



# AN ABSTRACT OF THE DISSERTATION OF

Gregory W. Wilson for the degree of Doctor of Philosophy in Oceanography presented  
on April 15, 2013.

Title: Data Assimilation for Prediction of Shallow Water Flows with Uncertain  
Bathymetry

Abstract approved: \_\_\_\_\_

H. Tuba Özkan-Haller

A new method is introduced for incorporating bathymetric uncertainty into predictions of nearshore and river flows (i.e., unstratified flows primarily forced by pressure and radiation stress gradients). The method involves the use of the ensemble Kalman filter (EnKF) as a parameter estimation scheme, where the parameter to be estimated is the spatial field of bathymetry. That is, bathymetry is treated as a slowly varying uncertain parameter in the model, which can be corrected via the assimilation of other available observations. The reason bathymetry is targeted is, as we show, it is often a limiting factor for accuracy in real-world modeling applications.

Results are shown using data from four field experiments. Two experiments involve measurements in the nearshore (surf zone) ocean at Duck, NC. There, we show that bathymetric uncertainty due to rapid bathymetric change (time scale of days), or simply lack of available measurements, can cause significant error in model predictions of waves and currents. We demonstrate the ability of the EnKF to reduce this error by correcting the bathymetry, which we then cross-validate using in-situ measurements. Specifically, the correction is achieved by assimilating in-situ observations of alongshore current and significant wave height, as well as (in a separate experiment) remote-sensing observations of alongshore current, wave celerity, and location of shoreline. Similarly in a river environment (Snohomish River, WA, and Kootenai River, ID), we demonstrate the EnKF using twin tests, assimilating pseudo-observations of currents from a variety of hypothesized platforms (fixed in-situ gages, passive drifters, and Doppler radar). Again, the EnKF is found to yield accurate estimates of bathymetry.

©Copyright by Gregory W. Wilson  
April 15, 2013  
All Rights Reserved

Data Assimilation for Prediction of Shallow Water Flows with  
Uncertain Bathymetry

by

Gregory W. Wilson

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of  
the requirements for the  
degree of

Doctor of Philosophy

Presented April 15, 2013  
Commencement June 2013



Doctor of Philosophy dissertation of Gregory W. Wilson presented on April 15, 2013.

APPROVED:

---

Major Professor, representing Oceanography

---

Dean of the College of Earth, Ocean, and Atmospheric Sciences

---

Dean of the Graduate School

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

---

Gregory W. Wilson, Author

## ACKNOWLEDGEMENTS

An early mentor once warned me that completing a PhD takes more than just aptitude, it requires an unwavering commitment to ones chosen scientific topic. On reflection, I would add (and I'm sure he would agree) that it takes the support of family, friends, and colleagues. Thankfully, I have had that in abundance. Working in the Coastal Imaging Lab has been a joy, and has given me lifelong friendships of which there are too many to list. As for my family (genetic and otherwise) back home, you have helped more than you know in the work shown here; you have been a rock where I can rest, and a place to hold my best values.

I also have had the honour of being surrounded by many role models in academia, who influenced the work in this dissertation as well as my philosophy in science and life. Boualem Khouider, Chris Garrett, Eric Kunze, and David Goodenough all took the time to put up with me in my formative years as a senior undergraduate, and gave me the foundations for surviving the commitment of a PhD. More recently, Tuba Özkan-Haller and Rob Holman have provided me with every opportunity and freedom to pursue new ideas, and find my own path — I didn't always agree with that approach (and probably neither did they) but in the end it has been invaluable. Robert Miller, my "sixth committee member", has given his time and insight on many occasions; his guidance was the genesis for much of the work on data assimilation presented here. And Alex Kurapov helped cement my understanding of data assimilation, assuring me that not only is it "real" science, but "science is whatever you are most interested in". To badly paraphrase Poincaré, all of the above people gave me my blueprint and my stones — I have simply stacked them up to make my own house. I can only hope they will see some of themselves in my work.

