AN ABSTRACT OF THE DISSERTATION OF

Amir Javaheri for the degree of Doctor of Philosophy in Civil Engineering presented on November 18, 2016.

Title: Assimilation of Multi-Sensor Data into Numerical Hydrodynamic Models of Inland Water Bodies.

Abstract approved:

_____

Meghna Babbar-Sebens

Numerical models are effective tools for simulating complex physical processes such as hydrodynamic and water quality processes in aquatic systems. The accuracy of the model is dependent on multiple model parameters and variables that need to be calibrated and regularly updated to reproduce changing aquatic conditions accurately. Multi-sensor water temperature observations, such as remote sensing data and in situ monitoring technologies, can improve model accuracy by providing benefits of individual monitoring technology to the model updating process. In contrast to in-situ temperature sensors, remote sensing technologies (e.g., satellites) provide the benefit of collecting measurements with better X-Y spatial coverage. However, the temporal resolution of satellite data is limited comparing to in-situ measurements. Numerical models and all source of observations have large uncertainty coming from different sources such as errors of approximation and truncation, uncertain model inputs, error in measuring devices and etc. Data assimilation (DA) is able to sequentially update the model state variables by considering the uncertainty in model and observations and estimate the model states and outputs more accurately. Data Assimilation has been proposed for multiple water resources studies that require rapid employment of incoming observations to update and improve accuracy of operational prediction models. The usefulness of DA approaches in assimilating water temperature observations from different types of monitoring technologies (e.g., remote sensing and

in-situ sensors) into numerical models of in-land water bodies (e.g., reservoirs, lakes, and rivers) has, however, received limited attention. Assimilating of water temperature measurements from satellites can introduce biases in the updated numerical model of water bodies because the physical region represented by these measurements do not directly correspond with the numerical model's representation of the water column. The main research objective of this study is to efficiently assimilate multi-sensor water temperature data into the hydrodynamic model of water bodies in order to improve the model accuracy. Four specific objectives were addressed in this work to accomplish the overall goal:

- Objective 1: Propose a novel approach to address the representation challenge of model and measurements by coupling a skin temperature adjustment technique based on available air and in-situ water temperature observations, with an ensemble Kalman filter (EnKF) based data assimilation technique for reservoirs and lakes.

- Objective 2: Investigate whether assimilation of remotely sensed temperature observations using the proposed data fusion approach can improve model accuracy with respect to in-situ temperature observations as well as remote sensing data.

- Objective 3: Investigate a global sensitivity analysis tool that combines Latin-hypercube and one-factor-at-a-time sampling to investigate the most sensitive model inputs and parameters in calculating the water age and water temperature of shallow rivers.

- Objective 4: Propose an efficient data assimilation framework to take the advantage of both monitoring technologies (e.g., remote sensing and in-situ measurements) to improve the model efficiency of shallow rivers.

Results showed that the proposed adjustment approach used in this study for four-dimensional analysis of a reservoir provides reasonably accurate surface layer and water column temperature forecasts, in spite of the use of a fairly small ensemble. Assimilation of adjusted remote sensing data using ensemble Kalman Filter technique improved the overall root mean square difference between modeled surface layer

temperatures and the adjusted remotely sensed skin temperature observations from 5.6 °C to 0.51 °C (i.e., 91% improvement). In addition, the overall error in the water column temperature predictions when compared with in-situ observations also decreased from 1.95 °C (before assimilation) to 1.42 °C (after assimilation), thereby, giving a 27% improvement in errors. In contrast, doing data assimilation without the proposed temperature adjustment would have increased this error to 1.98 °C (i.e., 1.5% deterioration). The most effective parameters to calculate water temperature were investigated and perturbed among the acceptable range to create the ensembles. Results show that water temperature is more sensitive to inflow temperature, air temperature, solar radiation, wind speed, flow rate, and wet bulb temperature respectively. Results also show in contrast to in-situ data assimilation, remote sensing data assimilation was able to effectively improve the spatial error of the model. Assimilation of in-situ observation improved the model efficiency at observation site. However, the model error increased by time and after less than two days, the model predictions of updated model were the same as base model before data assimilation. Hence, a maximum acceptable error between model and measurements was defined based on the application of model. Remote sensing data were assimilated into the model as they become available to improve the model accuracy for the entire river. In-situ data were also assimilated into the model when the error between model and observations exceeds the maximum error. Results showed that by assimilation of in-situ data one to three times per day, the average daily error reduced up to 58% comparing to situation that in-situ data were assimilated only once. In addition, the average spatial error reduced from 2.59 °C to 0.66 °C after assimilation of remote sensing data.

Assimilation of Multi-Sensor Data into Numerical Hydrodynamic Models of Inland
Water Bodies

by
Amir Javaheri

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented November 18, 2016
Commencement June 2017

Doctor of Philosophy dissertation of Amir Javaheri presented on November 18, 2016.

APPROVED:

_____

Major Professor, representing Civil Engineering

_____

Head of the School of Civil and Construction Engineering

_____

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

_____

Amir Javaheri, Author

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Professor Meghna Babbar-Sebens, for her longstanding support of my study and research. She has always been a great advisor and teacher. I would have never achieved this goal without her guidance. I would like to thank my committee members for all the supports and knowledge during the development of this research. Lastly, I would like to thank my amazing family for their love, unwavering support and constant encouragement.

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF TABLES

CHAPTER 1. Introduction

1.1. Problem statement

Water temperature is a crucial factor that affects water quality of in-land water bodies, such as lakes, reservoirs, and rivers, because of its significant impact on mixing processes and rate kinetics of multiple contaminants (Babbar-Sebens et al., 2013). Vertical stratification occurs because of water temperature difference. It also directly affects the vertical exchanges of mass, energy and momentum within the water column (Piccolroaz et al. 2013). Water temperature has been also identified as a critical factor that significantly affects the biological activity and metabolic rates of aquatic organisms living in the rivers or lakes. All aquatic species have a preferred water temperature range and if temperature exceeds too much from this range, major mortality will happen (Chu, 2003; Todar, 2012). The dissolved-oxygen in water is also highly affected by water temperature.

Sequential data assimilation methods can be used to update the initial condition of water quality numerical models as new observations become available (Evensen, 2007). The numerical model and measurements have large uncertainty coming from different sources such as errors of approximation and truncation, uncertain model inputs, error in measuring devices and etc. Ensemble Kalman Filter (EnKF), as a sequential data assimilation technique, is able to update the model state variables by considering the uncertainty in model and observations and estimate the states and outputs more accurately. EnKF has been successfully applied to various water resources problems such as studies that investigate the assimilation of temperature (Keppenne and Rienecker, 2003; Javaheri et al. 2016). Water temperature from different sources (e.g., satellite data and in-situ measurements) can be assimilated into model to update the state variables. However, assimilating multi-sensor water temperature measurements can introduce biases in the updated numerical model of water bodies because the physical region represented by these measurements do not directly correspond with the numerical model's representation of the water column. Also, it is not clear whether assimilation of data from a specific source would improve the model comparing to other measurements with different physical domain.

1.2. Goals and objectives

The overall goal of this research is to investigate how assimilation of multi-sensor data from different monitoring technologies would affect the model accuracy in numerical hydrodynamic models of in-land water bodies (e.g., rivers, lakes, reservoirs). To accomplish the overall goals, four specific objectives were addressed as follow:

- **Objective 1:** Proposes a novel approach to address the representation challenge of model and measurements by coupling a skin temperature adjustment technique based on available air and in-situ water temperature observations, with an ensemble Kalman filter (EnKF) based data assimilation technique.

- **Objective 2:** Investigate whether assimilation of remotely sensed temperature observations using the proposed data fusion approach can improve model accuracy with respect to in-situ temperature observations as well as remote sensing data.

- **Objective 3:** Investigate a global sensitivity analysis tool that combines latin-hypercube and one-factor-at-a-time sampling the most sensitive model inputs and parameters in calculating the water age and water temperature of shallow rivers.

- **Objective 4:** Propose an efficient data assimilation framework to take the advantage of both monitoring technologies (e.g., remote sensing and in-situ measurements) to improve the model efficiency of shallow rivers.

1.3. Outline

The organization of this research is as follow. Chapter 2 describes a novel strategy for adjusting remotely sensed skin temperature before it can be used for assimilation in numerical hydrodynamic model of reservoirs or lakes. Proposed methodology was applied to Eagle Creek Reservoir in central Indiana to investigate whether assimilation of adjusted remotely sensed temperature observations can also improve model accuracy with respect to in-situ temperature observations. Chapter 3 investigates a global sensitivity analysis to determine the most sensitive parameters in calculating water temperature and water age of shallow rivers. Results from this chapter, help to create the ensemble members more accurately for the next chapter. Chapter 4 uses the finding from previous chapter and presents a methodology to effectively assimilate the multi-

sensor observations into the hydrodynamic model of lower Klamath River to sequentially improve the model accuracy for the entire river. Finally, chapter 5 concludes the main findings from this research.

CHAPTER 2. From skin to bulk: An adjustment technique for assimilation of satellite-derived temperature observations in numerical models of small inland water bodies

Amir Javaheri[1], Meghna Babbar-Sebens[1] and Robert N. Miller[2]

[1]School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, USA.
[2]College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA.

2.1. Abstract

Data Assimilation (DA) has been proposed for multiple water resources studies that require rapid employment of incoming observations to update and improve accuracy of operational prediction models. The usefulness of DA approaches in assimilating water temperature observations from different types of monitoring technologies (e.g., remote sensing and in-situ sensors) into numerical models of in-land water bodies (e.g., lakes and reservoirs) has, however, received limited attention. In contrast to in-situ temperature sensors, remote sensing technologies (e.g., satellites) provide the benefit of collecting measurements with better X-Y spatial coverage. However, assimilating water temperature measurements from satellites can introduce biases in the updated numerical model of water bodies because the physical region represented by these measurements do not directly correspond with the numerical model's representation of the water column. This study proposes a novel approach to address this representation challenge by coupling a skin temperature adjustment technique based on available air and in-situ water temperature observations, with an ensemble Kalman filter based data assimilation technique. Additionally, the proposed approach used in this study for four-dimensional analysis of a reservoir provides reasonably accurate surface layer and water column temperature forecasts, in spite of the use of a fairly small ensemble. Application of the methodology on a test site - Eagle Creek Reservoir - in Central Indiana demonstrated that assimilation of remotely sensed skin temperature data using the proposed approach improved the overall root mean square difference between modeled surface layer temperatures and the adjusted remotely sensed skin temperature observations from 5.6 ^{o}C to 0.51 ^{o}C (i.e., 91% improvement). In addition, the overall error in the water column temperature predictions when compared with in-situ observations also decreased from 1.95 °C (before assimilation) to 1.42 °C (after assimilation), thereby, giving a 27% improvement in errors. In contrast, doing data assimilation without the proposed temperature adjustment would have increased this error to 1.98 °C (i.e., 1.5% deterioration).

Keywords

Data assimilation; Ensemble Kalman filter; Hydrodynamic model; In-situ data; Remote sensing data

2.2. Introduction

Management of water quality in in-land water bodies, such as lakes and reservoirs, is critical for minimizing human and ecological risks from contaminants. In the 1999 Drinking Water Infrastructure Needs Survey conducted by the Environmental Protection Agency (EPA), it was reported that an investment of $150 billion in drinking water systems over a 20-year period will be needed to ensure clean and safe drinking water [1]. Water column temperature is among the primary factors that affect water quality because of its significant impact on mixing processes and rate kinetics of multiple contaminants [2]. Vertical stratification caused by temperature affects vertical exchanges of mass, energy and momentum within the water column [3]. Hence, numerical water quality models that predict spatio-temporal changes in water column temperature have become critical tools for water managers operating lakes and reservoirs for human use and consumption [4-6]. However, the prediction efficiency of such models is dependent on multiple parameters and state variables that must be calibrated using all available data. While calibration techniques help improve the consistency between the model predictions and historic observations [7-11], they are not enough to ensure that a well-calibrated model will continue to predict accurately under operational and evolving conditions in water bodies.

For time dependent water quality numerical models, sequential data assimilation techniques can be used towards continued updating of an operational model's state variables, as and when new observations of the changing system become available [12]. A data assimilation process also has the potential to reduce uncertainty in prediction by integrating real-time observations from a variety of monitoring technologies [13]. However, using diverse data sources for updating water quality models requires an understanding of both strengths and limitations of individual data sources that may impact the assimilation process [2], [13]. For example, Satellites provide spatially-dense observations of skin temperature in water bodies. But they cannot be used to obtain sub-surface parameters, and have limited temporal resolution. On the other hand, in-situ sensors can provide multiple temporal observations at multiple sub-surface

depths to complement remote sensing data, even though in-situ sensors may not have a good X-Y coverage. With respect to remote sensing data, several studies in different fields have incorporated them in data assimilation. These include studies investigating the assimilation of temperature ([14-16]), aquatic contaminants ([17-21]), soil moisture ([22-26]), subsurface soil moisture ([27]), and snow cover ([28-31]). However, there are limited papers that report whether assimilation of remotely-sensed water surface temperature observations would produce updated models that are also accurate with respect to in-situ water column measurements. Babbar-Sebens et al. [2] recently used a variational data assimilation approach that used remotely sensed water temperature to correct the initial condition of the hydrodynamic model. Their results demonstrated that assimilation of remotely sensed data derived from Landsat-5 Thematic Mapper (TM) satellite reduced the overall error from 20.9% to 15.9%, when the model forecasts were compared to tests datasets derived from the same satellite. However, when they compared the water column temperature of model layers in the original and updated model with the in-situ measurements at different depths, it was found that the model error had actually increased by 50%, specifically from 1.8 $^\circ$C (before assimilation) to 2.7 $^\circ$C (after assimilation) [2]. While Babbar-Sebens et al. (2013) did not provide any solution to resolving this discrepancy between performance of updated models estimated by different data sources, their study highlighted the need for an adjustment method that would enable a direct comparison between skin temperature measurements obtained from satellites and the bulk temperatures of water column layers simulated by the numerical water quality model. The relationship between skin temperature and bulk temperature has been investigated by some previous studies [3], [32-35]. Hook et al. (2003) used four monitoring stations permanently deployed on Lake Tahoe, California–Nevada to compare the surface skin temperature and bulk temperature. They found that during the diurnal cycle, there is a noticeable difference between the bulk and skin temperatures which is related to strong solar radiation and low wind speeds at the site in the morning [33]. Donlon et al. (2002) using the remotely sensed sea surface temperature (SST) and high quality in-situ data obtained from radiometer systems considered the relationship between the SST_{skin}, the subsurface SST at depth (SST_{depth}), and the surface wind speed. They found that for wind speed lower than 6

m/s, there is a complicated relationship between SST_{skin} and SST_{depth} during the day because of the stratification of upper layers of ocean, while at night the skin layer is usually cooler. Also for wind speed higher than 6 m/s, there is a cool bias of -0.17 ± 0.07 K rms for both day and night conditions [32].

In this study, we have proposed a novel strategy for adjusting remotely sensed skin temperature collected in an operational setting. The adjustment enables incoming satellite-derived data to be converted into a representative bulk temperature of surface layer, before it can be used for assimilation in numerical hydrodynamic and water quality models. Additionally, coupling of the adjustment and assimilation techniques takes advantage of easily available daily air temperature, and the intermittently available in-situ observations that may or may not be collected on days and times when the remote sensing observations are obtained. The scientific merit of this work is that it provides an efficient adjustment and bias correction technique for extrapolating remotely sensed skin temperature observations to temperatures representing surface layers of numerical models that simulate in-land water bodies. Another scientific and practical merit is that the overall approach is able to produce an updated model with reasonably accurate forecasts in spite of the use of a small ensemble size during data assimilation. This is especially important for the purpose of high computational efficiency, when data assimilation is performed under operational settings.

The main research objective of this study is to investigate whether assimilation of remotely sensed temperature observations using the proposed data fusion approach can also improve model accuracy with respect to in-situ temperature observations. The proposed data fusion approach was tested using a hydrodynamic-temperature model of Eagle Creek Reservoir (ECR) in central Indiana, and involved the steps below:

(1)   Adjust the remotely sensed water skin temperature observations by using daily air temperatures and intermittently available in-situ water column temperatures, and, thereby, resolve the problem of bias arising due to sampling depths of remote sensing observations and in-situ measurements.

(2)   Assimilate the water temperature using adjusted remotely sensed water skin temperature from multi-spectral Landsat-5 TM band into the hydrodynamic-

temperature model, and update the model's initial conditions using an ensemble Kalman filter data assimilation framework.

(3) Compare the predictions from the model updated using the proposed methodology with incoming new in-situ observations to validate whether the updated model also produces more accurate sub-surface water column forecasts.

The remainder of the paper is organized in the following sections: Section 2.2 (Methodology) describes the study area and data collection, simulation model, sources of observations, relationship between remote sensing and in-situ observations and data assimilation algorithm, Section 2.3 describes the results of various experiments conducted in this study, and finally Section 2.4 provides concluding remarks.

2.3. Methodology

2.3.1. Study area and data collection

Eagle Creek Reservoir (ECR), located northwest of Indianapolis, Indiana, was used as the test site in this study (Figure 2.1). It lies within the Eagle Creek watershed and has a catchment area of about 419 km^2. The Eagle Creek Reservoir was originally constructed in 1967 for flood control, and is also used as the main drinking water supply for metropolitan Indianapolis and for recreational activities [36]. However, this reservoir tends to be in eutrophic to hypereutrophic states, with high concentration of phosphorous and blue algae [37]. The reservoir is small and shallow with total area of 5.1 km^2 at normal depth (240.79m above sea level) and mean depth of 5.7 m. The reservoir is divided into three main zones based on physical characteristics: the quarry (0.57 km^2), the northern basin (2.07 km^2), and the southern basin (2.97 km^2). The quarry has negligible interaction with the northern and southern basins. A land bridge separates the southern and northern basins, and water exchanges between them through an approximately 50 meter opening. The main streams flowing into the northern and southern basins of the reservoir include Eagle Creek, Bush Creek, Fishback Creek, and School Branch.

In this study, we used the hydrodynamic and temperature model built by Babbar-Sebens et al. (2013), described in more detail in Section 2.2.3. This model was developed using data obtained from different sources. Bathymetry data for the reservoir

and the water column temperature observations at multiple in-situ sites (see Figure 2.1 and Section 2.2.4 for details) were measured by Center of Earth and Environmental Sciences (CEES), Indiana University Purdue University Indianapolis (IUPUI) (http://www.cees.iupui.edu/). Hourly precipitation and air temperature data used for the model was obtained from National Climatic Data Center (http://www.ncdc.noaa.gov/) and Eagle Creek Airpark/Airport (53842) station, and the hourly solar radiation for 2008 was obtained from Indiana State Climate Office (http://climate.agry.purdue.edu/climate/). Inflows to the reservoir were estimated from the calibrated SWAT model of entire Eagle Creek Watershed (ECW) [38] and outflow measurements were obtained from a United States Geological Survey (USGS) gauge Station #03353460 at Clermont (1 km downstream from reservoir). Daily pool elevation data was obtained from USGS gauge Station #03353450 in the ECR. Since evaporation in ECR is unknown, measurements of a local water utility company at Carmel, 27 km (17 miles) east of ECR were used as evaporation rates.

Figure 2.1: Eagle Creek Reservoir. Left: aerial photograph showing the reservoir in green. The dark green feature at northeastern corner of the reservoir is the abandoned

quarry. Right: Gridded representation of the reservoir, with bottom topography shown in color. Units are meters above sea level. Black grid squares depict in-situ water quality observation sites and circles are monitoring regions encompassing the observation sites: 1- Dam region, 2 - Intake region, 3 - Marina region, and 4 - Eagle Creek inlet region.

2.3.2. Proposed approach for adjustment and assimilation of remotely-sensed temperature observations

Figure 2. 2 illustrates the overall components and the steps of the methodology proposed in this study. Orange trapezoids show the different types of observations used for the proposed assimilation methodology, green boxes display the process steps, and the blue ellipsoid shows the output. When using the method under operational conditions, Steps 1-11 in the method are repeated whenever new remote sensing data becomes available. The first set of steps – i.e., Steps 1-3 – are for calibrating the Air2Water model, which is then used to adjust the satellite-derived temperature observation to make them equivalent to EFDC numerical model's surface layer. Steps 4-7 are next conducted in order to complete the data assimilation procedures, and involve the use of Ensemble Kalman Filter technique and the adjusted remote sensing temperature measurements. After assimilation is over, Steps 8-10 are performed for validating the newly updated EFDC model with the new incoming in-situ observations, and for testing if assimilation with surface data also improved temperature predictions in the entire water column. Finally, in Step 11, it is tested if new remote sensing data are available, and re-starts the steps 1-11 whenever the condition is met. While, specific models and data used in the Methodology are described in this section, additional detail discussions on individual steps of the Methodology along with the results are described in the Results and Discussion section.

Since the study site used in this study was also used in previous work by Babbar-Sebens et al. (2013), it is worth noting the main differences between the previous study and this study. Specifically, unlike the previous work, in this study (1) the skin temperature was adjusted using the proposed approach to estimate an equivalent surface layer temperature that would be comparable to numerical model's surface layer predictions, (2) the ensemble Kalman filter approach which is more suitable for time

dependent problems has been implemented, and finally, (3) spatial bias from the model predictions was removed.

