Real Time Flow Estimation in Channel Networks using Lagrangian Data

by

Qingfang Wu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering-Civil and Environmental Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Alexandre Bayen, Chair
Professor Mark Stacey
Professor Roberto Horowitz

Fall 2013

Real Time Flow Estimation in Channel Networks using Lagrangian Data

Copyright Fall 2013

by

Qingfang Wu

Abstract

Real Time Flow Estimation in Channel Networks using Lagrangian Data

by

Qingfang Wu

Doctor of Philosophy in Engineering-Civil and Environmental Engineering

University of California at Berkeley

Professor Alexandre Bayen, Chair

The Sacramento-San Joaquin River Delta in California becomes inadequate in fresh water resources, while the water demand in California keeps increasing. Large-scale numerical flow models, for example Delta Simulation Model II (DSM2) and River, Estuary, And Land Model (REALM), used as crucial water resources management tools, are capable of providing critical information about tidal forcing and salinity transport in the bays and channels of the Delta. Reliable flow estimation and prediction of these models, however, largely depend on an accurate representation of open boundary conditions and initial conditions, which are usually calibrated against historical data sets acquired from Eulerian measurements near the boundaries.

In large watershed, unfortunately, these measurements have demonstrated many intrinsic limitations, for example small spatial coverage and sparse temporal sampling. Also, existing Eulerian sensors have many recorded failures, such as broken gauges, sensor drifts, process leaks, improper measuring devices, and many other random sources. More importantly, if the hydraulic system is radically altered, as in the case of extensive levee failures, the historical data sets can be of limited usage.

In this dissertation, a sensing-modeling system featuring rapidly deployed Lagrangian drifters is developed. The system is capable of predicting regional flows and transport in the Delta in a real-time mode, without dependence on historical data.

Lagrangian data is obtained when floating drifters move along with the flow and report their locations. The data provides instant information about the flow, including flow advections and eddy dispersions, and is further assimilated into underlying shallow water equation (SWE) models to characterize the flow state.

Different approaches to facilitate the flow estimation have been investigated in this dissertation. First, a variational assimilation method (Quadratic Programming) is applied to a 1D SWE model (Linearized Saint-Venant equations). The assimilation method poses the problem of estimating the flow state in a channel network as a quadratic programming by minimizing a quadratic cost function – the norm of the difference between the drifter observations and the model velocity predictions – and expressing the constraints in terms of linearized equalities and inequalities. The problem is then efficiently solved using a fast and robust algorithm. The approach is easy to implement and low in computation costs.

Later, a sequential assimilation method (Ensemble Kalman Filtering) is implemented to a 2D SWE model (depth-integrated Navier-Stokes equations). The assimilation method involves a series of state analysis and updates, where the observed Lagrangian data is incorporated into the state one step at a time to incrementally correct the model prediction. The implementation of this method demands powerful computation ability, and is achieved on high-performance computing clusters at NERSC.

To assess the performance of the proposed data assimilation methods, we investigated a distributed network of channels, subject to quasi-periodic tidal forcing, in the Sacramento-San Joaquin River Delta. Field operational experiments were carried out with a fleet of over 70 floating drifters, deployed within approximately $0.55\,\mathrm{km}^2$ of the river network. During the experiments, more than 325,000 GPS readings were taken from the floating drifters and collected, in real time, onto a central server. It is the first experiment of this kind conducted at such scale, where high-density Lagrangian data have been collected in a real river environment and successfully assimilated over a full tidal cycle.

It is demonstrated that both of the proposed assimilation methods (i.e., QP in 1D SWE model and EnKF in 2D SWE model) can handle the Lagrangian data with sufficient accuracy. In many practical cases, the 1D flow estimation is adequate for water resource management to retrieve critical flow characteristics in a prompt and efficient manner. In the case of complex channel geometry, however, the 2D flow estimation is vital to describe the hydraulic system.

To

Yiyi

Contents

List of Figures **vii**

List of Tables **x**

1 Introduction **1**

1.1 Background and Motivation 1

1.2 Overview of Flow Estimation 2

1.2.1 Flow Measurement Techniques 2

1.2.2 Existing Flow Models 4

1.2.3 Lagrangian Measurements 4

1.2.4 Data Assimilation 5

1.3 Floating Sensor Network 6

1.4 Research on Flow Modeling 6

1.5 Outline of the Dissertation 7

2 Parameter Identification for 1D Shallow Water Equations **9**

2.1 Introduction 9

2.2 System Model and Problem Statement 10

2.2.1 The Saint-Venant Model 10

2.2.2 Linearized Saint-Venant Model 11

2.2.3 Problem Statement 12

2.3 Saint-Venant Distributed Transfer Matrix 15

2.3.1 Uniform Flow Case: Upstream Flow Variables as Boundary Conditions . . 15
2.3.2 Uniform Flow Case: Input Variables as Boundary Conditions 16
2.3.3 Non-Uniform Flow Case: Backwater Approximation 16
2.4 Proposed Approach 18
2.5 Application to the Sacramento Delta 21
2.5.1 Parameter Identification of a River Reach 21
2.5.2 Spectral Analysis 22
2.5.3 Parameter Identification 23
2.5.4 Sensitivity 24
2.5.5 Validation 25
2.5.6 Predictability 25
2.6 Summary 26

3 Channel Network and Data Reconciliation 28

3.1 Introduction 28
3.2 Proposed Method 29
3.2.1 General Considerations 29
3.2.2 Linear Channel Network Model 30
3.2.3 Data Reconciliation 33
3.3 Application to the Sacramento River 34
3.3.1 Description of the System and Assumptions 34
3.3.2 Modal Decomposition of the Measured Data 36
3.3.3 Hydraulic Model of the Sacramento River and Georgiana Slough 38
3.3.4 Data Reconciliation of the Measured Data 40
3.3.5 Method Validation 42
3.4 Summary 53

4 Lagrangian Data in 1D Inverse Modeling 54

4.1 Introduction 54

4.2 Hydrodynamic Model in Tidal Channel Network 55
4.2.1 One-dimensional Numerical Scheme 55
4.2.2 Linear Network Model 57
4.3 Variational Data Assimilation Using Quadratic Programming 58
4.3.1 General Considerations 58
4.3.2 Notations 58
4.3.3 Decision Variables 59
4.4 Twin Test 60
4.4.1 Experiment Protocol 60
4.4.2 2D Shallow Water Equations and Numerical Forward Simulation 61
4.5 Case Study in the Delta 63
4.5.1 Introduction to the Experiment Field 63
4.5.2 Drifter Data Generation 64
4.5.3 Validation Results 65
4.6 Summary 66

5 Variational Data Assimilation in 1D Channel Networks 71

5.1 Introduction 71
5.2 Operational Experiment in the Sacramento River and Georgiana Slough 72
5.2.1 Experiment Field 72
5.2.2 Active Sensors 72
5.2.3 Android Drifters 75
5.3 Data Observation and Drifter Trajectory Filtering 76
5.3.1 Characteristics of Lagrangian Data 76
5.3.2 Overview of the Drifter Data from May 9th experiment 77
5.3.3 Trajectory Filtering 78
5.3.4 Average Velocity across the Channel 83
5.4 Formulation of the Data Assimilation Method 84
5.4.1 “Pseudo-Lagrangian” Data Assimilation 84

5.4.2 Observations and Cost Function 84
5.4.3 Quadratic Program Formulation of Data Assimilation 86
5.5 Data Assimilation Results 86
5.5.1 Overview 86
5.5.2 Assimilation with Lagrangian Data using Different Filtering Windows . . . 88
5.5.3 Assimilation with Both Lagrangian and Eulerian Data 89
5.5.4 Validation with Reference Drifters 89
5.5.5 Discussions 93
5.6 Summary 93

6 Real-Time Lagrangian Data Assimilation using EnKF in 2D SWE 96
6.1 Introduction 96
6.2 REALM Model 97
6.2.1 Introduction to REALM 97
6.2.2 2D Shallow Water Equations 97
6.3 Ensemble Kalman Filter 99
6.3.1 Kalman Filter and Extended Kalman Filter 99
6.3.2 Ensemble Kalman Filter 101
6.4 Application to Experimental Data 102
6.4.1 Model Error 102
6.4.2 Instability 103
6.5 Implementation of the Floating Sensor Network System 104
6.5.1 Communication Architecture 104
6.5.2 High Performance Computing 105
6.6 Data Assimilation and Results 107
6.6.1 Estimated State 107
6.6.2 Sensitivity Studies 109
6.7 Summary 112

7 Conclusion **113**

7.1 Summary of the Research 113

7.2 Highlights of the Work 115

7.3 Suggestions for Future Work 115

Bibliography **117**

List of Figures

1.1 *Delta Tidal Flows.* 2

2.1 *Backwater curve approximation.* 18

2.2 *DWR satellite stations along San Joaquin River.* 22

2.3 *Stage Analysis at SJL Station.* 23

2.4 *Parameter Identification: model output v.s. measurement.* 24

2.5 *Sensitivity analysis for the application case.* 25

2.6 *Model Validation: model output v.s. measurement.* 25

2.7 *Model Predictability during 12/18/2006 and 2/18/2007: model output v.s. measurement.* 26

3.1 *Flow compatibility of channel junctions.* 31

3.2 *Experimental area in the Sacramento River and the Georgiana Slough.* 35

3.3 *Spectral analysis of the discharge at the SDC Station.* 37

3.4 *Relative percent error between measurement data and modal representions as a function of the number of modes chosen for the decomposition.* 37

3.5 *Reconciliated boundary condition data vs. measured data.* 43

3.6 *Reconciliated boundary condition data vs. measured data.* 44

3.7 *The difference between Reconciliated data and measured data.* 45

3.8 *Sensors in the experiment area.* 46

3.9 *Validation of the model output with measurement using USGS measurements.* 47

3.10 *Validation of the model output with measurements using reconciliated BC.* 49

3.11 *Error distribution between the reconciliated data and USGS measurements (perturbation discharge at DLC is added on purpose in the data set).* 51

3.12 *Error distribution between reconciliated and measured data (Discharge at DLC is removed from the data reconciliation process).* 52

4.1 *Spectral analysis of the discharge at the SDC Station.* 60

4.2 *Data Assimilation Flow Diagram.* 61

4.3 *Experiment area in the Sacramento River (*1*), the Cache Slough (*2*), the Steamboat Slough (*3*), the Sutter Slough (*4*), the Minor Slough (*5*), the Delta Cross Channel (*6*) and the Georgiana Slough (*7*).* 64

4.4 *Drifter trajectories and their release positions.* 68

4.5 *Comparison of the estimated boundary condition with USGS measurements at the boundaries of the domain.* 69

4.6 *Validation of the model outputs with USGS and ADCP measurements inside of the domain.* 70

5.1 *Annotated map of the Walnut Grove experimental area, for the May 9th field operational test.* 73

5.2 *Overview of the experimental timeframe, the two grey boxes denote the time domain during which the drifter data was collected.* 73

5.3 *Android drifter usage during the May 9th experiment. Photo credit: Berkeley Lab - Roy Kaltschmidt.* 74

5.4 *CAD drawing of active drifting sensor. Photo Credit: Carlos Oroza* 75

5.5 *Android drifter with complete parts list used for the study. Photo credit: Jonathan Beard* 76

5.6 *Trajectory of drifter on 05/09/2012.* 79

5.7 *Drifter passing by GSS station 3 times on 05/09/2012.* 80

5.8 *Details of the second pass-by.* 81

5.9 *Spatial-Time window to remove perturbations from Lagrangian drifters.* 82

5.10 *Velocity profile in a cross-section, the average velocity across the river cross section* $\overline{u}(X,t)$ *is calculated with the water column mean velocity* $u_{column}(X,t)$ *at each cell Eq.* (5.3). 84

5.11 *Snapshots of the Android drifter fleet during the May 9th, 2012 experiment, at two selected time in the field operational test.* 87

5.12 *Flow chart for the data assimilation. Drifter data is devided into two parts which include observation (used for the data assimilation) and validation.* 88

5.13 *Assimilated flow velocity near USGS stations (a) GES (b) GSS using 30-second space-time window.* 90

5.14 *Assimilated flow velocity near USGS stations (a) GES (b) GSS using 150-second space-time window.* 91

5.15 *Assimilated flow velocity near around USGS stations (a) GES (b) GSS using filtered drifter data from 150-second space-time window and Eulerian depth sensor.* . . 92

5.16 *Assimilated flow velocity at the locations of the reference drifters using filtered drifter data from 150-second space-time window and Eulerian depth sensor.* 94

6.1 *Communication architecture, showing the flow of data from drifters in the field to the database server and computation servers via the GSM service.* 104

6.2 *A schematic description of the REALM/EnKF assimilation system.* 106

6.3 *Dataflow from drifters in the field to the Carver cluster at NERSC.* 107

6.4 *Quiver plots of assimilation results for flow fields during May 9 experiment.* 108

6.5 *Exploration of impact of inflation factor* 109

6.6 *Exploration of impact of ensemble size with an inflation factor of* 1.2. 110

6.7 *EnKF data assimilation results v.s. USGS measurements.* 111

List of Tables

2.1 Parameters of the San Joaquin River. 23

3.1 Parameters for the Sacramento River and the Georgiana Slough. 40

3.2 Max-value, ρ-value and E-value for reconciliated data and measured data. 42

3.3 ρ-value and E-value for model validation with measured boundaries. 48

3.4 ρ-value and E-value for model validation after reconciliation. 48

3.5 χ^2 value for different cases. 50

3.6 ρ-value and E-value for modal validation (with and without discharge @ DLC). . . 53

4.1 Dominant Tidal Modes in the Delta. 59

4.2 Max-value, ρ-value and E-value for modeled data and measured data. 66

5.1 ℓ^2-value for assimilated flow velocity and Eulerian sensors. 93

Acknowledgements

I would like to give my deepest gratitude to my advisor, Professor Alexandre Bayen, for his foreseeing guidance, impressive enthusiasm, and continuous encouragement. I am forever indebted to Professor Bayen, who has been supporting and mentoring me for all aspects of my professional and personal development throughout my PhD studies.

I would like to thank Professors Mark Stacey and Roberto Horowitz for being my doctoral committee member. I would also thank Professors Alexandre Chorin, Tina Chow, Steven Glaser for being my qualifying committee member.

I gratefully thank my mentors at UC Berkeley and LBNL. Professors Mark Stacey and Tina Chow helped me through the Environmental Engineering curriculum. Professor Xavier Litrico enlightened me on inverse modeling. Drs. Eli Ateljevich and Peter Schwartz guided me into the REALM development.

I would thank many LBNL scientists, Professors Katherine Yelik and Phillip Colella in particular, who kindly provided me valuable computing resources in the research. I also thank Professor Armen Der Kiureghian for his support.

I thank the staff members from the Civil and Environmental department, Shelley Okimoto, Sanjay Govindjee, Jenna Tower, for their help through the curriculum and requirements. I also thank the staff members at CCIT, Joe Butler, Saneesh Apte, Daniel Edwards, Luis Torres, for their support in the Floating Sensor Project.

I am very lucky to work in the Bayen lab at UC Berkeley. All the colleagues are friendly, cooperative and willing to help. Andrew Tinka, Jonathan Beard, Kevin Weekly, and Carlos Oroza provided experimental drifter data. Julie Percelay and Olli-Pekka Tossavainen implemented the Ensemble Kalman Filtering algorithm. Christian Claudel, Sebastien Blandin, Dengfeng Sun, Saurabh Amin, Dan Work, Leah Anderson, Samitha Samaranayake, and many more, are so nice to work with.

Finally, from the depth of my heart, I would thank my family, for their profound love and enduring support.

Chapter 1

Introduction

1.1 Background and Motivation

The Sacramento-San Joaquin Delta (the Delta) is the critical part of the California water resource system and provides water supply to approximately two-thirds of the state population. The delta is a complex network of over $1150\ km$ of tidally-influenced channels and sloughs that are tidally driven by the San Francisco Bay. The San Joaquin River, with a length of $530\ km$, is the second-longest river in the Delta area. Fresh water enters the delta from the north, and interacts with high salinity water from the San Francisco Bay through the San Joaquin River. The transport and interaction of fresh and sea water establish the quality of water supply for municipal consumption, which is mainly extracted from the South Delta (Figure 1.1).

The flow and transport in the Delta are not fully understood yet, due to the tidal-driven flows, the complex geometry, and the high dispersion [107]. It is broadly accepted that the highly dispersive environment in the Delta is created by the different phasing of tidal flows in intersecting channels and dominated by tidal trapping [87]. In this transport and dispersive system, the interaction of tidal motions within the channels and watersheds of the Delta results in salt movement into the Delta. To quantify the extent of water quality and salt intrusion, critical facts and parameters, such as the freshwater flow, the effective dispersion coefficient for salt, and in-Delta operations, need to be well-defined and meant to capture the effects of tidal dispersion processes.

As the hub of the California water delivery system, the Delta requires an efficient management and response system for water distribution and unanticipated events, such as levee breaks, sudden change in fresh water, etc. Without full knowledge of what is happening in the hydraulic system, policy makers are forced to make conservative operational decisions to preserve water quality and secure water supply. This overly conservative approach may lead to substantial inefficiencies in

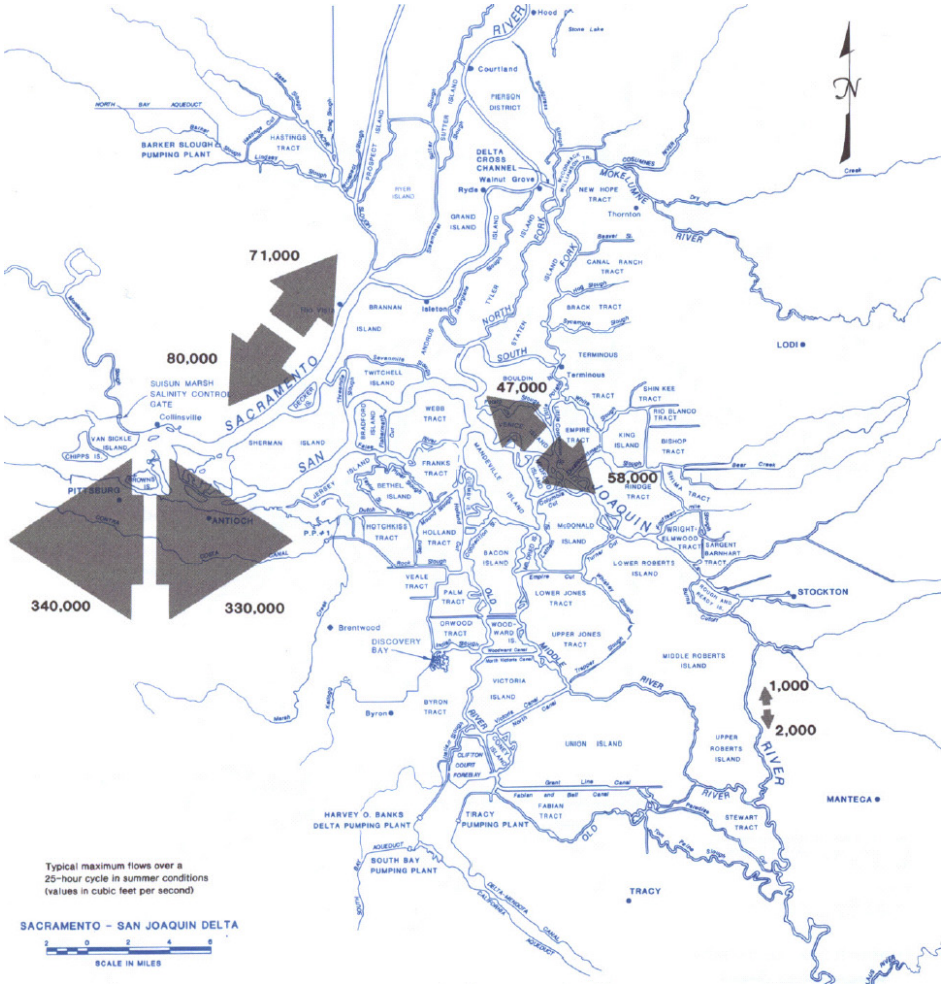


Figure 1.1: *Delta Tidal Flows.*

water distribution and unnecessary economic loss.

1.2 Overview of Flow Estimation

An efficient flow management demands accurate and prompt flow measurements and estimation at representative locations in the channel network, along with an effective and robust flow model capable of providing fast and reliable predictions.

1.2.1 Flow Measurement Techniques

Conventional flow measurement techniques are classified as either remote or in-situ sensing with regard to the subject flow they are measuring.

Remote sensing refers to techniques in which a sensing device is not in direct contact with the flow, and the measurements are usually taken from distance. One example of remote sensing techniques is satellite imagery, using microwave radars or synthetic aperture radars [18], [135], [91], [21]. Other widely-applied techniques include coherent microwave systems and ultra high frequency (UHF) radar from helicopters, riverbanks, bridges, or cableways [122], [152], [83].

In contrast, in-situ sensing refers to techniques where a sensing device is in direct contact with the flow. It can be further categorized into Eulerian or Lagrangian techniques (the terminologies are borrowed from hydrodynamics, which are defined with respect to different reference frames). Eulerian techniques utilize sensors that are fixed to an external reference frame (e.g., the river bank) and take measurements when the flow passes by. Lagrangian techniques feature sensors that drift freely along with the flow and conduct measurements as they move through it.

The primary data collected by a Lagrangian sensor is its position over time, although some Lagrangian sensors are capable of characterizing other flow properties, e.g., temperature, salinity, dissolved constituents, etc. A well-designed Lagrangian sensor should behave like an "ideal particle" in the flow, enabling a direct quantification of the flow velocity from the time series of its position. In hydrodynamics, such sensor devices are often called *drifters*. The design and fabrication of drifters have always been improved along with the advancement of the positioning and communication technologies.

The first drifter that could actively report its position was the "Swallow float", which was first operated in 1950s [141]. It was a neutrally buoyant device that would drift approximately 1,000 meters along a river while emitting acoustic pulses at a constant rate. These pulses were received by hydrophones pre-installed on the bank, and thus the float was readily located during its drifting. Development of drifters with acoustic communication capabilities continued in the 1960s and 1970s [61]. In 1978, the Argos satellite system [28] started to provide a global location and data uplink service, enabling oceanographic researchers to develop drifters that could continuously report their position and transfer sensor data during the mission. In the mid-1980s, many oceanographic drifters leveraged the Argos system, including the Davis (a.k.a. Coastal Dynamics Experiment drifter) [41], the Ministar (a.k.a. World Ocean Climate Experiment drifter) [114], and the Low Cost Tropical Drifter [17]. In recent years, many sensor networks have been developed for aquatic sensing missions. Some representative work includes the AMOUR project at MIT [45], the NEPTUS framework of AUVs at LSTS in Portugal [140], the submersible pneumatic drogues built at UCSD [66], the Slocum underwater drifters at MBARI [85], and the SmartBay sensor network project in Galway Bay [132]. The Floating Sensor Network project at UC Berkeley [1], to be introduced in later sections, designs and builds drifters primarily for riverine and estuarine environments.

1.2.2 Existing Flow Models

Large-scale numerical flow models, such as DSM2 [22] and REALM [11], sponsored by the California Department of Water Resources (DWR), have been used as crucial water resources management tools, providing critical information about tidal forcing and salinity transport in the bays and channels of the Delta. These models are based on mechanistic analysis of the hydrodynamics at different timescales.

A number of factors may affect the performance of these state-of-the-art models, such as mesh generation, flow discretization, and choice of numerical solver. More importantly, the performance of the models largely relies on the determination of open boundary conditions and the parameter calibrations against historical data sets.

Traditionally, the open boundary conditions can be obtained either via Eulerian observations near the boundaries, for example tidal gauge data, or through satellite data retrieval. Unfortunately, these measurements at large watershed have their intrinsic limitations, specifically small coverage and sparse sampling [105]. Furthermore, installed Eulerian sensors are proven to have many failures, such as broken gauges, process leaks, sensor drifts, improper use of measuring devices, and other random sources.

As the tidal phase differences in the Delta channels are remarkable, the existing flow models are usually calibrated against historical data sets in a timescale of months. Such a dependence on historical data sets, however, would make a modeling system less responsive to abrupt events, for example a massive levee breach or large adjustments in freshwater flow, in which historical data provides limited information for the rapidly changed situation, making it very difficult, if not entirely impossible, to estimate or predict the flow state.

1.2.3 Lagrangian Measurements

In the last two decades, techniques using surface and subsurface Lagrangian buoys have significantly advanced. Lagrangian data, in particular those collected from surface drifters, provide instant information about the flow, and can be used to describe flow advections and eddy dispersions. For this reason, Lagrangian measurements have been highly valued and extensively used in numerous meteorological and oceanic systems [8], [124], [84]. They are also used in river hydraulics [106], and in traffic modeling (i.e., mobile sensors) [69].

A major difficulty of utilizing the Lagrangian data acquired in field is that the data is often affected by local flow perturbations, and therefore exhibits so “noisy” that it can essentially mask any important flow characteristics. Data assimilation technique, as briefly introduced in the next section,

provides an effective measure to exclude these perturbations from flow measurements. By the use of the data assimilation technique, the measured data largely compensate for poorly-identified parameters of underlying physical model, such as boundary conditions, initial conditions, and physical processes that are not incorporated into the model [116], [103], [9], [24], [64].

Addressing these limitations, we present an efficient and reasonable velocity field based on the Lagrangian data collected from the large-scale drifter experiment using the data assimilation techniques implemented in this dissertation. It is one of the first successes of using Lagrangian data in a hydrodynamic system, especially in a complex tidal-driven channel network. This topic is of particular interest under conditions that Eulerian sensors are sparsely located, or flow conditions are changed rapidly, such as flood or dam break events.

1.2.4 Data Assimilation

Data assimilation is the process by which observations are incorporated into an underlying physical model of a real system. The assimilation results are considered reliable, not only because they are close to the observations at the required spatial and temporal scales, but also because they are constrained by certain dynamical or statistical relationships defined in the real physical world.

Most data assimilation methods available nowadays can be categorized into two groups: variational assimilation methods [112] or sequential assimilation methods [149]. Variational assimilation methods, such as Quadratic Programming [109], 3D-Var [169], 4D-Var [125], adjoint approach [16] etc., perform a single optimization on all the observed data to minimize a cost function. Sequential assimilation methods, on the other hand, conduct a series of state estimates and updates to integrate the observed data into the state estimate one step at a time. Example of sequential assimilation methods include: the Kalman filter [81], [126], which recursively processes noisy input data and eventually provides a statistically optimal state estimate of the underlying system; the Extended Kalman filter [4], which updates the estimate of the mean and covariance of the state by computing the Jacobian matrices of the state update equations; the Ensemble Kalman filter [50], [148], [52], which tracks the evolution of a number of random samples to update the state estimates; and the Unscented Kalman filter [151], which deterministically generates a minimal ensemble of samples and accurately tracks the mean and covariance.

Data assimilation can also be approached in other ways, for example optimal statistical interpolation [104], and Newtonian relaxation (a.k.a. nudging) [120].

1.3 Floating Sensor Network

The Floating Sensor Network project at UC Berkeley aims at designing and implementing an integrated observational and modeling system to provide capabilities to respond to uncertain events in hard-to-map waterways and allow management of the system at different timescales [145]. The ultimate goal of the project is to monitor real-time flows and predict transport in subsections of the Delta, and furthermore improve the efficiency of Delta operations.

The observational system consists of a set of Lagrangian drifters (developed at UC Berkeley) and existing Eulerian (fixed) sensors installed by the U.S. Geological Survey (USGS) and the California Department of Water Resources (DWR). These Lagrangian drifters can communicate both among themselves and with a base station. As a drifter floats through a river, the GPS data that it gathers gives a snapshot of the water flow conditions, conducting in-situ measurements of Lagrangian velocity and salinity, temperature, or other chemicals of interest. Leveraging commercial cellular networks could drastically cut the ongoing tasks of water flow monitoring and expand coverage to the thousands of miles of natural and man-made channels in the Delta which are currently under-monitored.

The modeling and data assimilation system includes (a) tidal boundary conditions approximation using Eulerian observations and (b) flow state estimation using real time Lagrangian measurements. In Part (a), the tidal boundary conditions is estimated by inverse approach to hydrodynamic modeling based on Eulerian observations from the interior or exterior of the model domain. These coarsely estimated boundary conditions are then used to drive a forward model and provide a preliminary description of the Delta flow state. The flow estimation and predication are further refined in Part (b), by incorporating Lagrangian drifter data collected and post-processed with data assimilation technique inside the model domain.

1.4 Research on Flow Modeling

This dissertation emphasizes on devoleping the modeling system of the Floating Sensor Network. The main research work is divided into five sequential phases. Each of them contributes a part of the real-time monitoring system and explores an aspect of research question or challenge of the Delta. The tasks focus on using enhanced data sets, statistical techniques, and numerical modeling to build up the flow modeling system:

- Explore the application and limitations of 1D shallow water equations (SWE) in tidal-driven channel using data from USGS stations (Phase 1).

- Present a methodology to detect and remove the field measurement errors for a channel network in the Delta (Phase 2).
- Incorporate Lagrangian drifter data measurements into a large channel network simulated by 1D shallow water equations, and find the optimal drifter deployment strategy for large channel networks (Phase 3).
- Improve River, Estuary, and Land Model (REALM), a new 2D hydrodynamic model initiated and funded by the California Department of Water Resources, for managing short-term operations in the Sacramento-San Joaquin Delta (Phase 4).
- Construct a real time flow monitoring and visualization system for the Delta by using the customized REALM model and the data collected by rapidly-deployed Lagrangian drifters (Phase 5).

1.5 Outline of the Dissertation

The dissertation is organized as follows:

Chapter 2 presents a technique to identify flow parameters for a simple river subjected to periodic tidal forces. The river is simplified as a single channel, and characterized with a 1D SWE model. The boundary conditions at upstream and downstream are determined from Eulerian measurements. Flow parameters are identified in the frequency domain with modal decomposition to reduce the computational costs. The chapter summarizes the research work in Phase 1, and is the basis of Eulerian flow estimations in the subsequent chapters.

Chapter 3 further develops the flow estimation method initiated in the previous chapter, featuring a 1D SWE model and Eulerian measurements. The hydraulic model is now applied to a channel network, which is closer to the reality in the Delta. Also, the measurement errors at boundaries are largely eliminated with data reconciliation technique using information redundancy. The chapter summarizes the research in Phase 2.

Chapter 4 proposes a novel method to retrieve the flow state of a channel network of arbitrary topography using Lagrangian data. The approach is formulated as a 1D Quadratic Programming (QP) problem based on minimizing the norm of the difference between Lagrangian data and modeled drifter trajectories, constrained by a 1D implicit linear channel network model. The proposed method is first tested in a twin experiment, in which the Lagragian data are actually generated by a 2D nonlinear shallow water model (TELEMAC) .

Chapter 5 further validates the proposed 1D QP method with real Lagrangian flow data acquired

in a drifter experiment in the Sacramento River near its junction with the Georgiana Slough. The above two chapters cover the research in Phase 3.

Chapter 6 designs and implements a real-time flow monitoring system to provide accurate descriptions of varying flow in the Delta, using a combination of Eulerian and Lagrangian computations. The underlying flow model is based on 2D shallow water equations. The data assimilation technique we chose here are Ensemble Kalman Filter. This chapter primarily reflects the research in Phase 5, and is the ultimate goal of the Delta flow study. In fact, all the previous phases are essentially the preparation work: Phases 1 and 2 estimate the boundary conditions of the channel network of interest using a simple 1D flow model and Eulerian sensors; Phase 3 tests the applicability of Lagrangian drifter data, and Phase 4 introduces 2D SWE model (REALM) into the system.

Chapter 7 summarizes the flow estimation research completed and outlines some suggestions for future work.

Chapter 2

Parameter Identification for 1D Shallow Water Equations

2.1 Introduction

Parameter identification for *partial differential equations* (PDEs) is a well-studied problem in open-channel hydraulic systems [167], [166], [86], [60]. The flow dynamics of these systems are governed by *shallow water equations* (SWE) which are nonlinear, hyperbolic PDEs, also known as the Saint-Venant equations [26], [168], [133], [58]. In numerous practical applications, these equations are linearized around some reference trajectory. If the linearization is done about a gradually varied steady state profile, the coefficients of the linearized PDE are spatially varying [97]. These coefficients are highly-nonlinear functions of the physical system parameters such as the canal geometry, the friction factor as well as the steady state boundary conditions. An accurate knowledge of these physical parameters is essential for understanding the flow dynamics in natural river systems [134] and for efficiently controlling the channel networks [99], [59], [96].

However, in most natural river systems, these parameters are unknown and have to be estimated from the flow variables measured by (fixed) Eulerian sensors [47] or (moving) Lagrangian sensors [116]. Even in the case of canal systems with non-rigid cross-sections, these parameters may widely vary over the time of operation. Furthermore, the sensors used for monitoring flow variables are often sparsely placed and have limited sensing abilities. It then becomes important to investigate techniques for accurate identification of the physical parameters of open-channel hydraulic systems.

In this chapter, the parameter identification in a linearized SWE subjected to periodic forcing is studied. This setting can model the effect of tidal excitations on the flow dynamics of long river

reaches located much upstream of the river-ocean confluence. In the current setting, the data of the problem is the flow variables that are monitored at the upstream and downstream ends as well as at selected intermediate locations. The parameters to be estimated are: the canal width, the bed slope, the roughness coefficient and the steady state upstream and downstream boundary conditions. The problem is to find the parameter values such that the estimated linear SWE describes the measured data as well as possible. The spatial dependence of these parameters is not considered in the current work.

A common approach to achieve this objective is to choose the parameter values that minimize the output least squares fits between the model predictions and the measured data [92]. However, this formulation involves minimization of a cost function, subjected to PDE constraints and it may be difficult to solve in a numerically efficient manner [47], [10]. In this chapter, tools from the frequency domain modeling of linear PDEs are used to convert the PDE constrained minimization problem to one without any PDE constraints. By modal analysis, the input data is expressed in terms of a small number of Fourier modes. The spatially dependent transfer matrix relating the input variables to the output variables is then used to obtain a parameterized prediction of the output response for the chosen set of modal inputs. The optimal parameters can be obtained by minimizing the squared deviations of the output predictions from the measured data. This minimization problem is considerably easier to solve in comparison to the original PDE constrained minimization problem. The work in this chapter has been published in [160].

2.2 System Model and Problem Statement

2.2.1 The Saint-Venant Model

The Saint-Venant equations are quasi-linear hyperbolic PDEs that describe the dynamics of one-dimensional flow in open-channel hydraulic systems [26], [39]. For a rectangular cross-section, these equations are given by:

$$TY_t + Q_x = 0 \tag{2.1}$$

$$Q_t + \left(\frac{Q^2}{TY} + \frac{gTY^2}{2}\right)_x + gTY\left(S_f - S_b\right) = 0 \tag{2.2}$$

for $(x,t) \in (0,X) \times \Re^+$, where X is the river reach m, $Q(x,t)$ is the discharge (m^3/s) across cross-section $A(x,t) = TY(x,t)$, $Y(x,t)$ is the stage or water-depth (m), T is the free surface width (m) (which is constant for rectangular cross-section), $S_f(x,t)$ is the friction slope (m/m),

S_b is the bed slope m/m, and g is the gravitational acceleration (m/s^2). Also, $V(x,t)$ is the average velocity (m/s) in the cross-section A defined by $V = Q/A$, and $P(x,t)$ is the wetted perimeter (m) defined by $P(x,t) = T + 2Y(x,t)$. The boundary conditions are $Q(0,t) = Q_0(t)$ and $Y(X,t) = Y_0(t)$. The initial conditions are given by $Q(x,0)$ and $Y(x,0)$ for $x \in [0,X]$. The friction slope is empirically modeled by the Manning-Strickler's formula

$$S_f = \frac{Q^2 n^2 (T+2Y)^{4/3}}{(TY)^{10/3}} \tag{2.3}$$

with n the Manning's roughness coefficient $(sm^{-1/3})$.

Under constant boundary conditions, equations (2.1,2.2) admit a *steady state solution*. Let the flow variables corresponding to the steady state condition be denoted by $Q_0(x)$, $Y_0(x)$ etc. where $x \in [0,X]$. The steady state equations are given by

$$\frac{dQ_0(x)}{dx} = 0 \tag{2.4}$$

$$\frac{dY_0(x)}{dx} = \frac{S_b - S_{f0}}{1 - F_0(x)^2} \tag{2.5}$$

with the wave celerity $C_0 = \sqrt{gY_0}$, the Froude number $F_0 = V_0/C_0$, and the steady state velocity $V_0 = Q_0/A_0$. While $Q_0(x) = Q_0 = Q_X$ by the first equation, the second equation is solved for $Y_0(x)$ with boundary condition in terms of downstream elevation $Y_0(X)$. In this chapter, we assume the flow to be *sub-critical*, i.e., $F_0 < 1$.

Remark 1. *In the case of uniform flow, the flow variables are constant along the length of the channel: the discharge $Q_0(x) = Q_0 = Q_X$ and the stage $Y_0(x) = Y_n$ (called the normal depth) can be deduced by solving the normal depth equation $S_{f0} = S_b$.*

2.2.2 Linearized Saint-Venant Model

Equation (2.2) of the Saint-Venant model is highly nonlinear in the flow variables Q and Y. Each term $f(Q,Y)$ in the Saint-Venant model can be expanded in Taylor series around the steady state flow variables $Q_0(x)$ and $Y_0(x)$. Considering only the first order perturbations: $f(Q,Y) \approx f(Q_0,Y_0) + (f_Q)_0\, q(x,t) + (f_Y)_0\, y(x,t)$, the first order perturbation in discharge (resp. stage) is given by $q(x,t) = Q(x,t) - Q_0(x)$ (resp. $y(x,t) = Y(x,t) - Y_0(x)$). From [93], the linearized Saint-Venant model for the perturbed flow variables q and y is:

$$T_0 y_t + q_x = 0 \tag{2.6}$$

$$q_t + 2V_0(x)q_x - \beta_0(x)q + \alpha_0(x)y_x - \gamma_0(x)y = 0 \tag{2.7}$$

with $\alpha_0(x)$, $\beta_0(x)$ and $\gamma_0(x)$ given by:

$$\alpha_0 = (C_0^2 - V_0^2)T_0 \tag{2.8}$$

$$\beta_0 = -\frac{2g}{V_0}\left(S_b - \frac{dY_0}{dx}\right) \tag{2.9}$$

$$\gamma_0 = gT_0\left[(1+\kappa_0)S_b - (1+\kappa_0-(\kappa_0-2)F_0^2)\frac{dY_0}{dx}\right] \tag{2.10}$$

with $\kappa_0 = 7/3 - 8Y_0/(3(2Y_0+T))$.

In the above equations, to emphasize that the free surface width T is uniform, it is denoted as T_0 and the dependence on x is omitted for readability. The upstream and downstream boundary conditions are respectively given by the upstream discharge perturbation $q(0,t)$ and the downstream stage perturbation $y(X,t)$. The initial conditions are given by $y(x,0) = 0$ and $q(x,0) = 0$ for all $x \in [0,X]$.

The linearized Saint-Venant model (2.6,2.7) can be written in the following form:

$$u_t = \mathcal{A}(x)u \tag{2.11}$$

where u is defined by the two-dimensional vector function $u(x,t) = (u_1(x,t), u_2(x,t))^T := (q(x,t), y(x,t))^T$ defined on $(x,t) \in \Re^+$, and $\mathcal{A}(x)$ denotes the linear operator given by:

$$\mathcal{A}(x) = \begin{pmatrix} 0 & \frac{1}{T_0} \\ \alpha_0(x) & 2V_0(x) \end{pmatrix}\frac{\partial}{\partial x} + \begin{pmatrix} 0 & 0 \\ \gamma_0(x) & \beta_0(x) \end{pmatrix} \tag{2.12}$$

The boundary conditions of (2.11) are given by

$$u_1(0,t), u_2(X,t) \tag{2.13}$$

and initial conditions are given by

$$u(x,0) = 0, \forall x \in [0,X] \tag{2.14}$$

2.2.3 Problem Statement

The boundary conditions $u_1(0,t) = q(0,t)$ and $u_2(X,t) = y(X,t)$ are the measured input variables, while the downstream discharge perturbation $u_1(X,t) = q(X,t)$ and the upstream stage perturbation $u_2(0,t) = y(0,t)$ are the *measured output variables*. The measured input and output variables are assumed to be known from a reference time 0 to a final time τ. Thus, for the linear PDE system (2.11), the measured data is the vector function $u(x,t)$ recorded at $x = 0$ and $x = X$.

The properties of the measured data depend on a variety of factors such as the environment, sensor characteristics, etc. In addition, it is assumed that $u(x,0) = 0$ for all $x \in [0, X]$.

The Manning roughness coefficient n, the free surface width T_0, the bed slope S_b, the steady state discharge Q_0 and the steady state downstream stage $Y_0(X)$ are the *unknown parameters* of the system (2.11). We define the parameter vector $\theta := (T_0, S_b, n, Q_0, Y_0(X))^T$ that ranges over the set:

$$\begin{aligned} \mathcal{D}_{\mathcal{A}} := &\left[\underline{T_0}, \overline{T_0}\right] \times \left[\underline{S_b}, \overline{S_b}\right] \times [\underline{n}, \overline{n}] \\ &\times \left[\underline{Q_0}, \overline{Q_0}\right] \times \left[\underline{Y}_0(X), \overline{Y}_0(X)\right] \end{aligned} \tag{2.15}$$

Remark 2. *For the case of uniform flow, since $Y_0(x) = Y_n$, the linear operator $\mathcal{A}(x)$ does not depend on the Manning's roughness coefficient n. Hence, the set of parameter vectors reduces to $\mathcal{D}^n_{\mathcal{A}} := \left[\underline{T_0}, \overline{T_0}\right] \times \left[\underline{S_b}, \overline{S_b}\right] \times \left[\underline{Q_0}, \overline{Q_0}\right] \times \left[\underline{Y_n}, \overline{Y_n}\right]$. The roughness coefficients are estimated using the normal depth equation.*

The linear operator $\mathcal{A}(x)$ can be parameterized by θ that ranges over the parameter set $\mathcal{D}_{\mathcal{A}}$. We define the set of all linear operators allowed by $\mathcal{D}_{\mathcal{A}}$ as $\mathbf{A}(x)$, which is given by:

$$\mathbf{A}(x) := \{\mathcal{A}(\theta; x) | \theta \in \mathcal{D}_{\mathcal{A}}\} \tag{2.16}$$

where $\mathcal{A}(\theta; x)$ denotes the linear operator for a given θ. To emphasize the dependence on θ of linearized Saint-Venant model, (2.11) is written as:

$$u_t = \mathcal{A}(\theta; x)u \tag{2.17}$$

We now define the forward and the inverse problems.

