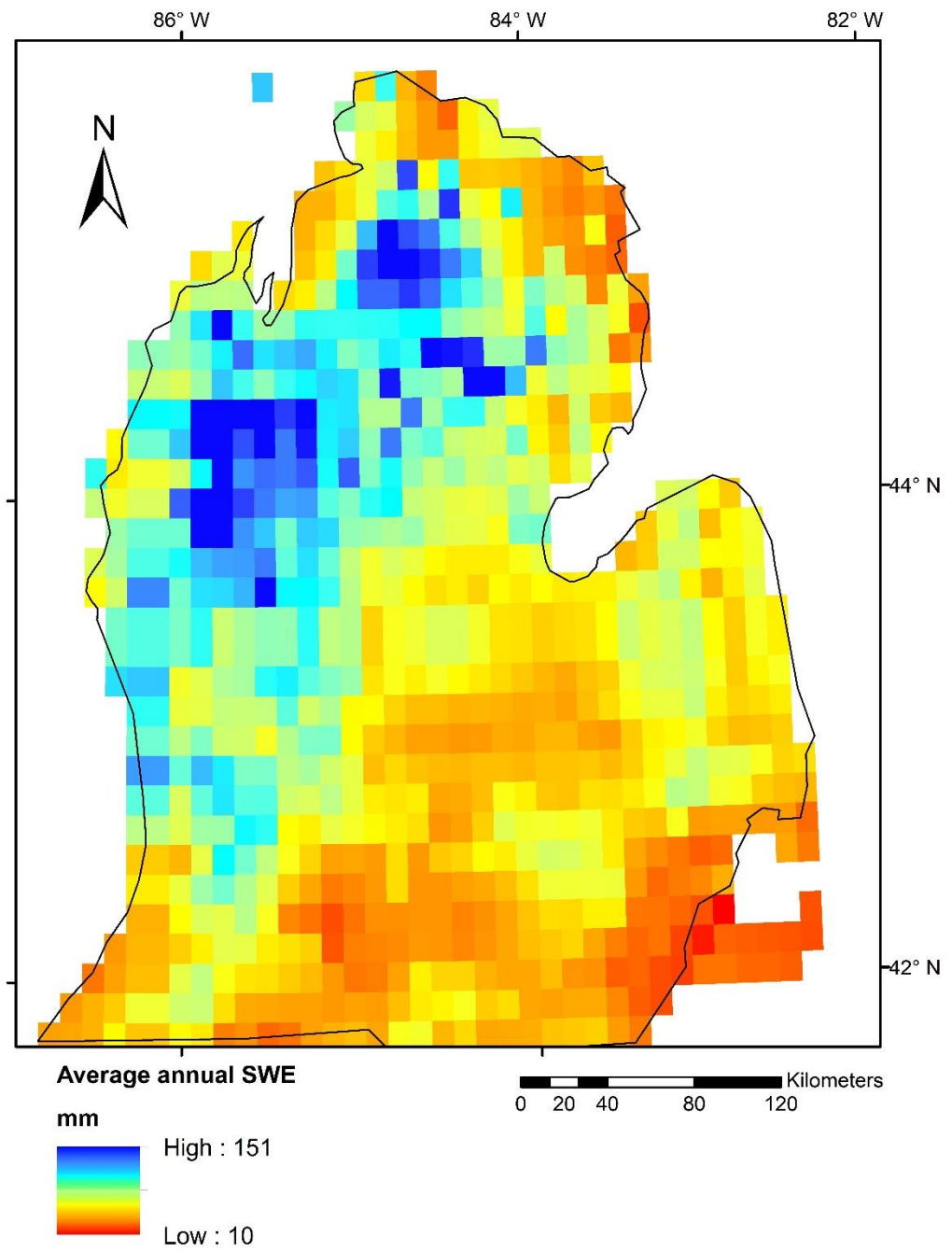
## 2. Average annual ET



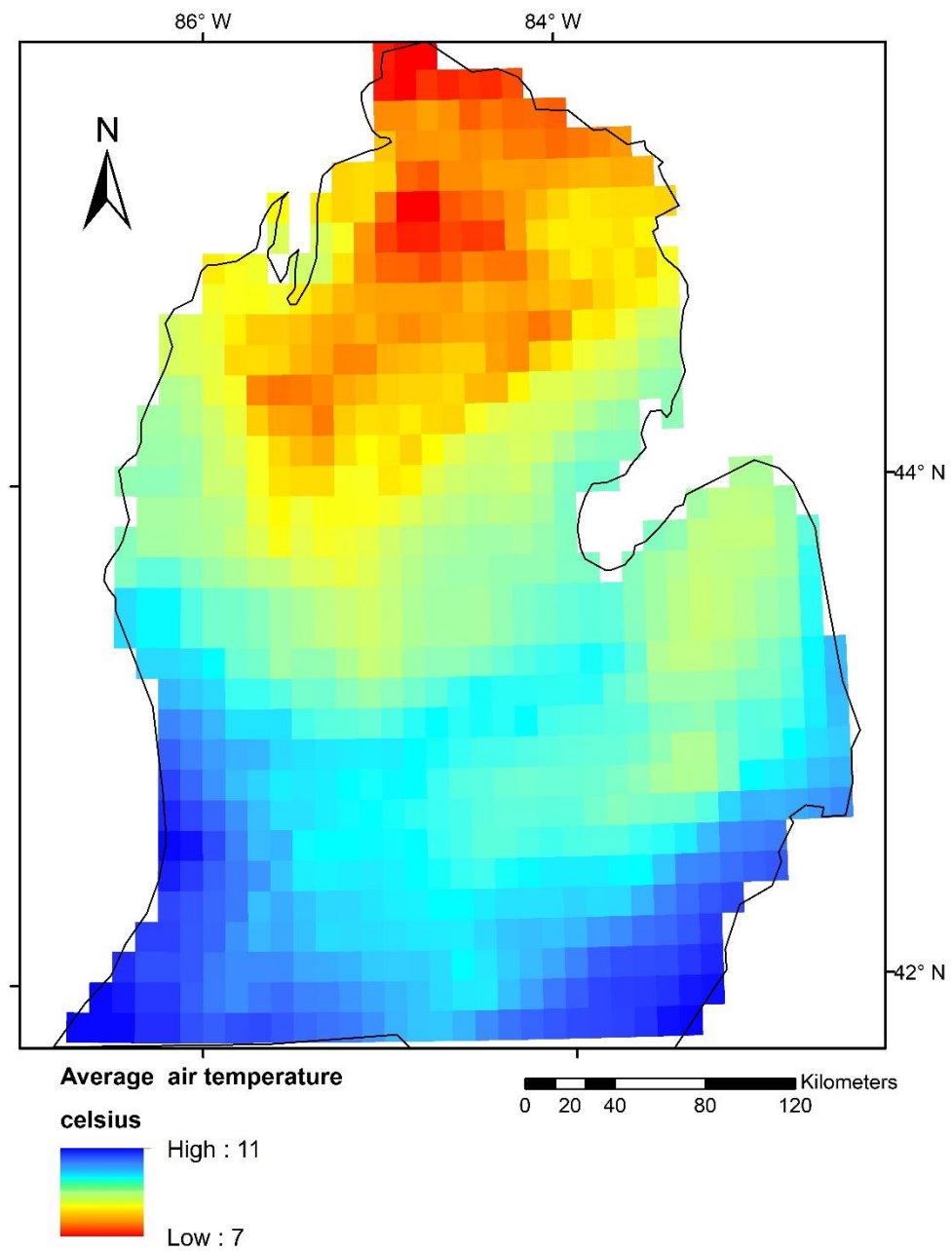
