Design and Optimization of Wireless-Sensor Networks for Real-Time Monitoring in the Sierra Nevada and Sacramento-San Joaquin Delta

by

Carlos A. Oroza

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, Berkeley

Committee in charge:

Professor Steven D. Glaser, Chair
Professor Alexandre M. Bayen
Professor Kristofer S.J. Pister

Spring 2017

Design and Optimization of Wireless-Sensor Networks for Real-Time Monitoring in the Sierra Nevada and Sacramento-San Joaquin Delta

Abstract

Design and Optimization of Wireless-Sensor Networks for Real-Time Monitoring in the Sierra Nevada and Sacramento-San Joaquin Delta

by

Carlos A. Oroza

Doctor of Philosophy in Engineering - Civil and Environmental Engineering

University of California, Berkeley

Professor Steven D. Glaser, Chair

California relies extensively on two pieces of natural infrastructure for water storage and conveyance: the Sierra Nevada mountains and the Sacramento-San Joaquin Delta. The mountains act as a natural reservoir, storing winter precipitation throughout the year and slowly releasing it during the summer when demand is high. Snowmelt is then routed through the Sacramento-San Joaquin Delta to the state water project, which distributes water throughout California. Despite the importance of these natural environments for the state's water resources, the in-situ infrastructure for water monitoring is limited. The existing distribution of real-time snow-sensors is present only in flat, low-elevation regions, which does not adequately capture the spatial variability of snow depth in complex terrain. To address this, the state relies on monthly synoptic snow surveys, which require trained surveyors to manually measure the snow depth at locations across the Sierra Nevada. This process provides dense in-situ data but is labor intensive and temporally sparse. It also does not capture other important variables, such as soil moisture and radiative forcing, which are necessary for next-generation hydrologic modeling. In the Delta, in-situ measurements are limited to Eulerian sensors, which measure the rate of flow going past a point on shore, but are unable to resolve finer resolution flow fields, or measure properties of the water, such as temperature, turbidity, and salinity. This dissertation explores problems related to the design and optimization of wireless in-situ sensor networks for monitoring the water balance in the Sierra Nevada and flow fields in the Sacramento-San Joaquin Delta. Topics include: optimizing wireless snow-sensor placements using LIDAR data and machine learning, constructing reliable wireless mesh networks in complex terrain, estimating the spatial variability of soil moisture in montane regions from in-situ sensors, and the design and controller optimization of wireless in-situ sensors in the Sacramento-San Joaquin Delta.

To my family

Contents

List of Figures

List of Tables

Acknowledgments

Pursing a PhD at Berkeley has been a truly exceptional experience thanks to the contributions of many people. I would first like to thank my principal advisor, Professor Steve Glaser for developing an exciting vision of what is possible in Civil Systems Engineering, guiding my research throughout my PhD, and lots of chats over coffee. I would like to thank Professor Alex Bayen and Andrew Tinka for introducing me to the Civil Systems program and inspiring me to pursue my graduate studies. Other faculty, graduate students, and postdocs have made innumerable contributions to my knowledge and research: a huge thank you to Professor Roger Bales, Professor Sally Thompson, Professor Kris Pister, Professor Scott Moura, Thomas Watteyne, Keoma Brun, Jon Beard, Alex Mead, Kevin Weekly, Ziran Zhang, Zeshi Zheng, and Sami Malek. Thanks also to the research and support staff who make everything happen, especially to Matt Meadows and Erin Stacy at the Southern Sierra Critical Zone Observatory. Finally, thank you to the funding agencies who supported my research throughout my graduate studies: the NSF CiBER IGERT Traineeship, the NSF Critical Zone Observatories, the California Department of Water Resources, and the Office of Naval Research.

Chapter 1

Introduction

1.1 Background

The Sierra Nevada Mountains and the Sacramento-San Joaquin Delta represent important pieces of natural infrastructure for delivering California's water resources. Like many western states, a substantial fraction of California's annual water supply is derived from snowpack in the Sierras melting in the spring [72]. The runoff is then routed through the Sacramento-San Joaquin Delta, providing water to more than 25 million people and irrigating crops valued at more than $30 billion per year [31]. Water management in these regions is difficult because multiple objectives must be balanced. For instance, the Delta is used for water conveyance, recreation, and provides a habitat for many native species of fish. The existing temperature and salinity gradients are variable, and are maintained by the interaction between spring snowmelt runoff and saltwater intrusion from the San Francisco Bay. The Delta levees protect a large amount of infrastructure, so large-scale failures are costly: the Department of Water Resources estimates that the costs associated with the 2004 Jones Tract failure could exceed 44 million dollars [73].

Improving water-resources management requires large-scale, real-time measurements of water resources in both regions. In the Sierras, real-time, in-situ measurements are presently limited to snow pillows which only sample flat, open terrain, but the distribution of snow cover can vary considerably as a function of topographic features. Synoptic snow surveys provide in-situ snow-depth measurements at representative snow courses but are labor intensive and conducted only at limited (monthly) time intervals (Figure 1.1). In the Delta, measurements are limited to Eulerian sensors at fixed locations, which only measure the flow past a given point. To address these limitations, wireless-sensor networks (WSNs) could potentially enable real-time monitoring of spatially distributed phenomena at a scale that was previously not feasible. This thesis explores topics related to the design and optimization of two such systems: wireless snow- and soil-moisture sensor networks in the Sierra Nevada and a network of robotic Lagrangian (moving measurement location) sensors for the Delta.

In the Sierras, the wireless snow-sensor networks provide in-situ measurements across

Figure 1.1: Conducting a synoptic snow survey. Measurements are taken by personnel at representative regions across the landscape.
Image source: USDA (flickr.com/photos/usdagov)

the range of independent variables affecting snow-depth variability, such as elevation, aspect, slope, and canopy coverage. A typical network consists of three components: sensor stations, a base station, and repeater nodes. Sensors measure snow depth, temperature, relative humidity, soil moisture/matric potential, and solar radiation. Repeaters route data from the sensors to a base station, which uploads real-time data to a server at UC Berkeley every fifteen minutes (Figure 1.2).

For sensing in the Delta, the Lagrangian sensor system is comprised of the robotic sensors, as well as the communication and visualization infrastructure required to transmit and display the data remotely in real time. Each robotic sensor consists of a cylindrical hull which contains the batteries, communications equipment, GPS, and onboard processors. Onboard sensors for salinity, temperature and depth extend from the base of the vehicle. The full backend system consists of an integrated computational support tool and platform linked to the floating sensors. Data are also sent back to a central server using the Global System for Mobile Communications (GSM) cell phone data network. The system enables the data from the sensors to be visualized in real time, giving the end-user live access to the sensor data.

The core contributions of this thesis are fourfold: (i) A sensor placement strategy for wireless snow observatories is presented and evaluated in three one-square-kilometer catchments. (ii) A data repeater placement strategy for mesh networks in complex terrain is evaluated at the Southern Sierra Critical Zone Observatory (SSCZO). (iii) The long-term accuracy of soil moisture estimated from in-situ sensors is evaluated using a six-year record

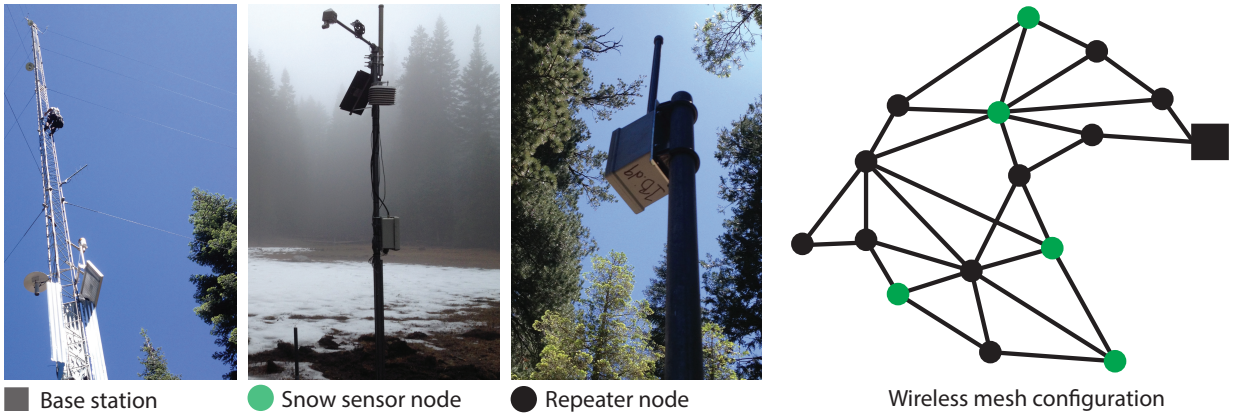

Figure 1.2: Components of a wireless snow-sensor network. A base station transmits real-time data using a satellite or cellular uplink. Snow-sensor nodes measure snow depth using ultrasonic sensors. Signal repeaters are placed throughout the region to link the sensor nodes to the base station. A multi-hop wireless mesh network is formed between all elements using 802.15.4e low-power radios.

at the Southern Sierra Critical Zone Observatory on the upper King's River. (iv) A design framework for in-situ Lagrangian sensing in the Delta is presented.

1.2 Document organization

The remaining chapters are organized as follows: Chapters 2-4 cover topics related to the optimization of sensor networks for monitoring hydrologic processes in the Sierra Nevada Mountains; Chapter 5 presents a design framework for an autonomous in-situ sensor network in the Sacramento-San Joaquin Delta; and Chapter 6 summarizes the contributions of the thesis and presents a number of opportunities for future research. A brief summary of each chapter is provided below:

Chapter 2 evaluates a sensor placement strategy for wireless snow observatories. Prior studies have shown that the spatial distribution of snow depth is controlled by a number of independent variables such as elevation, slope, aspect, and distribution of vegetation. Data on each of these variables is increasingly available at high resolution from airborne laser scanning of remote regions (LIDAR) or aerial imagery (e.g. Google Earth). The chapter describes a means of utilizing these datasets in order to optimize the distribution of snow-sensors prior to field deployment. First, sensor locations that are representative of the distribution of independent variables (slope, aspect, elevation, and canopy) are identified using a Gaussian mixture model. Second, the spatial variability and uncertainty of snow depth is estimated using a Gaussian process. Sensors are then added to the highest uncertainty regions in order to minimize the error of the estimator. This process is evaluated in three

one-square-kilometer catchments in the Southern Sierra, Tuolumne Basin, and Merced Basin in the Sierra Nevada.

Chapter 3 presents a data repeater placement strategy for wireless snow observatories. Once wireless snow-sensors are in place (as described in Chapter 2), repeaters must be placed throughout the WSN observatory to form a wireless mesh network that routes sensor data back to a real-time data uplink (satellite or cellular modem). One of the central challenges of establishing new wireless mesh networks in complex terrain is the difficulty of designing network topologies to ensure real-time data are delivered year-round. Given the large set of potential placements and complex terrain, it is difficult for field teams to determine how best to place a limited set of data repeater nodes to ensure a redundant topology is formed. The chapter outlines a strategy by which LIDAR data can be combined with packet delivery ratio (PDR) measurements in order to guide the selection of repeater placements during field deployment.

The snow-sensor networks described in Chapters 2 and 3 are typically designed to capture soil-moisture variability as well as snow-depth variability. In Chapter 4, the long-term accuracy of soil moisture estimation from in-situ sensors is quantified using a 6-year daily record of soil moisture from the Southern Sierra Critical Zone Observatory. Five explanatory variables (northness, soil texture, topographic wetness index, elevation, and location type) are used in an ensemble regression tree algorithm (Random Forest) in order to estimate soil moisture variability. The daily accuracy is quantified using the out-of-bag error for the Random Forest algorithm.

Chapter 5 describes a design framework for an autonomous fleet of Lagrangian sensors for in-situ monitoring in the Sacramento-San Joaquin Delta. A major limitation of existing Eulerian sensors is they only measure flow past a given point; they are unable to resolve finer scale flow fields or measure distributed and varying water properties like salinity or temperature at high resolution. This could be addressed by designing sensors that are deployed directly into the river. The chapter describes a design framework for a possible version of this sensor. The functional requirements and design parameters for the sensor are described using an Ishikawa diagram, and the results of the design process and sensor tests are described in detail.

Chapter 6 summarizes the findings in each chapter and outlines possibilities for future research, including integrating real-time data with remote sensing and other data sources.

Chapter 2

Sensor Placement Strategy for Wireless Snow Observatories

2.1 Introduction

This chapter presents a sensor-placement strategy for snow-sensors in wireless hydrologic observatories. A major challenge of hydrologic science in montane regions relates to estimating the spatial variability of snow cover [7, 101, 63, 38]. Multiple independent variables govern the distribution of snow cover, including elevation, slope, aspect, and the distribution of canopy, [28, 74, 59, 44]. Non-stationary effects such as climate warming and changes in vegetation structure may significantly alter the timing and magnitude of storage and runoff in these watersheds [36, 30]. Existing regression-based hydrologic models, which use statistical relations from historical hydrographs to predict runoff and inform allocation decisions [83, 88, 84] will have limited accuracy as conditions deviate from historical norms and thus may prove to be inadequate for predictions in water management.

Recent research has focused on improving hydrologic models by assimilating remote-sensing and in-situ measurements with distributed energy-balance models to better estimate storage and runoff [38]. These methods use well-developed remote-sensing [79, 89, 22, 81, 24] and energy-balance models [62, 60, 16] to estimate snow and snowmelt processes across basins. In-situ measurements for these methods are presently limited to snow pillows and snow courses, which largely sample flat, open terrain [68], yet the distribution of snow cover can vary considerably as a function of topographic features. To address this, in-situ sensor measurements can be deployed to capture the mean and variance of the snow depth, which can be used to inform models that use these statistics as inputs (e.g. [27]). Alternatively, individual sensor measurements can be used together with distributed models to estimate snow distribution across un-instrumented regions [9, 25, 26, 29, 43].

Prior studies have investigated the feasibility of using wireless-sensor networks to distribute representative snow measurements over a broader landscape [50, 87]. These studies demonstrated that wireless-sensor networks can be configured to provide simultaneous mea-

surements of snow-depth distribution, solar forcing, and subsurface exchange across a 1-km^2 region. Based on these findings, 14 additional 1-km^2 area networks were deployed across the American River Basin to develop a real-time water information system [51]. Existing deployment methods for wireless-sensor networks require extensive field surveys in order to identify sampling regions for a limited budget of sensor nodes. These practices are resource intensive, provide no guarantees of accuracy, and are unsustainable if wireless-sensor networks are going to see larger-scale adoption for real-time monitoring.

Although snow-depth statistics are site specific [37], and are often unknown prior to deployment, the distribution of independent physiographic variables is often known at high resolution (e.g. from LIDAR surveys [20]). Sensors could be deployed in representative locations based on the remotely-sensed data. Then, site-specific parameters for a snow-depth estimator could be inferred from the initial deployment in order to determine whether additional sensors are needed. This process would be well-suited to the field of machine learning, which includes "unsupervised" algorithms (which identify patterns in the independent variables without observing the dependent variable), and "supervised" or "active" algorithms (which employ limited observations of the independent variable in order to improve the estimator). A number of recent studies have applied machine learning algorithms to topics in hydrology, such as runoff and streamflow estimation [93, 61], evapotranspiration modeling [99], streamflow forecasting [86], assessment of the contamination of groundwater [52] and estimation of needs for reservoir releases [53, 98].

This chapter evaluates a two-step machine-learning method to identify optimal sensor locations for catchment-scale snow-depth observatories. First, a Gaussian mixture model (an unsupervised algorithm) was used to identify representative sampling locations in a LIDAR-derived feature space (i.e., the multivariate distribution of independent variables that exist within the catchment). Second, a Gaussian process was used to estimate the catchment-scale distribution of snow-depth and model uncertainty, then additional sensors were placed in high-uncertainty regions (i.e., supervised updates). The aim of this process is to determine the distribution of sensors that will minimize the RMSE and bias of the Gaussian process estimate throughout the catchment. The aims of the present study are to (i) determine how many sensors are needed in the unsupervised step in order to estimate the catchment-scale parameters for the estimator (i.e. the regression weights and autocorrelation for each independent variable), (ii) assess how many additional sensors are needed in the supervised step to optimally instrument the catchment, (iii) compare the accuracy of the snow-depth model using placements determined with the algorithm to an existing network and to randomized placements, and (iv) determine the spatial and temporal transferability of the method by evaluating the accuracy of the algorithm across multiple regions and over multiple LIDAR flights.

2.2 Methods

Study areas and data collection

The accuracy of the algorithm was evaluated in three 1-km^2 catchments. The first study site is located in the Southern Sierra Critical Zone Observatory (SSCZO) (37° 4' N, 119° 11' W), within the rain-snow transition of the Sierra Nevada near Fresno, California. In water year (WY) 2010 (October 1st, 2009 - September 30th, 2010), a 23-node wireless-sensor network was installed (Figure 2.1), spanning a 1.5 km transect in a forested headwater catchment where approximately 50-60% of annual precipitation falls as snow (Figure 9 in [8]). The catchment covers elevations between 1900-2100 m a.s.l., which are centrally distributed at 2000 m a.s.l. The mean and median slopes in the catchment are 11.6 and 10.2 degrees, respectively. Though a range of aspects exist within the catchment, the distribution of "northness" is weighted towards 180 degrees, as the catchment faces SW, towards the Central Valley. The canopy distribution is bimodal: most of the site is either open (penetration fraction = 1) or dense (penetration fraction = 0). For the present study, sensor locations in the existing wireless network were measured with a Trimble GPS (10-cm horizontal accuracy).

Node locations in the existing network are clustered such that the distribution of sensor measurements can be used to represent the catchment-scale mean and variance. Particular attention was paid to the effect of canopy, with clusters of sensors measuring the drip-edge, under-canopy, and open regions at multiple locations in the catchment. Although these measurements are spatially proximate, they represent 23 distinct measurements at dissimilar points in the feature space when used in the Gaussian-process estimator. Figure 2.1 shows the overall sensor distribution and typical network structure. The right-hand panel shows a cluster of sensors in the network that were placed to capture drip-edge-to-open gradient in the NE region of the catchment. Other clusters in the network are designed to capture under-canopy effects as well as gradients of slope, aspect, and elevation. The existing network provides a representative sampling of canopy and aspect, but significantly under-represents high and low elevations (elevations between 1850-1950 m and 2020-2100 m are not covered), as well as high values of slope (slopes above 15 degrees are not covered).

Physiographic variables (Figure 2.2) and a snow on/snow off raster (Figure 2.3) were gathered from the NSF Open Topography database, opentopography.org [accessed 1 February, 2016]. Elevation, slope, and aspect extracted from LIDAR data were processed in ArcMap 10.2. On average, 9.21 points per square meter were used to generate the 1-m^2 DEM, canopy, and snow-depth rasters. Elevation information was stored as a point cloud in raw LIDAR data and the points of ground returns were gridded, averaged, and smoothed in order to create a high-resolution digital elevation model (DEM). Slope and aspect were calculated at 1 m^2 resolution from the gradient of the DEM in the longitudinal and latitudinal directions in the nine adjacent grid cells. The "northness" component of aspect was used in the present study (0-180 degrees), as N/S differences affect snow cover due to differences in solar forcing.

The snow-depth raster was calculated by differencing the snow on/snow off surveys from March 14th, 2010 [40]. The snow depth is Gaussian-distributed between 0 and 3 m (mean

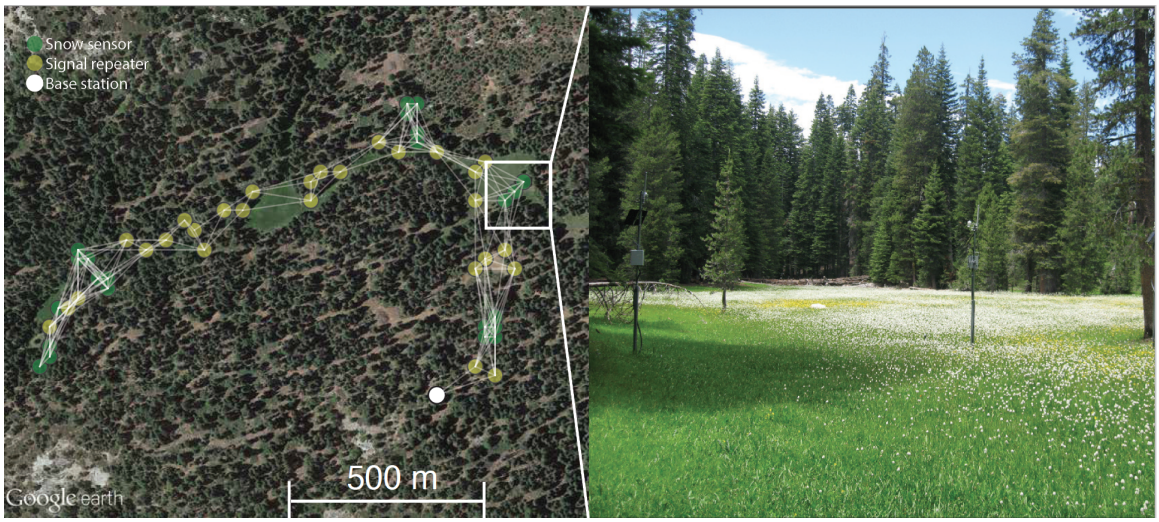

Figure 2.1: An existing wireless-sensor network at the Southern Sierra Critical Zone Observatory. Clusters of ultrasonic point measurements of snow depth are distributed across a 1.5-km transect in a 1-km^2 catchment. Data from each cluster is relayed every 15 minutes through a network of wireless elements to a base station with a real-time data uplink.

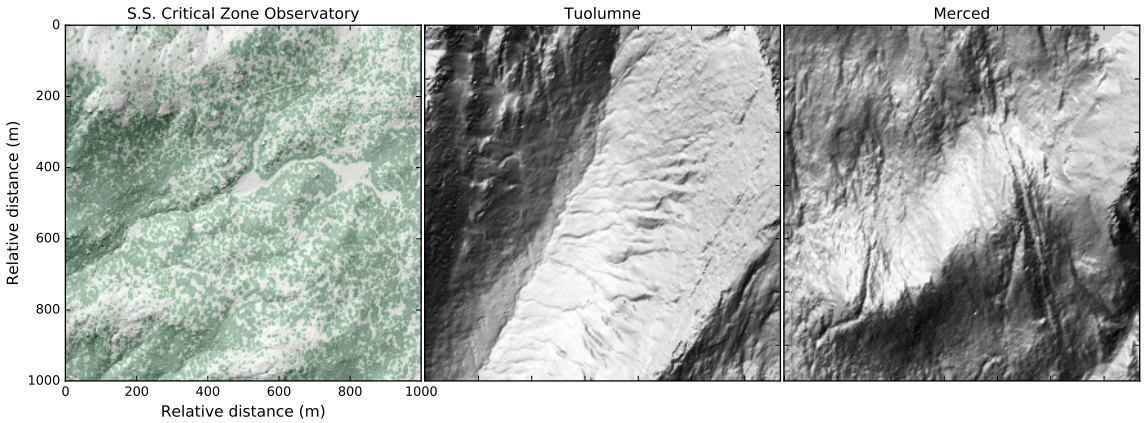

Figure 2.2: Hillshade maps showing the LIDAR-derived digital elevation models at each of the three sites used in the present study. At the Southern Sierra Critical Zone Observatory (left-hand panel), the distribution of canopy is overlain in green. The raster at the Southern Sierra Critical Zone Observatory was derived from LIDAR data in [40]. The rasters at the Tuolumne and Merced locations were derived from the JPL Airborne Snow Observatory [80].

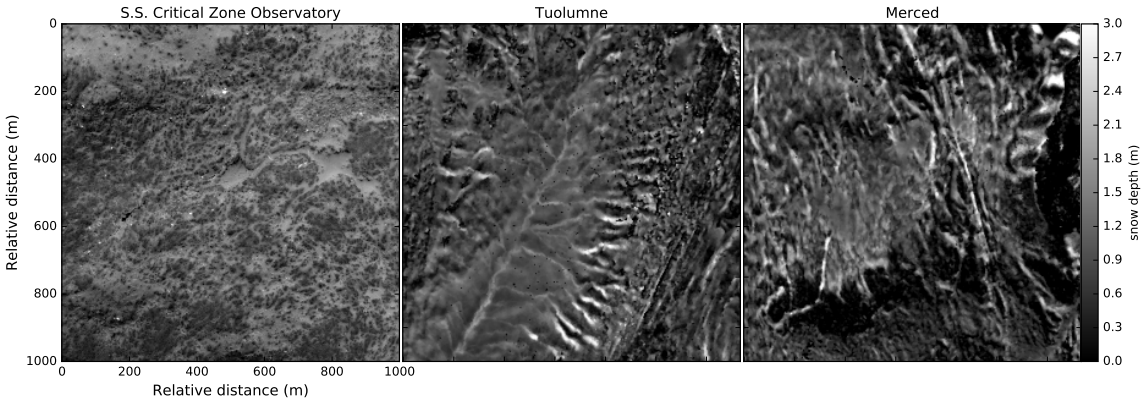

Figure 2.3: LIDAR-derived snow depth in each of the 1-km^2 catchments was used to determine optimal sensor locations for each site. At Tuolumne and Merced, these are the first rasters available for WY2014. The accuracy of the placements in the remaining LIDAR surveys is shown in Figures 2.9 and 2.10.

= 1.42, standard deviation = 0.50). A comparison against ground-truth surveys conducted during the LIDAR surveys showed 0.1 m of vertical error in the snow-depth raster. There were no returns in 28.5% region due to dense canopy. In these regions, trees were segmented from the LIDAR canopy height model using the variable-area local maxima algorithm and a typical pine tree height-diameter ratio [97, 109]. Of the regions with no returns, 2.3% were tree trunks. The snow depth at these points was set to zero and the remaining regions were gap-filled using bi-linear interpolation. The interpolation scheme was based on a prior analysis of snow-depth variability at a nearby instrument cluster [8], which determined that under-canopy measurements were typically 30 cm lower than drip-edge. Finally, the canopy penetration fraction from the LIDAR point cloud was calculated based on the methods presented in [112]. The raw and gap-filled LIDAR data are shown in Figure 2.4.

Sites for the long-term analysis were selected from within the Airborne Snow Observatory (ASO) [80]. The ASO mission runs LIDAR surveys every other week (beginning the last week of March) in the Merced and Tuolumne river basins of the Sierra Nevada. For the present study a 1 km^2 catchment was selected from within each basin. Both sites were selected from above the tree line. This provided two advantages for the present study: it permitted an evaluation of the transferability of the algorithm to a new environment (i.e. not densely forested), and meant that the snow rasters did not have to be interpolated under dense canopy. The accuracy of the algorithm could therefore be evaluated based solely on measured values. The Tuolumne site is located at (38.044591, -119.427359 degrees) at 3100 m a.s.l. and is due east from the Hetch Hetchy reservoir. The Merced site is located at (37.737945, -119.302983 degrees), 3450 m a.s.l. The 3 m resolution LIDAR rasters of site characteristics and snow depth are shown in Figures 2.2 and 2.3, respectively. At the Tuolumne catchment, the elevation range is 3085 m - 3340 m, the aspect is predominantly

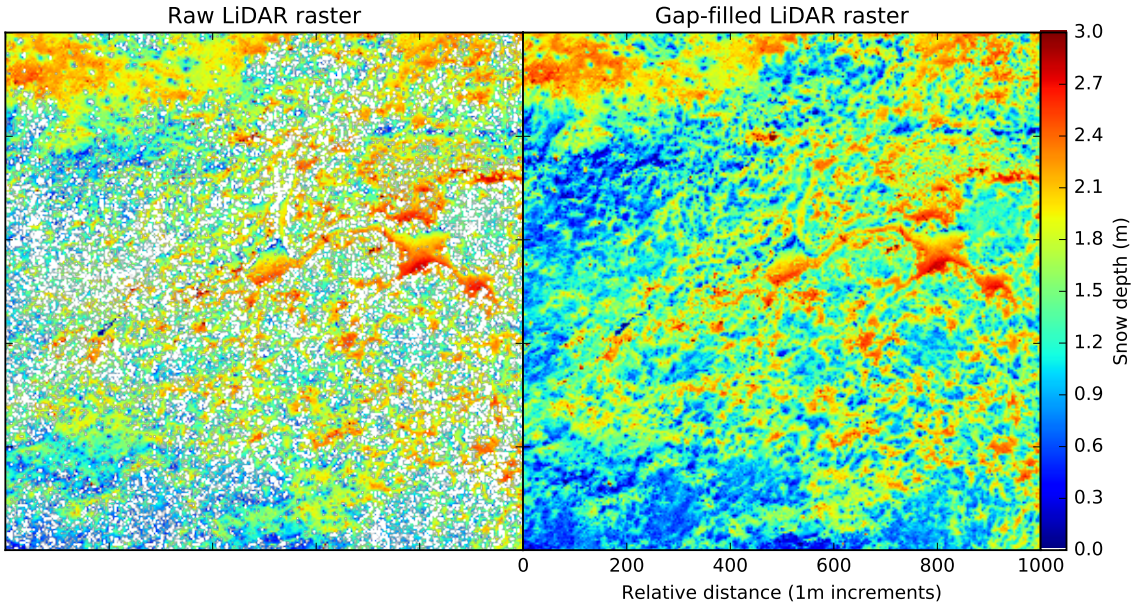

Figure 2.4: Raw (left) and gap-filled (right) LIDAR snow rasters at the Southern Sierra Critical Zone Observatory. Dense canopy resulted in no returns for under-canopy measurements in 26.2% of the catchment. These points were bi-linearly interpolated from adjacent measurements (described in Section 2.2). The snow rasters in Tuolumne and Merced did not require gap-filling.

west- and east-facing, and the mean and median slopes are 19.3 and 18.2 degrees, respectively. The Merced catchment spans 3270 m - 3680 m, the aspect is predominately south-facing, and the mean and median slopes are 20.0 and 17.5 degrees, respectively. Dates for the LIDAR surveys at each site are listed in Table 2.2. Based on the in-situ measurements from the Gin Flat snow pillow (located between the Merced and Tuolumne basins at 2150m), the differences in snow depth between the seven LIDAR scenes were primarily attributable to: accumulation (between scenes 1-2), melt (2-3), melt (3-4), accumulation (4-5), melt (5-6), accumulation and then melt (6-7).

Identification of representative sampling locations

In the first step of the proposed method, the distribution of sensors that is most-representative of the LIDAR-derived feature space is determined using a Gaussian mixture model, which assumes that a feature space (i.e., the combined $\mathbb{R}^6$, $\mathbf{x} = [x^{\text{lat}}, x^{\text{lon}}, x^{\text{slope}}, x^{\text{aspect}}, x^{\text{elevation}}, x^{\text{canopy}}]$ LIDAR data from Section 2.2) is a product of a finite number of latent (unobserved) components (i.e., sensors). The sensor's ability to observe each point in the feature space is represented using a multivariate normal distribution (Equation 2.1). This is the parametric expression of each component of the mixture. The expected value of such a normal distribu-

tion is the sensor's location in the feature space. Multiple Gaussian distributions (multiple sensor locations) were combined and weighted with mixing parameters from an ensemble of M mixture elements in Equations 2.2 and 2.3. The combined ability of all sensors to observe all points in the feature space is represented using a likelihood function (Equation 2.4).

Formally, the Gaussian mixture model is a linear superposition of D-dimensional multivariate normal distributions, $\mathcal{N}$, with expected value μ and covariance Σ applied to data $\mathbf{x}$, which are vectors including the four LIDAR derived variables. Each component is defined as:

$$\mathcal{N}(\mathbf{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma|^{1/2}}\exp\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\} \tag{2.1}$$

and their superposition, with mixing parameters, π_m, with m being an index denoting a single mixture component, is defined as:

$$p(x) = \sum_m = 1^M \pi_m \mathcal{N}(\mathbf{x}|\mu_m,\Sigma_m) \tag{2.2}$$

subject to:

$$\sum_{m=1}^{M}\pi_m = 1 \tag{2.3}$$

The complete log-likelihood function, evaluated over all the LIDAR-derived measurements, N is given by:

$$\ln\ \mathrm{p}(\mathbf{x}_n|\pi,\mu,\Sigma) = \sum_{n=1}^{N}\ln\Big\{\sum_{m=1}^{M}\pi_m\mathcal{N}(\mathbf{x_n}|\mu_m,\Sigma_m)\Big\} \tag{2.4}$$

The Expectation Maximization (EM) algorithm [64, 82] was used to retrieve the sensor locations. This is an iterative process in which the algorithm tries to recover the most-likely parameter estimates for the mixture of multivariate $\mathcal{N}_m$ to explain the data. A spherical covariance function was used and the model weights, covariance, and means were updated with each iteration. Once the maximization step no longer increases the log-likelihood, the process terminates and the optimal sensor locations have been found. Like many gradient-based optimization methods, EM converges to local minima. Therefore, 100 initializations from randomized starting points were used to select the result with the maximum likelihood. This process was parallelized onto four computational cores using a subsampled feature space (1-in-16 point subsampling to make the evaluation of many starting points computationally tractable). Then a nearest-neighbor search was performed through the full feature space (i.e. not subsampled) in order to find the physical location that most closely matches the features of each Gaussian mixture model mean.

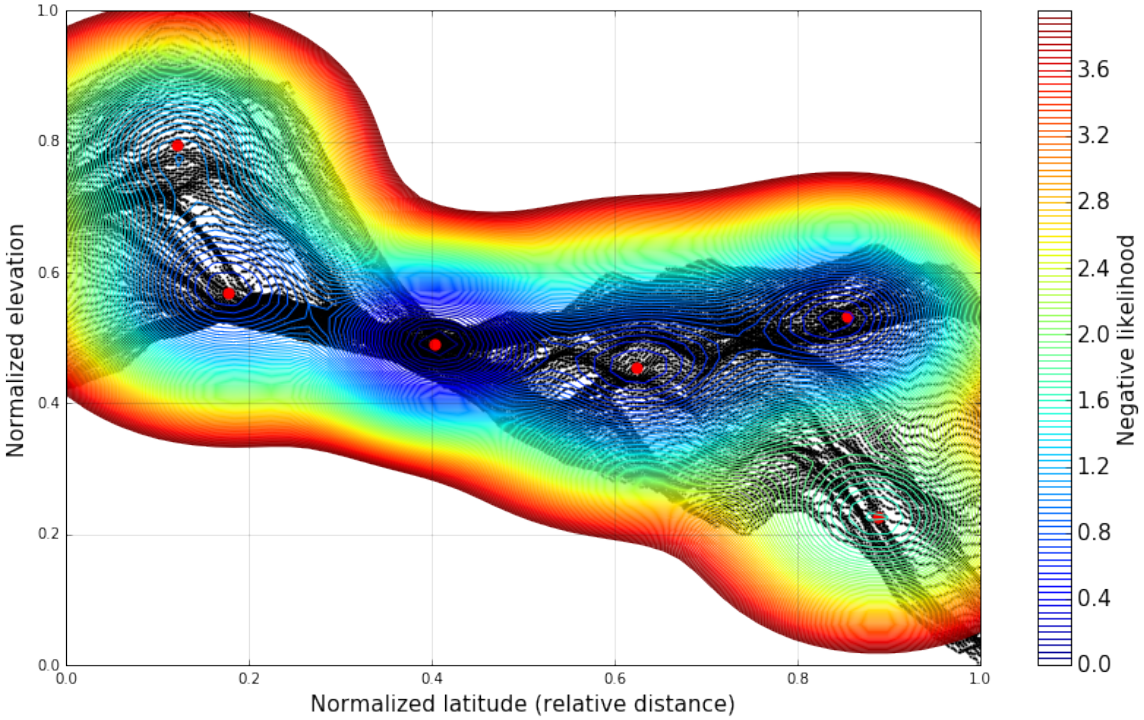

Figure 2.5: Representative sensor locations were determined using a Gaussian mixture model in the space of independent variables from Figure 2.2. Black points represent a 2-dimensional projection of the $\mathbb{R}^6$ LIDAR data onto the latitude-elevation plane. Optimal sensor locations (indicated by red points) are the expected values of the latent Gaussians. The likelihood function (contour lines) quantifies how well each point in the LIDAR-derived feature space is observed given the locations and covariance of each sensor (see Equation 2.4). The optimal parameters (expected value and covariance) for the model were determined using the EM algorithm.