Definition 1. *(Forward simulation). For a given θ, find u satisfying* (2.17) *with boundary conditions* (2.13) *and initial conditions* (2.14)*.*

Definition 2. *(Parameter identification). Given the set of linear operators $\mathbf{A}(x)$, recover the parameter set θ that* optimally *fits the model* (2.17)*, with boundary conditions* (2.13) *and initial conditions* (2.14)*, to the measured data $u(X,t)$ and $u(0,t)$ recorded for all $t \in [0, \tau]$.*

Often, the forward simulation problem is simply called the *forward problem* and the parameter identification problem is called the *inverse problem*. The identified parameter vector in the problem will be denoted as $\hat{\theta}$. In the definition of *inverse problem*, the optimality of the identified parameter vector $\hat{\theta}$ with respect to the given data needs to be defined. It should be noted that once the parameter identification problem is solved for $\hat{\theta}$, the forward simulation problem can be solved to obtain the *predicted output variables*: $\hat{u}_1(X, t|\theta) = \hat{q}(X, t|\theta)$ and $\hat{u}_2(0, t|\theta) = \hat{y}(0, t|\theta)$. The

notation $\hat{u}_1(X,t|\theta)$ indicates the prediction of $u_1(X,t)$ by the PDE model under appropriate initial and boundary conditions and knowledge of the parameter vector θ.

We define a scalar valued cost function as:

$$\begin{aligned}&\mathcal{J}(\theta;\tau;u(0,\cdot),u(X,\cdot))\\&=\int_0^\tau \frac{w_1}{u_{1,\text{norm}}^2}\left(u_1(X,t)-\hat{u}_1(X,t|\theta)\right)^2 dt\\&+\int_0^\tau \frac{w_2}{u_{2,\text{norm}}^2}\left(u_2(0,t)-\hat{u}_2(0,t|\theta)\right)^2 dt\end{aligned} \tag{2.18}$$

where w_1 and w_2 denote the weighing factors for the cost function, and $u_{1,\text{norm}}$ and $u_{2,\text{norm}}$ denote normalizing coefficients of the cost function. Here, it is assumed that the data $u(0,t)$ and $u(0,X)$ is recorded for $t \in [0,\tau]$.

It is important to note that equation (2.18) implies that for any given θ and τ, the evaluation of the cost function would require solving the forward problem, i.e., to solve (2.17) with boundary conditions (2.13) and initial conditions (2.14). If $\phi(x,t|\theta)$, $(x,t) \in [0,X]\times[0,\tau]$ is the solution of the forward problem, the predictions can be obtained by assigning $\hat{u}_1(X,t|\theta)=\phi(X,t|\theta)$ and $\hat{u}_2(0,t|\theta)=\phi(0,t|\theta)$ for $t\in[0,\tau]$. Thus, the parameter identification problem can be stated as follows:

Problem 1. *Solve the minimization problem:*

$$\hat{\theta}=\arg\min_{\theta\in\mathcal{D}_{\mathcal{A}}} \mathcal{J}(\theta;\tau;u(0,\cdot),u(X,\cdot)) \tag{2.19}$$

with $\mathcal{J}(\theta;\tau;u(0,\cdot),u(X,\cdot))$ *given by* (2.18) *and subject to the PDE constraint:* $\hat{u}_1(X,t|\theta)=\phi(X,t|\theta)$ *and* $\hat{u}_2(0,t|\theta)=\phi(0,t|\theta)$*, where* $\phi(x,t|\theta)$ *is the solution of the forward problem.*

The resulting estimate of the linearized Saint-Venant system is given by:

$$u_t=\mathcal{A}(\hat{\theta};x)u \tag{2.20}$$

with boundary conditions $u_1(0,t)$ and $u_2(X,t)$ and zero initial conditions.

Remark 3. *If* K *copies of the input and output variables are recorded, the cost function can then be written as:*

$$\begin{aligned}&\mathcal{J}_K(\theta;\tau;u^{1:K}(0,\cdot),u^{1:K}(X,\cdot))\\&=\frac{1}{K}\sum_{k=1}^{K}\mathcal{J}_K(\theta;\tau;u^k(0,\cdot),u^k(X,\cdot))\end{aligned}$$

where the superscript i indicates the index of data. In this chapter, K equals to 1 and thus will be omitted.

2.3 Saint-Venant Distributed Transfer Matrix

This section outlines the frequency domain modeling of the linearized Saint-Venant equations used later in the dissertation.

2.3.1 Uniform Flow Case: Upstream Flow Variables as Boundary Conditions

For the case of uniform flow, the application of Laplace transform to the linear PDE system (2.11) leads to an *ordinary differential equation* (ODE) in variable x, with the Laplace variable s (see [93] for details). Solving this ODE with boundary conditions $u_1(0,s)$ and $u_2(0,s)$ leads to an open-loop distributed transfer matrix for the case of uniform flow, $\mathcal{G}^v(x,s) = \left(g^v_{ij}(x,s)\right)$, relating the flow variables at any point x of the river reach $u(x,s)$ to the upstream flow variables $u(0,s)$:

$$\begin{pmatrix} u_1(x,s) \\ u_2(x,s) \end{pmatrix} = \begin{pmatrix} g^v_{11}(x,s) & g^v_{12}(x,s) \\ g^v_{21}(x,s) & g^v_{22}(x,s) \end{pmatrix} \begin{pmatrix} u_1(0,s) \\ u_2(0,s) \end{pmatrix} \tag{2.21}$$

with

$$g^v_{11}(x,s) = \frac{\lambda_2 e^{\lambda_1 x} - \lambda_1 e^{\lambda_2 x}}{\lambda_2 - \lambda_1} \tag{2.22}$$

$$g^v_{12}(x,s) = -T_0 s\, \frac{e^{\lambda_2 x} - e^{\lambda_1 x}}{\lambda_2 - \lambda_1} \tag{2.23}$$

$$g^v_{21}(x,s) = \frac{\lambda_1 \lambda_2}{T_0 s}\, \frac{e^{\lambda_2 x} - e^{\lambda_1 x}}{\lambda_2 - \lambda_1} \tag{2.24}$$

$$g^v_{22}(x,s) = \frac{\lambda_2 e^{\lambda_2 x} - \lambda_1 e^{\lambda_1 x}}{\lambda_2 - \lambda_1} \tag{2.25}$$

Here, λ_1 and λ_2 are the eigenvalues of the ODE system, given by:

$$\lambda_i(s) = \frac{2T_0 V_0 s + \gamma_0}{2\alpha_0} + (-1)^i \frac{\sqrt{4C_0^2 T_0^2 s^2 + 4T_0(V_0\gamma_0 - \alpha_0\beta_0)s + \gamma_0^2}}{2\alpha_0} \tag{2.26}$$

2.3.2 Uniform Flow Case: Input Variables as Boundary Conditions

Equation (2.21) can be modified to the case when the boundary conditions are $u_1(0,s)$ and $u_2(X,s)$. The distributed transfer matrix $\mathcal{G}^u(x,X,s) = \left(g^u_{ij}(x,X,s)\right)$ relating the flow variables at any point x of the river reach $u(x,s)$ to the measured input variables $u_1(0,s)$ and $u_2(X,s)$ (see Section 2.2.3) of the river reach is given by the equation:

$$\begin{pmatrix} u_1(x,s) \\ u_2(x,s) \end{pmatrix} = \begin{pmatrix} g^u_{11}(x,X,s) & g^u_{12}(x,X,s) \\ g^u_{21}(x,X,s) & g^u_{22}(x,X,s) \end{pmatrix} \begin{pmatrix} u_1(0,s) \\ u_2(X,s) \end{pmatrix} \tag{2.27}$$

with

$$g^u_{11}(x,X,s) = \frac{\lambda_2 e^{\lambda_1 x+\lambda_2 X} - \lambda_1 e^{\lambda_2 x+\lambda_1 X}}{\lambda_2 e^{\lambda_2 X} - \lambda_1 e^{\lambda_1 X}} \tag{2.28}$$

$$g^u_{12}(x,X,s) = -T_0 s\, \frac{e^{\lambda_2 x} - e^{\lambda_1 x}}{\lambda_2 e^{\lambda_2 X} - \lambda_1 e^{\lambda_1 X}} \tag{2.29}$$

$$g^u_{21}(x,X,s) = \frac{\lambda_1\lambda_2}{T_0 s}\, \frac{e^{\lambda_2 x+\lambda_1 X} - e^{\lambda_1 x+\lambda_2 X}}{\lambda_2 e^{\lambda_2 X} - \lambda_1 e^{\lambda_1 X}} \tag{2.30}$$

$$g^u_{22}(x,X,s) = \frac{\lambda_2 e^{\lambda_2 x} - \lambda_1 e^{\lambda_1 x}}{\lambda_2 e^{\lambda_2 X} - \lambda_1 e^{\lambda_1 X}} \tag{2.31}$$

where, λ_1 and λ_2 are given by (2.26).

2.3.3 Non-Uniform Flow Case: Backwater Approximation

Following the method proposed in [136], and further modified by [94], the backwater curve defined by equation (2.5) is approximated by two straight lines as represented in Figure (2.1). The river reach is then decomposed into two parts: a uniform part and a backwater part. The abscissa of the limit between the parts is denoted x_1. Let $x_u = x$ denote the location in the uniform part, $x_b = x - x_1$ denote the location in the backwater part. $X_u = x_1$ denote the length of the uniform part and $X_b = X - x_1$ denote the length of the backwater part. Let $\mathcal{G}^u(x_u, X_u, s)$ and $\mathcal{G}^b(x_b, X_b, s)$ respectively denote the transfer matrices for the uniform and the backwater parts. These matrices have the same form as the transfer matrix in the uniform case (see equation (2.27)).

The transfer matrix for the non-uniform case is represented as $\mathcal{G}^n(x,X,s) = (g^n_{ij}(x,X,s))$, with

$$\begin{pmatrix} u_1(x,s) \\ u_2(x,s) \end{pmatrix} = \begin{pmatrix} g^n_{11}(x,X,s) & g^n_{12}(x,X,s) \\ g^n_{21}(x,X,s) & g^n_{22}(x,X,s) \end{pmatrix} \begin{pmatrix} u_1(0,s) \\ u_2(X,s) \end{pmatrix} \tag{2.32}$$

The entries of the transfer matrix for the non-uniform case are given by:

- $x < x_1$:

$$
\begin{aligned}
g^n_{11}(x, X, s) &= g^u_{11}(x, x_1, s) + \frac{g^u_{12}(x, x_1, s) g^u_{11}(x_1, x_1, s) g^b_{21}(0, X_b, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{12}(x, X, s) &= \frac{g^u_{12}(x, x_1, s) g^b_{22}(0, X_b, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{21}(x, X, s) &= g^u_{21}(x, x_1, s) + \frac{g^u_{22}(x, x_1, s) g^u_{11}(x_1, x_1, s) g^b_{21}(0, X_b, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{22}(x, X, s) &= \frac{g^u_{22}(x, x_1, s) g^b_{22}(0, X_b, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)}
\end{aligned}
$$

- $x > x_1$:

$$
\begin{aligned}
g^n_{11}(x, X, s) &= \frac{g^b_{11}(x_b, X_b, s) g^u_{11}(x_1, x_1, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{12}(x, X, s) &= g^b_{12}(x_b, X_b, s) + \frac{g^b_{11}(x_b, X_b, s) g^u_{12}(x_1, x_1, s) g^b_{22}(0, X_b, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{21}(x, X, s) &= \frac{g^b_{21}(x_b, X_b, s) g^u_{11}(x_1, x_1, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{22}(x, X, s) &= g^b_{22}(x_b, X_b, s) + \frac{g^b_{21}(x_b, X_b, s) g^u_{12}(x_1, x_1, s) g^b_{22}(0, X_b, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)}
\end{aligned}
$$

- $x = x_1$:

$$
\begin{aligned}
g^n_{11}(x, X, s) &= \frac{g^u_{11}(x_1, x_1, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{12}(x, X, s) &= \frac{g^u_{12}(x_1, x_1, s) g^b_{22}(0, X_b, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{21}(x, X, s) &= \frac{g^b_{21}(0, X_b, s) g^u_{11}(x_1, x_1, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)} \\
g^n_{22}(x, X, s) &= \frac{g^b_{22}(0, X_b, s)}{1 - g^u_{12}(x_1, x_1, s) g^b_{21}(0, X_b, s)}
\end{aligned}
$$

In the following sections, the notation $\mathcal{G}^n(\theta; x, X, s)$ will be used to emphasize the non-uniform transfer matrix for parameter vector θ.

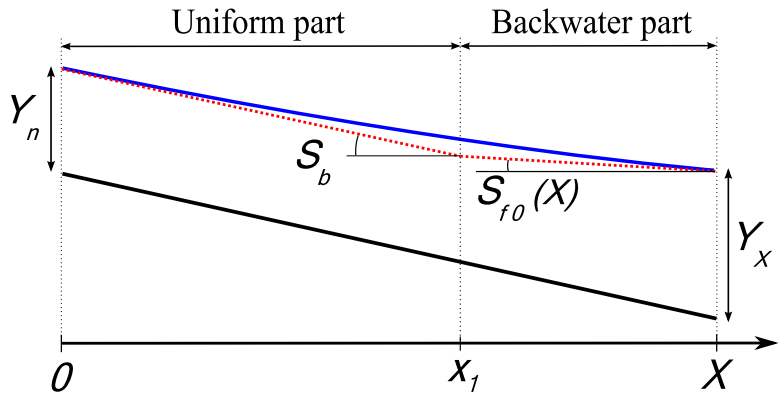


Figure 2.1: *Backwater curve approximation.*

2.4 Proposed Approach

Recall the cost function (2.18):

$$\begin{aligned}
&\mathcal{J}(\theta;\tau;u(0,\cdot),u(X,\cdot)) \\
&= \int_0^\tau \frac{w_1}{u_{1,\text{norm}}^2}\left(u_1(X,t)-\hat{u}_1(X,t|\theta)\right)^2 dt \\
&+ \int_0^\tau \frac{w_2}{u_{2,\text{norm}}^2}\left(u_2(0,t)-\hat{u}_2(0,t|\theta)\right)^2 dt
\end{aligned}$$

and that parameter estimation Problem (1) is to minimize the above cost function with respect to the admissible range of parameters $\mathcal{D}_\mathcal{A}$ subject to the PDE constraint that $\hat{u}_1(X,t|\theta)=\phi(X,t|\theta)$ and $\hat{u}_2(0,t|\theta)=\phi(0,t|\theta)$, where $\phi(x,t|\theta)$ is the obtained by solving (2.17) with boundary conditions (2.13) and initial conditions (2.14). This section uses the input-output transfer matrix $\mathcal{G}^n(\theta;x,X,s)$ in Section 2.3.3 to obtain a more direct formulation of the parameter estimation problem in which the cost function explicitly incorporates the constraint.

The fundamental idea behind the proposed approach is to decompose the input variables $u_1(0,t)$ and $u_2(X,t)$ into a finite sum of dominant oscillatory modes. In the case of a river reach influenced by the ocean at the downstream end, these modes can be thought as the principle modes of excitation produced by the tidal forcing. Under the assumption of N most dominant oscillatory modes, the input variables can be expressed as:

$$u_1(0,t) \approxeq \sum_{k=0}^{N}\left[d_k^{(1,0)}e^{j\omega_k t}+\overline{d_k^{(1,0)}}e^{-j\omega_k t}\right] \tag{2.33}$$

$$u_2(X,t) \approxeq \sum_{k=0}^{N}\left[d_k^{(2,X)}e^{j\omega_k t}+\overline{d_k^{(2,X)}}e^{-j\omega_k t}\right] \tag{2.34}$$

with $d_k^{(1,0)}$ and $d_k^{(2,X)}$, $k = 0, \ldots, N$, are respectively the Fourier coefficients of the spectral decomposition of $u_1(0,t)$ and $u_2(X,t)$. ω_k is the set of frequencies used for the modal decomposition.

Now, the non-uniform transfer matrix $\mathcal{G}^n(\theta; x, X, s)$ (see (3.4)) can be used to compute the output *predictions* $\hat{u}_1(X,t|\theta)$ and $\hat{u}_2(0,t|\theta)$ that go in the definition of the cost function:

$$\hat{u}_1(X,t|\theta) = \sum_{k=0}^{N} \left[\alpha_k^{(1,X)}(\theta) e^{j\omega_k t} + \overline{\alpha_k^{(1,X)}(\theta)} e^{-j\omega_k t} \right] \tag{2.35}$$

$$\hat{u}_2(0,t|\theta) = \sum_{k=0}^{N} \left[\alpha_k^{(2,0)}(\theta) e^{j\omega_k t} + \overline{\alpha_k^{(2,0)}(\theta)} e^{-j\omega_k t} \right] \tag{2.36}$$

with coefficients in the equations given by

$$\begin{aligned}
\alpha_k^{(1,X)}(\theta) &= d_k^{(1,0)} g_{11}^n(\theta; x, X, j\omega_k) + d_k^{(2,X)} g_{12}^n(\theta; x, X, j\omega_k) \\
\overline{\alpha_k^{(1,X)}(\theta)} &= \overline{d_k^{(1,0)}} g_{11}^n(\theta, x_1, X, j\omega_k) + \overline{d_k^{(2,X)}} g_{12}^n(\theta; x, X, j\omega_k) \\
\alpha_k^{(2,0)}(\theta) &= d_k^{(1,0)} g_{21}^n(\theta; x, X, j\omega_k) + d_k^{(2,X)} g_{22}^n(\theta; x, X, j\omega_k) \\
\overline{\alpha_k^{(2,0)}(\theta)} &= \overline{d_k^{(1,0)}} g_{21}^n(\theta; x, X, j\omega_k) + \overline{d_k^{(2,X)}} g_{22}^n(\theta; x, X, j\omega_k)
\end{aligned}$$

Note that the number of modes N in equations (2.33, 2.34) and (2.35, 2.36) is the same because the PDE is non-dispersive. It should be noted that $g_{ij}^n(\theta; x, X, -j\omega_k) = \overline{g_{ij}^n}(\theta; x_1, X, j\omega_k)$.

Similarly, the measured output variables $u_1(X,t)$ and $u_2(0,t)$ can be expressed as an infinite sum of response modes

$$u_1(X,t) = \sum_{k=0}^{\infty} \left[\alpha_k^{(1,X)} e^{j\omega_k t} + \overline{\alpha_k^{(1,X)}} e^{-j\omega_k t} \right] \tag{2.37}$$

$$u_2(0,t) = \sum_{k=0}^{\infty} \left[\alpha_k^{(2,0)} e^{j\omega_k t} + \overline{\alpha_k^{(2,0)}} e^{-j\omega_k t} \right] \tag{2.38}$$

Substituting the expressions in equations (2.35), (2.36), (2.37), (2.38) in the expression for cost

function (2.18), we obtain:

$$
\begin{aligned}
&\mathcal{J}(\theta;\tau;u(0,\cdot),u(X,\cdot)) \\
&\approxeq \hat{\mathcal{J}}\left(\theta;N;\left\{d_k^{(1,0)},d_k^{(2,X)}\right\}_{k=0}^{N},\left\{\alpha_k^{(1,X)},d_k^{(2,0)}\right\}_{k=0}^{\infty}\right) \\
&= 2\tau\sum_{k=0}^{N}\left(\frac{w_1}{u_{1,\text{norm}}^2}\left|\alpha_k^{(1,X)}(\theta)-\alpha_k^{(1,X)}\right|^2\right) \\
&+ 2\tau\sum_{k=0}^{N}\left(\frac{w_1}{u_{2,\text{norm}}^2}\left|\alpha_k^{(2,0)}(\theta)-\alpha_k^{(2,0)}\right|^2\right) \\
&+ 2\tau\sum_{k=N+1}^{\infty}\left(\frac{w_1}{u_{1,\text{norm}}^2}\left|\alpha_k^{(1,X)}\right|^2+\frac{w_2}{u_{2,\text{norm}}^2}\left|\alpha_k^{(2,0)}\right|^2\right)
\end{aligned}
\tag{2.39}
$$

where, the integration has been carried explicitly over $[0,\tau]$, with τ multiple of the smallest period of the N most dominant modes. Renormalizing for optimization purposes, the modified cost function $\bar{\mathcal{J}}$ is defined as:

$$
\begin{aligned}
&2\tau\sum_{k=0}^{N}\left[w_1\left|\frac{\alpha_k^{(1,X)}(\theta)-\alpha_k^{(1,X)}}{\alpha_{k,n}^{(1,X)}}\right|^2\right] \\
&+ 2\tau\sum_{k=0}^{N}\left[w_2\left|\frac{\alpha_k^{(2,0)}(\theta)-\alpha_k^{(2,0)}}{\alpha_{k,n}^{(2,0)}}\right|^2\right] \\
&+ 2\tau\sum_{k=N+1}^{\infty}\left[w_1\left|\frac{\alpha_k^{1,X}}{\alpha_{k,n}^{(1,X)}}\right|^2+w_2\left|\frac{\alpha_k^{2,0}}{\alpha_{k,n}^{(2,0)}}\right|^2\right]
\end{aligned}
\tag{2.40}
$$

with $\alpha_{k,n}^{(1,X)}$, $\alpha_{k,n}^{(2,0)}$ the normalizing coefficients of the Fourier coefficients. Note that choosing the constant normalizing coefficients gives more weight to the dominant frequencies, while choosing the coefficients of the response output modes as the normalizing coefficients gives same weight to all the selected dominant frequencies.

It is noted that the third term in the expression of cost function in (2.40) does not depend on parameter vector θ, and hence not affect the minimization of the cost function. Therefore, the parameter estimation problem can be expressed as:

$$
\hat{\theta}=\arg\min_{\theta\in\mathcal{D}_{\mathcal{A}}}\tilde{\mathcal{J}}\left(\theta;N;\left\{d_k^{(1,0)},d_k^{(2,X)}\right\}_{k=0}^{N},\left\{\alpha_k^{(1,X)},\alpha_k^{(2,0)}\right\}_{k=0}^{N}\right) \tag{2.41}
$$

with

$$\tilde{\mathcal{J}}\left(\theta; N; \left\{d_k^{(1,0)}, d_k^{(2,X)}\right\}_{k=0}^{N}, \left\{\alpha_k^{(1,X)}, \alpha_k^{(2,0)}\right\}_{k=0}^{N}\right)$$
$$= 2\tau \sum_{k=0}^{N} \left[w_1 \left| \frac{\alpha_k^{(1,X)}(\theta) - \alpha_k^{(1,X)}}{\alpha_{k,n}^{(1,X)}} \right|^2 \right.$$
$$\left. + \, w_2 \left| \frac{\alpha_k^{(2,0)}(\theta) - \alpha_k^{(2,0)}}{\alpha_{k,n}^{(2,0)}} \right|^2 \right] \tag{2.42}$$

It now becomes an unconstrained optimization problem and can be solved using standard non-linear programming.

2.5 Application to the Sacramento Delta

In this section, we apply the approach described above to identify parameters of a river reach in the Sacramento-San Joaquin Delta.

2.5.1 Parameter Identification of a River Reach

The Sacramento-San Joaquin Delta in California, USA, is a complex network of over $1150\ km$ of tidally-influenced channels and sloughs. The San Joaquin River, with a length of $530\ km$, is the second-longest river in the Delta area. The field of interest for our experiment is the San Joaquin River reach between DWR and USGS stations (*SJL* and *SJG*), as shown in Figure 2.2. The available data is the discharge perturbation, the stage perturbation for the stations *SJL*, *SJG* and the stage perturbation at *BDT*. The direction of the mean flow is from *SJL* to *SJG*. Measurements are available at every 900 seconds. The data was collected between $11/16/2006$ and $12/17/2006$.

This study makes the following simplifying assumptions:

- the flow is one dimensional;
- the wavelength of tidal forcing is long relative to the water stage;
- lateral and vertical accelerations are negligible;
- pressure distribution is hydrostatic;
- the effects of sediment deposition and scour are negligible; hence, the channel geometry is fixed;

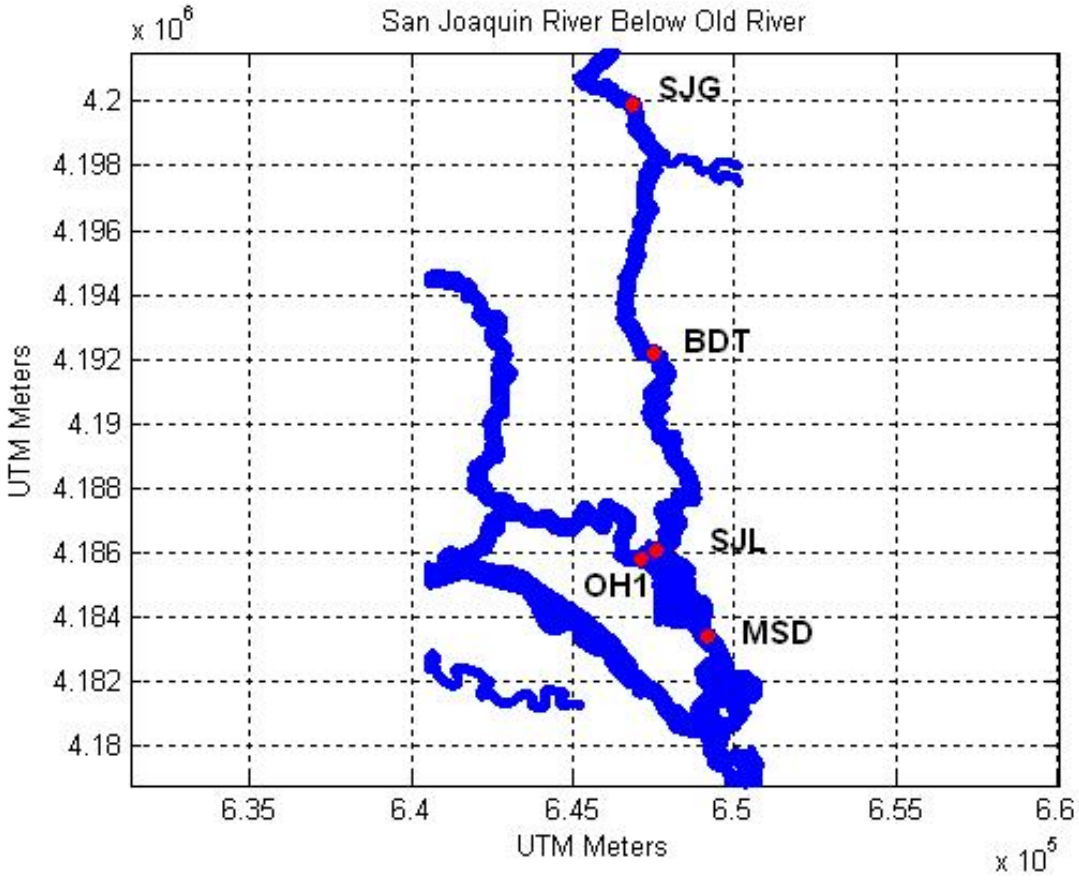


Figure 2.2: *DWR satellite stations along San Joaquin River.*

- the channel geometry can be averaged by a rectangular cross-section;
- the bed slope is small; water surface across any cross-section is horizontal;
- friction can be modeled by Manning's formula, and the non-uniformity has a small effect on the friction loss.

2.5.2 Spectral Analysis

Figure 2.3 shows the spectral analysis for the total stage at station SJL. The power spectrum is cutoff at $0.02dB/Hz$ to determine the 27 dominant frequencies. The figure also compares the measured stage and the stage generated from the first 27 modes. The results indicate that 27 modes are enough to capture the signal; the amplitude at 0 Hz is actually the nominal stage. There are three dominant tidal frequencies in the system: $\omega_1 = 0.0001407\ s^{-1}$ (or period $12.4\ hrs$), corresponding to the M2 tide generated by the Moon; $\omega_2 = 0.0000727\ s^{-1}$ (or period $24\ hrs$) corresponding to the K1 tide generated by the Sun and a $\omega_3 = 0.000068\ s^{-1}$ (or period $25\ hrs$) tide. Similar arguments hold for other measured variables.

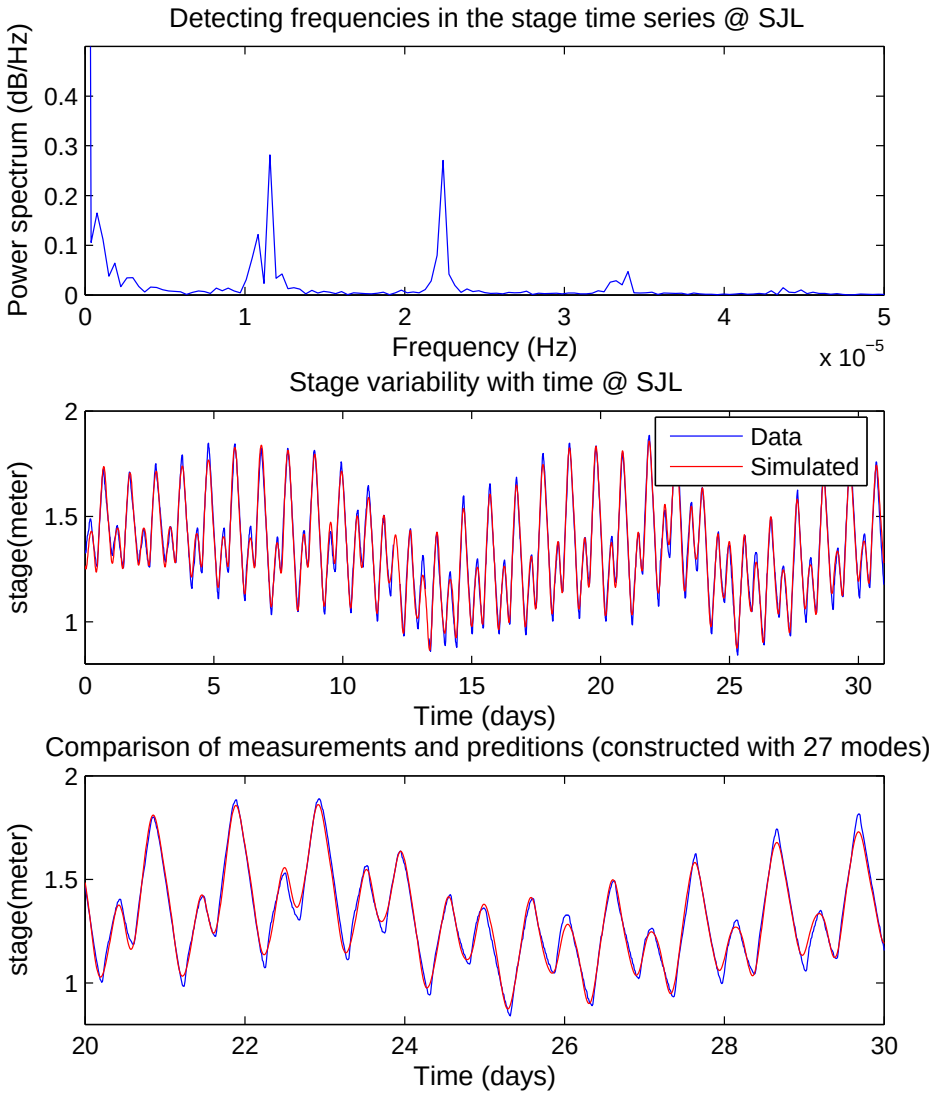


Figure 2.3: *Stage Analysis at SJL Station.*

2.5.3 Parameter Identification

This study chooses $\alpha_{k,n}^{(1,X)} = \alpha_k^{(1,X)}$, $\alpha_{k,n}^{(2,0)} = \alpha_k^{(2,0)}$ so that each significant frequency has the same weight. The weighing factors are w_1, w_2 are chosen to be 1. The constrained optimization function in MATLAB is used to identify the parameters listed in Table 2.1. This optimization process converges very quickly. The optimal cost function $\tilde{\mathcal{J}}$ eventually converges to 0.093, leading to the desired estimates. The values of T_0, S_b, and n are acceptable.

Q_0	Y_X	T_0	S_b	n
11.5 $\mathrm{m^3 s^{-1}}$	4.04 m	47.00 m	0.0002	0.048 $\mathrm{m^{-1/3} s}$

Table 2.1: Parameters of the San Joaquin River.

The estimated values of Q_0 and Y_X are not as good as the others, as the expected values are respectively $7.34\ m^3/s$ and $10.0\ m$. This can be partly explained by recalling that the transfer matrix for the non-uniform flow $\mathcal{G}^n(\theta; x, X, s)$ was derived using the two-pool approximation (see Section 2.3.3). A further improvement in the non-uniform transfer matrix might improve accuracy of the predicted steady state boundary conditions.

The comparison of the predicted and measured outputs (i.e., the stage perturbation at SJL and the discharge perturbation at SJG) is shown in Figure 2.4. It is observed that the estimated model accurately predicts the measured output values.

It should be noted, however, that the model underestimates the stage measurements and overestimates the discharge measurements. This is likely due to the rectangular cross-section assumption which leads to over-estimation of the free surface width $\hat{T}_0$, and consequently, under-estimation of the stage in order to satisfy discharge data.

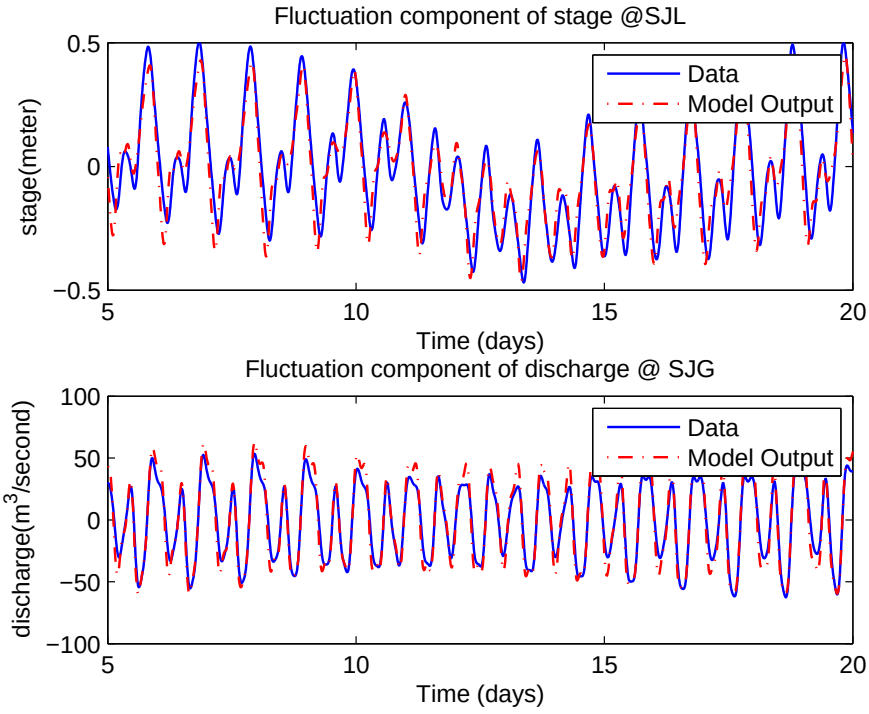


Figure 2.4: *Parameter Identification: model output v.s. measurement.*

2.5.4 Sensitivity

The graphs in Figure 2.5 present the variations of the cost function $\tilde{\mathcal{J}}$ with respect to each parameter. For each graph, one parameter varies, others remain constant at the values corresponding to

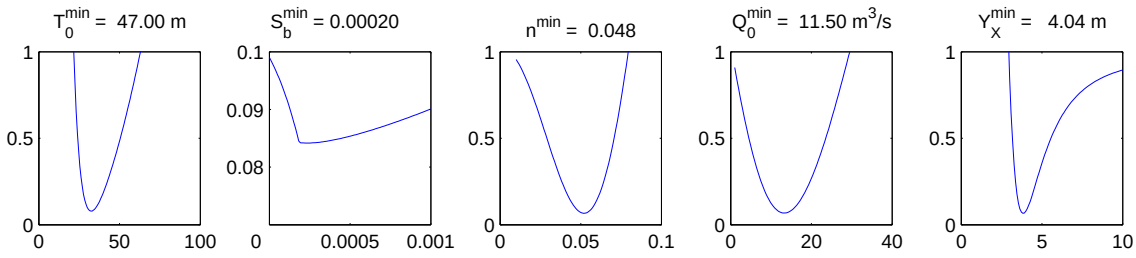


Figure 2.5: *Sensitivity analysis for the application case.*

the minimum of $\tilde{\mathcal{J}}$. A flat curve signifies that any other value of the parameter would involve the same value of $\tilde{\mathcal{J}}$. In the present case, the minimum on each graph is clearly defined, so we can confirm that the objective function and the model are sensitive enough to satisfactorily estimate the parameters.

2.5.5 Validation

The stage time series at BDT station, along with the given parameters in Table 2.1, are used to validate the model. As shown in Figure 2.6, the model output gives an accurate prediction of the validation data. For the validation case too, the predicted stage under-estimates the measured stage.

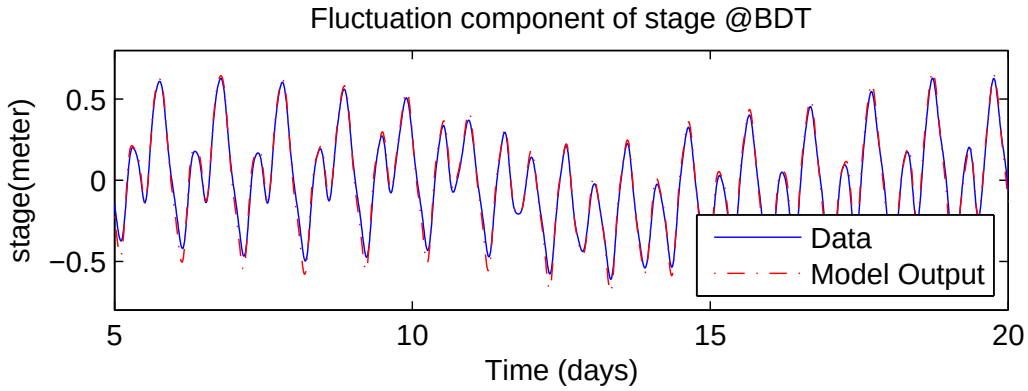


Figure 2.6: *Model Validation: model output v.s. measurement.*

2.5.6 Predictability

The data collected between 12/18/2006 and 2/18/2007 were used to further test the predictability of the model, with the identified parameters given in Table 2.1. Following the steps described in

the previous sections, the stage outputs at BDT station have been predicted, using the discharge data at SJL station and the stage data at SJG station during the time interval. Comparing with the observed stage data, it is clear that the model successfully characterizes the flow, with a good reflection of the stage flunctuation tendency in the time domain (Figure 2.7).

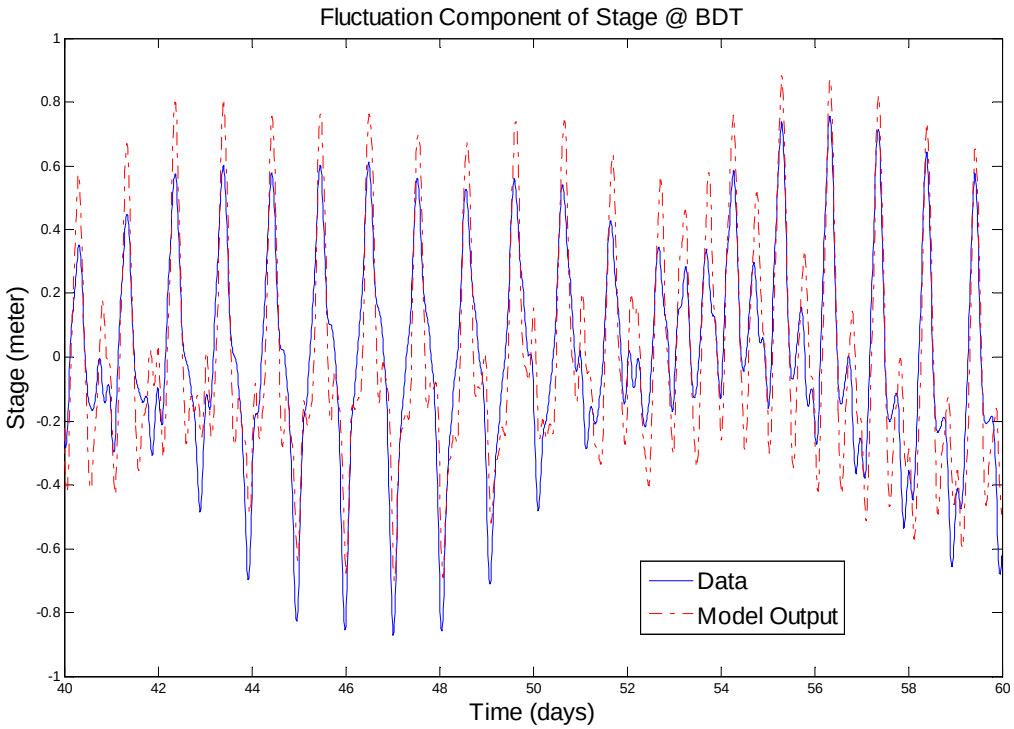


Figure 2.7: *Model Predictability during 12/18/2006 and 2/18/2007: model output v.s. measurement.*

2.6 Summary

This chapter investigates a parameter identification problem for hydraulic systems governed by first-order, linear hyperbolic shallow water equations subjected to periodic tidal forcing. The problem is posed as a PDE constrained optimization problem with data of the problem given by the measured input and output variables at the boundary of the domain.

Using the frequency domain modeling techniques, such as the modal decomposition and approximation of transfer matrix for the non-uniform steady state case, the output response can be expressed in terms of the spectral coefficients of the input excitation and the transfer matrix coef-

ficients evaluated at appropriate points. By considering a finite number of dominant oscillatory modes of the input, an accurate representation of the output is obtained. This converts the original PDE constrained optimization problem to one without any constraints. The optimal parameters can be identified using standard nonlinear programming.

The utility of the proposed approach is illustrated by considering a river reach in the Delta, subjected to tidal forcing. The parameter identification problem is to estimate the average free-surface width, the bed slope, the friction coefficient and the steady-state boundary conditions. It is shown that the estimated model gives an accurate prediction of the flow variables at an intermediate location within the reach.

In this chapter, all the measurements are considered error free. In the next chapter, we will consider the case that the number of measurements are larger than needed, and data reconlilation technique will be impleamented to account for measurement error.

Chapter 3

Channel Network and Data Reconciliation

3.1 Introduction

In hydraulic systems, numerous factors could lead to measurement errors, for example broken gauges, process leaks, sensor drifts, improper use of measuring devices, and other random sources [2]. Data reconciliation, an effective method to tune-up the measurement data [129], [15], [37], [3], has been applied in many engineering fields [44], [12], [127], [162].

The flow variables in a hydraulic system are related to each other by a mathematical model. Sensing on this hydraulic system is done using Eulerian USGS sensors. These Measurements are subject to errors, both random and systematic, so that the laws of conservation of mass and energy are not obeyed. In order to record the performance of the process, these measurements are adjusted in order that they conform to the conservation laws and any other constraints imposed upon them. This procedure is known as data reconciliation. The objective of data reconciliation is to use information redundancy to eliminate errors in real-time measurements [37], [17], [31], [30].