### 3. Average NDVI



#### 4. Average rainfall

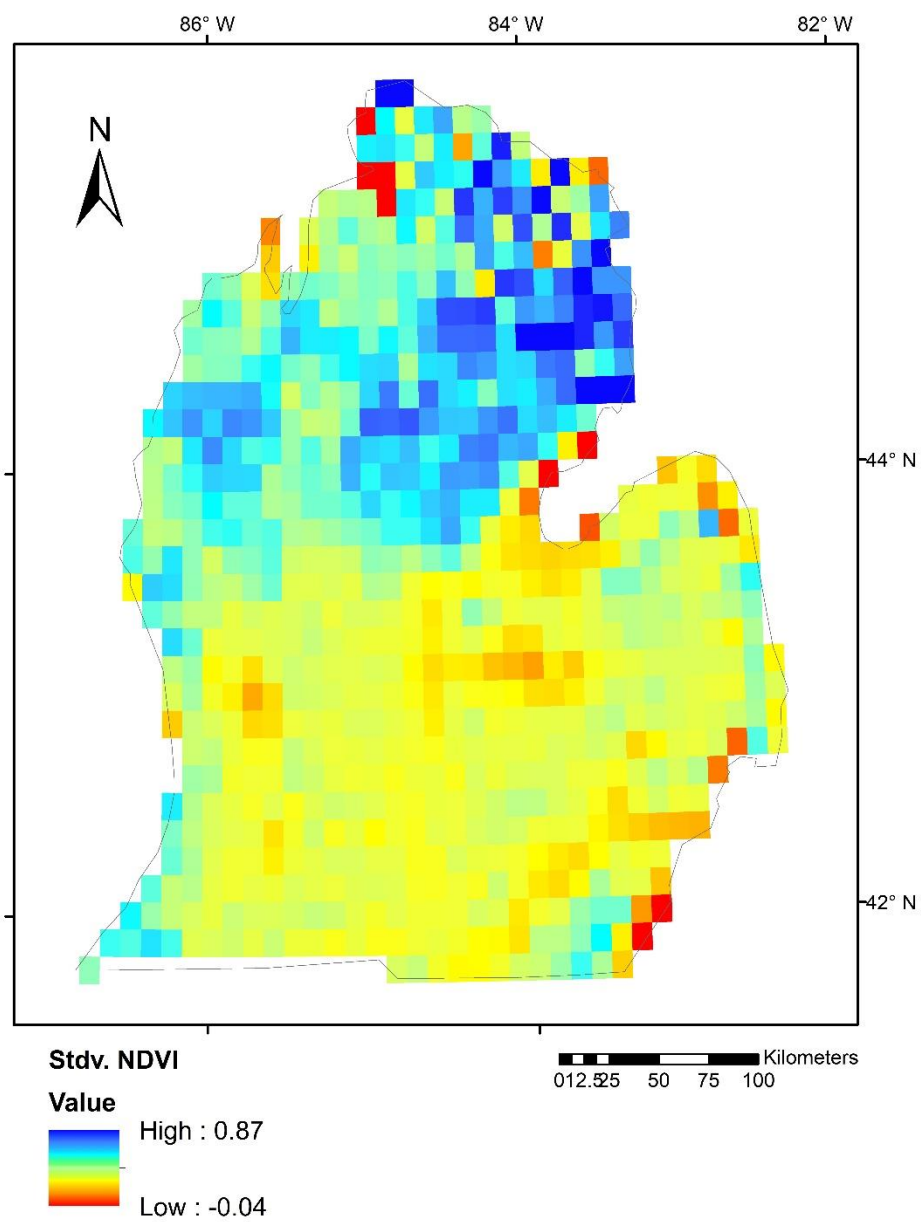