Figure 2. 2: Flowchart summarizing the methodological steps

2.3.3. EFDC Simulation model

The Environmental Fluid Dynamic Code (EFDC) [39], used in Steps 5, 7, and 9 of Figure 2. 2, and calibrated by Babbar-Sebens et al. (2013) for the Eagle Creek Watershed, was used in our study. EFDC is a widely used hydrodynamics models that has been applied to over 100 water bodies including rivers, lakes, reservoirs, wetlands, estuaries, and coastal ocean regions [40-44]. The EFDC model is able to solve the three dimensional, hydrostatic, free surface, turbulent averaged equations of motions for a variable density fluid. The EFDC uses dynamically coupled transport equations for turbulent kinetic energy, turbulent length scale, salinity and temperature, and supports stretched or sigma vertical coordinates and Cartesian or curvilinear orthogonal horizontal coordinates. The model is also able to solve an arbitrary number of Eulerian transport transformation equations for dissolved and suspended materials as well as complimentary Lagrangian particle transport transformation scheme. The continuity equation used by EFDC is given by

$$\frac{\partial H}{\partial t} + \frac{\partial (Hu)}{\partial x} + \frac{\partial (Hv)}{\partial y} + \frac{\partial w}{\partial z} = Q_H \qquad (2.1)$$

where H = water depth, u and v = horizontal velocity components in x and y direction respectively, w = vertical velocity component in z direction; QH = the volumetric source and sink term concerning evaporation and rainfall. The conservation of momentum equations are given as

$$\partial_t(Hu) + \partial_x(Huu) + \partial_y(Hvu) + \partial_z(wu) - fHv = -H\partial_x(g\xi + p + p_{atm}) +$$
$$(\partial_x h - z\partial_x H)\partial_z p + \partial_z(H^{-1}A_v\partial_z u) + Q_u \qquad (2.2)$$

$$\partial_t(Hv) + \partial_x(Huv) + \partial_y(Hvw) + \partial_z(wv) - fHu = -H\partial_y(g\xi + p + p_{atm}) +$$
$$(\partial_y h - z\partial_y H)\partial_z p + \partial_z(H^{-1}A_v\partial_z u) + Q_v \qquad (2.3)$$

where f = Coriolis factor, p = the water column hydro-static pressure; p_{atm} = the kinematic atmospheric pressure; A_v = vertical turbulent momentum diffusion coefficients, and Q_u and Q_v = momentum source-sink terms. The transport equations for temperature and salinity are

$$\partial_t(HT) + \partial_x(HuT) + \partial_y(HvT) + \partial_z(wT) = \partial_z(H^{-1}A_v\partial_z T) + Q_T \qquad (2.4)$$
$$\partial_t(HS) + \partial_x(HuS) + \partial_y(HvS) + \partial_z(wS) = \partial_z(H^{-1}A_v\partial_z S) + Q_S \qquad (2.5)$$

where Q_s and Q_T are horizontal diffusion and thermal sources and sinks. A number of alternatives are in place in the model to simulate general discharge control structures such as weirs, spillways and culverts [39].

The Courant-Friedrichs-levy condition (CFL condition) was used to test the suitability of grid sizes and time steps. The final grid (Figure 2.1) selected by Babbar-Sebens et al (2013) was built using Cartesian horizontal coordinate and sigma stretch vertical coordinate. It included expanding grids with minimum grid size 40 m to maximum 60 m and sigma scale of 0.2 for each layer. The expanding factor of 1.005 was used to expand grid sizes from the focal point (water intake). Time steps of the numerical model was chosen to be two seconds considering the model stability and the computational burden.

2.3.4. Temperature Data

One of the main objectives of this study is to assimilate remotely sensed water skin temperature into the hydrodynamic numerical model of in-land freshwater water body and to investigate the model accuracy with respect to in-situ temperature observations (i.e. Steps 1-4, 6-7, and 9-10). In order to achieve this goal, we used the remotely sensed water skin temperature data that Babbar-Sebens et al. (2013) had derived from the Landsat 5 TM satellite. In their study, in-situ measurements had been collected at 54 different (X, Y) locations (Figure 2.1) throughout the reservoir, and consisted of multiple measurements at several depths (starting from 25 cm below the surface to the bottom of reservoir) at each location. Four TM (band 6) thermal images were obtained from the satellite on dates Aug 7, Aug 23, Sept. 24, and Oct. 10 2008, and at local times 16:09, 16:08, 16:07, and 16:07 hours, respectively. The spectral radiance data from the satellites were converted into water skin temperature based on the Planck's law:

$$T(\lambda) = \frac{K_2}{ln\left[\frac{K_1}{L(\lambda)}+1\right]} \tag{2.6}$$

where λ is the wave length; K_1 and K_2 are the calibration constants as 607.76 Watts m^{-2} ster^{-1} μm^{-1} and 1260.56 Kelvin respectively; L is the spectral radiance in Watts m^{-2} ster^{-1} μm^{-1}. To assimilate data obtained from satellite, 300 random sampling locations in the reservoir were identified and then divided into two groups for training and testing. Out of these locations, 150 random locations were selected as training locations

to implement the ensemble Kalman filter data assimilation and remaining 150 locations were used for testing the efficiency of data assimilation process. In-situ observations are available at six times between August 14 and September 30 (as also shown in Table 2.1), at one location in each of the four monitoring regions (Figure 2.1). The instrumental error reported for the YSI probes that were used to measure the water column temperature at sampling locations has been reported to be $\pm0.15°C$ [2]. These in-situ measurements were used as unbiased references to adjust the remotely sensed skin data as well as the hydrodynamic model calibration and validation.

Table 2.1: Collection dates of remotely sensed and in-situ observations during the data assimilation period

	Jul 17	Jul 30	Aug 7	Aug 14	Aug 20	Aug 23	Aug 27	Sep 3	Sep 16	Sep 24	Sep 30	Oct 10	Oct 16
Remotely sensed data			✓			✓				✓		✓	
In-situ data	✓	✓		✓	✓		✓	✓	✓		✓		✓

2.3.5. Air2Water Model

Figure 2.3: Discrepancies between temperature observations from different sensors and temperature represented in the modeling domain. T_{air} represents air temperature collected at weather station, T_{skin} represents skin temperature derived from remote sensing measurements, $T_{depth\ n,m}$ represent the in-situ measurements taken at depth m in layer n, and $T_{EFDC\ n}$ represent the bulk temperature of the numerical model's n^{th} layer.

As discussed earlier, the numerical hydrodynamic model of the reservoir simulates the water column temperature at discrete depths (depending on the vertical discretization of the model) shown by the layer temperatures $T_{EFDC\ 1}$, $T_{EFDC\ 2}$, ... $T_{EFDC\ n}$ in Figure 2.3, whereas, remotely sensed observations used for data assimilation represent the skin temperature of the lake's surface (i.e. T_{skin} in Figure 2.3). Hence, to make an accurate comparison between the model outputs and observations, any existing bias would need to be removed and the remote sensing data would need to be converted into an equivalent bulk temperature value (i.e. which would be equivalent to $T_{EFDC\ 1}$) that represents the depth of surface layer of the numerical model. In-situ observations (i.e. $T_{depth\ n,m}$) are also usually collected at specific depths in the water column, and are not directly comparable to the remotely sensed skin temperature (T_{skin}) or the numerical

model's surface layer temperature ($T_{EFDC\ 1}$). Additionally, remotely sensed and in-situ data for most applications, including this study, may not be collected simultaneously. This makes it further complicated to relate the skin temperature and surface layer temperature of the reservoir. In this study, we used the method presented by Piccolroaz et al. (2013) to find the relationship between the water skin temperature and the air temperature above the reservoir (i.e., Steps 1-2). The skin temperature predicted by their method was then converted into a predicted surface layer bulk temperature (T_{adj}) by identifying the bias between the predicted skin temperature and a surface layer bulk temperature estimated from in-situ observations. We used the average of in-situ observations in the surface layer (i.e. $T_{depth\ 1,1}$ to $T_{depth\ 1,m}$) to estimate the surface layer's observed bulk temperature. Piccolroaz et al. (2013) developed a simple model called Air2Water to relate the temperature of lake skin (shown by T_{skin} in Figure 2.3) to air temperature (shown by T_{air} in Figure 2.3). This model is based on ordinary differential equation with at least 4 parameters (p_3, p_4, p_5 and p_6) as shown by Equation 2.7 and up to 8 parameters (not used in this study) that need to be calibrated using air and water temperature measurements. They suggested equations 2.7-2.9 that account for all the heat exchanges with the atmosphere and the deeper layer of the lake.

$$\frac{dT_{skin}}{dt} = \frac{1}{\delta}\left\{ p_1 \cos\left[2\pi\left(\frac{k}{k_{yr}} - p_2\right)\right] + p_3 + p_4(T_{air} - T_{skin}) + p_5 T_{skin}\right\} \tag{2.7}$$

$$\delta = exp\left(\frac{T_r - T_{skin}}{p_6}\right), \qquad T_{skin} \geq T_r \tag{2.8}$$

$$\delta = exp\left(\frac{T_{skin} - T_r}{p_7}\right) + exp\left(\frac{-T_{skin}}{p_8}\right), \qquad T_{skin} < T_r \tag{2.9}$$

where p_i are model parameters, k is time, k_{yr} is the duration of the year in suitable time units, T_{air} (Figure 2.3) is the air temperature (ºC), T_{skin} is the skin water temperature (ºC) and T_r represents the deep-water temperature and can be chosen depending on the thermal regime of the lake [3]. In the case of cold monomictic lakes, T_r is considered as the maximum surface temperature measured during the year. In the case of warm monomictic lakes, T_r is assumed to be the minimum surface temperature registered during the year. Finally, for the case of dimictic lakes, T_r can be assumed to be equal to the temperature of maximum density (4 ºC). The Eagle Creek reservoir can be considered as warm monomictic, so T_r is equal to minimum surface temperature registered during the year (2 ºC). Piccolroaz et al. (2013) also found that not all the

parameters in the Air2Water model are equally significant, and the number of parameters can be decreased from eight to four by eliminating the parameters *p1*, *p2*, *p7*, and *p8*. Piccolroaz et al. (2013) found that the first term in Equation 2.7 was insignificant compared to the second and third term, and hence this term (including parameters *p1* and *p2*) was omitted. Also, since equation 2.9 was not valid for Eagle Creek Reservoir, parameters *p7* and *p8* were deemed unnecessary. It is worth mentioning here that Air2Water model provided one estimate of water skin temperature for the entire reservoir, since only one air temperature T_{air} value was available for the entire X-Y domain of the reservoir. On the other hand, unique remote sensing observations were available for every spatial grid cell (Figure 2.1) in the reservoir, and in-situ observations were present at 4 X-Y locations (in the 4 monitoring regions of Figure 2.1). But since the standard deviations in spatial distribution of skin temperature derived from remote sensing measurements and the surface layer bulk temperatures estimated from in-situ observations in the four regions were found to be small enough (See Table 2.2 for details), the lack of variability in T_{air} measurements across the X-Y domain of this small reservoir was considered to have insignificant effect on the proficiency of Air2Water model.

Table 2.2: Standard deviations in spatial observations of data used for Air2Water model (^{o}C)

	Jul 17	Jul 30	Aug 7	Aug 14	Aug 20	Aug 23	Aug 27	Sep 3	Sep 16	Sep 24	Sep 30	Oct 10	Oct 16
Remotely sensed data			0.4			0.34				0.28		0.35	
In-situ data	0.17	0.11		0.15	0.27		0.38	0.41	0.06		0.3		0.21

2.3.6. Ensemble Kalman Filter algorithm

The Kalman filter [45], a sequential data assimilation method (Figure 2.4), is a commonly used data assimilation technique that was initially developed to update the state variables of linear systems as new observations became available [46]. Although the classical Kalman filter has been used for nonlinear problems [47-49], it is more applicable for state estimation of linear system under Gaussian noise [50-51]. Extended Kalman filter (EKF) has been proposed for updating the model state variables of nonlinear problems by assuming that these state variables are differentiable. This technique is based on linearization about the current mean and covariance [52]. However, this method is not applicable for fully nonlinear systems and leads to unstable results [53]. Ensemble Kalman filter (EnKF), another data assimilation technique, was introduced by Evensen [53] to deal with non-linear models, when the assumption of linearity cannot be used. It overcomes these limitations by using Monte Carlo simulations of the probability distribution of the state. EnKF is widely used in high order and nonlinear models such as hydrological models and weather forecasting models, when large numbers of observations are available [54]. In this study, we used the Ensemble Kalman Filter as the data assimilation method because of the non-linearity in the EFDC numerical model. Multiple studies have also used this method in the past to update hydrologic and hydrodynamic models (e.g., [55-63]).

Data assimilation systems incorporate the model prediction and observations in the presence of random, zero-mean error. However, they can be affected by biases from different sources [64]. The bias generally occurs due to systematic errors rather than random errors. It may be caused by inaccurate data, by inaccuracies in the observation operators that are used to express the relationship between observations and model states, by the assimilation methodology, and by incorrect physical parameterizations or boundary conditions [64, 65]. Thus, the bias existing in the datasets and models should be removed before starting data assimilation algorithm. However, identifying sources of biases is not a trivial task, and is a subject of active research in the data assimilation field [64].

Figure 2.4: Schematic diagram of the sequential data assimilation system

The ensemble Kalman filter algorithm was used in Step 7 of the proposed methodology (Figure 2. 2) to assimilate adjusted remotely sensed skin temperature observations into the hydrodynamic model. In the ensemble Kalman Filter, the hydrodynamic numerical model is represented using Equation 2.10:

$$X_k = F(X_{k-1}, u_{k-1}) + w_k \tag{2. 10}$$

And measurements are represented using Equation 2.11:

$$Y_k = H(X_k) + v_k \tag{2. 11}$$

where X denotes the vector of state variables, Y denotes the vector of measurements, u_k is the input vector, w_k and v_k are stationary zero-mean white noises with covariance matrices of Q_k and R_k for process and observations respectively, F represents the nonlinear model, H is the linear transformation which relates the state variables to observations, and the subscript "k" denotes the time step. If an ensemble of n predicted state variables is available, X^f can be written as

$$X_k^f = \left(X_k^{f_1}, \ldots, X_k^{f_n} \right) \tag{2. 12}$$

where superscript "f_i" represent the i-th forecast ensemble member. The closest station reporting solar radiation is 63 miles northwest from ECW which increases the uncertainty in data. Hence, the solar radiation and water initial temperature were considered as the main sources of error in the EFDC model, and were used to generate the ensemble. The ensemble of initial temperature and solar radiation were created by perturbing these variables. The average of ensemble is defined by

$$\overline{X}_k^f = \frac{1}{n}\sum_{i=1}^{n} X_k^{fi} \tag{2.13}$$

Since true states are not known, we estimate them using the average of realizations in the ensemble. Then the error matrix can be estimated by

$$P_k^f = \frac{1}{n-1}\langle(X_k^f - \overline{X}_k)(X_k^f - \overline{X}_k)^T\rangle \tag{2.14}$$

The error matrix is then used to calculate the Kalman gain matrix by

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1} \tag{2.15}$$

The updated state vector (X^a) is taken to be a linear combination of the forecast and the observations. The observations should be treated as random variables to get consistent error propagation in the ensemble Kalman filter [66]. Hence, the actual measurements were used as reference and random noise with zero mean and covariance R was added to measurements. The updating equation is given by:

$$X_k^a = X_k^f + K_k(Y_k - H_k X_k^f) \tag{2.16}$$

Even though it is not needed for the updating step, it is straightforward to calculate the error covariance of the updated field by

$$P_k^a = (I - K_k H_k)P_k^f \tag{2.17}$$

2.3.7. Undersampling

The accuracy of ensemble Kalman filter is highly dependent on the ensemble size. However, due to computational costs especially in complex models, it is not optimal to generate large ensembles. On the other hand, if the number of ensemble members is relatively small, the system may be undersampled and the ensemble may not accurately estimate the statistics of the system. Undersampling causes three main problems such as inbreeding, filter divergence and spurious correlations [67]. Inbreeding happens when the error covariance is underestimated due to small ensemble size. Inbreeding is a possible reason of filter divergence and development of spurious correlations [68-69]. Filter divergence appears when state variables are not able to be accurately updated by the observations. If the covariance is underestimated then more weighting in the Kalman gain is given to the observations and less to the model predictions. Conversely,

if the covariance is overestimated, there is a large uncertainty in the model prediction, and so less weighting is given to model prediction in Kalman gain and more weighting is given to the observations [67]. Spurious correlations can occur when all the observations seem to effect each of the state variables [70], even when the state variables are physically far from the observations. As a consequence, the state variables may get updated by the observations that have no underlying physical relationship between them [67]. The most common methods for dealing with these problems are covariance inflation and covariance localization [69].

2.3.7.1 Covariance inflation

The purpose of inflation is to increase the underestimated covariance matrix. Equation 2.18 is used to inflate the deviation of the forecast error with respect to their mean [71]:

$$X_k^f \leftarrow r(X_k^f - \overline{X}_k^f) + \overline{X}_k^f \tag{2.18}$$

where $\leftarrow$ denotes the replacement of previous value and *r* is the inflation factor which is slightly greater than 1. The optimal inflation factor is related to ensemble size and may vary between 1.01 and 1.07 [69], and [72]. Several methods have been proposed to estimate the inflated forecast and observational error covariance matrices [73-74]. The adjusted forms of forecast and observational error covariance matrices are $\lambda_k P_k$ and $\mu_k R_k$ respectively. Wu et al. 2013 suggested to use the second-order least squares (SLS; Wang and Leblanc, 2008) to estimate λ_k and μ_k [75].

$$\lambda_k = \frac{\mathrm{Tr}(d_k^T H_k P_k H_k^T d_k)\mathrm{Tr}(R_k{}^2) - \mathrm{Tr}(d_k^T R_k d_k)\mathrm{Tr}(H_k P_k H_k^T R_k)}{\mathrm{Tr}(H_k P_k H_k^T H_k P_k H_k^T)\mathrm{Tr}(R_k{}^2) - \mathrm{Tr}(H_k P_k H_k^T R_k)2} \tag{2.19}$$

$$\mu_k = \frac{\mathrm{Tr}(H_k P_k H_k^T H_k P_k H_k^T)(d_k^T R_k d_k) - \mathrm{Tr}(d_k^T H_k P_k H_k^T d_k)\mathrm{Tr}(H_k P_k H_k^T R_k)}{\mathrm{Tr}(H_k P_k H_k^T H_k P_k H_k^T)\mathrm{Tr}(R_k{}^2) - \mathrm{Tr}(H_k P_k H_k^T R_k)2} \tag{2.20}$$

where $d_k \equiv Y_k - H\overline{X}_k^f$

Then, the updated states can be computed as

$$X_k^a = X_k^f + \lambda_k P_k^f H_k^T (H_k \lambda_k P_k^f H_k^T + \mu_k R_k)^{-1}(Y_k - H_k X_k^f) \tag{2.21}$$

2.3.7.2 Covariance localization

The process of trimming off the error covariance at a specified distance is called covariance localization [67-70]. It is achieved by applying the correlation function with

local support, r, to the covariance matrix by using Schur product. The correlation function is defined as [76]:

$$\rho = \begin{cases} -\frac{1}{4}\left(\frac{|z|}{c}\right)^5 + \frac{1}{2}\left(\frac{|z|}{c}\right)^4 + \frac{5}{8}\left(\frac{|z|}{c}\right)^3 - \frac{5}{3}\left(\frac{|z|}{c}\right)^2 + 1, & 0 \leq |z| \leq c, \\ \frac{1}{12}\left(\frac{|z|}{c}\right)^5 - \frac{1}{2}\left(\frac{|z|}{c}\right)^4 + \frac{5}{8}\left(\frac{|z|}{c}\right)^3 + \frac{5}{3}\left(\frac{|z|}{c}\right)^2 - 5\left(\frac{|z|}{c}\right) + 4 - \frac{2}{3}\left(\frac{c}{|z|}\right) & c \leq |z| \leq 2c, \\ 0, & 2c \leq |z|. \end{cases} \quad (2.22)$$

where z is the Euclidean distance between grid points in the model and c is a length scale that the correlation reduces to zero at a distance of $2c$. The length scale can be estimated as $c = \sqrt{\frac{10}{3}} l$, where l is any cut of length.

After localization, the Kalman gain can be written as:

$$K_k = (\rho_o P_k^f) H_k^T \left[H_k (\rho_o P_k^f) H_k^T + R_k \right]^{-1} \quad (2.23)$$

Which is approximately

$$K_k = \left[\rho_o (P_k^f H_k^T) \right] \left[\rho_o (H_k P_k^f H_k^T) + R_k \right]^{-1} \quad (2.24)$$

Finally, the updated state variables can be estimated by

$$X_k^a = X_k^f + \left[\rho_o (\lambda_k P_k^f H_k^T) \right] \left[\rho_o (H_k \lambda_k P_k^f H_k^T) + \mu_k R_k \right]^{-1} \left(Y_k - H_k X_k^f \right) \quad (2.25)$$

2.4. Results and discussion

2.4.1. Hydrodynamic Model Calibration

In this study, before the proposed adjustment-assimilation approach was used for the data available from August 1st to October 10th, 2008 time period, the EFDC model was calibrated using the in-situ data available in the period January 1st, 2008 to July 31st, 2008. We used calibrated model of Babbar-Sebens et al. (2013) that has root mean square errors of 0.029 meter and 1.3 °C in pool elevation and temperature, respectively. It took approximately 20 hours on Intel(R) Xeon(R) CPU E3-1240 V2 @ 3.40 GHz processors to run the numerical model with 2 second time steps for the entire year and execute the assimilation steps for the days when remote sensing observations were available.

2.4.2. Lake superficial layer temperature and air temperature relationship

Note that using the Air2Water model with even four parameters can be challenging when the availability of remote sensing data is highly limited. For example, on August 7 when only one remote sensing data was available, Air2Water model was calibrated using one observation and led to overfitting. However, as new observations become available, the robustness of Air2Water model calibration can be expected to increase because the Air2Water model is then calibrated using the new observations as well as the previous ones (Figure 2. 2, step 1). As also discussed earlier, only one air temperature observation was available for each time when the four remote sensing datasets were collected. Hence, for calibration purposes, an average of all the skin temperature observations at various X-Y locations of the reservoir and collected at the same time was calculated to represent T_{skin} in the Air2Water model for that time. Table 2.3 shows the calibrated parameters p_3 to p_6 for each time that Air2Water model was updated. The calibrated model was then used to predict water skin temperature from air temperature, including on the days when in-situ data were available. The blue curve in Figure 2.5 (a-d) shows the predicted daily water skin temperature estimated by the four Air2Water models calibrated at different times. The predicted skin temperature (T_{skin}) was then compared with the average (shown as purple triangles in Figure 2.5) of all in-situ observations in EFDC model's surface layer (i.e. average of $T_{depth1,1}$ to $T_{depth1,m}$ in Figure 2.3). Finally it was concluded that skin temperatures of the lake predicted by the four Air2Water models had values 3.1 ℃, 3.7 ℃, 3.3 ℃ and 3.3 ℃ cooler than the average bulk temperature of top layer at the same X-Y location. So, the biases in remotely sensed observations that were estimated at the different data assimilation times using the Air2Water model and the available in-situ observations were removed, before the actual data assimilation steps were conducted. The remotely sensed skin temperature observations were adjusted in the following manner (Figure 2. 2, step 4):

$$T_{adj_t} = T_{skin_t} + Bias_{observation_t} = T_{RS_t} + Bias_{observation_t} \qquad (2.\,26)$$

where T_{adj} is adjusted skin temperatures, T_{RS} is remotely sensed skin temperatures (or, T_{skin} in Figure 2.3), and t is time. The green curve in Figure 2.5 illustrates the adjusted

skin temperatures (T_{adj}) at each time when remote sensing data becomes available and the Air2Water model is re-calibrated.