Finally, many people contributed observational data, computer code, and technical assistance which supported the work presented here. It would take pages to list each individual contribution, but I wish to sincerely thank all of the following people: Chris Chickadel, Derek Fong, Gabriel Garcia, Sarah Giddings, Mick Haller, Kent Hathaway, David Honneger, Adam Keen, Pavel Sakov, John Stanley, Jim Thompson, and John Warner.

Funding for the work presented here was provided by the Office of Naval Research, grants N00014-02-1-0198, N00014-07-1-0852, N00014-09-1-0121, and N00014-10-1-0932.

## CONTRIBUTION OF AUTHORS

The manuscripts comprising Chapters Two and Four of this dissertation include Robert A. Holman as a co-author. Dr. Holman's contribution included advice on the design and writing of both manuscripts, as well as remote sensing data collection and data analysis in Chapter Four.

# TABLE OF CONTENTS

	<u>Page</u>
1 General Introduction	1
2 Data Assimilation and Bathymetric Inversion in a 2DH Surf Zone Model	5
2.1 Introduction . . . . .	6
2.2 Observations . . . . .	8
2.3 Methodology . . . . .	9
2.3.1 Theory for Bathymetric Inversion . . . . .	10
2.3.2 Prior Bathymetric Ensemble . . . . .	12
2.3.3 Forward Model . . . . .	13
2.3.4 Observational Error Estimates . . . . .	15
2.3.5 Underlying Assumptions and Optimality . . . . .	16
2.4 Results . . . . .	17
2.4.1 Physical Setting . . . . .	18
2.4.2 Conceptual Interpretation: 1DH Case Study . . . . .	19
2.4.3 Assimilation During 2DH Flow . . . . .	25
2.4.4 Skill Statistics . . . . .	29
2.5 Discussion . . . . .	31
2.5.1 Assimilation Skill . . . . .	31
2.5.2 Bathymetric Input Sensitivity . . . . .	33
2.5.3 Effect of Sampling Scheme . . . . .	33
2.5.4 Assimilation of Other Observational Data Types . . . . .	34
2.6 Conclusions . . . . .	36
3 Ensemble-Based Data Assimilation for Estimation of River Depths	39
3.1 Introduction . . . . .	40
3.2 Methods . . . . .	41
3.2.1 Bathymetric Inversion Method . . . . .	41
3.2.2 Assumed Known Information . . . . .	42
3.2.3 Specification of Prior . . . . .	42
3.3 Verification . . . . .	44
3.3.1 Idealized Test Case: Straight-Channel With a Bump . . . . .	44
3.3.2 Test Cases With Realistic Bathymetry . . . . .	49
3.4 Discussion: Applicability to Real Observations . . . . .	55
3.5 Summary . . . . .	58

## TABLE OF CONTENTS (Continued)

	<u>Page</u>
4 Surf Zone Bathymetry and Circulation Predictions via Data Assimilation of Remote Sensing Observations	60
4.1 Introduction . . . . .	60
4.2 Observations . . . . .	63
4.2.1 Experiment . . . . .	63
4.2.2 In-Situ Data . . . . .	63
4.2.3 Optical Remote Sensing . . . . .	65
4.2.4 Infrared Remote Sensing: Particle Image Velocimetry . . . . .	70
4.2.5 Radar Remote Sensing: Wave Celerity and Direction . . . . .	72
4.3 Modeling and Bathymetry Inversion System . . . . .	73
4.3.1 Forward Model . . . . .	74
4.3.2 Mathematical Statement of Inverse Problem . . . . .	76
4.3.3 Covariance Modeling Using Ensembles . . . . .	77
4.3.4 Update Step . . . . .	78
4.3.5 Observation Operator . . . . .	79
4.3.6 Observation Error Covariance . . . . .	80
4.3.7 Ensemble Resampling and Covariance Inflation . . . . .	81
4.4 Results . . . . .	83
4.4.1 Experiment Setup . . . . .	83
4.4.2 Estimated Bathymetry . . . . .	85
4.4.3 Improved Prediction of Currents . . . . .	90
4.4.4 Observation Impact and Quantitative Skill Assessment . . . . .	92
4.5 Discussion . . . . .	96
4.5.1 Effects of Model Error . . . . .	96
4.5.2 Representation of Posterior Uncertainty . . . . .	97
4.5.3 Computational Efficiency . . . . .	98
4.6 Summary . . . . .	100
5 General Conclusion	102
Bibliography	104
Appendices	115
A Continuous Ranked Probability Score . . . . .	116
B Skill Statistics for $x > 250$ m . . . . .	118
C Global Wave Forecasting Model . . . . .	120

# LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
2.1	Top: plan-view map of observational array. Bottom: side-view, with depth transect from 22 September survey, showing typical bathymetric profile and sensor positions in water column. Still water level ranged from -0.5 m to 1.5 m NGVD during the experiment. . . . .	9
2.2	Top: leading modes in bathymetric EOF decomposition (normalized to unit variance). Bottom: percent of total variance for leading ten modes of bathymetric EOF decomposition. . . . .	13
2.3	Conditions observed for the period from 18 October to 21 October during the SD97 experiment: offshore significant wave height $H_{mo}$ , peak frequency $f_p$ , mean wave angle $\theta_0$ (Kuik <i>et al.</i> , 1988) (positive is from the north), and maximum alongshore current in SPUV array $v_{max}$ (positive is towards the north). Shaded regions represent times when bathymetric surveys were being performed (see Figure 2.4). Dashed lines correspond to 20 October, 1130 EST and 1530 EST (sections 2.4.2 and 2.4.3). . . . .	18
2.4	Black: bathymetric transects collected by CRAB on (from left to right) 18, 20, and 21 October. Colors: interpolated surveyed bathymetry $z_b$ (hotter colors represent shallower water). . . . .	19
2.5	Prior (blue), posterior (red), and measured (black) data for 1DH case study, 20 October 1130 EST. Dashed lines represent $\pm$ one standard deviation. Black circles correspond to data from fixed instruments (SPUV); black line is CRAB transect collected at $y = 843$ m, on 20 October, 1036–1040 EST. Still water level was 1.16 m NGVD. . . . .	21
2.6	Solid lines: scaled representers. Dashed lines: corresponding approximation based on extrapolation with $C_{hh}$ (see text). Crosses indicate location of measurement for each representer. $\hat{r}_{hv}$ has been negated for comparison with Figure 2.5. . . . .	23
2.7	Prior (blue), measured (black) and posterior data when assimilating $H_{mo}$ only (green), $v$ only (magenta), or both (red). . . . .	24
2.8	Scaled representers for correction of $H_{mo}$ . Crosses indicate location of measurement for each representer. $\hat{r}_{Hv}$ has been negated for comparison with Figure 2.5. . . . .	25