5. Average annual snow water equivalent (SWE)

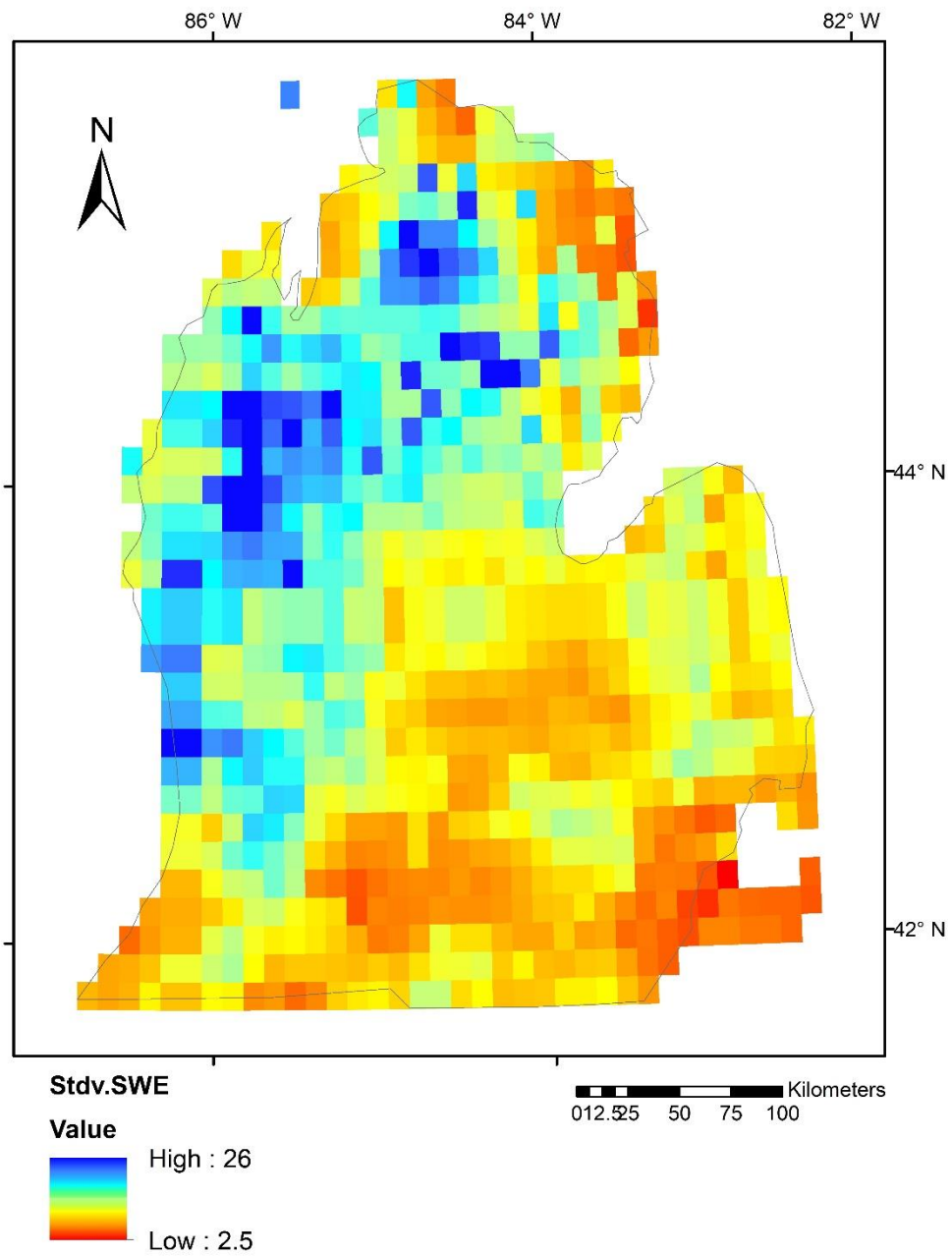