Figure 2.5: Comparison of predicted water skin temperature by Air2Water model calibrated at four different times using all available remote sensing (RS) measurements at that time (solid blue line), air temperature (dashed orange line), RS measurements (red circles), average bulk temperature of top layer (violet triangles), and adjusted skin temperatures (solid green line).

Table 2.3: Estimated Air2Water model parameters

	7-Aug	23-Aug	24-Sep	10-Oct
p_3 [$^\circ$C d^{-1}]	2.07×10^{-2}	1.77×10^{-2}	2.09×10^{-2}	1.99×10^{-2}
p_4 [d^{-1}]	5.12×10^{-3}	4.88×10^{-3}	5.51×10^{-3}	5.62×10^{-3}
p_5 [d^{-1}]	-2.31×10^{-3}	-2.31×10^{-3}	-2.38×10^{-3}	-2.43×10^{-3}
p_6 [$^\circ$C]	4.5	4.3	4.4	4.5

2.4.3. Remote sensing data assimilation

The period August 1st to October 10th, 2008 was selected to investigate how model predictions are updated via remotely sensed data assimilation and how the model performs with respect to in-situ observations. The remote sensing data are available at

August 7th, August 23rd, September 24th, and October 10th, while in-situ observation were provided at different days (August 14th, August 20th, August 27th, September 3rd, September 16th, and September 30th). Therefore, errors were reported for the two types of observations on different days.

The calibrated EFDC model was set up to restart at August 1st. Then the model was run from day 213th (August 1st) to the first observation which is at day 219th at 16:00 (Figure 2. 2, step 5). The Air2Water model was also calibrated using the remote sensing observation at August 7th, and the bias from observation was removed by comparing the Air2Water model results with in-situ observations at July 16th and July 30th (as shown in Figure 2.5a). On comparing the adjusted observations and model outputs at this time, a spatial bias in errors was observed over the reservoir (Figure 2.6). This bias was, therefore, removed before implementing ensemble Kalman filter because the filter assumes a zero-mean white noise error [63]. Figure 2.6 illustrates the difference between model predictions and adjusted skin temperatures along the reservoir. Figure 2.6.a displays the bias in the model outputs for August 7th. It indicates that the error is increasing from north to south with a sudden decrease near the bridge. The bias was removed by a linear trend (Eq. 2.27) as shown in Figure 2.6.a.

$$Bias = -0.054(j - 1) \qquad 1 < j < 75 \qquad\qquad (2. 27.a)$$

$$Bias = -0.118(j - 76) \qquad 76 < j < 120 \qquad\qquad (2. 28.b)$$

where j (y direction in Figure 2.6) is grids coordinate. The average of error along the i-coordinate (x direction in Figure 2.6) between the model outputs and adjusted temperatures for the observations at August 23rd and September 24th was constant along the reservoir with a sudden change near the bridge. The errors were 3.3 °C and 1.2 °C at southern basin and 4.2 °C and 2.1 °C at northern basin for second and third observations respectively (Figure 2.6.b and 6.c). For the last observation at October 10th, the average error was 1.2 °C in southern and northern basins. These spatial biases were removed by subtracting the absolute errors from the model outputs (Figure 2. 2, step 6). The ensemble Kalman filter data assimilation was then applied using the unbiased EFDC outputs and adjusted remotely sensed skin temperatures (Figure 2. 2, step 7).

Figure 2.6: Left panel: i and j coordinates of reservoir grid. Right panel: Difference between bulk temperatures derived from EFDC and adjusted temperatures from satellite retrievals at 300 points on the lake at four different times, a) August 7, b) August 23, c) September 24, and d) October 10. Red lines show the average of error that reduced from the model outputs.

Figure 2.7.a, 8.a, 9.a, and 10.a compare the remotely sensed data before and after removing the bias from the observations versus model outputs for August 7th, August 23rd, September 24th, and October 10th respectively. The root mean square errors were improved by 3.13 °C for August 7th, 3.2 °C for August 23rd, 3 °C for September 24th, and 2.8 °C for October 10th. Figure 2.7.b, 8.b, 9.b, and 10.b present the model outputs before and after removing the spatially bias from the model outputs against unbiased remotely sensed observations. The root mean square errors for the days that remote sensing data are available were reduced by 1.86 °C, 3.08 °C, 1.12 °C, and 0.48 °C for Figure 2.7.a, 8.a, 9.a, and 10.a respectively. Figure 2.7.c, 8.c, 9.c, and 10.c illustrate the unbiased model outputs before and after remotely sensed data assimilation for the testing points comparing with unbiased remotely sensed observations. After ensemble

Kalman filter data assimilation, the RMSE was reduced for the first to the fourth observations by 0.21 °C, 0.4 °C, 0.66 °C, and 0.35 °C respectively. Table 2.4 presents the summary of root mean square errors (°C) for the remote sensing data assimilation frame work. The error values were reported for the calibrated model without removing biases and before data assimilation, after skin temperature adjustment, after removing the spatial bias from model outputs and after implementing the remotely sensed data assimilation for training and testing locations. Results show that remotely sensed data assimilation can reduce the error of training locations as well as testing locations.

Figure 2.7: Model outputs vs. remote sensing observations for training locations (a) before (red crosses), and after (green diamonds) removing bias from the observations, (b) before (green diamonds), and after (orange circles) removing spatially bias from the model outputs, and (c) before (orange circles), and after (blue triangles) remote sensing data assimilation for August 7, 2008.

Figure 2.8: Model outputs vs. remote sensing observations for training locations (a) before (red crosses), and after (green diamonds) removing bias from the observations, (b) before (green diamonds), and after (orange circles) removing spatially bias from the model outputs, and (c) before (orange circles), and after (blue triangles) remote sensing data assimilation for August 23, 2008.

Figure 2.9: Model outputs vs. remote sensing observations for training locations (a) before (red crosses), and after (green diamonds) removing bias from the observations, (b) before (green diamonds), and after (orange circles) removing spatially bias from the model outputs, and (c) before (orange circles), and after (blue triangles) remote sensing data assimilation for September 24, 2008.

Figure 2.10: Model outputs vs. remote sensing observations for training locations (a) before (red crosses), and after (green diamonds) removing bias from the observations, (b) before (green diamonds), and after (orange circles) removing spatially bias from the model outputs, and (c) before (orange circles), and after (blue triangles) remote sensing data assimilation for October 10, 2008.

Table 2.4: Summary of root mean square errors (°C) of EFDC model with respect to remote sensing derived temperature observations

Date	Original Calibrated Model[a]	Original Calibrated Model[b]	Adjusted Calibrated Model[c]	RSDA_TA[d]	RSDA_TA[e]
7 – Aug	5.73	2.6	0.74	0.43	0.61
23 – Aug	7.2	4	0.92	0.48	0.65
24 – Sep	5.32	2.31	1.19	0.51	0.70

10 - Oct	4.23	1.43	0.95	0.58	0.68
Mean error	5.62	2.59	0.95	0.51	0.66

Original Calibrated Model[a]: Raw remote sensing derived temperature observations vs. original calibrated model.

Original Calibrated Model[b]: Adjusted remote sensing derived temperature observations vs. original calibrated model

Adjusted Calibrated Model[c]: Adjusted remote sensing derived temperature observations vs. adjusted calibrated model obtained by removing spatial bias from original calibrated model

RSDA_TA[d]: Training error (using 150 training X-Y locations) of updated model after the proposed temperature adjustment and remote sensing data assimilation steps (Steps 1-7 in Figure 2. 2).

RSDA_TA[e]: Testing error (using 150 testing X-Y locations) of updated model after the proposed temperature adjustment and remote sensing data assimilation steps (Steps 1-7 in Figure 2. 2).

2.4.4. Effect of remotely sensed data assimilation on model accuracy with respect to in-situ temperature observations

After assimilating adjusted remotely sensed skin temperatures into EFDC model, the updated model was run until the times when new in-situ observations became available (Figure 2. 2, step 8 and 9). All new incoming in-situ data that were collected after the most recently assimilated remote sensing measurements, and before the next set of available remote sensing observations, were used to validate the newly updated model (Steps 8, 9, and 10 in Figure 2. 2). This process was continued until the last set of observations that were available (Figure 2.4). The errors were calculated between in-situ measurements and model outputs for each of the three scenarios where (a) data assimilation was not conducted, (b) data assimilation was conducted but without any temperature adjustment, and (c) data assimilation was conducted with adjusted temperatures. The errors were calculated for the time period covering the data assimilation window (Figure 2. 2, step 10). Figure 2.11 compares the in-situ observations with the temperature vertical profiles predicted by models before and after data assimilation with adjusted satellite observations. The results in this figure illustrate temperature profiles for four different regions (Figure 2.1), and for one of the times

when in-situ data was collected after each remote sensing observation. The figure shows that data assimilation with adjusted satellite temperatures also improved the modeled water column temperatures, which the previous study by Babbar-Sebens et al. (2013) was not able to accomplish.

Table 2.5 lists the statistical summary of root mean square errors in water column temperatures at different X-Y locations and for the days that in-situ observations are available. The errors were calculated for model predictions before data assimilation (Before RSDA), after data assimilation with remotely sensed satellite observations that were not adjusted using Air2Water model (After RSDA), and after data assimilation with satellite observations that were adjusted using Air2Water model (RSDA_TA). The overall average of errors across time and space (last row in Table 2.5) show that data assimilation with unadjusted satellite-derived temperature data increased the overall error slightly from 1.95 °C to 1.98 °C. This is similar to the observation made by Babbar-Sebens et al. (2013) because even though the remotely sensed data assimilation significantly reduced the error of surface temperature in comparison to remote sensing observations (e.g., Table 2.4), the water column error increased from 1.95 °C to 1.98 °C (i.e. 1.5% increase) in contrast. As discussed earlier, the primary reason for this discrepancy in errors is the difference in depths represented by remote sensing and in-situ observations. However, after implementing the data assimilation with adjusted remotely sensed satellite observations (RSDA_TA), the water column temperature error had an overall decrease to 1.42 °C (i.e. 27% improvement).

The errors decreased for all the stations, except for stations measured on August 20th. Results in Figure 2.11 show that the original calibrated hydrodynamic model tended to predict water temperature warmer than the in-situ observations. The proposed coupled approach on temperature adjustment and data assimilation approach was able to generate models with updated initial conditions whose temperature values were lower than the original calibrated model. Hence, the updated initial conditions led to forecasted temperatures that were also cooler than what the original calibrated model would have forecasted. While this decreased the forecast error with respect to in-situ observations on most of the days, the error increased on August 20th. This could be attributed to multiple reasons, including the possibility that the model inputs for this

day were inaccurate. Moreover, the actual root mean square errors (Columns 4-6 in Table 2.5) are close in scale to the estimated $\sqrt{variance}$ (or, Estimated Error in Column 7 of Table 2.5) for all the stations. This indicates that the small ensemble size did not significantly deteriorate the performance of the data assimilation method.

The shaded rows in Table 2.5 show the stratified water column and white rows show the well-mixed layers. There is no observable consistency between the model error and the reservoir stratification. Additionally, on comparing these results with those reported by Babbar-Sebens et al. (2013), the reliability of proposed data assimilation framework is illustrated. The data assimilation framework used by Babbar-Sebens et al. (2013) showed that even though remotely sensed data assimilation improved the model prediction with respect to remote sensing data by 26%, the updated model's performance with respect to the in-situ observations in the water column worsened by 50 %. However, in present study, the model's performance was improved with respect to both remote sensing and in-situ observations by 91% (i.e., RMSE of 5.62 °C to 0.51 °C in Table 2.4) and 27% (i.e., 1.95 °C to 1.41 °C in Table 2.5), respectively.

Figure 2.11: Temperature vertical profiles at (a) August 14, (b) September 3, (c) September 90, and (d) October 16 in 4 different regions. Overall error metrics for other days and regions are presented in Table 2.5. The blue densely-dashed horizontal lines are the water surface elevation, the solid brown horizontal lines are the bottom elevation, the grey curves with circle markers are in-situ measurements, the red dashed curves with no markers are modeled before data assimilation, and green solid curves with no markers are modeled after data assimilation with adjusted remotely sensed observations.

Table 2.5: Statistical summary of remote sensing data assimilation (RSDA) before and after temperature adjustment (RSDA_TA) versus estimated square root of variance of water column temperatures averaged across depth. Shaded rows represent the stratified water column and white rows show the well-mixed water column. All errors are calculated with respect to in-situ observations in the vertical profile for a specific (I, J) location (see left panel of Figure 2.6).

Region ID (Figure 2.1)	Grid Coordinates (I, J)	Date	RMSE (°C)			
			Actual Error Before RSDA	Actual Error After RSDA	Actual Error After RSDA_TA	Estimated Error
1	(41,12)	14-Aug	2.0	1.6	1.0	1.0
2	(36,67)	14-Aug	1.5	2.2	1.0	1.1
3	(36,98)	14-Aug	2.2	1.1	1.4	1.1
4	(32,118)	14-Aug	1.9	2.0	1.4	1.1
1	(41,13)	20-Aug	1.4	3.3	1.6	1.0
2	(37,65)	20-Aug	0.8	3.2	1.3	1.1
3	(36,97)	20-Aug	1.5	3.5	1.9	1.0
4	(34,121)	20-Aug	1.3	2.6	1.5	1.2
1	(43,14)	27-Aug	1.6	3.4	1.5	1.0
2	(40,65)	27-Aug	2.3	2.2	1.7	1.1
3	(38,97)	27-Aug	2.7	1.1	1.6	1.0
4	(34,121)	27-Aug	4.7	2.2	2.7	1.2
1	(40,12)	3-Sep	0.4	3.7	0.3	1.0
2	(42,65)	3-Sep	1.9	2.1	1.7	1.1
3	(35,105)	3-Sep	2.1	2.5	1.9	1.2
4	(34,120)	3-Sep	2.8	2.6	2.5	1.1

1	(40,11)	16-Sep	1.8	2.1	1.2	1.0
2	(41,67)	16-Sep	1.8	0.7	0.8	1.1
3	(37,98)	16-Sep	1.6	0.3	1.4	1.1
4	(33,121)	16-Sep	2.7	1.7	2.2	1.1
1	(43,12)	30-Sep	1.7	1.9	0.7	1.0
2	(39,65)	30-Sep	1.8	1.0	1.0	1.1
3	(38,99)	30-Sep	2.5	0.7	1.8	1.1
4	(34,121)	30-Sep	3.7	2.2	2.2	1.2
1	(41,12)	16-Oct	1.9	1.5	1.3	1.0
2	(37,65)	16-Oct	1.8	1.8	1.3	1.2
3	(38,97)	16-Oct	1.2	1.0	0.9	1.1
4	(34,121)	16-Oct	0.4	0.8	0.3	0.8
Composite statistics (average of errors across time and space)			1.95	1.98	1.42	1.1

2.5. Conclusion

The limitations in using remotely sensed data assimilation in a numerical hydrodynamic model of inland reservoirs was investigated in this study. A data assimilation frame work was proposed to investigate how use of remotely-sensed inland lake temperatures in data assimilation will affect model accuracy with respect to in-situ temperature observations using ensemble Kalman filter. The results of the study highlighted the following findings:

a) The hydrodynamic model provides the water column temperature of the lake, while the remote sensing technology can only represent the skin temperature of water column. Hence, to make a consistent comparison between model outputs and remotely sensed data, we needed to find the relationship between water skin and bulk temperature of the reservoir. Air2Water model was implemented to

predict the water skin temperature of the reservoir. The predicted skin temperatures were calibrated using remotely sensed observations and then compared with in-situ measurements as new observations became available. Finally, the bias from remotely sensed temperatures was removed using the achieved relationship.

b) Comparing model outputs with adjusted remote sensing skin temperature showed the spatial biased over the reservoir for the days when remote sensing data were available. Due to assumption of zero-mean error in implementing ensemble Kalman filter, the spatial biases were removed by subtracting the absolute errors from the model outputs.

c) Among 300 sampling locations, 150 locations were used as training locations for estimating model errors during the data assimilation process and the other 150 locations were used as testing locations to estimate the model errors after assimilation was completed. After implementing ensemble Kalman filter, the model prediction was improved for the training locations as well as testing locations for all the observations.

d) The model predictions after remotely sensed data assimilation were compared with in-situ observation to investigate the model accuracy with respect to in-situ temperature observations. After data assimilation without temperature adjustment, the overall error worsened by 1.5%. However, after implementing data assimilation with temperature adjustment, the error was reduced by 27%.

2.6. Acknowledgement

2.7. References

[1] United States. (2009). Water on tap: What you need to know. Washington, D.C: U.S. Environmental Protection Agency, Office of Water.

[2] Babbar-Sebens, M., Li, L., Song, K., & Xie, S. (2013). On the Use of Landsat-5 TM Satellite for Assimilating Water Temperature Observations in 3D

Hydrodynamic Model of Small Inland Reservoir in Midwestern US. Advances in Remote Sensing, 2(3), 214-227.

[3]   Piccolroaz, S., Toffolon, M., & Majone, B. (2013). A simple lumped model to convert air temperature into surface water temperature in lakes. *Hydrology and Earth System Sciences*, 17(8), 3323-3338.

[4]   United States Environmental Protection Agency (USEPA), "Total Maximum Daily Load for Dissolved Oxygen and Nutrients to Mashapaug Pond," Rhode Island, 2002.

[5]   Jin, K. R., & Ji, Z. G. (2004). Case study: modeling of sediment transport and wind-wave impact in Lake Okeechobee. *Journal of hydraulic engineering*, 130(11), 1055-1067.

[6]   Tetra Tech, "Hydrodynamic and Water Quality Modeling Report for Lake Lanier," *Georgia*, 2009.

[7]   Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research*, 28(4), 1015-1031.

[8]   Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, 76(3), 501-521.

[9]   Duan, Q., H. V. Gupta, S. Sorooshian, A. N. Rousseau, and R. Turcotte (2003), Preface, in Calibration of Watershed Models, *Water Sci. Appl. Ser.*, vol. 6, edited by Q. Duan et al., p. v, AGU, Washington, D. C.

[10] Gupta V.K, Sorooshian S, Yapo PO. Towards improved calibration of hydrological models: multiple and noncomensurable measures of information. *Water Resour. Res.* 28(4), 1015-1031.

[11] Sorooshian, S., Duan, Q., & Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resources Research*, 29(4), 1185-1194.

[12] Evensen, G. (2007). Data assimilation (p. 279). New York: Springer.

[13] Moradkhani, H. (2008). Hydrologic remote sensing and land surface data assimilation. *Sensors*, 8(5), 2986-3004.

[14] Ezer, T., & Mellor, G. L. (1997). Simulations of the Atlantic Ocean with a free surface sigma coordinate ocean model. *Journal of Geophysical Research: Oceans* (1978–2012), 102(C7), 15647-15657.

[15] Keppenne, C. L., & Rienecker, M. M. (2003). Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. *Journal of Marine Systems*, 40, 363-380.

[16] Troccoli, A., & Haines, K. (1999). Use of the temperature-salinity relation in a data assimilation context. *Journal of Atmospheric and Oceanic Technology*, 16(12), 2011-2025.

[17] Seo, D. J., Koren, V., & Cajina, N. (2003). Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *Journal of Hydrometeorology*, 4(3), 627-641.

[18] Madsen, M. N., Martnack, J. N., Skotner, C., & Berg, A. (2006, September). Water quality surveillance and early warning in surface waters-integration of mathematical models and on-line monitoring. *In Proceedings of the 7th International Conference on Hydroinformatics* (pp. 1-23).

[19] Dorigo, W. A., Zurita-Milla, R., de Wit, A. J., Brazile, J., Singh, R., & Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International journal of applied earth observation and geoinformation*, 9(2), 165-193.

[20] Voutilainen, A., Pyhälahti, T., Kallio, K. Y., Pulliainen, J., Haario, H., & Kaipio, J. P. (2007). A filtering approach for estimating lake water quality from remote sensing data. *International journal of applied earth observation and geoinformation*, 9(1), 50-64.

[21] Mao, J. Q., Lee, J. H., & Choi, K. W. (2009). The extended Kalman filter for forecast of algal bloom dynamics. *Water research*, 43(17), 4214-4224.

[22] Ottlé, C., & Vidal-Madjar, D. (1994). Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-MOBILHY region. *Journal of Hydrology*, 158(3), 241-264.

[23] Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., & Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, 34(12), 3405-3420.

[24] Galantowicz, J. F., Entekhabi, D., & Njoku, E. G. (1999). Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. Geoscience and Remote Sensing, *IEEE Transactions* on, 37(4), 1860-1870.

[25] Pauwels, V. R., Hoeben, R., Verhoest, N. E., & De Troch, F. P. (2001). The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. *Journal of Hydrology*, 251(1), 88-102.

[26] Ines, A. V., Das, N. N., Hansen, J. W., & Njoku, E. G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. *Remote Sensing of Environment*, 138, 149-164.

[27] Olioso, A., Chauki, H., Courault, D., & Wigneron, J. P. (1999). Estimation of evapotranspiration and photosynthesis by assimilation of remote sensing data into SVAT models. *Remote Sensing of Environment*, 68(3), 341-356

[28] Slater, A. G., & Clark, M. P. (2006). Snow data assimilation via an ensemble Kalman filter. *Journal of Hydrometeorology*, 7(3), 478-493.

[29] Andreadis, K. M., & Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, 29(6), 872-886.

[30] Dong, J., Walker, J. P., & Houser, P. R. (2005). Factors affecting remotely sensed snow water equivalent uncertainty. *Remote Sensing of Environment*, 97(1), 68-82.

[31] Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., & Bayr, K. J. (2002). MODIS snow-cover products. *Remote sensing of Environment*, 83(1), 181-194.

[32] Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., & Murray, M. J. (2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, 15(4), 353-369.

[33] Hook, S. J., Prata, F. J., Alley, R. E., Abtahi, A., Richards, R. C., Schladow, S. G., & Pálmarsson, S. (2003). Retrieval of lake bulk and skin temperatures using Along-Track Scanning Radiometer (ATSR-2) data: A case study using Lake Tahoe, California. *Journal of Atmospheric and Oceanic Technology*, 20(4), 534-548.

[34] Jessup, A. T., & Branch, R. (2008). Integrated ocean skin and bulk temperature measurements using the Calibrated InfraRed In-situ Measurement System (CIRIMS) and through-hull ports. *Journal of Atmospheric and Oceanic Technology*, 25(4), 579-597.

[35] Hulley, G. C., Hook, S. J., & Schneider, P. (2011). Optimized split-window coefficients for deriving surface temperatures from inland water bodies. *Remote Sensing of Environment*, 115(12), 3758-3769.