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
2.9 Prior mean velocity, bathymetry, and wave height on transect $y = 828$ m, for 20 October, 1530 EST. Red arrows in left-hand plot are observed velocity; blue arrows are modeled velocity, plotted at even gridpoints; scale arrow in upper left is 50 cm/s. Colors in right-hand plot are model $z_b$ (still water level was 0.17 m NGVD). Solid and dashed lines in the lowermost plot are prior mean and standard deviation of $H_{mo}$ , and red circles are measured $H_{mo}$ . . . . .	26
2.10 As in Figure 2.9, but for posterior (updated) fields. . . . .	27
2.11 Across-shore transect ( $y = 828$ m) of prior (blue), posterior (red), and measured (black) $z_b$ (still water level was 0.17 m NGVD). Dashed lines represent $\pm$ one standard deviation. . . . .	28
2.12 Maps from scaled representer sub-matrix $\hat{r}_{hH}$ (top) and $\hat{r}_{hv}$ (bottom), for select measurement locations (white dots). Thick black contour is 0 cm, and subsequent contours are plotted at 5 cm intervals. . . . .	29
2.13 As in Figure 2.10, but for posterior (updated) fields using alongshore transect sampling scheme. Assimilated observations of $v$ are marked by white crosses, and observations of $H_{mo}$ are marked by white circles. . . . .	34
2.14 As in Figure 2.13, but for across-shore sampling scheme. . . . .	35
3.1 Schematic of 1D test bathymetry $h(x)$ , equation (3.9), showing the location of the river bottom (or negative-depth, $-h(x)$ ) as a function of along-channel distance ( $x$ ). The flow velocity $u(x)$ is assumed depth-uniform, and is in the $+x$ direction. . . . .	45
3.2 Example estimates of $h(x)$ using equation (3.2), for various values of bump size $h_b$ ; corresponding true $h(x)$ are plotted as dashed lines. . . . .	46
3.3 Estimated depth correction $\Delta h = h^f - h$ from equation (3.2), plotted as a function of prior velocity error $\Delta u = u^f - u_{meas}$ , at the location $x = l/2$ . True relationship (using $u = Q/h$ ) is plotted as dashed line; differences between true and estimated relationship are attributed to nonlinearity (see text). . . . .	47
3.4 Left: convergence of root-mean-square error (rmse) for increasing ensemble size; average (solid) and $\pm$ one standard deviation (dashed), computed over 10 realizations of each ensemble size. Right: example estimates using various ensemble sizes $N$ . . . . .	48

## LIST OF FIGURES (Continued)

<u>Figure</u>		<u>Page</u>
3.5	Left: depth estimates of a bump with $h_b = 1$ m, for $L_x = 5$ m (blue), 50 m (green), 100 m (red), and 200 m (cyan). True depth is plotted in black, and measurement locations are marked by circles (measurement sample spacing is $dx_{\text{obs}} = 12.5$ m). Right: root-mean-square (rms) difference between true and estimated bathymetry, using same bump geometry, for various decorrelation lengths $L_x$ , and various measurement sample spacings $dx_{\text{obs}}$ . Ten realizations of 500-member ensembles are computed for each combination of $L_x$ and $dx_{\text{obs}}$ ; solid line represents average rms error, and dashed lines are $\pm$ one standard deviation. . . . .	48
3.6	Prior bathymetry (a) and measured bathymetry (b) for Snohomish River test case. Inset shows along-channel-average of prior bathymetry, plotted as a function of across-channel position. . . . .	53
3.7	Posterior bathymetry for Snohomish River test case, with various sampling schemes: (a), assimilating 2D velocities, $\sim 10$ m grid; (b), assimilating across-channel velocity only, $\sim 10$ m grid; (c), assimilating along-channel velocity only, $\sim 10$ m grid; (d), assimilating across- and along-channel velocities at 5 m spacing on a simulated drifter track; and (e), assimilating isolated point measurements of across- and along-channel velocities. For (d) and (e), the locations of measurements are indicated by black dots. Bottom plot shows across-channel-averaged depths, compared to truth. . . . .	54
3.8	Posterior standard deviation (i.e. uncertainty) of bathymetry for Snohomish River test case, after assimilation of measurements using various sampling schemes. Labels (a)–(d) are as in Figure 3.7. . . . .	56
3.9	Prior (a), posterior (b), and measured (c) bathymetry, for Kootenai River, assimilating 2D velocities on $\sim 10$ m grid. Inset shows along-channel-average of prior bathymetry, plotted as a function of across-channel position. . . . .	57
4.1	Map of 2010 field experiment, showing locations of in-situ and remote sensing data. Location of shoreline and FRF pier are plotted as black lines for reference (offshore is in the positive- $x$ direction). In-situ data consist of 8m-array pressure gages ( $\circ$ ), co-located FRF pressure and current-profiling instruments ( $+$ ), and acoustic Doppler current profiler ( $\times$ ). Remote sensing data, from Argus tower ( $\otimes$ ) and radar tower ( $\oplus$ ), consist of Optical Current Meter (blue), Infrared PIV (red), CB1-optical (green), CB1-radar (magenta), and shoreline (yellow). . . . .	64