## 6. Average monthly air temperature

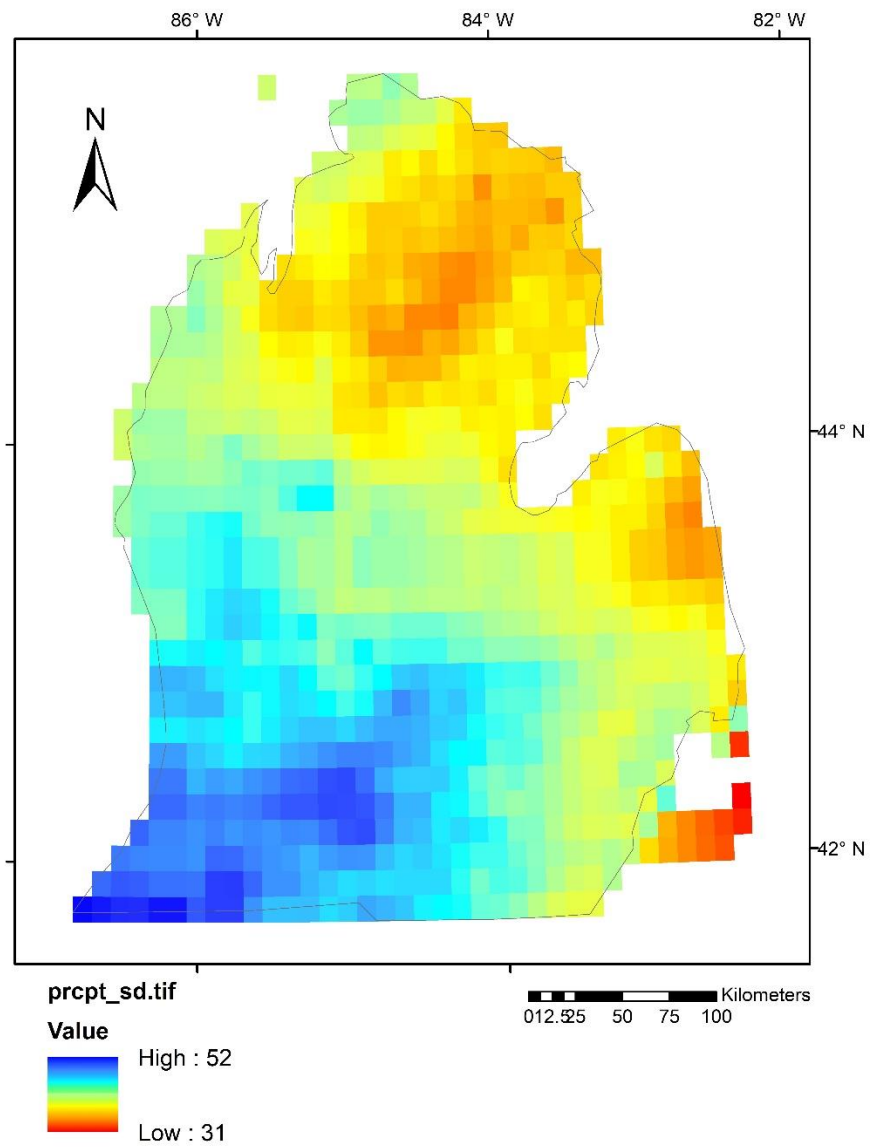