Snow-depth model

In the second step, the LIDAR-derived snow-depth measurements at the locations proposed by the Gaussian mixture model were used to estimate the distribution of snow depth across the catchment using a Gaussian process [85]. A Gaussian process is a regression technique that predicts a dependent variable (in this case, the snow depth, Y) using a set of independent variables, X, which are expected to be informative for its prediction. It establishes a covariance-based model that, using known input-output relationships (the training data at the measuring stations locations, $\{\mathbf{x}_n, y_n\}_{n=1}^{N}$), is able to predict the snow depth on new, unseen locations where the inputs can be obtained. In this application, the physiographic variables (slope, aspect, elevation, and canopy) were used as independent variables in the estimation (i.e., for a single point in $\mathbb{R}^4$, $\mathbf{x} = [x^{\text{slope}}, x^{\text{aspect}}, x^{\text{elevation}}, x^{\text{canopy}}]$). The Gaussian

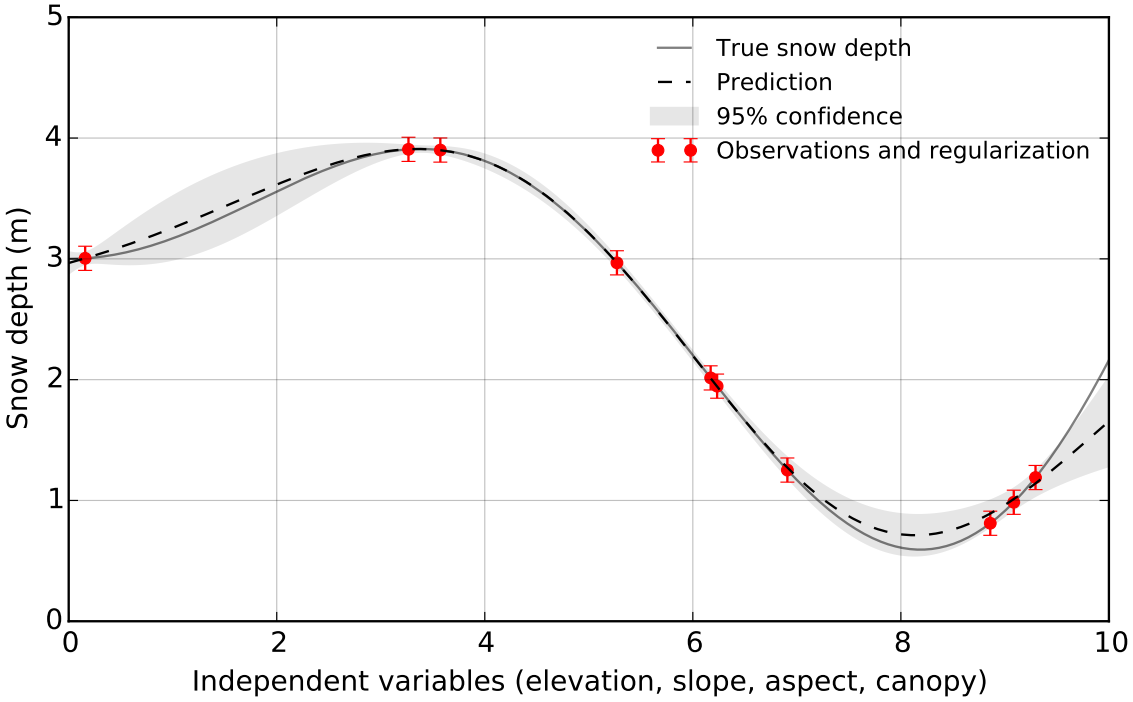

Figure 2.6: Illustration of a Gaussian process shown in one of the four dimensions. Uncertainty in the observations (0.1 m in LIDAR data) is quantified with a regularization at each point (Equation 2.7). The confidence of the prediction at intermediate points is a function of the covariance parameters estimated from measurements at the sensor locations. The estimated uncertainty is employed by the supervised updating process.

process (Equation 2.5) combines the N point measurements $\mathbf{X}^N$ from each sensor station and the four LIDAR-derived physiographic variables to estimate the mean snow depth, $\bar{Y}$, using a covariance function K,

$$Y \sim GP(\bar{Y}, K), \tag{2.5}$$

where the $\sim$ symbol means 'is distributed following'. An illustration of the Gaussian process is shown in Figure 2.6: given the observed data and uncertainty (sensor stations, in red) and a prior over the shape and parametrization of the functions which are likely to be observed (e.g., all functions are Gaussians with mean $\bar{Y}$ and covariance K), the Gaussian process estimates a posterior distribution with a predictive average (dashed line) and variance (the uncertainty of the prediction – the grey envelope). Note that close to the observed data, the variance approaches zero. Readers interested in the mathematical details may refer to Chapter 2 in the book by [85]. In the present study, a squared exponential covariance function (K_{se} in Equation 2.6) was used, which depends on four variables: $\mathbf{x}$ and $\mathbf{x}'$ are two points in the domain, d is the distance between them, and l is the characteristic length scale

(computed individually for each independent variable in the normalized feature space). This covariance was used because the snow depth was assumed to vary smoothly with respect to the input variables. The autocorrelation in the model is controlled by the l parameter. The algorithm has no prior information about this parameter; it must be estimated only from the point measurements at each sensor. Regression weights and autocorrelation are estimated using a constrained optimization by linear approximation algorithm in the software package SciPy (fmin cobyla) [82]. The error at each measurement (10 cm based on the LIDAR error) was quantified using the regularization in Equation 2.7, where y_i is the measured snow depth at point i, and σ_i is the measurement variance at y_i.

$$K_{se}(x, x') = exp(-\frac{||x - x'||^2}{2l^2}) = exp(-\frac{||d||^2}{2l^2}) \tag{2.6}$$

$$r_i = \left[\frac{\sigma_i}{y_i}\right]^2 \tag{2.7}$$

Supervised updates

In addition to estimating the spatial mean of the snow depth, the Gaussian process estimates the distribution of model uncertainty from the covariance matrix of the posterior distribution, which is a function of the independent variable weights and estimated autocorrelation. This provides a basis for placing additional sensors with the aim of minimizing uncertainty throughout the catchment. Using the measurements from the sensors placed in Section 2.2, the snow depth was modeled using the Gaussian process, and a new feature space corresponding to the highest 1% and 10% of model uncertainty was computed. The Gaussian mixture model was then used in each feature space to select additional sensor locations that both reduce the uncertainty of the model and are unique with respect to the combination of topographic variables they sample. The additional placements are considered supervised updates because they rely on observations of the dependent variable (i.e., the estimated autocorrelation and regression weights in the snow-depth model) to determine optimal locations. In the machine-learning community, the procedure is known as "active learning" [91]. They are distinct from the placements in Section 2.2, which are unsupervised in the sense that no observations of the dependent variable (snow depth) are used in their determination. Sensors placed in regions of high uncertainty will reduce the uncertainty throughout the domain at points with similar combinations of physiographic features.

Model evaluation

The error in the snow-depth model was determined by differencing the predicted snow depth under each scenario from the LIDAR-derived snow-depth raster (Figure 2.3). The accuracy of the snow-depth model under each placement scenario was computed using two standard metrics: RMSE (Equation 2.8) and Bias (Equation 2.9). In each equation, n is the number of points in the model, $\hat{y}_i$ is the model prediction at point i, and y_i is the true snow depth.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{2.8}$$

$$\text{Bias} = \sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{2.9}$$

Optimal number of sensors

The number of sensors required in the unsupervised step was determined by examining the error in the snow-depth model in a range of placement scenarios. In the first set of scenarios, the unsupervised placements ranged between 2 and 23 sensors. The optimal number of sensors was taken to be when the marginal improvement in RMSE was less than 10%. Given the placements from the unsupervised process, sensors were added using the process described in Section 2.2, up to the budget of 23 sensors. Finally, with the optimal number of sensors determined, the snow-depth estimate resulting from the proposed placements was compared to an equivalent number of randomly-chosen, but spatially-distributed sensors. The snow-depth model under 100 randomized configurations was evaluated to determine the expected accuracy of the randomized placements.

Long-term accuracy

Only one LIDAR survey was available for the region covered by the existing wireless sensor network, which is insufficient to assess the long-term accuracy of the algorithm. To address this, two regions within the JPL Airborne Snow Observatory were chosen (described in Section 2.2) that had multiple LIDAR surveys during WY2014. First, the unsupervised placements were selected based on the independent variables at each site (described in Section 2.2). Second, the supervised update step (described in Section 2.2) was performed in the first snow-depth raster of WY2014. The accuracy of the model estimated from the sensors in seven additional LIDAR surveys for each catchment was quantified using the bias and RMSE in each LIDAR survey. Results were then compared to the average of 100 random configurations of sensors. Finally, the ability of the sensors to represent the catchment-scale mean and standard deviation of snow depth was evaluated in each of the LIDAR surveys.

2.3 Results

Using the methods described in Section 2.2, six sensors were selected in the unsupervised step for the SSCZO catchment. The converged Gaussian mixture model for this configuration is shown in Figure 2.5. Optimal sensor placements (the expected values of the latent Gaussians) are shown as red markers. The likelihood function (quantifying how well the space of independent variables is observed under the current sensor configuration), is shown

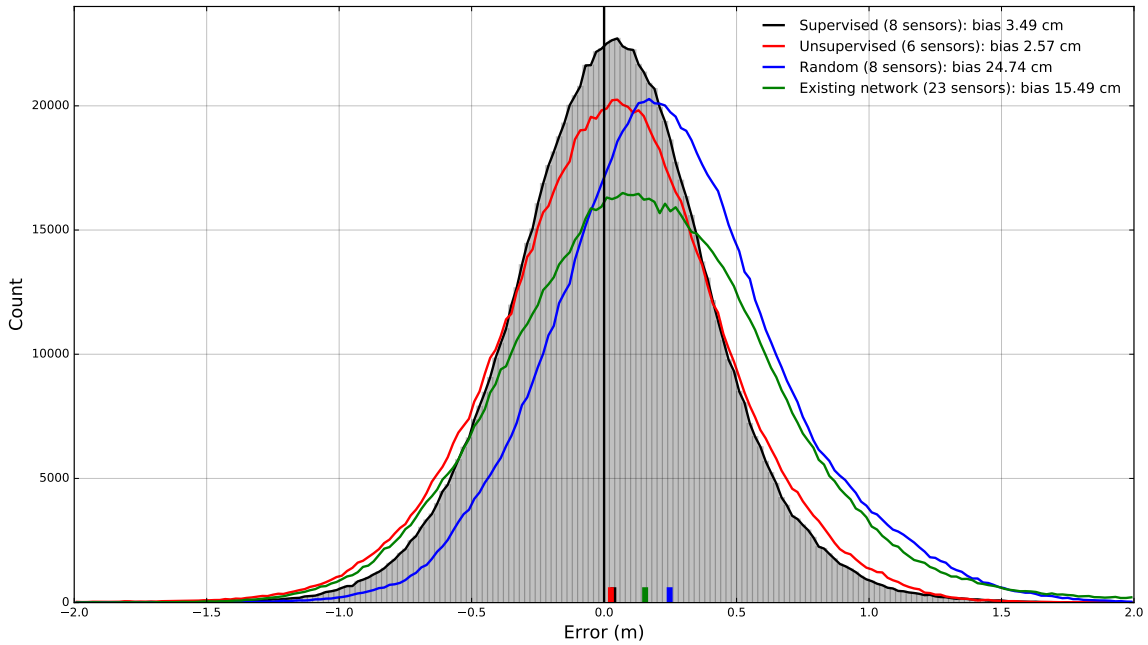

Figure 2.7: Histograms showing the distribution of error in the snow-depth model under each sensor configuration from Figure 2.8. The placements proposed by the machine learning algorithms have lower bias and RMSE than the eight random sensors and the existing network. The RMSE is lowered in the supervised update, but the bias slightly increases compared to the unsupervised placements.

as contour lines. Points that are well observed are shown in blue; poorly observed points are shown in red. As it is not possible to show the full six-dimensional feature space, Figure 2.5 illustrates a two-dimensional projection of the output. Adding more than six sensors in the unsupervised step did not significantly improve the accuracy of the algorithm (see Table 2.1). With 6 sensors, the accuracy of the algorithm was 43.23 cm; at 23 sensors, this was reduced only to 42.90 cm. Intermediate values were variable, and in one case as high as 73.88 cm (at 14 sensors). Two sensors were then placed in the supervised step (using the 10% uncertainty threshold), reducing the RMSE to 38.29 cm. The effect of the update is shown in Figure 2.7: the error becomes more tightly distributed around zero, but the bias remains essentially unchanged (it increases by slightly less than one centimeter).

In the 100-sample evaluation of randomized placements of eight sensors, the average RMSE was 63.7 cm. The distribution was heavy-tailed: most of the results were clustered between 40-75 cm RMSE, but there were eight outcomes with greater than 100 cm RMSE. A configuration corresponding to one of the average runs is shown in the middle panel of Figure 2.8. The spatial distribution is similar to the proposed placements, however the output of snow-depth model reveals a slight over-estimation of the snow-depth, particularly in the SW regions of the catchment (right-hand panel in Figure 2.8), resulting in a positive bias for the

Table 2.1: RMSE, bias with increasing number of sensors in the unsupervised step at SSCZO.

Sensors:	Bias (cm)	RMSE (cm)
5	35.58	277.5
6	2.57	43.2
7	7.47	52.8
8	9.00	63.8
9	0.16	59.3
10	-0.23	62.7
11	-9.73	52.6
12	-14.15	46.0
13	-9.11	56.0
14	-3.52	73.9
15	-4.67	63.4
16	-3.02	47.1
17	-4.32	66.3
18	3.58	52.4
19	-16.79	65.2
20	4.67	43.3
21	-8.25	43.9
22	-9.74	46.6
23	5.41	42.9

estimator (blue line in Figure 2.7).

The snow-depth model estimated from the existing 23 placements has higher error (RMSE 53.0 cm, bias 15.49 cm) than the estimates from the machine-learning method. The accuracy of the snow-depth model is high (less than 15-cm error) near the sensor clusters. However, the error is very high (greater than 1-m error) in the NW and SW region, where the snow-depth model over-estimates the true snow depth. This produces a more heavy-tailed error distribution than the proposed or random placements, and an overall overestimate of the true snow depth within the catchment (green line in Figure 2.7). Despite this, the true mean (1.42 m) of the snow-depth raster are better estimated by the sensors in the existing network. The mean and standard deviation of the sensor measurements is 1.76/0.49, 1.54/0.52, and 1.48/0.57 m for the proposed, random, and existing network respectively.

In the Tuolumne ASO site, 11 sensors were placed in the unsupervised step, and 4 were added in the supervised step (using the 1% uncertainty threshold). The bias and RMSE of the 15 locations proposed by the algorithm in the first LIDAR scene are higher than at SSCZO (left-hand panel, Figure 2.9). This negative bias persists and decreases in magnitude in the remaining LIDAR surveys (top line in Table 2.2) and the RMSE is bounded between 64.2 cm and 70.6 cm in the remaining surveys. The bias and RMSE in the average of the 100 of the random trials are worse throughout the remaining LIDAR surveys (second

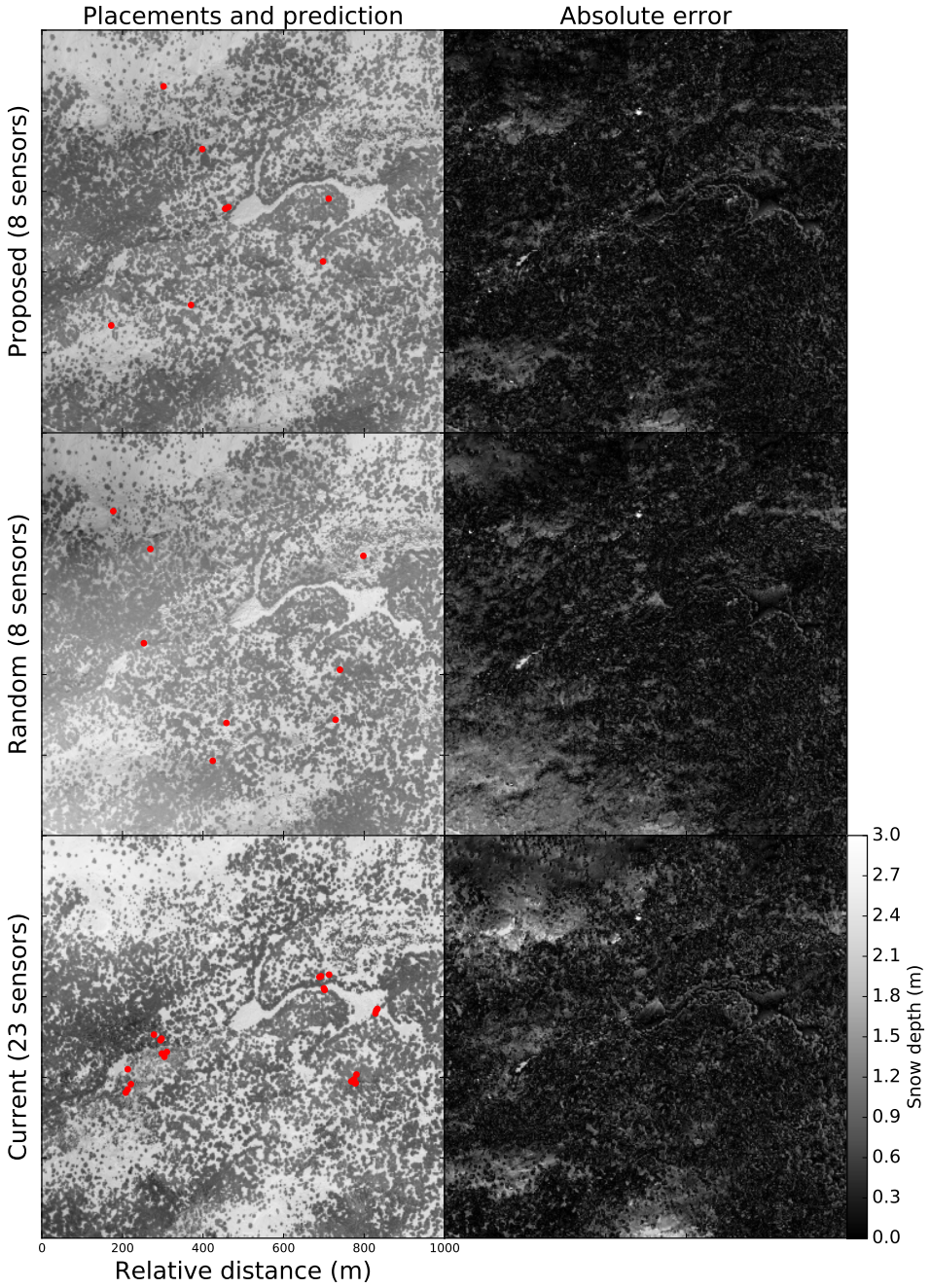

Figure 2.8: Predicted snow depth (left-hand panels) and error (right-hand panels) using three sensor configurations: proposed (top), random (middle), existing (bottom). The resolution of the snow-depth model is 1 m^2, the scale in the left-hand panel is 0-3.0 m. The error of each model is computed relative to the LIDAR snow-depth raster in Figure 2.3. The distribution of errors is compared in Figure 2.7.

Table 2.2: Long-term error and snow-depth statistics in ASO LIDAR data (all values cm).

Tuol	3-23-14	4-07-14	4-13-14	4-20-14	4-28-14	5-02-14	5-11-14
Bias:	-12.2/44.4	-15.3/47.3	-9.9/43.4	-13.9/35.4	-3.5/50.4	-5.6/27.3	-10.3/37.8
RMSE:	60.9/89.0	70.6/108.2	64.5/97.5	66.4/80.3	65.5/105.6	64.2/74.9	64.2/76.4
Mean:	115/101/136	153/130/172	127/105/145	110/75/125	124/97/142	94/69/115	85/59/107
Std	58/30/88	65/35/109	62/35/98	64/44/105	65/41/102	61/37/96	61/40/95

Merced	3-24-14	4-06-14	4-14-14	4-23-14	4-29-14	5-03-14	5-12-14
Bias:	0.4/-1.5	-2.4/-6.8	6.1/9.0	0.8/17.1	-5.5/12.8	0.6/6.4	5.2/12.6
RMSE:	58.0/70.1	73.1/82.5	65.1/84.0	62.5/87.7	66.6/74.5	61.6/74.0	59.4/81.1
Mean:	76/84/67	114/117/100	73/80/68	67/75/65	77/77/68	71/75/66	54/61/48
Std	58/49/48	69/64/63	61/52/49	61/54/52	66/60/58	63/59/52	57/49/46

number in top two lines of Table 2.2) and there is greater variability in both quantities. Neither the proposed nor the random placements capture the small-scale variability along the drainages in the center of the catchment (see spatial distribution of error in the right-hand panels, Figure 2.9). Although the machine-learning algorithm consistently outperforms the average of the random placements in estimating location-specific error, the correspondence between the measured snow mean and the snow variance is inconsistent (bottom two lines in Table 2.2). The mean and standard deviation are generally underestimated by the proposed placements and overestimated by the random placements.

In the Merced ASO site (Figure 2.10), 13 sensors were placed in the unsupervised step, and 1 was added in the supervised step (using the 10% uncertainty threshold). The RMSE and bias of the 14 locations proposed by the algorithm are lower than at Tuolumne (top lines in the "Merced" subsection of Table 2.2). The bias starts near zero and does not deviate above 6.1 cm. By contrast, the bias and RMSE resulting from the average of the random trials are high and variable throughout the remaining LIDAR surveys (second number in the top lines the "Merced" subsection of Table 2.2). Again, much of the small-scale variability in the catchment is not captured by either the random or proposed placements. The relationship between the mean and standard deviation of the sensor measurements compared to the catchment-scale statistics is again inconsistent (bottom two lines the "Merced" subsection of Table 2.2).

2.4 Discussion

Although individual point measurements of snow depth are poor estimators of the local mean, these results indicate that an automated process can be used to find a limited combination of representative placements that can be used to estimate the catchment-scale snow cover. Eight placements in strategic locations produced a better catchment-scale estimate than 23 placements in the existing network, likely due to the under-representation of high values of slope and elevation and the closer spatial distribution of the existing sensor clusters. The long-term analysis indicated that the method is transferable to other types of environments, and that the accuracy of the sensors placement determined from the first LIDAR survey

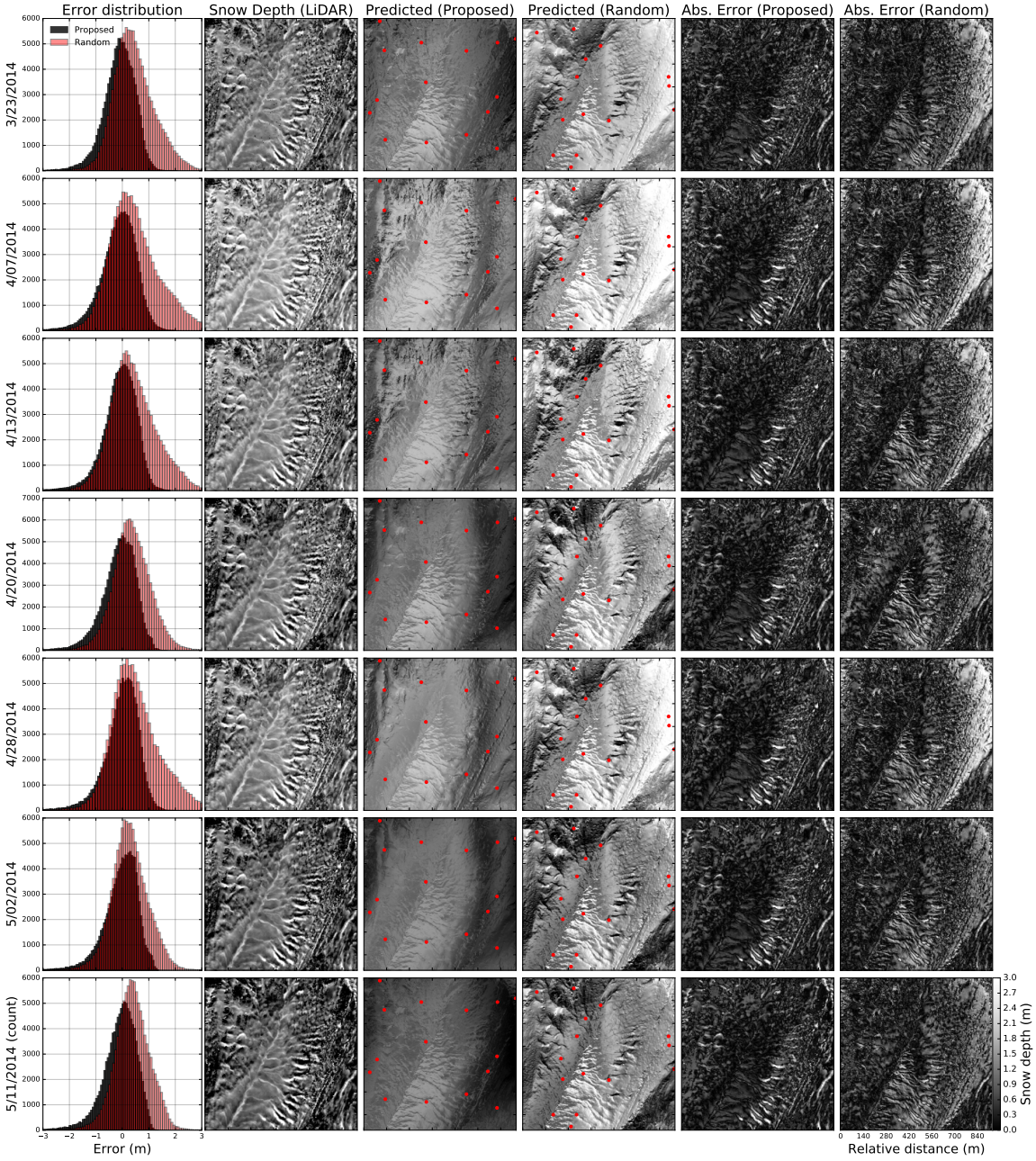

Figure 2.9: Snow-depth model accuracy (comparing proposed placements to the average of 100 random placements) in the remaining 2014 LIDAR surveys at Tuolumne. Sensor placements are overlain as red points.

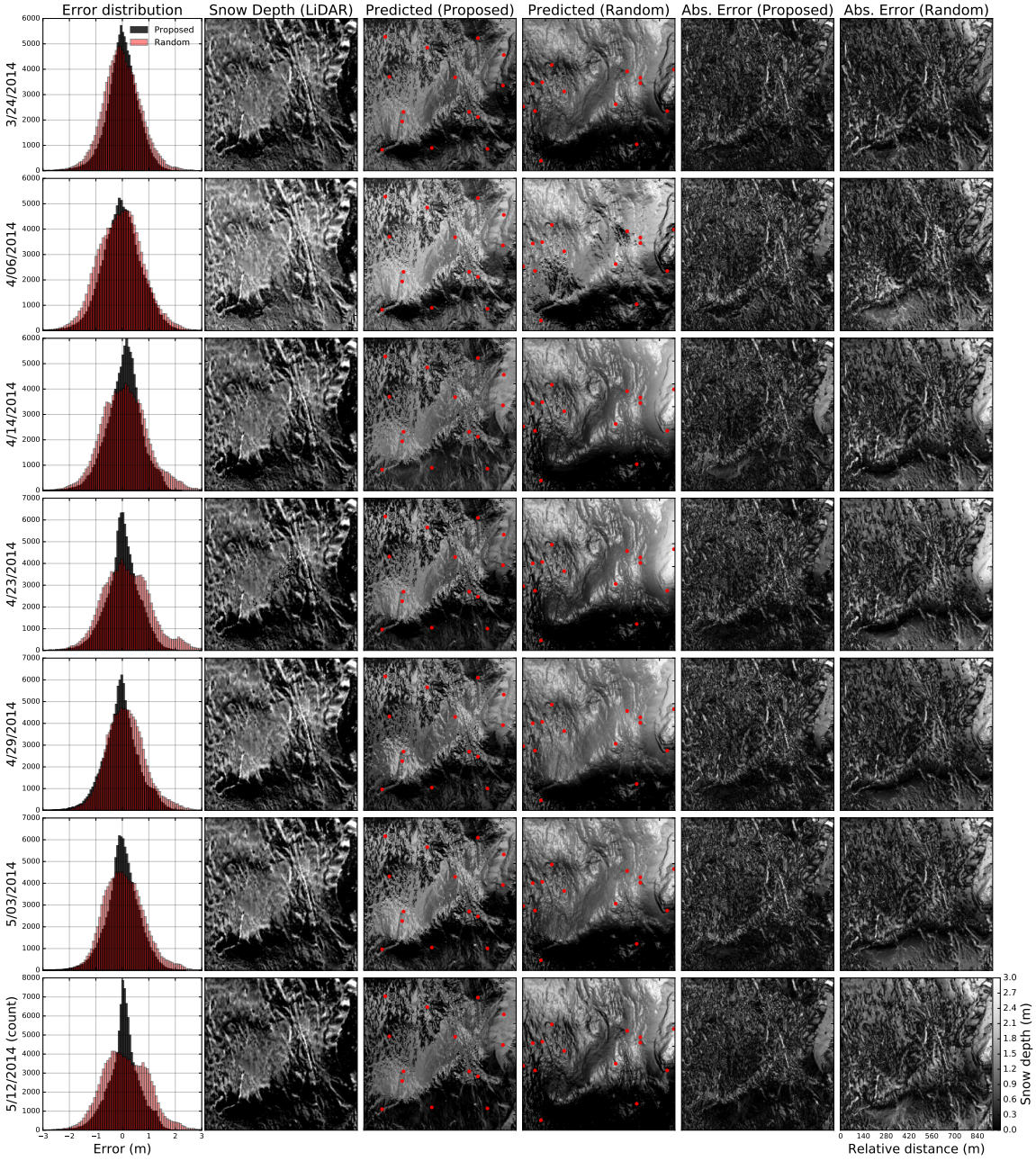

Figure 2.10: Snow-depth model accuracy (comparing proposed placements to the average of 100 random placements) in the remaining 2014 LIDAR surveys at Merced. Sensor placements are overlain as red points.

persists in the remaining surveys. This was not true of the random placements, which showed greater variability in the bias and RMSE in the remaining surveys.

At the ASO catchments (Tuolumne and Merced), the RMSE of the snow-depth model was higher than at the SSCZO, despite the additional sensors placed by the algorithm. This is likely due to the greater exposure of the terrain above the tree line in the ASO sites. At the SSCZO, the highest errors are in exposed (i.e. un-vegetated) areas in the NW, SW, and meadows (right-hand panels, Figure 2.8). At the ASO sites, the highest errors are distributed throughout the catchments (right-hand panels in Figures 2.9 and 2.10). The dense canopy closures at SSCZO likely constrain the wind redistribution effects to smaller regions within the catchment, thereby reducing the overall error and required number of sensors. It should also be noted that many of the under-canopy measurements in the SSCZO raster were interpolated. This may have affected the relative error and required number of sensors between the ASO and SSCZO sites by producing a more homogeneous distribution of under-canopy snow depth.

The present study is conducted in small catchments sampled at very high spatial resolution, whereas typical hydro-metrological studies are conducted at lower resolution (e.g. spatial resolutions on the order of 50m - 100m). Running the algorithm at lower resolution will likely affect the accuracy of the spatial snow cover model and placements determined by the algorithm. If the above analysis is repeated with the LIDAR data averaged to 50m resolution, the accuracy of the model increases and fewer sensors are placed by the algorithm. The variance explained by the low-resolution model in the first snow rasters for SSCZO, Merced, and Tuolumne is 56%, 59%, and 36% with 8, 10, and 8 sensors, respectively. Using the high-resolution data, the variance explained 45%, 6%, and 18% with 8, 14, and 15 sensors, respectively. This is likely due to small-scale variability being averaged out at lower resolutions. The temporal persistence of the accuracy is also observed in the 50m data: all snow cover models chosen by the algorithm had lower RMSE than the average random configuration. The range of model accuracies in the present study is consistent with the findings in [26], which analyzes the accuracy of multiple snow cover models in three 1-km^2 catchments. The study finds the models explain between 6.8% and 31% of the snow depth variance, depending on the specific catchment and statistical model.

The accuracy of the high-resolution model in the present study may be improved by adding independent variables that capture the effects of small-scale terrain features and wind redistribution. None of the algorithms in the ASO catchments accurately modeled the accumulation of snow depth near sharp transitions in the DEM, (e.g. along the drainages in Figure 2.9, and along the ridges of Figure 2.10). Similarly, none of the algorithms captured the increasing, S-facing gradient of snow depth across the meadow in the Southern Sierra Critical Zone Observatory, which was likely a result of wind redistribution. This effect has been observed in a number of prior studies [108, 67], and directional variables have been suggested to account for directional redistribution of snow (e.g., in [67]). Given the dense canopy throughout the catchment in the present study, a directional variable was not included in the feature space. In catchments with less dense distributions of canopy, it would likely be necessary. Care should be taken in mixed open/forested regions that the directional bias

measured in open regions is not translated to the regions with dense canopy.

It would be beneficial to evaluate the proposed method in a vegetated region over long time spans in which there are regular under-canopy measurements (from an in-situ network or synoptic survey) and multiple LIDAR flights. The accuracy of the algorithm may exhibit greater temporal variability in such environments, since the influence of canopy is different throughout the year (affecting snow interception during the accumulation phase, and long-wave radiation enhancement during the ablation phase). The method should also be evaluated in alpine catchments that exhibit lower temporal persistence in the spatial snow cover. The coefficients of determination between the first raster and successive rasters at Tuolumne and Merced are: 0.82, 0.81, 0.74, 0.80, 0.76, 0.73 and 0.93, 0.93, 0.86, 0.91, 0.90, 0.87, respectively. The long-term bias and RMSE may exhibit greater variability in catchments with lower temporal persistence.