## LIST OF FIGURES (Continued)

<u>Figure</u>		<u>Page</u>
4.2	Example shoreline detected from optical imagery on September 13, 1200 EST. Shoreline (red line) is defined as shoremost maximum of variance image (a); time-exposure (b) and snapshot (c) images are also shown for reference. . . . .	66
4.3	Left (a): example data product for Optical Current Meter (OCM): September 13, 0830 EST. Red arrows represent spatially-smoothed measured alongshore currents, with across-shore currents derived using continuity equation; yellow arrows are a numerical model prediction using measured bathymetry; magenta scale arrow in lower right represents 0.5 m/s; background is an Argus time-exposed image from the same time period. Right (b): comparison between OCM alongshore current, and in-situ (AquaDopp) measurements of depth-averaged current. . . . .	67
4.4	Left (a): comparison between wavenumber from optical CB1k product, and an estimate from cross-spectral analysis of in-situ pressure gage pairs. Dots represent in-situ gage pair $(x, y) = (233, 940), (375, 939)$ m; crosses represent $(375, 939), (446, 938)$ m. Right (b): comparison between wave angle CB1a and peak wave angle from analysis of in-situ gage data. Gages are located at $(x, y) = (233, 940)$ m (dots) $(375, 939)$ m (crosses), and $(446, 938)$ m (plusses). . . . .	69
4.5	Left (a): example data product for Infrared Particle Image Velocimetry: September 13, 1800 EST. Red arrows represent measured currents; yellow arrows are a numerical model prediction using surveyed bathymetry; magenta scale arrow in lower right represents 0.5 m/s; background is a time-exposed image during the same period (unfortunately not very informative due to low light). Right (b): comparison of IR-PIV estimates of alongshore current, and in-situ (AquaDopp) measurements of depth-averaged current. . . . .	71
4.6	As in Figure 4.4, but for CB1-radar. . . . .	72

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
<p>4.7 Conditions observed during experiment, vs. time in EST. The start/end time of the September 13 test case presented in section 4.4 is marked by vertical lines. Plots (a)–(c) show observed wave conditions in 8 meters depth: significant wave height <math>H_{m0}</math>, and integrated frequency and directional wave energy density (shown as shading in (b) and (c), normalized to unit energy for each time), with peak frequency and direction marked by solid lines. Wave directions are measured counterclockwise from the positive-<math>x</math> axis to the direction waves are coming from. Plot (d) shows observed tidal elevations relative to the NAVD88 vertical datum. . . . .</p>	84
<p>4.8 Number of remote sensing observation data points, vs. time in EST. The start/end time of the September 13 test case presented in section 4.4 is marked by vertical lines. Legend indicates the type of observation; the total number of observations during the September 13 test period is listed in parentheses. . . . .</p>	85
<p>4.9 Plot (a): background bathymetry <math>h^b</math> used to initialize the ensemble assimilation system: a climatological average of bathymetry observed at the field site. Plot (b): updated bathymetry after assimilating all observations except wave angle, over 24 assimilation cycles (ending 1830 EST). Plots (c,d): surveyed bathymetry from September 6 (c) and September 15 (d), where individual survey data points are plotted as yellow dots, and interpolated depths are shown as colors (white area in upper-right of Sep. 15 bathymetry is due to a gap in the data). Colorbar with contour marks at far right applies for all plots, and refers to depth relative to the NAVD88 vertical datum; in the case of the updated bathymetry (b), the same colors and contours are used, but color transparency is scaled to represent the posterior estimate of bathymetric uncertainty. . . . .</p>	86
<p>4.10 Transects from initial (blue) and final (red) estimated bathymetry (as in Figures 4.9a and 4.9b, respectively), and <math>\pm</math> one standard deviation (dashed lines), compared to measured bathymetry on September 6 (green) and September 15 (black). Alongshore transect (a) is from <math>x = 150</math> m, located over the nearshore terrace; across-shore transects (b,c) are from <math>y = 690, 870</math> m. Dots represent measured data from within 2 meters of the alongshore transect in plot (a), and within 10 meters of the across-shore transects in plots (b,c); green and black lines represent smoothed 2D interpolation of raw survey data, as in Figures 4.9c and 4.9d. . . . .</p>	88