This chapter presents theoretical results applicable to data reconciliation for tidally forced networks of open channels. Using the modal decomposition techniques, we are able to transform dynamic constraints into static constraints in the frequency domain, and subsequently obtain a static data reconciliation problem, which is easier to resolve and can lead to accurate results. Generally, this static data reconciliation problem is to minimize the measurement errors while satisfying the static constraints of the proposed model.

A spatially-dependent transfer matrix is constructed and applied to a channel network considered in this chapter, relating a selected set of model inputs to the output variables. The transfer matrix is a function of channel width, channel length, bed slope, mean discharge, mean stage and Manning coefficient. This set of parameters needs to be chosen carefully to characterize the geometry of the

channels, as the uncertainty of the parameters would contribute to the errors in the model output.

With this linear model in the frequency domain, the static data reconciliation problem is shown to reduce to a quadratic problem. The objective function used in the present study is a weighted L^2-norm of the difference between the measured and reconciliated data. The linear network model constructed serves as the constraints in the optimization problem. A closed-form optimal solution is obtained, resulting in a set of reconciliated boundary data consistent with both the linear network model and the statistical assumptions on measurement errors. Subsequently, we apply the reconstructed boundary conditions to the linear network model to obtain an accurate forward simulation of the flow within of the network domain.

3.2 Proposed Method

3.2.1 General Considerations

The general class of hydraulic system studied in the present chapter is a distributed network of channels subject to quasi-periodic tidal forcing. Sensing on this hydraulic system is done using fixed Eulerian USGS sensors, subject to measurement errors. The motivation of the work is to derive the "most likely" flow conditions from the measured data available in a given period of time (at least a few weeks to catch the tidal period), and forecast the future flow variables. The goal of the method is thus not real time estimation, as traditionally done in data assimilation [148], but forecast based on measured forcing.

The flow variables are related to each other by a mathematical model. Therefore, if the measurement data was error free, it would satisfy the model. Because the number of points at which the variables are measured is usually larger than needed to fully prescribe the model, there is "information redundancy" in the system. Once information redundancy exists, data reconciliation can be implemented to account for measurement error.

The ultimate goal of data reconciliation is to use such information redundancy in a system to have the data self-corrected using the model. An effective data reconciliation method allows the detection of any inconsistent or biased measurements, and furthermore provides corrected values (namely estimated measurements).

It should be noted that any information redundancy is model-specific. We therefore need to first construct a "good" hydraulic model to characterize the flow system, as described in the following section.

3.2.2 Linear Channel Network Model

Transfer Matrix Representation of Linearized Saint-Venant Model

To facilitate the mathematical analysis, we re-write the linearized Saint-Venant equations as follows:

$$\frac{\partial}{\partial t}\begin{pmatrix} q(x,t) \\ y(x,t) \end{pmatrix} = \left(\mathcal{A}(x)\frac{\partial}{\partial x} + \mathcal{B}(x)\right)\begin{pmatrix} q(x,t) \\ y(x,t) \end{pmatrix} \qquad (x,t) \in [0,X] \times [0,+\infty) \tag{3.1}$$

where,

$$\mathcal{A}(x) = \begin{pmatrix} -2V_0 & -\alpha_0 \\ -\frac{1}{T_0} & 0 \end{pmatrix} \qquad \mathcal{B}(x) = \begin{pmatrix} \beta_0 & \gamma_0 \\ 0 & 0 \end{pmatrix} \tag{3.2}$$

The application of Laplace transform to the linear PDE system (3.1) leads to the following *ordinary differential equation* (ODE) in the variable x, with a complex parameter s.

$$\frac{d}{dx}\begin{pmatrix} q(x,s) \\ y(x,s) \end{pmatrix} = \mathcal{A}^{-1}(x)\left[\mathcal{B}(x) - s\mathbb{I}_2\right]\begin{pmatrix} q(x,s) \\ y(x,s) \end{pmatrix} \tag{3.3}$$

Following the method initiated in [136], and further modified in [94], a transfer matrix $\mathcal{G}(x,X,s) = (g_{ij}(x,X,s))$ for the non-uniform channel relates the boundary conditions and intermediate flow variables, and is defined as:

$$\begin{pmatrix} q(x,s) \\ y(x,s) \end{pmatrix} = \mathcal{G}(x,X,s)\begin{pmatrix} q(0,s) \\ y(X,s) \end{pmatrix} \tag{3.4}$$

$\mathcal{G}(x,X,s)$ is a function of channel length X, average discharge Q_0, average downstream depth Y_X, average width T_0, bed slope S_b, and Manning coefficient n of the channel. The upstream and downstream boundary conditions are the upstream discharge perturbation $q(0,s)$ and the downstream stage perturbation $y(X,s)$, respectively. Because of the distributed nature of the system, this transfer function also depends on the coordinate x in the channel, since it relates inputs $q(0,s)$ and $y(X,s)$ to the state of the system $q(x,s)$ and $y(x,s)$ at any x in the channel.

Transfer Matrix Model for Channel Networks

The model (3.4) can be readily applied to tidally driven channel networks. The problem of interest can be stated as follows, and is illustrated in Figure 3.1. Given a set of "external" boundary

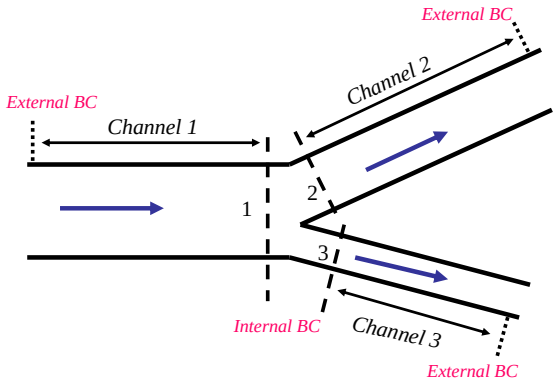


Figure 3.1: *Representation the system of interest for estimating internal boundary conditions (Internal BC) using data reconciliation on external boundary conditions (External BC): data is given at the three external conditions of channel* 1, 2 *and* 3*; the state of the system is computed at three internal locations of the system labeled* 1, 2 *and* 3*.*

conditions of a network, at which we have measurements, reconstruct flow conditions at "internal" locations (also referred as boundary conditions). This type of problem appears in our data assimilation work, in which we need estimates of boundary conditions at locations where fixed sensors are not available. The fundamental approach to build a network model is as follows:

- **Step 1**: Decompose the channel network into individual channel reaches, and apply the linear model (3.4) to each branch. For each of the river reach indexed by i, the flow variables $q_i(x, s)$, $y_i(x, s)$ denote the perturbed discharge and stage in the frequency domain respectively, X_i denote the length of the channel. The junction of the river reach is defined as the node of the channel network.

- **Step 2**: Impose the internal boundary conditions at every junction to ensure flow compatibility. Considering a simple river junction illustrated in Figure 3.1: for each channel i, we will have the upstream boundary condition $y_i(0, s)$, $q_i(0, s)$, and downstream boundary condition $y_i(X_i, s)$, $q_i(X_i, s)$. If these boundary conditions are at the inside nodes of the channel network, they are called internal boundary conditions, otherwise they are labeled as external boundary conditions. The linear relationships of hydraulic internal boundary conditions at a junction are specified by equations of mass and energy conservation. Assuming no change in storage volume within the junction, the continuity equation can be expressed by:

$$q_1(X_1, s) = q_2(0, s) + q_3(0, s)$$

  When the flows in all the branches meeting at a junction are subcritical, the equation for

energy conservation can be approximated by a kinematic compatibility condition as:

$$y_1(X_1, s) = y_2(0, s) = y_3(0, s)$$

where X_i is the downstream point of each channel i, and 0 is the upstream point of each channel i.

- **Step 3**: Assemble the equations for each individual channel and interior junctions together to model the entire network. The flow variables at the boundary of each channel are represented by a linear relationship:

$$M(s)Z(s) = 0 \tag{3.5}$$

where $Z(s)$ is the concatenated vector of all $[q_i(0, s), q_i(X_i, s), y_i(0, s), y_i(X_i, s)]^T$, where $i = 1, \cdots, N$, $Z(s)$ is thus the vector comprising the discharge and stage variables at the upstream and downstream ends of all channels; $M(s)$ is a matrix of appropriate dimension encoding the previous constraints.

- **Step 4**: Evaluate the unmeasured flow variables inside the channel network.

a) Specify the interior boundaries. In a channel network system, we define a subjective subset of boundary conditions ($Z_{givenBC} \subset Z$), which leads to a unique solution of model (3.5). This subset should satisfy: $\textit{dim}(Z_{givenBC}) = \textit{dim}(Z) - \textit{Rank}(M)$. All the other unknown boundary variables (interior and external), denoted as $Z_{otherBC} = Z \setminus Z_{givenBC}$, are therefore estimated with model (3.5). Model (3.5) now has the form:

$$Z_{otherBC} = R(s)Z_{givenBC} \tag{3.6}$$

where $R(s)$ is a matrix of appropriate size. Given $Z = \begin{pmatrix} Z_{givenBC} \\ Z_{otherBC} \end{pmatrix}$, $M(s) = \Big[R(s)| - \mathbb{I}\Big]$.

b) Estimation of the perturbed discharge and stage along the channel. It is achieved by a simple application of transfer function analysis:

$$\begin{pmatrix} q_i(x, s) \\ y_i(x, s) \end{pmatrix} = \begin{pmatrix} g_{i,11}(x, X_i, s) & g_{i,12}(x, X_i, s) \\ g_{i,21}(x, X_i, s) & g_{i,22}(x, X_i, s) \end{pmatrix} \begin{pmatrix} q_i(0, s) \\ y_i(X_i, s) \end{pmatrix} \tag{3.7}$$
$$i = 1, \cdots, N$$

where, $\mathcal{G}_i(x, X_i, s) = (g_{i,jk}(x, X_i, s))$ is the distributed transfer matrix based on the information of channel i.

3.2.3 Data Reconciliation

In practice, the measured data called Y_m is normally a superset of the data required to uniquely define the system. i.e., $Z_{givenBC} \subset Y_m \subseteq Z$. When this is the case, we can use the information redundancy and apply data reconciliation to detect and handle the measurement errors.

Using modal decomposition, we are able to convert the dynamic model (3.1) to a "static" model, in which the measurable variables are linked by a algebraic relationship in the frequency domain:

$$P(s)Y(s) = 0 \tag{3.8}$$

where $Y(s) = [Y_1, Y_2, Y_3, \cdots] \subseteq Z(s)$ is a vector of noise free measurements, and where $P(s)$ is a sub-matrix of $M(s)$ with the appropriate dimension.

It is assumed that the measurements are independent and subject to an additive noise. The measured data Y_m is composed of the "ideal" measurements vector Y and a noise vector ϵ :

$$Y_m = Y + \epsilon \tag{3.9}$$

This noise vector ϵ is assumed to follow a Gaussian distribution with zero mean and weight matrix $W = \text{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2)$. Here, σ_i represents the noise standard deviation for each measurement. The matrix W is chosen (positive definite), and its components are picked to have the error to be minimized in a weighted L^2 norm.

The objective of data reconciliation is to obtain estimated values $\hat{Y}$ close to the measurements Y_m while satisfying the "static" linear model (3.8). This can be formulated as an optimization problem with linear constraints. The cost function to minimize is the weighted quadratic error between the measurements Y_m and the reconciliated data $\hat{Y}$. The constraints are given by the model (3.8). The reconciliation problem in the spectral domain now becomes a least square problem with linear constraints:

$$\begin{aligned} \textbf{min.} \quad & f = (\hat{Y} - Y_m)^T W^{-1} (\hat{Y} - Y_m) \\ \textbf{s.t.} \quad & P\,\hat{Y} = 0 \end{aligned} \tag{3.10}$$

We hereby use the method suggested by [70] to solve the above data reconciliation problem. The constrained optimization problem is transformed into a corresponding unconstrained problem [20],

using the Lagrange multiplier vector ν. The Lagrangian of the problem reads:

$$L(\hat{Y}, \nu) = (\hat{Y} - Y_m)^T W^{-1} (\hat{Y} - Y_m) + 2\nu^T P\hat{Y} \tag{3.11}$$

In order to obtain the unknown variables, take partial derivatives and set them to zero:

$$\begin{aligned} \frac{\partial L}{\partial \hat{Y}} &= 2(\hat{Y} - Y_m)^T W^{-1} + 2\nu^T P = 0 \\ \frac{\partial L}{\partial \nu} &= 2\ P\hat{Y} = 0 \end{aligned} \tag{3.12}$$

Rewrite the above equations as:

$$\begin{pmatrix} W^{-1} & P^T \\ P & \mathbf{0}_{\dim(\nu)\times\dim(\nu)} \end{pmatrix} \begin{pmatrix} \hat{Y} \\ \nu \end{pmatrix} = \begin{pmatrix} W^{-1} Y_m \\ \mathbf{0}_{\dim(\nu)\times 1} \end{pmatrix} \tag{3.13}$$

Thus,

$$\hat{Y} = \begin{pmatrix} \mathbf{I}_{\dim(Y_m)} & \mathbf{0}_{\dim(Y_m)\times\dim(\nu)} \end{pmatrix} \begin{pmatrix} W^{-1} & P^T \\ P & \mathbf{0}_{\dim(\nu)\times\dim(\nu)} \end{pmatrix}^{-1} \begin{pmatrix} W^{-1} Y_m \\ \mathbf{0}_{\dim(\nu)\times 1} \end{pmatrix} \tag{3.14}$$

where, matrices **I** and **0** are Identity Matrix and Zero Matrix of appropriate size. Note that this program can be solved numerically easily, with standard optimization software such as CPLEX, MATLAB, CVX, etc [20].

The reconciliated measurements $\hat{Y}$ can then be used to obtain the desired internal boundary conditions using equations 3.5, 3.6, 3.7.

Remark 4. *The proposed approach thus consists in assuming high confidence in the model, and finding the "best" estimation, i.e., the estimation which minimizes measurement error. This is a standard procedure in data reconciliation.*

3.3 Application to the Sacramento River

3.3.1 Description of the System and Assumptions

The area of interest for our experiment is located around the junction of the Sacramento River and the Georgiana Slough, as shown in Figure 3.2. This network is monitored by a static sensor infrastructure subject to usual problems of inaccuracy and measurement errors for interested sensing

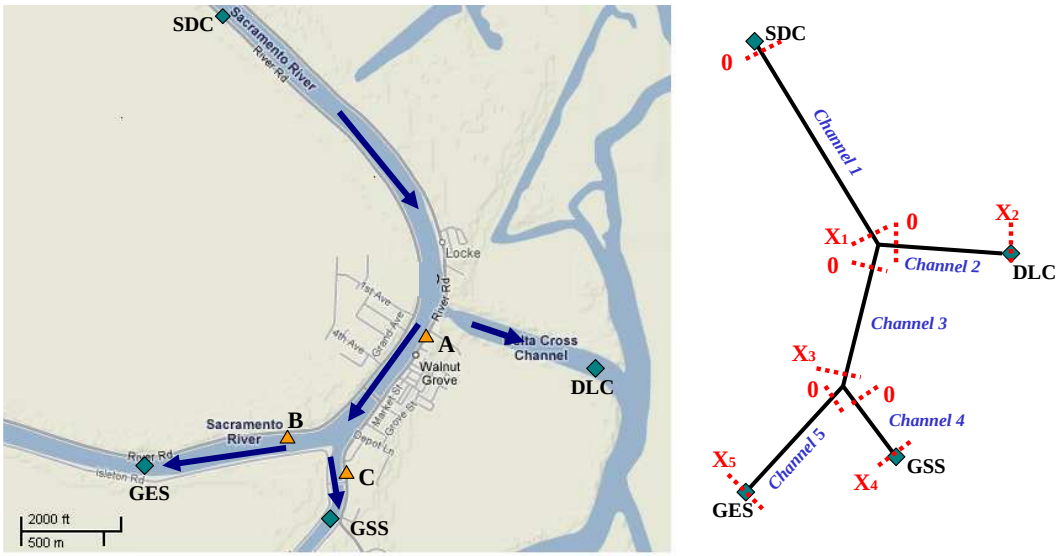


Figure 3.2: *Experimental area in the Sacramento River and the Georgiana Slough.*

systems. Most of the time, the direction of mean river flow is from north to south, as indicated with arrows. During the tidal inversion, the water flows in the opposite way.

Four USGS stations, namely *SDC*, *DLC*, *GES*, and *GSS*, are located at the external boundaries of this deployment field. The stations are marked as squares in Figure 3.2. Both discharge and stage are collected every 900 seconds at these stations. Note that in the USGS measurement system, only the stage are measured directly, and the discharge data is estimated by a rating curve, which is a relation between stream stage and streamflow. The relation of stream stage to streamflow is always changing, and need to be calibrated frequently. It would introduce errors if the rating curve has not been validated in time. More detailed information of the USGS stations can be found at http://ga.water.usgs.gov/edu/measureflow.html.

The field data was collected between 10/23/2007 and 11/13/2007. The raw field data is noisy, and the measurement errors are assumed to follow a normal Gaussian distribution. In addition, the following simplifications for the flow model have been made in this study:

- The flow can be represented by a one dimensional model.
- The channel geometry is fixed, as the effects of sediment deposition and scour are negligible during the experiment period.
- The channel geometry can be modeled by a rectangular cross-section.
- The lateral and vertical accelerations are negligible.

- The pressure distribution is hydrostatic.
- There is no significant jump along the bathymetry of the channel, and the bed slope is smooth and small.
- The water surface across any cross-section is horizontal.

These assumptions have been verified in practice during experimental field deployments performed by our lab. The model parameters are the average free surface width T_{0i}, the average bottom slope S_{bi}, the average Manning's coefficient n, the average discharge Q_i, and the average downstream stage Y_{Xi} for each channel i ($i = 1, \cdots, 5$). These parameters are known to us experimentally.

Based on measurements available to us at the *SDC*, *DLC*, *GSS*, and *GES*, the field data at three intermediate locations in the channel network (marked in triangles in Figure 3.2) are chosen to assess the accuracy of the method.

3.3.2 Modal Decomposition of the Measured Data

Since both the discharge and stage are measured at the four USGS stations (*SDC*, *DLC*, *GSS*, and *GES*), the measured flow variable vector Y_m is:

$$\begin{aligned} Y_m =& [q_1^m(0,t), y_1^m(0,t), q_2^m(X_2,t), y_2^m(X_2,t), q_4^m(X_4,t), y_4^m(X_4,t), \\ & q_5^m(X_5,t), y_5^m(X_5,t)]^T \end{aligned} \tag{3.15}$$

where m stands for measured. The fundamental idea is to decompose the measured variables Y_m into a finite sum of N dominant oscillatory modes. In the case of a channel network influenced by the ocean at the downstream end, these modes are essentially the dominant modes produced by tidal forcing. The measured variables are therefore expressed using modal decomposition:

$$Y_m = \sum_{k=0}^{N} \left[D_k e^{j\omega_k t} + \overline{D_k} e^{-j\omega_k t} \right] \tag{3.16}$$

where,

$$D_k = \left[d_k^{(1,1,0)}, d_k^{(1,2,0)}, d_k^{(2,1,X_2)}, d_k^{(2,2,X_2)}, d_k^{(4,1,X_4)}, d_k^{(4,2,X_4)}, d_k^{(5,1,X_5)}, d_k^{(5,2,X_5)} \right]^T \tag{3.17}$$

$D_k = [d_k^{(\alpha,\beta,\gamma)}]^T$ are the Fourier coefficients of the spectral decomposition of Y_m, where α, β, γ respectively denote the channel number, discharge/stage variable, and location of each channel reach. ω_k's are the set of frequencies used for modal decomposition.

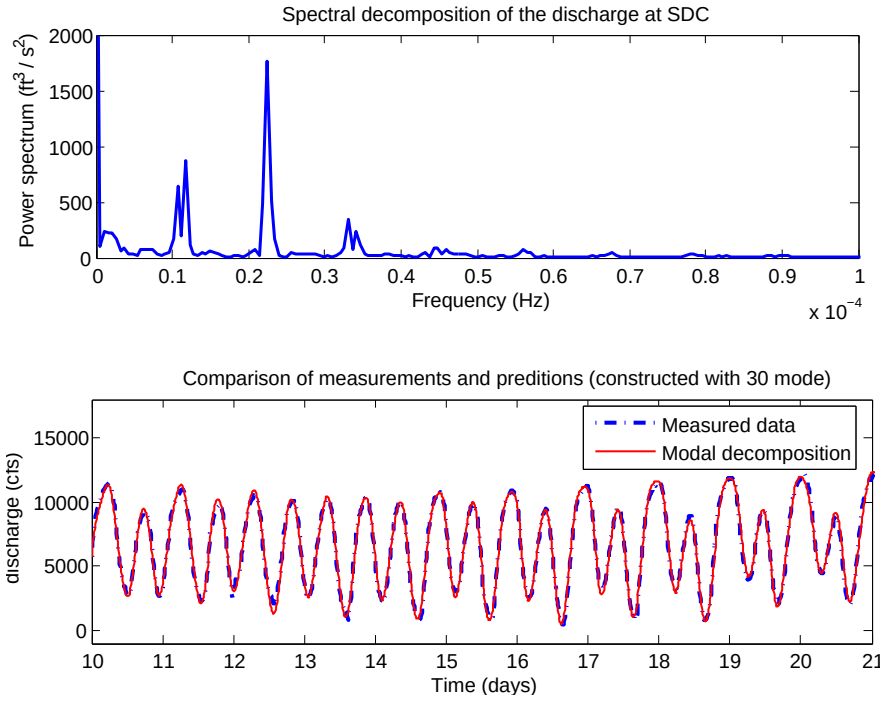


Figure 3.3: *Spectral analysis of the discharge at the SDC Station.*

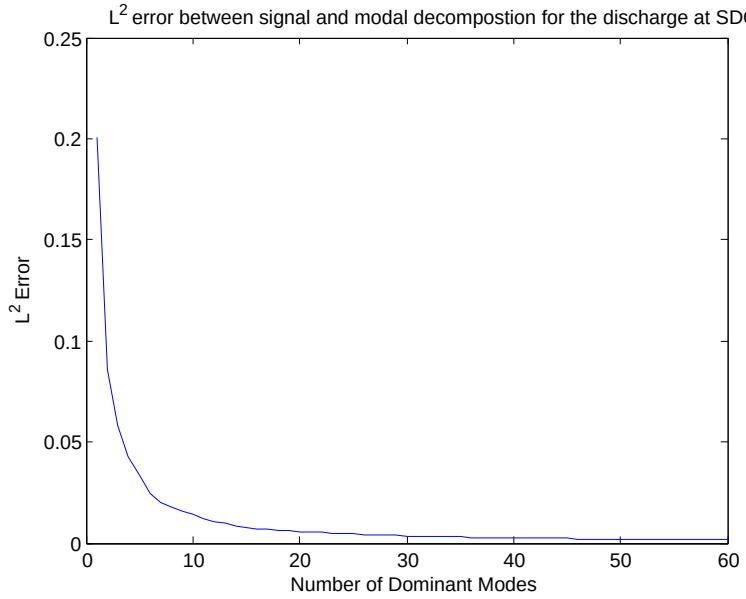


Figure 3.4: *Relative percent error between measurement data and modal represention as a function of the number of modes chosen for the decomposition.*

Figure 3.3 shows the spectral analysis for the discharge data at station *SDC*. There are three dominant tidal frequencies in the system: $\omega_1 = 2.31 \times 10^{-5}\ s^{-1}$ (or period $12.4\ hrs$ tide, corresponding to the M2 tide generated by the moon), $\omega_2 = 1.16 \times 10^{-5}\ s^{-1}$ (or period $24\ hrs$ tide, corresponding to the K1 tide generated by the sun) and a $\omega_3 = 1.11 \times 10^{-5}\ s^{-1}$ (or period $25\ hrs$ tide). The power

spectrum is cut-off at $70ft^3/s^2$ to determine the 30 dominant frequencies. The second plot in Figure 3.3 and Figure 3.4 indicate that 30 modes are sufficient to capture the signal. The amplitude at 0 Hz is essentially the nominal stage. Similar arguments hold for the other measurements.

3.3.3 Hydraulic Model of the Sacramento River and Georgiana Slough

The open channel network system in this study consists of five individual channels, as shown in Figure 3.2. For each channel, the discharge and stage at upstream and downstream are related by a non-uniform transfer matrix:

$$\begin{pmatrix} q_i(X_i,s) \\ y_i(0,s) \end{pmatrix} = \begin{pmatrix} g^n_{i,11}(X_i,X_i,s) & g^n_{i,12}(X_i,X_i,s) \\ g^n_{i,21}(0,X_i,s) & g^n_{i,22}(0,X_i,s) \end{pmatrix} \begin{pmatrix} q_i(0,s) \\ y_i(X_i,s) \end{pmatrix}$$
$$i = 1, \cdots, 5. \tag{3.18}$$

The linear relationships between internal boundary conditions at the two junctions are:

$$\begin{aligned} y_1(X_1,s) &= y_2(0,s) \\ y_2(0,s) &= y_3(0,s) \\ q_1(X_1,s) &= q_2(0,s) + q_3(0,s) \\ y_3(X_3,s) &= y_4(0,s) \\ y_4(0,s) &= y_5(0,s) \\ q_3(X_3,s) &= q_4(0,s) + q_5(0,s) \end{aligned} \tag{3.19}$$

A total of twenty flow variables $q_i(x,s)$, $y_i(x,s)$ (for $x = 0$ or X_i, $i = 1, 2, \cdots, 5$) are included in the system (3.18) and (3.19). These flow variables are related in a linear model $M(s)Z(s) = 0$ equation (3.5), with

$$\begin{aligned} Z(s) =& [q_1(0,s), y_1(0,s), q_1(X_1,s), y_1(X_1,s), q_2(0,s), y_2(0,s), q_2(X_2,s), y_2(X_2,s), \\ & q_3(0,s), y_3(0,s), q_3(X_3,s), y_3(X_3,s), q_4(0,s), y_4(0,s), q_4(X_4,s), y_4(X_4,s), \\ & q_5(0,s), y_5(0,s), q_5(X_5,s), y_5(X_5,s)]^T. \end{aligned}$$

Here, $M(s)$ is a 16 by 20 matrix, which encodes the 16 equations comprised of (3.18) (five channels) and (3.19) (internal boundary conditions).

Since rank $(M(s)) = 16$, given four boundary flow variables $Z_{givenBC} \subset Z$, all the other sixteen boundary flow variables $Z_{otherBC} = Z \setminus Z_{givenBC}$ can be uniquely determined by the sixteen equations set (3.18) (3.19).

Assume that the four known external boundary conditions of the network are: the discharge at *SDC*: $q_1(0,s)$, the stage at *DLC*: $y_2(X_2,s)$, the stage at *GSS*: $y_4(X_4,s)$ and the stage at *GES*: $y_5(X_5,s)$. All the other boundary flow variables can be solved by equation (3.6). More specifically,

$$
\begin{aligned}
Z_{givenBC} =& [q_1(0,s), y_2(X_2,s), y_4(X_4,s), y_5(X_5,s)]^T \\
Z_{otherBC} =& [y_1(0,s), q_1(X_1,s), y_1(X_1,s), q_2(0,s), y_2(0,s), q_2(X_2,s), q_3(0,s), y_3(0,s), \\
& q_3(X_3,s), y_3(X_3,s), q_4(0,s), y_4(0,s), q_4(X_4,s), q_5(0,s), y_5(0,s), q_5(X_5,s)]^T
\end{aligned}
$$

$$
R(s) = R_1(s)^{-1} R_2(s)^T
$$

$R_1(s) =$

$$
\begin{pmatrix}
0 & 1 & -g_{1,12}(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -g_{1,22}(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -g_{2,11}(s) & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -g_{2,21}(s) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & g_{3,11}(s) & 0 & -1 & g_{3,12}(s) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & g_{3,21}(s) & 0 & -1 & g_{3,22}(s) & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g_{4,11}(s) & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g_{4,21}(s) & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g_{5,11}(s) & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g_{5,21}(s) & 1 & 0 \\
0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & -1 & 0 & 0
\end{pmatrix}
$$

$$R_2(s) = \begin{pmatrix} g_{1,11}(s) & g_{1,21}(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_{2,12}(s) & g_{2,22}(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_{4,12}(s) & g_{4,22}(s) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_{5,12}(s) & g_{5,22}(s) & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Table 3.1: Parameters for the Sacramento River and the Georgiana Slough.

Channel	Q_{0i} (m^3s^{-1})	$Y_{Xi}(m)$	$T_{0i}(m)$	S_{bi}(m/km)	n($m^{-1/3}$s)	$X_i(m)$	$C_0(m/s)$	X_i/C_0(s)
$i = 1$	186.7	5.6	115	-0.04	0.0323	2800	7.42	377.4
$i = 2$	83.9	4.1	110	-0.09	0.0323	2000	6.30	317.5
$i = 3$	113.1	7.7	110	-0.04	0.0323	1300	8.71	149.3
$i = 4$	58.1	4.0	56	-0.19	0.0323	600	3.40	176.5
$i = 5$	65.2	5.3	89	-0.04	0.0323	1600	7.19	222.5

The parameters of the model are listed in Table 3.1. The mean discharge (Q_{0i}) of Channels 1, 2, 4, and 5 are computed using the measured discharge at *SDC*, *DLC*, *GSS*, *GES*, respectively. It is clear that the measurement data are inconsistent, since $Q_{01} \neq Q_{02} + Q_{04} + Q_{05}$. To partially compensate for the measurement error, the mean discharge at Channel 3 is set to be: $Q_{03} = [(Q_{01} - Q_{02}) + (Q_{04} + Q_{05})]/2$.

3.3.4 Data Reconciliation of the Measured Data

Let us assume that the measured variables Y_m are independent and subject to a Gaussian distributed noise. Based on the static model (3.6), the measurable variables are linked by a static relationship of the following form:

$$P(s)Y(s) = 0 \tag{3.20}$$

where

$$P(s) = \begin{bmatrix} R(s)_{1,1} & R(s)_{1,2} & R(s)_{1,3} & R(s)_{1,4} & -1 & 0 & 0 & 0 \\ R(s)_{6,1} & R(s)_{6,2} & R(s)_{6,3} & R(s)_{6,4} & 0 & -1 & 0 & 0 \\ R(s)_{13,1} & R(s)_{13,2} & R(s)_{13,3} & R(s)_{13,4} & 0 & 0 & -1 & 0 \\ R(s)_{16,1} & R(s)_{16,2} & R(s)_{16,3} & R(s)_{16,4} & 0 & 0 & 0 & -1 \end{bmatrix}$$

given $Y(s) \setminus Z_{givenBC}$ is the first, sixth, thirteenth and sixteenth element of $Z_{otherBC}$.

Now, combining the solution of the data reconciliation problem (3.14) with the static model (3.20), reconciliated measurements $\hat{Y}$ can be calculated.

Assume that $\hat{Y}$ is in the form:

$$\hat{Y} = \sum_{k=0}^{N} \left[B_k e^{j\omega_k t} + \overline{B_k} e^{-j\omega_k t} \right] \tag{3.21}$$

where $B_k = [b_k^{(\alpha,\beta,\gamma)}]^T$ is the Fourier coefficients vector of the spectral decomposition of $\hat{Y}$, and α, β, γ represent the channel number, discharge/stage variable, location of each channel reach respectively:

$$B_k = \left[b_k^{(1,1,0)}, b_k^{(1,2,0)}, b_k^{(2,1,X_2)}, b_k^{(2,2,X_2)}, b_k^{(4,1,X_4)}, b_k^{(4,2,X_4)}, b_k^{(5,1,X_5)}, b_k^{(5,2,X_5)} \right]^T \tag{3.22}$$

For specific dominant ω_k ($k = 1, \cdots, N$), the coefficient vector B_k in the equation (3.21) is calculated by equation (3.14):

$$B_k = \begin{pmatrix} \mathbf{I}_{8,8} & \mathbf{0}_{8,4} \end{pmatrix} \begin{pmatrix} W^{-1} & P(s)^T \\ P(s) & \mathbf{0}_{4,4} \end{pmatrix}^{-1} \begin{pmatrix} W^{-1} D_k \\ \mathbf{0}_{4,1} \end{pmatrix} \tag{3.23}$$

The reconciliated boundary condition data is shown and compared to measured data in Figures 3.5, 3.6 and 3.7. For clarity, the mean flow has been subtracted from the plots in the interest of magnifying the display scale. The reconciliated data is very close to the measurements in these figures.

The difference between the reconciliated data and measurements is further evaluated in Table 3.2. Three primary evaluation measures are analyzed here:

- The maximum value is the maximum difference between the reconciliated and measured data at the same time steps.
- The coefficient of efficiency E is defined as [110]:

$$E = 1 - \left[\frac{\sum_{i=1}^{N} (\hat{u}_i - u_i)^2}{\sum_{i=1}^{N} (u_i - \overline{u_i})^2} \right] \tag{3.24}$$

  where u_i is the flow variable of interest (for example q_i or y_i in this study), $\hat{u}_i$ is the reconciliated/modeled flow variable, $\overline{u_i}$ is the mean of u_i, for $i = 1$ to N measurement events. If the measured data is perfect, $E = 1$. If $E < 0$, the corresponding measurement is not reasonable and must be excluded from the modeling procedure.

- The last statistic evaluation of the analysis is the correlation coefficient ρ, given by:

$$\rho = \frac{\sum_1^N (u_i - \overline{u_i})(\hat{u}_i - \overline{\hat{u}_i})}{\sqrt{\sum_1^N (u_i - \overline{u_i})^2 \sum_1^N (\hat{u}_i - \overline{\hat{u}_i})^2}} \tag{3.25}$$

where $\overline{\hat{u}_i}$ represents the mean of reconciliated flow for $i = 1\ to\ N$ measurement events.

Table 3.2: Max-value, ρ-value and E-value for reconciliated data and measured data.

Variable	USGS Station	Max-value	E-value	ρ-value
Discharge	SDC	23.66 m^3/s	0.9930	0.9975
	DLC	28.23 m^3/s	0.9368	0.9883
	GES	13.00 m^3/s	0.9968	0.9985
	GSS	18.41 m^3/s	0.9368	0.8369
Stage	SDC	0.05 m	0.9889	0.9947
	DLC	0.12 m	0.9504	0.9759
	GES	0.07 m	0.9847	0.9935
	GSS	0.05 m	0.9938	0.9989

3.3.5 Method Validation

We use existing USGS sensors placed in the Delta as measurement points, and deploy our own sensors at selected locations to produce data used for the validation. The method is validated with measurements at existing USGS fixed sensor stations (see Figure 3.8), used as measurement points (see exact location in Figure 3.2).

Deployable Berkeley sensors (see Figure 3.8) were placed at locations A, B, C on the map of Figure 3.2. Location A is at the downstream of the junction of Sacramento River and Delta Cross Channel; Location B is at the downstream of *GSS* branch; Location C is at the downstream of Sacramento branch. Without loss of generality, the discharge at Location A, along with the stage data at three locations, are used to test the method. The measurements were conducted between 11/01/2007 and 11/12/2007.

Model Validation with Measured Boundary Conditions

Following the steps described in Section 3.2, the flow variables at the boundaries of each branch $Z_{otherBC}$ are calculated using Equation (3.6). Here, the measured discharge at *SDC*, stage at *DLC*, *GES*, *GSS* are used as $Z_{givenBC}$. The flow variables along each branch are estimated using the non-uniform transfer matrix (Equation (3.7)). The simulation results are shown in Figure 3.9.

Model calibration and validation are further evaluated using E-value and ρ-value. Table 3.3 summarizes the values of ρ and E in the validation sets of our channel flow model.

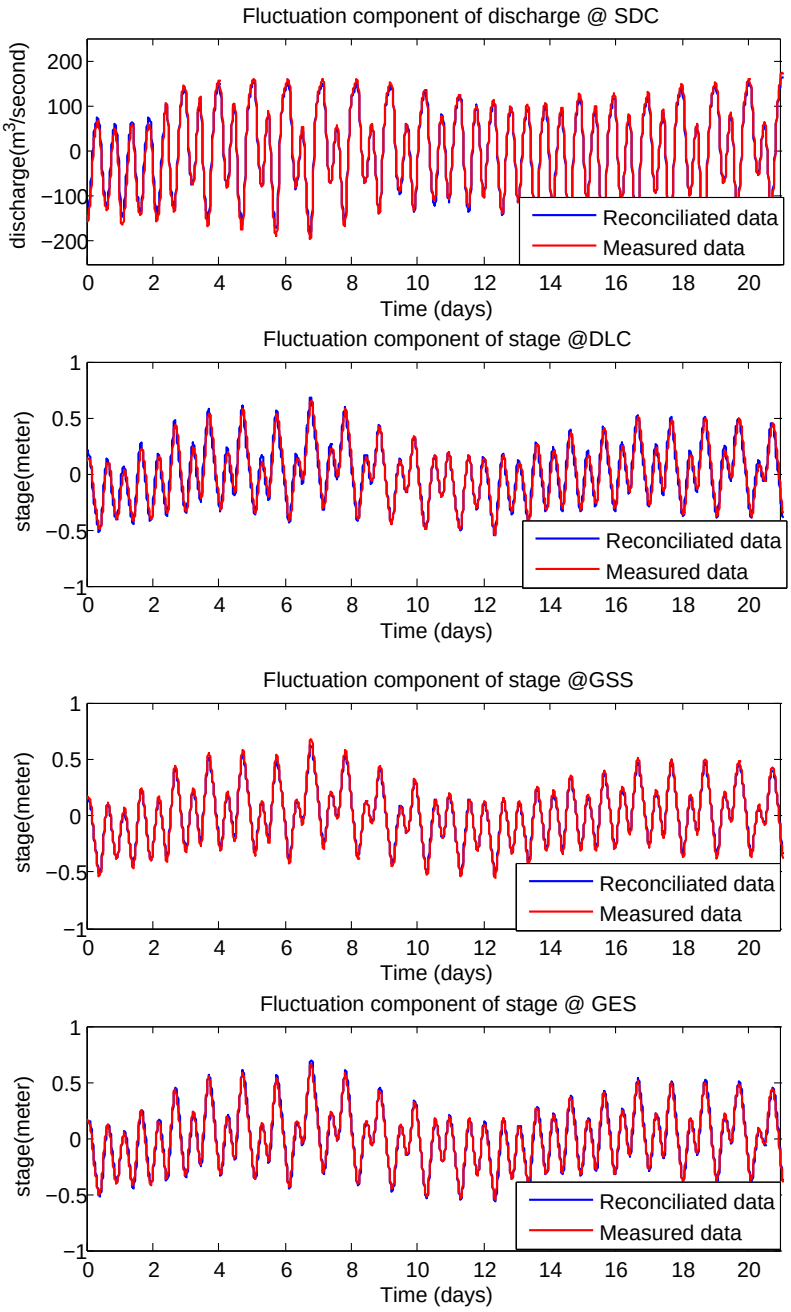


Figure 3.5: *Reconciliated boundary condition data vs. measured data.*

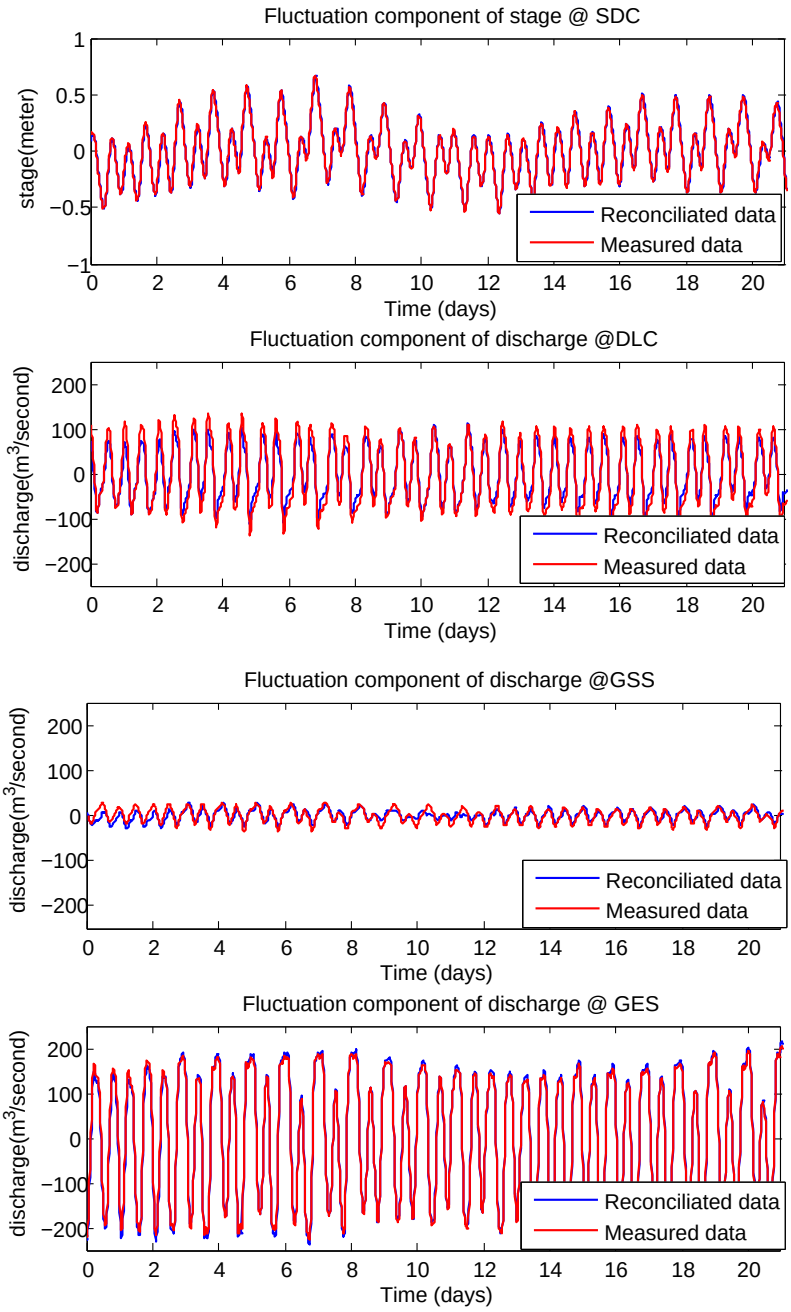


Figure 3.6: *Reconciliated boundary condition data vs. measured data.*

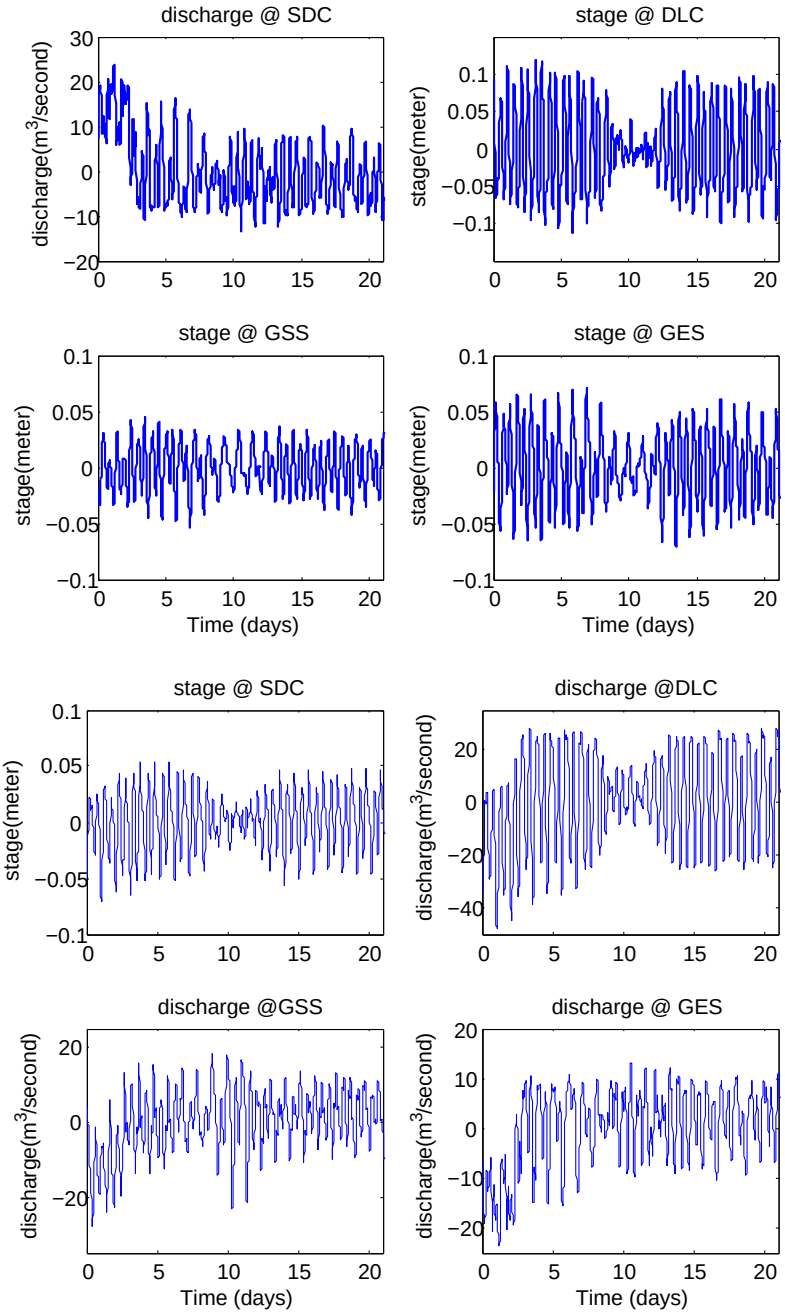
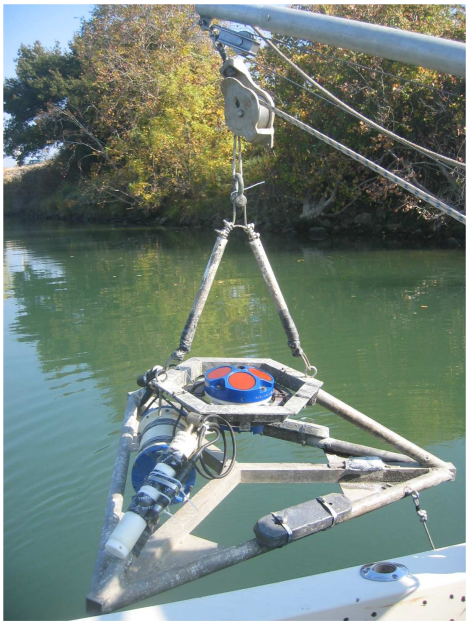