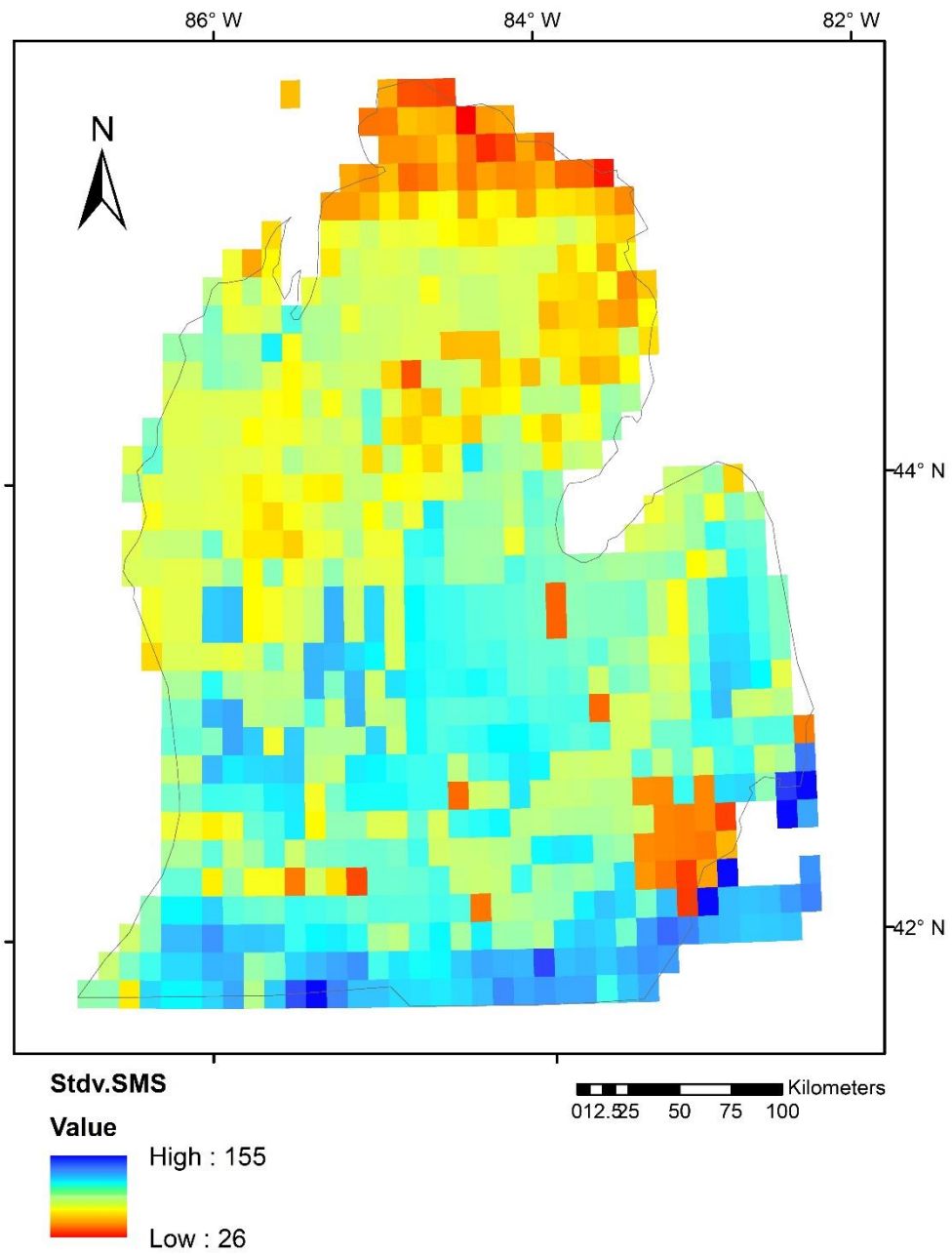
## 7. Standard deviation of monthly NDVI