Recent studies have demonstrated that mutual-information-based sensor placement strategies can outperform placements based on Gaussian process uncertainty [56]. This approach was not used in the present study as the mutual information algorithm is NP-complete with complexity $O(kn^4)$ (where k is the number of sensors and n is the number of possible locations). The fourth-order dependence on the number of potential locations can yield computationally intractable problems when using LIDAR data. The number of candidate placements in the numerical studies in [56] is on the order of 100-200, compared to over 1,000,000 in the present study. The mutual information-based strategy may be feasible to implement at lower spatial resolutions or by reducing the complexity using sub-modularity (discussed in [56]). It therefore warrants further research for this application.

There are practical considerations in observatory design that are not considered in the present study. Regions of the catchment may be inaccessible due to terrain attributes and other access constraints. This can be addressed by defining a set of inaccessible placements in the feature space from Figure 2.5. If the optimal sensor location is determined to be on an inaccessible grid element, a search to the nearest viable point in the feature space will be output as the optimal point. The proposed algorithm also requires a greater spatial distribution than the clustering approach. Recent field deployments have indicated that 1-km-scale wireless-sensor networks can be deployed using existing hardware in a variety of terrains. If the spatial extent is limited, the spatial coordinates could be removed from the unsupervised step, and an algorithm could determine the most-representative, spatially proximate distribution of nodes.

It should be noted that the "supervised updating" step in the present study would require two field deployments: one to gather data to estimate the distribution of uncertainty throughout the catchment (which also requires knowledge of the catchment-scale independent-variable distributions), and a second to add sensors in high-uncertainty regions. In practice, the marginal gain from the supervised updates may not be worth the marginal cost, though in this analysis, the supervised placements outperform un-supervised placements. In the present study, the first LIDAR flight was used to estimate the uncertainty and perform supervised placements. The selected locations may change depending on which raster is used, as the autocorrelation and regression weights will change throughout the year.

Given additional years of LIDAR data, it may be better to perform supervised placements based on a long-term analysis of the uncertainty within the catchment.

The true optimal number of nodes per site should be determined by considering the marginal value of the improved information. Combined with an understanding of the marginal cost of each additional placement, this would determine an optimal number of sensor nodes by comparing marginal cost and value functions. This approach would capture site-specific cost/value tradeoffs for each sensor network.

2.5 Conclusion

The research reported here suggests that a machine-learning algorithm can be used to identify snow-sensor locations in catchment-scale observatories prior to field deployment. In the three regions considered in the present study, the placements determined from the algorithm exhibited higher accuracy and less bias than an existing sensor network and an equivalent number of randomly selected locations. The accuracy of the algorithm was found to be consistent when the temporal transferability was evaluated in fourteen LIDAR rasters within the Airborne Snow Observatory.

The aim of the present study was to determine sensor locations to be used together with a snow-depth regression model to estimate the catchment-scale distribution. Without the Gaussian process estimate, the mean and standard deviation of the sensor measurements alone was not consistently more accurate than random or existing placements. Therefore the proposed algorithm is likely better suited to situations in which the snow depth is to be modeled at the catchment scale, as opposed to distributed models which take the mean and standard deviation (or variance) as input.

Chapter 3

Repeater Placement Strategy for Mesh Networks in Complex Terrain

3.1 Introduction

The prior chapter evaluated a sensor placement strategy for wireless snow-depth observatories. Once these sensor locations are determined, signal repeaters are needed to fill in data links too far apart to carry from sensor station to sensor station. These placements must be made to maximize the resiliency of the network by ensuring multiple data path redundancy allowing for the case of one or more elements failing during the winter season. The total number of placements is limited by cost and link capacity of the base station manager, so it is desirable to take advantage of terrain features to minimize the number of nodes placed.

This problem is not unique to snow-sensor networks: wireless sensor networks are increasingly being used to make distributed measurements in a number of fields [4]. It is often difficult to access these networks, so the initial deployment must ensure a redundant mesh topology to guarantee year-round access to the real-time data. Existing methods for optimizing wireless mesh topologies employ simplifying assumptions of a flat environment and symmetric, planar models of node connectivity [66]. In complex terrain, these assumptions do not hold: small movements of nodes in the wireless network can drastically affect the connectivity of adjacent nodes due to changes in terrain attributes (e.g. path intersection with terrain and vegetation).

In practice, these networks must be structured in the field in real time, using feedback from the field teams and diagnostic networking software [50]. Nodes are simultaneously reconfigured by personnel at strategic locations until the network is fully connected and each link has sufficient signal strength. The current approach is limited in that it relies on personnel with incomplete knowledge of the overall network structure to make their best guess for the local placements, without considering the effect on the broader network topology. The amount of time required for the network statistics to stabilize is on the order of many hours to days, limiting the number of combinations that can reasonably be evaluated.

Given the large geographical area to be covered, this process is time intensive and can result in sub-optimal network topology formation, resulting in data loss [87].

This chapter presents a four-step methodology for structuring redundant wireless mesh networks prior to field deployment using high-resolution remote sensing of the deployment site (LIDAR data): (i) packet delivery ratio (PDR) is measured on four channels at representative locations across the deployment region; (ii) the PDR data are combined with LIDAR (airborne laser scanning) elevation and canopy data to train an algorithm that estimates the PDR between any new pair of points in the domain; (iii) the link model is used to evaluate the connectivity of thousands of potential repeater placements; (iv) a minimal set of repeaters is found using a fast approximation algorithm for the Steiner tree problem and redundancy is added by adding repeaters at each articulation point to maximize connectivity.

The chapter is divided into four sections: Section 3.2 describes how the PDR measurements and LIDAR were collected; Section 3.3 describes the link model; Section 3.4 describes the topology optimization component; Section 3.5 concludes the chapter with a comparison of the algorithm's placements to the manually configured network described in [50].

3.2 PDR measurements and LIDAR data

First, LIDAR data were collected for the region around the SSCZO network from the prior chapter. These data were used to make 1 m^2 digital elevation and canopy models. The LIDAR survey was conducted from August 5th to August 15th, 2010. Point-cloud data were processed and classified by the National Center for Airborne Laser Mapping (NCALM), which is hosted on the NSF OpenTopography server. The study area was clipped from the point-cloud dataset for calculating the above-one-meter vegetation density. The point clouds were first grouped as ground returns or vegetation returns by using their classification attribute. Second, the elevation difference between each point was calculated by iterating through all vegetation data points and measuring the height to the nearest neighbor among all the ground returns. Finally, the vegetation data were filtered such that only the points above 1 m were retained. This process was repeated for the 1-10 m height range. Binary rasters were created for both vegetation layers, which were combined to create the final vegetation model. Thus, the final model is more heavily weighted for vegetation in the 1-10 m range, which is more likely to affect signal strength.

PDR measurements were then collected within the existing network. Antennas (L-COM HGV-2404U) were placed on lightweight poles that were measured to be the same height as the poles in the existing network (3 m). The data were collected on 4 channels (2.405, 2.425, 2.445, and 2.465 GHz) using the "radiotest" module for Dust Networks chips (github.com/dustcloud/smartmeshsdk/tree/master/app/RangeTest). The transmitter mote is set to continuously send a 40B packet every 5ms. The transmitter and receiver were placed in strategic locations throughout the observatory, with the intention of sampling across a representative range of distances, vegetation, and angle. Each transmitter and receiver loca-

Figure 3.1: Packet delivery ratio was measured using L-COM HGV-2404U 4dBi antennas at representative locations near the existing network.

tion was measured with a handheld GPS unit. Figure 3.1 shows the transmitter and receiver measurement setup, along with a cluster of nodes in the existing network.

Figure 3.2 shows where all measurements were taken, along with the average PDR along each path. The top panel shows a hillshade map of the 1 m^2 digital elevation model. The lower panel shows the same measurements with the canopy model. Darker green regions in the figure show where the canopy is denser in the 1-10m height range.

Collecting these measurements provided some intuition for how terrain properties affect the signal strength. First, when there was line of sight between the two antennas, PDR was high on all channels even for long links. For example, Figure 3.2 shows a link on the far right which received [639, 876, 929, 881] packets on each channel for the 2.405, 2.425, 2.445, and 2.465 GHz channels, respectively. The maximum number of packets received was 973, which was taken as 100% PDR. Removing line of site at large distances strongly affected PDR. Moving the receiver by 5 m (thereby introducing more vegetation along the path), caused the packets received to go to zero on all channels. Second, PDR was found to be extremely variable when there was significant vegetation along the path. While longer links were observed, high PDR measurements were rarely seen for links longer than approximately 50 m when dense vegetation was present on the path. Finally, angle was not observed to have a strong effect on PDR. This was tested by tilting the receiver to 0, 30, 60, and 90 degrees at a fixed distance. Angles up to 60 degrees did not result in significant degradation with regard to how many packets were received: [910, 965, 947, 894] for 60 degrees versus [877, 908, 869, 854] for zero degrees. At 90 degrees (antenna approximately 2 m off the ground, flat), a PDR decline was observed: [376, 447, 690, 264] packets were received in this configuration. There was also not a significant PDR drop when the transmitter and receiver were mounted

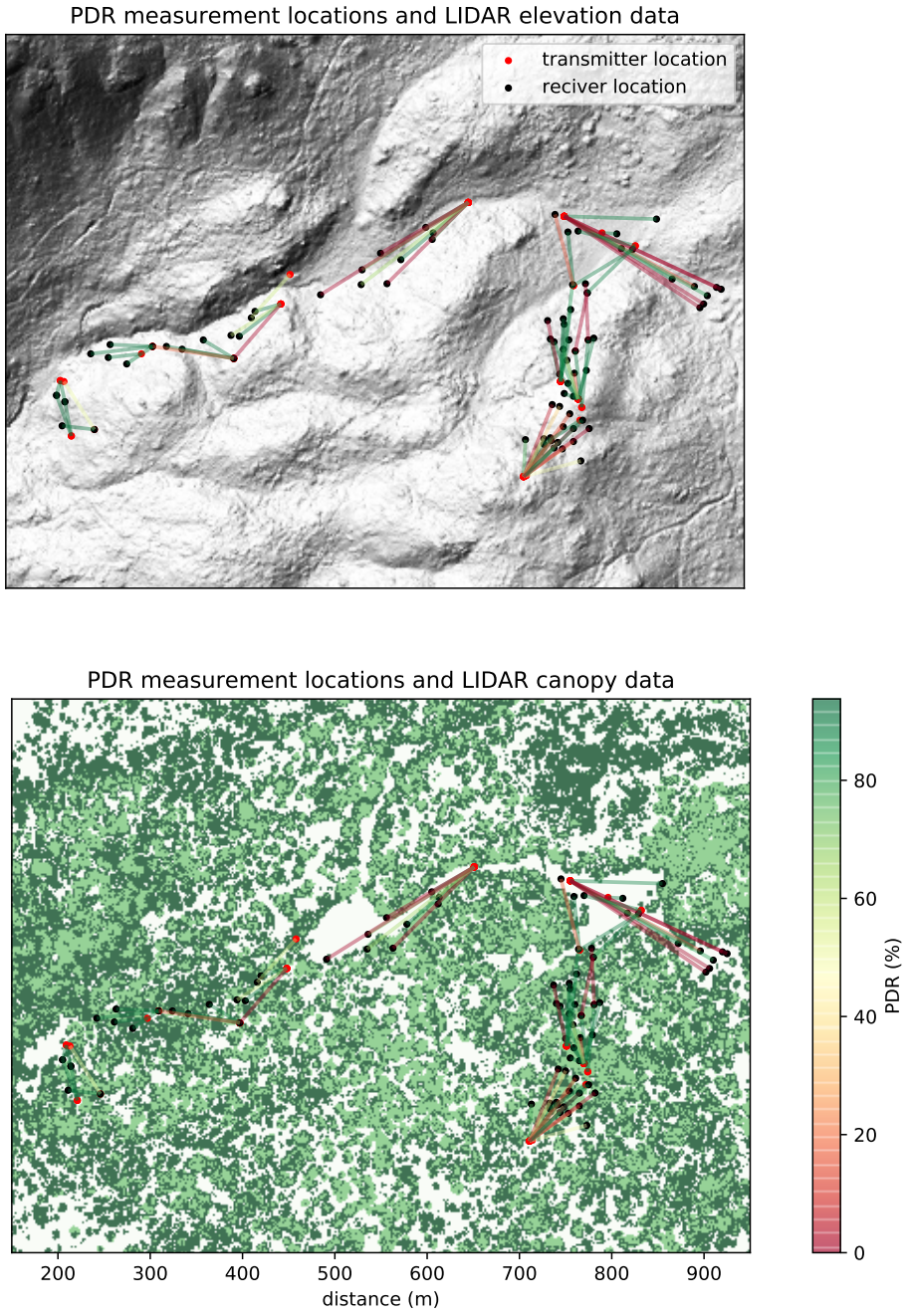

Figure 3.2: Transmitter and receiver locations shown with measured PDR and LIDAR elevation model (top) and canopy data (bottom).

vertically, but separated by a large angle. These results were likely specific to the 4dBi antennas used, which have a 50 degree vertical beam width. This characteristic makes these types of antennas a good choice for deployment in complex terrain. The following section attempts to capture the intuition developed during the data collection by training a PDR model based on the observed data.

3.3 PDR model

After collecting the LIDAR and PDR data from Section 3.2, the LIDAR data were used to compute path properties of each measured point based on the GPS locations of the transmitter and receiver. Given the intuition from Section 3.2, vegetation and distance were used as the primary path properties affecting the path links. First, the three-dimensional Euclidian between the two points was computed from DEM. Then, Bresenham's line algorithm [15] was used to compute all the grid cells between the transmitter and receiver. The amount of canopy was integrated along the path to calculate the total amount of vegetation on the path. Terrain intersection was computed by comparing the straight line of sight of the transmitter and receiver against the DEM profile between. If the two intersected within a 0.5 m tolerance, the link was set to zero PDR. A subset of the measurements and LIDAR path features is shown in Table 3.3.

Given the PDR measurements and path properties, a PDR model was created using a K-Nearest-Neighbor (KNN) regression scheme. Given an input space, x (the LIDAR-derived path features), a KNN regressor predicts a new value, $\hat{Y}$:

$$\hat{Y} = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \tag{3.1}$$

where k is the number of neighbors, and y_i is one of the measured PDR values. The number of neighbors to use in the regressor is tuned using 3-fold cross validation. The model was implemented in Scikit Learn version 18.1 using the *KNeighborsRegressor* module and *GridSearchCV* for cross validation. Negative mean squared was used as the scoring metric, and Euclidian distance was used as a distance metric. The search space consisted of between two and thirty neighbors. The optimal number of neighbors was determined to be seventeen.

Figure 3.3 shows the output of the KNN regressor for every 1m^2 grid cell in the domain, given a transmitter at the center of the image. The image illustrates how the algorithm captures the intuition gained from the PDR measurements from Section 3.2: good links can exist at long distances (dark green regions) when canopy does not intersect the path. If canopy intersects the terrain, PDR is not likely to be high at links greater than approximately 50 m (yellow regions). Where the path intersects the terrain, it is very unlikely a good link will form (red regions). The image illustrates how anisotropic path properties can be in complex terrain. The next section takes advantage of this property to find a minimal set of repeaters which connects the network.

Path distance, m	Path vegetation	2.405 GHz (Packets received)	2.425	2.445	2.465
71.20	0.0	973	899	905	892
100.07	16.0	879	916	902	864
135.35	6.0	944	834	894	902
160.21	41.0	135	237	161	22
177.30	38.0	639	876	929	881
182.12	91.0	0.0	1	0.0	2
187.50	94.0	0.0	0.0	0.0	0.0
178.44	48.0	0.0	0.0	0.0	0.0
177.27	78.0	0.0	0.0	0.0	0.0
46.88	51.0	902	914	901	923
42.63	27.0	902	856	842	847
62.58	70.0	882	942	915	901
74.31	87.0	914	928	922	897
61.70	78.0	0.0	0.0	0.0	0.0
64.73	66.0	0.0	0.0	0.0	0.0
64.04	5.0	938	897	925	887
72.86	51.0	839	941	939	894
63.08	57.0	530	565	590	659
36.62	67.0	908	912	920	909
62.71	70.0	479	533	573	486
36.77	46.0	927	921	930	895

Table 3.1: A subset of PDR measurements and LIDAR-derived path features

3.4 Topology optimization

The topology optimization consists of three steps: (1) determining the potential connectivity across the region of interest; (2) finding a minimal set of repeaters which connects all the sensor stations to the base station; (3) adding repeaters to maximize the connectivity of the mesh, up to the allowable budget of repeaters. First, repeaters were placed throughout the domain at 10 m intervals. Second, the PDR model from Section 3.3 was used to determine where links were likely to form. The PDR for a good link was required to be 75% or better. Third, a minimal set of repeaters was found to connect the sensor stations to the base station with at least 2-vertex-connectivity (the removal of any node will not disconnect the network) and 3-edge-connectivity (every node has at least 3 good neighbors) in the network topology. Finding minimal network topologies with guarantees on high-order connectivity has been shown to be MAXSNP-hard (no known Polynomial-Time Approximation Scheme exists) [92]. Therefore, an efficient heuristic algorithm was employed in order to find a minimal set of repeaters. A fast approximation algorithm for the Steiner tree problem [55] was used to find a minimal set of repeaters that fully connects the network (1-Vertex connectivity). This

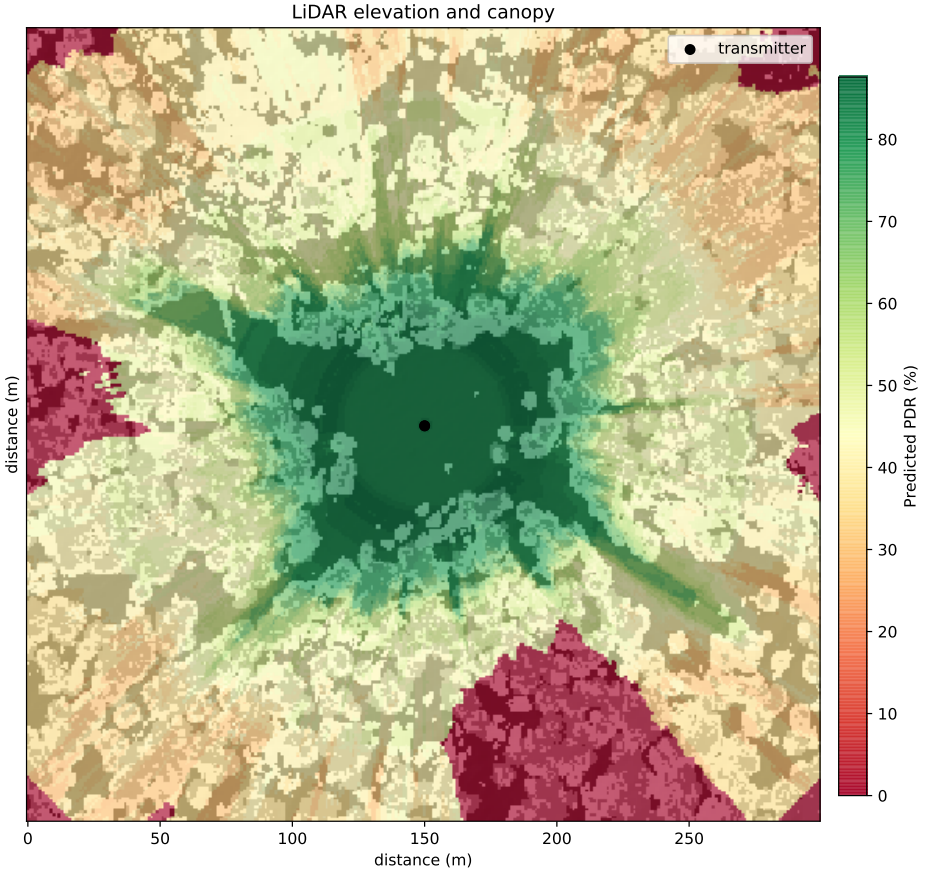

Figure 3.3: Predicted PDR at every 1 m^2 grid cell based on PDR measurements and LIDAR data given a transmitter at (150,150). Only dark-green links would be considered in the topology optimization component.

algorithm has low time complexity ($O(\|S\|\|V\|^2)$, for V total nodes, with S sensor stations. At each articulation point in the graph, additional repeaters were added to maximize the edge-connectivity to the adjacent nodes, until the final network is at least 2-vertex connected and 3-edge connected. An example set of initial placements is shown in Figure 3.5.

3.5 Results and discussion

The layout and number of repeaters determined by the algorithm is similar to the final network layout determined by the field team in [50]. Thirty-three repeaters were placed by the algorithm, compared to thirty-four in [50]. Where there is dense vegetation, the longest links selected by the algorithm are approximately 50 m, which accords with the findings and recommendations in [50]. Note that the algorithm takes advantage of longer line-of-sight

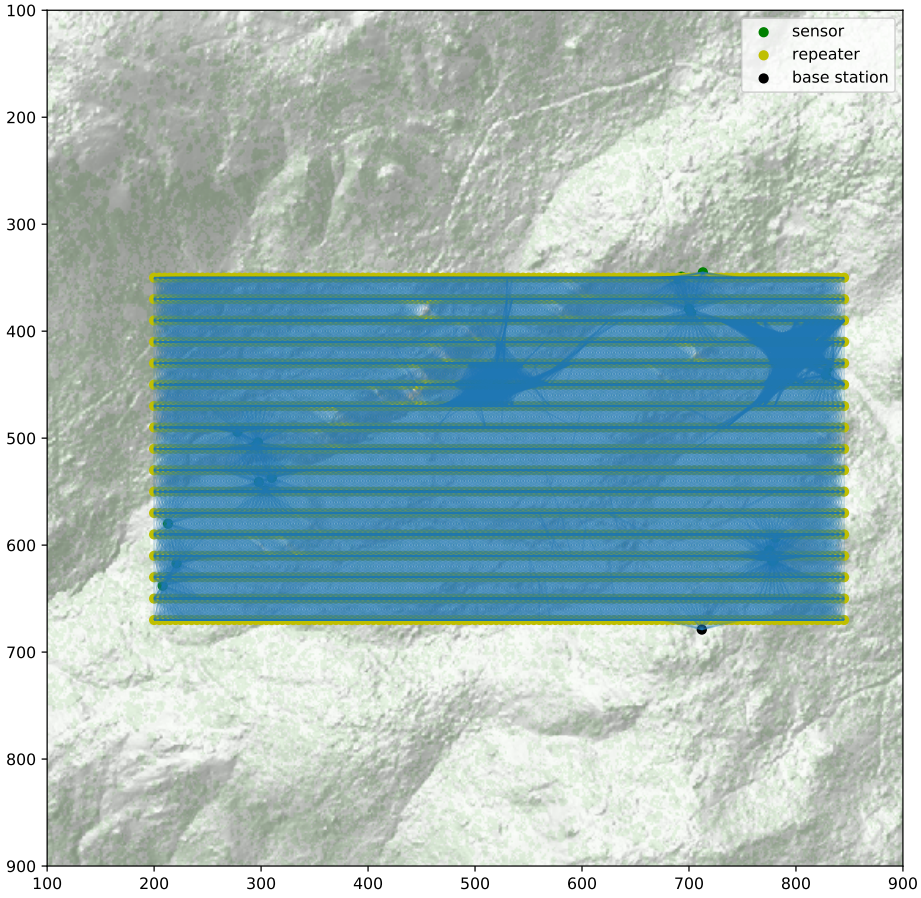

Figure 3.4: The connectivity is evaluated by placing repeaters throughout the region and evaluating potential network links.

pathways between the trees in the meadows to reduce the required number of repeaters.

These results indicate that the algorithm can effectively search through complex terrain to determine a reasonable network topology in order to guide the field team towards an initial set of placements. The estimated PDR should not be taken as a guarantee of path quality, however. In the present study PDR was found to vary considerably, particularly within vegetation. Moving links by just a few meters could result in PDR declining to zero despite similar path distance and vegetation characteristics. This also accords with the findings in [50]: the authors observed large temporal variability in the number of neighbors and path quality, which were not directly attributable to distance or vegetation. These effects may be attributable to factors such as destructive interference, multi-path fading, and changing environmental characteristics, such as relative humidity. It should also be noted that in seasonally snow-covered environments, surface characteristics can change dramatically as snow accumulates and melts, which is likely to affect long-term PDR.

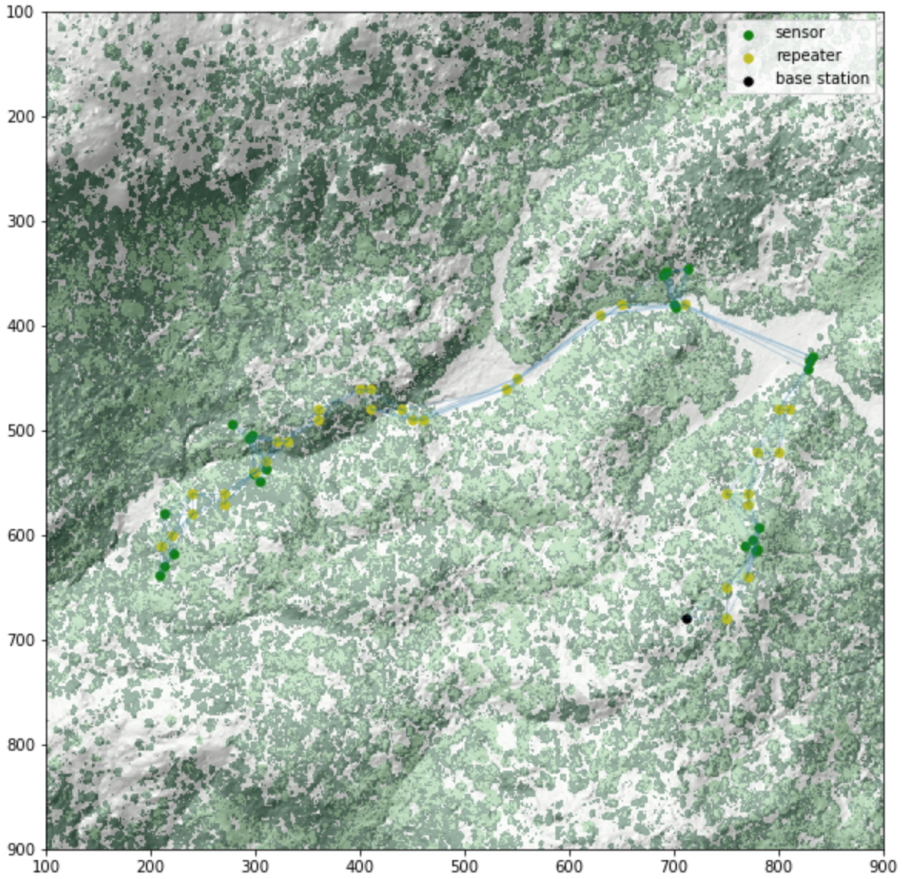

Figure 3.5: Repeater placements determined by the algorithm.

It will likely never be possible to predict PDR at any given time in even the simplest environments [104]. The factors affecting PDR in this and prior studies are inherently unpredictable, which underscores the importance of deploying redundant network topologies to ensure data continues to be received in the event of single node failures. The link predictor in the present study is only approximately 70% accurate in predicting good links in the cross-validated dataset. Therefore, it will likely always be necessary to have a field team validate the placements and make adjustments. Despite this, the algorithm and data presented in this chapter may provide a useful starting point for field teams to determine how many repeaters are required for a given environment. It also provides fixed starting points for field teams so they can place the network, find the few links that are not working and adjust them, rather than continuously rearranging the network, which prevents the mesh from forming optimally.

Chapter 4

Long-term accuracy of soil moisture estimated from in-situ sensors

4.1 Introduction

The prior two chapters discussed methods for optimizing wireless snow observatories. Many of these observatories are also equipped with soil moisture sensors, which provides a more complete estimate of the water balance at each node location. This chapter evaluates the long-term accuracy of estimating soil moisture from in-situ sensors in complex terrain. Soil moisture variability is controlled by multiple factors, including soil texture, controls on water routing, and aspect [70, 23, 110, 71, 105]. In montane regions, basin-scale estimation of soil moisture is confounded by complex topography, heterogenous subsurface properties, snow-soil interactions [106], and spatial variability of snow depth [69]. Though passive microwave monitoring of soil moisture enables remote observation of surface soil moisture, it is too coarse to capture spatial variability in the complex terrain of montane regions [7]. Remote-sensing techniques also only capture surface (typically 10 cm or above) soil-moisture storage [77, 103]. Observing only surface storage is insufficient for understanding controls on ecological functioning, as tree roots in these regions are present beyond 100 cm depth [8].

Recent advances in low cost sensor networks are enabling the deployment of more spatially extensive in-situ soil moisture measurements at a range of depths. Studies have suggested that an integrated sampling strategy could extrapolate in-situ measurements to uninstrumented regions in a basin [7, 46]. Developing a strategic and systematic approach for in-situ observation networks requires a long-term understanding of the accuracy of these methods as well as the controls on soil moisture. In seasonally snow-covered, semi-arid regions such as the Sierra Nevada, interannual patterns of precipitation and snowpack are highly variable [42]. This variability is exemplified by the multi-year drought in the Sierra Nevada that began in fall 2011, which has significantly lowered precipitation and snowpack-water storage, affecting water availability for local ecosystems and downstream stakeholders (California DWR Cooperative Snow Survey, cdec.water.ca.gov). The effect of these changes

on soil moisture availability in the rain/snow transition is of particular interest, as spatial patterns of soil-water storage affect catchment-scale water fluxes, including summer base flow in streams [10], biogeochemical processes and ecosystem health. Long-term changes in climate may also alter hydrologic processes in snow-dominated regions [11].

As the amount and timing of water entering the soil exhibit considerable inter-annual variability, predictors designed to extrapolate in-situ soil-water storage in average water years may not work well for years that are significantly wetter or drier. While there have been limited reports of catchment-scale soil moisture variability in these regions [106, 8], a long-term study of the attributes controlling patterns has not been reported. It is presently unclear how long-term variability will affect the predictive accuracy, and relative independent variable ranking, of soil moisture estimation. A long-term understanding of inter-annual differences in these features would benefit modeling, sampling strategies, and our understanding of ecological processes.

The aims of the present study were to: (i) quantify the accuracy of an ensemble regression tree algorithm (Random Forest) in estimating spatial soil moisture patterns over time at uninstrumented locations from in-situ measurements; (ii) identify which landscape attributes are most informative for predicting intra-annual and inter-annual patterns of soil moisture; (iii) determine whether dry years alter the estimator accuracy or temporal trends of independent variable ranking; and (iv) identify differences in the physical processes affecting soil-water storage in wet, dry, and average water years.

4.2 Methods

A 6-year dataset of spatially distributed water-balance measurements at the Southern Sierra Critical Zone Observatory (SSCZO) in the Kings River basin was used to quantify the long-term accuracy of soil moisture estimation in a montane environment. Soil moisture measurements at 10-, 30-, and 60-cm depths were co-located with snow-depth sensors, designed to capture gradients of physiographic features found to affect soil moisture (slope, aspect, elevation, and vegetation canopy cover) [8]. The data included a very wet water year (2011), years with near-average precipitation (2009-2010), and a record-dry period (2012-2014).

Site description and data collection

The Southern Sierra CZO is located in the Kings River basin of the Southern Sierra Nevada (Figure 4.1). It is situated in a mixed-conifer forest east of Fresno, California and contains sensors distributed across the rain-snow transition. The region receives mainly rain below 1500-m and snow above 2200-m elevation [8]. Soils are weakly developed and formed from decomposed granite [18]. Higher-elevation soils have a hard soil-bedrock interface, whereas soils at lower elevations have a deeper paralithic contact [8].

Data for the present study were collected from eighty-one soil-moisture sensors in the SSCZO deployed at twenty-seven water-balance nodes, which were placed into distinct higher-

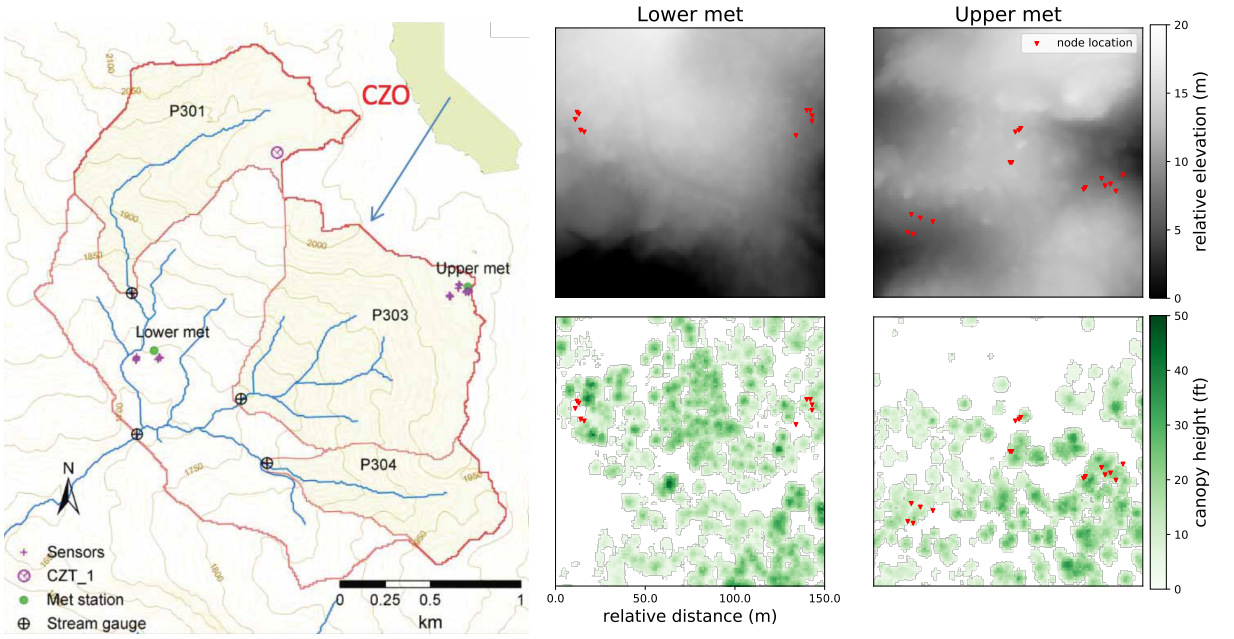

Figure 4.1: Study site location and sensor distribution. Left-hand panel shows the providence catchments, within the Critical Zone Observatory and the locations of the upper and lower clusters of sensor nodes within the basin. Right-hand panels show the distribution of sensors at upper and lower met with respect to the 1-m^2 LIDAR DEM (top panels), and canopy (bottom panels).

and lower-elevation clusters. Each node consists of an ultrasonic snow-depth sensor (Judd communications) and 3 to 4 co-located soil moisture and temperature sensors (Decagon EC-TM) at 10-, 30-, 60-, and 90-cm depths. The 90-cm measurements were not used in the present study as they are only installed at a subset of nodes, owing to limitations on hand excavation for installations in rocky saprolite. Data from 2 sensors (10cm south lower ppde and 60cm north lower open) were not used due to unreasonable values in the level two data. The higher-elevation site (abbreviated UM), had a cluster of 17 nodes, and lower-elevation (LM) site 10 nodes. The distribution of nodes was designed to capture gradients in slope, aspect, and canopy distribution. The following abbreviations for each node are used in the remainder of the article: (F/S/N) refers to nodes with flat, south-facing, and north-facing aspects, respectively; (uc/de/open) refers to nodes that are under canopy, at the drip edge, and open; and tree types are abbreviated AC: Abies concolor, PP: ponderosa pine, QK: black oak (Quercus kelloggii), PL: sugar pine (Pinus lambertiana) and CD: incense cedar (Calocedrus decurrens). Daily soil moisture storage, $S(t)_i$ for each node, i was determined by depth integrating the volumetric water content (VWC) data:

$$S(t)_i = \theta_{10cm}\Delta z_{0-10cm} + \theta_{30cm}\Delta z_{10-30cm} + \theta_{60cm}\Delta z_{30-60cm} \tag{4.1}$$

Topographic properties at each node (elevation, slope, aspect) were derived from a LIDAR digital elevation model of the region from the NSF open topography database (*opentopography.org*, accessed Sept 2016). The grid size of the computed variables was 1 m (derived from an average of 11.65 returns per square meter) [5]. Node locations were measured with a Trimble GPS (horizontal accuracies between 0.6 and 1.4 m).