Figure 3.7: *The difference between reconciliated data and measured data.*

(a) USGS Sensor station



(b) UC Berkeley deployable ADCP sensor

Figure 3.8: *Left: USGS Sensor station at GSS, used as a measurement sensor. Right: Deployable ADCP sensor, used in Section 3.3.5 for gathering the validation data (three of them were deployed between 11/01/2007 and 11/12/2007 in order to gather the data for this study).*

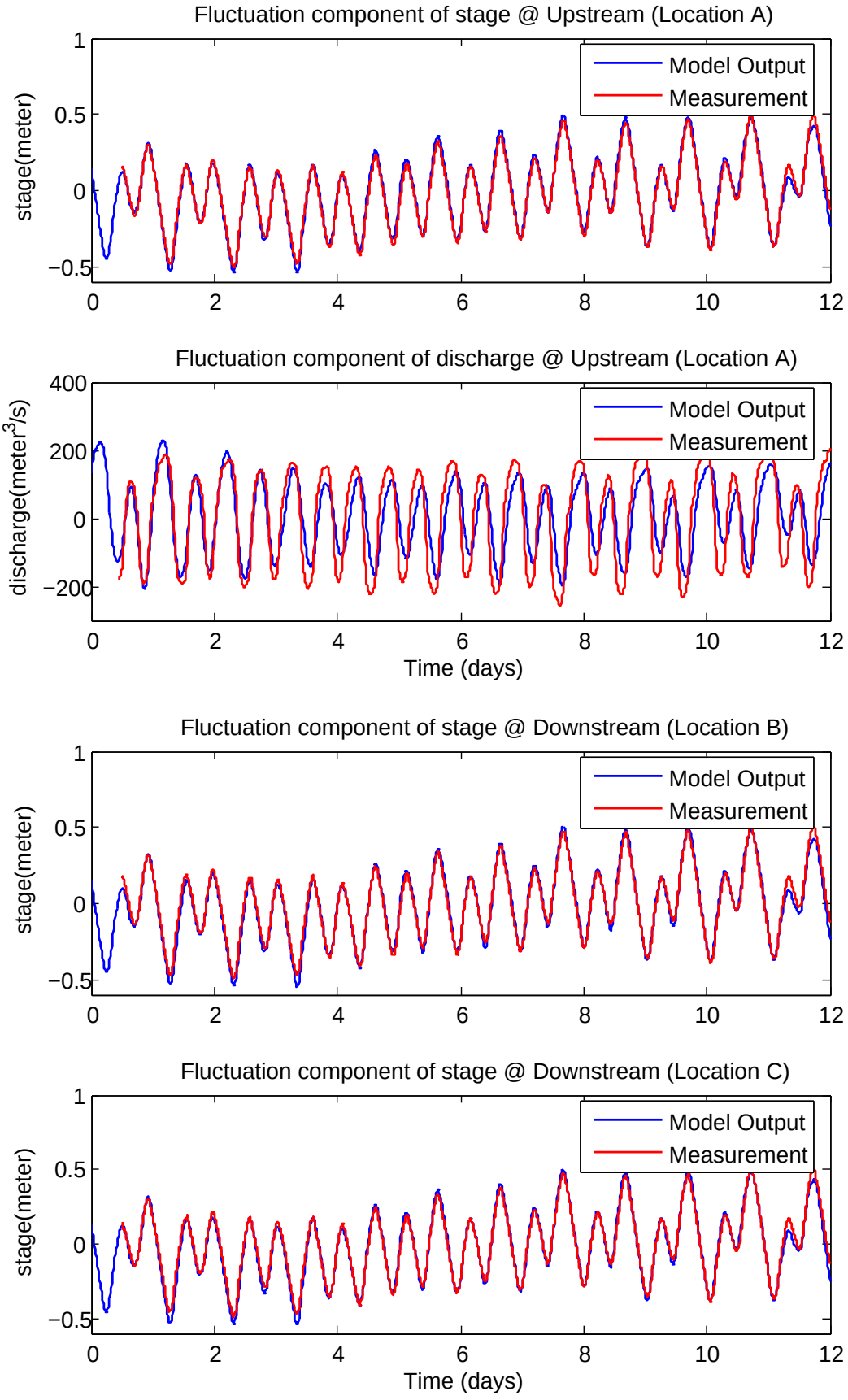


Figure 3.9: *Validation of the model output with measurement using USGS measurements.*

Table 3.3: ρ-value and E-value for model validation with measured boundaries.

Location	A	A	B	C
Variables	discharge	stage	stage	stage
E	0.7219	0.9820	0.9796	0.9807
ρ	0.8555	0.9922	0.9916	0.9927

From Figure 3.9 and Table 3.3, it is clear that the discharge at Location A is estimated with the least precision, as the characteristics of either the phase or the amplitude suffer a significant offset. Although we have precise stage estimation at Locations A, B, and C, the model does not provide enough information to characterize the flow in the experiment area.

Modal Validation with Reconciliated Boundary Conditions Based on all Measurements

We will use reconciliated data shown in Figure 3.5 and 3.6 as $Z_{givenBC}$. The flow variables $Z_{otherBC}$, $q_i(s,x)$, $y_i(x,s)$ are calculated using Equations (3.6) and (3.7). The simulation results are shown in Figure 3.10.

The values of ρ and E are listed in Table 3.4. Both ρ-values and E-values are close to unity.

Table 3.4: ρ-value and E-value for model validation after reconciliation.

Location	A	A	B	C
Variables	discharge	stage	stage	stage
E	0.9775	0.9643	0.9768	0.9612
ρ	0.9895	0.9876	0.9897	0.9875

Table 3.4 and Figure 3.10 thus indicate that the proposed model reconciliation approach provides a higher accuracy in the flow estimation.

Exclusion of Erroneous Sensors

Data reconciliation method enables us to detect and exclude sensors with erroneous measurements and further estimate the boundary conditions using the remaining properly-working sensors. Based on known results in data reconciliation, when the reconciliated data is close to the "true" data, the difference between the reconciliated and measured data must follow a Gaussian distribution with a zero mean. If this condition is not satisfied, the sensor is deemed to be malfunctioning and not suitable for measurements. In order to test the performance of our method, we first intentionally add a large (10 times in magnitude) perturbation to the discharge data measured at DLC, and conduct the standard data reconcliliation procedure described in preceding sections. The probability density function of the difference between the reconciliated data and measured data ($\hat{Y} - Y_m$) is

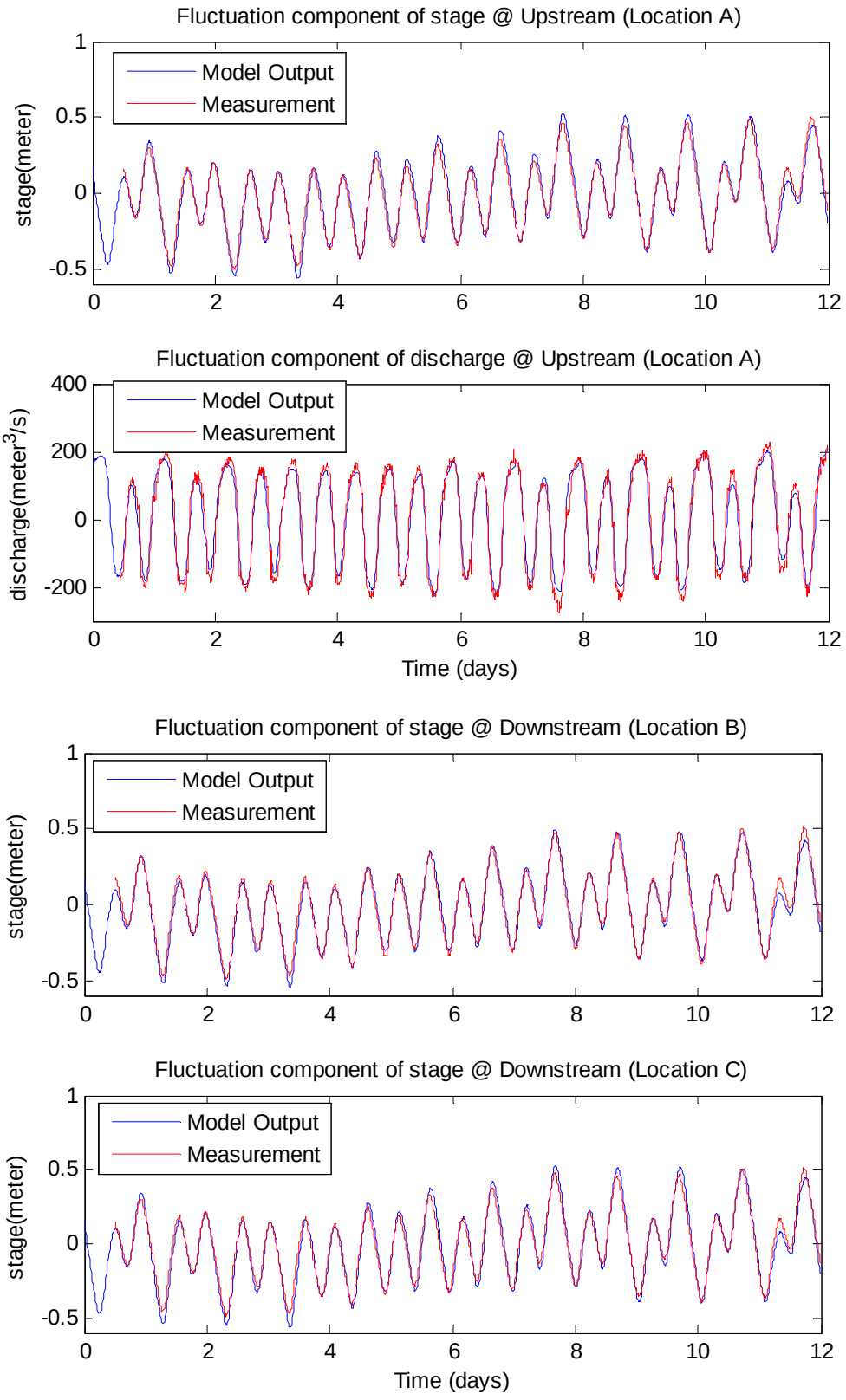


Figure 3.10: *Validation of the model output with measurements using reconciliated BC.*

calculated and compared in Figure 3.11. A Pearson's chi-square (χ^2) test is further applied to assess whether this probability distribution differs from a theoretical Gaussian distribution [121] [12]. The Pearson's χ^2 statistic is calculated as:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{3.26}$$

where,

$$\begin{aligned}
\chi^2 &= \text{test statistic that asymptotically approaches a } \chi^2 \text{ distribution} \\
O_i &= \text{observed frequency} \\
E_i &= \text{theoretical (Gaussian) frequency} \\
n &= \text{the number of possible outcomes of each event}
\end{aligned}$$

From the data (see the subplot in Figure 3.11 in particular), it is obvious that the difference between the measured and reconciliated discharge at station DLC does not follow the Gaussian distribution, implying that the sensor is malfunctioning and should be excluded from the data sets.

We then remove the discharge at DLC from the measurement data, and repeat the data reconciliation procedure. The probability density functions are showed in Figure 3.12. The difference between the reconciliated data and USGS measurements is then more likely to follow a standard Gaussian distribution.

In summary, the χ^2 values are evaluated and assembled in Table 3.5. Note that a small χ^2-value indicates that the observation data distribution is likely to follow a normal distribution.

Table 3.5: χ^2 value for different cases.

Variables	with wrong DLC-discharge data	without DLC discharge data
discharge at SDC	695.55	157.80
stage at SDC	180.20	226.03
discharge at DLC	1200.64	N/A
stage at DLC	369.80	234.73
discharge at GSS	171.36	256.62
stage at GSS	106.60	105.22
discharge at GES	675.58	219.09
stage at GES	159.71	144.54

Furthermore, we use the same three locations (A, B, C) in the experiment domain to validate the reconciliated data. Table 3.6 compares the ρ-value and E-value for the two cases: with the erroneous sensor and without the erroneous sensor.

It is rather interesting to note that modal output is not affected by the intentionally perturbed sensor

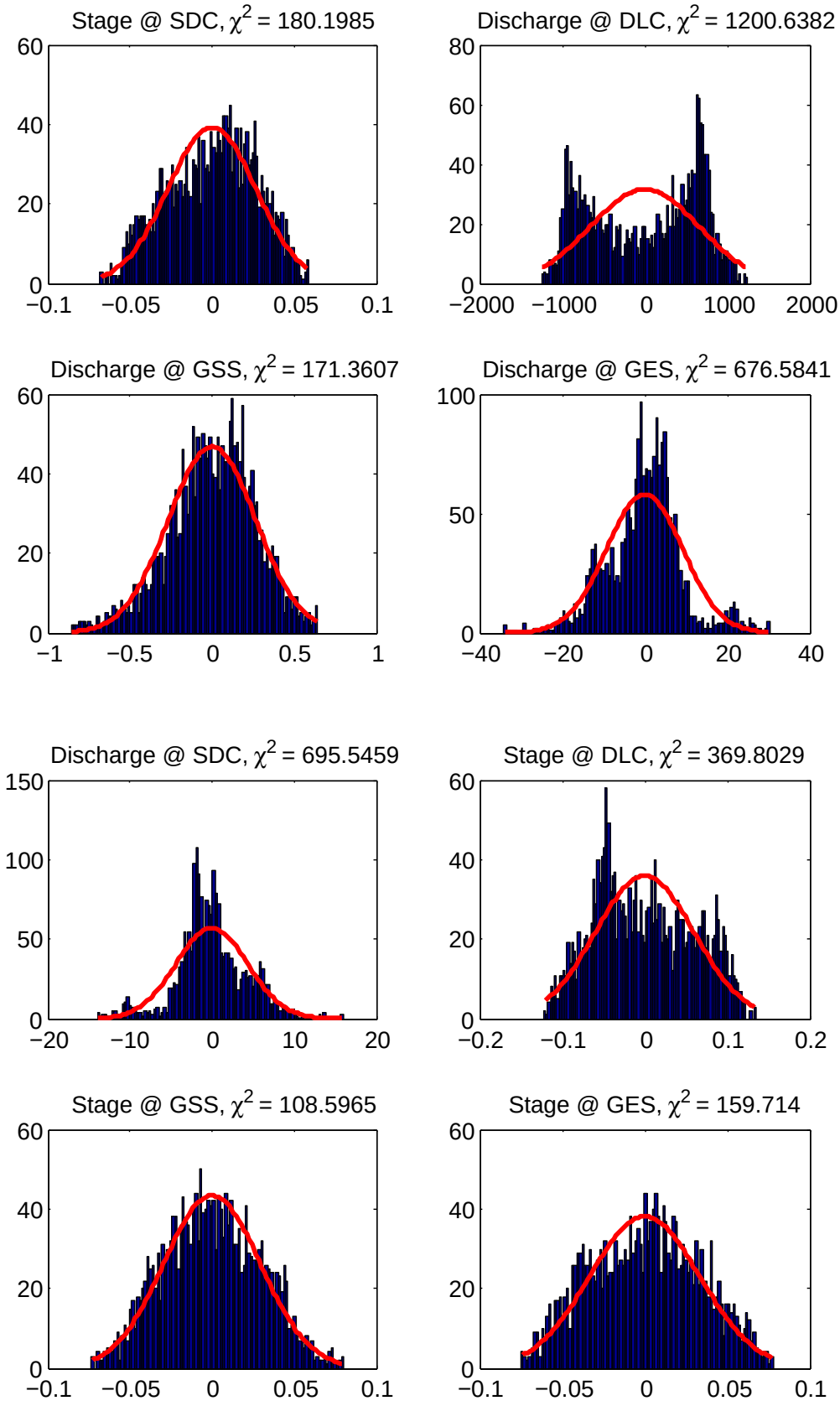


Figure 3.11: *Error distribution between the reconciliated data and USGS measurements (perturbation discharge at DLC is added on purpose in the data set).*

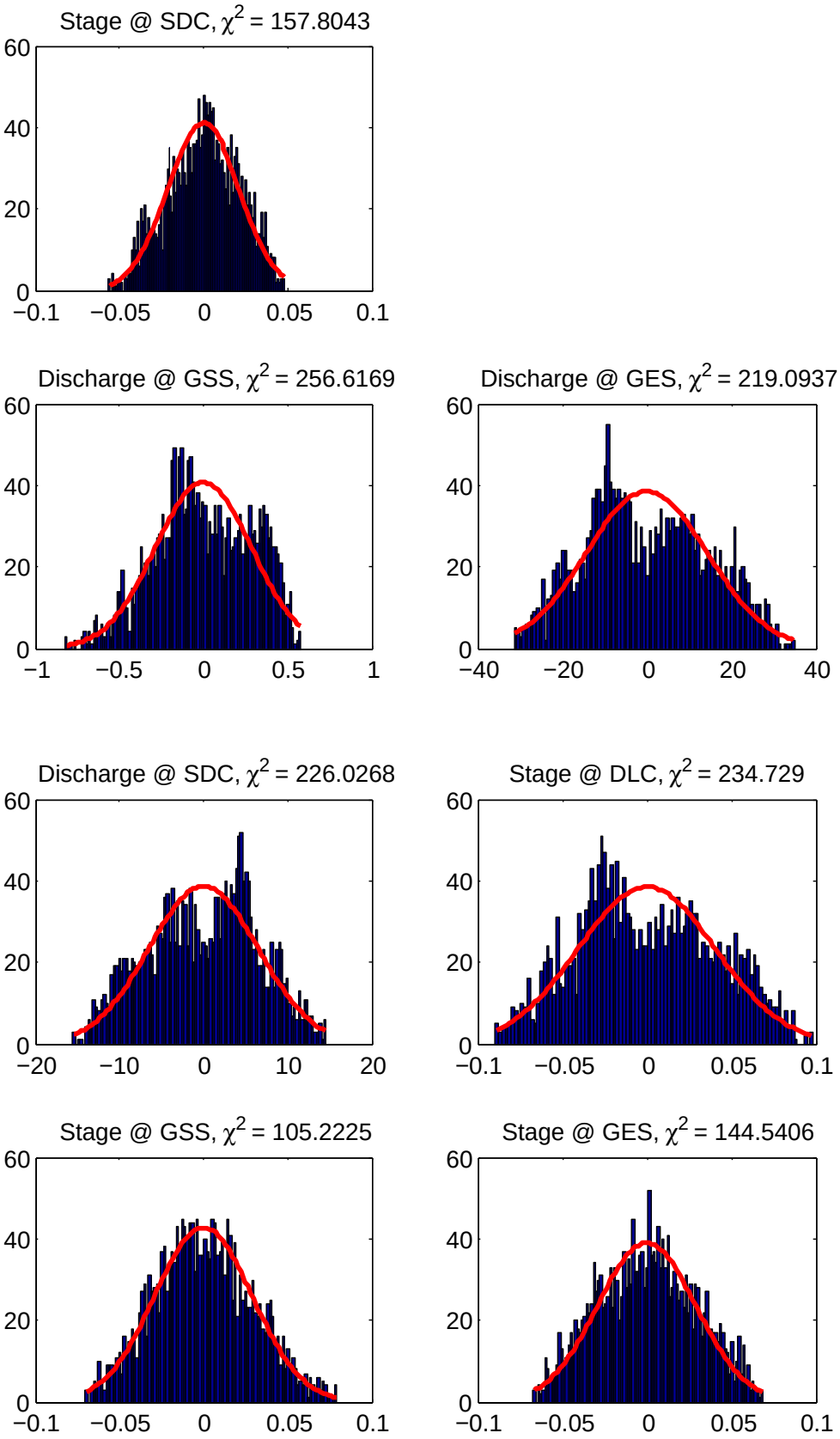


Figure 3.12: *Error distribution between reconciliated and measured data (Discharge at DLC is removed from the data reconciliation process).*

Table 3.6: ρ-value and E-value for modal validation (with and without discharge @ DLC).

Location	A	A	B	C
Variables	discharge	stage	stage	stage
E (with wrong sensor)	0.9777	0.9599	0.9762	0.9567
E (without wrong sensor)	0.9676	0.9651	0.9788	0.9611
ρ (with wrong sensor)	0.9889	0.9867	0.9894	0.9867
ρ (without wrong sensor)	0.9893	0.9892	0.9908	0.9891

data, meaning that the data reconciliation method is robust enough to provide satisfactory boundary conditions even if one of the sensors is malfunctioning.

3.4 Summary

This chapter further discusses flow variable estimations for an open channel network governed by the linearized Saint-Venant equations and subject to periodic forcing. The discharge at the upstream end of the system and the stage at the downstream end of the system are defined as the model inputs; the flow properties at selected internal locations, as well as the other external boundary conditions, are defined as the outputs. Both inputs and outputs are affected by noise and we use the model to estimate this data.

A spatially-dependent LSWE transfer matrix in the frequency domain is constructed to relate the model input and output for the non-uniform steady state case. Modal decomposition allows the output response to be expressed in terms of the spectral coefficients of the input variables and the transfer matrix coefficients evaluated at appropriate locations.

Data reconciliation in this case is reduced to a static least-square minimization problem in the frequency domain, and enables an efficient reconstruction of noisy boundary measurements. Subsequently, the flow properties at any location in the system can be accurately constructed from the input measurements.

The approach proposed in this study has been applied to a channel network in the Sacramento-San Joaquin Delta, using four USGS fixed sensors as measurement points. The flow prediction has been successfully validated at three intermediate locations of the channel system, with deployed sensors from UC Berkeley.

This method is now applied to short-term forecast of internal conditions in the Georgiana Slough and Sacramento River, which we use for our experimental drifter deployments. These internal conditions are particularly useful for our data assimilation and inverse modeling studies featuring Lagrangian data, as will be elaborated in the subsequent chapters.

Chapter 4

Lagrangian Data in 1D Inverse Modeling

4.1 Introduction

In this chapter, a quadratic programming (QP) based method initiated in [139] is developed to determine the open boundary conditions in tidal channel networks by using Lagrangian measurements of the flow. More specifically, we derive the velocity field in a channel network solely from the position information collected by drifters. The proposed method is to minimize the norm of the difference between the drifter observations and model velocity predictions, subject to the constraints given by discretized linear equations. One of the major contributions of this chapter is to pose the problem of estimating the open boundary conditions of a channel network as a quadratic program by minimizing a quadratic cost function and expressing the constraints in terms of linearized equalities and inequalities. The proposed quadratic program can be solved using fast and robust algorithms, and it is capable of providing reliable open boundary conditions for any flow simulations [163].

To assess the performance of the proposed QP method, we investigate a distributed network of channels in the Delta, subject to quasi-periodic tidal forcing. The main obstacle of applying a linear model in the channel networks is the well-known tidal trapping phenomenon [54]. The trapping mechanism makes water elevation and velocity not in phase, inducing the flow dispersion and eddy diffusion at the junctions of channels. The drifter trajectory at these junctions, because of the turbulent mixing processes, usually display a stochastic "spaghetti-like" shape, which is indicative of slow currents. Another contribution of the chapter is to successfully assimilate this chaotic drifter data, and, as a result, the channel network system can be adequately simulated using one-dimensional linearized Saint-Venant model.

4.2 Hydrodynamic Model in Tidal Channel Network

Remark 5. *In natural channels, the shape, size, and slope may vary along the stream length x. In the case of non-uniform flow, the flow variables vary along the length of the channel, i.e., the velocity $V_0(x) \neq V_0 \neq V_X$ and the stage $Y_0(x) \neq Y_0 \neq Y_X$. This non-uniform flow can be best approximated using a backwater profile model [94] [95].*

4.2.1 One-dimensional Numerical Scheme

The Preissman implicit finite difference scheme [25] is applied to linearized Saint-Venant equations:

$$f(x,t) \approx \frac{\theta}{2}(f_{j+1}^{k+1} + f_j^{k+1}) + \frac{1-\theta}{2}(f_{j+1}^{k} + f_j^{k}) \tag{4.1}$$

$$\frac{\partial f}{\partial x} \approx \theta \frac{f_{j+1}^{k+1} - f_j^{k+1}}{\Delta x} + (1-\theta)\frac{f_{j+1}^{k} - f_j^{k}}{\Delta x} \tag{4.2}$$

$$\frac{\partial f}{\partial t} \approx \frac{f_{j+1}^{k+1} + f_j^{k+1} - f_{j+1}^{k} - f_j^{k}}{2\Delta t} \tag{4.3}$$

where $f(x,y)$ is the flow variables (either v or y in our case), $\theta \in (0,1)$ is a time weighting coefficient, j denotes the space step and k the time step. This scheme has the advantage of allowing non-equidistant grids Δx and is unconditionally stable as long as $\theta > 0.5$. This enables a more flexible schematization of the river, especially in the case of strongly varying cross sections. The time step is a function of the required accuracy only and can be chosen freely.

The discretization form of Linearized Saint-Venant equations can be therefore written as:

$$
\begin{aligned}
\frac{y_{j+1}^{k+1}+y_j^{k+1}-y_{j+1}^k-y_j^k}{2\Delta t} = \\
-Y_0(x)\left[\theta\frac{v_{j+1}^{k+1}-v_j^{k+1}}{\Delta x}+(1-\theta)\frac{v_{j+1}^k-v_j^k}{\Delta x}\right] \\
-V_0(x)\left[\theta\frac{y_{j+1}^{k+1}-y_j^{k+1}}{\Delta x}+(1-\theta)\frac{y_{j+1}^k-y_j^k}{\Delta x}\right] \\
-\frac{dY_0(x)}{dx}\left[\frac{\theta}{2}(v_{j+1}^{k+1}+v_j^{k+1})+\frac{1-\theta}{2}(v_{j+1}^k+v_j^k)\right] \\
+\alpha_0(x)\left[\frac{\theta}{2}(y_{j+1}^{k+1}+y_j^{k+1})+\frac{1-\theta}{2}(y_{j+1}^k+y_j^k)\right]
\end{aligned} \tag{4.4}
$$

$$
\begin{aligned}
\frac{v_{j+1}^{k+1}+v_j^{k+1}-v_{j+1}^k-v_j^k}{2\Delta t} = \\
-V_0(x)\left[\theta\frac{v_{j+1}^{k+1}-v_j^{k+1}}{\Delta x}+(1-\theta)\frac{v_{j+1}^k-v_j^k}{\Delta x}\right] \\
-g\left[\theta\frac{y_{j+1}^{k+1}-y_j^{k+1}}{\Delta x}+(1-\theta)\frac{y_{j+1}^k-y_j^k}{\Delta x}\right] \\
-\beta_0(x)\left[\frac{\theta}{2}(v_{j+1}^{k+1}+v_j^{k+1})+\frac{1-\theta}{2}(v_{j+1}^k+v_j^k)\right] \\
+\gamma_0(x)\left[\frac{\theta}{2}(y_{j+1}^{k+1}+y_j^{k+1})+\frac{1-\theta}{2}(y_{j+1}^k+y_j^k)\right]
\end{aligned} \tag{4.5}
$$

Using the above discretization of the linearized Saint-Venant equations (4.4) (4.5), the linear form model for a single channel i can be represented as:

$$E_{k,i}X_{k+1,i} = A_{k,i}X_{k,i} + B_{k,i}U_{k,i} \tag{4.6}$$

where $X_{k,i}$ is the state variable

$$X_{k,i} = (v_{k,i,1}, y_{k,i,1}, \cdots, v_{k,i,l_i}, y_{k,i,l_i})^T \tag{4.7}$$

$U_{k,i}$ is boundary conditions at time $k\Delta t$

$$U_{k,i} = (v_{k,i,1}, y_{k,i,l_i})^T \tag{4.8}$$

where l_i denotes the downstream point of each channel i, and 1 is the upstream point of each channel i. $E_{k,i}$, $A_{k,i}$ and $B_{k,i}$ are matrices constructed by assembling equations (4.4) and (4.5) above. $v_{k,i,j}$ and $y_{k,i,j}$ are respectively the velocity and stage perturbation at location $j\Delta x$ at time $k\Delta t$ in channel i.

4.2.2 Linear Network Model

Consider a simple river junction illustrated in Figure 3.1. The linear equations of hydraulic internal boundary conditions at a junction are specified by equations of mass and energy conservation. Following the approach as specified in the previous chapter, and assuming no change in storage volume within the junction, the continuity equation can be explicitly expressed as:

$$v_{k,1,l_i} \cdot T_1 = v_{k,2,1} \cdot T_2 + v_{k,3,l} \cdot T_3 \tag{4.9}$$

When the flows in all the branches meeting at a junction are subcritical, the equation for energy conservation can be approximated by a kinematic compatibility condition as:

$$y_{k,1,l_i} = y_{k,2,1} = y_{k,3,1} \tag{4.10}$$

The equations are assembled for each individual channel and interior junctions together to model the entire network. The flow variables inside the domain are represented by a linear relationship:

$$E_k X_{k+1} = A_k X_k + B_k U_k \tag{4.11}$$

where X_k is the concatenated vector of $X_{k,i}$ and U_k is the boundary conditions of the channel network system.

The boundary conditions of (4.15) are given by

$$U_k = \left[u(k,i,j)|_{\partial\Omega_{\text{upstream}}}, y(k,i,j)|_{\partial\Omega_{\text{downstream}}}\right] \tag{4.12}$$

and initial conditions given by

$$X_0 = 0 \tag{4.13}$$

The linear network model parameters are the average free surface width $T_{0,i}$, the average bottom slope $S_{b,i}$, the average Manning's coefficient n, the average velocity $V_{0,i}$, and the average downstream stage $Y_{l_i,i}$ for each channel i ($i = 1, \cdots, 3$). These parameters can be determined experimentally.

4.3 Variational Data Assimilation Using Quadratic Programming

4.3.1 General Considerations

There are two main categories of data assimilation methods: variational methods based on optimal control theory and statistical methods based on optimal statistical estimation. We choose a variational method and incorporate it into the linear network model described in the previous sections, since it is practically efficient, demands less computational resources while having a proven record of being reliable if the dynamic model is properly calibrated.

In this section, open boundary condition estimation is formulated using the information of velocity and position measurements provided by a number of drifters which are released in a channel network. Following standard procedure in variational data assimilation, the cost function used in this chapter is constructed to minimize the difference between the measured velocity at the location of the drifters and the velocity predicted by the model. With the linear model constraints, the problem can be formulated as a QP and solved efficiently. Furthermore, with the assumption that tidal flow variables can be expressed by dominant oscillatory modes, the number of estimation variables is substantially reduced.

4.3.2 Notations

We employ the traditional notation of variational data assimilation in discrete time and space [123]:

- X_k : Vector of state variables (v, y) for each mesh point at time $k\Delta t$.
- Y_k : Vector of observed variables at time $k\Delta t$.
- R_k : Covariance matrix of the observation error at time $k\Delta t$.
- H_k : Observation operator, which projects the state vector X_k into the observation subspace containing Y_k.

We deploy D passive drifters in the network to collect Lagrangian velocity measurements in the system, and estimate boundary conditions by minimizing the ℓ^2-norm of the error between the observed data and the corresponding model predictions:

$$\mathcal{J} = \sum_k (Y_k - H_k[X_k])^T R_k^{-1} (Y_k - H_k[X_k]) \tag{4.14}$$

This positive semi-definite quadratic cost function is constrained by:

$$E_k X_{k+1} = A_k X_k + B_k U_k \tag{4.15}$$

In this way, the variational data assimilation problem can be posed as a Quadratic Program:

$$\begin{aligned} \min \quad & \frac{1}{2} X^T P X + q^T X \\ \text{s.t.} \quad & GX \leq h \\ & FX = b \end{aligned} \tag{4.16}$$

where X is the concatenated vector of X_k from time 0 to the final time step; P is a symmetric matrix reducing the ℓ-norm of the error and q is vector containing the information of Y_k, H_k and R_k; F and b are the block diagonal matrix of A_k and B_k. Normally G and h are 0, and the QP can be solved by a linear system. In our case, we may impose heuristic inequality constraints to reduce the search space.

4.3.3 Decision Variables

The decision variables of the QP problem (4.16) are the flow variables at the open boundaries. When expressed in the time domain, the number of decision variables would be equal to the number of boundaries times the number of time steps. Using spectrum analysis, flow variables in a tidal system can be modeled with seven dominant tidal modes, as shown in Figure 4.1.

Table 4.1: Dominant Tidal Modes in the Delta.

Tide	Tide Period T_l(hours)	Tide Frequency $\omega_l = \frac{2\pi}{T_l}$ $(rad \cdot s^{-1})$
K1	23.9345	$7.2921 \cdot 10^{-5}$
M2	12.4206	$1.4082 \cdot 10^{-4}$
MK3	8.1771	$2.1344 \cdot 10^{-4}$
M4	6.2103	$2.8104 \cdot 10^{-4}$
M6	4.1202	$4.2360 \cdot 10^{-4}$
O1	25.8193	$6.7598 \cdot 10^{-5}$
N2	12.6584	$1.3788 \cdot 10^{-4}$

These dominant tidal modes are also listed in Table 4.1. Thus, any flow variables at the boundaries can be evaluated as:

$$u(k\Delta t) \approx \sum_{l=0}^{7} \left[d_l e^{j\omega_l k\Delta t} + \overline{d_l} e^{-j\omega_l k\Delta t} \right] \tag{4.17}$$

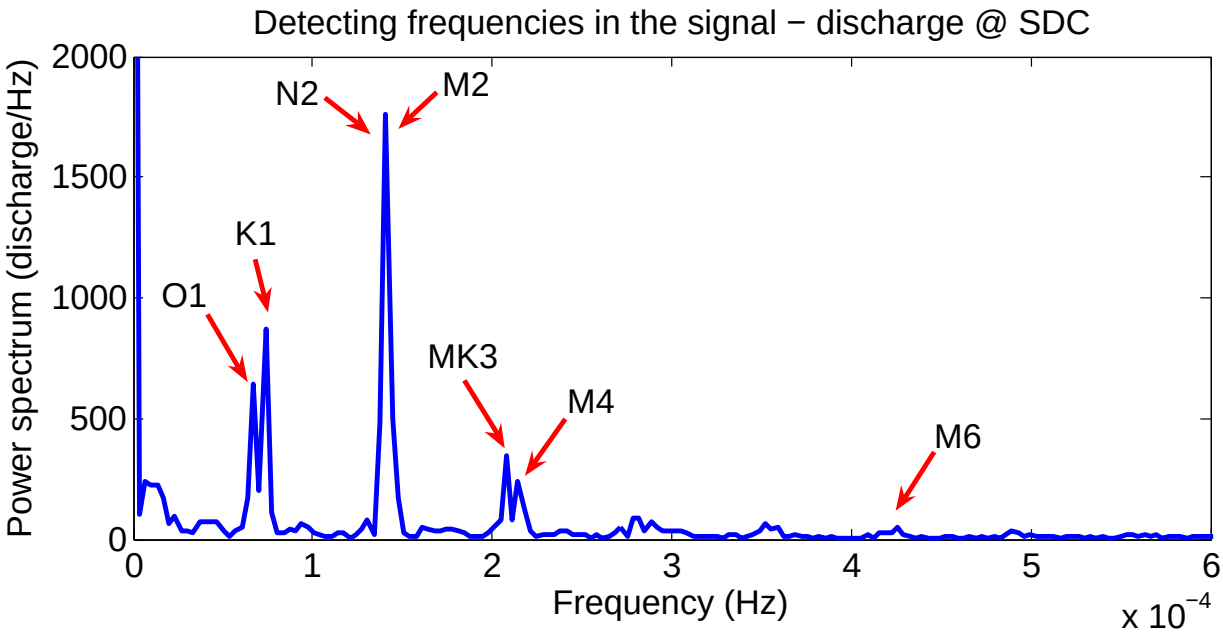


Figure 4.1: *Spectral analysis of the discharge at the SDC Station.*

where $\omega_l = \frac{2\pi}{T_l}$ is the frequency associated with one of the seven dominant tidal periods. The decision variables of this inverse modeling problem are the unknown coefficients d_l corresponding to specified tidal frequencies for each boundary to be estimated. In this way, the number of decision variables is substantially reduced, which speeds up the convergence of QP process.

4.4 Twin Test

4.4.1 Experiment Protocol

The following paragraphs describe a twin test to validate the proposed method. The intuitive way is to assimilate field Lagrangian data into our linear model and compare the estimated boundary conditions with Eulerian measurements at the boundaries. The Lagrangian drifter data in the chapter is generated by using TELEMAC-2D [130], a fully nonlinear shallow water equation (SWE) solver, featuring an unstructured triangular grid mesh and finite element analysis. The virtually simulated drifter data will be replaced by field measurements collected by GPS equipped drifters in subsequent chapters.

A set of fixed Eulerian USGS sensors (see Figure 3.8(a)) on this hydraulic system is employed as the boundary conditions for model simulation, and a finite number of passive drifters are released virtually in the experiment domain. During the inverse modeling process, only these simulated

drifter data are used to re-construct open boundary conditions, which are then compared with the initial boundary setting. Another set of USGS Eulerian sensors, along with the deployed fixed Acoustic Doppler Current Profiling (ADCP) instrumentation (see Figure 3.8(b)) and Water Pressure Sensors, are used to validate the flow characteristics inside the experiment domain. The flow chart of the experiment process is shown in Figure 4.2.

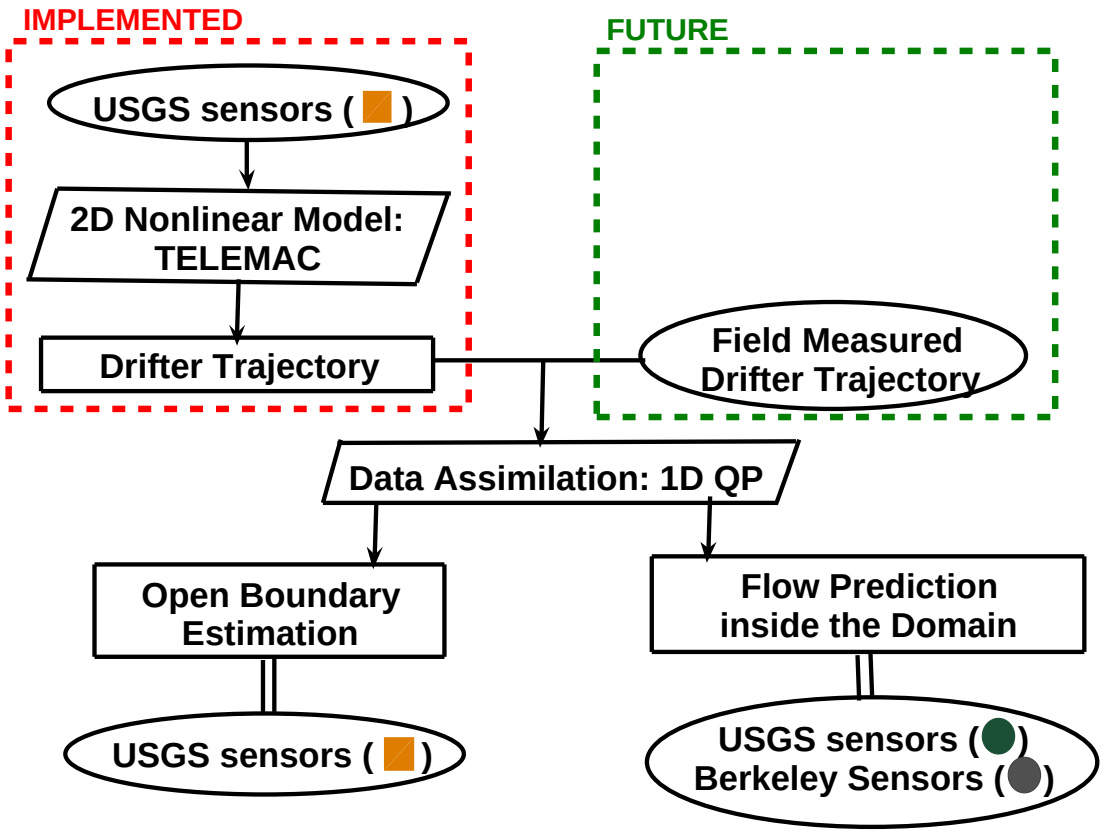


Figure 4.2: *Data Assimilation Flow Diagram. Note: Different shapes are used to represent external data (◯), procedure(◇) and calculated data (□); different lines stand for computations (→) and comparisons (=); different marker colors and shapes indicate the data are measured by sensors at different locations, which will be explained in Figure 4.3.*

4.4.2 2D Shallow Water Equations and Numerical Forward Simulation

In this section, we will set the forward simulation and introduce the Lagrangian measurements.