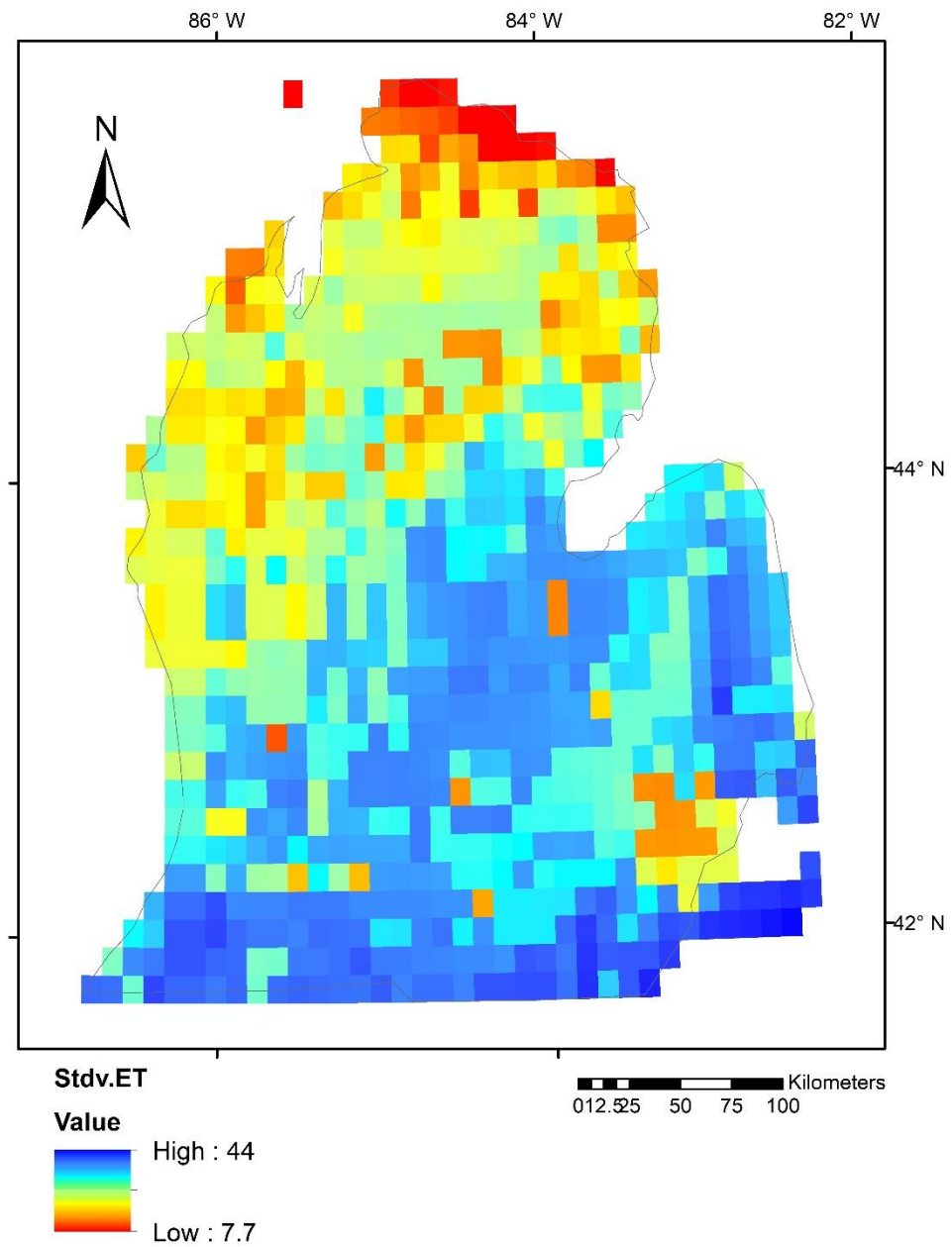
8. Standard deviation of monthly snow water equivalent