Soil moisture estimation

To determine the long-term accuracy of soil moisture estimation from the in-situ measurements, an ensemble regression-tree algorithm (Random Forest, see [14]) was applied to the daily storage data as well as volumetric water content at each layer individually. Classification and regression tree (CART) algorithms can be used to build predictors when independent variables are a mix of continuous and categorical features (e.g. topographic-wetness index and location type in the present study). A single regression tree is known to overfit data, particularly when the number of features and number of samples are similar in magnitude [32]. To address this, ensemble tree algorithms such as Random Forest combine estimates from multiple CART models to arrive at an estimate of the true output. These methods have seen recent adoption in a variety of fields, such as remote sensing [34] and upscaling eddy-covariance measurements [49].

Five independent variables were used in the estimator for the present study: soil texture, topographic wetness, elevation, northness, and location type. Soil-texture properties were extracted from a prior survey, [8], and are shown in Table 4.1 for reference. Note that texture of the lower-elevation nodes (LM) is significantly finer than that at the higher-elevation nodes. The topographic-wetness index (TWI) was derived from the LIDAR elevation raster using the equation in [12],

$$\mathrm{TWI} = \ln\frac{a}{\tan(b)} \tag{4.2}$$

where a is the upslope contributing area per unit contour and tan(b) is the local slope. Topographic wetness was processed using the built-in module available at opentopography.org. Elevation was represented as an integer-based categorical variable representing the high and low clusters to minimize the collinearity with other features. Northness was computed from the LIDAR slope and aspect rasters:

$$\mathrm{northness} = \sin(\mathrm{slope}) \times \cos(\mathrm{aspect}) \tag{4.3}$$

Finally, location type was encoded as one of three categorical variables representing drip edge, under canopy, and open. The node properties are summarized in Table 4.1.

The regression was implemented using the *Random Forest* module in Scikit-learn version 18.1. One hundred trees were used in the ensemble. A new regression was performed each day in the six-year study period for depth-integrated soil-water storage as well as for the VWC in the 10-, 30-, and 60-cm layers individually. Average soil texture was used when

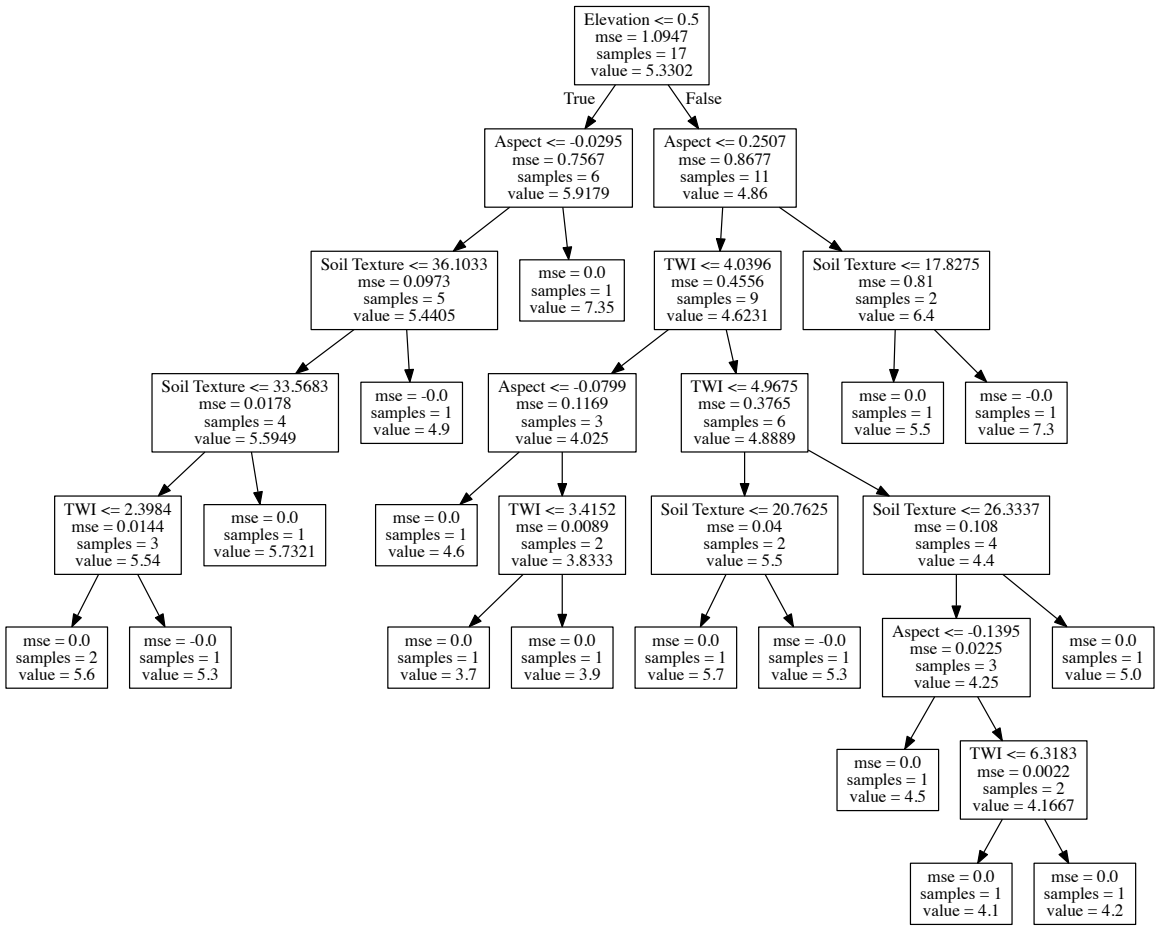

Figure 4.2: Illustration of the ensemble tree-based regression algorithm for depth-integrated storage. One of the 100 estimators is shown. The feature importances are derived from the relative rank of the variable averaged across all trees.

Table 4.1: Properties for each node. Soil texture (clay% + silt%) shown for 10-, 30-, and 60-cm layers. Upper met soils are weakly developed, lower met are well developed, containing more clay and silt.

Node	10 cm (texture)	30 cm	60 cm	TWI	Elevation	Northness
North upper acde	29.87	25.30	27.02	5.70	1979.25	0.22
North upper acuc	24.98	23.75	22.52	7.14	1978.99	0.24
North upper cdde	23.78	21.32	17.00	6.79	1977.37	0.36
North upper cduc	21.04	17.98	9.85	7.29	1978.38	0.31
North upper open	25.93	21.88	21.36	4.44	1974.63	0.18
North upper plde	24.14	21.98	22.79	4.67	1977.21	0.08
North upper pluc	23.25	21.96	20.42	6.85	1977.92	0.26
South upper acde	24.32	21.58	13.93	5.42	1974.90	-0.26
South upper acuc	23.84	19.75	16.50	3.46	1974.85	-0.18
South upper open	23.75	18.92	15.20	3.62	1976.78	-0.14
South upper qkde	23.41	19.27	17.55	3.49	1976.91	-0.20
South upper qkuc	22.70	19.92	17.67	4.51	1977.30	-0.25
Flat upper acde	19.61	14.35	13.46	5.69	1983.15	0.03
Flat upper acuc	16.23	11.64	11.00	3.64	1983.17	0.02
Flat upper open	22.00	23.24	25.15	3.20	1982.62	0.06
Flat upper ppde	29.07	24.10	24.47	5.50	1982.60	-0.02
Flat upper ppuc	30.13	25.58	26.02	3.48	1982.63	-0.02
North lower acde	29.08	27.74	26.64	2.84	1732.00	-0.11
North lower acuc	32.61	31.02	28.71	6.24	1732.08	-0.01
North lower open	23.00	20.41	20.44	1.96	1733.14	-0.09
North lower cduc	32.37	33.21	34.35	1.48	1732.49	-0.05
North lower cdde	36.57	34.23	31.45	6.02	1732.80	-0.02
South lower cdde	35.88	37.50	38.66	1.73	1737.03	-0.12
South lower cduc	35.11	37.21	35.29	1.77	1737.26	-0.11
South lower open	38.50	34.03	41.88	4.77	1737.45	-0.15
South lower ppde	36.03	32.59	32.86	4.94	1738.47	-0.18
South lower ppuc	37.77	35.37	33.71	3.39	1738.33	-0.27

applying the algorithm to the depth-integrated soil storage. The predictive accuracy of the algorithm was quantified with the coefficient of determination (R^2) using the out-of-bag error for the Random Forest (*oob score* in Scikit-learn). Where there were gaps in the daily measurements, the data were excluded from the algorithm for that day. The relative importance of each independent variable was determined from the *feature importances* attribute of the predictor, which determines the relative contribution of each independent variable based on how frequently it is used in the CART estimator, averaged over all trees in the ensemble [82].

4.3 Results

The soil moisture storage in average precipitation (WY 2009), wet (WY 2011), and dry (WY 2014) years is shown in Figure 4.3. The results are discussed with respect to the wet-up, high-flux, and dry-down phases, which are labeled on the figure.

Soil moisture estimation

During the six-year study period, the predictor accuracy and the relative contribution of each physiographic variable differed considerably (Figure 4.4). The top panel shows the daily data for precipitation, soil-moisture storage, and snow depth. This panel illustrates the inter- and intra-annual variability of these quantities. The second panel shows the predictive accuracy (R^2 of predicted and measured data) of the ensemble regression tree. Each lower panel shows the relative importance of each physiographic variable in the ensemble.

The accuracy is highest during the high-flux and early dry-down periods of wet years (2010-11). At the end of the dry-down period in all years, the coefficient of determination declines to near zero. The low accuracy period typically lasts for one to two months, before a new influx of precipitation restores the predictive accuracy on the wet-up period. During the high-flux period of each year, large changes to the snowpack state result in a temporary decrease in accuracy (e.g., R^2 decreases from 0.4 to 0.05 during an early season melt event in 2009 and from 0.6 to 0.2 in 2010). Drought conditions exhibit extended periods of low accuracy: during WY 2012 and 2014 the R^2 is near-zero for four months (July-October). By WY 2014, dry conditions coincide with lower accuracy even during the high-flux period.

Soil texture is the highest-rank predictor of soil-water storage during high-flux periods, with elevation and northness contributing as well. The importance of texture declines during early dry-down, with that of elevation and northness increasing. Topographic wetness becomes important as dry-down progresses. During this late dry-down period, northness and location type exhibit some importance, but less than soil texture. The importance of topographic wetness increases during the dry-down process, after the snowpack has finished melting, however the out-of-bag R^2 is low. During dry years the importance of topographic wetness remains high for a longer portion of the water year. The relative importance peaks during the initial distribution of snow and during the initial dry-down period. The impor-

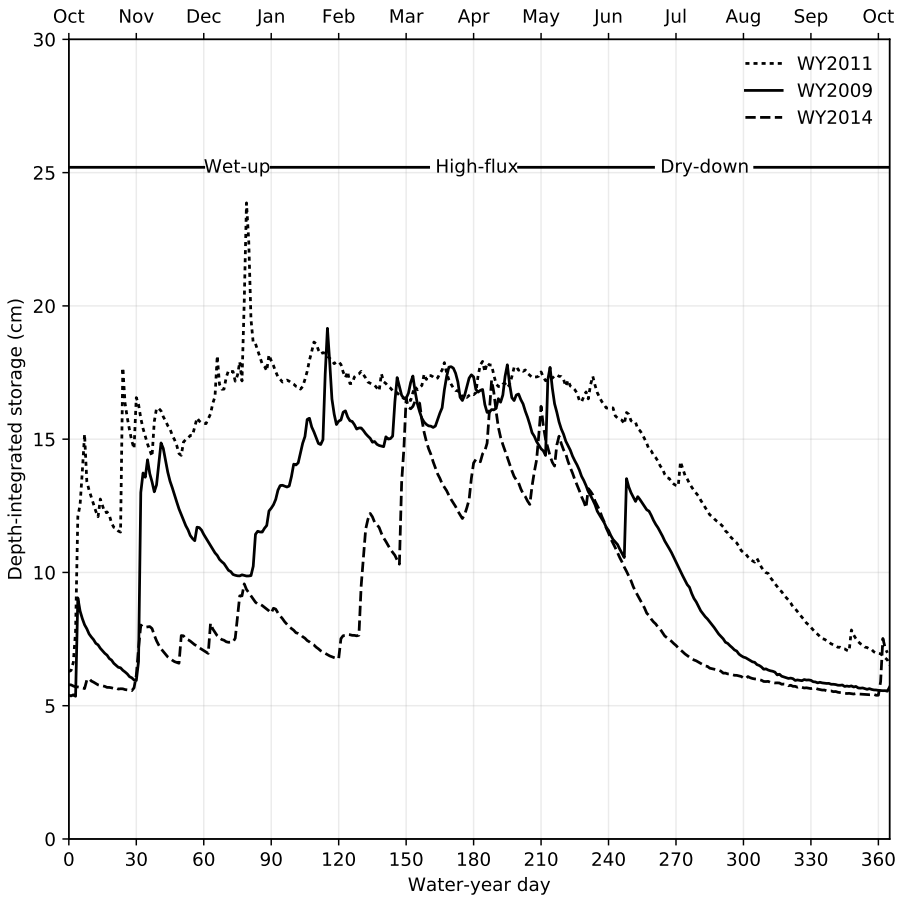

Figure 4.3: Mean depth-integrated soil-water storage for average-precipitation (WY2009), wet (WY2011), and dry (WY2014) years in three different periods of the water year.

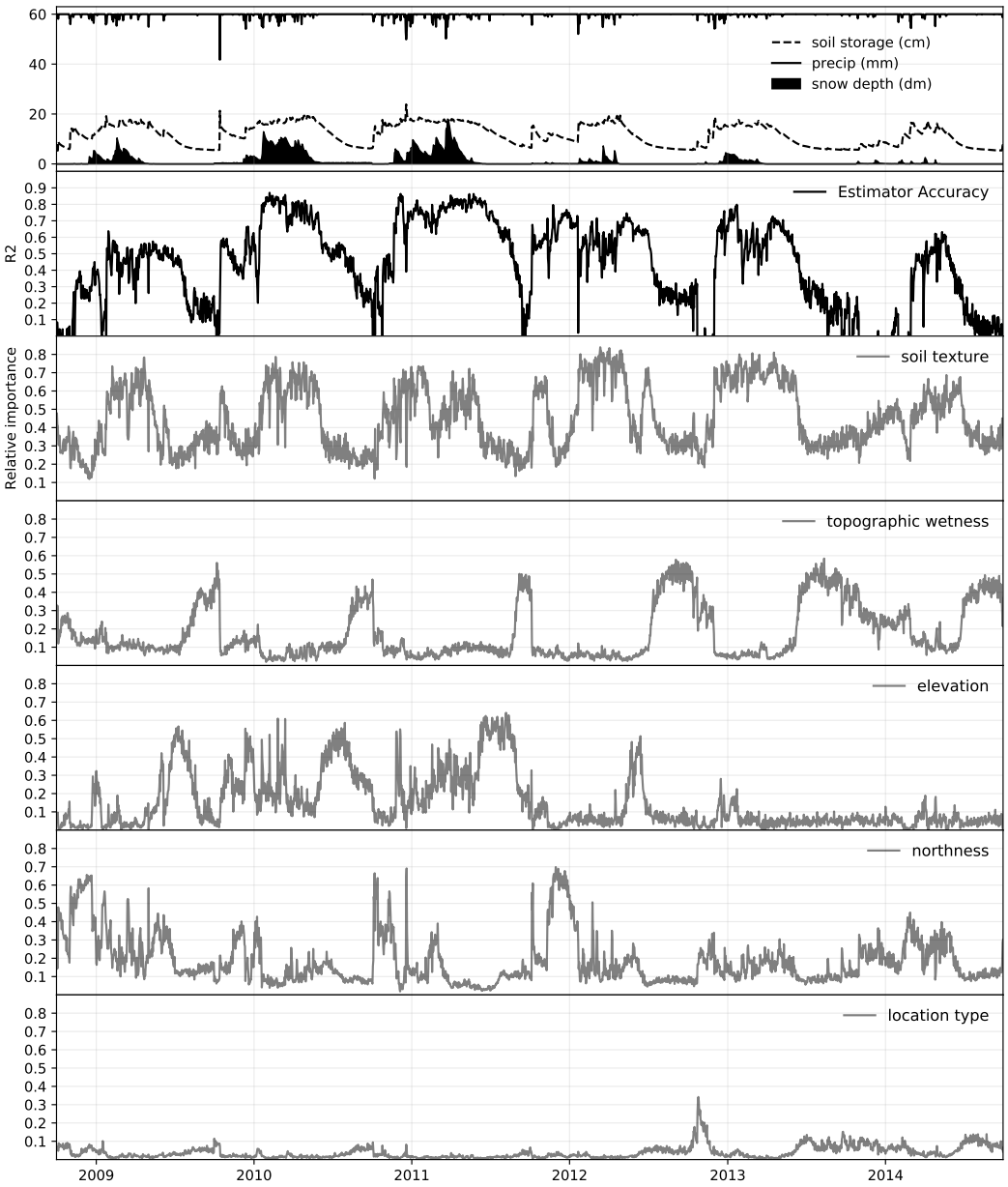

Figure 4.4: Ensemble regression-tree predictive accuracy and independent variable ranking for six water years. Top panel shows daily precipitation, mean soil-moisture storage, and mean snow depth. Second panel shows predictive accuracy (coefficient of determination) of ensemble prediction algorithm. Lower panels show relative importance of each predictive variable, values for which sum to 1.0.

tance of elevation peaks just as the last of the snowpack is melting, slightly earlier than topographic wetness. In dry years, the contribution of elevation declines entirely. The importance of northness typically peaks early in the water year, just before and during the early accumulation of snow pack. Finally, location type exhibits minimal importance in any year. The peak in its importance in late 2012 should be disregarded as it corresponds to a zero-accuracy period for the estimator.

When the predictor is applied to VWC data at each depth individually (Figures 4.5, 4.6, and 4.7), there are large variations in predictability and relative feature ranking. The accuracy is higher for the 60-cm layers than for the 30- or 10-cm layers. Surface (10-cm layers) are most predictable during the high-flux and dry-down phases of wet years, whereas subsurface (60-cm layers) are most predictable during the wet-up and high-flux phases. The accuracy in the intermediate (30-cm layers) is consistently lower than the other layers, showing extended periods of zero R^2. Northness, elevation and soil texture are the highest-rank predictors for the 10-cm layers, whereas only elevation, topographic wetness and soil texture are strong predictors for the 60-cm layer. In the 60-cm layer, the duration of the high-importance for elevation declines, while the importance of topographic wetness increases.

Differences in the physical processes affecting soil-water storage in each water year

In dry years, differences in the wet-up period were associated with drier lower soil layers, which are not affected by initial inputs of precipitation, illustrated in Figure 4.8. The figure shows the soil moisture storage for all layers of a single node, a drip-edge node near a ponderosa pine in the upper met station. Conditions at the end of the dry-down period are nearly identical in wet and dry years. Lower-layer soils are all just below 0.1 VWC. The 10-cm soils are drier than the lower layers in both years (around 0.03 VWC). In wet years, early season precipitation affects all soil layers, resulting in a highly correlated response across the soil column. In dry years, shallow soil layers respond to precipitation, but lower-soil layers are not affected. Deeper soils do not begin to store water until well into the high-flux period. Similar patterns are seen across many of the other vertical soil moisture transects (data not shown).

Differences during the high-flux period were associated with a transition to non-continuous snow cover during dry years. Unlike the wet-up period, which was characterized by dry lower-layer soils, upper soil layers in the high-flux period show greater declines due to multiple winter-spring melt events. The coincident increases in soil temperature, diurnal soil temperature fluctuations, and correlated ET increases, suggest that some of the storage is potentially being partitioned into evaporation rather than infiltration (Figure 4.9). The lower three panels show mean snow, soil moisture, soil temperature, and patterns for wet and dry years; the top panel shows coincident evapotranspiration patterns during dry years.

Physical differences in the dry-down period are minimal. Initial storage conditions in dry years are lower, resulting in a more uniform distribution of soil moisture across the

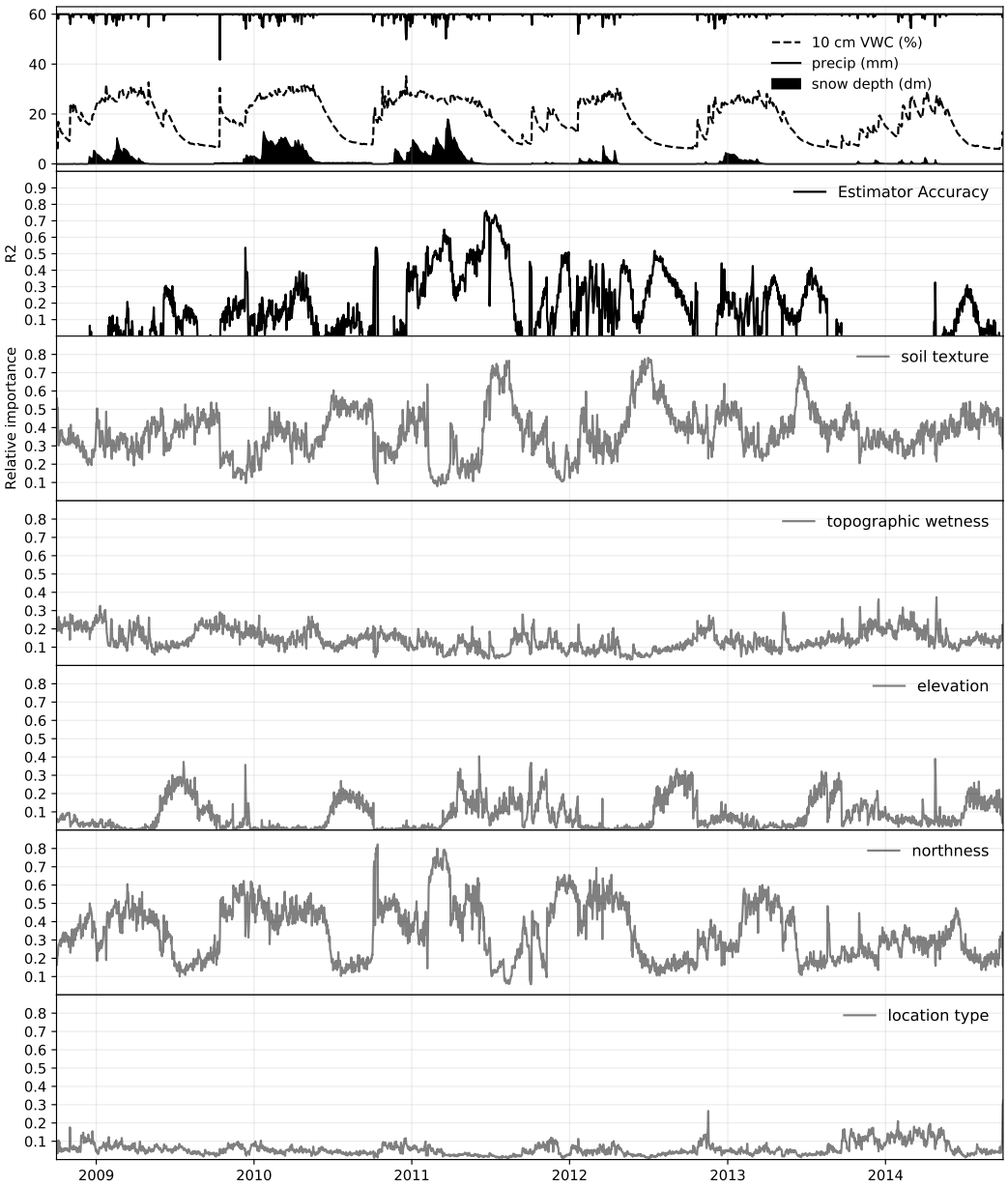

Figure 4.5: Ensemble regression-tree predictive accuracy and independent variable ranking for six water years at 10 cm depth. Top panel shows daily precipitation, mean 10-cm VWC data, and mean snow depth. Second panel shows predictive accuracy (coefficient of determination) of ensemble prediction algorithm. Lower panels show relative importance of each predictive variable, values for which sum to 1.0.

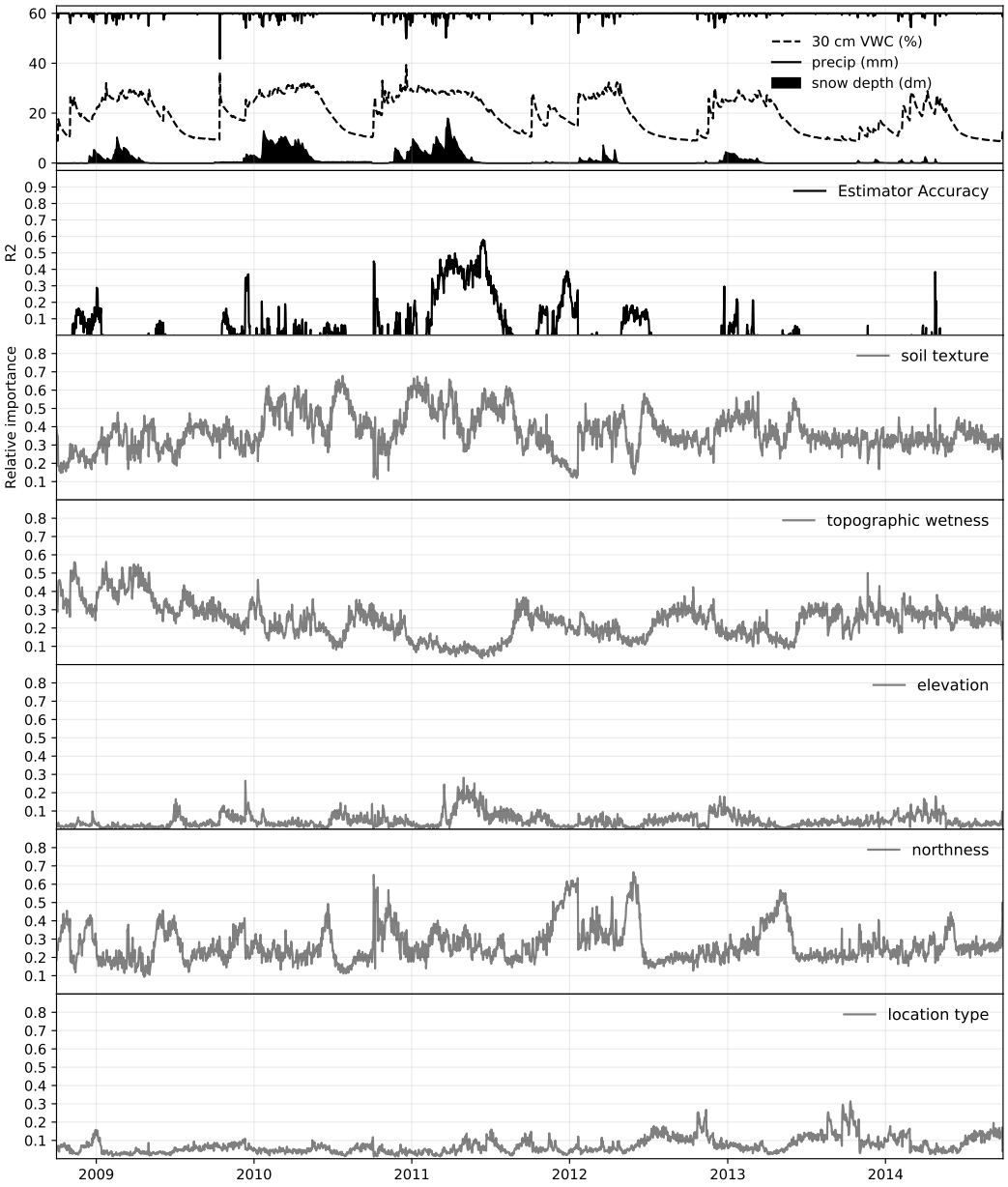

Figure 4.6: Ensemble regression-tree predictive accuracy and independent variable ranking for six water years at 30 cm depth. Top panel shows daily precipitation, mean 30-cm VWC data, and mean snow depth. Second panel shows predictive accuracy of ensemble prediction algorithm. Lower panels show relative importance of each predictive variable, values for which sum to 1.0.

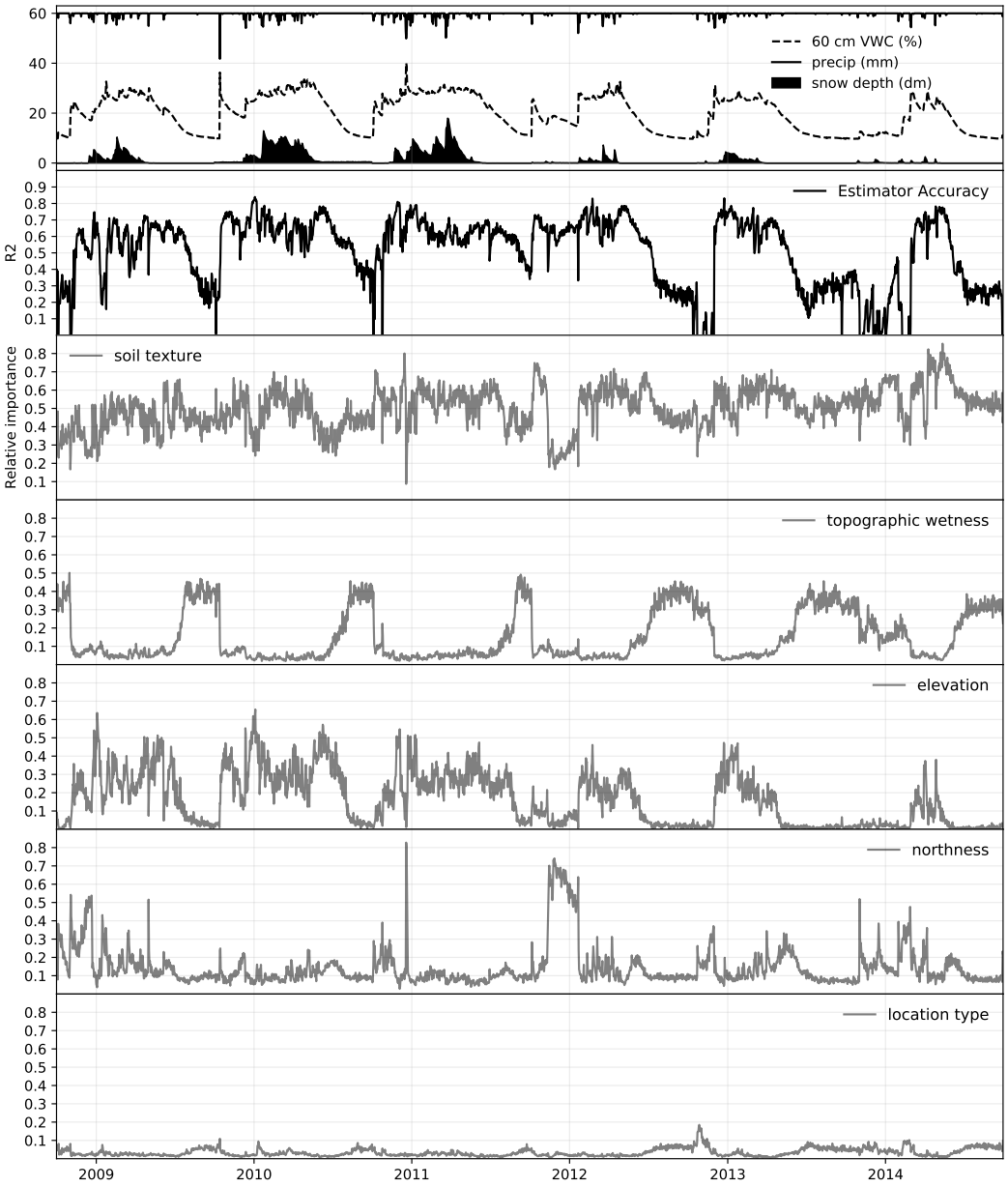

Figure 4.7: Ensemble regression-tree predictive accuracy and independent variable ranking for six water years at 60 cm depth. Top panel shows daily precipitation, mean 60-cm VWC data, and mean snow depth. Second panel shows predictive accuracy of ensemble prediction algorithm. Lower panels show relative importance of each predictive variable, values for which sum to 1.0.