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
<p>4.11 Difference between estimated bathymetry and raw bathymetric survey data. Top plots (a,c) use data from the September 6 survey, showing differences before (a) and after (b) data assimilation; bottom plots (c,d) are the same, but using data from the September 15 survey. Positive differences indicate an overprediction of depth. Black contours are the estimated depths as in Figure 4.9. . . . .</p>	89
<p>4.12 Measurements and forward model predictions for September 13, 1800 EST (low tide). Background shading represents X-band radar backscatter averaged over a 17 minute collection period; high backscatter (green) corresponds to breaking waves, and/or surface roughness due to the presence of a rip current (Haller <i>et al.</i>, 2013). Red arrows are IR-PIV measurements of currents (scale arrow in lower-right is 50 cm/s), which confirm the presence of a rip current at <math>y \approx 900</math> m. Red line is optical-based shoreline position. Yellow arrows are predicted time-averaged currents from a forward model run using (a) the estimated bathymetry after assimilation of data from 0700–1200 EST, before any rip current was observed, and (b) the estimated bathymetry after assimilation of data up to and including 1800 EST. . . . .</p>	91
<p>4.13 Estimated bathymetry from data-denial experiments (see text), plotted as in Figure 4.9b. Each result is from assimilating shoreline observations and one other observation type: (a), alongshore currents (OCM and IR-PIV); (b), wavenumber (CB1k and CB1k-radar); (c), wave angle (CB1a and CB1a-radar); (d), shoreline only. . . . .</p>	93
<p>4.14 As in Figure 4.13, but assimilating shoreline plus (a) OCM currents, and (b) IR-PIV currents. Note the different observational footprint of the two products, and the resulting differences in estimated bathymetry. . . . .</p>	95
<p>4.15 Root-mean-square difference in estimated bathymetry for various ensemble sizes in a simplified assimilation system (see text), compared to a reference run with 300 ensemble members. In each case, differences are computed over the domain <math>50 &lt; x &lt; 500</math> m and <math>0 &lt; y &lt; 1000</math> m, and over each of the 24 assimilation cycles. Three realizations (indicated by circles) were performed for each ensemble size, to account for differences due to the stochastic nature of the filter. Note, the rms-difference computed for two runs with 300 ensemble members each was 4.8 cm, which defines an approximate “noise floor” for the statistic. . . . .</p>	99

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
4.16 Bathymetry estimate as in Figure 4.9b, but using 1/4 ensemble size (50 members rather than 200). . . . .	99

# LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1 Model accuracy statistics before and after assimilation, for sensors in $x < 250$ m. $\epsilon$ is rms difference between model and observations, CRPS is the Continuous Ranked Probability score (see text, and appendix A), and $S$ is a skill score (equation (2.10)). The calculation of CRPS for $u$ and $h$ assume observational error standard deviations of 6.7 cm/s (same as for $v$ ) and 10 cm, respectively. . . . .	30
3.1 Accuracy statistics for Snohomish River bathymetry estimates, comparing several measurement sampling schemes. Labels in first column are: Prior, no assimilation; Pts, assimilating isolated point observations of 2D velocity; $u$ , assimilating observations of along-channel (but not across-channel) velocity on a $\sim 10$ m grid; $v$ , assimilating observations of across-channel (but not along-channel) velocity on a $\sim 10$ m grid; $u, v$ , assimilating observations of 2D velocity on $\sim 10$ m grid; and Drifter, assimilating 2D velocity sampled along a simulated drifter track. The statistics are root-mean-square error (rmse) and squared-correlation ( $r^2$ ); separate columns show statistics computed using all model gridpoints as the sample (“Full-Domain”), and using only the measurement locations as the sample (“Obs.-Points”). . . . .	52
3.2 Accuracy statistics for Kootenai River test cases (cf. Table 3.1 for labeling conventions). . . . .	56
4.1 Skill statistics, comparing September 6 and September 15 raw bathymetric survey data (Figures 4.9c and 4.9d, resp.) to estimates from data-denial experiments presented in Figure 4.13 (labels (a)–(d) are following Figure 4.13), as well as the initial estimate before data assimilation (“Initial”, Figure 4.9a) and the estimate when assimilating all observational data except wave angles (“All”, Figure 4.9b). The statistic $\epsilon$ is defined as the root-mean-square difference between surveyed and estimated bathymetry, and $r^2$ is the squared correlation. Both statistics are computed over two regions: Region “A” is the area where both alongshore current and wavenumber were observed (union of red and blue polygons in Figure 4.1); Region “B” is the union of all observed areas (union of all polygons in Figure 4.1). . . . .	94

## LIST OF APPENDIX FIGURES

<u>Figure</u>	<u>Page</u>
A.1 pdf (left) and cdf (right) representations of the prediction (blue) and observation (red). . . . .	117

## LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
B.1 As in Table 1, but for $x > 250$ m. . . . .	119
C.1 Description of grids used by wave forecasting model. . . . .	120

# Data Assimilation for Prediction of Shallow Water Flows with Uncertain Bathymetry

## 1. General Introduction

In 1871, Adhémar Jean Claude Barré de Saint-Venant published a manuscript introducing the mathematical treatment of one-dimensional shallow water hydrodynamics, the equations which now bear his name (Barré Saint-Venant, 1871). At the time, Saint-Venant’s physical reasoning would not have been considered groundbreaking (having been developed by Euler over a century beforehand), and his mathematical derivation was not particularly difficult. However, the monumental impact of Saint-Venant’s equations stems from the fact that shallow water hydrodynamics applies to a huge breadth of societally relevant problems. One example is the propagation of tides in an inlet or estuary, the original motivation for Saint-Venant’s work. Indeed, a century later the same approach was used for an analysis of flooding on the Thames River including the city of London (Bowen & Pinless, 1974), to determine the level of coastal protection needed, and inform the design and operation of the Thames Barrier. Another example is the prediction of nearshore currents: the importance of currents in amphibious naval operations during World War II led to an increased scientific interest in nearshore prediction (Sverdrup & Munk, 1946; Galvin, 1967), and ultimately to a mathematical description using shallow water theory<sup>1</sup> (Longuet-Higgins, 1970a; Bowen, 1969), which forms the foundation of most contemporary nearshore models. It is a testament to the importance of shallow water hydrodynamics, and hence to the insight of Saint-Venant himself, that his equations have touched so many applications in engineering and science.

Following the development of basic mathematical descriptions for shallow water flows, researchers have since turned to the study of increasingly detailed physics in the same general setting, namely unstratified nearshore/estuarine/river environments (the subject

---

<sup>1</sup>It is interesting to note the parallels between the development of nearshore hydrodynamics and that of Saint-Venant’s equations. In both cases, proposed models from the engineering community for solving practical problems (Galvin, 1967) were ultimately replaced by a more general theory using shallow water equations. In the nearshore, a key insight was the introduction of the radiation stress due to surface gravity waves (Longuet-Higgins & Stewart, 1964).