Two-dimensional Shallow Water Equations

The governing hydrodynamic equations for forward simulation are [130]:

$$\frac{\partial h}{\partial t} + \vec{u} \cdot \nabla h + h \nabla \cdot \vec{u} = 0 \tag{4.18}$$

$$\frac{\partial u}{\partial t} + \vec{u} \cdot \nabla u = -g \frac{\partial \eta}{\partial x} + F_x + \frac{1}{h} \nabla \cdot (h \nu_t \nabla u) \tag{4.19}$$

$$\frac{\partial v}{\partial t} + \vec{u} \cdot \nabla v = -g \frac{\partial \eta}{\partial y} + F_y + \frac{1}{h} \nabla \cdot (h \nu_t \nabla v) \tag{4.20}$$

The friction forces are given by the Manning law:

$$F_x = -\frac{1}{\cos \alpha} \frac{g n^2}{h^{4/3}} u \sqrt{u^2 + v^2} \tag{4.21}$$

$$F_y = -\frac{1}{\cos \alpha} \frac{g n^2}{h^{4/3}} v \sqrt{u^2 + v^2} \tag{4.22}$$

where h is the total depth of water, $\vec{u} = (u, v)$ is the velocity in the domain, g is the gravity acceleration, η is the free surface elevation, ν_t is the coefficient of turbulence diffusion, α is the bed slope of river bottom, and n is the Manning coefficient. The boundary condition and initial condition are given by:

$$u(x, y, t)|_{\partial \Omega_{\text{land}}} = 0, v(x, y, t)|_{\partial \Omega_{land}} = 0 \tag{4.23}$$

$$(u(x, y, t), v(x, y, t))|_{\partial \Omega_{\text{upstream}}} = f(x, y, t) \tag{4.24}$$

$$\eta(x, y, t)|_{\partial \Omega_{\text{downstream}}} = g(x, y, t) \tag{4.25}$$

$$u(x, y, 0) = u_0, v(x, y, 0) = v_0, h(x, y, 0) = h_0, \tag{4.26}$$

where $\partial \Omega$ represents the boundaries of our computational domain and f, g are known functions.

Lagrangian Drifters

The deployed drifters is modeled as passive Lagrangian tracers. In this framework, the drifters move along local flow streamlines, obeying the following equations:

$$\frac{dx_D(t)}{dt} = u[x_D(t), y_D(t), t] \tag{4.27}$$

$$\frac{dy_D(t)}{dt} = v[x_D(t), y_D(t), t] \tag{4.28}$$

with the drifter initial conditions

$$x_D(t) = x_{D,0}, y_D(t) = y_{D,0} \tag{4.29}$$

Numerical Solution

The numerical solutions of the 2D shallow water equations and drifter positions are computed using a commercial hydrodynamic software TELEMAC-2D [130]. TELEMAC-2D applies a streamline upwind Petrov-Galerkin based finite element solver for hydrodynamic equations. The turbulence and mixing process at the estuaries are taken into account.

To generate the drifter data, a forward simulation is run from time t_0 to time t_1 with given boundary conditions to stabilize the flow. At t_1, drifters are released randomly inside the domain and their trajectories are simulated using a Runge-Kutta method and the velocity field provided by the nonlinear shallow water forward simulation. The data assimilation process estimates the boundary conditions, which are compared with the previously given boundary conditions, as well as the flow variables at intermediate locations within the watershed.

4.5 Case Study in the Delta

4.5.1 Introduction to the Experiment Field

The area of interest for our experiment covers the Sacramento River, the Cache Slough, the Steamboat Slough, the Sutter Slough, the Minor Slough, the Delta Cross Channel, and the Georgiana Slough, as shown in Figure 4.3. Most of the time, the direction of mean river flow is from north to south, as indicated with arrows. During the tidal inversion, the water flows in the opposite way.

Ten USGS stations, namely *HWB*, *RYI*, *SRV*, *HWV*, *SUT*, *SSS*, *SDC*, *DLC*, *GES*, and *GSS*, are scatterly located in this experiment field. The stations are marked as orange diamonds and green circles in Figure 4.3. Both velocity and stage are collected every 900 seconds at these stations. Note that in the USGS measurement system, only the stages are measured directly. The velocity data is estimated by a rating curve, which is a relation between stream stage and stream flow. The relation of stream stage to stream flow is in continuous change, and needs to be calibrated frequently. It would introduce errors if the rating curve has not been validated in time.

The field data was collected between 0:00am and 8:30am on 11/12/2007. The same simplifications as detailed in Section 3.3.1 have been made for the flow model in this study.

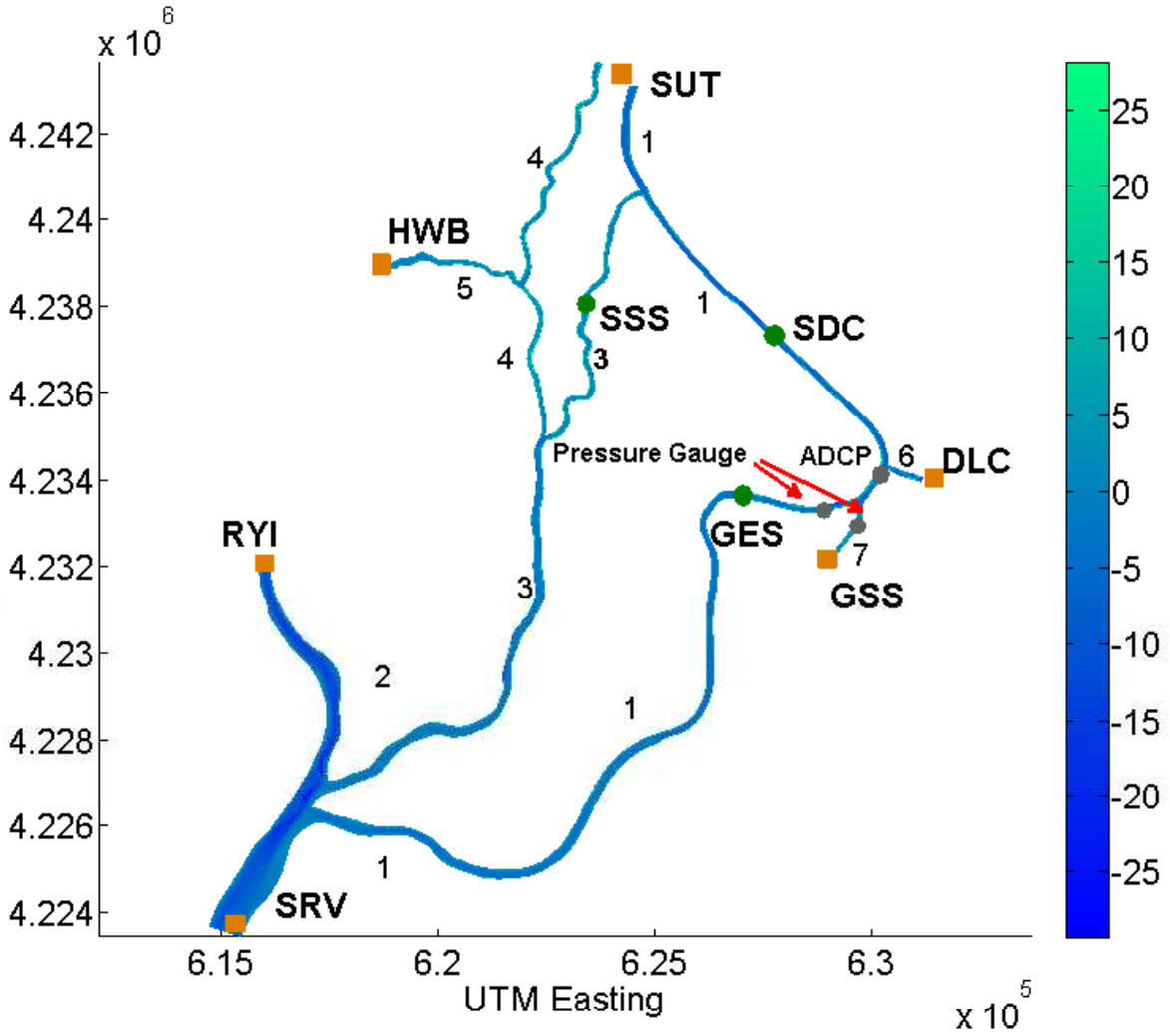


Figure 4.3: *Experiment area in the Sacramento River* (1)*, the Cache Slough* (2)*, the Steamboat Slough* (3)*, the Sutter Slough* (4)*, the Minor Slough* (5)*, the Delta Cross Channel* (6) *and the Georgiana Slough* (7)*.*

4.5.2 Drifter Data Generation

TELEMAC-2D conducts a nonlinear flow simulation using velocity data measured at USGS stations *SRV*, *RYI*, *GSS*, and stage data measured at *DLC* and *SUT*. The geometry of the area is complex; thus we use an unstructured finite element mesh (41375 nodes, 74983 triangular elements). The bottom friction is modeled using Manning's law. The Manning coefficient is chosen as a constant 0.02, both in time and space, corresponding to a straight gravel bottom [38]. The turbulence process is included such that the flow streamline at the estuaries are similar to the reality.

The simulation runs for two and a half hours before the release of the drifters so that a steady state is reached. The drifters are virtually released from 2:30AM to 6:30PM on November 11, 2007. This time period was chosen to capture the highly varying flow in Sacramento Delta. We release a total

of 39 drifters during the experiment (6 hours). The first thirteen drifters are released at 2:30AM on the centerline of selected sub-channels. The other two sets of thirteen drifters are released at 4:30AM and 6:30AM, respectively. Drifter positions are recorded every 60 seconds until the end of the experiment at 8:30AM. Figure 4.4 shows the drifter trajectories and the snapshots of the drifter positions corresponding to the three releases of the drifters.

4.5.3 Validation Results

Following the method described in Section 4.3, we assimilate the drifter data generated by TELEMAC (as described in Section 4.5.2) to reconstruct the boundary conditions at *SRV*, *RYI*, *GSS*, *DLC* and *SUT*. The QP problem is expressed in AMPL, the optimization modeling language, and solved with CPLEX. The assimilation process takes approximately 65 minutes to calculate all 13 sub-channels with a 2.33 GHz Pentium dual core processor.

The reconstructed boundary condition data is shown and correlated to measured data in Figures 4.5. The estimated data is very close to the measurements.

Without loss of generality, the flow variables measured by USGS sensor *GES* (marked as green circle in Figure 4.3), along with the velocity and stage data recorded by selective deployable UC Berkeley sensor (marked as grey circle in Figure 4.3), are used to validate the flow model. The simulation results are shown in Figure 4.6.

The difference between the modeled data and measurements is further evaluated in Table 4.2. Three primary evaluation measures are analyzed here:

- The maximum value is the maximum difference between the estimated and measured data at the same time step.
- The coefficient of efficiency E is defined as [110]:

$$E = 1 - \left[\frac{\sum_{i=1}^{N} (\hat{u}_i - u_i)^2}{\sum_{i=1}^{N} (u_i - \overline{u_i})^2} \right] \tag{4.30}$$

  where u_i is the flow variable of interest (for example v_i or y_i in this study), $\hat{u}_i$ is the modeled flow variable, $\overline{u_i}$ is the mean of u_i, for $i = 1$ to N measurement events. If the measured data is perfect, $E = 1$. If $E < 0$, the corresponding measurement is not credible and must be excluded from the modeling procedure.

- The correlation coefficient ρ is given by:

$$\rho = \frac{\sum_1^N (u_i - \overline{u_i})(\hat{u}_i - \overline{\hat{u}_i})}{\sqrt{\sum_1^N (u_i - \overline{u_i})^2 \sum_1^N (\hat{u}_i - \overline{\hat{u}_i})^2}} \tag{4.31}$$

where $\overline{\hat{u}_i}$ represents the mean of model estimated flow for $i = 1\ to\ N$ measurement events. If the measured data is perfect, $\rho = 1$.

Table 4.2: Max-value, ρ-value and E-value for modeled data and measured data.

Variable	USGS Station	Max-value	E-value	ρ-value
Velocity	GES	0.05 m/s	0.9930	0.9975
	SDC	0.04 m/s	0.9368	0.9883
	SSS	0.055 m/s	0.9968	0.9985
	ADCP	0.07 m/s	0.9435	0.8923
Stage	GES	0.05 m	0.9889	0.9947
	SDC	0.12 m	0.9504	0.9759
	SSS	0.07 m	0.9847	0.9935
	Pressure Sensor I	0.06 m	0.8479	0.8743
	Pressure Sensor II	0.06 m	0.9345	0.8734

Figure 4.6 and Table 4.2 indicate that the proposed approach possesses a good flow estimation accuracy.

4.6 Summary

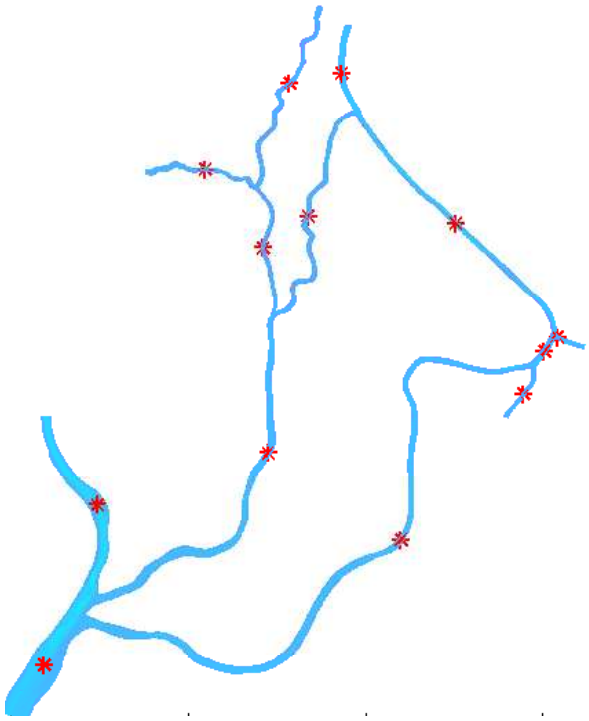
In this chapter, we present a boundary condition estimation method for complex channel networks using Lagrangian data. The channel network system is governed by first-order linearized Saint-Venant equations, subject to periodic forcing. The flow estimation is posed as a PDE constrained optimization problem with the objective of minimizing the norm of the difference between the observed variables and model outputs. After linearizing and discretizing the governing equations using an implicit discretization scheme, linear constraints are constructed which leads to a quadratic programming formulation of the estimation problem.

A major advantage of the 1D QP formulation is that it requires low computational cost, making the method applicable to many vast and complex hydraulic networks. Modal decomposition allows the estimated outputs to be expressed in terms of dominant tidal frequencies. This greatly reduces the number of decision variables, and consequently lowers the computation complexity.

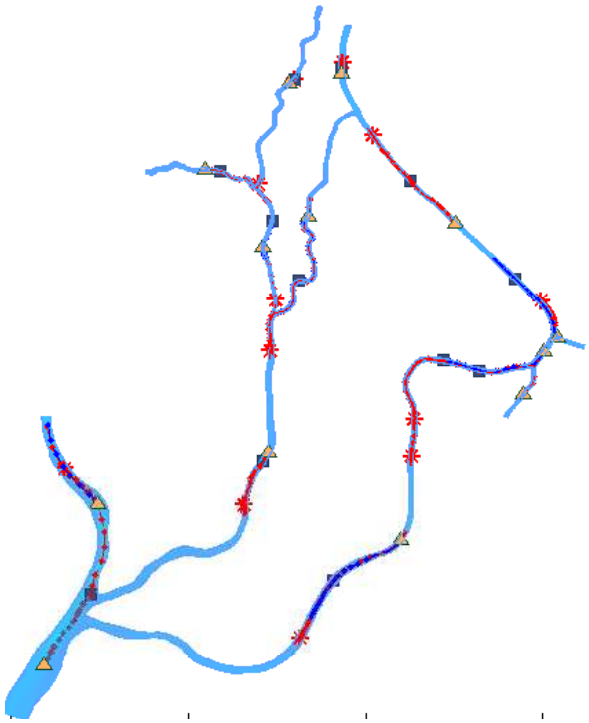
The performance of the method has been validated in a twin experiment in the delta, in which the drifter data are generated by a 2D nonlinear shallow water model (TELEMAC).

In the next chapter, we further examine the effectiveness of Lagrangian data in flow estimation, using real drifter data collected from GPS-equipped devices deployed in the Sacramento-San Joaquin

Delta. The effects of different number of drifters and deployment strategies will also be discussed.

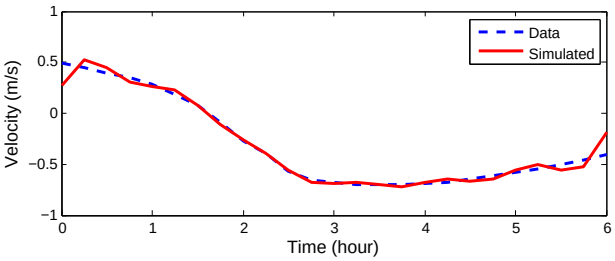


(a) Time Step 0

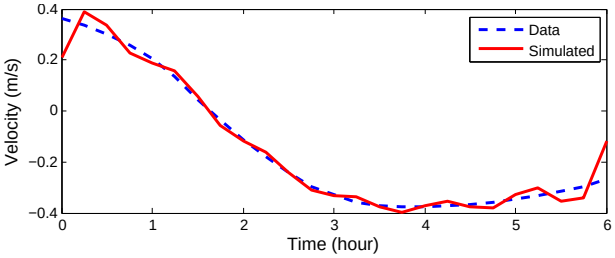


(b) Time Step 240

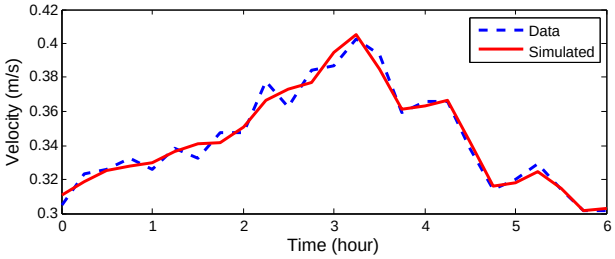
Figure 4.4: *Drifter trajectories and their release positions.* 13 *drifters are released at time step* 0 *(*$*$*) and* 13 *drifters at time step* 240 *(*Δ*). The drifter positions are recorded every 60 second until the experiment ends.*



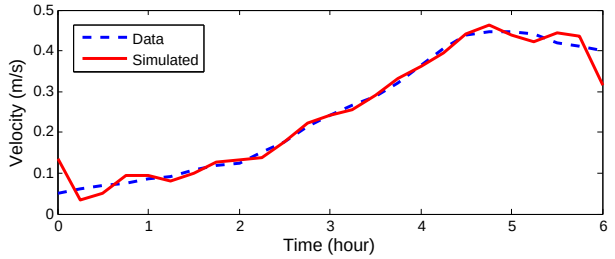
(a) Velocity Varying with Time at USGS station: RYI



(b) Velocity Varying with Time at USGS station: SRV

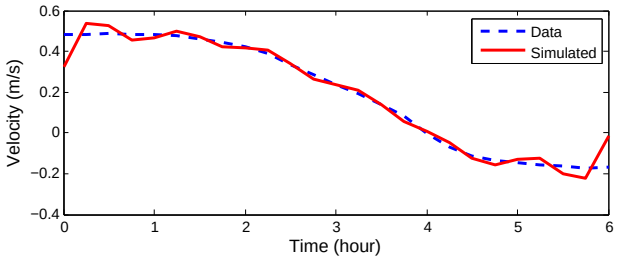


(c) Velocity Varying with Time at USGS station: GSS

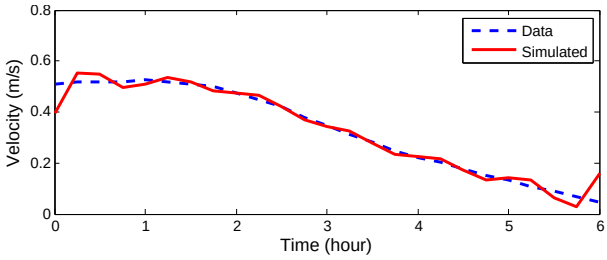


(d) Velocity Varying with Time at USGS station: DLC

Figure 4.5: *Comparison of the estimated boundary condition with USGS measurements at the boundaries of the domain.*



(a) Velocity Varying with Time at USGS station: GES



(b) Velocity Varying with Time at UC Berkeley sensor: ADCP

Figure 4.6: *Validation of the model outputs with USGS and ADCP measurements inside of the domain.*

Chapter 5

Variational Data Assimilation in 1D Channel Networks

5.1 Introduction

To assess the performance of the proposed QP method developed in the previous chapter, we investigated a distributed network of channels, subject to quasi-periodic tidal forcing, in the Sacramento-San Joaquin River Delta [164]. A field operational test was carried out with a fleet of 70 surface drifters, deployed within approximately $0.55\,\mathrm{km}^2$ of the river network. During the approximately 4 hour experiment, 325,000 GPS readings were taken from the surface drifters and collected, in real time, onto a central server. It is the first experiment of this kind conducted at such scale, where high-density Lagrangian data have been collected in a real river environment.

We demonstrate that the proposed QP approach can successfully handle these drifter data to obtain accurate estimations, and, as a result, the channel network system is adequately simulated using one-dimensional linearized Saint-Venant model.

The ramifications of this result are that robust and accurate estimations for the flow in a channel network can be obtained quickly without introducing an unacceptable amount of inaccuracy due to simplifying the model. Thus, flow estimations could be provided even as fast as real-time and give water resource management a useful tool to understand the river system.

5.2 Operational Experiment in the Sacramento River and Georgiana Slough

5.2.1 Experiment Field

On May 9, 2012, an experimental deployment of 70 drifters was conducted at the junction of the Sacramento River and the Georgiana Slough near Walnut Grove, California.

Figure 5.1 shows the spatial domain of the experiment. The Sacramento River is the larger channel including Labels A and B, while the Georgiana Slough splits off from the Sacramento River at F and continues south toward Label C.

The flow along the Sacramento River was approximately 0.46 m/s (1.5 ft/s) in the outgoing (from northeast to southwest) direction. This is the non-inverted tidal condition. Figure 5.2 shows the water velocity at two USGS stations: GES, GSS over time during the experiment.

The original plan was to deploy all of the drifters from the Walnut Grove Public Dock (Label A in Figure 5.1), allow them to drift through the junction, retrieve them at downstream points B and C, and then recycle them at point D and E for the rest of the experimental run.

Unfortunately, there was a significant underwater construction operation at the junction (box F) that day, requiring a mid-experiment change of plans: drifters were initially released from A and picked up around F, redeployed at D and E, and then cycled from B and C. Drifter retrievals and redeployments were performed by two boat teams of three workers (one pilot and two drifter retrievers), as seen in Figure 5.3.

5.2.2 Active Sensors

The design of the active "drifters" was motivated by a number of functional requirements concerning actuation, form factor, and mission life [118], [143].

Under certain circumstances, for example if the drifters need to stay within a study area or avoid some obstacles in water, it is desirable to have the drifters "actuated" when needed. For this purpose, two differential drive motors are installed on each active drifter, enabling it to move cross-stream at a maximum forward velocity of 0.3 m/s. An onboard compass and a proportional-integral-derivative (PID) controller can stabilize the drifter along a desired heading. In addition, an onboard buoyancy control system enables the drifter to "dive" up to 5 m in water, if needed, making it capable of measuring underwater qualities.

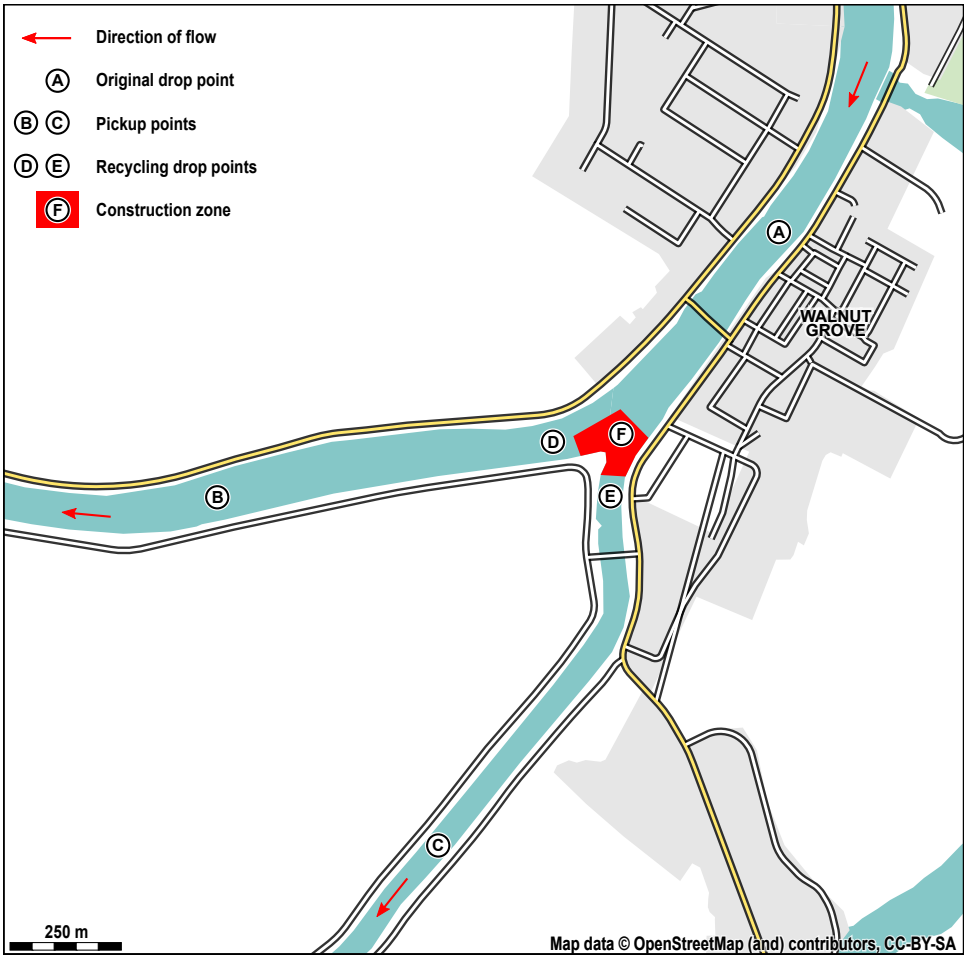


Figure 5.1: *Annotated map of the Walnut Grove experimental area, for the May 9th field operational test.*

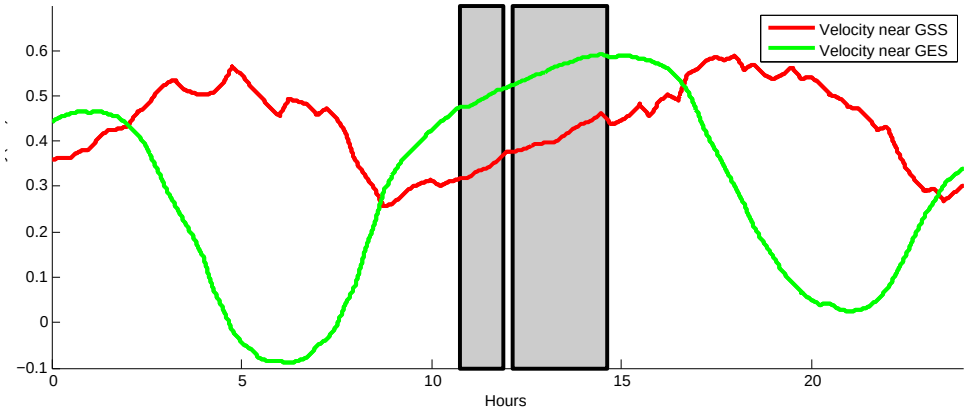


Figure 5.2: *Overview of the experimental timeframe, the two grey boxes denote the time domain during which the drifter data was collected.*

(a) Deployment

(b) Passive sensing

(c) Retrieval

(d) Redeployment

Figure 5.3: *Android drifter usage during the May 9th experiment. Photo credit: Berkeley Lab - Roy Kaltschmidt.*

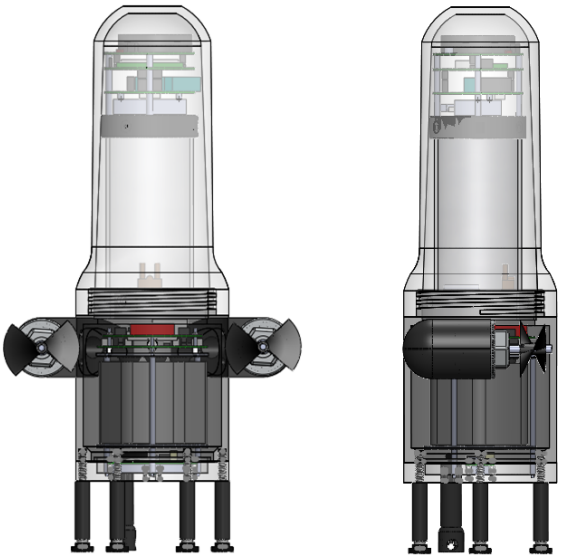


Figure 5.4: *CAD drawing of active drifting sensor. Photo Credit: Carlos Oroza*

The drifter is powered with an 11.1V, 10.4 amp-hour Li-Ion battery, enabling up to 72 hours of mission life with a 10 percent duty cycle.

The active drifter has enabled two communication modules: a Motorola G24 OEM GSM module for direct communication with the data process server, and a Digi XBee-PRO 802.15.4 module for short-range communication among the drifters and the field team.

The drifter hull is a vertically oriented cylinder, designed to be hydrostatically stable (with a center of buoyancy located above its center of mass), and to have a mass to volume ratio that keeps it low in the water (only the antennas extend above the waterline). The hull is constructed from high-performance Acetyl Homopolymer, making it suitable for most aquatic environments.

The manufacturing process for the active sensors is easily scalable to large quantities. So far, over 40 active sensors have been produced and have been deployed in the Sacramento-San Joaquin Delta region.

5.2.3 Android Drifters

In order to increase the amount of collected data through higher fleet numbers (requiring lower production cost per unit), and to achieve greater reliability and improve manufacturability, it is desirable to avoid custom electronics and take advantage of the positioning and communication features of modern mobile phones. This leads to the "Android drifters" [14] (or *passive sensors*) that have been recently developed at UC Berkeley.

1: Duct tape handle/flag
2: PVC hull (filter canister)
3: Motorola DEFY
4: Rubber bands
5: PVC phone holster
6: Foam disk
7: Lithium Ion battery
8: If-found placard
9: Aluminum disk
10: Delrin base

Figure 5.5: *Android drifter with complete parts list used for the study. Photo credit: Jonathan Beard*

Each Android drifter is an inexpensive assembly consisting of a mobile phone running the Android operating system, a lithium-ion battery to extend the phone life, a waterproof enclosure based around a water filter canister, and supporting mechanical parts (Figure 5.5).

The core functionality of the Android app, a custom application written for the Android operating system, is to transmit time-stamped GPS positions to a remote server over the cellular network. The app user interface features the ability to start and stop the service of transmitting and logging, a display of current GPS data and orientation, and a menu to set the drifter ID, measurement frequency and server address.

It should be noted that operating passive drifters in river environments demands greater personnel involvement for deployment, retrieval, and protection.

5.3 Data Observation and Drifter Trajectory Filtering

5.3.1 Characteristics of Lagrangian Data

Lagrangian data has been widely studied in the last two decades, in particular in the field of atmospheric and oceanic sciences [102], [55], [71]. These types of data from both the surface drifters and subsurface floats have been intensively used in two categories: (a) to describe the main statistics of the general circulation by inferring the mean flow structure [27], and (b) to specify the second-order statistics and transport properties [150]. Compared to the traditional Eulerian sensors, Lagrangian sensors have several advantages: First, the devices are inexpensive to build and main-

tain. Secondly, Lagrangian data, especially the surface drifter data, presumably contains records of the flow motion at all scales. Thirdly, with the advancement in GPS technology, it provides more and more detail information about the dynamic flows.

However, raw Lagrangian data is difficult to use and interpret, due to the well-known fact that Lagrangian motion is often affected by local flow perturbations, which are caused by various physical processes, such as turbulence, surface wind, vertical mixing, etc. For example, one of the main obstacles of assimilating Lagrangian data into a hydrodynamic model in channel networks is the well-known tidal trapping phenomenon [54]. The trapping mechanism makes water elevation and velocity out of phase, inducing flow dispersion and eddy diffusion at the junctions of channels. The drifter trajectory at these junctions, due to the turbulent mixing processes, usually displays a stochastic "spaghetti-like" shape, which is indicative of slow currents. However, these small-scale features are not fully addressed by the numerical models, and furthermore cause instability in the data assimilation process. Lagrangian data acquired from the sensors therefore need to be pre-processed for "smoother" trajectories to exclude any measurement error due to hardware malfunctioning, or any small-scale perturbations.

Another important issue is that Eulerian and Lagrangian behaviors are not strictly related to each other. It is not unusual to have regular Eulerian behavior, e.g., a time-periodic velocity field, co-exist with Lagrangian chaos or vice-versa. Moreover, unknown details, such as small scale processes or surface wind forcing, pose further difficulties for processing Lagrangian data. As a result, it is challenging to infer the Eulerian velocity field from the Lagrangian measurements.

Addressing these limitations, we present an efficient and reasonable velocity field based on the Lagrangian data collected from the May 9th experiment using the proposed data assimilation technique derived in Section 4.3. It is one of the first successes of using Lagrangian data in a hydrodynamic system, especially in a complex tide-driven channel network. This topic is of particular interest under conditions that Eulerian sensors are sparsely located, or flow conditions are changed rapidly, such as flood or dam break events.

5.3.2 Overview of the Drifter Data from May 9th experiment

The GPS-based drifter data was collected via the *LocationManager* service of the Android operating system [79]. Under normal conditions (i.e., clear sky and the GPS is locked onto more than four satellites), the operating system will provide the user application with a new position estimate every one second. The GPS coordinates are provided by the operating system as a pair of double-digit (64-bit floating point) values representing global latitude and longitude in decimal degrees. Immediately upon receiving these coordinates, the location is transformed into UTM coordinates. Our customized software transmits the latest GPS coordinates every five seconds, rounds them to

the nearest centimeter, and stores them on an onboard memory.

Figure 5.6 shows the GPS measurements collected by Drifter 37 during the field operational test. Figure 5.7 shows a different view of the data where only points that are near the GSS USGS station are plotted. Furthermore, in Figure 5.8, we plot the points during the second travel that Drifter 37 passes by the GSS station. The drifter data taken close to the GSS station is of interest because it can be directly compared against the velocity measurements being take by that station. Measurements taken further away could be at points where the local river velocity does not physically match the velocity of the GSS station.

It is noted that the GPS coordinates provided by the Android operating system appear to be quantized to approximately 1.88 meters in the x direction and 1.1 meters in the y direction. Evidence of this is seen by the deviations of the x and y coordinates, which seem to be discrete levels, or "quantized". A possible explanation for the y direction is that the fifth decimal place in a decimal representation of latitude corresponds to 1.11 meters for this region. Unfortunately, we were unable to find the same relationship for the precision in the x direction (the fifth decimal place for longitude is 0.88 meters). This fact is most likely due to the architecture of the GPS-capturing system of the Android phone. The discretization error can be mitigated by a filter which has low pass characteristics (such as averaging multiple samples) [138]. Therefore, the filtering step discussed in the following section is imperative to smooth out any discretization error.

The drifters transmit their data via a socket interface over the cell phone's GSM connection to our Internet server. The coordinates are both stored and transmitted as ASCII-encoded decimal numbers, along with the drifter's identifier, and a "valid" flag indicating whether the drifter is floating or in storage. The Internet server parses and stores the received data in a *SQLite* database for later retrieval. The server software also forwards the data, via a socket connection, to a computing cluster where data analysis can be performed [155].