[36] Lobugeois, F. (2009). Development of A 3D Hydrodynamic Model For Eagle Creek Reservoir. IUPUI 09/2009.

[37] Tedesco, L.P., Pascual, D.L., Shrake, L.K., Hall, R.E., Casey, L.R., Vidon, P.G.F., Hernly, F.V., Salazar, K.A., Barr, R.C., Ulmer, J., Pershing, D., (2005). Eagle Creek Watershed Management Plan: An Integrated Approach to Improved Water Quality. Eagle Creek Watershed Alliance, CEES Publication 2005-07, IUPUI, Indianapolis, 182p.

[38] Piemonti, A. D., Babbar-Sebens, M., & Jane Luzar, E. (2013). Optimizing conservation practices in watersheds: Do community preferences matter? *Water Resources Research*, 49(10), 6425-6449.

[39] Tetra Tech. (2007). The Environmental Fluid Dynamics Code, User Manual, US EPA Version 1.01.

[40] Hamrick, J. M., & Wu, T. S. (1997). Computational design and optimization of the EFDC/HEM3D surface water hydrodynamic and eutrophication models. *In Next generation environmental models and computational methods* (pp. 143-161). Society for Industrial and Applied Mathematics, Philadelphia, PA.

[41] Lu, S., and Davie, S.R., (2006). Charleston Harbor system 3-dimensional modeling, *Proceedings of the Ninth International Conference on Estuarine and Coastal Modeling*, Charleston, SC, pp. 13-31.

[42] Xia, M., Xie, L., And Pietrafesa, L.J., (2007). Modeling of the Cape Fear River Estuary Plume, *Estuaries and Coasts*, 30(4), 698-709.

[43] Liu, Z., Hashim, N. B., Kingery, W. L., Huddleston, D. H., & Xia, M. (2008). Hydrodynamic modeling of St. Louis Bay estuary and watershed using EFDC and HSPF. *Journal of Coastal Research*, 107-116.

[44] Zhou, J., Falconer, R. A., & Lin, B. (2014). Refinements to the EFDC model for predicting the hydro-environmental impacts of a barrage across the Severn Estuary. *Renewable Energy*, 62, 490-505.

[45] Kalman R. (1960). New approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1), 35-45.

[46] Maybeck, P. S. (1982). *Stochastic models, estimation, and control* (Vol. 3). Academic press.

[47] Beneš, V. E. (1981). Exact finite-dimensional filters for certain diffusions with nonlinear drift. Stochastics: *An International Journal of Probability and Stochastic Processes*, 5(1-2), 65-92.

[48] Daum, F. E. (1986). Exact finite-dimensional nonlinear filters. Automatic Control, *IEEE Transactions* on, 31(7), 616-622.

[49] Krener, A. J. & Duarte, A. (2004). A Hybrid Computational Approach to Nonlinear Estimation. *Proc. Conf. Dec. Contr.*, Kobe, Japan, pp. 1815-1819.

[50] Kitanidis, P. K., & Bras, R. L. (1980). Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. *Water Resources Research*. 16(6), 1025-1033.

[51] Kitanidis, P. K., & Bras, R. L. (1980). Real-time forecasting with a conceptual hydrologic model: 2. Applications and results. *Water Resources Research*, 16(6), 1034-1044.

[52] Gelb, A. (1974). *Applied optimal estimation*. MIT press.

[53] Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* (1978–2012), 99(C5), 10143-10162.

[54] Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., & Ridley, A. (2006). What is the ensemble Kalman filter and how well does it work? *In American Control Conference*, (pp. 6-pp). IEEE.

[55] Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28(2), 135-147.

[56] Moradkhani, H., Hsu, K. L., Gupta, H., & Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, 41(5).

[57] Miller, R. N., & Cane, M. A. (1989). A Kalman filter analysis of sea level height in the tropical Pacific. *Journal of Physical Oceanography,* 19*,* 773–790*.*

[58] Miller, R. N., Busalacchi, A. J., & Hackert, E. C. (1995). Sea surface topography fields of the tropical Pacific from data assimilation. *Journal of Geophysical Research: Oceans* (1978–2012), 100(C7), 13389-13425.

[59] Crow, W. T., & Wood, E. F. (2003). The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Advances in Water Resources*, 26(2), 137-149.

[60] Entekhabi, D., Nakamura, H., & Njoku, E. G. (1994). Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *Geoscience and Remote Sensing, IEEE Transactions on*, 32(2), 438-448.

[61] Gallus Jr, W. A., & Segal, M. (2000). Sensitivity of forecast rainfall in a Texas convective system to soil moisture and convective parameterization. *Weather and forecasting*, 15(5), 509-525.

[62] Li, X., Koike, T., & Pathmathevan, M. (2004). A very fast simulated re-annealing (VFSA) approach for land data assimilation. *Computers & Geosciences*, 30(3), 239-248.

[63] Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P., Njoku, E. G., & Owe, M. (2007). Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research: Atmospheres* (1984–2012), 112(D9).

[64] Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3323-3343.

[65] Dee, D. P., & Da Silva, A. M. (1998). Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 124(545), 269-295.

[66] Burgers, G., Jan van Leeuwen, P., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6), 1719-1724.

[67] Petrie, R. (2008). Localization in the ensemble Kalman Filter. Master's thesis, Univ. of Reading, Reading, UK.

[68] Houtekamer, P. L., & Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3), 796-811.

[69] Hamill, T. M., Whitaker, J. S., & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11), 2776-2790.

[70] Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, 129(12), 2884-2903.

[71] Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12), 2741-2758.

[72] Whitaker, J. S., & Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7), 1913-1924.

[73] Wang, X., & Bishop, C. H. (2003). A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the atmospheric sciences*, 60(9), 1140-1158.

[74] Miyoshi, T. (2011). The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Monthly Weather Review*, 139(5), 1519-1535.

[75] Wu, G., Zheng, X., Wang, L., Zhang, S., Liang, X., & Li, Y. (2013). A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation. *Quarterly Journal of the Royal Meteorological Society*, 139(672), 795-804.

[76] Gaspari, G., & Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554), 723-757.

CHAPTER 3. Global sensitivity analysis of water age and water temperature of shallow rivers

Amir Javaheri[1], Meghna Babbar-Sebens[1], Julie Alexander[2], Sascha Hallett[2], and Jerri Bartholomew[2]

[1]School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, USA.

[2]Department of Microbiology, Oregon State University, Corvallis, Oregon, USA.

3.1. Abstract

Many watersheds in the Pacific North West of North America are anthropogenically manipulated, with system-wide impacts. Identification of abiotic parameters that can be altered in managed rivers would inform associated management actions. Water age and water temperature are among the main factors that could help decision makers identify the required management actions to reduce fish diseases and other ecological impacts. We conducted a global sensitivity analysis of water temperature and water age of shallow rivers to a comprehensive set of hydraulics and meteorological parameters. We applied an analysis technique, which combined Latin-hypercube and one-at-a-time sampling method, to the hydrodynamic numerical model of the shallow Lower Klamath River in California, USA. This effective method requires a relatively small number of model runs and is able to globally assess the sensitivity of the model to all of the parameters. The response of water temperature and water age to the bathymetry data was also investigated using the Monte Carlo sampling method. We found that flow rate and bottom roughness are the most effective parameters in calculating the water age. Water temperature is more sensitive to inflow temperature, air temperature, solar radiation, wind speed, flow rate, and wet bulb temperature respectively. Although other parameters such as heat exchange parameters are less effective on river water temperature, they are important in model calibration.

Keywords

Latin-Hypercube; One at a time; Sensitivity analysis; Water temperature; Water age

3.2. Introduction

In the Pacific North West of North America, many major rivers are managed. In the Klamath River, which flows through Oregon and California to the Pacific Ocean, discharge is regulated by a series of reservoirs and dams. Changes to the natural hydrograph post-manipulation are associated with economic and ecological impacts including the health of native fishes.

Ceratonova shasta (*C. shasta*), a myxosporean parasite, is known as a significant source of mortality for salmonid fishes in the Klamath River. The parasite is fully

aquatic, with immotile, waterborne infectious stages that cycle through sessile invertebrate polychaete worms as well as migrating salmonids. The annual prevalence of *C. shasta*-infection in out-migrating juvenile Chinook salmon has been estimated at up to 84 percent in some years (True et al. 2016). Infection can cause significant mortality in juvenile salmonids when river temperatures reach 15-18° C in late spring and following low magnitude peak winter and spring flows (Bartholomew and Foott 2010; Hallett et al. 2012; Jordan 2012). Reduced risk of *C. shasta* infection in salmonids has been observed following high magnitude winter and spring discharge (Hallett et al. 2012; True et al. 2011). Low river water temperature is one mechanism that may explain the reduced risk of infection under these conditions because cooler water temperatures slow parasite proliferation in both hosts.

Releasing cooler water from the lowermost dam is generally considered as one of the main management actions to reduce downstream river water temperature. The threshold for parasite density that would causes high mortality of salmonids especially Chinook salmon is ~10 spores per liter (Hallett et al. 2012). Hence, releasing dam water also reduces the parasite density. Travel time is another important factor that has the potential to help decision makers identify the management actions to reduce the disease risk. Travel time will inform epidemiological models and monitoring efforts, addressing specific questions including, (i) how long parasites may be expected to stay in the system after dam release, (ii) when and how much water would need to be released from dam to achieve a certain decrease in water temperature and decrease in parasite density, (iii) when is the best time to collect water samples to document temperature and parasite density before, during and after a water release event.

In previous studies, atmospheric conditions have been reported to be significant drivers of river temperature (Edinger et al. 1974; Ward 1985; Stefan and Preud'homme 1993). Other researchers have also recognized flow discharge into the river as a significant driver of the river water temperature (Morse 1972). The water temperature is expected to be lower for higher flow rates (Grant 1977; Hockey et al. 1982). There are several studies that have evaluated the sensitivity of river temperature to hydraulic and meteorological conditions (Wu 1992; Sinokrot and Stefan 1994; Gu and Li 2002). Sinokrot and Stefan (1994) introduced the air temperature and shortwave solar

radiation as the most important parameters that affect the water temperature. Gu and Li (2002) found that sensitivity of river temperature to flowrate and inflow temperature to be as important as atmospheric conditions. They reported that air temperature, flowrate, relative humidity, and inflow temperature are the most sensitive parameters in calculating the maximum temperature, and slope and bottom roughness have no effect on the daily mean temperature and small effect on maximum water temperature. They considered the river slope to be a constant value for the entire river. However, in large scale numerical models, the river is divided into several meshes, and, hence, the bottom elevation and slope of each cell should be specified separately. Estimating spatially varying slopes requires an accurate bathymetry dataset for the entire length of the river. Light Detection And Ranging (LIDAR) is one approach to get high resolution topographic data to represent the land surface. However, only a specific type of LIDAR data, called green LIDAR, can penetrate the water and collect the bathymetry data. In absence of green LIDAR, bathymetry could be collected by *in-situ* data collection techniques. However, *in-situ* bathymetry data has lower resolution comparing to LIDAR. Another limitation is that when acoustical sounding is used to determine the water depth in shallow rivers with high vegetation, the acoustic signal reaches the vegetation instead of riverbed (Rogala 1999). Hence, *in-situ* bathymetry data can have uncertainty in its values, thereby requiring its effects to be considered in hydrodynamic models.

The time elapsed since a water particle moving from its boundary to a specific location is called water age (Delhez et al. 1999). Water age represents the travel time when the initial water age is set to zero. Shen and Haas (2004) calculated the mean age and residence time of a released substance from the tributaries to the tidal York River, and found that water age mostly depends on the river discharge. Gong et al. (2009) developed a three-dimensional model to evaluate the effect of wind on travel time by using the concept of water age in the tidal Rappahannock River. They found that water age distribution is highly affected by the local wind in estuaries, and depends significantly on the interaction of wind, buoyancy forcings, and on the pre-status of the circulation (Gong et al. 2009). However, sensitivity of water age to these parameters in shallow rivers is not clear. Additionally, the sensitivity of water age to the bathymetry

data has not been evaluated in previous studies. Hence, there is still the need to investigate the effect of a comprehensive set of parameters on the water age in shallow rivers.

Even though there are several studies that present the response of water temperature and water age to meteorological and hydraulic conditions, more research is still needed to completely investigate the effect of a comprehensive set of model inputs and parameters. Since, model calibration is usually time consuming and requires a large amount of effort, knowledge of sensitive parameters would help model developers. The sensitivity of a model to calibration parameters such as surface (water/air and bed/water) heat exchange variables has not been entirely explored. Furthermore, due to a relatively large uncertainty in bathymetry data, the sensitivity of models to this parameter needs to be investigated. In this study, we built a three-dimensional hydrodynamic model to model the Lower Klamath River. We used a global sensitivity analysis tool that combines Latin-hypercube and one-factor-at-a-time sampling to investigate the most sensitive model inputs and parameters in calculating the water age and water temperature of shallow rivers. Monte Carlo sampling method was also implemented to evaluate the sensitivity of model to bathymetry data.

3.3. Study site

The Klamath River basin is located in southern Oregon and northern California. It is divided into upper basin and lower basin at Iron Gate Dam (IGD) at river kilometer 304. The modeled reach for this study begins at IGD and continues 100 km downstream to the Seiad Valley (river kilometers 304 to 204). This reach incorporates a high infection zone likely to be influenced by a prescribed release of dam water. The lower Klamath basin is generally covered by forest except the drainages of Shasta and Scott River which is used mostly for agriculture and rangeland. Unlike the upper basin, the lower basin is steep and rocky with a complex terrain. Two main tributaries join the main-stem: the Shasta and Scott Rivers, draining from the Shasta and Scott basins respectively. However, there are smaller tributaries from the lower Klamath basin that drain into the river, including Beaver Creek, Horse Creek, and Grider Creek (Figure 3.1).

Figure 3.1: Map of the Klamath River showing the study area from Iron Gate Dam to Seiad Valley. Red dots show the USGS discharge stations at IGD, Shasta, Scott, and Seiad Valley stations. Green dots show the stations at Tree of Heaven (TH), Beaver Creek (BC), and Community Center (CC) where bathymetry data have been measured. Yellow dots show the weather stations at Collins Baldy (CLB) and Slater Butte (SRB). Blue arrows also show the major tributaries joining the Klamath River.

3.4. Methodology

3.4.1. Data collection

Data required for building this model included flow measurements obtained from the United States Geological Survey (USGS) gauge Stations #11516530 at IGD, #11517500 at Shasta River, #11519500 at Scott River, and #11520500 at Seiad Valley; atmospheric data including the precipitation, air temperature, relative humidity, wind speed, and wind direction collected at Collins Baldy (CLB) and Slater Butte (SRB) stations. The water temperature measurements were collected from Karuk Tribe Water Resources measurements at IGD, Seiad Valley, and Shasta River. Bathymetry data were obtained from a survey conducted by Department of the Interior, Bureau of Reclamation (USBR) with support from the USGS. They used two boats equipped with a multibeam Acoustic Doppler current profiler (ADCP) interfaced with Global Positioning System (GPS). However, there are 6 areas with total length of 11.5 km that boats could not collect the data due to gaps in GPS coverage and data collection issues (aeration, shallow depth, etc.). Woolpert, Inc. gathered LIDAR data from the Link Dam to Happy Camp with 0.91 m (3 ft) resolution which totally covers the study site. Since LIDAR used for collecting these data cannot penetrate the water, collected data

represent the water surface elevation instead of the bathymetry. Water depth is also gathered from USBR for the entire length of study site. The bathymetry was generated by subtracting the water depth from water surface elevation for areas with missing data. However, the water elevation and water depth were not been collected on the same date, so the generated bathymetry needed to be adjusted by comparison with measured data. Accurate bathymetry with error tolerance of approximately ±0.03 m were collected using GPS and echo sounder at Tree of Heaven, Beaver Creek, and Community Center sites (Wright et al. 2014). Woolpert also collected a set of bathymetric cross sections for 4.8 km starting from I-5 to the Shasta River. These two sets of data were used for bathymetry adjustment of missing areas. Furthermore, the data collected by boats were not dense across the river. The nearest neighbor interpolation was used to generate the bathymetry data for missing areas in lateral direction.

Table 3.1: Data used for bathymetry generation and adjustment with their coverage.

Data	Region	Process
LIDAR	IGD to Seiad Valley	Bathymetry generation
Bathymetry	IGD to Seiad except 6 areas (total length of 11.5 km)	Bathymetry generation
Water depth	IGD to Seiad Valley	Bathymetry generation
Bathymetry	Tree of Heaven, Beaver Creek and Community Center site	Bathymetry adjustment
Cross sections	4.8 km upstream of Shasta river	Bathymetry adjustment

3.4.2. Numerical model

3.4.2.1. Hydrodynamics

The Environmental Fluid Dynamic Code (EFDC) was used to create a three-dimensional hydrodynamic model of the lower Klamath River in this study. This model has been applied to several studies to simulate the flow. EFDC contains three functional modules including hydrodynamics, water quality, and sediments-contaminants. It

supports Cartesian and curvilinear orthogonal horizontal coordinates and stretched or sigma vertical coordinates. EFDC solves the following continuity and momentum equations (Eqs. 3.1-3.3) given by

$$\frac{\partial H}{\partial t} + \frac{\partial (Hu)}{\partial x} + \frac{\partial (Hv)}{\partial y} + \frac{\partial w}{\partial z} = Q_H \tag{3.1}$$

$$\partial_t(Hu) + \partial_x(Huu) + \partial_y(Hvu) + \partial_z(wu) - fHv = -H\,\partial_x(g\xi + p + p_{atm}) + (\partial_x h -$$
$$z\,\partial_x H)\,\partial_z p + \partial_z(H^{-1}A_v\,\partial_z u) + Q_u \tag{3.2}$$

$$\partial_t(Hv) + \partial_x(Huv) + \partial_y(Hvw) + \partial_z(wv) - fHu = -H\,\partial_y(g\xi + p + p_{atm}) + (\partial_y h -$$
$$z\,\partial_y H)\,\partial_z p + \partial_z(H^{-1}A_v\,\partial_z u) + Q_v \tag{3.3}$$

where H is water depth, u and v are horizontal velocity components in x and y direction respectively, w is vertical velocity component in z direction, Q_H is the volumetric source and sink term concerning evaporation and rainfall, f is Coriolis factor, p is the water column hydro-static pressure, p_{atm} is the kinematic atmospheric pressure, A_v is vertical turbulent momentum diffusion coefficients, and Q_u and Q_v are momentum source-sink terms. The transport equations for temperature is (Tetra tech 2007)

$$\partial_t(HT) + \partial_x(HuT) + \partial_y(HvT) + \partial_z(wT) = \partial_z(H^{-1}A_v\,\partial_z T) + Q_T \tag{3.4}$$

where T is temperature, and Q_T is the source and since term.

3.4.2.2. Heat exchange

Surface and sediment heat exchange can be formulated as equations 3.5 and 3.6 respectively.

$$H_n = H_s + H_a + H_c + H_e - (H_{sr} + H_{ar} + H_{br}) \tag{3.5}$$
$$H_{sw} = -K_{sw}(T_w - T_s) \tag{3.6}$$

where H_n is the net rate of heat exchange, H_s is the short-wave solar radiation, H_a is the long-wave solar radiation, H_c is the heat conduction, H_e is evaporative heat loss, H_{sr} the reflected short-wave radiation, H_{ar} is the reflected long-wave radiation, H_{br} is the back radiation from the water surface, H_{sw} the rate of sediment/water heat exchange, K_{sw} is the coefficient of sediment/water heat exchange, T_w is the water temperature, and T_s is the sediment temperature.

3.4.2.3. Water age

Water age can be calculated based on tracer and age concentration (Deleersnijder et al. 2001):

$$\frac{\partial c(t,\vec{x})}{\partial t} + \nabla\big(uc(t,\vec{x})\big) - K\nabla c(t,\vec{x}) = 0 \tag{3.7}$$

$$\frac{\partial \alpha(t,\vec{x})}{\partial t} + \nabla\big(u\alpha(t,\vec{x})\big) - K\nabla\alpha(t,\vec{x}) = c(t,\vec{x}) \tag{3.8}$$

where $\vec{x}$ is a coordinate, u is the velocity in space and time, c is the tracer concentration, K is the diffusivity tensor, and α is the age concentration.

3.4.3. Model set up

Curvilinear model grids were used in this study with five cells across the river and three vertical layers. The longitudinal grid resolution ranged from 85 m to 150 m. Boundaries include the flows from IGD, Shasta River, Beaver Creek, Horse Creek, Scott River, Grider Creek, and the open boundary at Seiad Valley. The model was run for three days with one second time steps.

3.4.4. Sensitivity method and effective parameters

A complete sets of parameters used in EFDC have been shown in Table 3.2 and Table 3.3. These parameters were used in the sensitivity analysis in order to investigate the model representation of water age and water temperature. LH-Oat (Van Griensven et al. 2005), a novel sampling method that is a combination of latin-hypercube and one-factor-at-a-time sampling method, was used in this study. This method is done by taking N Latin Hypercube sample points (user defined) with a probability of occurrence equal to $1/N$ and varying each LH sample point by changing the parameters one at a time. This efficient method requires a total of $N \times (P + 1)$ runs where P is the total number of parameters. The partial effect $S_{i,j}$ (in percent) of each parameter is estimated as

$$S_{i,j} = \left| \frac{100 \times \left(\frac{M(e_1,\ldots,e_i\times(1+f_i),\ldots,e_P)-M(e_1,\ldots,e_i,\ldots,e_P)}{[M(e_1,\ldots,e_i\times(1+f_i),\ldots,e_P)+M(e_1,\ldots,e_i,\ldots,e_P)]/2} \right)}{f_i} \right| \tag{3.9}$$

where M is the model function, i and j refer to the parameters and LH point respectively, e_i refers to model parameters, and f_i is a fraction by which parameters are changed. Finally, each parameter can be ranked as the largest effect to the smallest effect by

giving rank 1 to a rank equal to the total number of parameters respectively. The parameters listed in Table 3.2 and Table 3.3 were changed randomly over the range of ± 50% considering five LH sampling points.