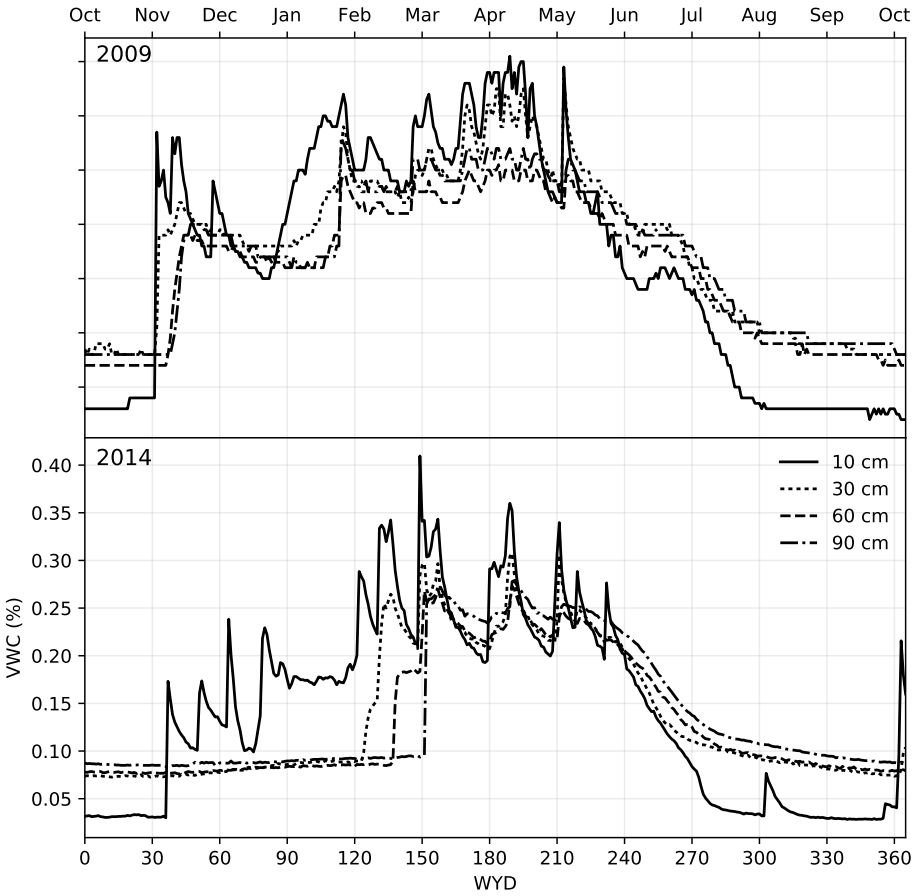

Figure 4.8: Soil moisture at a single node in wet (2009) and dry (2014) years. In dry years, initial precipitation has minimal effects on lower levels of soil moisture, resulting in a de-coupling of storage in the upper and lower layers.

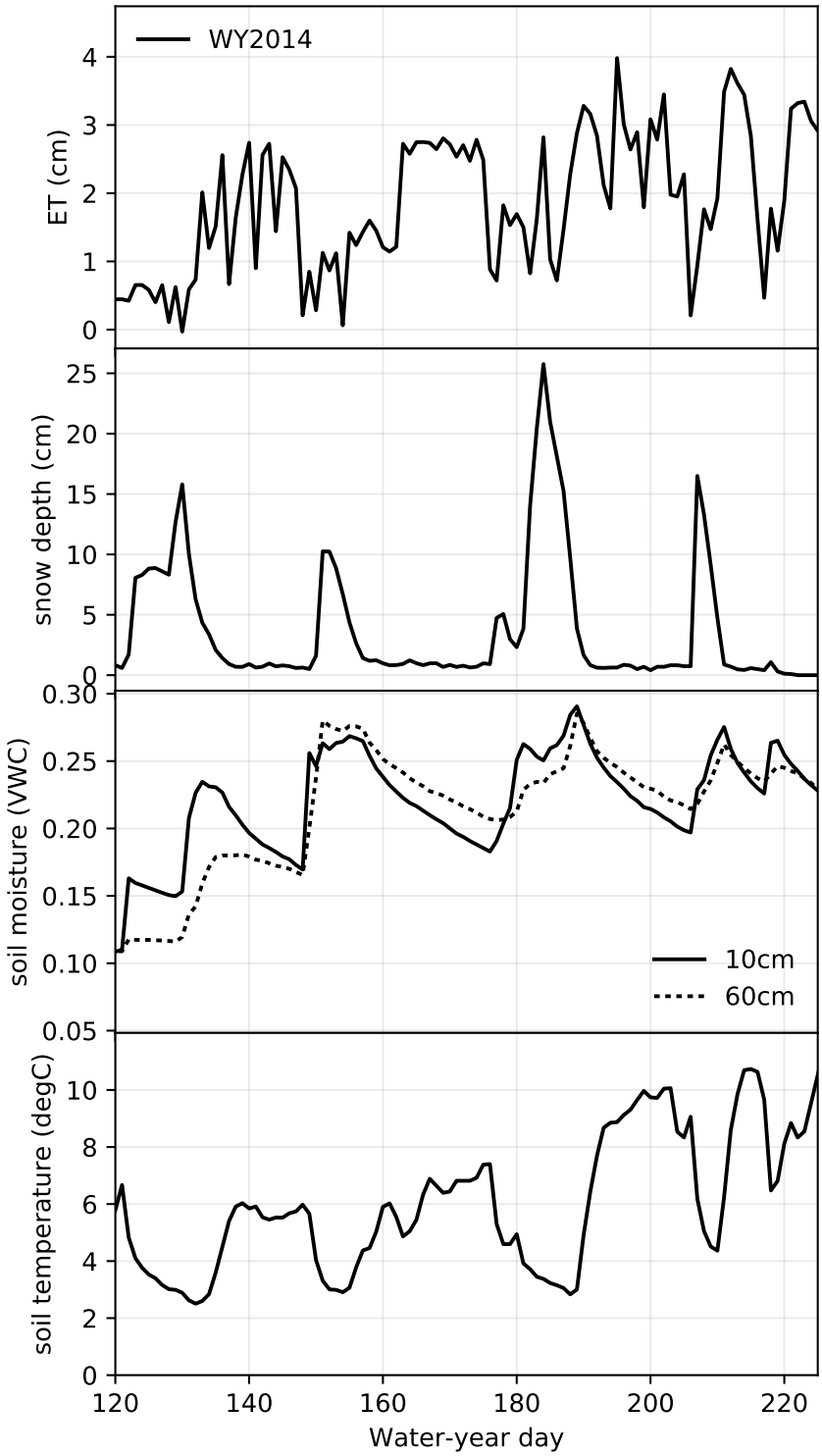

Figure 4.9: Snow-off conditions during the typical high-flux period of WY 2014 resulted in soil moisture storage decline, increased soil temperature, and coincident increases in ET.

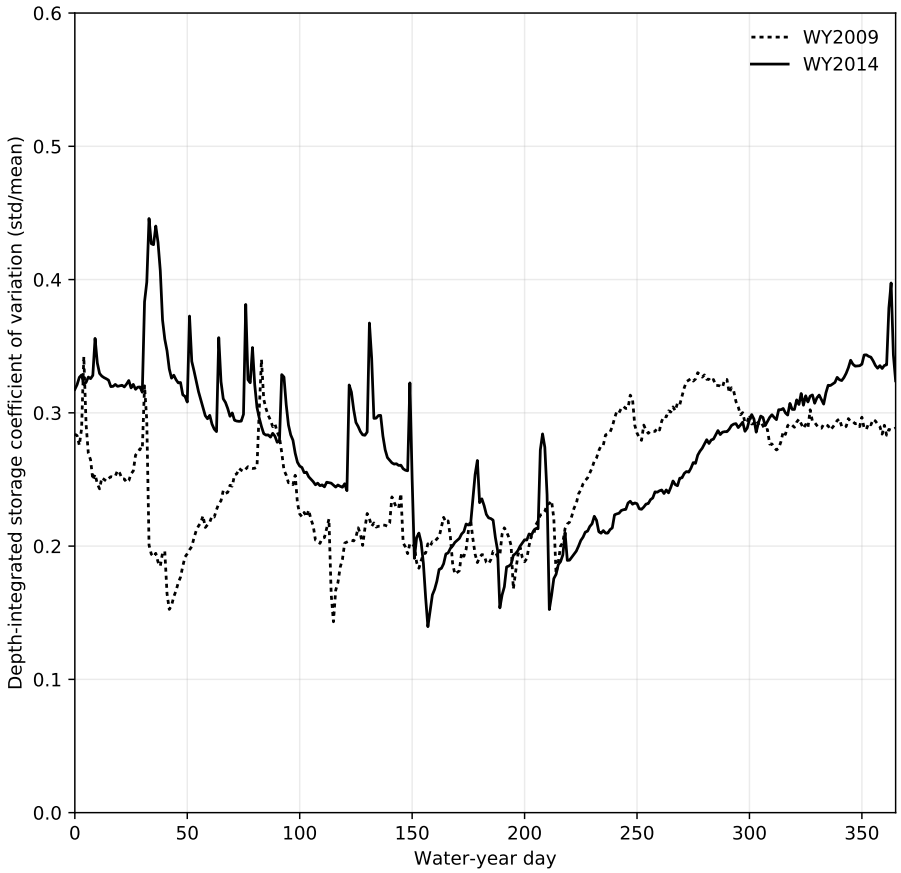

Figure 4.10: Coefficient of variation for wet and dry years (WY2009 and 2014, respectively). Dry years are characterized by greater variability during the wet-up period, and lower variability during the high-flux and dry-down periods.

region during the dry-down process. This is reflected in the coefficient of variation of all nodes (Figure 4.10). The typical water year exhibited a steep increase in the coefficient of variation during the summer months. The dry year exhibited a more gradual increase.

4.4 Discussion

The variability of the predictor accuracy and independent variable ranking in the present study accord with the results in [105] and [106], which found significant seasonal variability

in the degree to which terrain indices explain the distribution of soil moisture. Topographic features exhibit predictable seasonal controls on soil moisture storage in typical water years, but are altered by drought conditions (Section 4.3). Across all layers, wetter soils tend to be more predictable, which accords with the findings in both prior studies. However, the short-term predictive accuracy can decrease during wet periods due to significant changes in the snowpack (e.g., melting during January 2009 and 2010). Also, soil moisture storage for large precipitation inputs during the wet-up period is not predictable. Given these findings, confidence in the statistical modeling of soil moisture storage must incorporate daily knowledge of the snowpack state, soil moisture state and precipitation. This underscores the importance of deploying complete water-balance instrument clusters to un-instrumented regions rather than soil-moisture sensors alone. The most consistently predictable period during all years occurred during the dry-down period, shortly after the snow depth declined to zero. The predictive accuracy remains high during the dry-down process for the spring and summer months, but declines thereafter.

Inter- and intra-annual differences in the independent variable ranking may be explained by different water-routing processes during each part of the water year. The northness terrain attribute was more highly ranked for surface (10cm) water storage, likely a result of radiative variability affecting soil-water storage near the surface. Deeper soil layers were more affected by topographic wetness and elevation, which may reflect different snow-soil interactions. The relative importance of elevation declined in dry years, perhaps due to the decreasing differences in snowpack between the upper- and lower-elevation sites.

These results suggest sensor-placement strategies must account for the relative importance of predictive features, which vary across seasons and depths. The distribution of surface soil moisture sensors should aim to capture differences in soil texture and northness, whereas subsurface sensors should aim to capture variability of topographic wetness, soil texture and elevational differences. Sampling strategies for dry years should aim to capture greater variability of topographic features such as topographic wetness and northness.

Section 4.3 revealed that extremely dry years are characterized by physical changes which alter soil moisture storage patterns across the basin. Dry lower-soil layers and minimal input precipitation resulted in a decoupling of lower- and upper-soil moisture storage in the wet-up period. In typical water years, tightly coupled soil layers make observations of the surface storage a good proxy for lower-level storage. In dry years, the upper- and lower-level soil moisture are less tightly coupled. Remote sensing tools, which measure only surface soil moisture may overestimate soil storage in the lower soil column during dry years, and may underestimate soil storage during transient snow-off conditions. During the wet-up period of dry years, remote sensing may therefore underestimate plant-available water in the soil.

4.5 Conclusion

Four conclusions can be drawn from the findings in this chapter: (1) An ensemble regression tree algorithm using topographic features and soil texture as independent variables

exhibits higher accuracy estimating depth-integrated water storage and soil moisture at various depths during wet years than during dry years. (2) Soil texture has consistently high feature importance across all soil layers. Other landscape attributes exhibit seasonal trends: northness peaks during the wet-up period, and elevation and topographic-wetness index peak during the late high-flux and early dry-down periods, respectively. (3) The relative rank of each predictor exhibits temporal variability with predictable seasonal trends during wet years, which do not persist during dry years. (4) Physical drivers of altered soil-moisture patterns in dry versus wet years included lower-than-average deep soil storage, and more frequent snow-off conditions, resulting in a decoupling of upper and lower layers of the soil column.

Chapter 5

Design Framework for Autonomous Lagrangian Sensors in the Sacramento-San Joaquin Delta

5.1 Introduction

This chapter describes the design methodology for a network of robotic Lagrangian floating sensors designed to perform real-time monitoring of water flow, environmental parameters, and bathymetry of shallow water environments (bays, estuarine and riverine environments). Unlike previous Lagrangian sensors which passively monitor water velocity, the sensors described in this chapter can actively control their trajectory on the surface of the water and are capable of inter-sensor communication. The addition of these functionalities enables Lagrangian sensing in obstacle-encumbered environments, such as rivers. The Ishikawa cause and effect design framework is used to ensure that the final system synthesizes the diverse operational and functional needs of multiple end user groups to arrive at a broadly applicable system design. A summary of potential applications for the system is given, including completed projects performed on behalf of the Department of Homeland Security, Office of Naval Research, and the California Bay-Delta Authority.

A Lagrangian sensor is an instrument used in the field of oceanography and hydrodynamics to monitor the currents and other physical properties of large-scale hydraulic systems. The instrument measures the water velocity by moving within the medium, along a trajectory of the flow (as opposed to an "Eulerian" sensor which measures the properties of the medium from a fixed location). It measures other physical properties, such as temperature, turbidity, and dissolved oxygen, with a variety of onboard sensors. Lagrangian sensors have seen extensive use in the oceanographic community since 1955 [96]. In the field, they are referred to as "drifters." Early drifters used acoustic communication to transmit data to researchers [35]. Their capabilities expanded in 1978 with the introduction of the Argos satellite service, which enabled remote communication of sensor data and location [17]. Ex-

amples of Argos-enabled units include the Costal Dynamics Experiment drifter [19], and the Low Cost Tropical Drifter [76]. In the oceanographic community, these sensors are usually "passive," meaning they have no actuation capabilities.

The design of oceanographic drifters generally consists of a buoyant spherical or cylindrical housing for the sensors and communications equipment, which is tethered to a drogue. The drogue extends into the water and is designed to capture the current by presenting a large, symmetric drag profile to oncoming flow. For example, the Low Cost Tropical Drifter's sensor housing is a PVC tube which is 11.4 centimeter diameter and 2.9 meters long. It is designed to be positively buoyant such that 0.4 meters of the cylinder are above the surface of the water. The drogue is attached by a 7.5 meter nylon tether which is 1.6 cm in diameter. The drogue itself is a perforated cylindrical tube which is 10 meters long and 50 centimeters in diameter. Similarly, the TRISTAR drifter consists of a 48 centimeter fiberglass sphere tethered to a 557 centimeter by 544 centimeter symmetric drogue.

A logical extension of the oceanographic research in Lagrangian sensing is to develop sensors for near-shore environments such as rivers and bays. It is important to instrument such environments because most of the renewable freshwater available for human use flows through rivers [78]. Lagrangian sensors can be used in this context to better monitor the flow of freshwater and the transport of constituents therein. Specific examples include assessing vulnerabilities to contaminate spills or infrastructure failure in critical water-resource regions, planning reservoir release and gate control policies to affect the intrusion of saltwater, and monitoring the effect of heavy agriculture use on freshwater supplies.

Simply scaling down the oceanographic design paradigm will not work for Lagrangian sensors operating in constricted domains. In such environments, the possible locations of the sensor represent more space relative to the boundaries of the environment, and sensors can easily become entangled on the shore. Also, because rivers are obstacle-encumbered and contain many branches, there are situations in which the sensor's motion must be actively controlled, rendering passive sensors less useful for monitoring missions. Finally, in rivers, sensors move through the experimental domain faster, requiring more frequent deployments and retrievals. Thus, a new sensor must be designed that retains a suitable form factor for Lagrangian sensing, but adds capabilities such as active control and inter-vehicle communication.

Fortunately, recent developments in sensor networks for aquatic sensing have made operations in constricted environments feasible. Examples include the AMOUR project at MIT [21], the NEPTUS Autonomous Underwater Vehicle (AUV) at LSTS in Portugal [95], submersible pneumatic drogues built at UCSD [41], the Slocum underwater drifters at MBARI [13] and the Smart Bay sensor network in Galway Bay [90]. In many cases, these projects have added the communication and actuation capabilities necessary for operations in constricted environments. Many of these Autonomous Underwater Vehicles (AUVs), such as the submarines used in LSTS, are not suited for Lagrangian sensing in rivers specifically. The overall size and shape of a submarine is not appropriate for Lagrangian sensing since it does not present a symmetric drag profile to the flow, and the deployments require specialized equipment.

This chapter details the design for the physical sensors in such a network. The use of a design methodology adapted from the Ishikawa "cause and effect" framework (discussed in detail below) has enabled the creation of an electro-mechanical system which balances the diverse needs of a wide range of end-user groups. This chapter presents an overview of the *Floating Sensor Network* system, a discussion of the design methodology used for the physical sensors, and a detailed explanation of how each component of the system evolved from the methodology. The chapter concludes with three case studies in which the design decisions were validated in real-world tests.

5.2 System Design Overview

The UC Berkeley *Floating Sensor Network* [1] is a system of robotic Lagrangian sensors designed for use in shallow water environments [3, 100]. The system is comprised of the robotic sensors shown in Figure 5.1, as well as the communication and visualization infrastructure required to transmit and display the data in real time.

Each robotic sensor consists of a cylindrical hull which contains batteries, communications equipment, GPS, and onboard processors. Onboard sensors for salinity, temperature and depth extend from the base of the vehicle.

Figure 5.2 shows a schematic of the full drifter-backend system, which consists of an integrated computational support tool and platform linked to the floating sensors. Data from the sensors can be intercommunicated between vehicles using 802.15.4 radios. Data is also sent back to a central server using the GSM cell phone data network. The system enables the data from the sensors to be assimilated and visualized in real time, giving the end user live access to the sensor data. It also assists field operations by providing a constant stream of vehicle diagnostics and position data.

The velocity measurements inferred from GPS can be combined with measurements from a static infrastructure maintained by the US Geological Survey [100]. The computations necessary to take data from the sensors in the field and create an estimate of the entire system is called data assimilation. This is an estimation problem characterized by sparse sensor data. Several techniques can be used to perform data assimilation using these streaming data measurements. For example, the velocity estimates can be assimilated with an Ensemble Kalman Filter algorithm to generate velocity estimates for the rest of the river system [94] [100]. This technique incorporates the sparse sensor data with a model of the hydrodynamic system, such as a two-dimensional shallow water equation, to arrive at an improved state estimate of the system. These estimates are sent back to the processors on the vehicles for path planning.

The vehicle positions, sensors readings, and state estimates from the Ensemble Kalman Filter algorithm are available online in real time through a visualization tool developed at UC Berkeley called DIVA, built on Google mapping technology (see Figure 5.3).

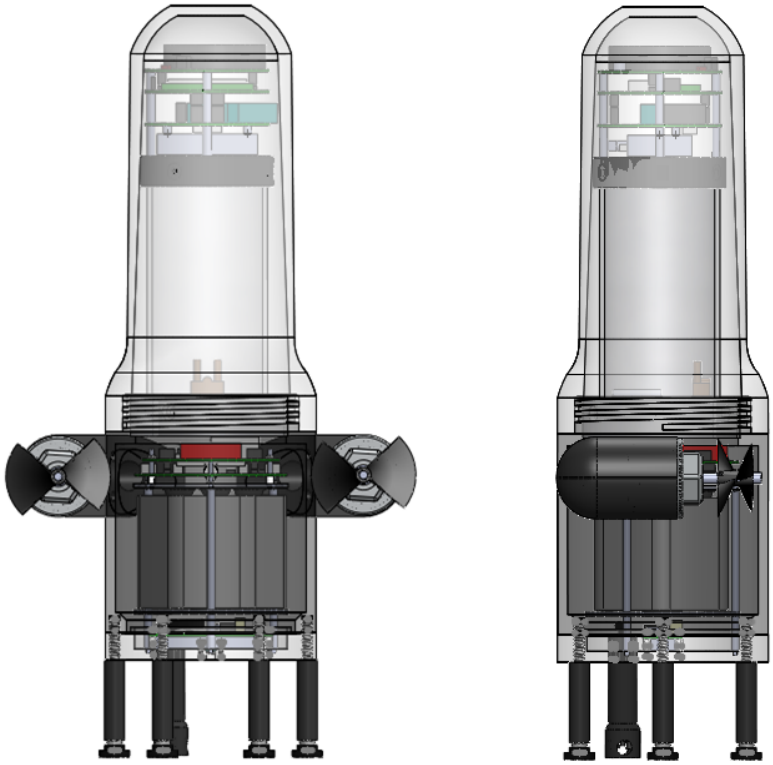

Figure 5.1: *Floating Sensor Network* vehicle design.

5.3 Design Methodology

During the design process of an electro-mechanical system with interrelated system components, it is important to be able to track and prioritize the desired functionalities of the final system. During the prototyping phase of the *Floating Sensor Network*, multiple prototype designs were rejected because they failed to address key system functionalities, which were overlooked in the design process. For example, Section 4.3 describes a prototype actuation system which failed to meet fundamental portability, serviceability and sensing requirements because they were not sufficiently managed during design. The prototype was developed without a formal design methodology, and its early failure to balance functional requirements underscores the utility of a system to manage the design process.

A general approach to prioritizing and interrelating system components during the design process has been adapted from the Ishikawa "cause and effect" process [48, 107]. In this approach, the desired system functionality is related through a cause and effect or "fishbone" diagram. The item in the center of the fishbone diagram is the desired outcome for the system, and the branches, or "causes" leading up to that effect are arranged by priority. In the diagram, components or "causes" with higher priority are closer to the desired outcome for the system. This approach facilitates prioritization of system functionality during the

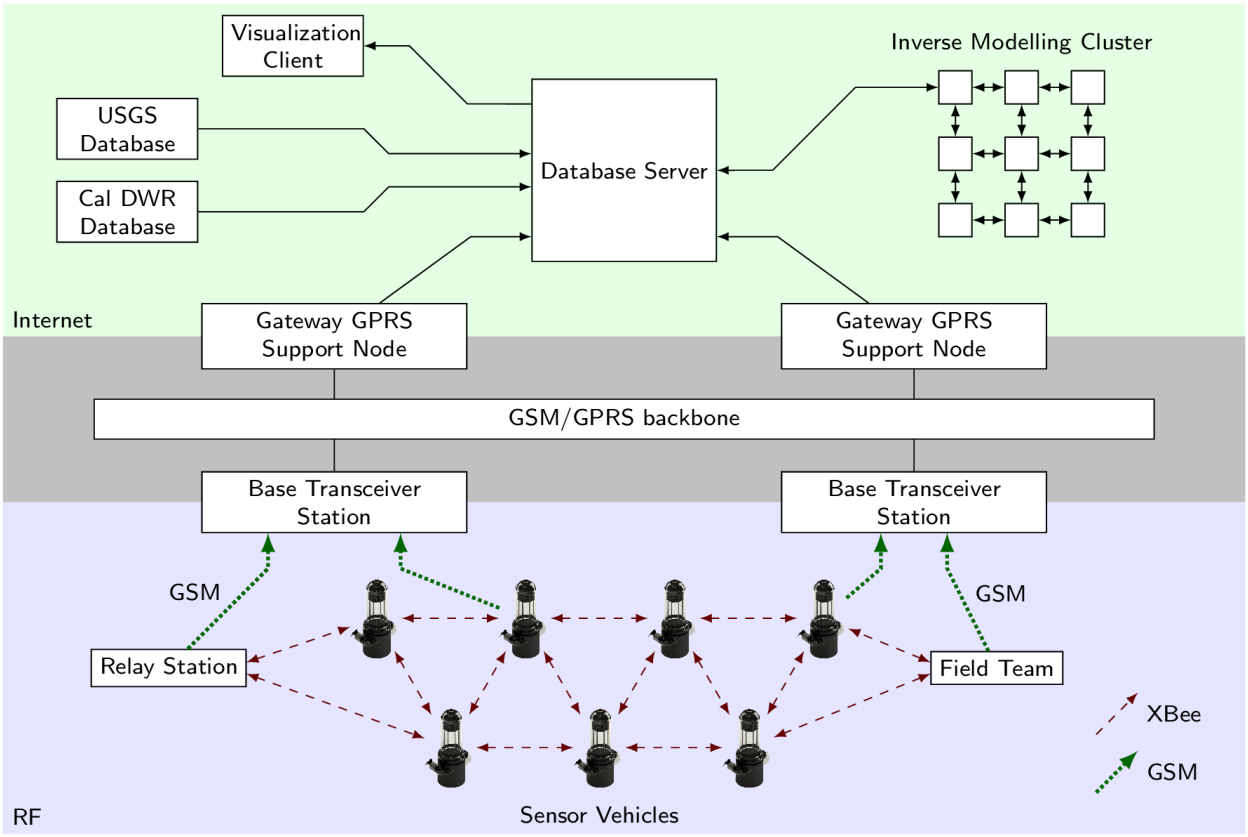

Figure 5.2: Schematic of the drifter-backend system, including floating sensors and backend computational infrastructure.

design process and guides the selection of specific system components. In this manner, once the general system requirements (or "functional requirements") are enumerated and prioritized, a second fishbone diagram can be created with specific "design parameters" in place of the functional requirements.

The Ishikawa diagram in Figure 5.4 presents the generic functional requirements of the *Floating Sensor Network*. Figure 5.5 presents an Ishikawa diagram with the same structure, but with each component replaced with specific "design parameters". For example, "sense location" has been replaced with GPS system, and "Lagrangian sensor" has been replaced with "vertical profile, symmetric drag". The following sections provide a more detailed explanation of how the design parameters for each component on the Ishikawa diagram affect the final system design.

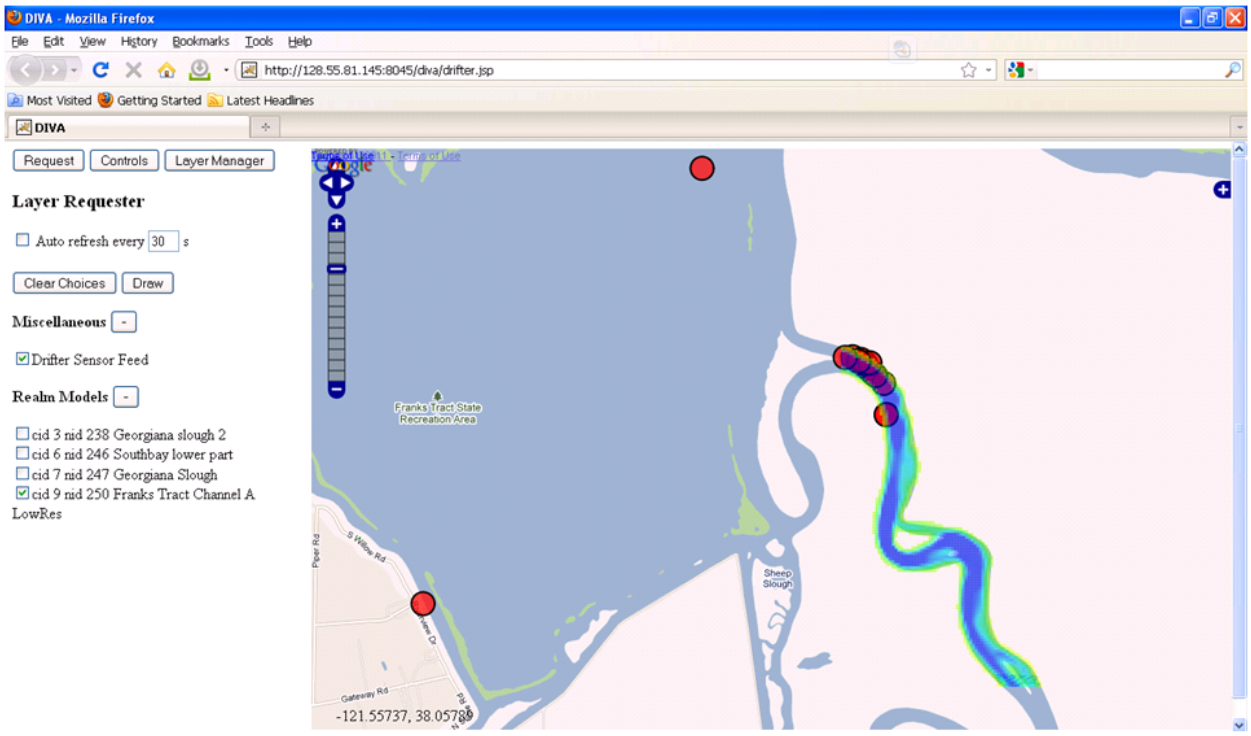

Figure 5.3: Vehicle position and water velocity computed from the Ensemble Kalman Filter algorithm running in the background server are available on the web in real-time.

5.4 Vehicle Design

Sections 4 and 5 follow the prioritization order from the Ishikawa diagram. They explain in detail how the functional requirements from Figure 5.4 evolved into the specific design parameters listed in Figure 5.5 during the design process.

Form Factor

As indicated by its position in the Ishikawa diagram, the form factor (or overall size and shape of sensor) is one of the most important considerations in the design process since it determines the vehicle's effectiveness as a Lagrangian sensor. Previous studies have indicated that appropriate Lagrangian sensors must maximize their cross sectional area to flow and should present a roughly symmetric drag profile [94]. Additionally, feedback from researchers and operators in the Office of Naval Research indicated that each sensor should be man-portable since the vehicles are usually deployed by a small team from a boat with limited space and/or payload.

To ensure that the vehicle lends itself to Lagrangian sensing, its hull is a vertically-oriented cylinder. It is designed to sit low in the water with mass and volume configuration

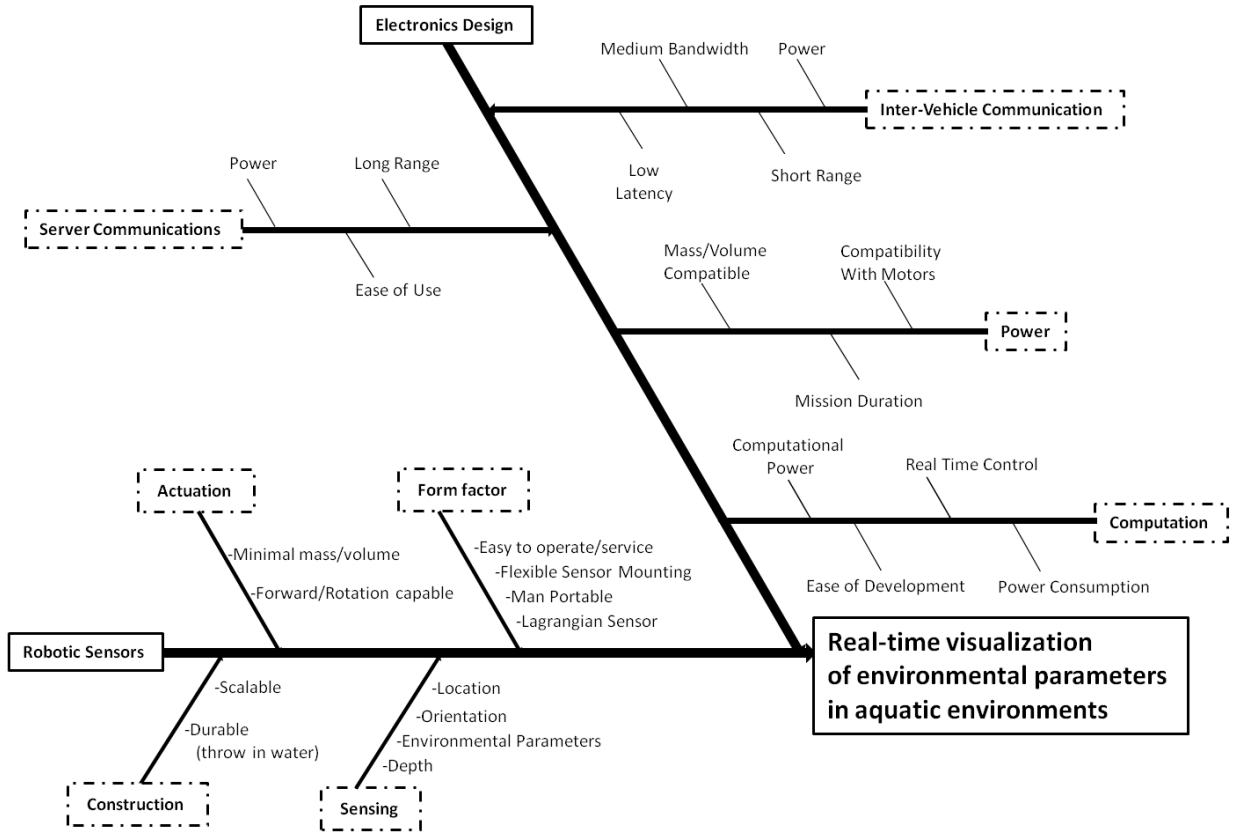

Figure 5.4: Complete Ishikawa diagram for design process with functional requirements.

that makes it hydrostatically stable. It also must have a sizable cross sectional area to ensure that it settles into the local flow as quickly as possible.

To ensure portability, size and weight constraints were enforced during the design process. Each vehicle was not to weigh more than seven kilograms, exceed twenty centimeters in diameter or fifty centimeters in length.

In order for the system to be broadly applicable for multiple end users, the vehicle must be able to carry a variety of immersed sensors. Therefore, a modular PVC sensor mounting plate was designed into the bottom of the vehicle as shown in Figure 5.6. It is 12 centimeters in diameter and 2 centimeters thick, ensuring ample room exists for a variety of threaded interfaces and o-ring seals.

Finally, to ensure that the vehicles are field serviceable, each functional system is a modular unit which can easily be removed and replaced. As shown in Figure 5.7, the vehicle consists of the following systems: propulsion pods, power electronics and sensors, and communications electronics. With the exception of the propulsion pods, each system can be removed and replaced without the use of hand tools.

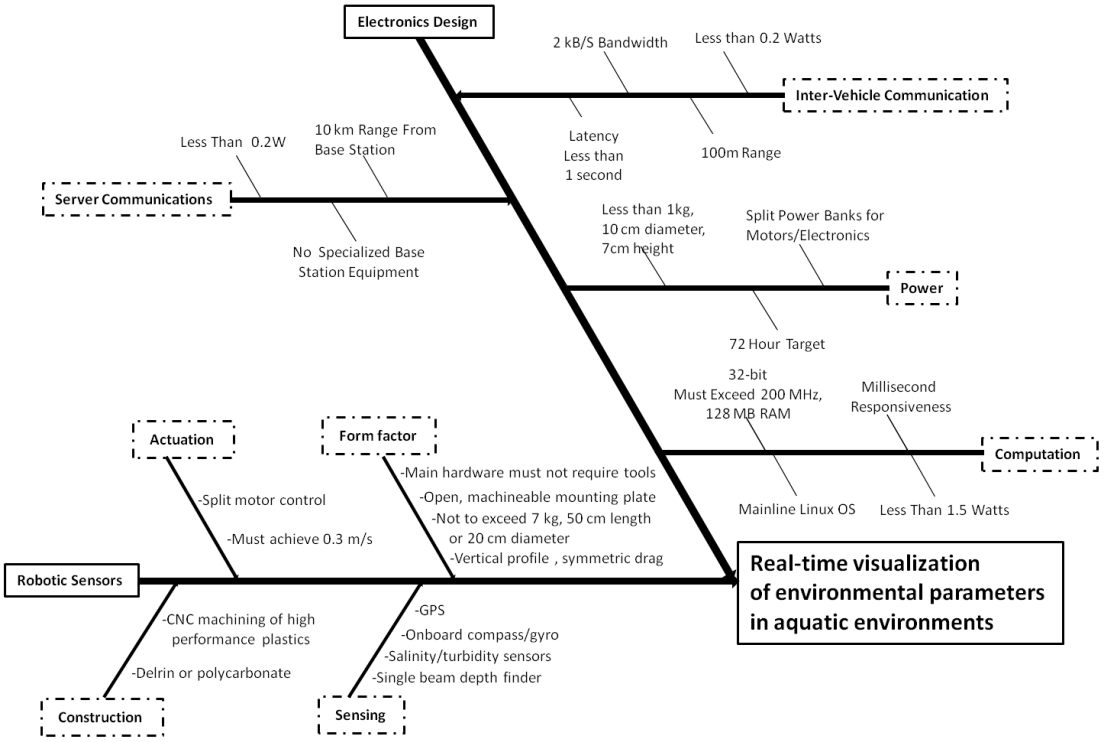

Figure 5.5: Complete Ishikawa diagram for design process with design parameters.