5.3.3 Trajectory Filtering

In order to properly use the data we collected from the May 9th experiment, we need to preprocess the data using a "Pseudo-Lagrangian" method (which will be elaborated later). More specifically, when the time interval between successive position measurements is smaller than the Lagrangian integral timescale T_ℓ, the flow velocity u can be approximated as finite difference of successive positions. However, in the case of experimental flows, such as the ones we are dealing with, the estimated velocity can not be used directly in the data assimilation system, as the velocity of the drifters is perturbed by many physical processes that can not be precisely simulated in the shallow water models. In this case, one should recognize that individual trajectories are largely unpredictable, and a statistical description is preferable. However, when a large number of

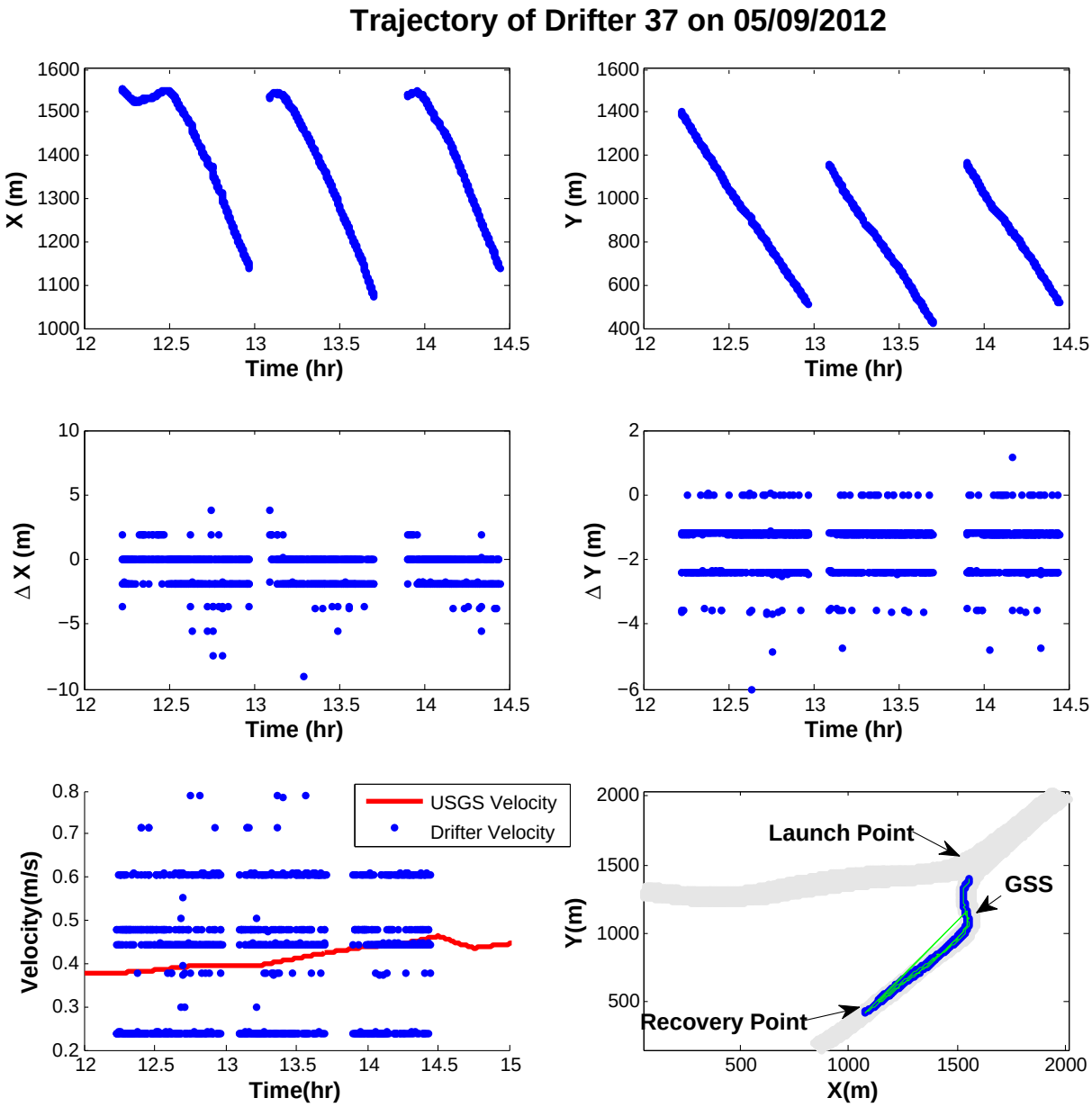


Figure 5.6: *Top row: x and y coordinates of the GPS measurements as a function of time, for drifter 37 during the field operational test. Middle row: Corresponding deviations of the x and y coordinates between successive measurements. These exhibit quantization due to the GPS-capturing system of the Android phones used. Lower row left: Corresponding velocities obtained by finite differences of successive measurements, compared against USGS measurements. Lower right: GPS measurements captured during the experiment, plotted on map of the domain, with USGS sensor station marked.*

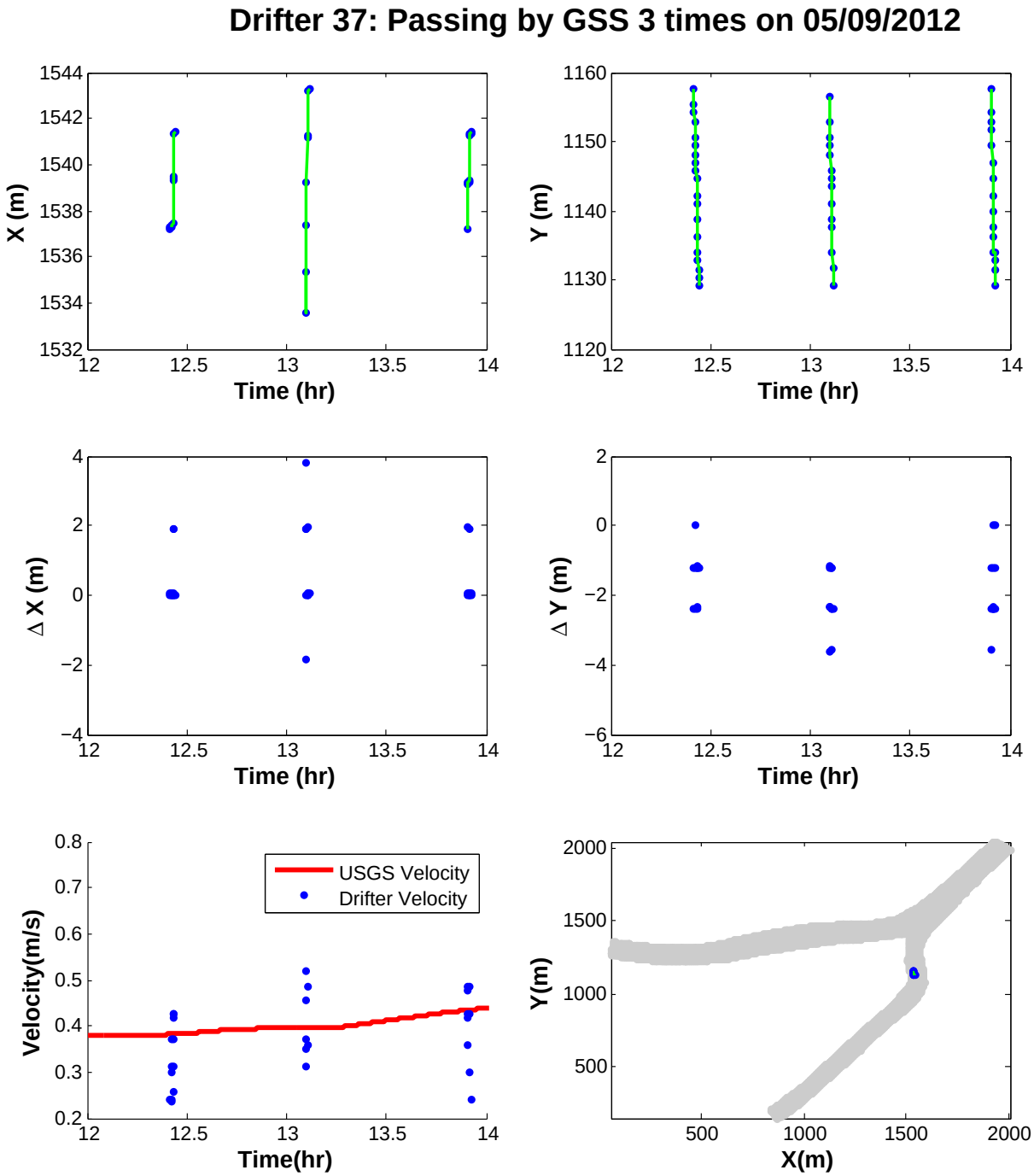


Figure 5.7: *The data of Figure 5.6, filtered to include only points in which Drifter 37 was near the USGS station. Top row: x and y tracks of the GPS measurements as a function of time. Middle row: corresponding increments in x and y between the successive GPS tracks. Lower row left: corresponding velocities obtained by finite differencing the position (compared to USGS measurements). Lower right: measurement locations for the above data.*

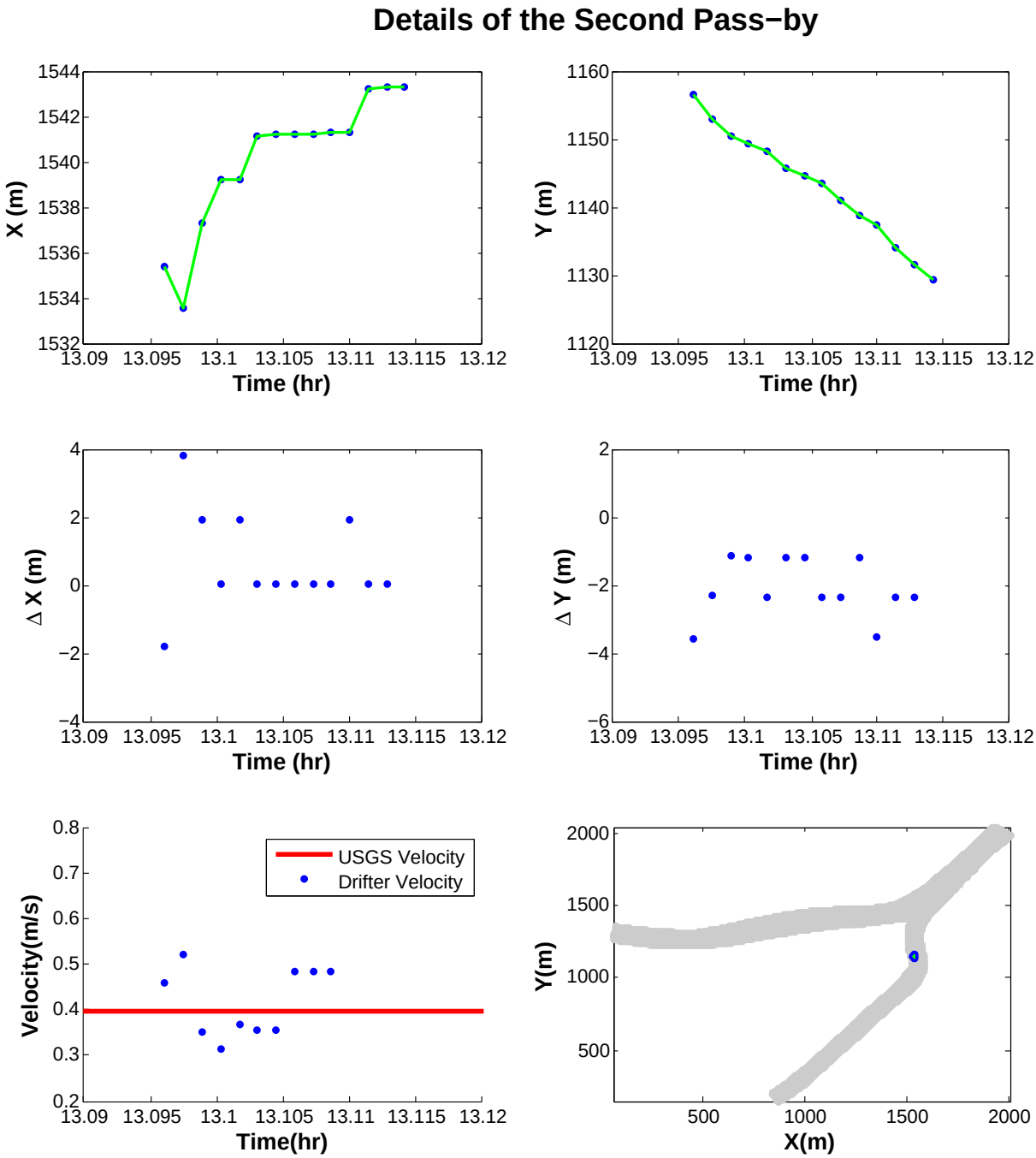


Figure 5.8: *The data of Figure 5.7, filtered to include only the second pass-by, near hour 13, of the experiment. Top row: x and y tracks of the GPS measurements as a function of time. Middle row: corresponding increments in x and y between the successive GPS tracks. Lower row left: corresponding velocities obtained by finite differencing the position (compared to USGS measurements). Lower right: measurement locations for the above data*

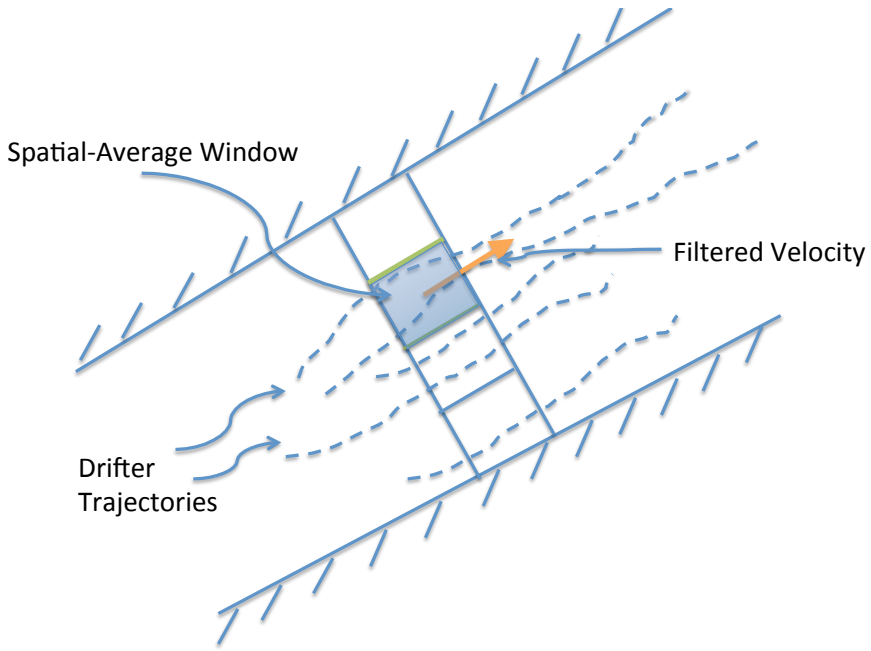


Figure 5.9: *Rectangle spatial-time window to remove perturbations from Lagrangian drifters. The shadowed box is the space-time window W, dash lines denote the drifter trajectories, and orange arrow denote the filtered velocity.*

trajectory observations are available, a space-time filter should be applied to the original data set to remove any small-scale perturbations and measurement errors.

In this section, we define a local space-time averaging filter. The filtered drifter velocity $u_{\text{filtered}}(X,t)$ is defined as the mean velocity observed at time t and location X in a space-time window $\mathcal{W} = \mathcal{W}_t \times \mathcal{W}_s$, where $\mathcal{W}_t$ denotes a temporal neighborhood of t and $\mathcal{W}_s$ denotes the spatial neighborhood of $X(t)$.

$$u_{\text{filtered}}(X,t) = \frac{1}{\mathcal{T}} \sum_{i=1}^{N_{\text{obs}}} \int_{\mathcal{W}_t} \frac{d}{dt} X_i^{\text{obs}} \mathbb{1}_{X_i^{\text{obs}} \in W_s} dt, \tag{5.1}$$

where

$$\mathcal{T} = \sum_{i=1}^{N_{\text{obs}}} \int_{\mathcal{W}_t} \mathbb{1}_{X_i^{\text{obs}} \in W_s} dt \tag{5.2}$$

It is important to select a suited bin size for the filter. On one hand, the window should be sufficiently large to ensure any small-scale perturbations being removed. On the other hand, however, it should be big enough to have the most important characteristics of the flow included. It should be noted the space-time windows may be overlapped along the flow, resulting in a smooth flow representation.

In Figure 5.9, a spatial window is drawn for Lagrangian drifters from time t^0 to time t^f. A number of drifters travel through this spatial window during the time window $\mathcal{W}_t$. The observed drifter

trajectories are represented with dotted lines. The velocities of all the drifters passing through the specific spatial window from time t^0 to time t^f are "averaged out", and the filtered velocity of these drifters is represented with a solid arrow. This filtered velocity is thus considered to represent the local flow velocity at the specific time, and later utilized in the flow computations described in subsequent sections.

5.3.4 Average Velocity across the Channel

Since our model applies one-dimensional shallow water equations, the velocity in the model system is defined as the average flow velocity across the channel, and thus the data we obtained from the previous section $u_{\text{filtered}}(X,t)$ need to be further refined.

Generally, for flows in open channels and natural rivers, actual velocities in a cross section varies from the highest value near the channel center to the lowest value near overbanks or river bottom. In a river discharge measurement protocol recommended by the U. S. Geological Survey (USGS), the mean water column velocity $u_{\text{column}}(X,t)$ in a shallow water system is determined by the average of velocities measured at the vertical locations 60% of the water depth. If the velocity profile follows the log-law-of-the-wall, the theoretical ratio of water column mean velocity over surface velocity would be 0.85. In practice, this ratio can be verified by an acoustic instrument, usually placed at fixed locations to measure the velocity profiles across the river. In previous experiments, the velocity profiles were measured with an Acoustic Doppler Current Profiler (ADCP), and the computed mean velocity to surface velocity ratio is in the range of 0.80 to 0.93 [42].

Once the water column mean velocity $u_{\text{column}}(X,t)$ is determined, the average velocity across the river cross section $\overline{u}(X,t)$ is readily evaluated by:

$$\overline{u}(X,t) = \frac{1}{\mathcal{A}} \int u_{\text{column}}(x,t) h(x,t) dx \tag{5.3}$$

where

$$\mathcal{A} = \int h(x,t) dx \tag{5.4}$$

$h(x,t)$ is the total water depth at the specific location x at time t.

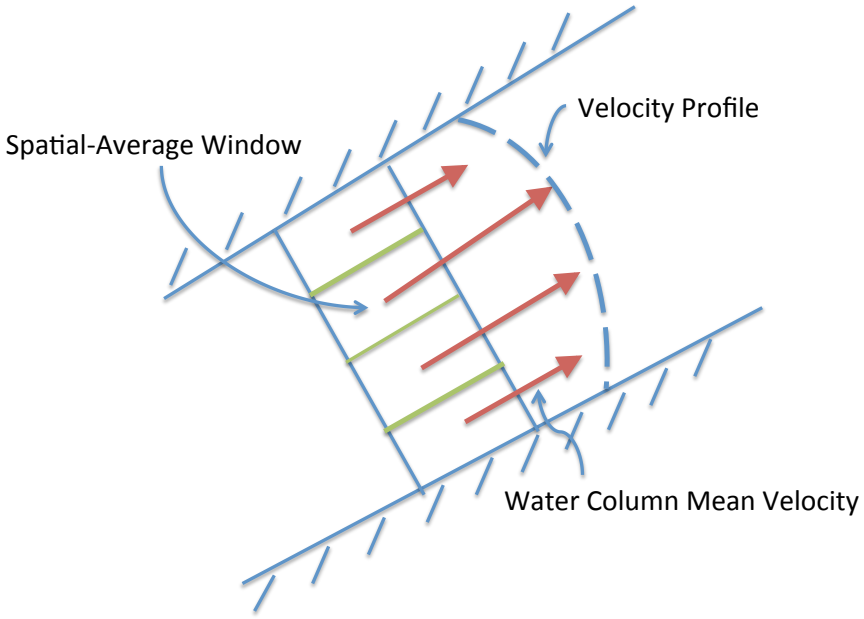


Figure 5.10: *Velocity profile in a cross-section, the average velocity across the river cross section* $\overline{u}(X,t)$ *is calculated with the water column mean velocity* $u_{column}(X,t)$ *at each cell Eq.* (5.3).

5.4 Formulation of the Data Assimilation Method

5.4.1 "Pseudo-Lagrangian" Data Assimilation

One of main issues to consider for assimilation of Lagrangian information into hydrodynamic models is to quantify the connection between Lagrangian measurements and Eulerian velocity. A simple and intuitive solution to this challenge is to approximate the Eulerian field by estimating the Eulerian velocity. Such a method is usually called "Pseudo-Lagrangian" Data Assimilation [103]. The method works well when the sampling period δt is much smaller than the Lagrangian correlation time scale T_l [63], [150].

5.4.2 Observations and Cost Function

Variational data assimilation is based on optimal control theory and consists in identifying the control vector that minimizes a cost function measuring the discrepancy between the state variable of the model and data obtained from the observation of the physical system [65], [116], [144].

In an ideal case in which both of the flow measurement devices are active and the measurement data are reliable, a cost function to minimize the difference between the most valid observation data and model state variables is constructed. The standardized framework was specified in [78] and notations are defined as follows:

- X_k: Concatenated vector of state variables (u, h) for all mesh points at time t_k.
- X_B: Background term vector.
- U_k: Vector of boundary conditions at time t_k.
- Y_k: Vector of observed variables at time t_k.
- B: Covariance matrix of the background error.
- R_k: Covariance matrix of the observation error at time t_k.
- H_k: Observation operator, which projects the state vector X_k into the observation subspace containing Y_k.

Our data assimilation strategy is to search for the initial state X_0 and boundary conditions U_k that minimize the l^2-norm of the difference between the state and observation variables and the difference between the initial state X_0 and the background term X_B:

$$\mathcal{J}(X_0, U_k) = (X_0 - X_B)^T B^{-1}(X_0 - X_B) + (Y_k - H_k[X_k])^T R_k^{-1}(Y_k - H_k[X_k]) \tag{5.5}$$

The objective function expressed in Equation (5.5) is a function of the initial state and boundary condition of the system. X_B, the background term, is a "first guess" about the state of system. The background term is usually derived from the historical data, from the model prediction, or from the assimilation result of the previous time step. The covariance matrix B and R_k represent the weight given to the background term and the observations. The observation operator, H_k, is usually non-linear in general variational data assimilation schemes. However, in a "Pseudo-Lagrangian" data assimilation process, where the observations and state variables both represent velocity, H_k will be a time-varying observation matrix. For simplicity, if we take the assimilation time step the same as the observation sampling time, H_k would be a $(0, 1)$ matrix, with element $i, j = 1$ if the drifter associated with measurement i was in the cell associated with the state variable j at time k.

The covariance matrices B and R_k essentially specify the weight given to the error terms in Equation (5.5). A reasonable choice of B and R_k is $B = b\mathbb{1}$ and $R_k = r\mathbb{1}$, where r should reflect the accuracy of the observations, and b is determined by the quality of the background term X_B.

It should be noted that the cost function used here is different from what is defined in the twin experiment [163], since the drifter data in the May experiment is acquired in a relatively short period of time (comparing with the numerical simulations in the twin experiment), and thus the time history data recorded is used directly to construct the cost function, instead of the modal decomposition in the frequency spectrum.

5.4.3 Quadratic Program Formulation of Data Assimilation

Generally, minimizing the cost function (5.5) is very costly when the direct model and the observation operator are not linear, for at least two reasons: a) each iteration of the adjoint method requires one integration of the full non-linear direct model and one integration of the adjoint linearized model; and b) the cost function is not necessarily convex, and the minimization process may converge to a local minimum, or may take considerable time to converge, or may not converge at all. The linearization of network constraints makes it possible to pose the data assimilation problem as a quadratic problem, with the positive semi-definite quadratic cost function (5.5):

$$\begin{aligned} &\text{minimize } \mathcal{J}(X_0) = \frac{1}{2} X^T P X + q^T X \\ &\text{subject to } GX \leq h \\ &\qquad\qquad AX = b \end{aligned} \tag{5.6}$$

X is the vertical concatenation of all state vectors from time t_{start} to the end of the data assimilation period t_{end}, and P and q are formed by expanding all the terms in Equation (5.5). Equation $AX = b$ represents the flow dynamics constrained by Equations (2.6)-(2.14). In order to reduce the search time, G and h are used to keep the search in a realistic set.

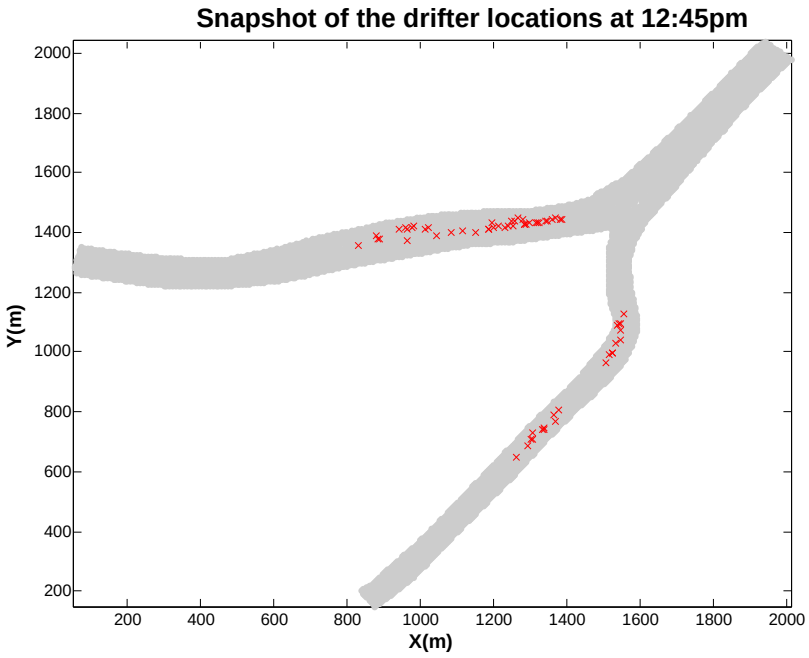
It should be noted that the computational cost of the solving the quadratic program above is very low. Furthermore, the uniqueness of an optimum X is not guaranteed, due to the rank-deficiency of the observation matrix H_k; however, all the linear subspaces from the set of optimal solutions by employing the background term X_B can be excluded [144].

5.5 Data Assimilation Results

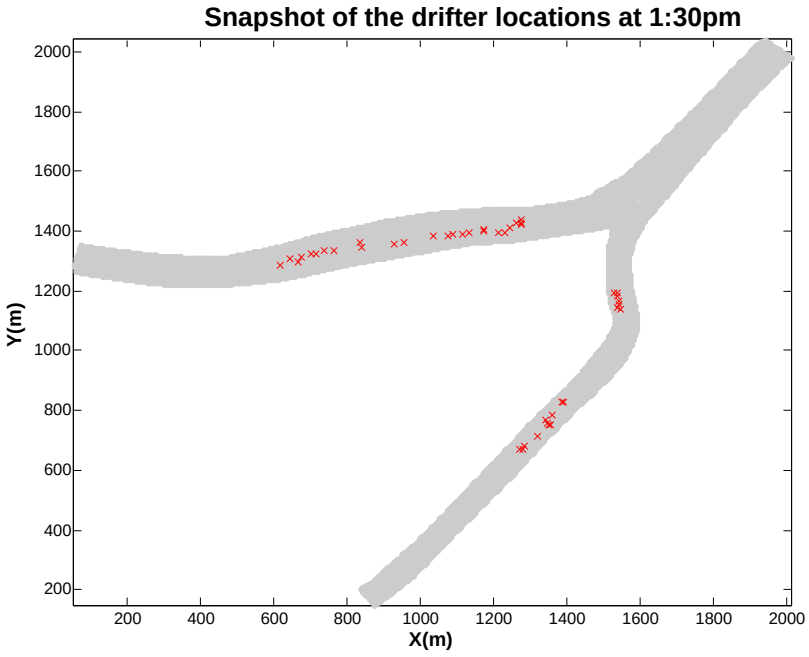
5.5.1 Overview

In this section, we present the data assimilation results of our May 9th, 2012 experiment. Figure 5.11 gives two snapshots of the Android drifters during the experiment to illustrate the coverage of the experiment.

Two sets of data are tracked in our experiment: The first one is the drifter trajectories provided by the Lagrangian drifters from the May 9th, 2012 experiment; the second set of data consists of water depth and velocity at certain locations in the experimental domain, acquired with classical USGS Eulerian sensors.



(a) 12:45



(b) 13:30

Figure 5.11: *Snapshots of the Android drifter fleet during the May 9th, 2012 experiment, at two selected time in the field operational test.*

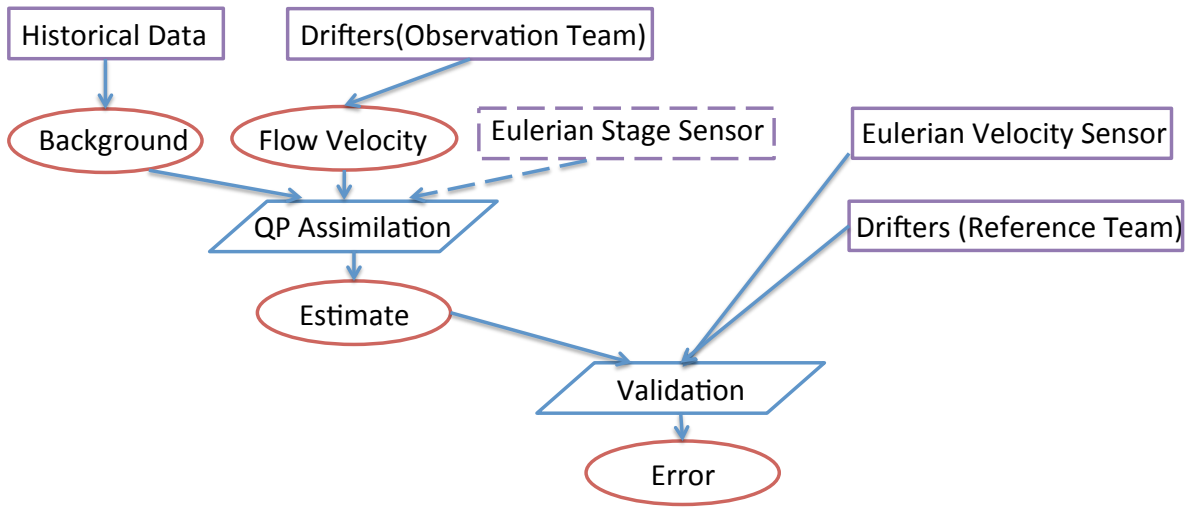


Figure 5.12: *Flow chart for the data assimilation. Drifter data is devided into two parts which include observation (used for the data assimilation) and validation.*

Figure 5.12 shows the flow chart of the data assimilation. Historical data from USGS is used to generate the background term for the QP process. The estimate of the state of the system is generated by assimilating the drifter data from the observational team either with or without the Eulerian stage information. Either the Eulerian velocity data provided by USGS or the drifter data from a Reference Group is used for validation only. The data assimilation is implemented with the optimization modeling language AMPL and solved with CPLEX.

The assimilation results validate not only the suitability of the proposed flow estimation method using Lagrangian data, but also the effectiveness of different data filtering windows. Furthermore, it is noted that the estimation quality of the flow state in the hydrodynamic system may be further improved if Eulerian information, if available, is also included in the data assimilation.

5.5.2 Assimilation with Lagrangian Data using Different Filtering Windows

The choice of the space-time window affects the results of the filtered data significantly. In our experiment, the spatial dimension of the windows is set to be 15 meters, and two different time windows of 30 seconds and 150 seconds are specifically selected. The larger filtering window (meaning more drifter information is "averaged") results in less fluctuations in the filtered Lagrangian data, and a smoother flow state. However, if the filtering window is too large, we might lose important local flow information.

Figure 5.13 and Figure 5.14 present the flow velocity near the USGS stations predicted from the flow field recovered from the assimilation results, using 30-second and 150-second filtering windows, respectively. We compare them with the velocity measured at two USGS stations for vali-

dation. Four test cases are summarized here, with respective filtering windows, using drifter data from the full or half fleet.

The plots indicate that assimilated flow velocity estimates become less "noisy" with a larger filtering window. This trend is consistent with that of the drifter data. Moreover, the signal quality is improved, and the signal magnitude is at the correct scale. We hypothesize that noise is either inherited from the Lagrangian data controlled by unknown physical processes, or affected by any erroneous drifter measurements, which can make the data assimilation process unstable.

Another interesting observation is the effect of different drifter numbers involved in the data assimilation. In general, the more drifters included in the assimilation, the less noisy and more plausible the flow estimation is. This is not surprising, since more drifters essentially carry more flow information into the assimilation process.

5.5.3 Assimilation with Both Lagrangian and Eulerian Data

We also have the local water depth measured at USGS stations included in the data assimilation, along with the filtered Lagrangian drifter trajectories.

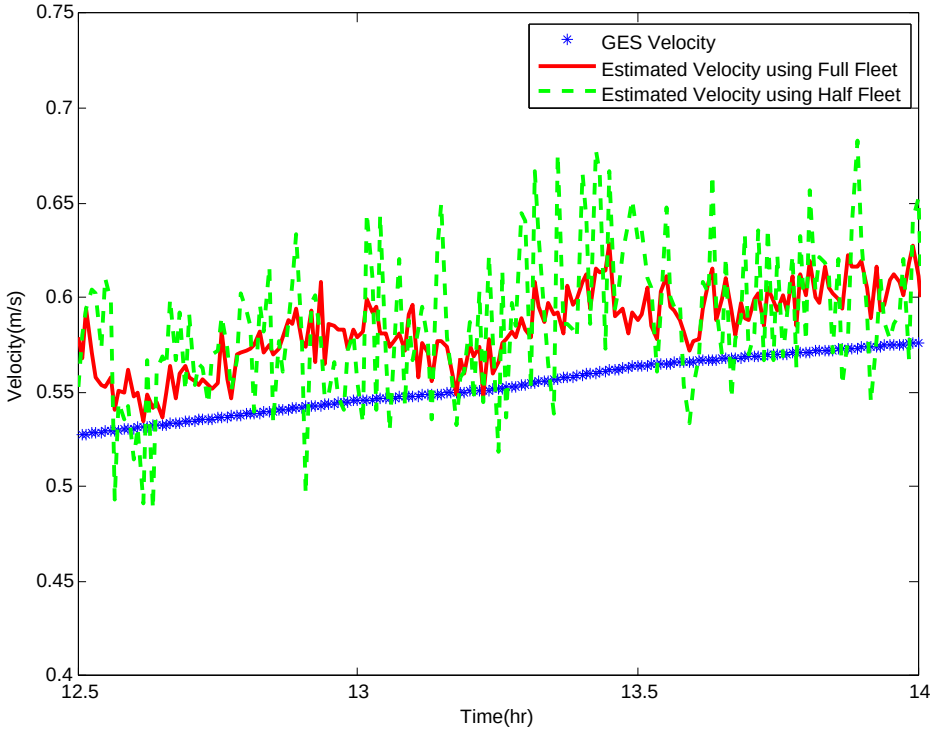
Figure 5.15 demonstrates a remarkable improvement, with respect to stability, of the flow field estimation, over the estimation results which only include drifter trajectories. The assimilated results are significantly closer to the Eulerian flow velocities at two USGS stations.

5.5.4 Validation with Reference Drifters

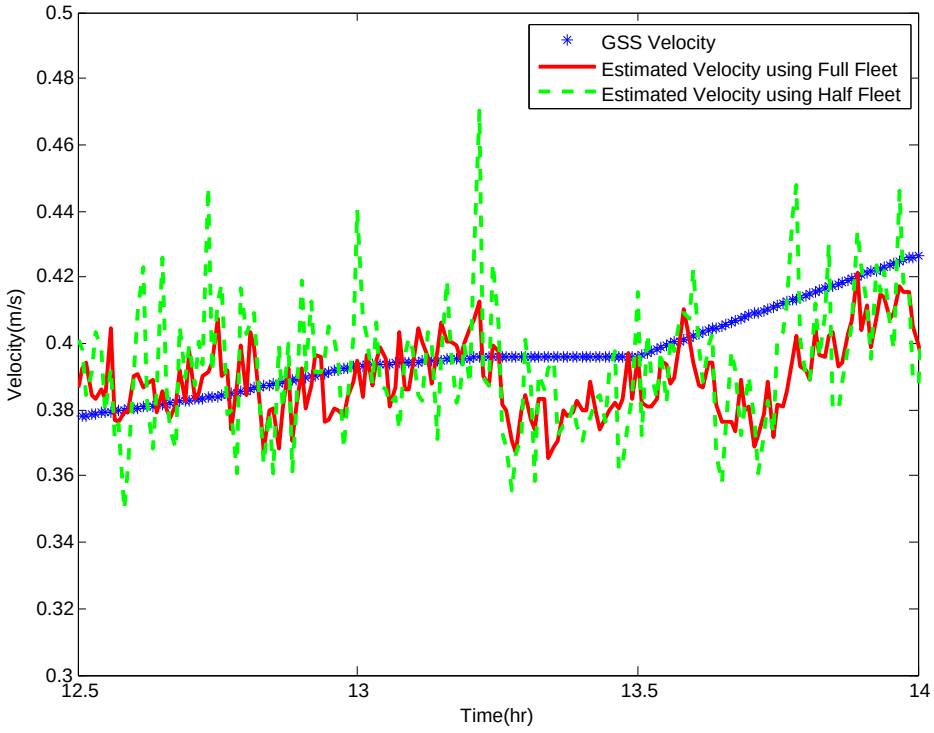
In the two previous cases, the assimilated flow velocity is correlated with the Eulerian velocity at USGS stations. It is possible, however, to validate the assimilation method with another set of Lagrangian drifter data as well, which provides an additional validation procedure, given that no data provides "ground truth".

We divide the Lagrangian drifters into two groups: one is for our regular data collection, and the flow state is essentially derived from these measurements; the other group, namely the Reference Group, is for validation purposes only, where the drifter velocity acquired from these devices is correlated with the assimilated flow results.

Figure 5.16 shows the assimilation results with the Lagrangian data inputs from half of the drifter fleet, as well as the drifter data collected from the other half fleet (i.e., the Reference Group). It shows a good agreement between the two data sets.

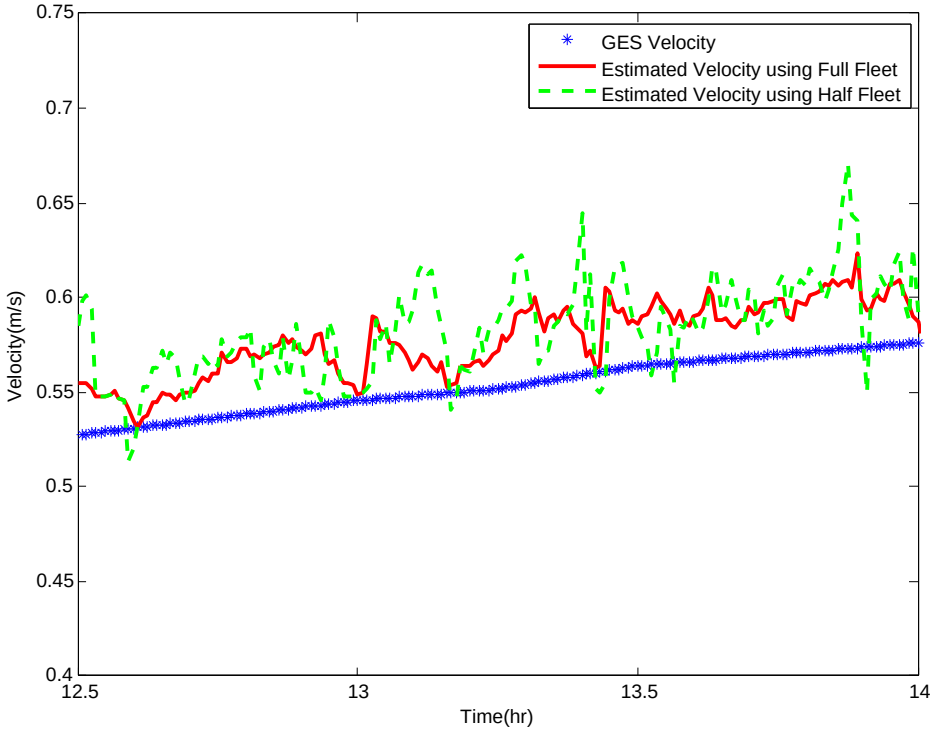


(a) Sacramento River.

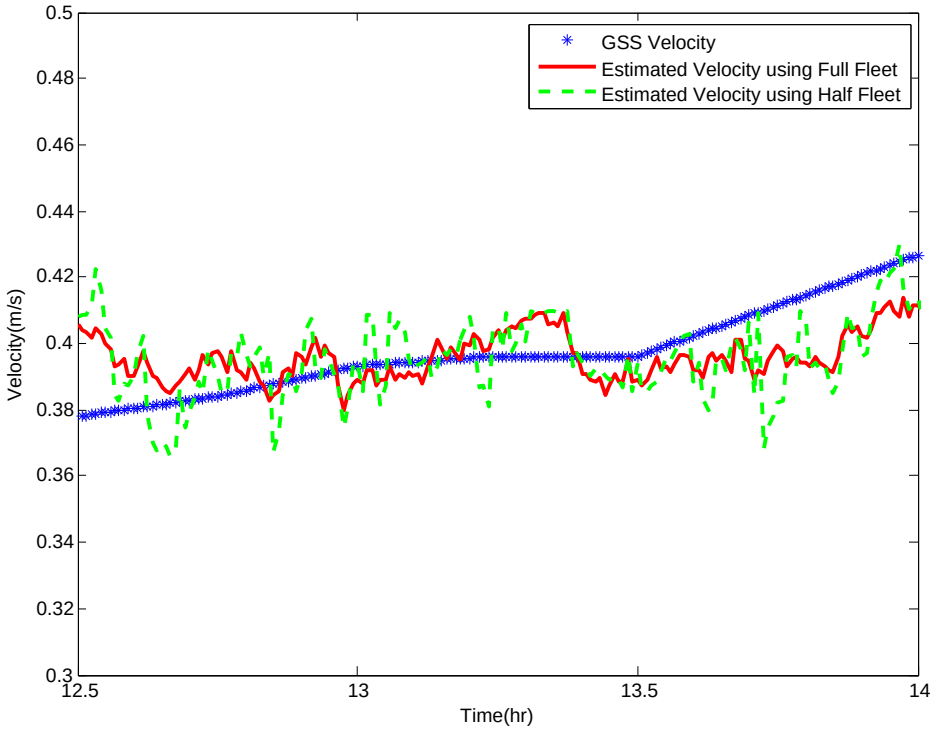


(b) Georgiana Slough.

Figure 5.13: *Assimilated flow velocity near USGS stations (a) GES (b) GSS using 30-second space-time window.*

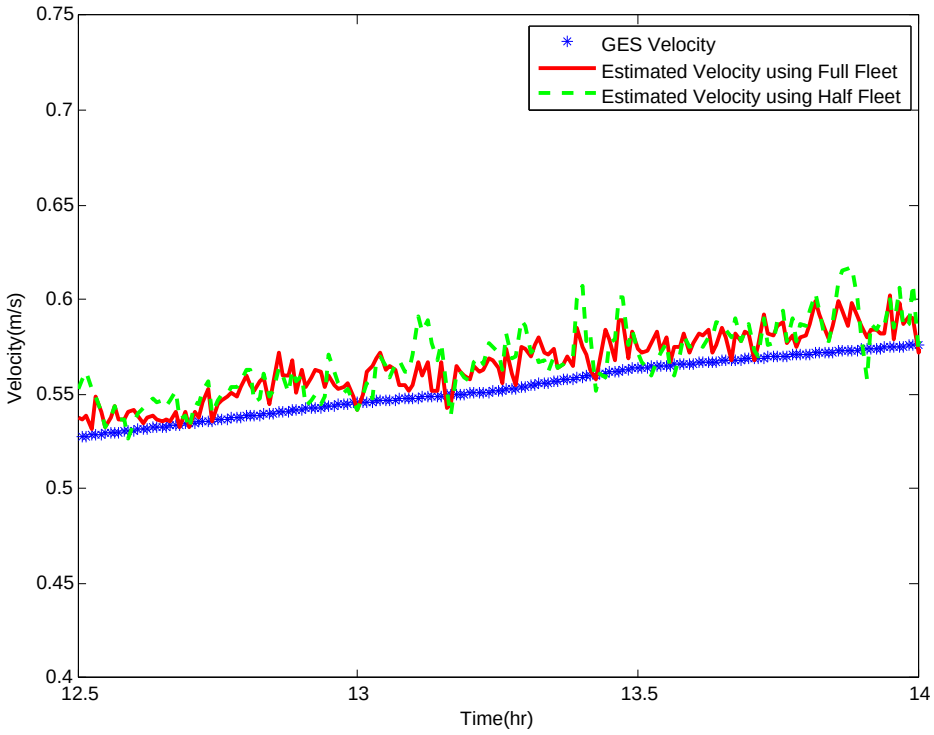


(a) Sacramento River

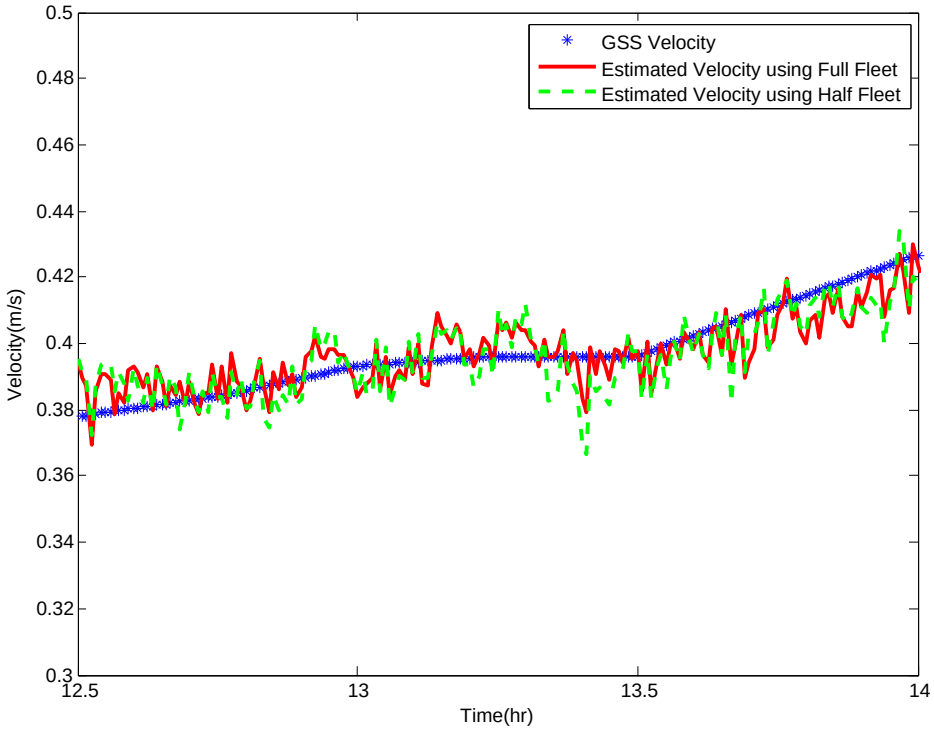


(b) Georgiana Slough

Figure 5.14: *Assimilated flow velocity near USGS stations (a) GES (b) GSS using 150-second space-time window.*



(a) Sacramento River



(b) Georgiana Slough

Figure 5.15: *Assimilated flow velocity near around USGS stations (a) GES (b) GSS using filtered drifter data from 150-second space-time window and Eulerian depth sensor.*

In addition, the assimilated results appear to have a smaller fluctuation than the measured data from the Reference Group. This is largely due to the fact that data assimilation process takes advantage of the information redundancy from inputs, and thus stabilize the flow features.

5.5.5 Discussions

The difference between the assimilated data and measurements is further evaluated in Table 5.1 by computing the relative error norm ℓ^2:

$$\ell^2 = \left(\frac{\sum_{i=1}^{N} (\hat{u}_i - u_i)^2}{\sum_{i=1}^{N} (u_i)^2} \right)^{\frac{1}{2}} \tag{5.7}$$

where u_i is the measured flow variable of interest (for example the flow velocity from Eulerian sensors in this study), $\hat{u}_i$ is the estimated flow variable, for $i = 1$ to N measurement events.

Table 5.1: ℓ^2-value for assimilated flow velocity and Eulerian sensors.