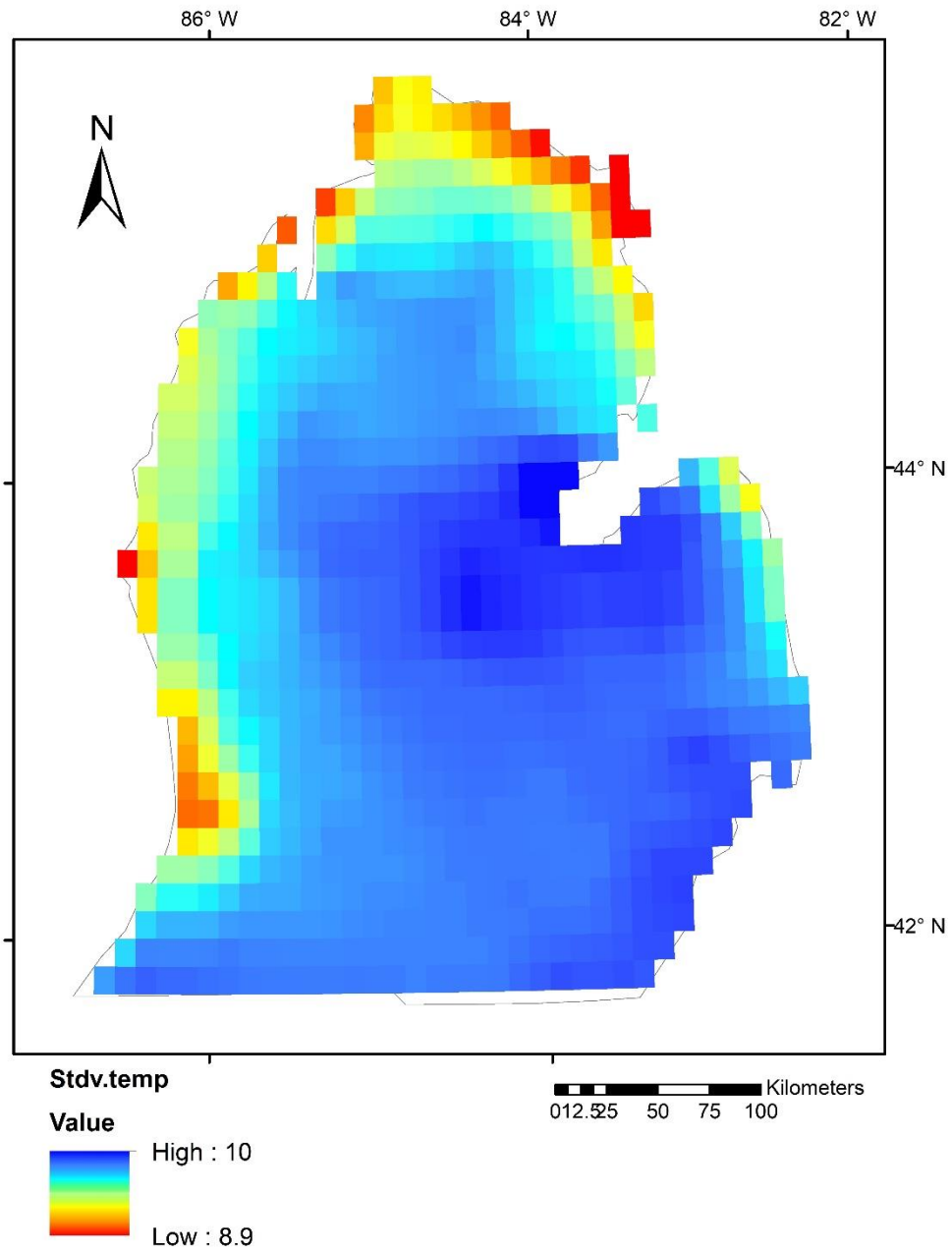
## 9. Standard deviation of monthly precipitation



10. Standard deviation of monthly soil moisture



11. Standard deviation of monthly evapotranspiration



12. Standard deviation of monthly temperature