Table 3.2: Parameters and parameters range used in water age sensitivity analysis

Name	Min	Max	Definition
Q	30	80	Flowrate
WINDS	0	10	Wind speed
WINDD	0	360	Wind direction
Z0	0.001	0.2	Bottom roughness
AHO	0.001	100	Horizontal momentum and mass diffusivity
AVO	10^{-6}	10^{-2}	Vertical eddy (kinematic) viscosity
ABO	10^{-9}	10^{-5}	Vertical molecular diffusivity

Table 3.3: Parameters and parameters range used in water temperature sensitivity analysis

Name	Min	Max	Definition
Q	30	80	Flowrate
TSER	5	25	Inflow temperature
PATM	945	960	Atmosphere pressure
TDRY	0	40	Dry atmosphere temperature
TWET	0	15	Wet bulb atmosphere temperature
EVAP	0	0.03	Evaporation rate
SOLSWR	0	1200	Solar short wave radiation at water surface
CLOUD	0	1	Fractional cloud cover
WINDS	0	10	Wind speed
WINDD	0	360	Wind direction
Z_0	0.001	0.2	Bottom roughness
SWRATNF	0	2	Fast scale solar short wave radiation attenuation coefficient
FSWRATF	0	1	Fraction of solar short wave radiation adsorbed in the top layer

DABEDT	0	1	Thickness of active bed temperature layer
HTBED1	0	1	Convective heat coefficient between bed and bottom water layer
HTBED2	0	0.5	Heat transfer coefficient between bed and bottom water layer
AHO	0.001	100	Horizontal momentum and mass diffusivity
AVO	10^{-6}	10^{-2}	Vertical eddy (kinematic) viscosity
ABO	10^{-9}	10^{-5}	Vertical molecular diffusivity

3.4.5. Model sensitivity to bathymetry data

The calibrated model of Klamath River for 2015 was obtained from previous model developed by Javaheri et al. (2016). Monte Carlo sampling approach was used for assessing the sensitivity of model to bathymetry data. This method samples from the possible range of the input values. The samples are usually generated randomly. Based on the estimated error of bathymetry from section 3.4.1, the bottom elevation of each cell was perturbed among the range of error to create the ensembles. The advantage of this method is that sensitivity and uncertainty analysis can be accomplished at once. Since this method is computationally expensive, two extreme event during 2015 were selected to evaluate the water temperature and water age sensitivity to bathymetry data. However, usually the maximum fish mortality in the lower Klamath River occurs in June when the temperature exceeds 15 °C and the number of fishes in the river are greatest. Hence, another scenario was added during this time to investigate the effect of bathymetry on water temperature.

3.5. Results and discussions

3.5.1. Bathymetry data generation

After subtracting the water depth from water surface elevation, it was concluded that water surface elevation provided by LIDAR data needed to be adjusted by 31 cm in order to minimize the error. After water surface elevation adjustment and

interpolation of data across the river, the overall root mean square error between generated bathymetry and actual data is less than 40 cm.

3.5.1. Water age (travel time) sensitivity

Table 3.3 shows the sensitivity rank of all the parameters for water age at Seiad Valley. Based on the partial effects, parameters were classified into three groups. Flow rate was classified as the most important parameter (global sensitivity rank equal to 1). Bottom roughness was classified as very important (global sensitivity rank equal to 2). Wind speed, wind direction, and vertical eddy viscosity were classified as slightly important (3 ≤ global sensitivity rank ≤ 5). Horizontal mass diffusivity and vertical molecular diffusivity have no impact on the water age, so they were ranked 7 (total number of parameters). Figure 3.3 shows the partial effect of each parameter colored based on the importance group.

Since flow rate is the most sensitive parameter, two different periods during 2015 were selected, one with a high flow rate (80 cms) and one with a low flow rate (25 cms), to investigate the water age sensitivity to bathymetry data. Figure 3.4 shows that average variance decreases by number of sample size. For sample size more than 200, the average variance converges to a constant value. It states that this model with more than 200 sample size could be representative of the bathymetry error in the system. Figure 3.5 illustrates the average travel time from Iron Gate dam to Seiad valley with the minimum and maximum bounds at to different times with various flowrate. The average difference between lower bound and upper bound is about 20 minutes and the maximum difference between bounds are 25 minutes for low flow and 40 minutes for high flow. Water age is slightly affected by bathymetry, however, flow rate does not impact the sensitivity of water age to bathymetry.

Figure 3.2: Estimated bathymetry (blue filled circles) vs measured data (red open triangles).

Table 3.4: Sensitivity rank of parameters for water age (parameters with no appearance of sensitivity get rank 7).

Parameter/Input	Sensitivity rank
Q	1
Z0	2
WINDS	3
WINDD	4
AVO	5
AHO	7
ABO	7

Figure 3.3: Partial effect of each parameter on water age for f = ±50.

Figure 3.4: Average of variance versus sample size for water age model.

Figure 3.5: Average travel time from Iron Gate dam to Seiad valley (solid line) with the minimum and maximum possible bounds (dashed line) for a) high flow during winter (80 cms), and b) low flow during summer (25 cms).

3.5.2. Water temperature sensitivity

Table 3.4 lists all the parameters and their sensitivity rank for predicting the water temperature at Seiad Valley. Based on the partial effects, we classified the parameters into five groups. Inflow temperature was classified as the most important parameter (global sensitivity rank equal to 1). Air temperature, solar radiation, wind speed, flowrate, and wet-bulb temperature were classified as very important ($2 \leq$ global sensitivity rank $\leq$ 6). Bottom roughness, fast scale solar short wave radiation attenuation coefficient, fraction of solar short wave radiation absorbed in the top layer, and evaporation rate were classified as important ($7 \leq$ global sensitivity rank $\leq$ 10). Heat transfer coefficient between bed and bottom water layer, vertical eddy viscosity, convective heat coefficient between bed and bottom water layer, and wind direction were classified as slightly important ($11 \leq$ global sensitivity rank $\leq$ 14). Finally, thickness of active bed temperature layer, fractional cloud cover, atmospheric pressure, horizontal mass diffusivity, and vertical molecular diffusivity were classified as not important (global sensitivity rank $\geq$ 15). Figure 3.6 illustrates the partial effect of each parameter highlighted based on the importance group. Figure 3.7 shows the correlation estimates between the various parameters, and it is seen that only dry temperature and wet bulb temperature are highly correlated (correlation value of 0.91). Since these two parameters are positively correlated, the uncertainty in model will be greater than if they were not correlated.

Table 3.5: Sensitivity rank of parameters for water temperature.

Parameter/Input	Sensitivity Rank	Parameter/Input	Sensitivity Rank
TSER	1	HTBED2	11
TDRY	2	AVO	12
SOLSWR	3	HTBED1	13
WINDS	4	WINDD	14
Q	5	DABEDT	15
TWET	6	CLOUD	16
Z0	7	PATM	17
FSWRATF	8	AHO	17

| SWRATNF | 9 | ABO | 17 |
| EVAP | 10 | | |

Figure 3.6: Partial effect of each parameter on water temperature for *f* = ±50.

	PATM	TDRY	TWET	EVAP	SOLSWR	WINDS
PATM	1	−0.15	−0.12	0.03	−0.12	−0.03
TDRY	−0.15	1	0.91	0.38	−0.22	0.03
TWET	−0.12	0.91	1	0.27	−0.21	0.06
EVAP	0.03	0.38	0.27	1	−0.19	0.04
SOLSWR	−0.12	−0.22	−0.21	−0.19	1	−0.03
WINDS	−0.03	0.03	0.06	0.04	−0.03	1

Figure 3.7: Correlation matrix of forcing parameters

Table 3.5 lists the most effective parameters on water temperature and their range for three periods, the warmest period, the coldest period, and the period with highest fish mortality during 2015 to investigate the water temperature sensitivity to bathymetry. Figure 3.8 represents the timely averaged variance versus the sample size. By increasing the sample size from 50 to 100, the average variance decreases. However, there is no significant change for sample size more than 100. Figure 3.9

illustrates the average water temperature at Seiad valley with the minimum and maximum bounds of possible results. It is seen that with 40 cm error in bathymetry, the average and maximum difference between upper bound and lower bound are 0.48 °C and 0.59 °C respectively during the winter (Figure 3.9.a). However, these differences are bigger for figure 8.b. The average and maximum difference between upper bound and lower bound are 1.2 °C and 1.7 °C respectively. During the summer (Figure 3.9.c), the average and maximum difference between upper bound and lower bound are 1.65 °C and 2.3 °C respectively. It represents that water temperature is more sensitive to bathymetry for low flows and warm meteorological conditions.

Table 3.6: The most effective parameters on water temperature and their range for selected days in order to investigate the water temperature sensitivity to bathymetry.

Parameter	Range of variations		
	11 Jan – 13 Jan	19 May – 21 May	4 Jul – 6 Jul
TSER (°C)	3 - 4	14 - 16	20 - 23
TDRY (°C)	1.6 - 5	8 - 16.7	20 - 33.3
SOLSWR (watt/m^2)	0 - 265	0 - 600	0 - 949
WINDS (m/s)	0 - 1.8	0 - 2.2	0 - 2.7
Q (cms)	48 - 51	32 - 33	24 - 25
TWET (°C)	(-1.9) - 0.3	1.2 - 4.7	6.1 - 11.9

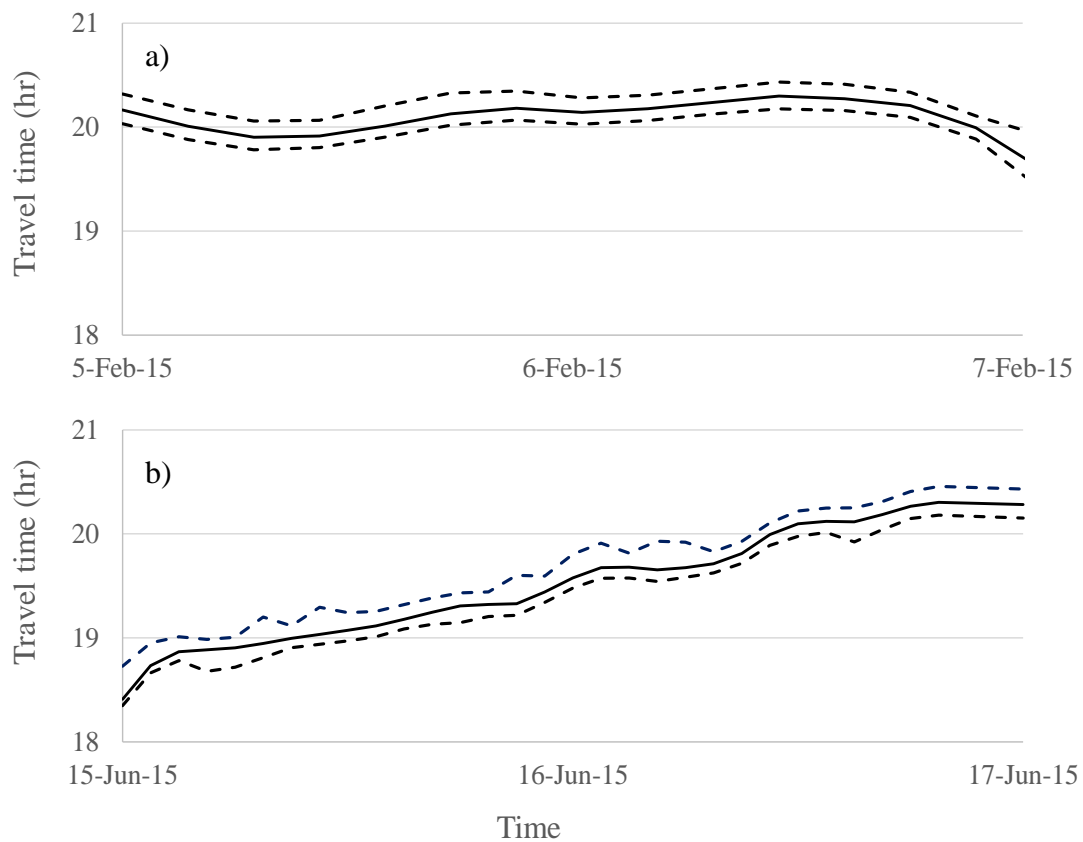**Figure 3.8: Average of variance versus sample size for water temperature model.**

Figure 3.9: Average water temperature at Seiad valley (solid line) with the minimum and maximum possible bounds (dashed line) for a) cold period (January), period with highest fish mortality (May), and c) warm period (July).

3.6. Conclusion

A three-dimensional hydrodynamic model of the lower Klamath River was developed to evaluate the most sensitive parameters in calculating the river water age and water temperature. A global sensitivity analysis tool, a combination of latin-hypercube and one-factor-at-a-time sampling, was implemented to estimate the partial sensitivity of

each parameter and rank them from the most important to not important. The results highlighted the following findings:

- Bathymetry data were generated (with overall error less than 40 cm) for the areas that data were not available.

- Flowrate and bottom roughness are the most sensitive parameter in estimating the water age. Even though wind speed, wind direction, and vertical eddy viscosity are sensitive parameters in estuaries (Gong et al., 2009), water age is not sensitive to these parameters in shallow rivers. Horizontal mass diffusivity and vertical molecular diffusivity have no impact on the water age.

- Travel time is slightly affected by bathymetry. Considering the error in generated bathymetry data, the average error between lower bound and upper bound of possible results is less than 20 minutes. Also, flowrate does not affect the travel time sensitivity to bathymetry.

- Inflow temperature was found to be the most effective parameter on river water temperature. Air temperature, solar radiation, wind speed, flowrate, wet-bulb temperature, bottom roughness, fast scale solar short wave radiation attenuation coefficient, fraction of solar short wave radiation absorbed in the top layer, and evaporation rate are the other important parameters in calculating the water temperature respectively. The other parameters are less important comparing to stated parameters.

- In contrast to Gu and Li (2002), we found that water temperature is sensitive to bathymetry especially during warm periods. Considering the estimate error in generated bathymetry, the average difference between lower bound and upper bound of possible results are 1.65 °C and 0.48 °C during the summer and winter respectively.

3.7. Acknowledgements

3.8. References

[1]   Bartholomew, J. L., Whipple, M. J., Stevens, D. G., and Fryer, J. L. (1997). "The life cycle of *Ceratomyxa shasta*, a myxosporean parasite of salmonids, requires a freshwater polychaete as an alternate host". *The Journal of parasitology*, 83(5), 859-868.

[2]   Bartholomew, J. L., and Foott, J. S. (2010). "Compilation of information relating to Myxozoan Disease effects to inform the Kamath Basin Restoration Agreement". Oregon State University and US Fish and Wildlife Service.

[3]   Deleersnijder, E., Campin, J.M., and Delhez, E.J.M., (2001). The concept of age in marine modeling. I. Theory and preliminary model results. *Journal of Marine Systems*, 28, 229–267.

[4]   Delhez, E.J.M., Campin, J.-M., Hirst, A.C., and Deleersnijder, E., (1999). Toward a general theory of the age in ocean modeling. *Ocean Modeling*, 1, 17–27.

[5]   Edinger, J. E., Brady, D. K., and Geyer, J. C. (1974). Heat exchange and transport in the environment. *Report No. 14, Electric Power Res. Inst*. Pub. No. EA-74-049-00-3, Palo Alto, CA, Nov. 1974, 125 pp.

[6]   Hallett, S. L., Ray, R. A., Hurst, C. N., Holt, R. A., Buckles, G. R., Atkinson, S. D., and Bartholomew, J. L. (2012). Density of the waterborne parasite *Ceratomyxa shasta* and its biological effects on salmon. *Applied and environmental microbiology*, 78(10), 3724-3731.

[7]   Gong, W., Shen, J., and Hong, B. (2009). The influence of wind on the water age in the tidal Rappahannock River. *Marine Environmental Research*, 68(4), 203-216.

[8]   Grant, P. J. (1977). Water temperatures of the Ngaruroro River at three stations. *Journal of Hydrology* (N.Z.) 16, 148-157.

[9]   Gu, R. R., and Li, Y. (2002). River temperature sensitivity to hydraulic and meteorological parameters. *Journal of Environmental Management*, 66(1), 43-56.

[10]   Javaheri, A., Babbar-Sebens, M., and Alexander, J. D. (2016). A Hydrodynamic and Water Quality Model of the Lower Klamath River, California, for the Management of Fish Mortality. *In World Environmental and Water Resources Congress* (pp. 492-502).

[11]   Jordan, M. S. (2012). "Hydraulic predictors and seasonal distribution of *Manayunkia speciosa* density in the Klamath River, California, with implications for ceratomyxosis, a disease of salmon and trout". M.S. Thesis,

Departments of Water Resources Science and Microbiology, Oregon State University, Corvallis, OR.

[12] Morse, W. L. (1972). Stream temperature prediction under reduced flow. *Journal of the Hydraulics Division* 98(HY6), 1031-1047.

[13] Rogala, J. T. (1999). *Methodologies employed for bathymetric mapping and sediment characterization as part of the Upper Mississippi River system navigation feasibility study*. US Army Engineer Waterways Experiment Station.

[14] Shen, J., and Haas, L., (2004). Calculating age and residence time in the tidal York River using three-dimensional model experiments. *Estuarine, Coastal and Shelf Science*, 61, 449–461.

[15] Sinokrot, B. A., and Stefan, H. G. (1994). Stream water-temperature sensitivity to weather and bed parameters. *Journal of Hydraulic Engineering*, 120(6), 722-736.

[16] Stefan, H. G., and Preud'Homme, E. B. (1993). Stream Temperature Estimation from Air Temperature. *Water Resources Bulletin*, 29, 27-45.

[17] Tetra Tech. (2007). The Environmental Fluid Dynamics Code, User Manual, US EPA 777 Version 1.01.

[18] True, K., Bolick, A., and Foott, J. S. (2011). FY 2010 Investigational Report: Myxosporean parasite (*Ceratomyxa shasta and Parvicapsula minibicornis*) annual prevalence of Infection in Klamath River basin juvenile Chinook salmon, April-August 2010. U.S. Fish and Wildlife Service California-Nevada Fish Health Center, Anderson, CA.

[19] True, K., Voss, A., and Foott, J. (2016). "Myxosporean Parasite (*Ceratonova shasta and Parvicapsula minibicornis*) Prevalence of Infection in Klamath River Basin Juvenile Chinook Salmon, April - July 2015. U.S. Fish & Wildlife Service California − Nev ada Fish Health Center, Anderson, CA.

[20] Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of hydrology*, 324(1), 10-23.

[21] Ward, J. V. (1985). Thermal characteristics of running waters. *Hydrobiologia*, 125, 31-46.

[22] Wright, K. A., Goodman, D.H., Som, N.A., and Hardy, T.B. (2014). Development of two-dimensional hydraulic models to predict distribution of *Manayunkia speciosa* in the Klamath River. *U.S. Fish and Wildlife Service*. Arcata Fish and Wildlife Office, Arcata Fisheries Technical Report Number TR 2014-19, Arcata, California.

[23] Wu, J. (1992). Variation of the heat transfer coefficient with environmental parameters. *Journal of physical oceanography*, 22(3), 293-300.

CHAPTER 4. Assimilation of multi-sensor water temperature observations into hydrodynamic model of shallow rivers

Amir Javaheri[1], Meghna Babbar-Sebens[1], Robert N. Miller[2], Julie Alexander[3], Sascha Hallett[3], and Jerri Bartholomew[3]

[1]School of Civil and Construction Engineering, Oregon State University, Corvallis, Oregon, USA.

[2]College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis, Oregon, USA.

[3]Department of Microbiology, Oregon State University, Corvallis, Oregon, USA.

4.1. Abstract

There is lack of knowledge in assimilation of multi-sensor water temperature observations into hydrodynamic model of shallow rivers to find, (a) how model accuracy would improve after assimilation of each measurements, and (b) how to assimilate these data into the prediction model to increase the model accuracy and decrease the computational burden. Mutli-sensor observations used in this study include in-situ observations with dense temporal resolution, and remote sensing measurements with better spatial coverage. The physical region represented by satellite data does not directly correspond with the numerical model's representation of the water column. Satellite data show the skin water temperature, while numerical models estimate the bulk temperature. Furthermore, for rivers narrower than the resolution of satellite data, temperature of each cell represent the weighted average temperature of land and water. These factors introduce biases in the updated numerical model of water bodies.

This study implements an efficient ensemble Kalman filter method using Latin-Hypercube sampling to assimilate unbiased multi-sensor water temperature observations into the hydrodynamic model of the lower Klamath River located in northern California. Results showed that assimilation of remote sensing data from Landsat-7 improved the model prediction for the entire river. The average spatial error reduced from 2.59 °C to 0.66 °C (i.e., 75% improvement). In-situ data assimilation reduced the error at observation location, however, updated model predicts the water temperature the same as un-updated model in less than two days. In the other hand, it is not computationally efficient to assimilate all the available data into the model. In order to overcome these challenges, in-situ data were sequentially assimilated into the model as the error exceeded the maximum allowed error (i.e., depends on model application). Results showed that by assimilation of in-situ data one to three times per day, the average daily error reduced up to 58% comparing to situation that in-situ data were assimilated only once.

Keywords

Data assimilation; Ensemble Kalman filter; In-situ data; Latin-Hypercube; Remote sensing data

4.2. Introduction

Water temperature has been identified as a critical factor that significantly affects the biological activity and metabolic rates of aquatic organisms living in the rivers or lakes. All aquatic species have a preferred water temperature range, and if temperature varies too much from this range major mortality can occur. While fishes (such as salmonids) live in cold water streams, other aquatic organisms (such as plants) prefer to live in warm water temperatures (Chu, 2003; Todar, 2012). Bartholomew and Foott (2010) found that if river temperatures reach 15-18 °C in late spring, infection can cause high mortality in juvenile salmonids. Furthermore, dissolved oxygen in water is related to water temperature, and is also a main factor for aquatic life. Water temperature also affects the water quality due to its significant effect on fate and transport of contaminants (Schnoor, 1996; Chapra, 2008; Babbar-Sebens et al., 2013).

Typical methods of water temperature monitoring include remote sensing measurements and in-situ data collection. Remote sensing data from satellites provide spatially-dense but temporally limited data, while and in-situ sensors collect temporally continuous data with limited spatial coverage. Remote sensing water surface temperature can be extracted from Landsat-7 Enhanced Thematic Mapper plus (ETM+) Band 6 with 60 m (resampled to 30 m) resolution and repeat cycle of 16 days. However, satellite data is limited by errors introduced by low resolution and impermeability to water. For rivers and stream narrower than the remote sensing grid cells, measured temperature represents the average temperature of river banks and water. Also, remote sensing data represents the skin temperature which does not directly correspond with the numerical model's representation of the water column (Donlon et al. 2002). It is also directly affected by solar radiation and wind and it could be cooler or warmer than the bulk temperature of water. While adjustment techniques have been proposed to remove the bias from water temperature satellite data for lakes, reservoirs and oceans (Javaheri et al., 2016a; Cohn et al., 1988; Reynold et., al 2002; Donlon et al., 2002), this relationship is not determined for shallow rivers.