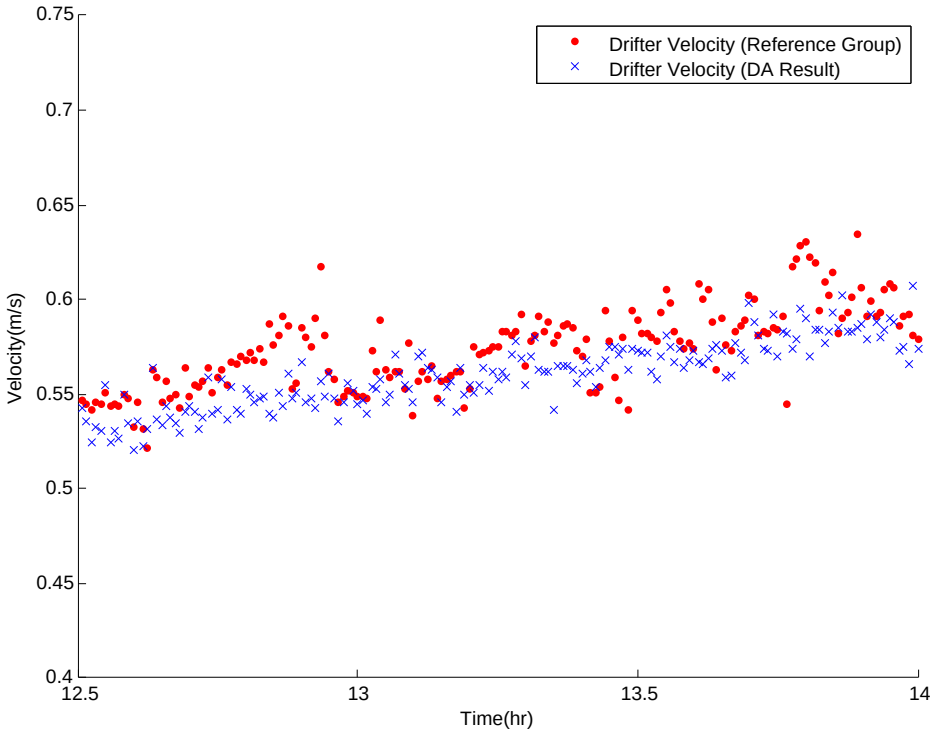
Location	Sacramento River		Georgiana Slough	
Filtering Window	Full Fleet	Half Fleet	Full Fleet	Half Fleet
30s Window (Drifter Only)	0.06	0.09	0.07	0.08
150s Window(Drifter Only)	0.05	0.06	0.06	0.07
150s window(Drifter+Stage)	0.02	0.03	0.03	0.03

Table 5.1 shows that the proposed approach possesses good flow estimation accuracy, especially when the Eulerian data are also included in the assimilation. Also, the drifter number contributes to the estimation accuracy as well. As we discussed before, more drifters essentially carry more flow information into the assimilation, and consequently improve the quality of flow estimation. Another interesting observation is related to the filtering time. The larger the time window (meaning the flow is observed in a coarser timeframe), the flow state exhibits more steady, and thus, not surprisingly, the flow estimation becomes more accurate.

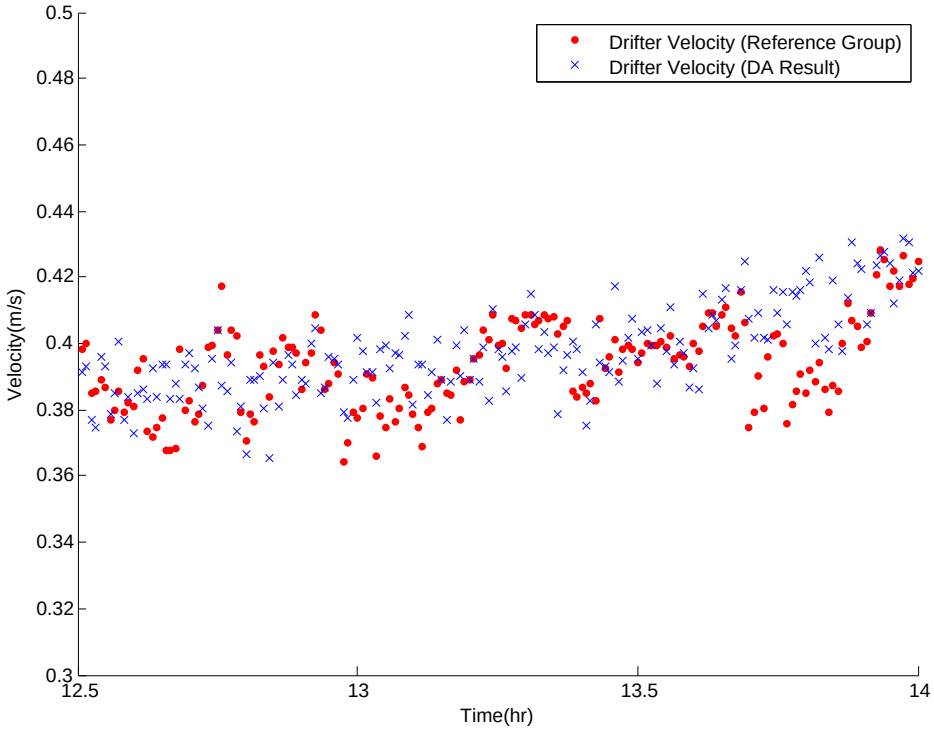
It is also noted that the mean values of the estimated velocity of Sacramento River in Figure 5.13, Figure 5.14, and Figure 5.15 are a bit off from the USGS measurements. This is likely due to the system error from GES measurements, or some consistent wind effect in the Sacramento River.

5.6 Summary

In this chapter, we presented the estimation of initial condition and boundary condition for complex channel networks using Lagrangian measurement data. This is the first successful application



(a) Sacramento River



(b) Georgiana Slough

Figure 5.16: *Assimilated flow velocity at the locations of the reference drifters using filtered drifter data from 150-second space-time window and Eulerian depth sensor.*

of variational Lagrangian data assimilation in the hydrodynamic system. The solution is formulated as a QP problem based on minimizing the difference between measured Lagrangian data and modeled drifter trajectories, constrained by a 1D implicit linear channel network model. The major advantage of the 1D QP formulation is that it requires low computational cost, making the method applicable to many vast and complex hydraulic networks. The effectiveness of the method has been validated with a field experiment in which the drifter data collected from GPS equipped drifters deployed in the Sacramento-San Joaquin Delta.

In the next chapter, we will extend the data assimilation techniques in a 2D SWE model (e.g., REALM), and apply statistical filtering to achieve a more accurate representation of flow state in tidal channel network. More experimental data from the Delta will be collected and processed to validate the approach.

Chapter 6

Real-Time Lagrangian Data Assimilation using EnKF in 2D SWE

6.1 Introduction

The ultimate research goal of the dissertation and work in general is to design and implement an effective real-time flow estimation system using the floating sensors.

In this chapter, Lagrangian flow data, acquired in terms of GPS coordinates from floating sensors, is further incorporated into a 2D flow model. Compared to 1D shallow water models used in the chapters 4-5, 2D models have their intrinsic merits of better describing the flow states and evolution. We choose a two-dimensional shallow water equation model whose initial implementation simulates flows in the entire San Francisco Bay and the Sacramento-San Joaquin Delta.

An ensemble Kalman filter (EnKF) technique is subsequently used here to integrate the uncertainties in the full nonlinear 2D model. We employ parallel computation and adaptive mesh refinement for rapid computation.

The present work focuses on the state estimation problem in river hydraulics [161]. The area of interest is a section of the Sacramento River near Walnut Grove, California. More specifically, Lagrangian measurement data from drifters is assimilated into a two-dimensional shallow water model with uncertain upstream and downstream boundary conditions using an ensemble Kalman filter (EnKF).

6.2 REALM Model

In this section, the governing equations for the river flow and the Lagrangian drifters are reviewed. The state-space model used in this work is also presented.

6.2.1 Introduction to REALM

REALM (River, Estuary, And Land Model) [11], [137] is a reliable, extensible open-source hydraulic model, employing high performance numerical algorithms to address difficult hydrodynamic, water quality, and biological questions arised from water management in the San Francisco Bay-Delta. The development of REALM is a collaboration between the California Department of Water Resources (DWR) and Lawrence Berkeley National Lab (LBNL).

This model is based on on a Cartesian grid, and features embedded boundary discretizations. The underlying shallow water equations in REALM are derived from the depth-integrated Navier-Stokes equations [53], [100], with a hydrostatic treatment of pressure, and a Boussinesq assumption [62] of baroclinic (salt-induced) density variation. The equations consist of the conservation of mass and momentum of water, as formulated in the following section.

6.2.2 2D Shallow Water Equations

The domain of interest is an estuary system consisting of bays and channels. At the present stage, the shoreline and the channel bottom are assumed to be time-independent. (Ultimately, the design anticipates a shoreline that evolves in time due to tides and flooding.) The water column is defined by a rigid boundary of known bathymetry on the bottom and a free surface on the surface. Appropriate boundary treatments make it feasible to represent a three-dimensional geometry in a two-dimensional flow model.

Conservation Equation

The conservation equation can be expressed in terms of the height of water column h, local velocities u and v, and salt concentration s as:

$$\frac{\partial U}{\partial t} + \frac{\partial F^x + F_d^x}{\partial x} + \frac{\partial F^y + F_d^y}{\partial y} = S \tag{6.1}$$

where the vector of conserved variables:

$$U = (h \quad hu \quad hv \quad hs)^T \tag{6.2}$$

represents the water mass height h, momentum hu and hv and mass of salt per unit area hs. The flux across cell faces in the x- and y-directions include the convective acceleration and gravity wave terms:

$$F^x = (hu \quad hu^2 + \frac{g\rho h^2}{2\rho_0} \quad huv \quad hus)^T \tag{6.3}$$

$$F^y = (hv \quad huv \quad hv^2 + \frac{g\rho h^2}{2\rho_0} \quad hvs)^T \tag{6.4}$$

In these equations, g denotes the gravitational constant, ρ_0 denotes the density of fresh water, and

$$\rho = \rho(s(x, y, t)) \tag{6.5}$$

denotes the density of salt water. The diffusive terms F_d^x and F_d^y include momentum diffusion due to eddy diffusivity and a more general dispersion term for salt:

$$F_d^x = (0 \quad -\epsilon h\frac{\partial u}{\partial x} \quad -\epsilon h\frac{\partial v}{\partial x} \quad -h\frac{\partial s}{\partial x})^T \tag{6.6}$$

$$F_d^y = (0 \quad -\epsilon h\frac{\partial u}{\partial y} \quad -\epsilon h\frac{\partial v}{\partial y} \quad -h\frac{\partial s}{\partial y})^T \tag{6.7}$$

The sources and sinks considered here include the pressure component due to changes in bottom elevation and friction:

$$S = (0 \quad -\frac{g\rho}{\rho_0}hb_x - \tau_x \quad -\frac{g\rho}{\rho_0}hb_y - \tau_y \quad 0)^T \tag{6.8}$$

where $b(x, y)$ is the elevation of the bed. τ_x and τ_y are components of a bottom stress, given by a zero-order closure relation using the Manning′ s friction coefficient:

$$\tau_x = \frac{\rho g h n^2}{\rho_0 h^{\frac{4}{3}}} u\sqrt{u^2 + v^2} \tag{6.9}$$

$$\tau_y = \frac{\rho g h n^2}{\rho_0 h^{\frac{4}{3}}} v\sqrt{u^2 + v^2} \tag{6.10}$$

Discretization

A finite volume discretization based on a Cartesian grid has been applied to the shallow water equations. A primary new feature of the spatial discretization is the calculation and interpolation of cell and face-averaged quantities based on the spatial variation in bathymetry. In this way, a three dimensional geometry can be be represented in a 2D domain. The computational scheme is espically useful when dealing with moving shorelines or for a mixed one, two, or three dimensional model.

A finite volume predictor-corrector scheme has been chosen for time update, which constructs accurate, upwinded estimates of the fluxes on cell faces and then updates cell average values accordingly. The solution algorithm was first used in [34], which was based on earlier work [33].

6.3 Ensemble Kalman Filter

6.3.1 Kalman Filter and Extended Kalman Filter

The goal of data assimilation is to determine a state estimate of the system that best fits a set of noisy observations and the underline models [131]. The standardized framework was specified in [88] and notations are defined as follows:

- x_k: m- dimensional vector representing the state of the system at a given time t_k, the components of x are the state variables at the different grid points.
- x_k^b: Background state of the system at a given time t_k.
- x_k^a: Analysis state of the system at a given time t_k.
- y_k: Vector of observations collected at time t_k.
- H_k: Observation operator , which defines the functional relation between x_k and y_k.
- R_k: Covariance matrix of observation error at time t_k.
- P_{k-1}^a: Analysis error covariance matrix.
- P_k^b : Background error covariance matrix.

For a time series of observations, we assume that the observations depend on x_k in a known way:

$$y_k = H_k x_k + \upsilon_k \tag{6.11}$$

where, υ_k is a Gaussian random variable with mean 0 and covariance matrix represented as R_k. When the dynamics and the observation operator are linear, the analysis state x_k^a can be obtained by finding the minimum of the cost function defined as follows:

$$\begin{aligned}\mathcal{J}(x_k) =&(x - x_k^b)^T (P_k^b)^{-1}(x - x_k^b) \\ &+ (y_k - H_k x_k)^T R_k^{-1}(y_k - H_k x_k)\end{aligned} \tag{6.12}$$

where the first term includes all the information collected up to time t_{k-1}, and the second term reflects the effects of observations collected at t_k. The Kalman filter method solves this least squares problem sequentially, here are the steps:

- Calculate the background state x_k^b by propagating the analysis state x_{k-1}^a from t_{k-1} to t_k using the dynamics:

$$x_k^b = M_{k-1,k} x_{k-1}^a \tag{6.13}$$

  where $M_{k-1,k}$ is the (linear) operator of the dynamics.

- Calculate the background error covariance matrix P_k^b by propagating the estimate of the analysis error covariance matrix P_{k-1}^a from t_{k-1} to t_k:

$$P_k^b = M_{k-1,k} P_{k-1}^a M_{k-1,k}^T \tag{6.14}$$

- The analysis error covariance matrix P_k^a is given by

$$P_k^a = (I - K_k H_k) P_k^b \tag{6.15}$$

  where the Kalman gain matrix K is defined by

$$K_k = P_k^b H_k^T (H_k P_k^b H_k^T + R_k)^{-1} \tag{6.16}$$

- The state estimate x_k^a is obtained by

$$x_k^a = x_k^b + K_k(y_k - H_k x_k^b) \tag{6.17}$$

In general, $M_{k-1,k}$ does not provide a perfect representation of the true dynamics. Thus, the effect of model errors is often quantified by adding a model error covariance matrix Q to the right side of (6.14). The model errors introduced during the forecast steps would contribute to the overall uncertainty in the background x_k^b.

Extended Kalman Filter methods [149] further expand the applicability of the Kalman filter to nonlinear systems, by using the nonlinear model in (6.13). This approach assumes that the overall

uncertainties in the state estimates are small, and consequently their evolution can be approximated by the linearized dynamics, although the model dynamics itself is nonlinear. It has been proven, however, that forecasting a high-dimensional nonlinear system using the tangent linear model can be very expensive in computational costs and may lead to unbounded linear instability of the filter [56].

6.3.2 Ensemble Kalman Filter

Variants of the ensemble Kalman filter (EnKF), initially proposed by Evensen [49], are now becoming the mainstream data assimilation techniques in numerous engineering fields. They have been widely used in atmospheric and oceanic sciences [108], [68], [51], in hydrological applications [128], [153], [46], [165], [29], and in other engineering applications [158], [13], [32].

The EnKF approaches improve Kalman filtering by replacing (6.14) with a much cheaper approach for the calculation of P_k^b: at time t_k, a n-member ensemble of initial conditions, $x_k^{a(i)}, i = 1, 2, ..., n$, is selected around the ensemble mean $\bar{x}_{k-1}^a$; then the members of the ensemble are propagated using the nonlinear model to generate a background ensemble $x_k^{b(i)}, i = 1, 2, ..., n$. Typically, the state space dimension m is orders of magnitude, which is much larger than the ensemble size n. Thus the EnKF estimates of $\bar{x}_k^b$ and P_k^b are calculated as:

$$\bar{x}^b = n^{-1} \sum_{i=1} x^{b(i)} \tag{6.18}$$

$$P^b = (n-1)^{-1} \sum_{i=1}^{n} (x^{b(i)} - \bar{x}^b)(x^{b(i)} - \bar{x}^b)^T \tag{6.19}$$

Although the EnKF significantly reduces the computational costs by implementing the smaller number of model integrations, there are a number of issues associated with any EnKF scheme:

- The rank of P^b in Kalman filter is m, while the rank of its estimate in EnKF method is $n-1$. This could be a problem when solving (6.15) and (6.16);
- EnKF is more sensitive to model errors than 3D-VAR and 4D-VAR, because the error statistics is propagated through many analysis cycles;
- The actual uncertainty of P^b is typically underestimated.

6.4 Application to Experimental Data

Numerous studies to date have tested EnKF systems under perfect model assumption with simulated observations. It is only within the last decade that the EnKF assimilation has been tested with experimental atmospheric observations [73], [156], [142], [146], [101]. Since 2005, an EnKF method has been operational in the Canadian Meteorological Centre [75]. When using the same model and observations, the performances of EnKF are about the same as the performances of a four-dimensional variational data assimilation (4DVAR) system.

The main difficulties in the applications of experimental data are the noisy observations, imperfect model, and system stability. More specifically,several issues need to be addressed to ensure a data assimilation process can run stably, such as:

- Uncertainties of observations.
- Differences between the forecast model and the real-world model.
- Imperfect forward observation operator.

6.4.1 Model Error

It is critical for a successful implementation of EnKF to account for model error in an appropriate manner [40], [43]. Neglecting model error in the EnKF would lead to ensemble members spread too small [74], [80].

Numerical model errors may result from physical and numerical approximations, e.g., approximate parameterizations of physical processes, numerical discretization, the failure to represent subgrid-scale events, etc. As an example, model error occurs when a three-dimensional system is approximated in a two-dimensional model. This would lead to discrepancies between the model and the reality, and adversely impact the performance of the estimated system. When preparing a data assimilation scheme for a non-perfect model and real observations, it would be overly optimistic to assume a perfect physical model [43]. In this case, the "true" forecast error covariance must be considered to reflect uncertainties.

A systematic Bayesian approach to recover from the modeling errors was proposed in [82] for stationary estimation problems, and in [76], [77] for non-stationary problems. This is called the (Bayesian) approximation error approach [115], which is based on approximating modeling errors with additive Gaussian noise processes.

Extending the work in [147], we explore the approach for characterizing the model error in the river

flow in channel networks on the basis of the Bayesian approximation error theory. The versatility of the theory enables us to model the error caused by numerous uncertainty sources simultaneously.

6.4.2 Instability

With a good estimation of model error, we can assume a suited physical model available in assimilation. In practice, however, an ensemble of vectors not globally orthogonalized has its tendency to collapse toward a small subspace. As a result, even for a perfect model, the background error covariance tends to be underestimated. These effects tend to underestimate the forecast error, and therefore give too little weight to the observations, which can subsequently lead to the divergence of the filter [5]. To avoid this problem, multiplicative inflations and additive random perturbations [35], [90] are usually used in order to separate and enlarge the ensemble perturbations.

Multiplicative Inflation

One common challenge for EnKF is the insufficient prior variance due to small ensembles, model error, and other algorithmic deficiencies. To address it, the amplitude of each perturbation vector is multipled by an artifical factor larger than 1. In this way, no direction change would be applied at the analysis step, but the amplitude of the forecast perturbation vectors are sufficiently large [6]. This is achieved by using multiplicative inflation of the prior ensemble [7], where the model forecast ensemble is adjusted:

$$X_{m,n}^{inf} = \sqrt{\lambda}(x_{m,n} - \bar{x}_m) + \bar{x}_m \tag{6.20}$$

where λ is called a covariance inflation factor [89].

Additive Random Perturbations

As an alternate solution to the rank-deficiency problem that causes filter divergence, introducing perturbations in new random directions would increase the local ensemble dimension [119], [117], enabling us to "refers" the ensemble at every analysis step [67], [36]. This method effectively avoids the excessive convergence of bred vectors into a small dimensional space [154].

6.5 Implementation of the Floating Sensor Network System

6.5.1 Communication Architecture

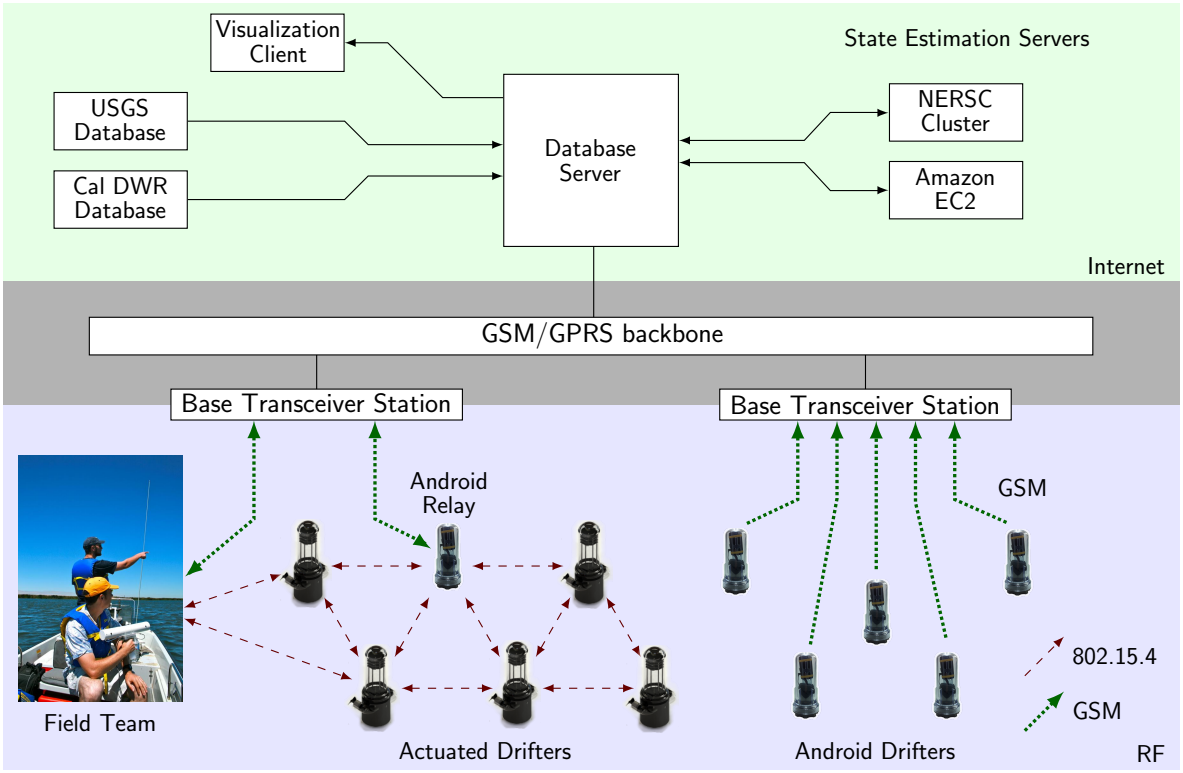


Figure 6.1: *Communication architecture, showing the flow of data from drifters in the field to the database server and computation servers via the GSM service.*

Figure 6.1 shows the communication links between various elements of the system. Data collected by the active and passive drifters is sent back to the database server using the *General Packet Radio Service* (GPRS) of GSM. The Android smartphone on board each passive sensor provides the necessary GPRS functionality. A Digi XBee-PRO 802.15.4 module on each active drifter enables short-range communication with other drifters and the field team. The XBee-PRO module conforms to the IEEE 802.15.4–2006 draft standard for low-power mesh networking. Our experience in outdoor environments shows that point-to-point links of 100 m are reliable, and we have seen connectivity at distances of 1 km. In order to bridge between the short-range networking and the database servers, we built 10 specialized Android drifters carrying a XBee-PRO module as well as an Android smartphone. These devices, called "*Relays*", were put in static locations around the experimental environment. They did not gather data themselves, but simply collected the data from the active drifters and transmitted it to the database server via GSM.

The database server acts as a central repository for gathered data and assimilation results. In addition to the Lagrangian data collected by the drifters, relevant data from USGS and California

Department of Water Resources (DWR) sensor stations is collected and stored. The sensor data is sent to the computational cluster at NERSC. Results from the assimilation process are stored on the same database server, and queried by the visualization application. In this way, we have successfully put the flow "on-line".

6.5.2 High Performance Computing

Implementing the proposed Ensemble Kalman Filter scheme with the REALM model requires a set of interface routines that exchange information between the model and database, as shown in Figure 6.2. The system starts from an initial ensemble of states that capture the uncertainties of the initial state estimate. The REALM features an underlying two-dimensional hydraulic model. Given one set of observations and an ensemble of initial states, the system assimilates the observations, updates the states, and determines how far to advance the REALM to accommodate the next set of observations. The REALM, used as the forecast model, then advances each ensemble member (either in turn or all-at-once) to the time of the next available observations. This forecast-update cycle continues as long as there are more observations to assimilate or the name list control is not met. After the assimilation processes are completed, a set of restart files (suitable to continue an experiment with more observations) and diagnostic files are generated and stored. The REALM/EnKF assimilation system is now enabled for multivariate assimilation of several data sets. This includes water depth and velocity measured from USGS stations, drifter trajectories measured by drifters developed at UC Berkeley.

A computer cluster used in HPC consists of a group of computers linked to each other in order to behave as a single system. The computers are interconnected via a fast network and each one contains homogeneous hardware and software. The main objective of HPC clusters is to use the processing power of multiple nodes in parallel. This parallelization requires a communication between the nodes while processing the tasks if the tasks are dependent.

These computations will be done using real time data acquired by the drifters in the water. The new real-time measurements provided by the drifters will be assimilated with the outputs of each flow model running on the client nodes. Once the data assimilation is done, each model's state is updated and sent back to its corresponding client to start a new cycle. The network communication is managed using a Java application. Figure 6.3 shows the data flow in the system.

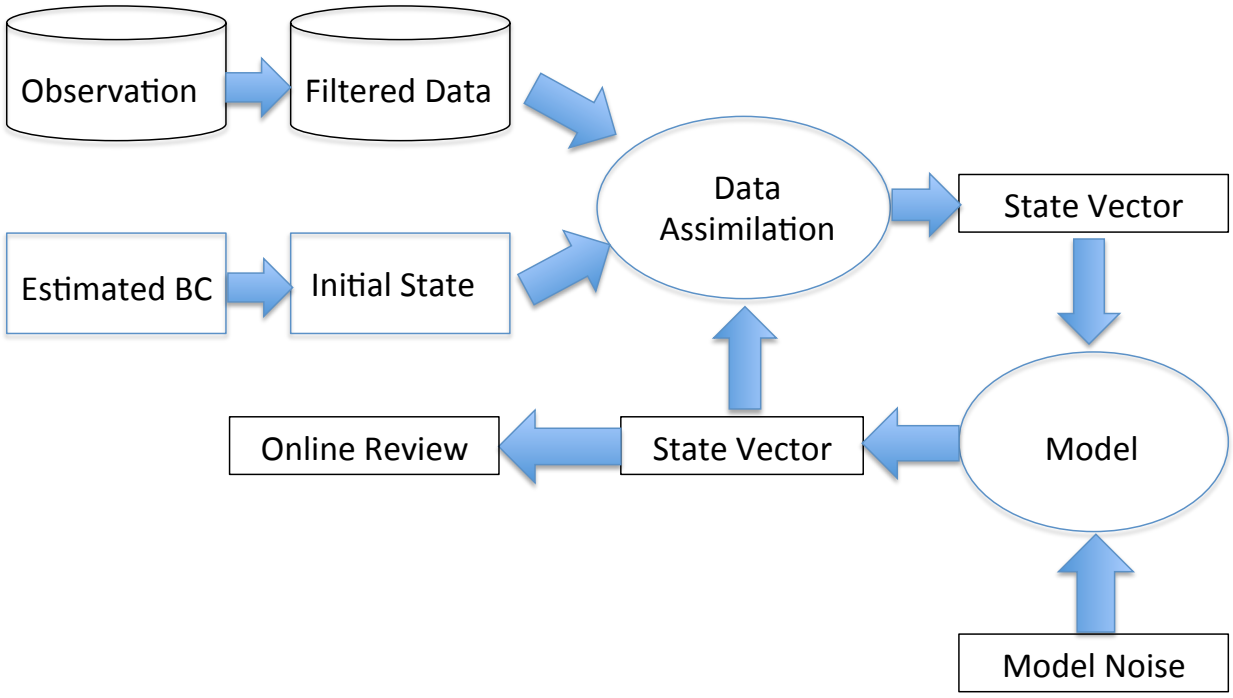


Figure 6.2: *A schematic description of the REALM/EnKF assimilation system. Observations are filtered, an ensemble of initial states are created. The states are converted to the form required by the REALM and advanced to the required time. Those states are converted to a form required by EnKF and the process repeats. When there are no more observations is met, a set of restart files and a set of diagnostic files are generated.*

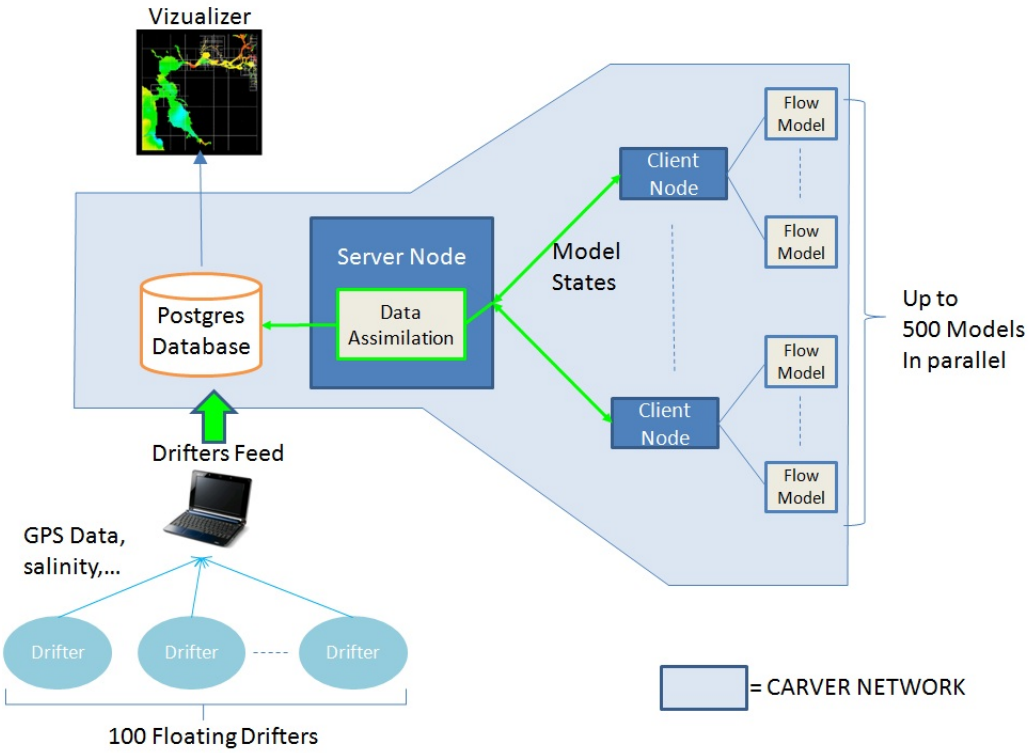


Figure 6.3: *Dataflow from drifters in the field to the Carver cluster at NERSC.*

6.6 Data Assimilation and Results

Assimilation experiments were conducted for 2 hours, using the flow data collected during the May 9, 2012 experiment. The EnKF was implemented with different number of ensemble members and inflation factors. We tested the system under different data configuration and model setups.

It should be noted that there were no USGS Eulerian measurements available that can be used as boundary conditions in the experimental domain. Therefore, we use the data reconciliation method developed in Chapter 3 to get an estimation of discharges, and pose it to be the first guess of the data assimilation process in this chapter.

The results of these experiments are presented and discussed below.

6.6.1 Estimated State

The forecasts and analysis from the assimilation experiments, using 100 ensemble members and an inflation factor of 1.2, are shown in Figure 6.4.

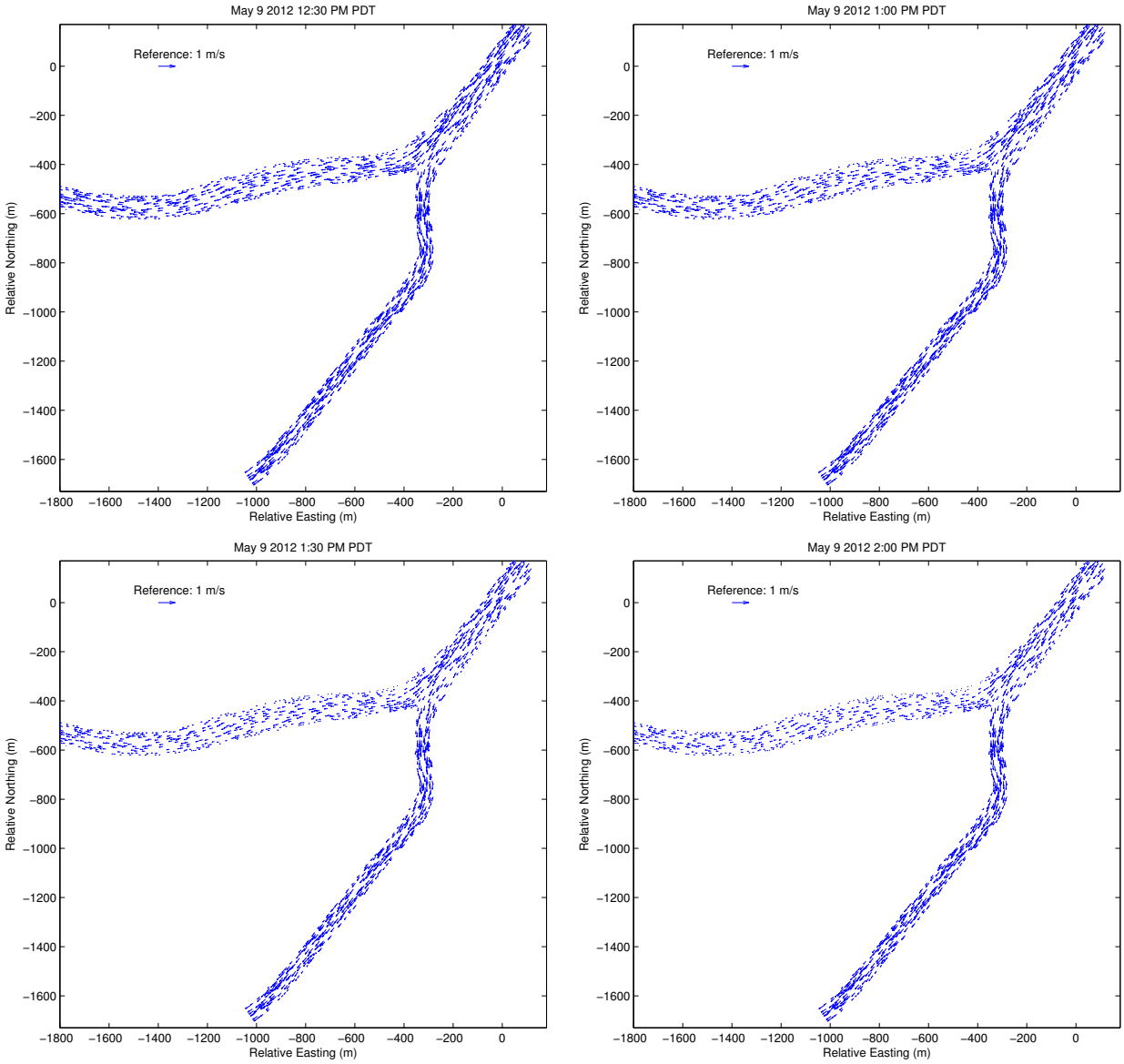


Figure 6.4: *Quiver plots of assimilation results for flow fields during May 9 experiment.*

6.6.2 Sensitivity Studies

The sensitivity of the assimilation system was studied by running the system with various inflation factors and ensemble sizes. The performance of the assimilation is mainly assessed by the RMS error between in-situ observations and model forecasts. As shown in Figure 6.5, the sensitivity of the assimilation system to the inflation factor was examined by testing three values: 1.1, 1.2, and 1.3. The results suggest that inflation plays an important role in the performance of the assimilation system. Increasing inflation from 1.1 to 1.2 improves the accuracy of the estimates, as inflation affects the spread of the ensemble, and accounts for uncertainties that could not be specified in the filter covariance [72]. However, increasing the inflation factor to 1.3 caused some of the ensemble members to diverge during their forward integration with the model.

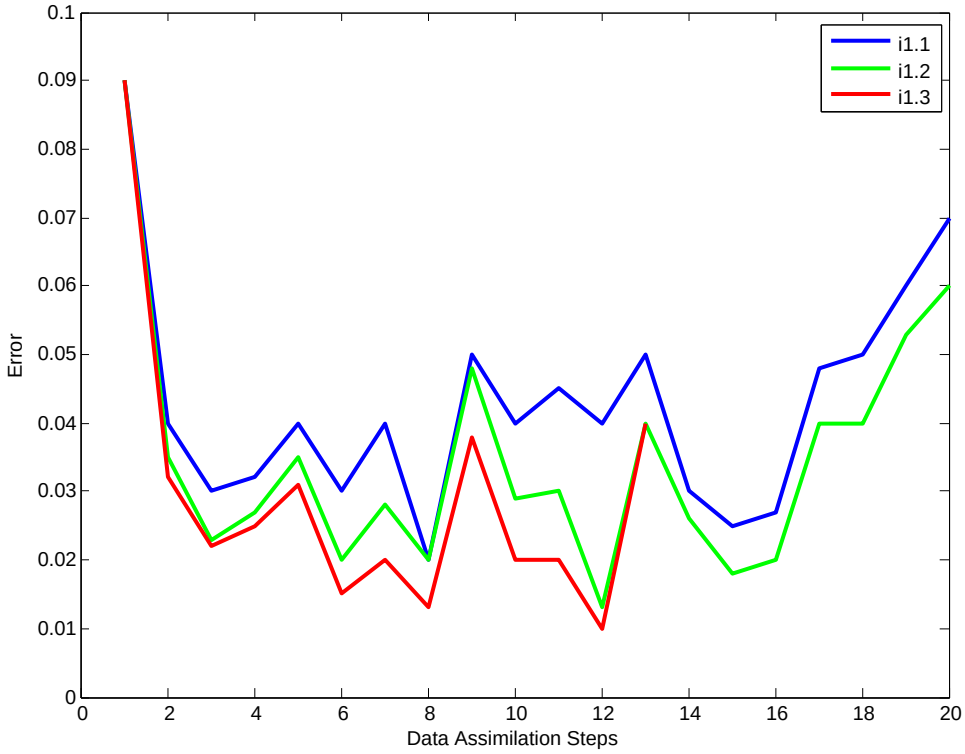


Figure 6.5: *The impact of inflation factor with an ensemble size of 50. Root mean square (RMS) between in-situ measurement and the assimilation solution. In the legends, "$i1.1$", "$i1.2$", and "$i1.3$" refer to the inflation factors of* 1.1,1.2, *and* 1.3, *respectively.*

The choice of the ensemble size must be large enough to account for the mean and spread of the prior distribution, and allow an accurate representation of the covariance between the observation and the prior state.

To study the sensitivity of the EnKF assimilation system to the ensemble size, the EnKF sheme was run with 50, 100, and 200 members. The root-mean-square (RMS) error between observations

and model forecasts are shown in Figure 6.6. In these runs, the inflation factor is 1.2.

In general, increasing the ensemble size tends to decrease the RMS error between model forecasts and in-situ observations. Estimation accuracy is clearly improved when the ensemble size increases from 50 to 100, although the solution obtained with 50 members is quite acceptable. Although increasing the ensemble size from 100 to 200 can also improve the results to some extent, the computational cost would be doubled to achieve this gain, and thus it is usually not recommended in any on-line data assimilations.

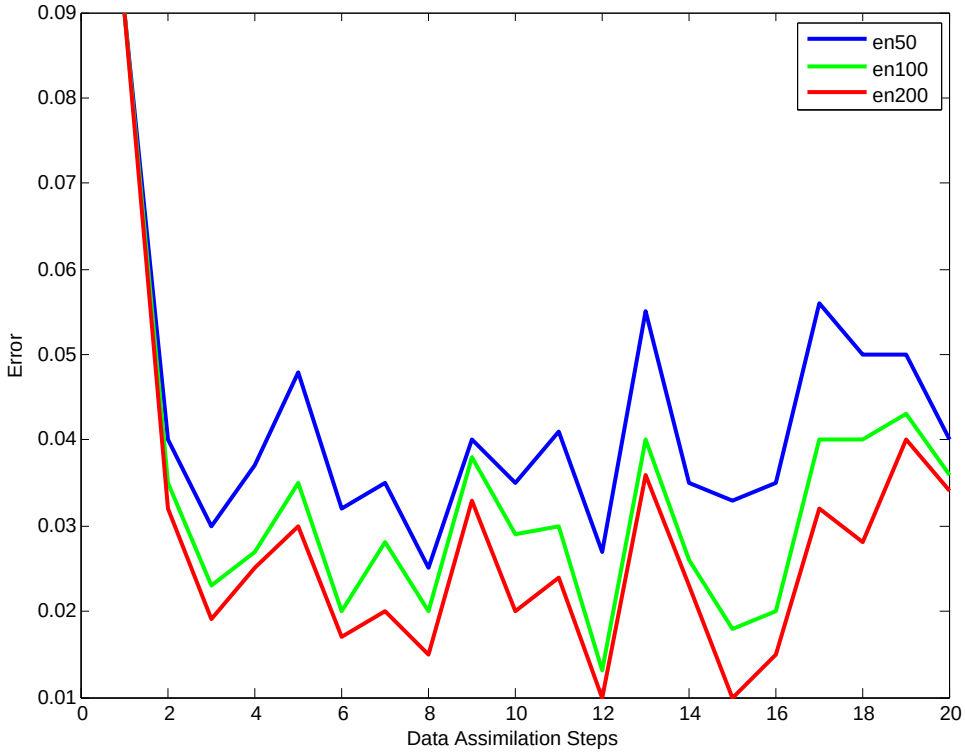


Figure 6.6: *Exploration of impact of ensemble size with an inflation factor of* 1.2.

We also studied the impact of the assimilated data on the system performance. To this end, we conducted assimilation experiments in which both depth data from USGS and velocity data from drifters were assimilated jointly. The results of these experiments are shown in Figure 6.7 for the forecast and analysis RMS errors. These assimilation runs used an ensemble size of 150 and an inflation factor of 1.2. This is consistent with known EnKF implementations in the literature [157], [159].