The height of the antennas for communication and sensing is an important design parameter dictating form factor. Early tests with prototypes indicated that the antennas for the GPS, GSM, and 802.15.4 radios all needed to be at least 2.5 centimeters above the water surface to guarantee a consistent signal.

Another parameter dictating form factor and overall dimensions was mission life. The external diameter of the main hull is fixed by the pre-fabricated housing chosen for the upper hull. The overall length of the vehicle into the water could be varied to accommodate larger or smaller batteries. The final vehicle length was chosen to accommodate a battery capable of 72 hours of mission life (see Section 5 for details) while not violating the man-portability functional requirement.

Finally, it was imperative that the mass be distributed so that the vehicle be hydrostatically stable in the desired orientation. Therefore the battery was located low in the vehicle to ensure that the center of mass was below the center of buoyancy. The motor pods were also located as close to the center of mass as possible in order to minimize any unnecessary moments acting on the vehicle during steady state operation.

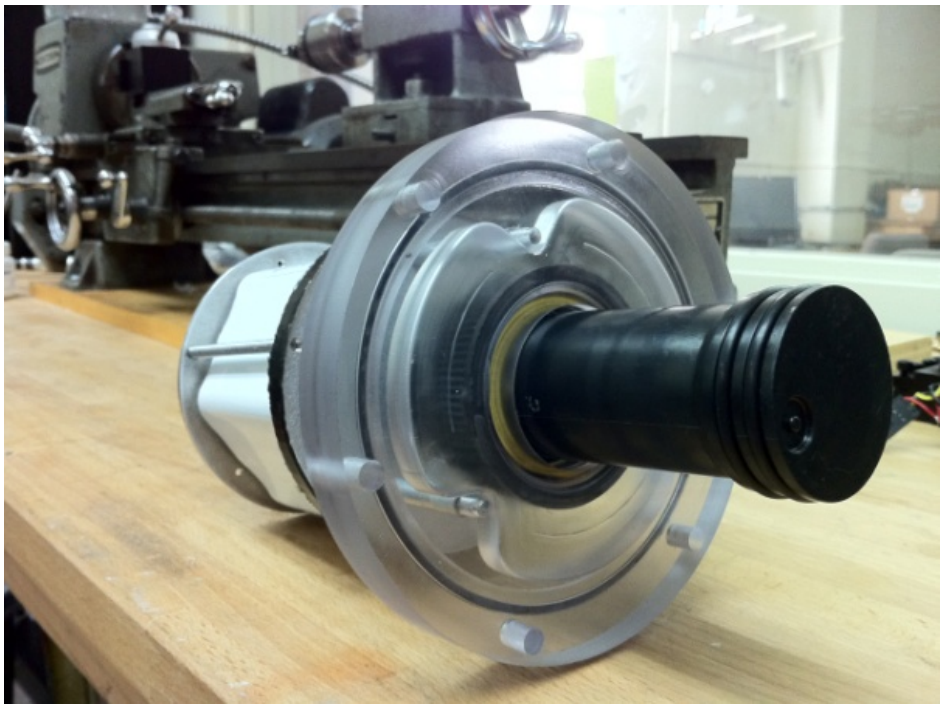

Figure 5.6: Side view of lower sensor mounting plate with single beam depth finder installed.

Sensing

The fundamental sensing mission of the vehicle is to measure the river velocity and to make measurements of water quality factors. Because velocity is not uniform across the water column, hydrodynamic models are needed to infer the velocity deep in the water, hence the necessity of the backend computing infrastructure discussed previously.

The water quality sensor currently used is the Omega CDE222 temperature and electro-conductivity sensor. An estimate of the salinity of the water can be made from these two quantities. The Omega CDE222 is normally a hand-held laboratory sensor; it is inexpensive, and its form factor is convenient for integrating directly into the drifter, making it a good choice for a representative water quality sensor.

The most convenient way to estimate position and velocity in outdoor environments at reasonable accuracies is a civilian GPS unit. Factors to consider when selecting an OEM GPS unit include power consumption, form factor, accuracy, and ease of integration. The Magellan AC-12 GPS unit is slightly larger than other available units, but has higher accuracy; 1.0 m *Circular Error Probable* (CEP) as opposed to 2.5 m CEP, which is the more usual accuracy of low-cost, commercial grade OEM GPS units. This GPS unit also provides an estimate of the velocity of the drifter. This is important both for the passive sensing function of the drifter as well as for control during active propulsion. Because of the non-trivial interaction between the thrust generated by the propellers, the rotational and linear

Figure 5.7: Modular structure of vehicle's internals: (from left to right) Delrin lower hull with removable motor pods, lower electronics assembly with battery and sensors, upper electronics assembly with microprocessors and guidance/location sensors, and clear PVC upper hull.

drag from the surrounding water, and the unknown velocity of the surrounding water, it would be very difficult to estimate the velocity of the drifter based on the control input to the propellers. It is far easier to directly estimate the velocity of the device using the GPS module.

Estimating the vehicle heading is very important for any guidance or motion control operations. An estimate of heading can be made from GPS readings alone, but at the low speeds at which the vehicle moves, these estimates would be inaccurate to the point of inefficacy. A separate sensor, the Honeywell HMC6532, was chosen to provide heading information. The HMC6532 is an integrated circuit package containing two magnetometers and logic circuitry to translate those readings into a heading measurement relative to magnetic north. It provides roughly 2° accuracy at 20 Hz, which is appropriate for the heading control tasks faced by the vehicle.

Vehicles which have been modified for use with the Office of Naval Research also include a Hummingbird single-beam depth finder for bathymetry mapping.

Actuation

The objective of adding propulsion capability to a Lagrangian sensor is to enable it to stay off the banks and clear of obstructions in shallow water. In the case of adaptive sampling, it also enables the drifters to position themselves near features of the water which the user might want to monitor. Initial experiments with a fleet of ten passive sensors in the Sacramento-San Joaquin Delta indicated that sensors got stuck after approximately one hour if left unattended. Therefore, the propulsion system was designed to re-orient the vehicle and enable it to perform brief course corrections to keep clear of river boundaries. Given that the vehicle must present a large cross sectional area to flow, it is not expected to achieve high velocities. A design speed target of 0.3 m/s was set to enable cross stream movement, but not upstream movement or station keeping.

Initial actuated prototypes featured a single propeller with a gear driven rotary pod to re-orient the propulsive force as shown in Figure 5.8. The gear train occupied excessive space inside the vehicle and it was determined that the man-portability functional requirement would have been violated for this design approach to work. Also with single motor actuation, stability of the controller became an issue quickly.

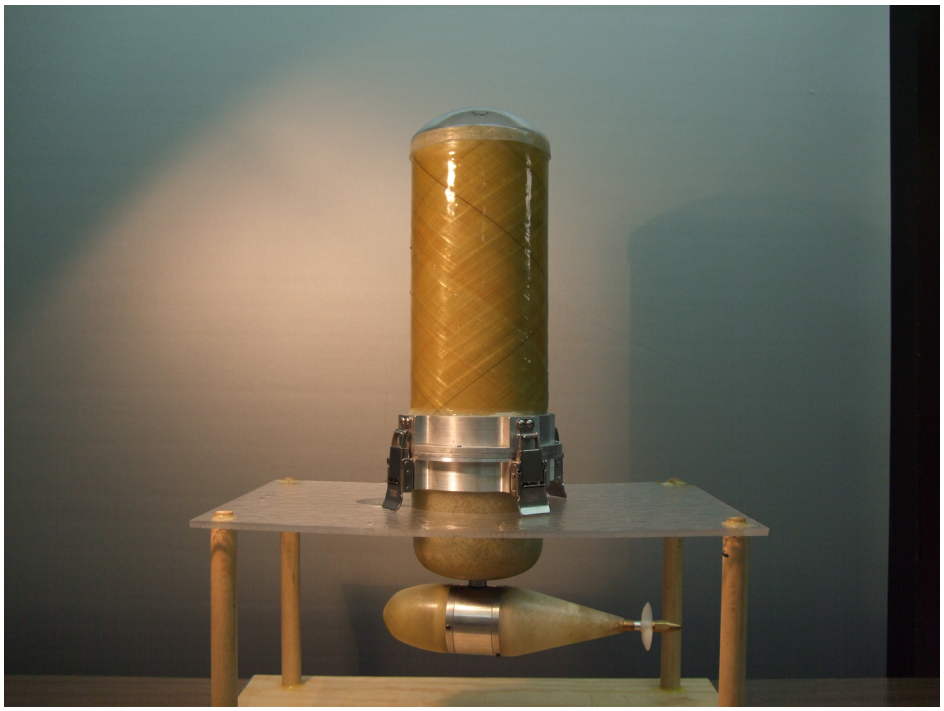

Figure 5.8: Single propeller design candidate which did not meet portability functional requirement and presented stability issues.

Therefore the vehicle was redesigned with split motor control. In this configuration, the vehicle has two motor/propeller modules, one on either side of the vehicle. This allows two

degrees of freedom for independent control of orientation and forward velocity. To achieve forward motion, the onboard computer compares the desired GPS waypoint to the current location of the vehicle. The computer calculates the bearing that the vehicle must maintain to reach the GPS waypoint, and the vehicle uses the electronic compass to drive itself toward the desired waypoint along the required heading. A feedback controller is used to control the two independent motors in order to maintain desired heading and speed, given feedback from the compass and GPS.

Angular Velocity Characterization

In order to investigate the suitability of different heading controllers for the sensor, a mathematical model of the vehicle's response to motor inputs was desired. A pre-production prototype was designed with split motor control and optical encoders on each motor shaft so that the angular response of the vehicle could be mapped to the difference in revolutions per minute (RPM) between the two motors. To generate data for the model, a high speed camera was used to track the vehicle's angular displacement as a function of time to known motor inputs.

The model is based on the law of dynamics for angular motion, i.e. that the net torque on an object equals its moment of inertia multiplied by its angular acceleration. In the equation below, θ represents the angular displacement of the vehicle, I is the moment of inertia, and τ is the torque.

$$\tau = I \cdot \ddot{\theta} \tag{5.1}$$

A net torque will act on the vehicle when there is a difference in RPM between the two motors, ΔRPM. There will also be a drag torque, which is assumed to be directly proportional to the angular velocity. Thus, the equation above becomes an ordinary differential equation in which ΔRPM acts as a control:

$$c_1 \Delta\text{RPM} - c_2\dot{\theta} = I \cdot \ddot{\theta} \tag{5.2}$$

In the equation above, c_1 and c_2 represent unknown constants to be determined by a polynomial approximation of the high speed camera data shown in Figure 5.9.

The vehicle response was modeled in Simulink. The block diagram of the plant is reproduced in Figure 5.10.

In the figure, "RPM" represents ΔRPM, "Theta" represents the angular displacement of the vehicle, and c_1 and c_2 are the constants defined previously.

Heading controller comparison

The creation of an accurate vehicle model enabled an investigation into the suitability of multiple heading controllers for the sensor. The function of the heading controller is to apply power to the motors in order to orient the sensor along a given trajectory in a stable

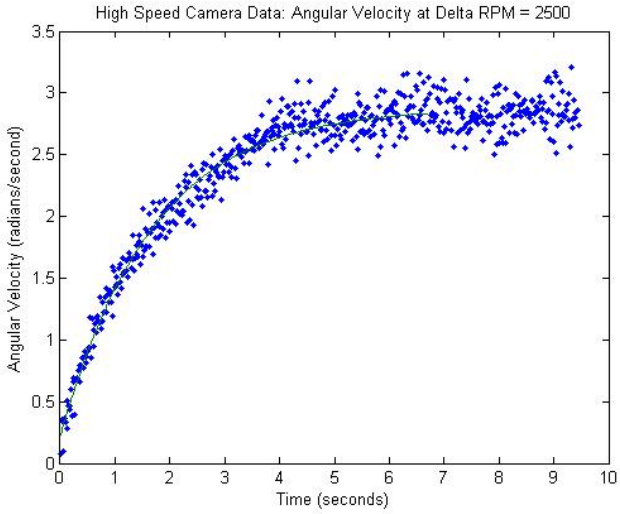

Figure 5.9: Angular velocity data for the vehicle model.

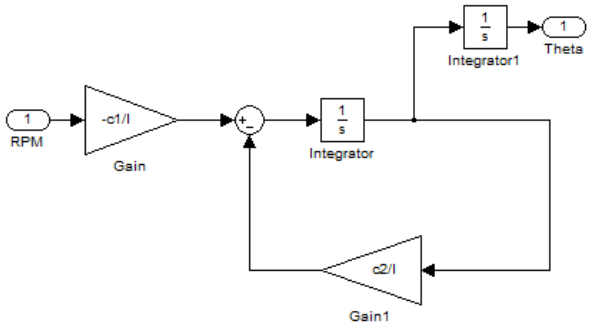

Figure 5.10: Vehicle plant for control simulation, designed in Simulink.

and time-efficient manner. Three types of controllers were investigated: proportional, PID, and model-based.

The simplest of these is a proportional controller. In this scheme, the magnitude of the control input is directly proportional to difference between the desired state and the current state. In this context, the desired state is the desired heading of the sensor and the current state is its current heading. The blue line in Figure 5.11 shows a simulation of a sensor starting at zero radians and turning to $\pi/2$ radians under proportional control. The simulation shows that the sensor overshoots the desired heading and oscillates before settling to the desired heading. This is expected because the form factor of the vehicle enables it to rotate rapidly. Therefore, the second motor is not capable of countering the sizable angular momentum of the vehicle as it approaches the desired heading, resulting in oscillations. The magnitude of the overshoot can be minimized by reducing the proportional control gain.

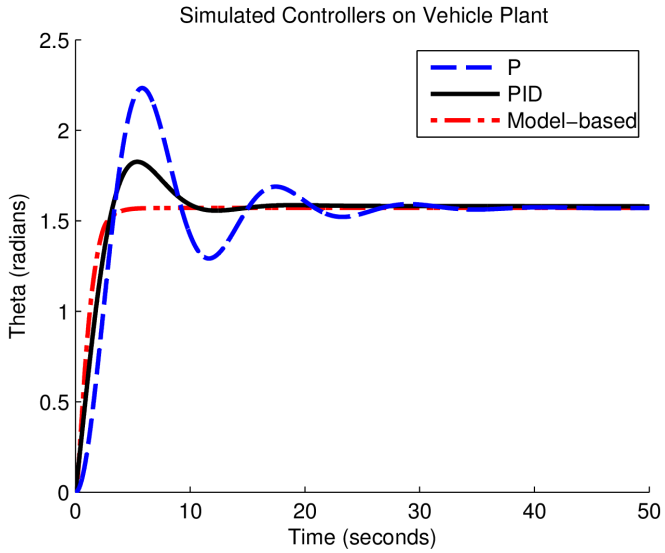

Figure 5.11: Simulink simulation of the effect of three proposed controllers on the vehicle plant.

This results in the sensor taking excessive time to reach the desired heading. Thus, it was deemed that proportional control would not be effective for this system.

In order to minimize the controller overshoot, a PID (proportional, integrator, derivative) controller was simulated. This control scheme adds terms for the integral and derivative to the calculation of the control effort. The derivative term prevents the overshoot by reducing the control effort if the sensor is rapidly approaching the desired heading. The integral term adds control effort as a function of time if the sensor dwells far from the desired heading. The relative control gains can be tuned heuristically or with automated methods. The black line in Figure 5.11 shows the result a PID controller applied to the system. It can be seen that the overshoot and oscillation is reduced in this scheme.

Finally, a model-based controller was simulated for the vehicle. In this scheme, the data from the vehicle model is uploaded to the vehicle's controller. This design has the advantage that it can more accurately predict when the vehicle is approaching the desired heading, minimizing the chances of overshoot. This controller worked well on prototypes and successfully minimized overshoot and settling time, as seen in the red line of Figure 5.11. The use of this controller requires that shaft encoders be installed on every production vehicle. It was determined that the added cost and complexity of this approach was not worth the marginal improvement in settling time compared to the PID controllers. Thus, the production fleet uses PID controllers.

Forward Velocity Characterization and Validation

Once the vehicle size requirements and control were determined, motors and propellers had to be sized to meet the target forward velocity of 0.3m/s. The vehicle drag coefficient estimate, C_d for forward motion is $C_d = 0.8$, based on calculations for an ideal finite cylinder [111]. Vehicle cross sectional area, A was determined by CAD program to be 0.032 m^2. These numbers are used to estimate the steady-state speed, v of the vehicle with respect to the water as it travels in a straight line with the sum of the two propeller forces operating at F_{prop}. Equating drag force to propulsive force (with water density, ρ):

$$F_{\text{drag}} = \tfrac{1}{2}\rho C_d A v^2 = F_{\text{prop}} \tag{5.3}$$

Solving for F_{prop} in this equation with $v = 0.3$ m/s, the desired propulsion force is 1.15 N or 0.58 N per motor. A test setup was developed consisting of a force load cell attached to a motor/propeller pod. The motor was driven by a speed controller at a range of speeds at 7.4 V, and the output force was logged as function of input power. A number of motor/propeller combinations were explored. The chosen combination produced the most force per unit input power near the 0.58 N target.

Figure 5.12: Motor force test platform: submerged motor unit is attached to an extended arm which pivots against a force transducer. Power is applied to the motor at a range of values, and output force is recorded in Matlab.

The actual speed of the vehicle was estimated during tests in an outdoor tank at the UC Davis Bodega Bay Marine Laboratory. By driving back and forth in the still water of the tank, while receiving GPS signals, the speed of the vehicle can be estimated. Two techniques were used: first, the GPS velocity signal itself was averaged over a run across the pool, providing an estimate of 0.264 m/s with a standard deviation of 0.036 m/s. An alternate method is to take a finite difference of the GPS positions, spaced 6 s apart: this

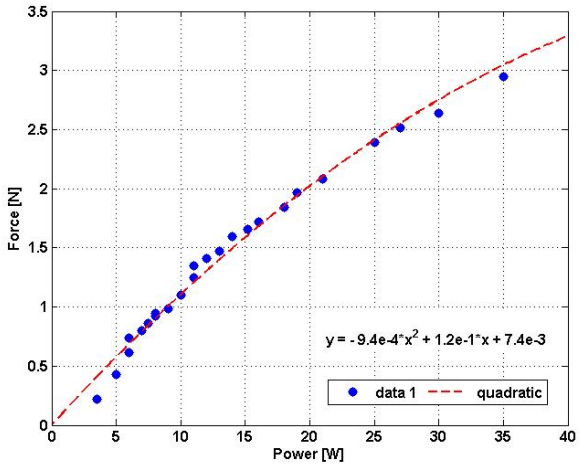

Figure 5.13: Output force as a function of input power for chosen motor/propeller.

method results in an estimate of 0.242 m/s with a standard deviation of 0.033 m/s. It is expected that the finite difference estimate would be lower, because the drifter does not travel in a perfectly straight line. Figure 5.14 shows the time series of the speed estimates by the two methods during a run across the pool, and Figure 5.15 shows the GPS positions gathered during the run.

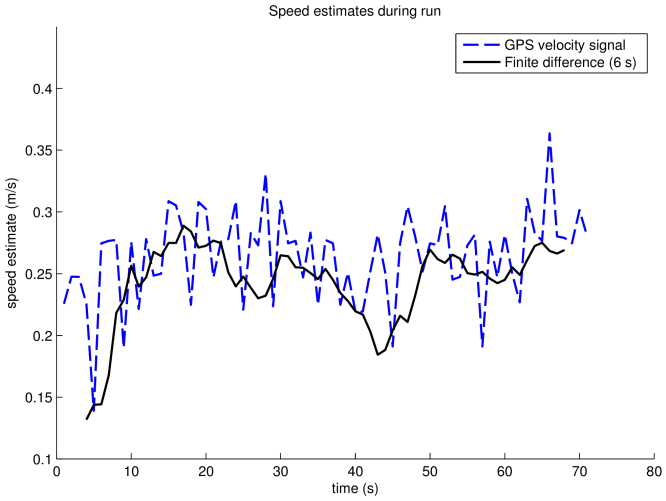

Figure 5.14: Vehicle velocity inferred from GPS signal.

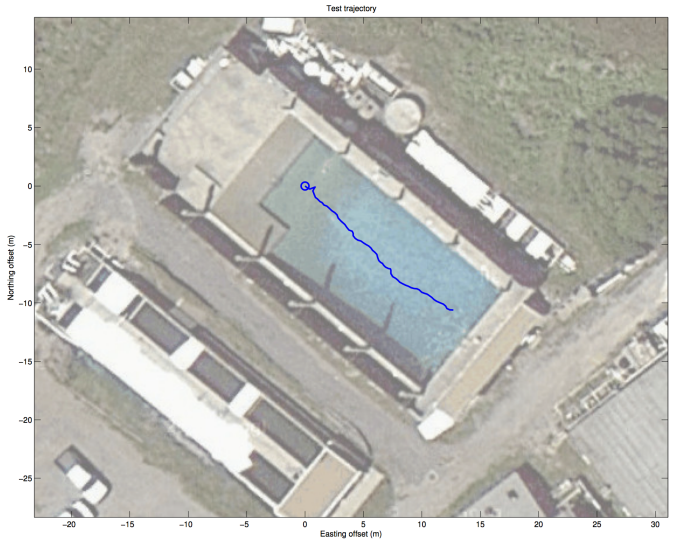

Figure 5.15: GPS trajectory from vehicle during velocity test.

An Alternate Approach: Mobile Phones as a Controller

A future version of the sensor may use the computational and sensing capabilities of a modern smartphone to act as a Lagrangian sensor. To control the motors and read the compass it is necessary to output PWM values to the speed controllers. This signal is fed to a speed controller which supplies the power necessary to drive a DC electric motor. This is enabled by an Android IOIO board which plugs into an Android phone via USB and allows the configuration of multiple pins for PWM out, digital IO etc. It also allows the user to access the compass and GPS data necessary for real-time control.

As can be seen in Figure 5.16, the control system hardware would consist of four primary components. The Android phone is connected to the IOIO board and is supplied with 11.1V DC from an external source. Pins 5 and 6 are configured for PWM out. The signals from these pins are wired with a ribbon cable into a custom-built speed control board. The IOIO board also has a 5V out which is wired to the speed control board, which is used to enable the motors. The speed controller board has a protection circuit, which enables users to disable the motors if necessary. The speed controllers are also powered from an external 11.1V source. The result is a system which controls the motors with feedback from compass readings, as can be seen in Figure 5.18.

In Figure 5.18, the desired heading of the vehicle is set at zero degrees. As the compass on the phone moves back and forth across zero degrees, the vehicle powers each motor separately to maintain the desired heading. The advantage of this approach would be dramatically lowered cost and a simplified user interface. However, this approach is limited in that there are heterogeneous properties of the phone, requiring tuning the controller for each phone, which is impractical for the fleet. It is necessary to test a control algorithm to find the tuning

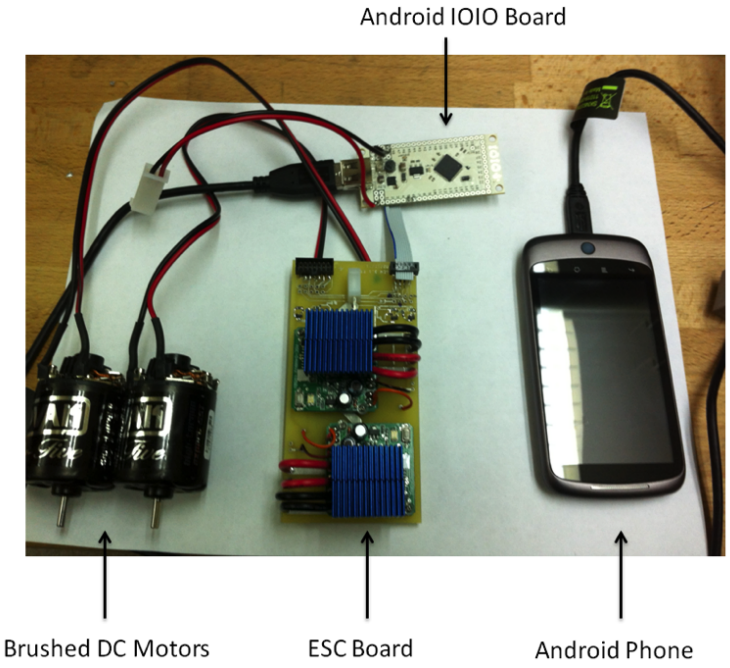

Figure 5.16: System components necessary to replace real time control system in actuated sensor.

Figure 5.17: Working system showing feedback control of motors from internal magnetometer.

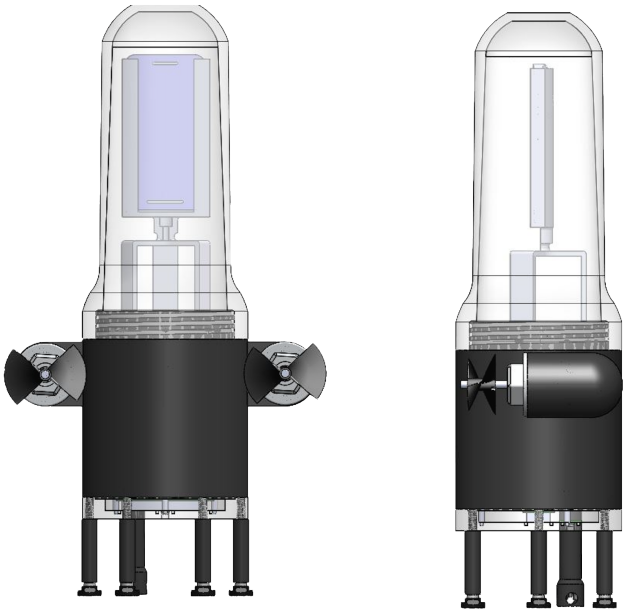

Figure 5.18: Modified active sensor with mobile phone.

parameters for waypoint control without explicitly being programmed.

The controller's performance could be quantified with the following cost function:

$$J(\theta) = \frac{1}{T - t_o} \int_{t_0}^{T} [r(t) - y(t, \theta)]^2 dt \tag{5.4}$$

In equation 5.4, t_0 represents the time at which the first peak occurs, T is the total time of the experiment, r is the desired heading, and y is the heading at any given time. Thus, this equation quantifies the total difference between the actual heading and the desired heading throughout the course of the experiment, except for the initial transient. A set of PID gains that quickly reaches the desired heading and does not overshoot drastically will therefore minimize the cost.

For this application, it is necessary to include the initial transient (since it represents the period during which the vehicle is driving itself toward the desired heading). Thus the cost function is as follows:

$$J(\theta) = \frac{1}{T} \int_{0}^{T} [r(t) - y(t, \theta)]^2 dt \tag{5.5}$$

As the experiments are run, an extremum-seeking (ES) algorithm [54] estimates the gradient of the cost function and modifies each of the PID gains to follow the gradient to a

local minimum of the cost. The algorithm is as follows:

$$\begin{aligned} \xi(k) &= -h\xi(k-1) + J(\theta(k-1)) \\ \hat{\theta}(k+1) &= \hat{\theta}_l(k) \\ \theta(k+1) &= \hat{\theta}_l(k+1) - \alpha_i \cos \omega_i(k+1) \end{aligned} \tag{5.6}$$

This is a discrete time implementation of a high pass filter on the cost signal, a integrator, a low pass filter and finally a perturbation signal that generates new PID gains. The purpose of each step is as follows: the high pass filter removes the DC component of the cost signal, which is then multiplied by a discrete time sinusoid to extract the portion of the cost that is due to perturbation of the parameter estimate (the gradient estimate). The gradient estimate is then integrated with step size γ and added to the perturbation signal to obtain new PID gains. Thus in the above equations, k represents the iteration number, h is the cutoff frequency of the high pass filter, γ is the integration gain, θ is a vector of the PID gains, i is the index of the θ vector, and $\hat{\theta}$ is an estimate of the PID gains. Finally, ω and α are the perturbation frequency and amplitude for the updated PID gains. Note that the perturbation frequency and amplitude are different for each PID gain.

This method follows the gradient of the cost function toward a local minimum.

$$\begin{aligned} \omega_i &= a^i \pi \\ 0 &< a < 1 \\ 0 &< h < 1 \end{aligned} \tag{5.7}$$

Initial Controller Parameters

The D and I gains are both set to zero and the P gain is increased from zero until the system starts to oscillate. The gain value at which this occurs is recorded as Ku and the time period of oscillation is recorded as Tc. The PID gains are then prescribed based on the following formula:

K_p	K_i	K_d
0.6 K_c	$2\frac{K_p}{T_c}$	$\frac{K_u P_u}{8}$

The ES algorithm is initialized with the following coefficients:

K_p	K_i	K_d
180	32.7	248

The response of the assumed plant is shown in Figure 5.19

This response is taken as a starting value for the ES algorithm and the system is run for one thousand iterations. As expected, the gain coefficients are iteratively modified to reduce the cost, shown in Figures 5.20 and 5.21.

The final gains are shown in the table below, and the system response to these gains is shown in Figure 5.22

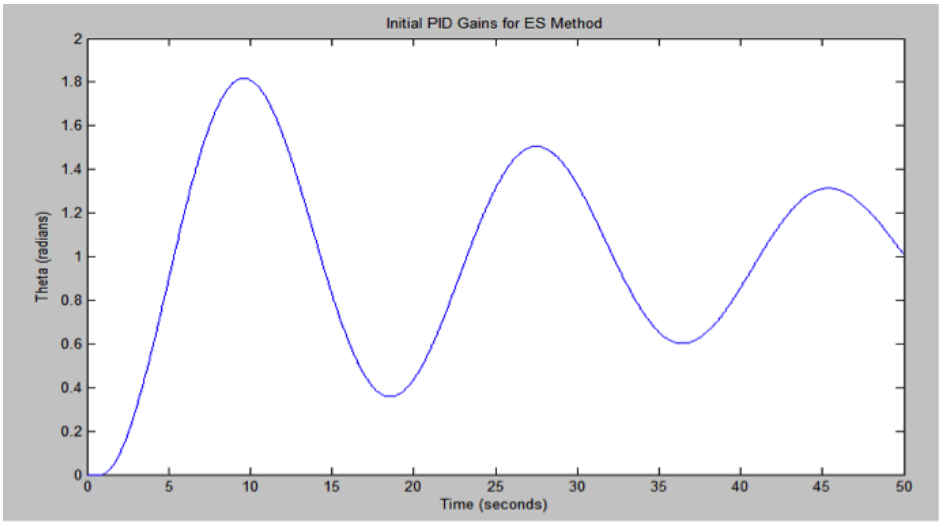

Figure 5.19: System response to initial control gains.

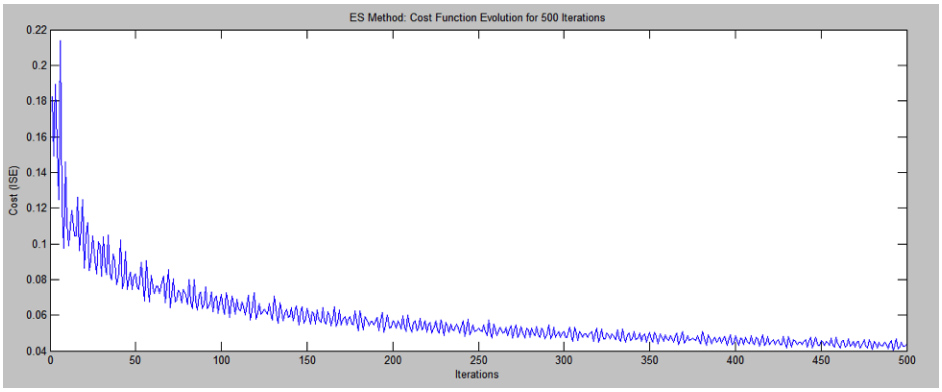

Figure 5.20: Evolution of cost function.

K_p	K_i	K_d
435	270	10

The results accord well with expectations. After one thousand iterations, the cost of the hand-tuned PID gains is over 2.6 times higher than for the ES gains. The ES response also minimizes target overshoot and stabilizes more quickly. After fifty seconds, the hand-tuned method still has not settled, whereas the ES method settles after thirty seconds. The model assumptions are then modified, doubling the moment of inertia and halving the power of the motors. The results are shown in figure 5.23.

As expected, increasing the power of the motors while decreasing the inertia requires a lower proportional gain, but similar I and D gains. The plant was also modified to model a time delay in real-time control computation. The resulting plant is shown in Figure 5.24.

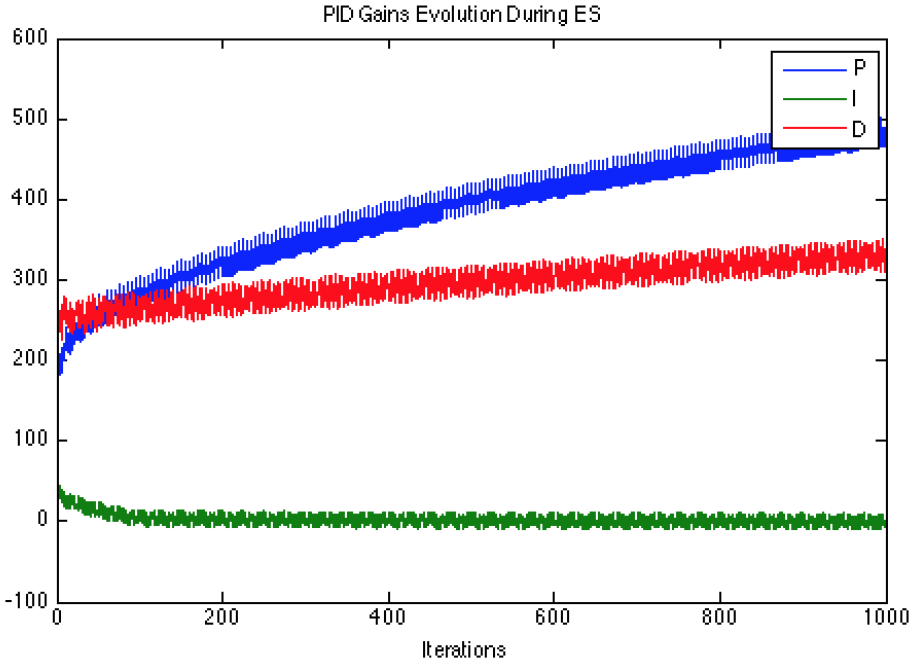

Figure 5.21: Evolution of PID coefficients.