The numerical model and measurements have large uncertainty coming from different sources such as errors of approximation and truncation, uncertain model inputs, error in measuring devices and etc. Data assimilation (DA) is able to sequentially update the model state variables by considering the uncertainty in model and observations and estimate the states and outputs more accurately. Data assimilation has been successfully used in several water resources problems such as water temperature prediction (Ezer and Mellor, 1997; Keppenne and Rienecker, 2003; Troccoli and Haines, 1999). Javaheri et al. (2016a) assimilated remote sensing water temperature obtained from Lansat-5 Thematic Mapper (TM) satellite into the hydrodynamic model of Eagle Creek Reservoir located in central Indiana. They found that adjustment of remote sensing data before assimilation into the model not only improves the model efficiency with respect to satellite data, but it also reduces the error comparing with in-situ data. Since the fate and transport and mixing processes of lake and rivers can be significantly different (e.g. in shallow rivers due to high mixing the temperature stratification is negligible), it is unknown whether similar adjustment-assimilation techniques would help improve prediction of numerical models of shallow rivers. Additionally, for rivers that have regular temperature measurements, an effective and efficient data assimilation framework is needed to assimilate heterogeneous data sources from different monitoring technology into the prediction model in order to (i) keep the model error below the maximum acceptable error, and (ii) decrease the computational cost by assimilating less data.

In this study, we have proposed a temperature data assimilation methodology that takes advantage of observations coming from different sources and an efficient ensemble Kalman Filter technique based on Latin hypercube sampling method. This methodology was applied to a hydrodynamic model of lower Klamath River located in northern California, in order to investigate the following specific research questions:

> (1) How does assimilation of in-situ water temperature observations into the hydrodynamic model of a river affect the model accuracy comparing with remote sensing data?

(2) How does assimilation of remote sensing data derived from satellite affect the model accuracy when results of updated model are compared with in-situ data?

(3) What is the most efficient way to take the advantage of both monitoring technologies and assimilate them into the prediction model to increase the overall model accuracy (spatial and temporal) and decrease computational costs?

This paper is organized as follow. Section 4.3 explains the case study and data collection, extracting water temperature from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) band 6, hydrodynamic model, and data assimilation formulation. In section 4.3, results for different scenarios after assimilation of in-situ data, remote sensing data, and both were discussed and finally Section 4.4 provides the conclusion.

4.3. Methodology:

4.3.1. Case study

The Klamath River is emanates from Cascade Mountains and flows from the Upper Klamath Lake through southern Oregon and Northern California and continues to the Pacific Ocean. The total length of river is about 410 km and it is divided into upper and lower Klamath River at Iron Gate dam (IGD). This study is focused on the lower Klamath River (Figure 4. 1) starting from IGD at river kilometer (RKM) 304 to the Seiad Valley (SV) (RKM 205). The lower Klamath basin is steep and rocky with a complex terrain generally covered by forest except the drainages of Shasta and Scott River. Among the several tributaries draining from the Lower Klamath Basin to the river, Shasta and Scott Rivers have the highest flow. Flows at IGD vary during the year; December and January usually have the highest flow and lowest flow occur from June to September.

Figure 4. 1: Lower Klamath River from Iron Gate dam to Seiad Valley. Red dots show the USGS discharge stations at IGD, Shasta, Scott, and Seiad Valley stations. Blue arrows also show the locations where major tributaries join the Klamath River. Water flows from IGD station to Seiad Valley station.

4.3.2. Temperature data

In-situ data has been collected every 15 minutes at Saied Valley station (Figure 4. 1). This data was used for model calibration as well as data assimilation. The remotely sensed water temperature data was obtained from Landsat 7 Enhanced Thematic Mapper Plus (ETM+) band 6. Remote sensing temperatures were selected at three different dates during the year with different temperatures (May 23rd, July 1st, and September 19th of 2015). The spectral radiance from satellite images need to be converted into at-satellite brightness temperature, *BT*, by Planck's law

$$BT = \frac{K_2}{ln\left[\frac{K_1}{L_\lambda}+1\right]} \tag{4.1}$$

where K_1 and K_2 are the calibration constants as 666.09 Watts m^{-2} ster^{-1} μm^{-1} and 1282.71 Kelvin respectively, and L_λ is spectral radiance which is estimated by

$$L_\lambda = M_L Q_{cal} + A_L \tag{4.2}$$

where M_L is band-specific multiplicative rescaling factor, Q_{cal} is quantized and calibrated standard product pixel values, and A_L is band-specific additive rescaling factor. The brightness temperature at satellite is converted to Land Surface Temperature (LST) by (Artis & Carnahan, 1982)

$$\text{LST} = BT/\big(1 + w \times (BT/P) \times ln(e)\big) \tag{4.3}$$

The emissivity of water is 0.98 and LST represents the skin water temperature of river. To assimilate remote sensing data, approximately 200 random sampling locations along the river were selected. Half of these sampling locations were used as training locations to implement the ensemble Kalman filter data assimilation and the other half were used for testing the efficiency of data assimilation method. Since cell size of temperature raster at some areas is bigger than the river length, the temperature of that cell represents the weighted average temperature of land and water which expects to be warmer than the water temperature. Also, remote sensing data represents the skin temperature of water which is affected by solar radiation and wind. Hence, the bias from satellite data should be removed before data assimilation.

4.3.3. Simulation model

The Environmental Fluid Dynamic Code (EFDC) is a three-dimensional surface water modeling system for hydrodynamic, water quality, and sediment-toxic contaminant simulations. The EFDC was originally developed at the Virginia Institute of Marine Science to model the estuarine system. However, it has been widely used to model the flow characteristics of different water bodies such as rivers, lakes, reservoirs, wetlands, and coastal ocean regions (Hamrick and Wu, 1997). EFDC solves the equations of continuity and momentum as well as transport equations to estimate the dynamics and temperature respectively (Tetra tech, 2007). It supports Cartesian and curvilinear orthogonal horizontal coordinates and stretched or sigma vertical coordinates (Craig, 2005). Grid size and time step were determined by considering the Courant-Friedrichs-levy condition (CFL condition). The model was built using curvilinear coordinate and sigma stretch vertical coordinate. The final model grid include 1270 longitudinal grids, 5 lateral grid, and 3 layers ($1270 \times 5 \times 3$). The model was run with 2 seconds time step considering the stability and the computational burden.

4.3.4. Ensemble Kalman filter algorithm

The ensemble Kalman filter (EnKF), introduced by Evensen (1994), was used as the data assimilation technique in this study. This method is effective for high order nonlinear models. If we denote X as vector of state variables and Y as vector of measurements, we may write out hydrodynamic numerical model, *F*, and measurements as:

$$X_k = F(X_{k-1}, u_{k-1}) + w_k \tag{4.4}$$

$$Y_k = H(X_k) + v_k \tag{4.5}$$

where u_k is the input vector, w_k is the uncertainty due to error in model, v_k is the vector of measurement errors, H is linearized observation transformation matrix, and the subscript "k" indicates the time step. We need to estimate the forecast error covariance in order to update the state variables. Since, the true values for state variables are unknown, we use the average of ensembles as true values.

$$P_k^f = \frac{1}{n-1} \sum_{i=1}^{n} (X_k^f - \overline{X}_k)(X_k^f - \overline{X}_k)^T \tag{4.6}$$

The updated state vector is taken to be a linear combination of model predictions and measurements:

$$X_k^a = X_k^f + K_k(Y_k - H_k X_k^f) \tag{4.7}$$

where K is Kalman gain and it can be estimates by:

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1} \tag{4.8}$$

In order to avoid the problems such as inbreeding, filter divergence and spurious correlations due to undersampling, covariance inflation and covariance localization methods were applied. Hence, the updating equation can be expressed by

$$X_k^a = X_k^f + [\rho_o(\lambda_k P_k^f H_k^T)][\rho_o(H_k \lambda_k P_k^f H_k^T) + \mu_k R_k]^{-1}(Y_k - H_k X_k^f) \tag{4.9}$$

where λ_k and μ_k are inflation factors for forecast and observation covariance respectively, and ρ is correlation function (Javaheri et al., 2016a). Random noise with zero mean and covariance R was added to measurements to get consistent error propagation in the ensemble Kalman filter (Burger et al., 1998).

4.3.4.1 Latin-hypercube sampling

Typically, EnKF uses a Monte Carlo approach to randomly sample the variables among their distribution range. In this study, instead of Monte Carlo, Latin-Hypercube sampling was selected to increase the efficiency. Latin-Hypercube sampling method (McKay et al., 1979; Iman and Conover, 1980; McKay, 1988) is based on Monte Carlo simulation but instead of random sampling approach, it uses a stratified sampling approach. This method divides the distribution of each parameter into *N* equal probability intervals (with a probability of occurrence equal to 1/*N*). Then, for each parameter, each interval is sampled once. This method is robust and requires less number of simulations comparing to Monte Carlo simulation which reduces the computational costs.

4.3.4.2 Estimation of ensembles spread

The performance of EnKF is highly dependent on the quality of ensemble generation. Ensembles are created by perturbing the forcing data. If ensembles have too much spread, the covariance matrix of model may be over predicted and if ensembles have too little spread the covariance matrix may be underestimated. A simple method was introduced by Anderson (Anderson, 2001) to evaluate the amount of forcing data perturbation. According to this method, root mean square error (RMSE) ratio, *Ra*, is the ratio of the time-averaged RMSE of the ensemble mean, *R1*, to the mean RMSE of the ensemble members, *R2* (Moradkhani et al., 2005).

$$Ra = \frac{R_1}{R_2} \tag{4.10}$$

$$R_1 = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\left[\left(\frac{1}{n}\sum_{i=1}^{n}\hat{y}_t^i\right) - y_t^i\right]^2} \tag{4.11}$$

$$R_2 = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{y}_t^i - y_t^i\right)^2} \tag{4.12}$$

where *n* is ensemble size and *T* is period of analysis. The Normalized RMSE Ratio (NRR) is calculated:

$$NRR = \frac{Ra}{E[Ra]} \tag{4.13}$$

where *E*[*Ra*] is the expected value of the *Ra* and can be estimated as (Murphy, 1988):

$$E[Ra] = \sqrt{\frac{(n+1)}{2n}} \tag{4.14}$$

Optimum value for NRR should be close to unity. NRR>1 indicates that ensemble spread is too small, while NRR<1 shows that ensemble spread is too big (Moradkhani et al., 2005).

4.4. Results

Calibrated model of Lower Klamath River from January 1st to April 30th was obtained from Javaheri et al. (2016b). The calibrated model predicts the water temperature of river with root mean square errors of 0.88 °C comparing with in-situ data collected at Seiad Valley station. The period of May 1st to September 31st was selected as data assimilation period.

4.4.1. Ensemble generation

The distribution of each input parameter was divided into 30 intervals and each interval was sampled only once. The most sensitive parameters in calculating the water temperature of shallow rivers include inflow temperature, air temperature, solar radiation, wind speed, flowrate, and wet-bulb temperature (Javaheri et al., to be submitted). These model inputs were perturbed to generate the ensemble members. Figure 4.2 illustrates the variation of normalized RMSE ratios for different perturbation factors. It was found that with 15% distribution range for input parameters and 20 ensemble members, NRR is 1.013 which is very close to unity.

Figure 4.2: Normalized RMSE ratio (NRR) against input perturbation.

4.4.2. Assimilation of in-situ temperature data

Observations at Seiad Valley were assimilated into the model at noon on May 23rd, July 1st, and September 19th (Figure 4.3), the same time that Lansat-7 has collected remote sensing data. Figure 4.3 compare the model predictions of model before and after in-situ data assimilation with in-situ observations at Seiad Valley. It also illustrates that depending on forcing data of each day, after 36 to 48 hours after in-situ data assimilation, updated model predicts the water temperature the same as un-updated model. Table 4.1 presents the RSME of model before data assimilation and different time periods after data assimilation. According to the results, the error decreased up to 48 hours after in-situ data assimilation for Figure 4.3.a and b. Model underestimates the minimum and maximum temperature at these days. After assimilation of in-situ data, updated model is run with warmer initial conditions that causes the error reduction. However, for September 19th, the error of updated model increased after 12 hours. At the time of data assimilation, model first overestimates the water temperature. It makes the model to be run with cooler initial conditions and error reduces for few hours. Then, model underestimates the water temperature so updated model with cooler initial condition would have higher error. Table 4.2 presents the error between modeled temperature, in-situ measurements, and remote sensing observations at the time that in-situ data were assimilated (12 pm, the same time that satellite data was collected). Results show that after assimilation of in-situ observations, the RMSE reduces at the observation site. However, in-situ data assimilation is not capable to efficiently

improve the spatial temperature predictions (Figure 4.4). The error reduction for May 23rd and September 19th is not noticeable. For July 1st, even though in-situ data assimilation reduced the error for the observation site, the error for the entire river increased. The main reason low spatial resolution and lack of enough observation along the river.

Table 4.1: The averaged root mean square error of model before and after in-situ data assimilation. $\overline{E_x}$ denotes the averaged error "x" hours after in-situ data assimilation.

Date	Root mean square error (°C)					
	Before in-situ DA			After in-situ DA		
	$\overline{E_{12}}$	$\overline{E_{24}}$	$\overline{E_{48}}$	$\overline{E_{12}}$	$\overline{E_{24}}$	$\overline{E_{48}}$
May 23rd	0.77	0.65	0.69	0.49	0.44	0.55
July 1st	0.74	0.59	0.54	0.44	0.39	0.4
September 19th	0.64	0.55	0.56	0.34	0.55	0.57

Table 4.2: Root mean square error of model before and after in-situ data assimilation comparing with in-situ and remote sensing observations at the assimilation time.

Date	Root mean square error (°C)			
	Comparing with in-situ observations for entire river		Comparing with RS observations at Seiad Valley	
	Before in-situ DA	After in-situ DA	Before in-situ DA	After in-situ DA
May 23rd	0.4	0.07	2.76	2.68
July 1st	0.51	0.18	3.29	3.91
September 19th	0.58	0.08	1.71	1.55

Figure 4.3: Depth-averaged water temperature observation at Seiad valley comparing with model predictions before and after in-situ data assimilation at 12 pm for a) May 23rd, b) July 1st, and c) September 19th. Black thick line shows the RS observation, red solid line shows the model prediction before in-situ data assimilation, and blue dashed line shows the updated model after in-situ data assimilation.

Figure 4.4: Longitudinal profile of water temperature observations comparing with model predictions before and after in-situ data assimilation at 12 pm for a) May 23rd, b) July 1st, and c) September 19th. Black thick line shows the RS observation, red solid line shows the model prediction before in-situ data assimilation, and blue dashed line shows the updated model after in-situ data assimilation.

4.4.3. Assimilation of remote sensing temperature data

The bias in remote sensing observations (from Lansat-7) was first removed by comparing with in-situ data at Seiad Valley station and then the unbiased observations were used for remote sensing data assimilation (RSDA). The results of updated model were then compared with in-situ observations to investigate how assimilation of remote sensing observation will affect the model efficiency comparing with future in-situ data (Figure 4.5). The RMSE of modeled temperature before remote sensing data assimilation and for different time periods after data assimilation was estimated (Table 4.3). The error reduced for all the assimilation dates, however, updated model after in-situ data assimilation was able to predict water temperature more accurately than updated model after remote sensing data assimilation for July 1st. The results of updated model after RSDA was then compared with satellite data (Figure 4.6). Table 4.4 summarizes the RMSE of modeled temperature after RSDS comparing with in-situ measurements, and remote sensing observations at the time that satellite data were assimilated (12 pm, the same time that satellite data was collected).

Table 4.3: The averaged root mean square error of model before and after Remote sensing data assimilation.

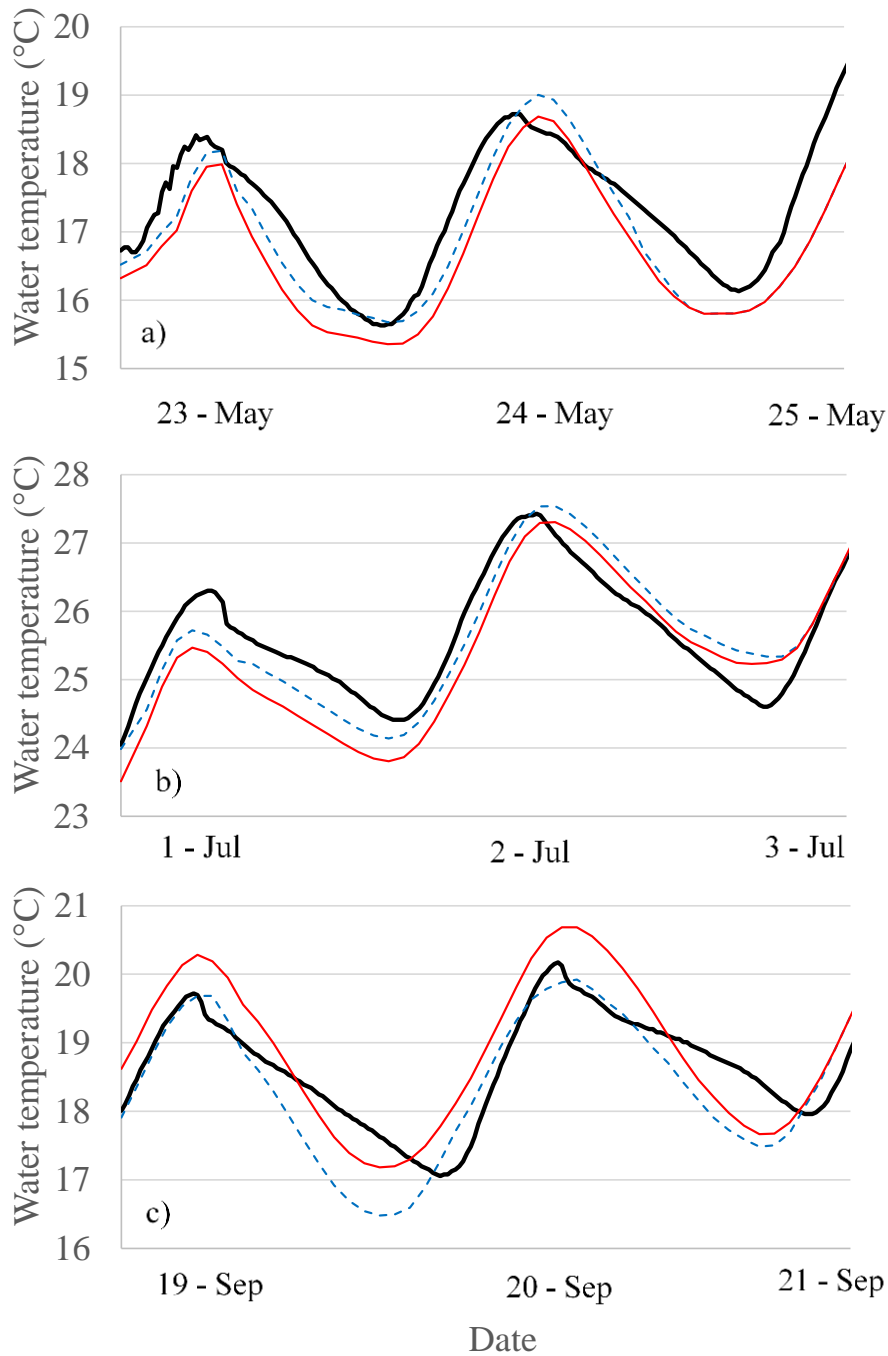| Date	Root mean square error (°C)					
	Before RSDA			After RSDA		
	$\overline{E_{12}}$	$\overline{E_{24}}$	$\overline{E_{48}}$	$\overline{E_{12}}$	$\overline{E_{24}}$	$\overline{E_{36}}$
May 23rd	0.77	0.65	0.69	0.4	0.38	0.48
July 1st	0.74	0.59	0.54	0.54	0.44	0.42
September 19th	0.64	0.55	0.56	0.43	0.47	0.47

Figure 4.5: Depth-averaged water temperature observation at Seiad valley comparing with model predictions before and after remote sensing data assimilation at 12 pm for a) May 23rd, b) July 1st, and c) September 19th. Black thick line shows the RS observation, red solid line shows the model prediction before in-situ data assimilation, and blue dashed line shows the updated model after in-situ data assimilation.
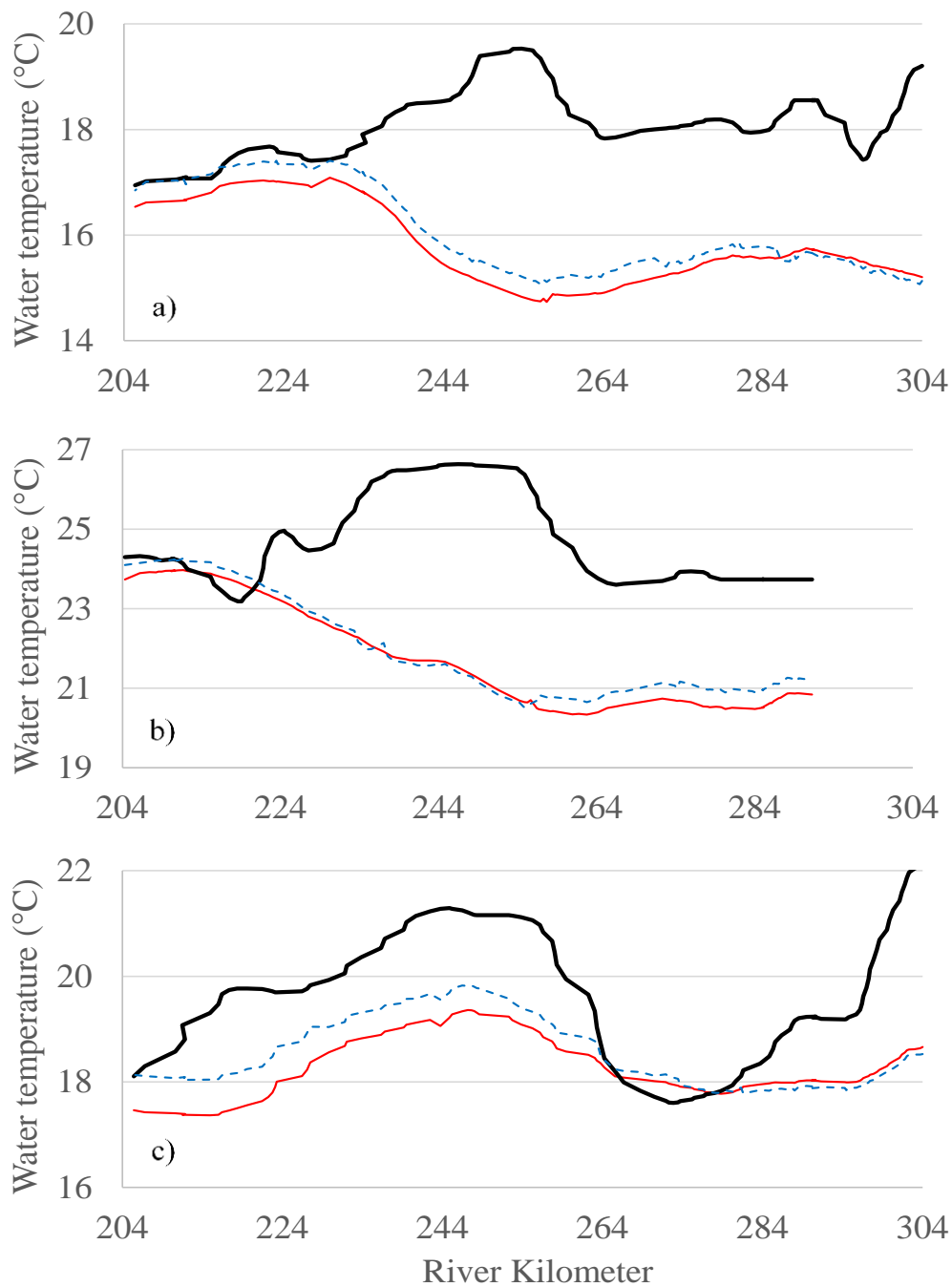
Figure 4.6: Longitudinal profile of water temperature observations comparing with model predictions before and after remote sensing data assimilation at 12 pm for a) May 23rd, b) July 1st, and c) September 19th. Black thick line shows the RS observation, red solid line shows the model prediction before in-situ data assimilation, and blue dashed line shows the updated model after in-situ data assimilation.