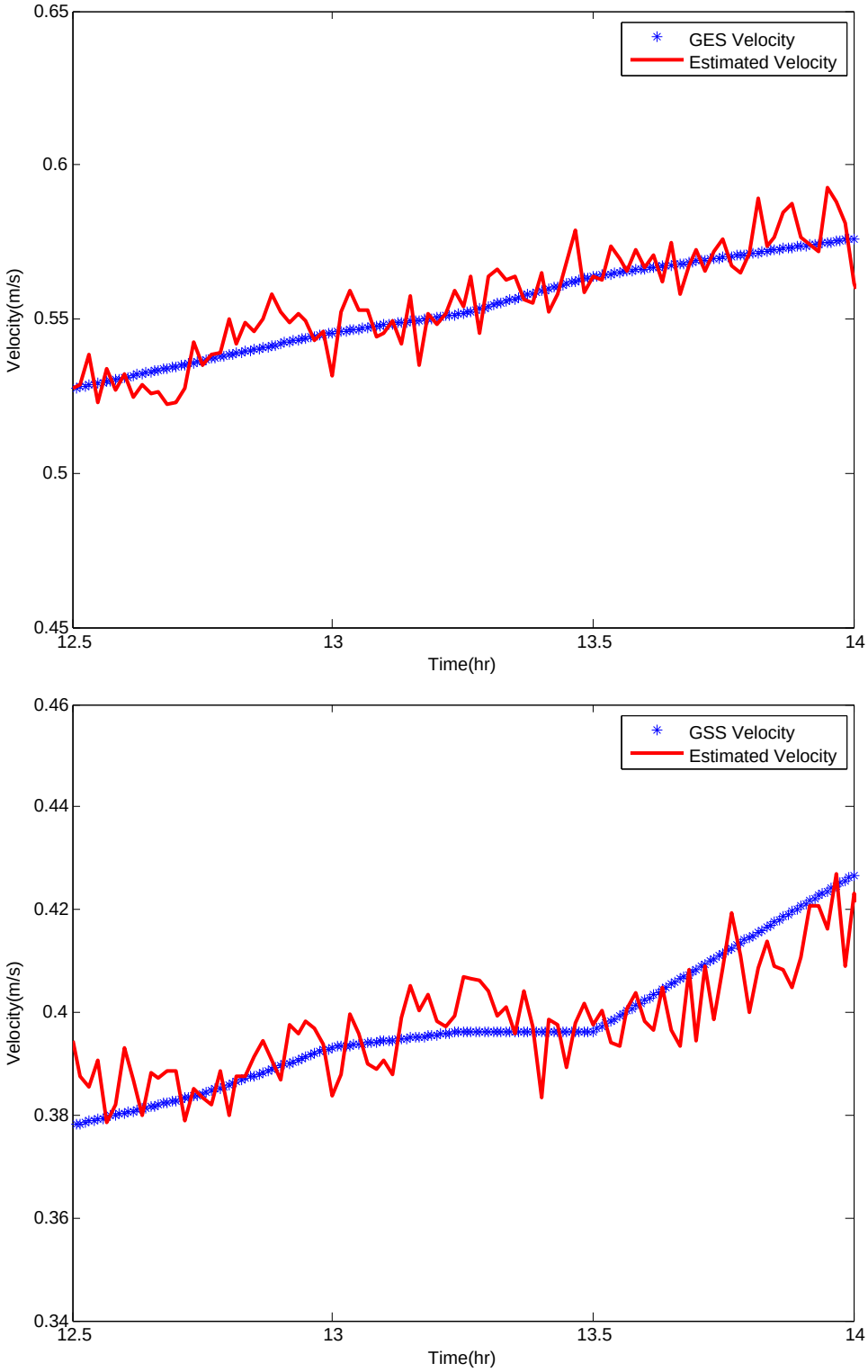


Figure 6.7: *EnKF data assimilation results v.s. USGS measurements.*

6.7 Summary

We have proposed a state estimation method for 2D shallow water equations in river hydraulics using Lagrangian measurement data. The method is based on the use of the ensemble Kalman filter.

Special attention has been paid to handle modeling errors that arise from the use of a simplified evolution model for the river flow and thus for the dynamics of the Lagrangian drifters. The modeling error in this study was modeled as a random variable with a Gaussian distribution. As the computation of the modeling error statistics is an intensive task, it should be performed prior to the actual estimation process.

It is concluded that when the modeling errors are taken into account, the estimates are more accurate than the state estimates obtained otherwise. Also, algorithms with ad-hoc error structures may suffer from stability problems, if the model errors are not adequately represented. Furthermore, extensive tuning may be needed to ensure stable assimilations.

Chapter 7

Conclusion

7.1 Summary of the Research

A prompt and accurate flow estimation of large, tidal-driving channel networks has been a persistent challenge in water resource management, especially in the Sacramento-San Joaquin Delta of California, where the water demand keeps increasing. In this dissertation, we present the framework of a sensing-modeling system that utilizes Lagrangian drifter data to characterize the flow state in the area of interest, with an emphasis on the implementation of data assimilation techniques. The system is capable of predicting regional flows and transport in the Delta in a real-time mode, without dependence on historical data.

We initiate the research by proposing a novel measure to reliably estimate the boundary conditions in a tidal-driven channel using Eulerian measurement data. The channel is represented with a 1D shallow water equation (SWE) model (Linearized Saint-Venant equations), and the flow parameters are identified in the frequency domain with modal decomposition to reduce computational costs. Later, this Eulerian-based boundary estimation method is extended to linear channel networks, giving the consideration of mass continuity and energy conservation at internal junctions. The estimation accuracy is substantially improved with data reconciliation. This feature, taking advantage of information redundancy in the data set, enables us to identify and exclude any erroneous measurements. As such, a fast and accurate representation of boundary conditions is particularly useful in subsequent real-time flow estimation, featuring Lagrangian measurements and data assimilation techniques.

Lagrangian data is obtained when floating drifters move along with the flow and report their locations. The data provides instant information about the flow, and is further assimilated into underlying SWE models. Two different approaches to facilitate the flow estimation have been investigated

in this dissertation.

- First, a variational assimilation method (Quadratic Programming) is applied to a 1D SWE model (Linearized Saint-Venant equations). The assimilation method poses the problem of estimating the flow state in a channel network as a quadratic programming by minimizing a quadratic cost function – the norm of the difference between the drifter observations and the model velocity predictions – and expressing the constraints in terms of linearized equalities and inequalities. The problem is then efficiently solved using a fast and robust algorithm. The approach is easy to implement and low in computation costs. The proposed method is initially tested in a twin experiment, in which the Lagrangian data are actually simulated by a 2D nonlinear shallow water model (TELEMAC), and later tested with experimental observations.

- Later, a sequential assimilation method (Ensemble Kalman Filtering) is implemented to a 2D SWE model (depth-integrated Navier-Stokes equations). The SWE model is implemented in a large hydraulic model named REALM. The assimilation method involves a series of state analysis and updates, where the observed Lagrangian data is incorporated into the state one step at a time to incrementally correct the model prediction. The boundary conditions previously estimated from Eulerian data provide an "initial" guess in the assimilation. The implementation of this method demands powerful computation ability, and is achieved on high-performance computing clusters at NERSC. It leads to a real-time flow monitoring system that is capable of providing accurate descriptions of varying flow in the Delta, using a combination of Eulerian and Lagrangian computations.

To assess the performance of the proposed data assimilation methods, we investigated a distributed network of channels, subject to quasi-periodic tidal forcing, in the Sacramento-San Joaquin River Delta. Field operational experiments were carried out with a fleet of over 70 floating drifters, deployed within approximately $0.55\,\mathrm{km}^2$ of the river network. During the experiments, more than 325,000 GPS readings were taken from the floating drifters and collected, in real time, onto a central server. It is the first experiment of this kind conducted at such scale, where high-density Lagrangian data have been collected in a real river environment and successfully assimilated over a full tidal cycle.

It is demonstrated that both of the proposed assimilation methods (i.e., QP in 1D SWE model and EnKF in 2D SWE model) can handle the Lagrangian data with sufficient accuracy. In many practical cases, the 1D flow estimation is adequate for water resource management to retrieve critical flow characteristics in a fast and efficient manner. In the case of complex channel geometry, however, the 2D flow estimation is vital to represent the hydraulic system.

7.2 Highlights of the Work

The majority of existing literature in open channel flow estimation focuses on the utilization of Eulerian measurement data, as it is readily to relate them to the flow properties. Lagrangian flow data, on the other hand, is much more difficult to interpret, causing rather limited application in flow estimation. This dissertation not only verifies the applicability of Lagrangian data in reality, but also indicates that, with proper data assimilation, Lagrangian flow data is capable of accurately characterizing flow in a fine scale.

Data assimilation using Lagrangian data has been studied for years. Most of the related literature, however, validates the method with numerical examples only. This dissertation presents the-first-of-its-kind assimilation using Lagrangian flow data from a real channel network.

Moreover, the dissertation explores various data assimilation options for open channel networks subject to tidal forcing: the underlying physical model can be either 1D (Linearized Saint-Venant Equation) or 2D (depth-integrated Navier-Stokes Equations); and the data assimilation method can be either variational (Quadratic Programming) or sequential (EnKF). The computational cost varies remarkably when implementing different data assimilation methods, and it is highly recommended to choose a suited assimilation method to efficiently process the Lagrangian data in specific flow estimations.

7.3 Suggestions for Future Work

The dissertation presents the main framework of a flow estimation system featuring Lagrangian drifter data. There are many aspects to be further improved:

- Current drifter design enables only the tracking of its positions. If integrated with more physical/chemical/biologic sensors (e.g., temperature, salinity, nutrients, etc.) on the drifter platform, it would become a much more powerful tool in flow studies. Accordingly, the underlying physical model used in data assimilation would become more sophisticated.
- Some data assimilation methods are rather sensitive to the model parameters chosen in the implementation. For example, the EnKF assimilation results only converge at some properly-selected spatial and time steps. The stability in assimilation methods is worth of investigation.
- In general, data assimilations would consume considerable computation resources, and the overall assimilation time tends to grow when more information is involved. It is therefore

important to optimize the size of drifter fleet, the resolution of the simulation model and choose a suited deployment strategy.

Bibliography

[1] Floating Sensors Network website. `http://float.berkeley.edu`, 2013.

[2] ALBUQUERQUE, J. S., AND BIEGLER, L. T. Data reconciliation and gross-error detection for dynamic systems. *AIChE Journal 42*, 10 (1996), 2841–2856.

[3] ALHAJ-DIBO, M., MAQUIN, D., AND RAGOT, J. Data reconciliation: A robust approach using a contaminated distribution. *Control Engineering Practice 16*, 2 (2008), 159–170.

[4] ANDERSON, B. D., AND MOORE, J. B. *Optimal filtering*. DoverPublications. com, 2012.

[5] ANDERSON, J. L. A local least squares framework for ensemble filtering. *Monthly Weather Review 131*, 4 (2003), 634–642.

[6] ANDERSON, J. L. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A 61*, 1 (2009), 72–83.

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

[8] ANTHES, R. A. Data assimilation and initialization of hurricane prediction models. *Journal of Atmospheric Sciences 31* (1974), 702–719.

[9] APTE, A., HAIRER, M., STUART, A., AND VOSS, J. Sampling the posterior: An approach to non-Gaussian data assimilation. *Physica D: Nonlinear Phenomena 230*, 1 (2007), 50–64.

[10] ATANOV, G. A., EVSEEVA, E. G., AND MESELHE, E. A. Estimation of roughness profile in trapezoidal open channels. *Journal of Hydraulic Engineering 125*, 3 (1999), 309–312.

[11] ATELJEVICH, E., COLELLA, P., GRAVES, D., LIGOCKI, T., PERCELAY, J., SCHWARTZ, P., AND SHU, Q. CFD modeling in the san francisco bay and delta. In *Proceedings of the Fourth SIAM Conference on Mathematics for Industry* (2009).

[12] BAGAJEWICZ, M. J., AND CABRERA, E. Data reconciliation in gas pipeline systems. *Industrial & engineering chemistry research 42*, 22 (2003), 5596–5606.

[13] BARRERO, O., BERNSTEIN, D. S., AND DE MOOR, B. L. Spatially localized Kalman filtering for data assimilation. In *American Control Conference, 2005. Proceedings of the 2005* (2005), IEEE, pp. 3468–3473.

[14] BEARD, J., WEEKLY, K., OROZA, C., TINKA, A., AND BAYEN, A. M. In *Networked Embedded Systems for Every Application (NESEA), 2012 IEEE 3rd International Conference on* (2012), IEEE, pp. 1–7.

[15] BEDJAOUI, N., LITRICO, X., KOENIG, D., RIBOT-BRUNO, J., AND MALATERRE, P.-O. Static and dynamic data reconciliation for an irrigation canal. *Journal of Irrigation and Drainage Engineering 134*, 6 (2008), 778–787.

[16] BENNETT, A., AND MCINTOSH, P. Open ocean modeling as an inverse problem: Tidal theory. *Journal of Physical Oceanography 12*, 10 (1982), 1004–1018.

[17] BITTERMAN, D., AND HANSEN, D. The design of a low-cost tropical drifter buoy. In *Proc. Marine Data Systems International Symp* (1986), pp. 575–581.

[18] BJERKLIE, D. M., MOLLER, D., SMITH, L. C., AND DINGMAN, S. L. Estimating discharge in rivers using remotely sensed hydraulic information. *Journal of Hydrology 309*, 1 (2005), 191–209.

[19] BOUTTIER, F., AND KELLY, G. Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Quarterly Journal of the Royal Meteorological Society 127*, 574 (2001), 1469–1488.

[20] BOYD, S., AND VANDENBERGHE, L. *Convex optimization*. Cambridge University Press, 2004.

[21] BREKKE, C., AND SOLBERG, A. H. Oil spill detection by satellite remote sensing. *Remote Sensing of Environment 95*, 1 (2005), 1–13.

[22] CALIFORNIA DEPARTMENT OF WATER RESOURCES. *Enhanced Calibration and Validation of DSM2 HYDRO and QUAL.*

[23] CASTAINGS, W., DARTUS, D., HONNORAT, M., LE DIMET, F.-X., LOUKILI, Y., AND MONNIER, J. Automatic differentiation: a tool for variational data assimilation and adjoint sensitivity analysis for flood modeling. In *Automatic Differentiation: Applications, Theory, and Implementations*. Springer, 2006, pp. 249–262.

[24] CHAI, T., CARMICHAEL, G. R., SANDU, A., TANG, Y., AND DAESCU, D. N. Chemical data assimilation of transport and chemical evolution over the pacific (trace-p) aircraft measurements. *Journal of Geophysical Research: Atmospheres (1984–2012) 111*, D2 (2006).

[25] CHAU, K. Application of the preissmann scheme on flood propagation in river systems in difficult terrain. *Hydrology in Mountainous Regions. I-Hydrological Measurements* (1990).

[26] CHOW, V. T. *Open-channel hydraulics*. McGraw-Hill, 1959.

[27] CHU, P. C., IVANOV, L. M., KORZHOVA, T. P., MARGOLINA, T. M., AND MELNICHENKO, O. V. Analysis of sparse and noisy ocean current data using flow decomposition. part II: Applications to Eulerian and Lagrangian data. *Journal of Atmospheric and Oceanic Technology 20*, 4 (2003), 492–512.

[28] CLARK, D. D. Overview of the argos system. In *OCEANS'89. Proceedings* (1989), vol. 3, IEEE, pp. 934–939.

[29] CLARK, M. P., SLATER, A. G., RUPP, D. E., WOODS, R. A., VRUGT, J. A., GUPTA, H. V., WAGENER, T., AND HAY, L. E. Framework for understanding structural errors (fuse): A modular framework to diagnose differences between hydrological models. *Water Resources Research 44*, 12 (2008).

[30] CLAUDEL, C. G. *Convex formulations of inverse modeling problems on systems modeled by Hamilton-Jacobi equations. Applications to traffic flow engineering*. PhD thesis, University of California, 2010.

[31] CLAUDEL, C. G., AND BAYEN, A. M. Convex formulations of data assimilation problems for a class of Hamilton-Jacobi equations. *SIAM Journal on Control and Optimization 49*, 2 (2011), 383–402.

[32] COLBURN, C., CESSNA, J., AND BEWLEY, T. State estimation in wall-bounded flow systems. part 3. the ensemble Kalman filter. *Journal of Fluid Mechanics 682* (2011), 289.

[33] COLELLA, P. Multidimensional upwind methods for hyperbolic conservation laws. *Journal of Computational Physics 87*, 1 (1990), 171–200.

[34] COLELLA, P., GRAVES, D. T., KEEN, B. J., AND MODIANO, D. A cartesian grid embedded boundary method for hyperbolic conservation laws. *Journal of Computational Physics 211*, 1 (2006), 347–366.

[35] CONSTANTINESCU, E. M., SANDU, A., CHAI, T., AND CARMICHAEL, G. R. Ensemble-based chemical data assimilation. i: General approach. *Quarterly Journal of the Royal Meteorological Society 133*, 626 (2007), 1229–1243.

[36] CORAZZA, M., KALNAY, E., YANG, S., ET AL. An implementation of the local ensemble Kalman filter in a quasi geostrophic model and comparison with 3D-Var. *Nonlinear Processes in Geophysics 14*, 1 (2007), 89–101.

[37] CROWE, C. M. Data reconciliation - progress and challenges. *Journal of Process Control 6*, 2 (1996), 89–98.

[38] CROWE, C. T., ELGER, D. F., ROBERSON, J. A., CROWE, C. T., AND CROWE, C. T. *Engineering fluid mechanics*, vol. 7. Wiley Hoboken, NJ, 2005.

[39] CUNGE, J. Practical aspects of computational river hydraulics. *Pitman Publishing Ltd. London,(17 CUN), 1980, 420* (1980).

[40] DALEY, R. Estimating model-error covariances for application to atmospheric data assimilation. *Monthly Weather Review 120*, 8 (1992), 1735–1746.

[41] DAVIS, R. E. Drifter observations of coastal surface currents during code: The method and descriptive view. *Journal of Geophysical Research: Oceans (1978–2012) 90*, C3 (1985), 4741–4755.

[42] DE SERRES, B., ROY, A. G., BIRON, P. M., AND BEST, J. L. Three-dimensional structure of flow at a confluence of river channels with discordant beds. *Geomorphology 26*, 4 (1999), 313–335.

[43] DEE, D. P. On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review 123*, 4 (1995), 1128–1145.

[44] DELTOUR, J.-L., CANIVET, E., SANFILIPPO, F., AND SAU, J. Data reconciliation on the complex hydraulic system of canal de provence. *Journal of Irrigation and Drainage Engineering 131*, 3 (2005), 291–297.

[45] DETWEILLER, C., VASILESCU, I., AND RUS, D. An underwater sensor network with dual communications, sensing, and mobility. In *OCEANS 2007-Europe* (2007), IEEE, pp. 1–6.

[46] DEVEGOWDA, D., ARROYO-NEGRETE, E., AND DATTA-GUPTA, A. Flow relevant covariance localization during dynamic data assimilation using EnKF. *Advances in Water Resources 33*, 2 (2010), 129–145.

[47] DING, Y., JIA, Y., AND WANG, S. S. Identification of Manning's roughness coefficients in shallow water flows. *Journal of Hydraulic Engineering 130*, 6 (2004), 501–510.

[48] ENGELEN, R. J., ANDERSSON, E., CHEVALLIER, F., HOLLINGSWORTH, A., MATRICARDI, M., MCNALLY, A. P., THÉPAUT, J.-N., AND WATTS, P. D. Estimating atmospheric CO2 from advanced infrared satellite radiances within an operational 4D-Var data assimilation system: Methodology and first results. *Journal of Geophysical Research: Atmospheres (1984–2012) 109*, D19 (2004).

[49] EVENSEN, G. 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 (1994), 10143–10162.

[50] EVENSEN, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean dynamics 53*, 4 (2003), 343–367.

[51] EVENSEN, G. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics 54*, 6 (2004), 539–560.

[52] EVENSEN, G., AND VAN LEEUWEN, P. J. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review 128*, 6 (2000), 1852–1867.

[53] FALCONER, R. An introduction to nearly horizontal flows. *Coastal Estuarial and Harbour Engineers, Reference Book, E and FN Spon Ltd., London* (1993), 27–36.

[54] FISCHER, H. B. *Mixing in inland and coastal waters*. Academic Press, 1979.

[55] FISHER, M., AND LARY, D. Lagrangian four-dimensional variational data assimilation of chemical species. *Quarterly Journal of the Royal Meteorological Society 121*, 527 (1995), 1681–1704.

[56] FUKUMORI, I., AND MALANOTTE-RIZZOLI, P. An approximate Kalman filter for ocean data assimilation: An example with an idealized gulf stream model. *Journal of Geophysical Research: Oceans (1978–2012) 100*, C4 (1995), 6777–6793.

[57] GEJADZE, I. Y., AND MONNIER, J. On a 2D 'zoom' for the 1D shallow water model: Coupling and data assimilation. *Computer Methods in Applied Mechanics and Engineering 196*, 45 (2007), 4628–4643.

[58] GLAISTER, P. Approximate riemann solutions of the shallow water equations. *Journal of Hydraulic Research 26*, 3 (1988), 293–306.

[59] GLANZMANN, G., VON SIEBENTHAL, M., GEYER, T., PAPAFOTIOU, G., AND MORARI, M. Supervisory water level control for cascaded river power plants. In *Proceedings of the Hydropower Conference, Stavanger, Norway* (2005).

[60] GOTTLIEB, J., DUCHATEAU, P., ET AL. *Parameter identification and inverse problems in hydrology, geology and ecology*. Kluwer Academic Publishers, 1996.

[61] GOULD, W. J. From swallow floats to argo - the development of neutrally buoyant floats. *Deep Sea Research Part II: Topical Studies in Oceanography 52*, 3 (2005), 529–543.

[62] GRAY, D. D., AND GIORGINI, A. The validity of the boussinesq approximation for liquids and gases. *International Journal of Heat and Mass Transfer 19*, 5 (1976), 545–551.

[63] GRIFFA, A. Applications of stochastic particle models to oceanographic problems. In *Stochastic modelling in physical oceanography*. Springer, 1996, pp. 113–140.

[64] GRIFFA, A. *Lagrangian analysis and prediction of coastal and ocean dynamics*. Cambridge University Press, 2007.

[65] GUNSON, J. R., AND MALANOTTE-RIZZOLI, P. Assimilation studies of open-ocean flows: Error measures with strongly nonlinear dynamics. *Journal of Geophysical Research 101*, C12 (1996), 28473–28.

[66] HAN, Y., DE CALLAFON, R. A., CORTÉS, J., AND JAFFE, J. Dynamic modeling and pneumatic switching control of a submersible drogue. In *ICINCO (2)* (2010), pp. 89–97.

[67] HARGREAVES, J., ANNAN, J., EDWARDS, N., AND MARSH, R. An efficient climate forecasting method using an intermediate complexity earth system model and the ensemble Kalman filter. *Climate Dynamics 23*, 7-8 (2004), 745–760.

[68] HENDRICKS FRANSSEN, H., AND KINZELBACH, W. Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research 44*, 9 (2008).

[69] HERRERA, J. C., WORK, D. B., HERRING, R., BAN, X. J., JACOBSON, Q., AND BAYEN, A. M. Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies 18*, 4 (2010), 568–583.

[70] HEYEN, G., DUMONT, M.-N., AND KALITVENTZEFF, B. Computer-aided design of redundant sensor networks. *Computer Aided Chemical Engineering 10* (2002), 685–690.

[71] HONNORAT, M., MONNIER, J., AND LE DIMET, F.-X. Lagrangian data assimilation for river hydraulics simulations. *Computing and Visualization in Science 12*, 5 (2009), 235–246.

[72] HOTEIT, I., PHAM, D.-T., AND BLUM, J. A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical pacific. *Journal of Marine Systems 36*, 1 (2002), 101–127.

[73] HOUTEKAMER, P., AND MITCHELL, H. L. Ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society 131*, 613 (2005), 3269–3289.

[74] HOUTEKAMER, P., MITCHELL, H. L., AND DENG, X. Model error representation in an operational ensemble Kalman filter. *Monthly Weather Review 137*, 7 (2009), 2126–2143.

[75] HOUTEKAMER, P. L., MITCHELL, H. L., PELLERIN, G., BUEHNER, M., CHARRON, M., SPACEK, L., AND HANSEN, B. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review 133*, 3 (2005), 604–620.

[76] HUTTUNEN, J., AND KAIPIO, J. Approximation error analysis in nonlinear state estimation with an application to state-space identification. *Inverse Problems 23*, 5 (2007), 2141.

[77] HUTTUNEN, J. M., AND KAIPIO, J. P. Approximation errors in nonstationary inverse problems. *Inverse Problems and Imaging 1*, 1 (2007), 77.

[78] IDE, K., COURTIER, P., GHIL, M., AND LORENC, A. Unified notation for data assimilation: operational, sequential and variational. *Practice 75*, 1B (1997), 181–189.

[79] INC., G. Using the location manager. `http://developer.android.com/training/basics/location/locationmanager.html`, 2013. Accessed: 2/19/2013.

[80] ISAKSEN, L., FISHER, M., AND BERNER, J. Use of analysis ensembles in estimating flow-dependent background error variance. In *Proceedings of the ECMWF Workshop on Flow-dependent aspects of data assimilation* (2007), pp. 11–13.

[81] JULIER, S. J., AND UHLMANN, J. K. New extension of the Kalman filter to nonlinear systems. In *AeroSense'97* (1997), International Society for Optics and Photonics, pp. 182–193.

[82] KAIPIO, J. P., AND SOMERSALO, E. *Statistical and computational inverse problems*, vol. 160. Springer, 2005.

[83] KISHORE KUMAR, K., JAIN, A., AND NARAYANA RAO, D. VHF/UHF radar observations of tropical mesoscale convective systems over southern india. In *Annales Geophysicae* (2005), vol. 23, Copernicus GmbH, pp. 1673–1683.

[84] LAVENDER, K. L., BRECHNER OWENS, W., AND DAVIS, R. E. The mid-depth circulation of the subpolar north atlantic ocean as measured by subsurface floats. *Deep Sea Research Part I: Oceanographic Research Papers 52*, 5 (2005), 767–785.

[85] LEONARD, N. E., PALEY, D. A., DAVIS, R. E., FRATANTONI, D. M., LEKIEN, F., AND ZHANG, F. Coordinated control of an underwater glider fleet in an adaptive ocean sampling field experiment in Monterey Bay. *Journal of Field Robotics 27*, 6 (2010), 718–740.

[86] LEONARD BECKER, W. W. Identification of parameters in unsteady open channel flows. *Water Resources Research 8*, 4 (1972), 956–965.

[87] LERCZAK, J. A., GEYER, W. R., AND CHANT, R. J. Mechanisms driving the time-dependent salt flux in a partially stratified estuary. *Journal of Physical Oceanography 36*, 12 (2006), 2296–2311.

[88] LERMUSIAUX, P., ROBINSON, A., HALEY, P., AND LESLIE, W. Advanced interdisciplinary data assimilation: Filtering and smoothing via error subspace statistical estimation. In *OCEANS'02 MTS/IEEE* (2002), vol. 2, IEEE, pp. 795–802.

[89] LI, H., KALNAY, E., AND MIYOSHI, T. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society 135*, 639 (2009), 523–533.

[90] LI, H., KALNAY, E., MIYOSHI, T., AND DANFORTH, C. M. Accounting for model errors in ensemble data assimilation. *Monthly Weather Review 137*, 10 (2009), 3407–3419.

[91] LIANG, S. *Quantitative remote sensing of land surfaces*, vol. 30. John Wiley & Sons, 2005.

[92] LIONS, J. L. Remarks on the theory of optimal control of distributed systems. *Control Theory of Systems Governed by Partial Differential Equations* (1977), 1–5.

[93] LITRICO, X., AND FROMION, V. Frequency modeling of open-channel flow. *Journal of Hydraulic Engineering 130*, 8 (2004), 806–815.

[94] LITRICO, X., AND FROMION, V. Simplified modeling of irrigation canals for controller design. *Journal of Irrigation and Drainage Engineering 130*, 5 (2004), 373–383.

[95] LITRICO, X., AND FROMION, V. Boundary control of linearized saint-venant equations oscillating modes. *Automatica 42*, 6 (2006), 967–972.

[96] LITRICO, X., AND FROMION, V. h_∞ control of an irrigation canal pool with a mixed control politics. *Control Systems Technology, IEEE Transactions on 14*, 1 (2006), 99–111.

[97] LITRICO, X., AND FROMION, V. *Modeling and control of hydrosystems*. Springer, 2009.

[98] LORENC, A. C. Modelling of error covariances by 4D-Var data assimilation. *Quarterly Journal of the Royal Meteorological Society 129*, 595 (2003), 3167–3182.

[99] MAREELS, I., WEYER, E., OOI, S. K., CANTONI, M., LI, Y., AND NAIR, G. Systems engineering for irrigation systems: Successes and challenges. *Annual Reviews in Control 29*, 2 (2005), 191–204.

[100] MARSHALL, J., ADCROFT, A., HILL, C., PERELMAN, L., AND HEISEY, C. A finite-volume, incompressible navier stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research: Oceans (1978–2012) 102*, C3 (1997), 5753–5766.

[101] MENG, Z., AND ZHANG, F. Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. part III: Comparison with 3D-Var in a real-data case study. *Monthly Weather Review 136*, 2 (2008), 522–540.

[102] MILLER, R. N., GHIL, M., AND GAUTHIEZ, F. Advanced data assimilation in strongly nonlinear dynamical systems. *Journal of the Atmospheric Sciences 51*, 8 (1994), 1037–1056.

[103] MOLCARD, A., GRIFFA, A., AND ÖZGÖKMEN, T. M. Lagrangian data assimilation in multilayer primitive equation ocean models. *Journal of Atmospheric and Oceanic Technology 22*, 1 (2005), 70–83.

[104] MOLCARD, A., PITERBARG, L. I., GRIFFA, A., ÖZGÖKMEN, T. M., AND MARIANO, A. J. Assimilation of drifter observations for the reconstruction of the Eulerian circulation field. *Journal of Geophysical Research 108*, C3 (2003), 3056.

[105] MOLCARD, A., POJE, A. C., AND ÖZGÖKMEN, T. M. Directed drifter launch strategies for Lagrangian data assimilation using hyperbolic trajectories. *Ocean Modelling 12*, 3 (2006), 268–289.

[106] MONNIER, J., HONNORAT, M., LAI, X., AND DIMET, F.-X. L. Assimilation of remote sensed data for river hydraulic simulations. *PAMM 7*, 1 (2007), 1100203–1100204.

[107] MONSEN, N. E., CLOERN, J. E., AND BURAU, J. R. Effects of flow diversions on water and habitat quality: Examples from california's highly manipulated sacramento-san joaquin delta. *San Francisco Estuary and Watershed Science 5*, 3 (2007).

[108] MORADKHANI, H., HSU, K.-L., GUPTA, H., AND SOROOSHIAN, S. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research 41*, 5 (2005), W05012.

[109] MORÉ, J. J., AND TORALDO, G. Algorithms for bound constrained quadratic programming problems. *Numerische Mathematik 55*, 4 (1989), 377–400.

[110] NASH, J., AND SUTCLIFFE, J. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology 10*, 3 (1970), 282–290.

[111] NAVON, I. A review of variational and optimization methods in meteorology. In *Festive volume of the international symposium on variational methods in geosciences. Elsevier Science Pub. Co. Developments in Geo-mathematics* (1986), vol. 5, pp. 29–34.

[112] NAVON, I. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of Atmospheres and Oceans 27*, 1 (1998), 55–79.

[113] NAVON, I. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of Atmospheres and Oceans 27*, 1 (1998), 55–79.

[114] NILLER, P. P., DAVIS, R. E., AND WHITE, H. J. Water-following characteristics of a mixed layer drifter. *Deep Sea Research Part A. Oceanographic Research Papers 34*, 11 (1987), 1867–1881.

[115] NISSINEN, A., KOLEHMAINEN, V. P., AND KAIPIO, J. P. Compensation of modelling errors due to unknown domain boundary in electrical impedance tomography. *Medical Imaging, IEEE Transactions on 30*, 2 (2011), 231–242.

[116] NODET, M. Variational assimilation of Lagrangian data in oceanography. *Inverse Problems 22*, 1 (2006), 245.

[117] OCZKOWSKI, M., SZUNYOGH, I., AND PATIL, D. Mechanisms for the development of locally low-dimensional atmospheric dynamics. *Journal of the atmospheric sciences 62*, 4 (2005), 1135–1156.

[118] OROZA, C., TINKA, A., WRIGHT, P. K., AND BAYEN, A. M. Design of a network of robotic Lagrangian sensors for shallow water environments with case studies for multiple applications. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* (2013).

[119] OTT, E., HUNT, B. R., SZUNYOGH, I., ZIMIN, A. V., KOSTELICH, E. J., CORAZZA, M., KALNAY, E., PATIL, D., AND YORKE, J. A. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A 56*, 5 (2004), 415–428.

[120] PANICONI, C., MARROCU, M., PUTTI, M., AND VERBUNT, M. Newtonian nudging for a richards equation-based distributed hydrological model. *Advances in Water Resources 26*, 2 (2003), 161–178.

[121] PLACKETT, R. L. Karl pearson and the chi-squared test. *International Statistical Review/Revue Internationale de Statistique* (1983), 59–72.

[122] PLANT, W. J., KELLER, W. C., AND HAYES, K. Measurement of river surface currents with coherent microwave systems. *Geoscience and Remote Sensing, IEEE Transactions on 43*, 6 (2005), 1242–1257.

[123] POLAVARAPU, S., TANGUAY, M., AND FILLION, L. Four-dimensional variational data assimilation with digital filter initialization. *Monthly Weather Review 128*, 7 (2000), 2491–2510.

[124] POULAIN, P.-M., LUTHER, D. S., AND PATZERT, W. C. Deriving inertial wave characteristics from surface drifter velocities: Frequency variability in the tropical pacific. *Journal of Geophysical Research 97*, C11 (1992), 17947–17.

[125] POWELL, B., ARANGO, H., MOORE, A., DI LORENZO, E., MILLIFF, R., AND FOLEY, D. 4D-Var data assimilation in the intra-americas sea with the regional ocean modeling system (roms). *Ocean Modelling 25*, 3 (2008), 173–188.

[126] RAFIEE, M., WU, Q., AND BAYEN, A. Kalman filter based estimation of flow states in open channels using Lagrangian sensing. In *Proceedings of the 48th Conference of Decision and Control* (Shanghai, Dec 2009).

[127] RAGOT, J., AND MAQUIN, D. Reformulation of data reconciliation problem with unknown-but-bounded errors. *Industrial & engineering chemistry research 43*, 6 (2004), 1530–1536.

[128] REICHLE, R. H., MCLAUGHLIN, D. B., AND ENTEKHABI, D. Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review 130*, 1 (2002), 103–114.

[129] REIMERS, C., WERTHER, J., AND GRUHN, G. Flowsheet simulation of solids processes: data reconciliation and adjustment of model parameters. *Chemical Engineering and Processing: Process Intensification 47*, 1 (2008), 138–158.

[130] REPORT EDF. *TELEMAC 2D. Version 5.2 – Principle note*, 2002.

[131] ROBINSON, A. R., AND WALSTAD, L. J. The harvard open ocean model: Calibration and application to dynamical process, forecasting, and data assimilation studies. *Applied Numerical Mathematics 3*, 1 (1987), 89–131.

[132] RYAN, J., NIC AONGHUSA, C., AND SWEENEY, E. Smartbay, ireland: Design and planning for a cabled ocean observatory off the west coast of ireland. In *OCEANS 2007* (2007), IEEE, pp. 1–4.

[133] SADOURNY, R. The dynamics of finite-difference models of the shallow-water equations. *Journal of the Atmospheric Sciences 32*, 4 (1975), 680–689.

[134] SANDERS, B. F. High-resolution and non-oscillatory solution of the st. venant equations in non-rectangular and non-prismatic channels. *Journal of hydraulic Research 39*, 3 (2001), 321–330.

[135] SCHOWENGERDT, R. A. *Remote sensing: models and methods for image processing*. Academic Press, 2006.

[136] SCHUURMANS, J., CLEMMENS, A., DIJKSTRA, S., HOF, A., AND BROUWER, R. Modeling of irrigation and drainage canals for controller design. *Journal of Irrigation and Drainage Engineering 125*, 6 (1999), 338–344.

[137] SHU, Q., AND ATELJEVICH, E. Adaptive mesh, embedded boundary model for flood modeling. In *Methodology for Flow and Salinity Estimates in the Sacramento-San Joaquin Delta and Suisun Marsh*. 32nd Annual Progress Report, 2011.

[138] STEWART, R., AND PFANN, E. Oversampling and sigma-delta strategies for data conversion. *Electronics & Communication Engineering Journal 10*, 1 (1998), 37–47.

[139] STRUB, I. S., PERCELAY, J., STACEY, M. T., AND BAYEN, A. M. Inverse estimation of open boundary conditions in tidal channels. *Ocean Modelling 29*, 1 (2009), 85–93.

[140] SUJIT, P., SOUSA, J., AND PEREIRA, F. UAV and AUVs coordination for ocean exploration. In *OCEANS 2009-EUROPE* (2009), IEEE, pp. 1–7.

[141] SWALLOW, J. C. A neutral-buoyancy float for measuring deep currents. *Deep Sea Research (1953) 3*, 1 (1955), 74–81.

[142] SZUNYOGH, I., KOSTELICH, E. J., GYARMATI, G., KALNAY, E., HUNT, B. R., OTT, E., SATTERFIELD, E., AND YORKE, J. A. A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus A 60*, 1 (2008), 113–130.

[143] TINKA, A., RAFIEE, M., AND BAYEN, A. M. Floating sensor networks for river studies. *IEEE Systems Journal 7* (2013), 36–49.

[144] TINKA, A., STRUB, I., WU, Q., AND BAYEN, A. M. Quadratic programming based data assimilation with passive drifting sensors for shallow water flows. *International Journal of Control 83*, 8 (2010), 1686–1700.

[145] TINKA, A., WU, Q., WEEKLY, K., OROZA, C., BEARD, J., AND BAYEN, A. M. Heterogenous fleets of active and passive floating sensors for river studies. *In preparation* (2013).

[146] TORN, R. D., AND HAKIM, G. J. Ensemble data assimilation applied to rainex observations of hurricane katrina (2005). *Monthly Weather Review 137*, 9 (2009), 2817–2829.

[147] TOSSAVAINEN, O.-P., PERCELAY, J., STACEY, M., KAIPIO, J. P., AND BAYEN, A. State estimation and modeling error approach for 2D shallow water equations and Lagrangian measurements. *Water Resources Research 47*, 10 (2011), W10510.

[148] TOSSAVAINEN, O.-P., PERCELAY, J., TINKA, A., WU, Q., AND BAYEN, A. M. Ensemble Kalman filter based state estimation in 2D shallow water equations using Lagrangian sensing and state augmentation. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on* (2008), IEEE, pp. 1783–1790.

[149] TUAN PHAM, D., VERRON, J., AND CHRISTINE ROUBAUD, M. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems 16*, 3 (1998), 323–340.

[150] VENEZIANI, M., GRIFFA, A., REYNOLDS, A. M., AND MARIANO, A. J. Oceanic turbulence and stochastic models from subsurface Lagrangian data for the northwest atlantic ocean. *Journal of Physical Oceanography 34*, 8 (2004), 1884–1906.

[151] WAN, E. A., AND VAN DER MERWE, R. The unscented Kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000* (2000), IEEE, pp. 153–158.

[152] WANG, C., WEN, B., MA, Z., YAN, W., AND HUANG, X. Measurement of river surface currents with UHF FMCW radar systems. *Journal of Electromagnetic Waves and Applications 21*, 3 (2007), 375–386.

[153] WANG, W.-C., CHAU, K.-W., CHENG, C.-T., AND QIU, L. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology 374*, 3 (2009), 294–306.

[154] WANG, X., BISHOP, C. H., AND JULIER, S. J. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Monthly Weather Review 132*, 7 (2004), 1590–1605.

[155] WEEKLY, K., TINKA, A., ANDERSON, L., AND BAYEN, A. M. Autonomous river navigation using the Hamilton-Jacobi framework for underactuated vehicles. in preparation.

[156] WHITAKER, J. S., HAMILL, T. M., WEI, X., SONG, Y., AND TOTH, Z. Ensemble data assimilation with the ncep global forecast system. *Monthly Weather Review 136*, 2 (2008), 463–482.

[157] WORK, D. B., BLANDIN, S., TOSSAVAINEN, O.-P., PICCOLI, B., AND BAYEN, A. M. A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress 2010*, 1 (2010), 1–35.

[158] WORK, D. B., TOSSAVAINEN, O.-P., BLANDIN, S., BAYEN, A. M., IWUCHUKWU, T., AND TRACTON, K. An ensemble Kalman filtering approach to highway traffic estimation using gps enabled mobile devices. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on* (2008), IEEE, pp. 5062–5068.

[159] WORK, D. B., TOSSAVAINEN, O.-P., JACOBSON, Q., AND BAYEN, A. M. Lagrangian sensing: traffic estimation with mobile devices. In *American Control Conference, 2009. ACC'09.* (2009), IEEE, pp. 1536–1543.

[160] WU, Q., AMIN, S., MUNIER, S., BAYEN, A. M., LITRICO, X., AND BELAUD, G. Parameter identification for the shallow water equation using modal decomposition. In *Decision and Control, 2007 46th IEEE Conference* (2007), IEEE, pp. 1584–1590.

[161] WU, Q., AND BAYEN, A. M. Real-time water resource monitering using Lagrangian drifters. *In preparation* (2013).

[162] WU, Q., LITRICO, X., AND BAYEN, A. M. Data reconciliation of an open channel flow network using modal decomposition. *Advances in Water Resources 32*, 2 (2009), 193–204.

[163] WU, Q., RAFIEE, M., TINKA, A., AND BAYEN, A. M. Inverse modeling for open boundary conditions in channel network. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on* (2009), IEEE, pp. 8258–8265.

[164] WU, Q., TINKA, A., WEEKLY, K., BEARD, J., AND BAYEN, A. M. Variational Lagrangian data assimilation in open channel networks. *In preparation* (2013).

[165] XIE, X., AND ZHANG, D. Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter. *Advances in Water Resources 33*, 6 (2010), 678–690.

[166] YEH, W. W. Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research 22*, 2 (1986), 95–108.

[167] ZACHMANN, D., DU CHATEAU, P., AND KLUTE, A. Simultaneous approximation of water capacity and soil hydraulic conductivity by parameter identification1. *Soil Science 134*, 3 (1982), 157–163.

[168] ZHOU, J. G., CAUSON, D. M., MINGHAM, C. G., AND INGRAM, D. M. The surface gradient method for the treatment of source terms in the shallow-water equations. *Journal of Computational Physics 168*, 1 (2001), 1–25.

[169] ZHU, J., ZHOU, G., YAN, C., FU, W., AND YOU, X. A three-dimensional variational ocean data assimilation system: Scheme and preliminary results. *Science in China Series D: Earth Sciences 49*, 11 (2006), 1212–1222.