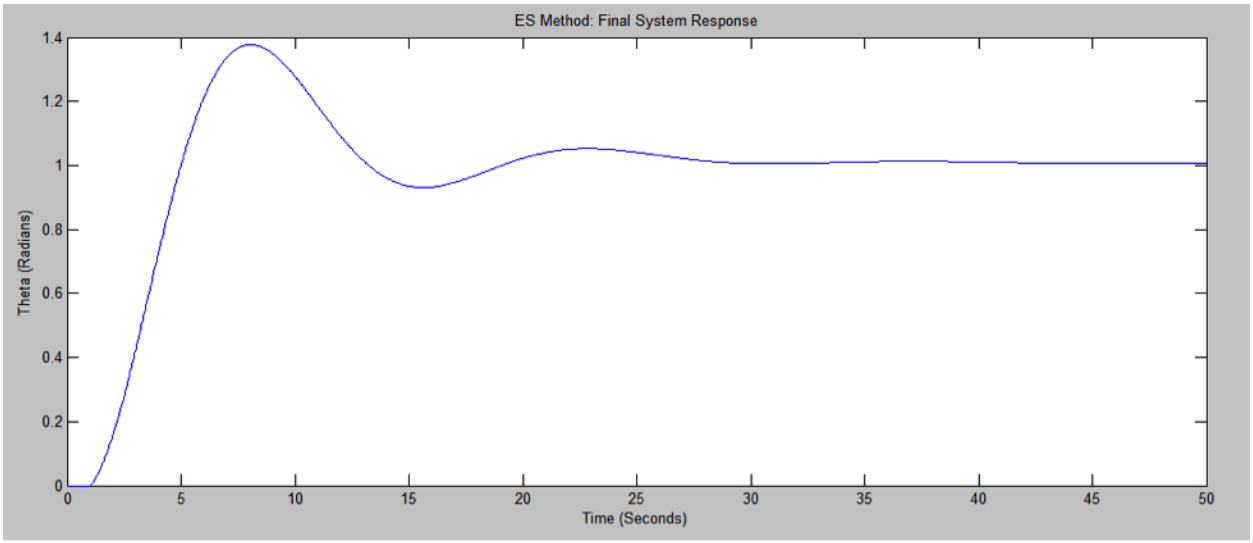

Figure 5.22: System response to control gains after ES algorithm.

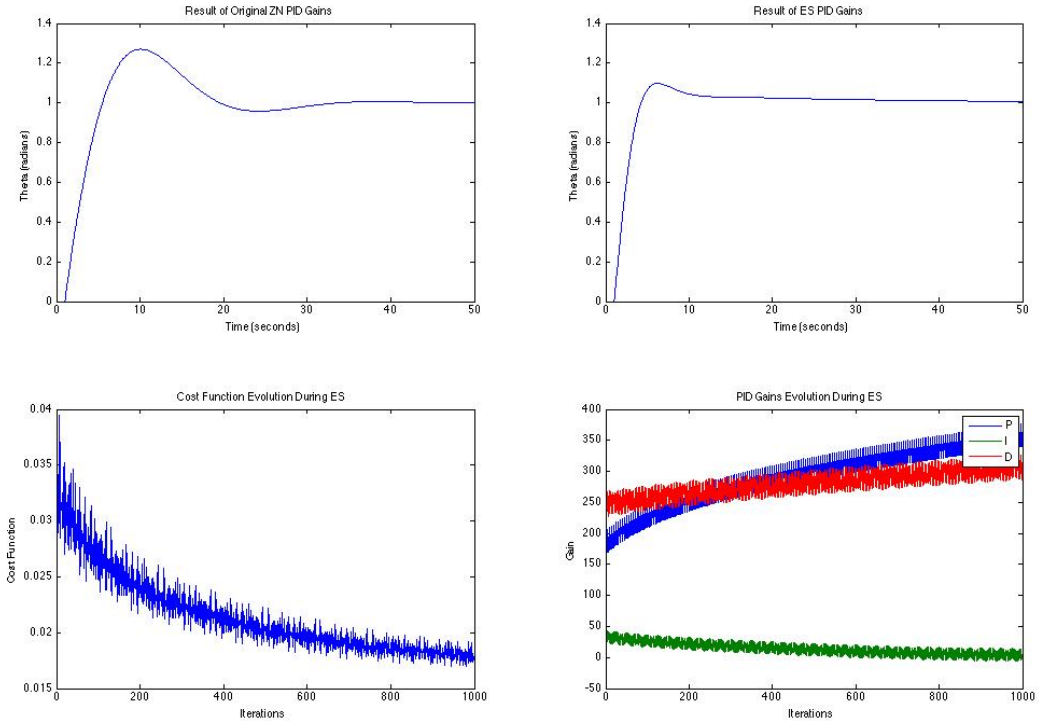

Figure 5.23: A drifter with twice the power and half the inertia.

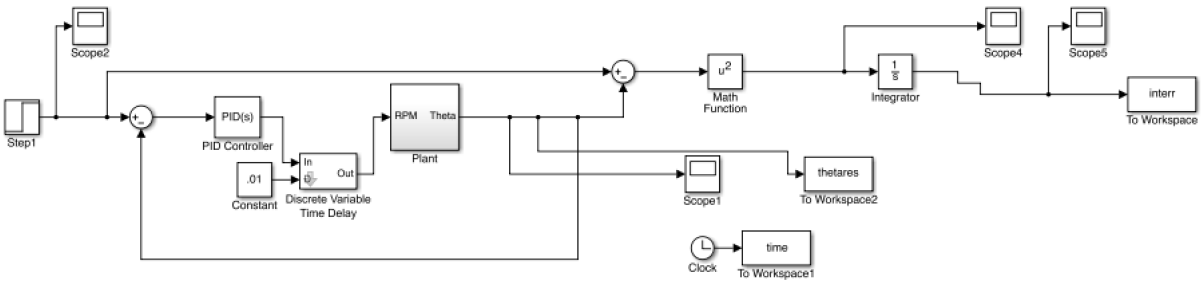

Figure 5.24: Plant with time delay.

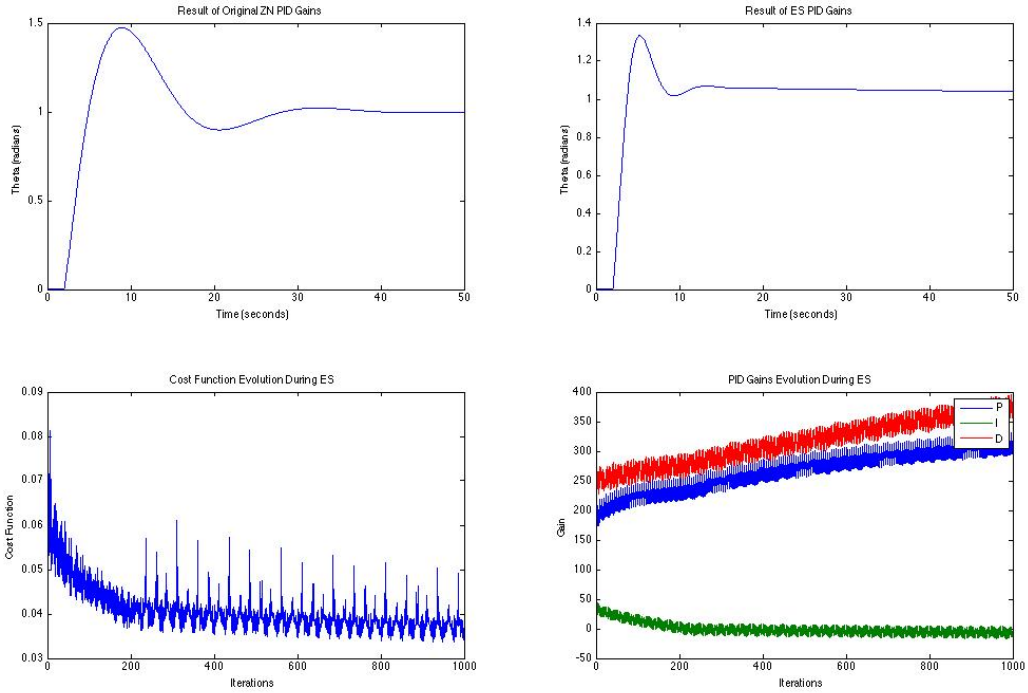

Figure 5.25: Drifter with one second time delay.

The results of the time delay system are shown in Figure 5.25. Even with time delay and heterogeneous fleet properties, the extremum-seeking algorithm was able to find a reasonable set of controller parameters. This indicates that the future cost of each sensor could be brought down using off the shelf smartphone components, even if fleets carry heterogenous payloads and real-time guarantees are not provided by the operating system.

Depth Controller

Many important features of hydrological systems exist under the surface of the water. Therefore, it is desirable that the sensor be able to control its motion along the water column to take measurements at known depths.

During the design process, the possibility of adding this capability was investigated within the requirements outlined in the Ishikawa diagram. A prototype sensor containing an adjustable ballast system was created. Ballast is added by drawing surrounding water into the vehicle through silicone tubes connected to a gear pump, as shown in Figure 5.26. The water is contained in a sealed reservoir in the upper hull of the vehicle.

A pressure sensor embedded in the lower PVC mounting plate is used to determine depth of the vehicle below the water surface, as shown in Figure 5.27. A proportional controller

Figure 5.26: Prototype sensor with buoyancy control.

uses feedback from the pressure sensor to add and remove water from the ballast reservoir to adjust the sensor depth.

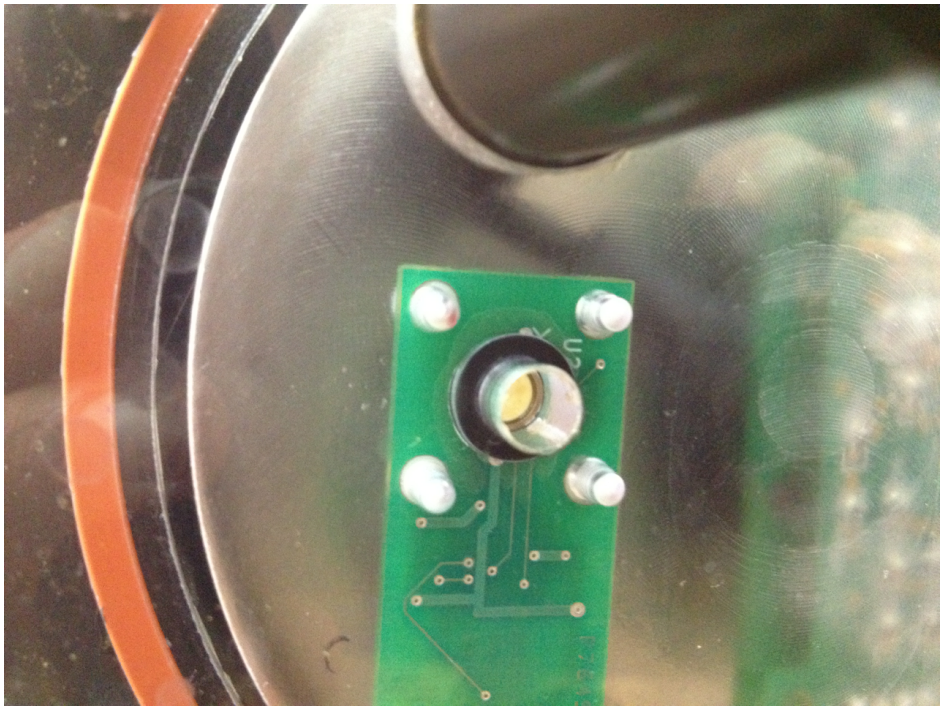

Figure 5.27: View of bottom PVC mounting plate with embedded pressure sensor.

Laboratory experiments in closed tanks revealed that this was a viable method of controlling depth. The sensors were able to perform multiple controlled dives to 5 meters depth. Integrating buoyancy into the final operational system presented non-trivial challenges.

First, the sensors lose GPS, GSM, and 802.15.4 connectivity when submerged. This presents two issues: the sensor's position is no longer known, and it is no longer able to communicate in the event of a failure. The first challenge could be addressed with a nine-degree of freedom inertial measurement unit (IMU). These units use filtered data from 3-axis accelerometers, gyroscopes, and magnetometers to estimate vehicle position from a known starting point. However these are expensive and would require additional computational complexity. The communication challenge could be solved by creating a separate pod for communication equipment that stays on the surface of the water while the main hull is submerged. This would complicate the vehicle design and impinge upon the man-portability requirement outlined in the Ishikawa process.

Second, the failure modes of the system as designed can result in complete loss of the vehicle. If any component of the buoyancy system fails, it will begin to flood the vehicle. This would damage the electronics and likely result in the vehicle becoming completely submerged before a field team could retrieve it. This could be mitigated by adding a leak sensor that would automatically trigger the pump to empty the reservoir at maximum rate. However, if

the pump jammed or a tube became obstructed, this safety mechanism would be rendered useless. Therefore, though the current production vehicles can be modified for buoyancy control, this capability is not used in field operations.

Construction

As shown in the Ishikawa diagram, there are a number of functional requirements for the hull construction of the sensor. The most important parameters require that it be weatherproof and waterproof. This requires that the material maintain its tolerances underwater within a wide range of temperatures. It is also important that the design be scalable, so that production can be rapidly increased after the first prototypes were produced.

Therefore, a high performance plastic, Acetyl Homopolymer (Delrin) was chosen for the main hull construction. It is well suited to CNC machining which allows scalability in manufacturing. It is also a material that can be injection molded at larger production quantities. Finally, it only absorbs 0.2% of its volume in water during 24 hour submersion and has an operating temperature range of -29 °C to 89 °C.

5.5 Electronics

Computation

The vehicles were designed to be used as a research platform for investigating implementations of distributed robotics, multivehicle control, environmental sensing, and data assimilation. The computational requirements for controlling the vehicle and fulfilling mission goals therefore vary widely between scenarios. The computational hardware was thus chosen to provide flexibility.

An embedded microprocessor system supporting the Linux operating system offered an efficient development process. Multitasking, file systems, databases, inter-process communications, and advanced memory management are examples of some of the useful features of a Linux operating system in this context.

One disadvantage of high-level systems such as a Linux-based embedded computer is that real-time control can be difficult to implement. This is because guaranteed timing of control tasks (which is essential for real-time control) is difficult to achieve in a general purpose operating system. A common solution, and one adopted here, is to split the control functionality between the high-level, non-real-time microprocessor and a low-level, real-time microcontroller to handle fast, time-critical control tasks.

The selected embedded microprocessor is the Gumstix Overo Water, a single-board computer using the Texas Instruments OMAP 3530 Applications Processor running at 720 MHz, 512 MB of RAM, and a 2 GB MicroSD card for storage. The Overo comes installed with mainline Linux 2.6.34 using the OpenEmbedded distribution and toolchain [57]. Earlier generations of Gumstix products featured a Gumstix microprocessor combined with an Atmel

Component	Voltage	Current	Duty cycle	Power
Overo	3.7 V	0.3 A	50%	0.51 W
G24	3.7 V	0.2 A	5%	0.04 W
XBee-PRO ZB	3.3 V	0.3 A	5%	0.04 W
Motors	7.4 V	2.4 A	10%	1.8 W

Table 5.1: Component power requirements.

ATMEGA128 microcontroller for low-level tasks such as motor control [39]. Using this architecture as inspiration led to the selection of the Atmel XMEGA128 [6] (a next-generation version of the ATMEGA) for the motor-control and other low-level control tasks.

Power

Rechargeable electrochemical batteries are the cheapest and most convenient way to store the electrical energy needed for the on-board electronics and motors. The power budget for the electrical systems is shown in Table 5.1. The energy storage requirement is set by the power budget and mission time; the major remaining choice for the battery system is the battery chemistry, which will determine the mass, volume, and material cost of the battery that can store the required power. A summary of battery chemistries and their energy densities is shown in Table 5.2. Lithium ion and lithium polymer batteries are the only ones that are compatible with the mass and volume constraints of the hull (as discussed in section 5.4).

The electronic components and the motors require different input voltages. Although it would be easy to incorporate a voltage regulating circuit to lower the voltage for the electronics, allowing all systems to share a common battery, there are reasons why it makes sense to keep the motor power separate from the electronics power. First, the inrush current when motors start can cause a voltage slump which could cause drop-outs in the electronics power; second, brushed DC motors can generate noise on the power lines that could adversely affect the electronics; third, if the motors are run so long as to exhaust the battery, and the electronics are on the same battery, the vehicle will no longer be able to gather data or transmit its location. The first and second problems could be mitigated by careful design of the voltage regulator circuitry, but it is easier to simply keep the two power sources separate.

The selected design was a hexagonal pack of 19 cylindrical lithium ion cells, with five cells dedicated to the electronics (3.7 V, 170 kJ, allowing 80 hours of electronics operation) and 14 cells dedicated to the motors (7.4 V, 480 kJ, allowing 74 hours of operation at the 10% duty cycle). See Table 1 for component power requirements.

Communications

There are three primary reasons to communicate with a drifting sensor in the field: (1) to discover the local water conditions for real-time sensing applications; (2) to track the sensor's position for retrieval, or to query its health and operational status (battery energy

Chemistry/format Example	Specific energy	Energy density
Lithium ion Panasonic NCR18650 [75]	840 kJ/kg	2100 kJ/L
Lithium polymer Sanyo UPF673791 [102]	760 kJ/kg	1800 kJ/L
Nickel-metal hydride Panasonic HHR110AAO [45]	200 kJ/kg	620 kJ/L
Lead-acid Panasonic LC-P0612P [58]	130 kJ/kg	370 kJ/L

Table 5.2: Representative battery capacities for various chemistries.

remaining, etc); (3) to share data between vehicles for multi-vehicle control applications or with the command center for remote actuation of the drifter.

These goals have different requirements for transmission range, bandwidth, and latency. Goals (1) and (2) above have very low bandwidth requirements; 0.01–0.1 kB/s, with up to 30 s latency, would be acceptable. These transmissions will need to be sent over distances of kilometers or greater. By contrast, goal (3) requires more bandwidth and lower latency; 2 kB/s with < 1 s latency. For multi-vehicle control applications the vehicles can be assumed to be relatively close: 100 m is a reasonable range. These diverse requirements can be best addressed with two separate communication networks.

The range requirements of the long-range communication system pose a challenge in the estuarine environment. The vehicles themselves might be up to 10 km away from their origin. Islands, levees, trees, and buildings are all interfering obstacles. The best self-contained solution would be to erect a communication tower on-site to minimize the fading through these obstacles. Although truck-mounted portable tower solutions exist, there is a more convenient option: the civilian *Global System for Mobile Communications* (GSM) network. Using a GSM module such as the Motorola G24 [33], the drifting sensor can use the GPRS (General Packet Radio Service) of the GSM protocol suite [2] to open a TCP connection to any server on the Internet. The guaranteed minimum bandwidth of a GPRS connection is 9.6 kB/s, but the latency is not guaranteed. Empirical tests show that latency of 1–5 s is common.

Due to the larger latency, the GSM solution is inappropriate for multi-vehicle control applications. At shorter ranges, it is reasonable to expect a clean line of sight between the vehicles, and so a low-power point-to-point radio system is appropriate. The emerging IEEE 802.15.4 standard for low-power wireless networks [47] defines protocols for low-powered radios to form mesh networks and transfer small quantities of data (appropriate for sensor networks, automation, or other embedded applications) over the 2.4 GHz ISM frequency band. Some radios that conform to a subset of the IEEE 802.15.4 drafts are branded as *ZigBee* radios. The Digi XBee-PRO ZB is a ZigBee radio that allows short-range, line-of-

sight, low-latency data communications between vehicles. One feature of the XBee-PRO ZB that distinguishes it from similar modules is the on-board power amplifier, which increases the transmit power to 50 mW, extending the transmission range of the system. Connectivity at distances of up to 1 km was observed in river environments when using these modules.

5.6 Case Studies

The *Floating Sensor Network* team tested the various capabilities of the system with completed projects involving the US Army Corps of Engineers, the Office of Naval Research, and the California Bay-Delta Authority. The following sections describe how these case studies were used to test the functional requirements of the system detailed in Sections 4 and 5.

US Army Corps of Engineers

One application envisioned for the *Floating Sensor Network* is responding to levee failures. Using the real time web interface described previously, it would be easy to pinpoint the location of a levee breach and track the movement of water out of the system. It would also be useful for characterizing the change in salinity, temperature and other important parameters after the breach has occurred.

To test this capability, the team was invited to participate in the 2009 Rapid Repair of Levee Breaches Demonstration in Stillwater, Oklahoma. The test was operated by the Department of Homeland Security and the US Army Corps of Engineers. A test levee was built at the Federal Agriculture Department's Hydrologic Engineering Research Unit shown in Figure 5.28.

The channel behind the levee was filled and the levee was breached, releasing 125 cubic feet of water every second. The sensors were deployed upstream and allowed to pass through the breach:

Despite a brief loss of GPS signal while the vehicle was submerged in the breach, the experiment demonstrated the robustness of the system for use in real world environments and validated the design parameter dictating antenna height.

In addition to testing the robustness of the sensor design, this offered an opportunity to test the data assimilation capabilities of the backend system. This experiment and the Extended Kalman Filter analysis was first described in [3]. For the experiment, drifters were deployed into the supply canal upstream of the levee breach, shown in Figure 5.30. The upstream boundary condition was the supply canal flow control, set to 1.42 m^3/s. The downstream boundary condition was a gate that could be raised or lowered to restrict the flow out of the experimental region. Drifters were released at approximately 30 s intervals near the upstream boundary in Figure 22(a). After traveling through the canal for approximately 400 s, they were individually retrieved in Figure 22(b). Figure 22(c) marks the location of the downstream control gate.

Figure 5.28: Test levee under construction.

Figure 5.29: Floating sensor about to pass through the breach.

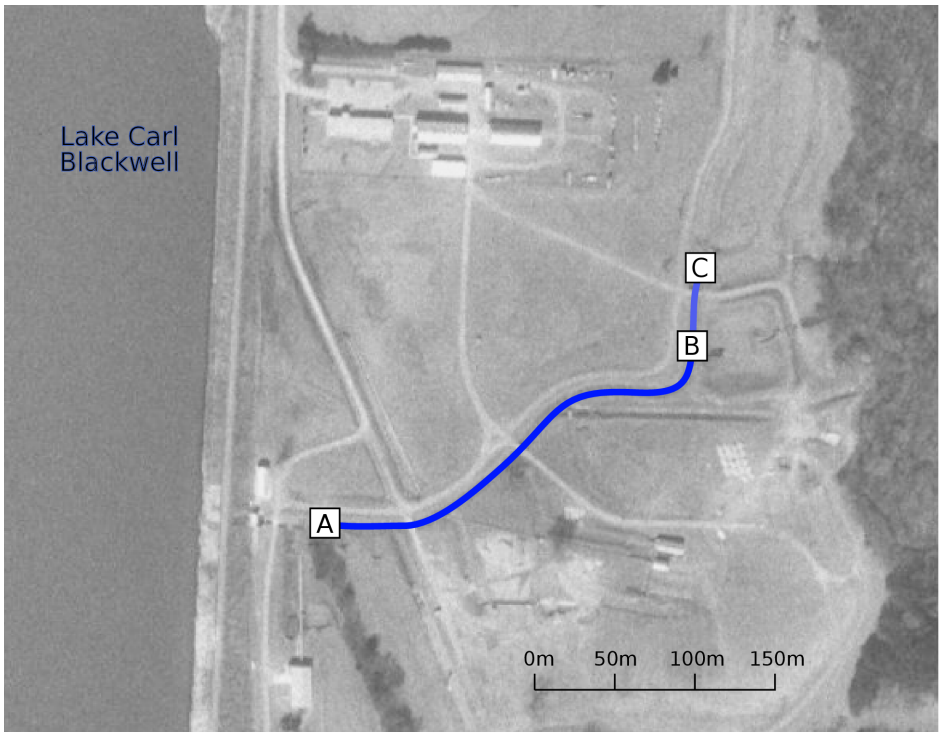

Figure 5.30: Annotated satellite image of supply canal.

A total of 20 runs were performed, and divided into five cycles of four runs each. Each run in the cycle had a different operation of the downstream control gate. During the first run, the gate remained open for the entire run. During the second run, the gate was closed as soon as the sixth drifter was released. During the third run, the gate remained closed. Finally, during the fourth run the gate was opened as soon as the final drifter was released. The cycle was then repeated.

The velocities of the sensors in one run are shown in Figure 5.31. The spikes in the velocity signal correspond to the sensor being thrown in the water. The velocity data from five sensors could then be assimilated with an Extended Kalman Filter (EKF), to estimate the flow in discretized 5 meter regions of the channel. Figure 5.32 shows the flow estimation in discretized "cells" using three techniques: forward simulation of the one-dimensional shallow water equation, and an EKF with and without estimating the bed slope.

The robustness of the sensor combined with the data assimilation system could be useful in the event of a levee failure in a region which supplies drinking water and is a fragile ecosystem for wildlife, such as the Sacramento San Joaquin Delta. On June 3, 2004, a levee breach occurred on the west levee of the Upper Jones Tract in the southern region of the Delta in San Joaquin County. Using the existing static infrastructure, it would be difficult to track changes near the breach. By contrast, the floating sensors could be immediately deployed by small boat teams in the affected region and would be able to upload data in

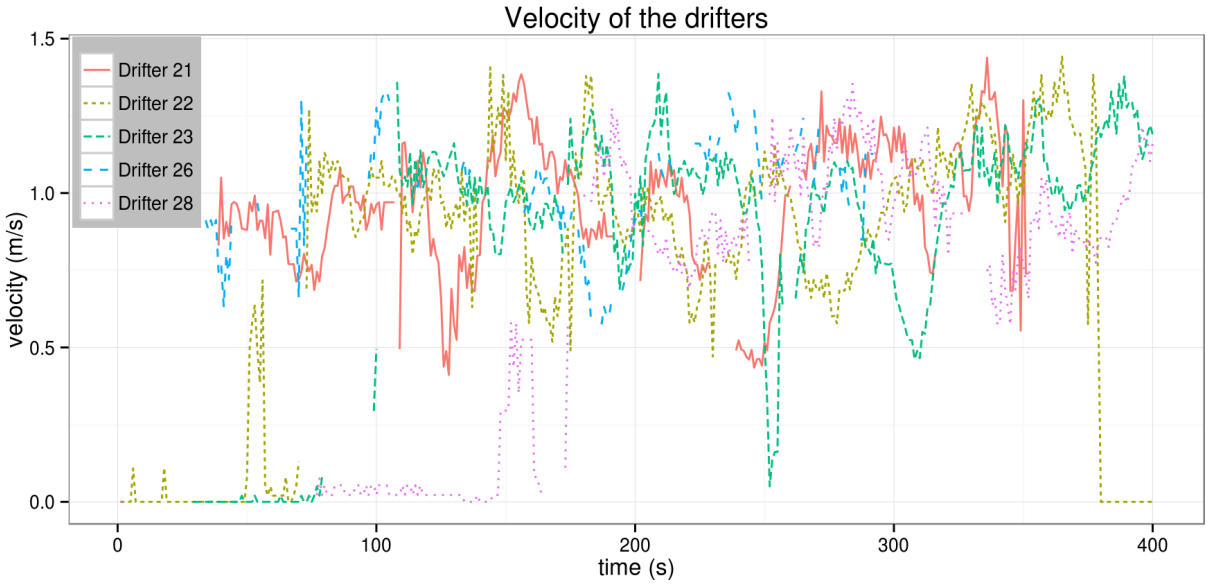

Figure 5.31: Measured sensor velocity during canal deployment. Source: [3]

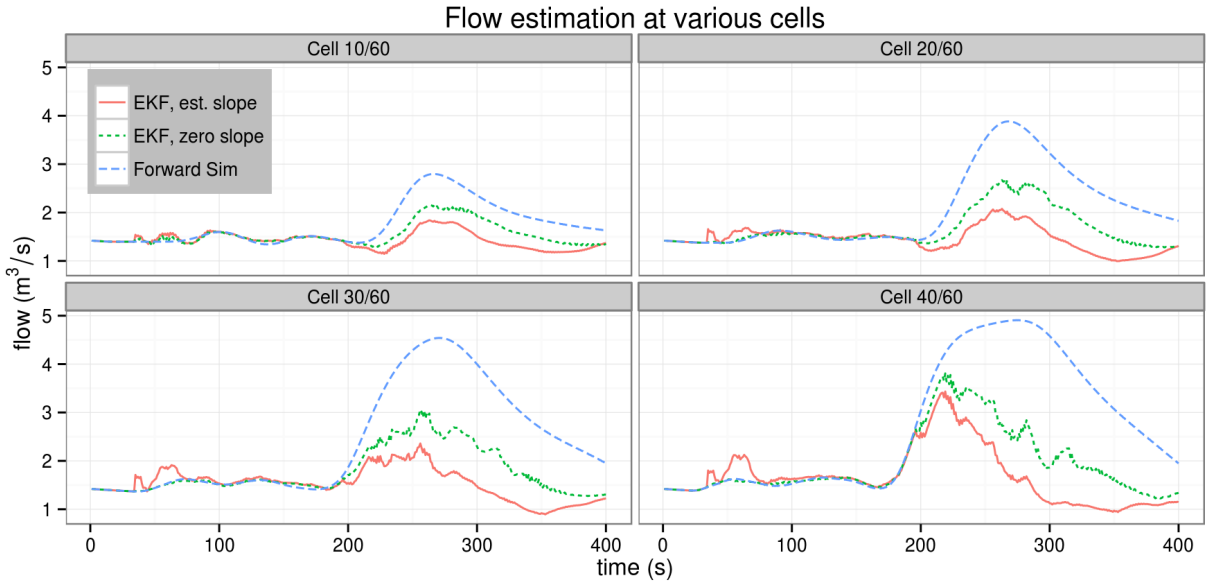

Figure 5.32: Estimated flow in discretized "cells" of the canal. Source: [3]

real time.

California Bay-Delta Authority

A number of experiments have been performed with the *Floating Sensor Network* to support research in the Sacramento San Joaquin Delta region performed on behalf of the California Bay-Delta Authority. The sensors have been used to track temperature changes near the Jones Tract failure and the implications for the delta smelt fish in the region. Other researchers have used the sensors to study the effect of submerged vegetation on water flow.

The sensors have also been used in the San Francisco Bay to study the exchange of water between the Bay and the rivers that feed into it. The sensors were deployed near the Mayfield Slough in the South San Francisco Bay. They were tracked over a full tidal cycle to determine if particles which get pulled into the Bay from the Slough during the ebb tide return during the flood tide, as shown in Figure 5.33. Unfortunately, high winds resulted in the data not matching the exact expectations of hydrodynamicists involved in the experiment, but the operation confirmed the ability of the system to operate unattended overnight and be retrieved. It also validated the design parameters governing battery life and mission time.

The experiment also provided confidence in the design parameters, indicating that multiple communication methods are necessary. During the middle of the night, the GSM signal unexpectedly failed (likely due to limited coverage in the center of the Bay). Without the capacity to communicate over the 802.15.4 radios, the vehicles would have been unrecoverable.

Office of Naval Research

The *Floating Sensor Network* was also explored as a platform for autonomous depth mapping of rivers for the Office of Naval Research. As part of phase one SIBR with SSCI, Inc., the team redesigned the sensors to be able to carry single beam depth finders.

SSCI developed a number of strategies in which the vehicles could leverage inter-vehicle communications to coordinate while operating in unstructured environments to upload depth maps in real time. Tests were done at Aquatic Park near UC Berkeley and the data was uploaded to Google Earth.

This application underscored the importance of design modularity. The sensor network was not designed with this application in mind but was able to accommodate the new application with limited changes to the system design thanks to the modular PVC sensor mounting plate and flexible electronics design. It also validated the computation design parameters such as the need to have a powerful processor for coordinated control tasks with a separate low level controller for real time control.

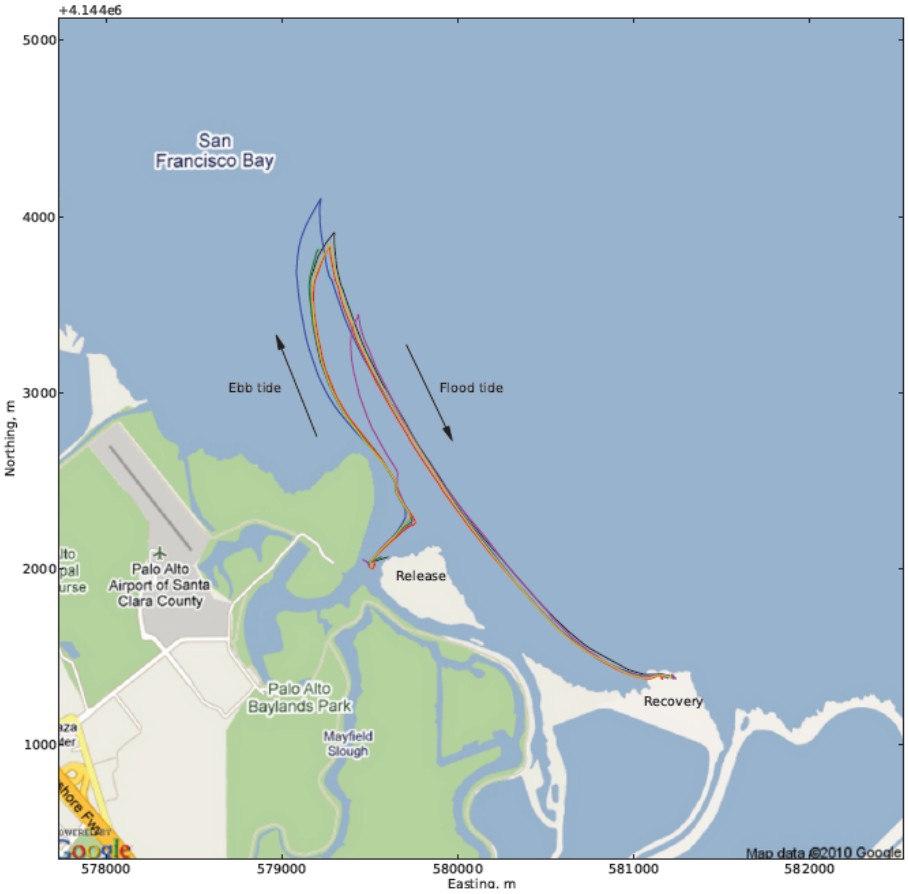

Figure 5.33: GPS traces from floating sensors during 24h San Francisco Bay experiment.