Table 4.4: Root mean square error of model before and after remote sensing data assimilation comparing with in-situ and remote sensing observations at the assimilation time.

Date	Root mean square error (°C)				
	Comparing with RS observations for entire river			Comparing with in-situ observations at Seiad Valley	
	Before RSDA	After RSDA		Before RSDA	After RSDA
		Training	Testing		
May 23rd	2.76	0.82	1.03	0.4	0.14
July 1st	3.29	0.95	1.08	0.51	0.27
September 19th	1.71	0.71	0.82	0.58	0.44

4.4.4. Assimilation of remote sensing and in-situ temperature data

We then assimilated both in-situ and remote sensing data into the model at appropriate time to keep the model updated as new observations (irrespective of source) become available. In-situ measurements at Seiad Valley station are available every 15 minutes, however, it not computationally efficient to assimilate them every 15 minutes. A maximum error between model predictions and observations was defined. When the error exceeds this value, new in-situ measurements are assimilated into the model. This maximum error depends on the operation of the temperature model in different fields. In this research, it was set to be 0.5 °C. Remote sensing data also assimilate when they are available. Satellite date collected at 12 pm on May 23rd, were assimilated into the model to improve the water temperature for the entire river, then the model was run until the error was exceeded from the maximum acceptable range. Then in-situ data were assimilated into the model (Figure 4.7.a). The same procedure was repeated for other days that remote sensing data are available (i.e. July 1st and September 19th; Figure 4.7.b and c respectively). It is seen that instead of assimilation of data every 15 minutes, two to five data assimilation during 48 hours is enough to keep the model error below the satisfactory range for selected dates. Table 4.5 summarizes the error

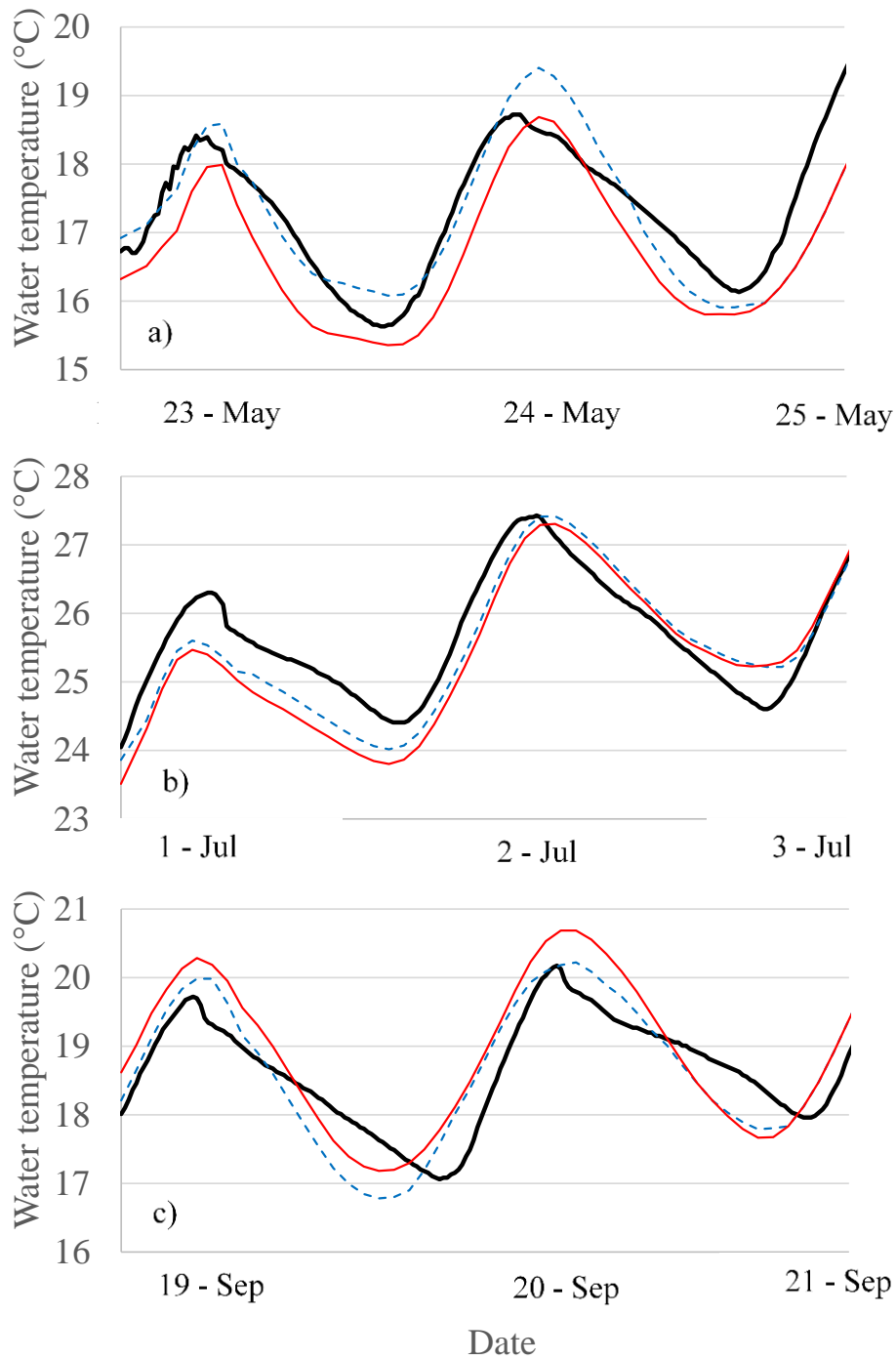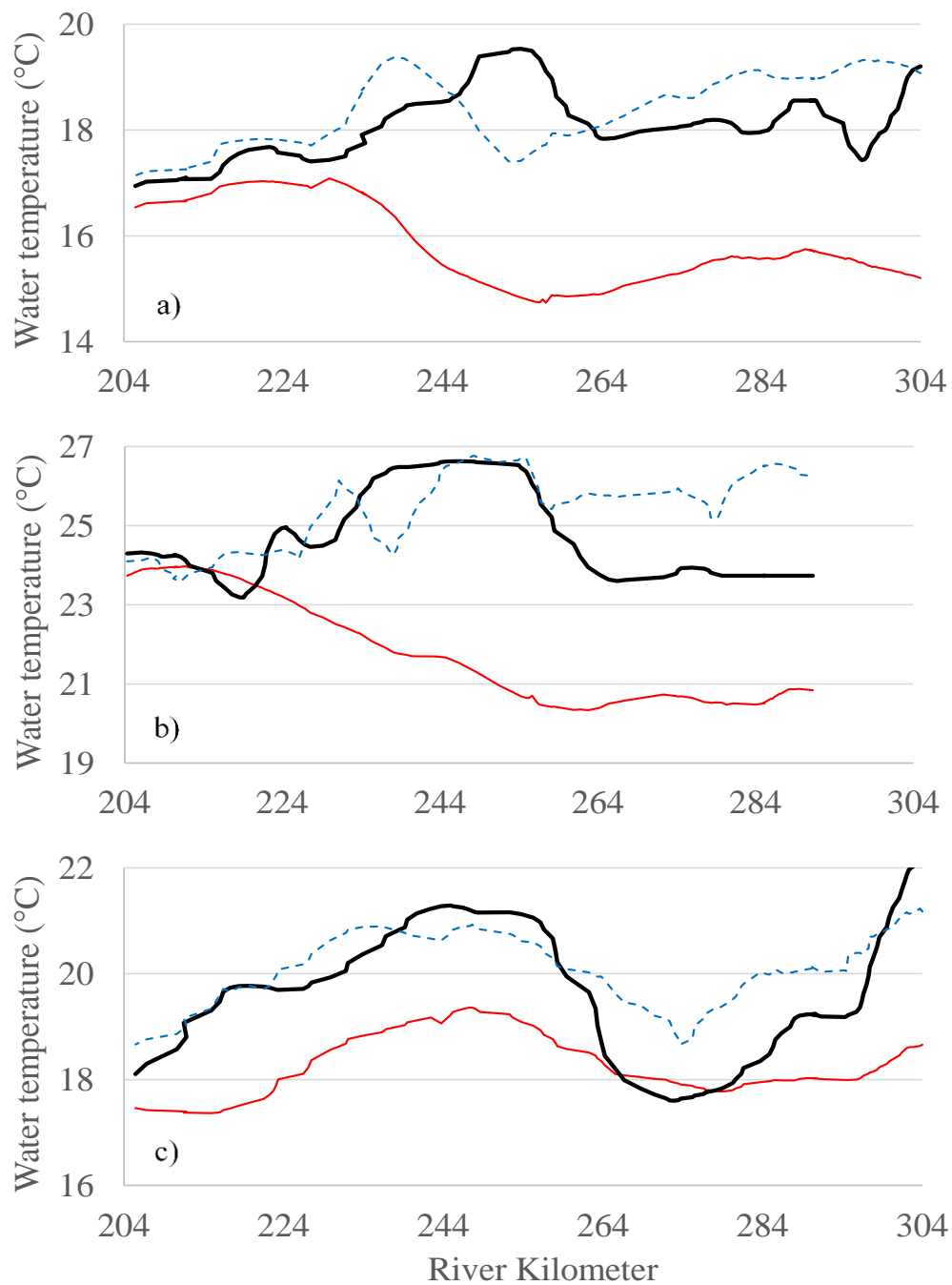values for the periods of 12, 24, and 48 hours after RSDA and while in-situ data were sequentially assimilated into the model.
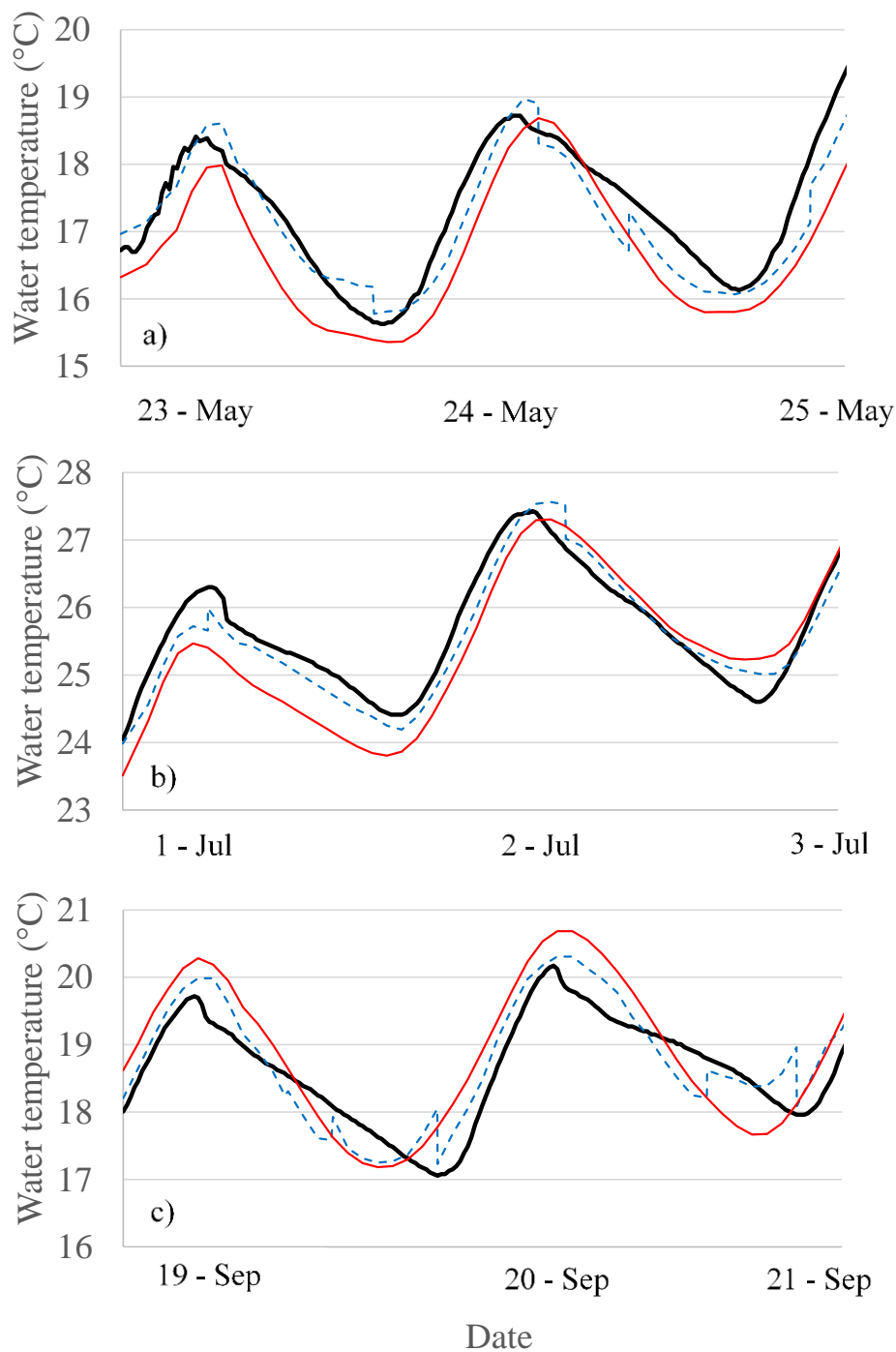

Figure 4.7: Depth-averaged water temperature observation at Seiad valley comparing with model predictions before data assimilation and after assimilation of remote sensing data at 12 pm for a) May 23rd, b) July 1st, and c) September 19th, while in-situ data were sequentially assimilated into the model. Black thick line shows the RS observation, red

solid line shows the model prediction before in-situ data assimilation, and blue dashed line shows the updated model after in-situ data assimilation.

Table 4.5: The averaged root mean square error of model before and after RSDA and sequentially assimilation of in-situ observations.

Date	Root mean square error (°C)					
	Before in-situ DA			After in-situ DA		
	$\overline{E_{12}}$	$\overline{E_{24}}$	$\overline{E_{48}}$	$\overline{E_{12}}$	$\overline{E_{24}}$	$\overline{E_{36}}$
May 23rd	0.77	0.65	0.69	0.23	0.30	0.31
July 1st	0.74	0.59	0.54	0.31	0.25	0.25
September 19th	0.64	0.55	0.56	0.33	0.39	0.38

4.5. Conclusion

A data assimilation framework using ensemble Kalman filter and Latin-hypercube sampling was implemented to update the initial condition of a three-dimensional hydrodynamic model of lower Klamath River. Observations from different monitoring technologies (i.e. in-site data, and remote sensing data from Landsat-7) were assimilated into the model to improve the model efficiency. Most effective parameters in water temperature calculation of shallow rivers were used to generate the ensemble members among the acceptable perturbation range. In contrast to in-situ data assimilation, remote sensing data assimilation was able to considerably improve the spatial error of the model. The average spatial error reduced from 2.59 °C to 0.66 °C. Assimilation of in-situ observation into the model improved the model efficiency at observation site. However, the model error increased by time and after less than two days, the model predictions of updated model were the same as base model before data assimilation. In order to decrease the model error by time and avoid the computational cost by assimilating all the available in-situ measurements, a maximum acceptable error between model and measurements was defined. In-situ observations were assimilated into the model when the error exceeded this threshold. Results showed that by assimilation of in-situ data one to three times per day, the average daily error reduced up to 58% comparing to situation that in-situ data were assimilated only once.

4.6. Acknowledgement

4.7. References:

[1] Chu, A. (2003). Prionitis lanceolata. In Monterey Bay Aquarium Research Institute: Marine Botany.

[2] Todar, K. (2012). Diversity of Metabolism in Procaryotes. In Online Textbook of Bacteriology.

[3] Schnoor, J. L. (1996). Environmental modeling: fate and transport of pollutants in water, air, and soil. John Wiley and Sons.

[4] Chapra, S. C. (2008). Surface water-quality modeling. Waveland press.

[5] Bartholomew, J. L., and Foott, J. S. (2010). Compilation of information relating to Myxozoan Disease effects to inform the Kamath Basin Restoration Agreement. *Oregon State University and US Fish and Wildlife Service*.

[6] Babbar-Sebens, M., Li, L., Song, K., and Xie, S. (2013). On the Use of Landsat-5 TM Satellite for Assimilating Water Temperature Observations in 3D Hydrodynamic Model of Small Inland Reservoir in Midwestern US. *Advances in Remote Sensing*, 2(3), 214-227.

[7] Javaheri, A., Babbar-Sebens, M., and Alexander, J. D. (2016). A Hydrodynamic and Water Quality Model of the Lower Klamath River, California, for the Management of Fish Mortality. *In World Environmental and Water Resources Congress* (pp. 492-502).

[8] Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., & Murray, M. J. (2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, *15*(4), 353-369.

[9] Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of climate*, *15*(13), 1609-1625.

[10] Ezer, T., and Mellor, G. L. (1997). Simulations of the Atlantic Ocean with a free surface sigma coordinate ocean model. *Journal of Geophysical Research: Oceans* (1978–2012), 102(C7), 15647-15657.

[11] Keppenne, C. L., & Rienecker, M. M. (2003). Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. *Journal of Marine Systems*, 40, 363-380.

[12] Troccoli, A., & Haines, K. (1999). Use of the temperature-salinity relation in a data assimilation context. *Journal of Atmospheric and Oceanic Technology*, 16(12), 2011-2025.

[13] Artis, D. A. and Carnahan, W. H. (1982). Survey of emissivity variability in thermography of urban areas. *Remote Sensing of Environment*, 12, 313 – 329.

[14] Sobrino, J. A., Jiménez-Muñoz, J. C., and Paolini, L. (2004). Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of environment*, 90(4), 434-440.

[15] Carlson, T. N. and Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index; *Remote Sens. Environ.*, 62 241–252.

[16] Hamrick, J. M., & Wu, T. S. (1997). Computational design and optimization of the EFDC/HEM3D surface water hydrodynamic and eutrophication models. In *Next Generation environmental models and computational methods* (pp. 143-161). Society for Industrial and Applied Mathematics, Philadelphia, PA.

[17] Tetra Tech, "Hydrodynamic and Water Quality Modeling Report for Lake Lanier," Georgia, 2009.

[18] Craig P M, 2005. EFDC-DS/EFDC-Explorer User Manual: 3D Hydrodynamic and Water Quality Modeling System. Dynamic Solutions International, LLC, Knoxville, TN. 1–10.

[19] Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* (1978–2012), 99(C5), 10143-10162.

[20] Burgers, G., Jan van Leeuwen, P., and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6), 1719-1724.

[21] McKay, M.D., Beckman, R.J., and Conover, W.J., (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21 (2), 239–245.

[22] Iman, R. L., & Conover, W. J. (1980). Small sample sensitivity analysis techniques for computer models. with an application to risk assessment.*Communications in statistics-theory and methods*, *9*(17), 1749-1842.

[23] McKay, M. D. (1988). Sensitivity and uncertainty analysis using a statistical sample of input values. *Uncertainty analysis*, 145-186.

[24] Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, *129*(12), 2884-2903.

[25] Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, *28*(2), 135-147.

[26] Murphy, J. M. (1988). The impact of ensemble forecasts on predictability.*Quarterly Journal of the Royal Meteorological Society*, *114*(480), 463-493.

[27] Javaheri, A., Babbar-Sebens, M., and Alexander, J. D. (2016b). A Hydrodynamic and Water Quality Model of the Lower Klamath River, California, for the Management of Fish Mortality. In *World Environmental and Water Resources Congress* (pp. 492-502).

[28] Javaheri, A., Babbar-Sebens, M., Alexander, J. D., Hallett, S. L., and Bartholomew, J. L. (To be submitted). Global sensitivity analysis of water age and water temperature of shallow rivers.

CHAPTER 5. Final Discussion

5.1. Conclusion

The limitations in assimilation of multi-sensor data into numerical hydrodynamic model of inland water bodies (e.g., reservoirs, lakes, and rivers) was investigated in this study. A data assimilation frame work using ensemble Kalman Filter was proposed for each water bodies to investigate how assimilation of water temperature from different monitoring technologies affects the model accuracy. Water temperature measurements used in this study include (i) in-situ temperature data with high temporal resolution and low spatial availability, and (ii) remote sensing data with better X-Y spatial coverage and low temporal resolution. The results of this research are categorized in two sections:

5.1.1. Water temperature data assimilation in reservoirs and lakes

Application of the methodology on Eagle Creek Reservoir, located in Central Indiana highlighted the following findings:

- An adjustment technique was proposed based on air and in-situ water temperature to remove the bias from satellite data. Air2Water model was implemented to predict the water skin temperature of the reservoir and then they were calibrated using remotely sensed observations. The predicted skin temperatures were then compared with in-situ measurements as new observations became available for bias correction.

- For the days that satellite data are available, model outputs were compared with adjusted remote sensing skin temperature. The spatial biases were removed by subtracting the absolute errors from the model outputs.

- Sampling locations were divided into two groups, training locations to estimate model errors during the data assimilation process, and testing locations to estimate the model errors after assimilation. After implementing ensemble Kalman filter, the model prediction was improved for the training locations as well as testing locations for all the observations.