5.7 Conclusion

This chapter has described how the Ishikawa design methodology can be applied to developing a new network of robotic sensors for autonomous water management in near-shore environments. Specifically, it describes how the design parameters for the system evolved from a more general set of functional requirements which are interconnected through a cause and effect diagram. As a result of this design and prototyping process, a number of general conclusions are reached for the system.

First, when designing a sensing solution for multiple end users, it is important to have a modular sensor mounting system. The case studies confirmed that there are a number of applications for the sensor network ranging from bathymetry mapping in remote environments to salinity tracking for environmental monitoring purposes. If the system is to be broadly applicable, it must be able to accommodate a diverse set of applications with minimal changes to the core system design.

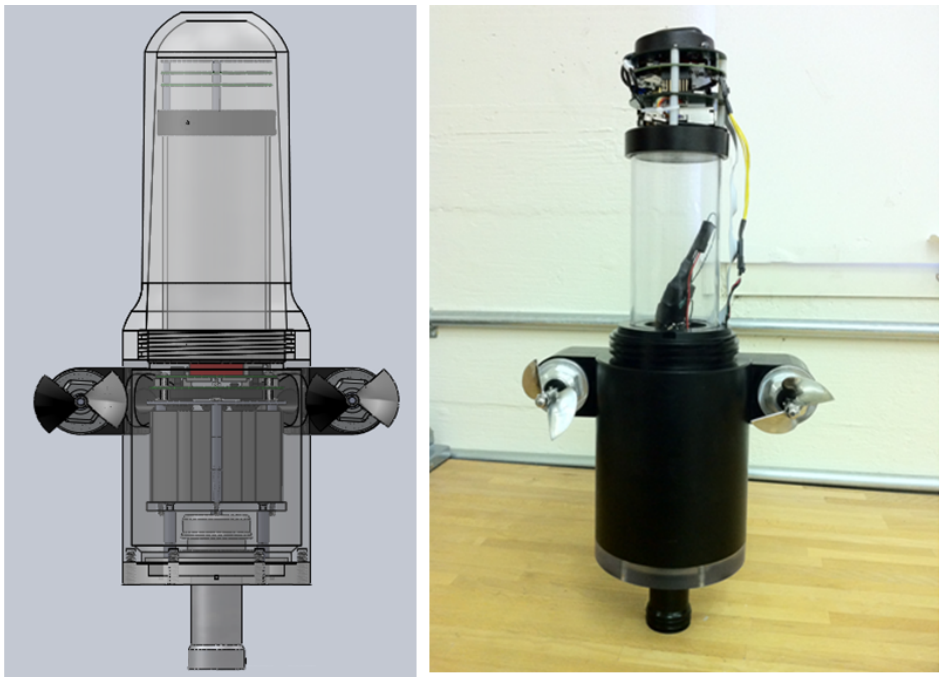

Figure 5.34: Floating sensor modified for single beam depth finder.

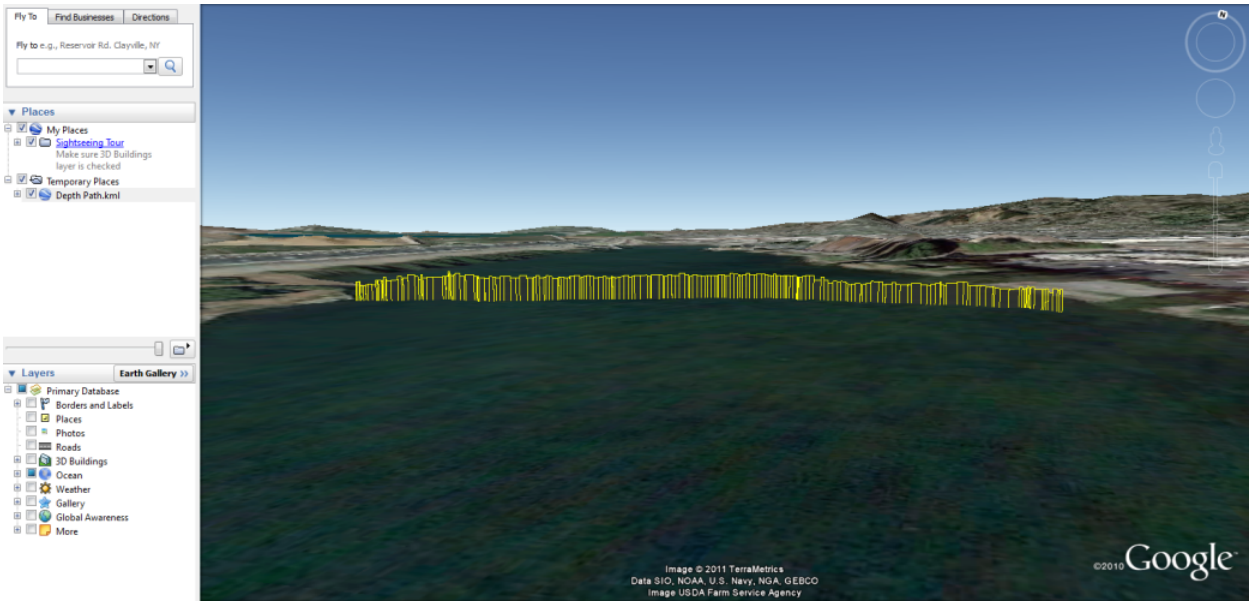

Figure 5.35: Automated depth survey from sensor path shown in Google Earth.

Second, the longer field tests have underscored the importance of redundant communication devices for retrieval. When GSM connectivity was lost during the 24-hour experiment in the San Francisco Bay, retrieval would have been impossible without having the ZigBee radios as a backup.

Third, the case studies have confirmed the need for actuation for Lagrangian sensors in near-shore environments. Field experiments which featured passive and active sensors deployed at the same time in the same location showed that passive sensors were only effective for 10 minutes before becoming entangled on shore whereas the active sensors operated for over an hour without becoming entangled.

Fourth, the 72-hour mission target seems to be sufficient for near-shore environments. Due to the small experimental domains of riverine networks, the sensors usually leave the experimental region in under 10 hours. The 24-hour test in the San Francisco Bay is the longest mission requested thus far.

Finally, while the vehicle form factor presents a reasonable tradeoff between man-portability, Lagrangian sensing, and active control capacity, further design studies could consider a more formal approach to optimizing vehicle size, battery life, and actuation mechanisms.

Chapter 6

Thesis Conclusions and Future Work

The prior chapters covered four topics related to the design and optimization of in-situ sensor networks for monitoring hydrologic phenomena in the Sierra Nevada Mountains and Sacramento-San Joaquin Delta. Chapter 2 evaluated a sensor placement strategy for wireless hydrologic observatories. Chapter 3 presented a repeater placement strategy for wireless mesh networks in complex terrain. Chapter 4 evaluated the long-term accuracy of estimating soil moisture from in-situ sensors, and Chapter 5 presented a design framework for an autonomous in-situ sensor network for the Sacramento San Joaquin Delta. The findings in each chapter raise a number of potential opportunities for future research, which are discussed in this concluding chapter.

The findings in Chapter 2 indicate that structuring catchment-scale observatories prior to field deployment using remotely-sensed data may be a feasible alternative to conducting field surveys. Further research could quantify the accuracy of the algorithm over longer time spans, particularly in forested environments. It would also be useful to investigate how increased information from these networks translates to value for downstream users. This would provide a more objective metric for the marginal value of information, which could be compared to the marginal cost of network establishment to determine the best locations for new catchment-scale observatories.

It may be possible to optimize the placement of the elements of the entire network (sensor and data repeater nodes) in tandem by combining the research reported in Chapters 2 and 3. The sensor locations in Chapter 2 were selected based on the nearest neighbor to the mean of each latent component in the Gaussian mixture model. Instead it may be possible to consider all the neighbors within a defined region near each latent component. The repeater placement algorithm could then be run using combinations of each of these potential placements, in order to find the distribution of nodes that produces a redundant mesh with the fewest number of repeater nodes.

The findings in Chapter 4 underscore the importance of in-situ measurements for monitoring soil moisture in montane regions, which feature topographic complexity, heterogenous soil and vegetation properties, and coupled snow-soil interactions. Future studies could evaluate these methods using more in-situ sensors deployed across greater elevation gradients in

the rain-snow transition. It would be informative to deploy sensors which capture gradients in variables not included in this study, particularly to investigate features that could better predict soil moisture in dry years. For instance, sampling over larger biomass gradients may better inform differences in water usage patterns by vegetation in wet and dry years. Future studies could also develop methodologies to synthesize the findings from this study into sensor placements and modeling strategies for basin-scale soil monitoring. Remote sensing, in-situ data, and deterministic modeling could then be combined to better estimate soil moisture storage in atypical water years.

It would be interesting to improve upon the drifter design presented in Chapter 5 in order to solve some of the limitations of the final sensor. While a mechanism for buoyancy control was present in the final design, there is no means to track the vehicle once it is submerged since the GPS loses signal. Further research is needed to determine how accurately the vehicles can be guided once they are submerged. One solution might be to employ the onboard IMU for underwater positioning. Another may be to design a GPS antenna that can be re-deployed to the surface. It would also be beneficial to design a sensor that addresses the mobility issues presented by the symmetric drag profile requirement for Lagrangian sensing. It may be possible to place a movable ballast that rotates the sensor into a horizontal position while the motors are engaged, but transitions to a vertical position while in use as a Lagrangian sensor.

Finally, it would be beneficial to consider how the dense, real-time measurements afforded by the systems described in this thesis can be combined with other data sources to meet future water challenges. Existing management policies often rely on analysis of historical data in order to plan and allocate water, implicitly assuming a stationary hydrologic system. This approach will likely have limited utility for future generations [65]. Due to rising temperatures, precipitation is more likely to fall as rain in the mountains, resulting in earlier and less controlled runoff. Significant high tides already put pressure on the levee networks, threatening below-sea level communities, agriculture, and freshwater availability. Sea-level rise will likely exacerbate these challenges. Managing water in a non-stationary environment presents a sizable multi-objective optimization problem: the needs of agriculture, hydropower, urban water users, and ecological communities must all be met simultaneously. An ideal approach for managing water in a non-stationary environment would integrate real-time measurements of the hydrologic system with remote sensing and distributed hydrologic models in order to better allocate water resources.

While an important first step in realizing this approach involves the design and optimization of systems that enable persistent real-time measurements in the environments that supply our water resources, future research should explore how best to connect these systems into an intelligent water grid. Sensor networks, computational resources, remote sensing tools are all progressing at a rapid pace, and we have only begun to realize the potential of combining these systems into an integrated water management infrastructure. Remote sensing data is increasingly available from sources such as the JPL Airborne Snow Observatory, which can be combined with in-situ measurements to better estimate large-scale hydrologic processes. As data on water usage in agriculture and urban areas is becoming increasingly

available, real-time measurements of both the supply and demand side are now within reach. Data from real-time sensor networks must be seamlessly integrated with large-scale hydrologic models and used for automated management and control of water usage along every step of the hydrologic cycle. Achieving this vision will require advances in large-scale optimization algorithms, in-situ sensor networks, standardization, and inter-operability. It will also require integrating research from beyond engineering, as water resources management is confounded by numerous social, political, and legislative factors. Information systems must be designed with these societal constraints in mind. Ultimately, it is insufficient to consider each of these systems in isolation, and intelligent water management will be facilitated by exchanging real-time knowledge throughout the water cycle.

Bibliography

[1] Floating Sensor Network website: `http://float.berkeley.edu`.

[2] *3GPP TS 23.060 version 10.4.0: "Digital cellular telecommunications system (Phase 2+); Universal Mobile Telecommunications System (UTMS); General Packet Radio Service (GPRS); Service description; Stage 2"*. European Telecommuniations Standards Institute. 2011.

[3] A. Bayen A. Tinka M. Rafiee. "Floating Sensor Networks for River Studies". In: *IEEE Systems Journal* (2012).

[4] Ian F Akyildiz et al. "Wireless sensor networks: a survey". In: *Computer networks* 38.4 (2002), pp. 393–422.

[5] S.P. Anderson, G. Qinghua, and E.G. Parrish. *Snow-on and snow-off Lidar point cloud data and digital elevation models for study of topography, snow, ecosystems and environmental change at Southern Sierra Critical Zone Observatory, California: Southern Sierra CZO, University of California at Merced, (digital media)*. 2012.

[6] *AVR XMEGA A3 Device Datasheet*. 8068S. Atmel Corporation. San Jose, California, Sept. 2010.

[7] Roger C. Bales et al. "Mountain hydrology of the western United States". en. In: *Water Resources Research* 42.8 (Aug. 2006), W08432. ISSN: 1944-7973. DOI: `10.1029/2005WR004387`. URL: `http://onlinelibrary.wiley.com/doi/10.1029/2005WR004387/abstract` (visited on 11/25/2014).

[8] Roger C Bales et al. "Soil moisture response to snowmelt and rainfall in a Sierra Nevada mixed-conifer forest". In: *Vadose Zone Journal* 10.3 (2011), pp. 786–799.

[9] Benjamin Balk and Kelly Elder. "Combining binary decision tree and geostatistical methods to estimate snow distribution in a mountain watershed". In: *Water Resources Research* 36.1 (2000), pp. 13–26.

[10] Lawrence E Band. "Effect of land surface representation on forest water and carbon budgets". In: *Journal of Hydrology* 150.2 (1993), pp. 749–772.

[11] Tim P Barnett, Jennifer C Adam, and Dennis P Lettenmaier. "Potential impacts of a warming climate on water availability in snow-dominated regions". In: *Nature* 438.7066 (2005), pp. 303–309.

[12] KJ Beven and MJ Kirkby. "A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant". In: *Hydrological Sciences Journal* 24.1 (1979), pp. 43–69.

[13] P. Bhatta et al. "Coordination of an underwater glider fleet for adaptive ocean sampling". In: *Proc. International Workshop on Underwater Robotics*. 2005.

[14] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[15] Jack E Bresenham. "Algorithm for computer control of a digital plotter". In: *IBM Systems journal* 4.1 (1965), pp. 25–30.

[16] Kaye Brubaker, Albert Rango, and William Kustas. "Incorporating radiation inputs into the snowmelt runoff model". In: *Hydrological Processes* 10.10 (1996), pp. 1329–1343.

[17] David D. Clark. "Overview Of the Argos System". In: *OCEANS '89. Proceedings*. Vol. 3. Sept. 1989, pp. 934–939. DOI: 10.1109/OCEANS.1989.586711.

[18] RA Dahlgren et al. "Soil development along an elevational transect in the western Sierra Nevada, California". In: *Geoderma* 78.3 (1997), pp. 207–236.

[19] Russ E. Davis. "Drifter observations of coastal surface currents during CODE: The method and descriptive view". In: *Journal of Geophysical Research* 90.C3 (1985), pp. 4741–4755. DOI: 10.1029/JC090iC03p04741.

[20] Jeffrey S Deems, Thomas H Painter, and David C Finnegan. "Lidar measurement of snow depth: a review". In: *Journal of Glaciology* 59.215 (2013), pp. 467–479. DOI: 10.3189/2013JoG12J154.

[21] Carrick Detweiller, Iuliu Vasilescu, and Daniela Rus. "An underwater sensor network with dual communications, sensing, and mobility". In: *OCEANS 2007-Europe*. IEEE. 2007, pp. 1–6. DOI: 10.1109/OCEANSE.2007.4302445.

[22] Jeff Dozier. "Spectral signature of alpine snow cover from the Landsat Thematic Mapper". In: *Remote Sensing of Environment* 28 (1989), pp. 9–22.

[23] Thomas Dunne and Richard D Black. "An experimental investigation of runoff production in permeable soils". In: *Water Resources Research* 6.2 (1970), pp. 478–490.

[24] Luca Egli et al. "Dynamics of snow ablation in a small Alpine catchment observed by repeated terrestrial laser scans". In: *Hydrological Processes* 26.10 (2012), pp. 1574–1585.

[25] Tyler A Erickson, Mark W Williams, and Adam Winstral. "Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States". In: *Water Resources Research* 41.4 (2005).

[26] Jennifer Erxleben, Kelly Elder, and Robert Davis. "Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains". In: *Hydrological Processes* 16.18 (2002), pp. 3627–3649.

[27] Richard Essery and John Pomeroy. "Implications of spatial distributions of snow mass and melt rate for snow-cover depletion: theoretical considerations". In: *Annals of Glaciology* 38.1 (2004), pp. 261–265.

[28] DA Faria, JW Pomeroy, and RLH Essery. "Effect of covariance between ablation and snow water equivalent on depletion of snow-covered area in a forest". In: *Hydrological Processes* 14 (2000), pp. 2683–2695.

[29] S. R. Fassnacht, K. A. Dressler, and R. C. Bales. "Snow water equivalent interpolation for the Colorado River Basin from snow telemetry (SNOTEL) data". In: *Water Resources Research* 39.8 (2003). 1208, ISSN: 1944-7973. DOI: 10.1029/2002WR001512. URL: http://dx.doi.org/10.1029/2002WR001512.

[30] Mark G Flanner et al. "Springtime warming and reduced snow cover from carbonaceous particles". In: *Atmospheric Chemistry and Physics* 9.7 (2009), pp. 2481–2497.

[31] D. Cayan D. Checkley R. Keeling L. Levin M. Ohman A. Gershunov D. Pierce R. Somerville D. Victor G. Franco and S. Moser. *Scripps Researchers Assess the Future of Climate in California.*

[32] Yoav Freund and Robert E Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting". In: *Journal of computer and system sciences* 55.1 (1997), pp. 119–139.

[33] *Motorola G24 Developers' Guide: Module Hardware Description.* Motorola. 2007.

[34] Pall Oskar Gislason, Jon Atli Benediktsson, and Johannes R Sveinsson. "Random forests for land cover classification". In: *Pattern Recognition Letters* 27.4 (2006), pp. 294–300.

[35] W. John Gould. "From Swallow floats to Argo–the development of neutrally buoyant floats". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 52.3–4 (2005), pp. 529–543.

[36] Michael L. Goulden and Roger C. Bales. "Mountain runoff vulnerability to increased evapotranspiration with vegetation expansion". en. In: *Proceedings of the National Academy of Sciences* 111.39 (Sept. 2014), pp. 14071–14075. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1319316111. URL: http://www.pnas.org/content/111/39/14071 (visited on 12/22/2014).

[37] T. Grünewald et al. "Statistical modelling of the snow depth distribution in open alpine terrain". In: *Hydrology and Earth System Sciences* 17.8 (2013), pp. 3005–3021. DOI: 10.5194/hess-17-3005-2013. URL: http://www.hydrol-earth-syst-sci.net/17/3005/2013/.

[38] Bin Guan et al. "Snow water equivalent in the Sierra Nevada: Blending snow sensor observations with snowmelt model simulations". en. In: *Water Resources Research* 49.8 (Aug. 2013), pp. 5029–5046. ISSN: 1944-7973. DOI: 10.1002/wrcr.20387. URL: http://onlinelibrary.wiley.com/doi/10.1002/wrcr.20387/abstract (visited on 11/25/2014).

[39] *Gumstix® robostix*. Gumstix, Inc. URL: http://www.gumstix.com/store/product_info.php?products_id=139.

[40] Qinghua Guo et al. "Effects of topographic variability and lidar sampling density on several DEM interpolation methods". In: *Photogrammetric Engineering & Remote Sensing* 76.6 (2010), pp. 701–712.

[41] Y. Han et al. "Dynamic Modeling and Pneumatic Switching Control of a Submersible Drogue". In: *7th Int. Conference on Informatics in Control, Automation and Robotics*. 2010, pp. 89–97.

[42] Robert F Harrington, Kelly Elder, and Roger C Bales. "Distributed snowmelt modeling using a clustering algorithm". In: (1995).

[43] Brian J Harshburger et al. "Spatial interpolation of snow water equivalency using surface observations and remotely sensed images of snow-covered area". In: *Hydrological Processes* 24.10 (2010), pp. 1285–1295.

[44] Kay Helfricht et al. "Interannual persistence of the seasonal snow cover in a glacierized catchment". In: *Journal of Glaciology* 60.223 (2014), pp. 889–904. DOI: 10.3189/2014JoG13J197. URL: http://www.ingentaconnect.com/content/igsoc/jog/2014/00000060/00000223/art00007.

[45] *HHR110AAO datasheet*. Panasonic. Jan. 2004.

[46] Paul R Houser et al. "Integration of soil moisture remote sensing and hydrologic modeling using data assimilation". In: *Water Resources Research* 34.12 (1998).

[47] *IEEE P802.15.4e/D0.01 Draft Standard for Information technology: Telecommunications and information exchange between systems: Local and metropolitan area networks: Specific requirements: Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs) Amendment 1: Add MAC enhancements for industrial applications and CWPAN*. IEEE P802.15 Working Group for Wireless Personal Area Networks (WPANs), Sept. 2009.

[48] Kaoru Ishikawa. "Cause and Effect Diagram". In: *Proceedings of the International Conference on Quality Control*. Tokyo, Oct. 1969, pp. 607–610.

[49] M Jung, M Reichstein, and A Bondeau. "Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model". In: *Biogeosciences* 6.10 (2009), pp. 2001–2013.

[50] Branko Kerkez et al. "Design and performance of a wireless sensor network for catchment-scale snow and soil moisture measurements". en. In: *Water Resources Research* 48.9 (Sept. 2012), W09515. ISSN: 1944-7973. DOI: 10.1029/2011WR011214. URL: http://onlinelibrary.wiley.com/doi/10.1029/2011WR011214/abstract (visited on 11/25/2014).

[51] B Kerkez et al. "Design and development of a wireless sensor network to monitor snow depth in multiple catchments in the American River Basin, California: Hardware selection and sensor placement techniques". In: *AGU Fall Meeting Abstracts*. Vol. 1. 2010, p. 07.

[52] Abedalrazq Khalil et al. "Applicability of statistical learning algorithms in groundwater quality modeling". In: *Water Resources Research* 41.5 (2005). W05010. ISSN: 1944-7973. DOI: 10.1029/2004WR003608. URL: http://dx.doi.org/10.1029/2004WR003608.

[53] Abedalrazq Khalil et al. "Sparse Bayesian learning machine for real-time management of reservoir releases". In: *Water Resources Research* 41.11 (2005). W11401. ISSN: 1944-7973. DOI: 10.1029/2004WR003891. URL: http://dx.doi.org/10.1029/2004WR003891.

[54] Nick J Killingsworth and Miroslav Krstic. "PID tuning using extremum seeking: online, model-free performance optimization". In: *IEEE control systems* 26.1 (2006), pp. 70–79.

[55] L Kou, George Markowsky, and Leonard Berman. "A fast algorithm for Steiner trees". In: *Acta informatica* 15.2 (1981), pp. 141–145.

[56] Andreas Krause, Ajit Singh, and Carlos Guestrin. "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies". In: *Journal of Machine Learning Research* 9.Feb (2008), pp. 235–284.

[57] M. Lauer. "Building Linux distributions with BitBake and OpenEmbedded". In: *Proceedings of the Free and Open Source Developers' European Meeting (FOSDEM 2005)*. Brussels, Belgium, 2005.

[58] *LC-P0612P datasheet*. Panasonic. June 2000.

[59] Michael Lehning, Thomas Grünewald, and Michael Schirmer. "Mountain snow distribution governed by an altitudinal gradient and terrain roughness". In: *Geophysical Research Letters* 38.19 (2011). L19504. ISSN: 1944-8007. DOI: 10.1029/2011GL048927. URL: http://dx.doi.org/10.1029/2011GL048927.

[60] Timothy Link and Danny Marks. "Distributed simulation of snowcover mass- and energy-balance in the boreal forest". In: *Hydrological Processes* 13.14 (1999), pp. 2439–2452. DOI: 10.1002/(SICI)1099-1085(199910)13:14/15.

[61] S. Londhe and S. Charhate. "Comparison of data-driven modelling techniques for river flow forecasting". In: *Hydrological Science Journal* 55.7 (2010), pp. 1163–1174. DOI: 10.1080/02626667.2010.512867.

[62] Danny Marks, Jeff Dozier, and Robert E Davis. "Climate and energy exchange at the snow surface in the alpine region of the Sierra Nevada: 1. Meteorological measurements and monitoring". In: *Water Resources Research* 28.11 (1992), pp. 3029–3042.

[63] JL McCreight et al. "Inference and uncertainty of snow depth spatial distribution at the kilometre scale in the Colorado Rocky Mountains: the effects of sample size, random sampling, predictor quality, and validation procedures". In: *Hydrological Processes* 28.3 (2014), pp. 933–957. DOI: 10.1002/hyp.9618.

[64] Geoffrey McLachlan and David Peel. *Finite mixture models*. John Wiley & Sons, 2004.

[65] P. C. D. Milly et al. "Stationarity Is Dead: Whither Water Management?" In: *Science* 319.5863 (2008), pp. 573–574. ISSN: 0036-8075. DOI: 10.1126/science.1151915. eprint: http://science.sciencemag.org/content/319/5863/573.full.pdf. URL: http://science.sciencemag.org/content/319/5863/573.

[66] Satyajayant Misra et al. "Constrained relay node placement in wireless sensor networks: Formulation and approximations". In: *IEEE/ACM Transactions on Networking (TON)* 18.2 (2010), pp. 434–447.

[67] Noah P Molotch and Roger C Bales. "Scaling snow observations from the point to the grid element: Implications for observation network design". In: *Water Resources Research* 41.11 (2005).

[68] Noah P Molotch and Roger C Bales. "SNOTEL representativeness in the Rio Grande headwaters on the basis of physiographics and remotely sensed snow cover persistence". In: *Hydrological Processes* 20.4 (2006), pp. 723–739.

[69] Noah P Molotch et al. "Ecohydrological controls on snowmelt partitioning in mixed-conifer sub-alpine forests". In: *Ecohydrology* 2.2 (2009), pp. 129–142.

[70] Ian D Moore, T Williams Norton, and Jann E Williams. "Modelling environmental heterogeneity in forested landscapes". In: *Journal of Hydrology* 150.2 (1993), pp. 717–747.

[71] ID Moore, GJ Burch, and DH Mackenzie. "Topographic effects on the distribution of surface soil water and the location of ephemeral gullies". In: *Transactions of the ASAE (USA)* (1988).

[72] Philip W Mote et al. "Declining mountain snowpack in western North America". In: *Bulletin of the American meteorological Society* 86.1 (2005), pp. 39–49.

[73] Jeffrey Mount and Robert Twiss. "Subsidence, sea level rise, and seismicity in the Sacramento-San Joaquin Delta". In: *San Francisco Estuary and Watershed Science* 3.1 (2005).

[74] KN Musselman, NP Molotch, and PD Brooks. "Effects of vegetation on snow accumulation and ablation in a mid-latitude sub-alpine forest". In: *Hydrological Processes* 22.15 (2008), pp. 2767–2776.

[75] *NCR18650 datasheet*. Panasonic. Feb. 2010.

[76] Pearn P. Niiler, Russ E. Davis, and Henry J. White. "Water-following characteristics of a mixed layer drifter". In: *Deep-Sea Research* Vol. 34, No. 11 (1987), pp. 1867–1881.

[77] Eni G Njoku et al. "Soil moisture retrieval from AMSR-E". In: *Geoscience and Remote Sensing, IEEE Transactions on* 41.2 (2003), pp. 215–229.

[78] Taikan Oki and Shinjiro Kanae. "Global Hydrological Cycles and World Water Resources". In: *Science* 313 (2006), pp. 1068–1072.

[79] Thomas H Painter et al. "Retrieval of subpixel snow-covered area and grain size from imaging spectrometer data". In: *Remote Sensing of Environment* 85.1 (2003), pp. 64–77.

[80] Thomas H Painter et al. "The Airborne Snow Observatory: Fusion of scanning Lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo". In: *Remote Sensing of Environment* 184 (2016), pp. 139–152.

[81] Anja Palli et al. "Spatial and temporal variability of snow accumulation using ground-penetrating radar and ice cores on a Svalbard glacier". In: *Journal of Glaciology* 48.162 (2002), pp. 417–424.

[82] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[83] Tom R Perkins, Thomas C Pagano, and David C Garen. "Innovative operational seasonal water supply forecasting technologies". In: *Journal of Soil and Water Conservation* 64.1 (2009), 15A–17A.

[84] A Rango and J Martinec. "Revisiting the Degree-Day Method for Snowmelt Computations". In: *JAWRA Journal of the American Water Resources Association* 31.4 (1995), pp. 657–669.

[85] C. E. Rasmussen and C. K. I. Williams. *Gaussian processes for machine learning*. MIT press, 2006.

[86] K. Rasouli, W. W. Hseih, and A. J. Cannon. "Daily streamflow forecasting by machine learning methods with weather and climate inputs". In: *Journal of Hydrology* 414-415 (2012), pp. 284–293. DOI: 10.1016/j.jhydrol.2011.10.039.

[87] Robert Rice and Roger C Bales. "Embedded-sensor network design for snow cover measurements around snow pillow and snow course sites in the Sierra Nevada of California". In: *Water resources research* 46.3 (2010).

[88] Eric A Rosenberg, Andrew W Wood, and Anne C Steinemann. "Statistical applications of physically based hydrologic models to seasonal streamflow forecasts". In: *Water Resources Research* 47.3 (2011).

[89] Walter Rosenthal and Jeff Dozier. "Automated mapping of montane snow cover at subpixel resolution from the Landsat Thematic Mapper". In: *Water Resources Research* 32.1 (1996), pp. 115–130.

[90] J. Ryan, N. Aonghusa, and E. Sweeney. "SmartBay, Ireland: Design and Planning for a Cabled Ocean Observatory off the West coast of Ireland". In: *OCEANS 2007*. IEEE. 2008, pp. 1–4. DOI: 10.1109/OCEANS.2007.4449225.

[91] B. Settles. *Active learning*. Morgan & Claypool, 2012.

[92] Hong Shen and Longkun Guo. "Efficient 2-approximation algorithms for computing 2-connected steiner minimal networks". In: *IEEE Transactions on Computers* 61.7 (2012), pp. 954–968.

[93] Dimitri P. Solomatine and Durga Lal Shrestha. "A novel method to estimate model uncertainty using machine learning techniques". In: *Water Resources Research* 45.12 (2009). W00B11. ISSN: 1944-7973. DOI: 10.1029/2008WR006839. URL: http://dx.doi.org/10.1029/2008WR006839.

[94] I. Strub et al. "Comparison of two data assimilation algorithms for shallow water flows, I". In: *Networks and Heterogeneous Media* 4 (June 2009), pp. 409–430. DOI: 10.3934/nhm.2009.4.409.

[95] P. B. Sujit, Joao Sousa, and Fernando L. Pereira. "UAV and AUVs coordination for ocean exploration". In: *OCEANS 2009-EUROPE*. IEEE. 2009, pp. 1–7. DOI: 10.1109/OCEANSE.2009.5278262.

[96] John C. Swallow. "A neutral-buoyancy float for measuring deep currents". In: *Deep Sea Research (1953)* 3.1 (1955), pp. 74–81. DOI: doi:10.1016/0146-6313(55)90037-X.

[97] Tyson L. Swetnam and Donald A. Falk. "Application of Metabolic Scaling Theory to reduce error in local maxima tree segmentation from aerial LiDAR". In: *Forest Ecology and Management* 323 (2014), pp. 158–167. ISSN: 0378-1127. DOI: http://dx.doi.org/10.1016/j.foreco.2014.03.016. URL: http://www.sciencedirect.com/science/article/pii/S0378112714001704.

[98] Andres M. Ticlavilca and Mac McKee. "Multivariate Bayesian Regression Approach to Forecast Releases from a System of Multiple Reservoirs". In: *Water Resources Management* 25.2 (2010), pp. 523–543. ISSN: 1573-1650. DOI: 10.1007/s11269-010-9712-y. URL: http://dx.doi.org/10.1007/s11269-010-9712-y.

[99] Alfonso F. Torres, Wynn R. Walker, and Mac McKee. "Forecasting daily potential evapotranspiration using machine learning and limited climatic data". In: *Agricultural Water Management* 98.4 (2011), pp. 553–562. ISSN: 0378-3774. DOI: 10.1016/j.agwat.2010.10.012. URL: http://www.sciencedirect.com/science/article/pii/S0378377410003331.

[100] O.-P. Tossavainen et al. "State estimation and modeling error approach for 2D shallow water equations and Lagrangian measurements". In: *Water Resources Research* (2011). DOI: 10.1029/2010WR009401.

[101] E. Trujillo and M. Lehning. "Theoretical analysis of errors when estimating snow distribution through point measurements". In: *The Cryosphere* 9.3 (2015), pp. 1249–1264. DOI: 10.5194/tc-9-1249-2015. URL: http://www.the-cryosphere.net/9/1249/2015/.

[102] *UPF673791 datasheet*. Sanyo. Mar. 2011.

[103] Wolfgang Wagner et al. "Soil moisture from operational meteorological satellites". In: *Hydrogeology Journal* 15.1 (2007), pp. 121–131.

[104] Thomas Watteyne et al. "Mitigating multipath fading through channel hopping in wireless sensor networks". In: *Communications (ICC), 2010 IEEE International Conference on*. IEEE. 2010, pp. 1–5.

[105] Andrew W Western et al. "Observed spatial organization of soil moisture and its relation to terrain indices". In: *Water resources research* 35.3 (1999), pp. 797–810.

[106] CJ Williams, JP McNamara, and DG Chandler. "Controls on the temporal and spatial variability of soil moisture in a mountainous landscape: the signature of snow and complex terrain". In: *Hydrology and Earth System Sciences* 13.7 (2009), pp. 1325–1336.

[107] J Wilson and P Wright. "Design of monocular head-mounted displays, with a case study on fire-fighting". In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 221.12 (2007), pp. 1729–1743. DOI: 10.1243/09544062JMES721.

[108] Adam Winstral and Danny Marks. "Simulating wind fields and snow redistribution using terrain-based parameters to model snow accumulation and melt over a semi-arid mountain catchment". In: *Hydrological Processes* 16.18 (2002), pp. 3585–3603.

[109] H.T. Wonn and K.L. O'Hara. "Height:Diameter Ratios and Stability Relationships for Four Northern Rocky Mountain Tree Species". In: *Western Journal of Applied Forestry* 16.2 (2001), pp. 87–94.

[110] Dan Zaslavsky and Gideon Sinai. "Surface hydrology: I—explanation of phenomena". In: *Journal of the Hydraulics Division* 107.1 (1981), pp. 1–16.

[111] M. M. Zdravkovich et al. "Flow past short circular cylinders with two free ends". In: *Journal of Fluid Mechanics* 203.1 (1989), pp. 557–575. DOI: 10.1017/S002211208900159X.

[112] Z. Zheng, P. B. Kirchner, and R. C. Bales. "Topographic and vegetation effects on snow accumulation in the southern Sierra Nevada: a statistical summary from lidar data". In: *The Cryosphere* 10.1 (2016), pp. 257–269. DOI: 10.5194/tc-10-257-2016. URL: http://www.the-cryosphere.net/10/257/2016/.