- Assimilation of remotely sensed skin temperature data using the proposed approach improved the overall root mean square difference between modeled surface layer temperatures and the adjusted remotely sensed skin temperature observations from 5.6 $^{\circ}$C to 0.51 $^{\circ}$C (i.e., 91% improvement).

- The overall error in the water column temperature predictions when compared with in-situ observations also decreased from 1.95 °C (before assimilation) to 1.42 °C (after assimilation), thereby, giving a 27% improvement in errors.

5.1.2. Water temperature data assimilation in rivers

A global sensitivity analysis technique which combined latin-hypercube and one at a time sampling method was applied to the hydrodynamic numerical model of the shallow Lower Klamath River in California. The results can be summarized as follow:

- Flowrate and bottom roughness are the most sensitive parameter in estimating the water age. Wind speed, wind direction, and vertical eddy viscosity slightly affect the water age. Horizontal mass diffusivity and vertical molecular diffusivity have no impact on the water age.

- Water age is not sensitive to bathymetry. Considering the error in generated bathymetry data, the average error between lower bound and upper bound of possible results is less than 20 minutes (less than 2%).

- Water temperature is most sensitive to inflow temperature. Air temperature, solar radiation, wind speed, flowrate, wet-bulb temperature, bottom roughness, fast scale solar short wave radiation attenuation coefficient, fraction of solar short wave radiation absorbed in the top layer, and evaporation rate are the other important parameters in calculating the water temperature respectively. The other parameters are less important comparing to stated parameters.

- Even though Gu and Li (2002) claimed that water temperature is not sensitive to bathymetry, results showed that the average difference between lower bound and upper bound of possible results is 1.65 °C during the summer. However, model sensitivity to bathymetry reduces for cold seasons.

The abovementioned results were used to generate the ensembles for ensemble Kalman Filter approach used in chapter 4. The ensembles were created by perturbing water, air

temperature, solar radiation, wind speed, flowrate, and wet-bulb temperature by 15%. The perturbation range was estimated using Normalized Room Mean Square Error Ration (NRR). Results showed that:

- Assimilation of in-situ data improved the model accuracy at observation site. However, the spatial error reduction is negligible.

- After less than two days, the updated model predicts the water temperature the same as un-updated model which shows the potential need for further assimilation of observations.

- Due to computational cost, it is not optimal to assimilate all the available in-situ data (every 15 minutes) into the model. A maximum error between model and observations was defined. As error exceeded the threshold, in-situ observations were assimilated into the model.

- Results showed that by assimilation of in-situ data one to three times per day, the average daily error reduced up to 58% comparing to situation that in-situ data were assimilated only once.

- Remote sensing data assimilation significantly improved the model prediction for the entire river. The average spatial error reduced from 2.59 °C to 0.66 °C.

REFERENCES

[1] Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, 129(12), 2884-2903.

[2] Anderson, J. L., and Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12), 2741-2758.

[3] Andreadis, K. M., and Lettenmaier, D. P. (2006). Assimilating remotely sensed snow observations into a macroscale hydrology model. *Advances in Water Resources*, 29(6), 872-886.

[4] Artis, D. A. and Carnahan, W. H. (1982). Survey of emissivity variability in thermography of urban areas. *Remote Sensing of Environment*, 12, 313 – 329.

[5] Babbar-Sebens, M., Li, L., Song, K., and Xie, S. (2013). On the Use of Landsat-5 TM Satellite for Assimilating Water Temperature Observations in 3D Hydrodynamic Model of Small Inland Reservoir in Midwestern US. *Advances in Remote Sensing*, 2(3), 214-227.

[6] Bartholomew, J. L., Whipple, M. J., Stevens, D. G., and Fryer, J. L. (1997). "The life cycle of *Ceratomyxa shasta*, a *myxosporean* parasite of salmonids, requires a freshwater polychaete as an alternate host". *The Journal of parasitology*, 83(5), 859-868.

[7] Bartholomew, J. L., and Foott, J. S. (2010). "Compilation of information relating to *Myxozoan* Disease effects to inform the Kamath Basin Restoration Agreement". *Oregon State University and US Fish and Wildlife Service*.

[8] Beneš, V. E. (1981). Exact finite-dimensional filters for certain diffusions with nonlinear drift. Stochastics: An International *Journal of Probability and Stochastic Processes*, 5(1-2), 65-92.

[9] Burgers, G., Jan van Leeuwen, P., and Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, 126(6), 1719-1724.

[10] Carlson, T. N. and Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index; *Remote Sens. Environ.*, 62 241–252.

[11] Chu, A. (2003). Prionitis lanceolata. In Monterey Bay Aquarium Research Institute: Marine Botany.

[12] Craig P M, 2005. EFDC-DS/EFDC-Explorer User Manual: 3D Hydrodynamic and Water Quality Modeling System. Dynamic Solutions International, LLC, Knoxville, TN. 1–10.

[13] Crow, W. T., and Wood, E. F. (2003). The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97. *Advances in Water Resources*, 26(2), 137-149.

[14] Daum, F. E. (1986). Exact finite-dimensional nonlinear filters. *IEEE Transactions on Automatic Control*, 31(7), 616-622.

[15] Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3323-3343.

[16] Dee, D. P., and Da Silva, A. M. (1998). Data assimilation in the presence of forecast bias. *Quarterly Journal of the Royal Meteorological Society*, 124(545), 269-295.

[17] Deleersnijder, E., Campin, J.M., and Delhez, E.J.M., (2001). The concept of age in marine modeling. I. Theory and preliminary model results. *Journal of Marine Systems*, 28, 229–267.

[18] Delhez, E.J.M., Campin, J.-M., Hirst, A.C., and Deleersnijder, E., (1999). Toward a general theory of the age in ocean modeling. *Ocean Modeling*, 1, 17–27.

[19] Dong, J., Walker, J. P., and Houser, P. R. (2005). Factors affecting remotely sensed snow water equivalent uncertainty. *Remote Sensing of Environment*, 97(1), 68-82.

[20] Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., and Murray, M. J. (2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate,* 15(4), 353-369.

[21] Dorigo, W. A., Zurita-Milla, R., de Wit, A. J., Brazile, J., Singh, R., and Schaepman, M. E. (2007). A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. *International Journal of Applied Earth Observation and Geoinformation*, 9(2), 165-193.

[22] Duan, Q., Sorooshian, S., and Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water resources research*, 28(4), 1015-1031.

[23] Duan, Q. Y., Gupta, V. K., and Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, 76(3), 501-521.

[24] Duan, Q., H. V. Gupta, S. Sorooshian, A. N. Rousseau, and R. Turcotte (2003), Preface, in Calibration of Watershed Models, Water Sci. Appl. Ser., vol. 6, edited by Q. Duan et al., p. v, AGU, Washington, D. C.

[25] Edinger, J. E., Brady, D. K., and Geyer, J. C. (1974). Heat exchange and transport in the environment. *Report No. 14, Electric Power Res. Inst*. Pub. No. EA-74-049-00-3, Palo Alto, CA, Nov. 1974, 125 pp.

[26] Entekhabi, D., Nakamura, H., and Njoku, E. G. (1994). Solving the inverse problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2), 438-448.

[27] Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* (1978–2012), 99(C5), 10143-10162.

[28] Evensen, G. (2007). Data assimilation (p. 279). New York: Springer.

[29] Ezer, T., and Mellor, G. L. (1997). Simulations of the Atlantic Ocean with a free surface sigma coordinate ocean model. *Journal of Geophysical Research: Oceans* (1978–2012), 102(C7), 15647-15657.

[30] Galantowicz, J. F., Entekhabi, D., and Njoku, E. G. (1999). Tests of sequential data assimilation for retrieving profile soil moisture and temperature from observed L-band radiobrightness. *IEEE Transactions on Geoscience and Remote Sensing*, 37(4), 1860-1870.

[31] Gallus Jr, W. A., and Segal, M. (2000). Sensitivity of forecast rainfall in a Texas convective system to soil moisture and convective parameterization. *Weather and forecasting*, 15(5), 509-525.

[32] Gaspari, G., and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554), 723-757.

[33] Gelb, A. (1974). *Applied optimal estimation*. MIT press.

[34] Gillijns, S., Mendoza, O. B., Chandrasekar, J., De Moor, B. L. R., Bernstein, D. S., and Ridley, A. (2006). What is the ensemble Kalman filter and how well does it work? *In American Control Conference*, (pp. 6-pp). IEEE.

[35] Gong, W., Shen, J., and Hong, B. (2009). The influence of wind on the water age in the tidal Rappahannock River. *Marine Environmental Research*, 68(4), 203-216.

[36] Grant, P. J. (1977). Water temperatures of the Ngaruroro River at three stations. *Journal of Hydrology* (N.Z.) 16, 148-157.

[37] Gu, R. R., and Li, Y. (2002). River temperature sensitivity to hydraulic and meteorological parameters. *Journal of Environmental Management*, 66(1), 43-56.

[38] Gupta V.K, Sorooshian S, Yapo PO. Towards improved calibration of hydrological models: multiple and noncomensurable measures of information. *Water Resour Res* 1998;34:751–63.

[39] Hall, D. K., Riggs, G. A., Salomonson, V. V., DiGirolamo, N. E., and Bayr, K. J. (2002). MODIS snow-cover products. *Remote sensing of Environment*, 83(1), 181-194.

[40] Hallett, S. L., Ray, R. A., Hurst, C. N., Holt, R. A., Buckles, G. R., Atkinson, S. D., and Bartholomew, J. L. (2012). Density of the waterborne parasite *Ceratomyxa shasta* and its biological effects on salmon. *Applied and environmental microbiology*, 78(10), 3724-3731.

[41] Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129(11), 2776-2790.

[42] Hamrick, J. M., and Wu, T. S. (1997). Computational design and optimization of the EFDC/HEM3D surface water hydrodynamic and eutrophication models. *In Next generation environmental models and computational methods* (pp. 143-161). Society for Industrial and Applied Mathematics, Philadelphia, PA.

[43] Hook, S. J., Prata, F. J., Alley, R. E., Abtahi, A., Richards, R. C., Schladow, S. G., and Pálmarsson, S. (2003). Retrieval of lake bulk and skin temperatures using Along-Track Scanning Radiometer (ATSR-2) data: A case study using Lake Tahoe, California. *Journal of Atmospheric and Oceanic Technology*, 20(4), 534-548.

[44] Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., and Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, 34(12), 3405-3420.

[45] Houtekamer, P. L., and Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3), 796-811.

[46] Hulley, G. C., Hook, S. J., and Schneider, P. (2011). Optimized split-window coefficients for deriving surface temperatures from inland water bodies. *Remote Sensing of Environment*, 115(12), 3758-3769.

[47] Iman, R. L., and Conover, W. J. (1980). Small sample sensitivity analysis techniques for computer models. with an application to risk assessment. *Communications in statistics-theory and methods*, *9*(17), 1749-1842.

[48] Ines, A. V., Das, N. N., Hansen, J. W., and Njoku, E. G. (2013). Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction. Remote Sensing of Environment, 138, 149-164.

[49] Javaheri, A., Babbar-Sebens, M., and Miller, R. N. (2016). From skin to bulk: An adjustment technique for assimilation of satellite-derived temperature observations in numerical models of small inland water bodies. *Advances in Water Resources*, *92*, 284-298.

[50] Javaheri, A., Babbar-Sebens, M., and Alexander, J. D. (2016b). A Hydrodynamic and Water Quality Model of the Lower Klamath River, California, for the Management of Fish Mortality. *In World Environmental and Water Resources Congress* (pp. 492-502).

[51] Javaheri, A., Babbar-Sebens, M., Alexander, J. D., Hallett, S. L., and Bartholomew, J. L. (To be submitted). Global sensitivity analysis of water age and water temperature of shallow rivers.

[52] Jessup, A. T., and Branch, R. (2008). Integrated ocean skin and bulk temperature measurements using the Calibrated InfraRed In-situ Measurement System (CIRIMS) and through-hull ports. *Journal of Atmospheric and Oceanic Technology*, 25(4), 579-597.

[53] Jin, K. R., and Ji, Z. G. (2004). Case study: modeling of sediment transport and wind-wave impact in Lake Okeechobee. *Journal of hydraulic engineering*, 130(11), 1055-1067.

[54] Jordan, M. S. (2012). "Hydraulic predictors and seasonal distribution of Manayunkia speciosa density in the Klamath River, California, with implications for ceratomyxosis, a disease of salmon and trout". M.S. Thesis, Departments of Water Resources Science and Microbiology, Oregon State University, Corvallis, OR.

[55] Kalman R. (1960). New approach to linear filtering and prediction problems. *Trans AMSE J Basin Eng*. 82D:35–45.

[56] Keppenne, C. L., and Rienecker, M. M. (2003). Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter. *Journal of Marine Systems*, 40, 363-380.

[57] Kitanidis, P. K., and Bras, R. L. (1980). Real-time forecasting with a conceptual hydrologic model: 1. Analysis of uncertainty. *Water Resources Research*. 16(6), 1025-1033.

[58] Kitanidis, P. K., and Bras, R. L. (1980). Real-time forecasting with a conceptual hydrologic model: 2. Applications and results. *Water Resources Research*, 16(6), 1034-1044.

[59] Krener, A. J. and Duarte, A. (2004). A Hybrid Computational Approach to Nonlinear Estimation. *Proc. Conf. Dec. Contr*., Kobe, Japan, pp. 1815-1819.

[60] Li, X., Koike, T., and Pathmathevan, M. (2004). A very fast simulated re-annealing (VFSA) approach for land data assimilation. *Computers and Geosciences*, 30(3), 239-248.

[61] Liu, Z., Hashim, N. B., Kingery, W. L., Huddleston, D. H., and Xia, M. (2008). Hydrodynamic modeling of St. Louis Bay estuary and watershed using EFDC and HSPF. *Journal of Coastal Research*, 107-116.

[62] Lobugeois, F. (2009). Development of A 3D Hydrodynamic Model For Eagle Creek Reservoir. IUPUI 09/2009.

[63] Lu, S., and Davie, S.R., (2006). Charleston Harbor system 3-dimensional modeling, *Proceedings of the Ninth International Conference on Estuarine and Coastal Modeling*, Charleston, SC, pp. 13-31.

[64] Madsen, M. N., Martnack, J. N., Skotner, C., and Berg, A. (2006, September). Water quality surveillance and early warning in surface waters-integration of mathematical models and on-line monitoring. *In Proceedings of the 7th International Conference on Hydroinformatics* (pp. 1-23).

[65] Mao, J. Q., Lee, J. H., and Choi, K. W. (2009). The extended Kalman filter for forecast of algal bloom dynamics. *Water research*, 43(17), 4214-4224.

[66] Maybeck, P. S. (1982). *Stochastic models, estimation, and control* (Vol. 3). Academic press.

[67] McKay, M. D. (1988). Sensitivity and uncertainty analysis using a statistical sample of input values. *Uncertainty analysis*, 145-186.

[68] McKay, M.D., Beckman, R.J., and Conover, W.J., (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21 (2), 239–245.

[69] Miller, R. N., and Cane, M. A. (1989). A Kalman filter analysis of sea level height in the tropical Pacific. *Journal of Physical Oceanography, 19, 773–790.*

[70] Miller, R. N., Busalacchi, A. J., and Hackert, E. C. (1995). Sea surface topography fields of the tropical Pacific from data assimilation. *Journal of Geophysical Research: Oceans* (1978–2012), 100(C7), 13389-13425.

[71] Miyoshi, T. (2011). The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Monthly Weather Review*, 139(5), 1519-1535.

[72] Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, 28(2), 135-147.

[73] Moradkhani, H., Hsu, K. L., Gupta, H., and Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, 41(5).

[74] Moradkhani, H. (2008). Hydrologic remote sensing and land surface data assimilation. *Sensors*, 8(5), 2986-3004.

[75] Morse, W. L. (1972). Stream temperature prediction under reduced flow. *Journal of the Hydraulics Division* 98(HY6), 1031-1047.

[76] Murphy, J. M. (1988). The impact of ensemble forecasts on predictability.*Quarterly Journal of the Royal Meteorological Society*, *114*(480), 463-493.

[77] Olioso, A., Chauki, H., Courault, D., and Wigneron, J. P. (1999). Estimation of evapotranspiration and photosynthesis by assimilation of remote sensing data into SVAT models. *Remote Sensing of Environment*, 68(3), 341-356.

[78] Ottlé, C., and Vidal-Madjar, D. (1994). Assimilation of soil moisture inferred from infrared remote sensing in a hydrological model over the HAPEX-MOBILHY region. *Journal of Hydrology*, 158(3), 241-264.

[79] Pauwels, V. R., Hoeben, R., Verhoest, N. E., and De Troch, F. P. (2001). The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. *Journal of Hydrology*, 251(1), 88-102.

[80] Petrie, R. (2008). *Localization in the ensemble Kalman Filter*. Master's thesis, Univ. of Reading, Reading, UK.

[81] Piemonti, A. D., Babbar-Sebens, M., and Jane Luzar, E. (2013). Optimizing conservation practices in watersheds: Do community preferences matter? *Water Resources Research*, 49(10), 6425-6449.

[82] Piccolroaz, S., Toffolon, M., and Majone, B. (2013). A simple lumped model to convert air temperature into surface water temperature in lakes. *Hydrology and Earth System Sciences*, 17(8), 3323-3338.

[83] Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P., Njoku, E. G., and Owe, M. (2007). Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *Journal of Geophysical Research: Atmospheres* (1984–2012), 112(D9).

[84] Rogala, J. T. (1999). *Methodologies employed for bathymetric mapping and sediment characterization as part of the Upper Mississippi River system navigation feasibility study*. US Army Engineer Waterways Experiment Station.

[85] Seo, D. J., Koren, V., and Cajina, N. (2003). Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *Journal of Hydrometeorology*, 4(3), 627-641.

[86] Shen, J., and Haas, L., (2004). Calculating age and residence time in the tidal York River using three-dimensional model experiments. *Estuarine, Coastal and Shelf Science*, 61, 449–461.

[87] Slater, A. G., and Clark, M. P. (2006). Snow data assimilation via an ensemble Kalman filter. *Journal of Hydrometeorology*, 7(3), 478-493

[88] Sinokrot, B. A., and Stefan, H. G. (1994). Stream water-temperature sensitivity to weather and bed parameters. *Journal of Hydraulic Engineering*, 120(6), 722-736.

[89] Sobrino, J. A., Jiménez-Muñoz, J. C., and Paolini, L. (2004). Land surface temperature retrieval from LANDSAT TM 5. *Remote Sensing of environment*, 90(4), 434-440.

[90] Sorooshian, S., Duan, Q., and Gupta, V. K. (1993). Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resources Research*, 29(4), 1185-1194.

[91] Stefan, H. G., and Preud'Homme, E. B. (1993). Stream Temperature Estimation from Air Temperature. *Water Resources Bulletin*, 29, 27-45.

[92] Tetra Tech. (2007). The Environmental Fluid Dynamics Code, User Manual, US EPA Version 1.01.

[93] Tetra Tech, (2009). Hydrodynamic and Water Quality Modeling Report for Lake Lanier," Georgia.

[94] Tedesco, L.P., Pascual, D.L., Shrake, L.K., Hall, R.E., Casey, L.R., Vidon, P.G.F., Hernly, F.V., Salazar, K.A., Barr, R.C., Ulmer, J., Pershing, D., (2005). Eagle Creek Watershed Management Plan: An Integrated Approach to Improved

Water Quality. Eagle Creek Watershed Alliance, CEES Publication 2005-07, IUPUI, Indianapolis, 182p.

[95] Todar, K. (2012). Diversity of Metabolism in Procaryotes. In Online Textbook of Bacteriology.

[96] Troccoli, A., and Haines, K. (1999). Use of the temperature-salinity relation in a data assimilation context. *Journal of Atmospheric and Oceanic Technology*, 16(12), 2011-2025.

[97] True, K., Bolick, A., and Foott, J. S. (2011). FY 2010 Investigational Report: Myxosporean parasite (*Ceratomyxa shasta* and *Parvicapsula minibicornis*) annual prevalence of Infection in Klamath River basin juvenile Chinook salmon, April-August 2010. U.S. *Fish and Wildlife Service California-Nevada Fish Health Center*, Anderson, CA.

[98] True, K., Voss, A., and Foott, J. (2016). "*Myxosporean Parasite* (*Ceratonova shasta* and *Parvicapsula minibicornis*) Prevalence of Infection in Klamath River Basin Juvenile Chinook Salmon, April - July 2015. U.S. Fish and Wildlife Service California – Nev ada Fish Health Center, Anderson, CA.

[99] United States. (2009). Water on tap: What you need to know. Washington, D.C: U.S. Environmental Protection Agency, Office of Water.

[100] United States Environmental Protection Agency (USEPA), "Total Maximum Daily Load for Dissolved Oxygen and Nutrients to Mashapaug Pond," Rhode Island, 2002.

[101] Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of hydrology*, 324(1), 10-23.

[102] Voutilainen, A., Pyhälahti, T., Kallio, K. Y., Pulliainen, J., Haario, H., and Kaipio, J. P. (2007). A filtering approach for estimating lake water quality from remote sensing data. *International journal of applied earth observation and geoinformation*, 9(1), 50-64.

[103] Ward, J. V. (1985). Thermal characteristics of running waters. *Hydrobiologia*, 125, 31-46.

[104] Wang, X., and Bishop, C. H. (2003). A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the atmospheric sciences*, 60(9), 1140-1158.

[105] Whitaker, J. S., and Hamill, T. M. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7), 1913-1924.

[106] Wright, K. A., Goodman, D.H., Som, N.A., and Hardy, T.B. (2014). Development of two-dimensional hydraulic models to predict distribution of Manayunkia speciosa in the Klamath River. *U.S. Fish and Wildlife Service*. Arcata Fish and Wildlife Office, Arcata Fisheries Technical Report Number TR 2014-19, Arcata, California.

[107] Wu, J. (1992). Variation of the heat transfer coefficient with environmental parameters. *Journal of physical oceanography*, 22(3), 293-300.

[108] Wu, G., Zheng, X., Wang, L., Zhang, S., Liang, X., and Li, Y. (2013). A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation. *Quarterly Journal of the Royal Meteorological Society*, 139(672), 795-804.

[109] Xia, M., Xie, L., And Pietrafesa, L.J., (2007). Modeling of the Cape Fear River Estuary Plume, *Estuaries and Coasts*, 30(4), 698-709.

[110] Zhou, J., Falconer, R. A., and Lin, B. (2014). Refinements to the EFDC model for predicting the hydro-environmental impacts of a barrage across the Severn Estuary. *Renewable Energy*, 62, 490-505.