of the present work). Examples include the study of bottom boundary layer processes (Smith & McLean, 1977; Grant & Madsen, 1979; Nielsen, 1992) and horizontal momentum mixing (Svendsen & Putrevu, 1994; Nadaoka & Yagi, 1998). Still others have looked to three-dimensional aspects which are not explicitly included in the traditional shallow water equations (Bathurst *et al.*, 1979; Garcez-Faria *et al.*, 2000). The ability to compute solutions at high resolution in space and time has also permitted the study of detailed flow behavior; some examples in nearshore oceanography include the generation and evolution of low-frequency eddies (Bowen & Holman, 1989; Özkan Haller & Kirby, 1999; Clark *et al.*, 2012), the response of the flow to a spatially varying seafloor (Slinn *et al.*, 2000; Wilson *et al.*, 2013), and the coupling between hydrodynamics and sediment transport (i.e., nearshore morphodynamics, Wright & Short (1984)). Similarly in fluvial dynamics there has been much interest in the understanding of flow in meandering and braided river channels, and, again, the resulting morphodynamics (Smith & McLean, 1984; Ikeda & Parker, 1989). As a result of these and other efforts, present day understanding of shallow water hydrodynamics<sup>2</sup> can be considered quite mature, and numerical models can now generate quantitative predictions of natural flows.

This ongoing development has, of course, been paralleled by continued interest in new real-world applications of such models. Two such applications have already been mentioned, above (tidal flooding in estuaries, and naval operations in the nearshore). Other recent examples include the prediction of hazardous currents for recreational beach users (Austin *et al.*, 2012), and the transport of tracers, such as bacteria/pollutants (Feng *et al.*, 2013), or larval organisms (Shanks *et al.*, 2010; Rilov *et al.*, 2008), along natural coastlines. Due to efforts to improve model physics, as above, the challenges faced in more-recent applications are often less related to the correctness of the model, and more related to the ability to correctly specify the physical environment. In short, model boundary conditions (and other inputs) can never be exactly specified in the real world, and this can have serious consequences for model accuracy. Hence, the complexity of coastlines, inlets, and rivers is tantalizing to some researchers (Coco & Murray, 2007), but can spell disaster for others seeking to make quantitative predictions or forecasts.

It is this problem — hydrodynamic prediction in poorly-constrained natural shallow water systems — that forms the motivation for the present work. We ask: how can we make meaningful predictions in situations where the model physics is well-established, but the detailed physical environment is unknown or uncertain? In particular, we focus on one important aspect of that problem: uncertainty of bathymetry. We present a new

---

<sup>2</sup>We now use this term loosely, not specifically referring to the shallow water equations.

approach, using methods from data assimilation (specifically the ensemble Kalman filter; Evensen (2006)) to incorporate bathymetric uncertainty into model predictions, and to control that uncertainty via the inclusion of additional observational data.

The second chapter of this dissertation introduces the application of data assimilation for the prediction of surf zone waves and currents. The surf zone, defined as the nearshore region where depth-limited wave breaking occurs, is a highly energetic and dynamic environment. Surf zone bathymetry is difficult to measure directly, and can change dramatically within a matter of days due to sediment transport. We present a real-world case study where the resulting bathymetric uncertainty dominates the sources of error for prediction of a surf zone flow. Assimilation of data (in-situ measurements of alongshore current and wave height) is subsequently shown to alleviate this model error by correcting bathymetry errors. We also delve into the basic workings of the assimilation method, and explore its ability to model interrelationships among variables in the presence of bathymetric uncertainty. This work appeared as a publication in the *Journal of Geophysical Research, Oceans* (Wilson *et al.*, 2010).

Chapter Three applies the same methodology to a different shallow water setting: that of a river or narrow estuary. Here, the method is again shown to be capable of modeling the relationship between uncertain bathymetry and uncertain model predictions of river currents. Again, the assimilation of observed currents is shown to correct errors in bathymetry, or, equivalently, to estimate bathymetry as an “inverse problem”. In this case, the investigation is done using twin-tests, to explore several possible methods of observation (fixed in-situ gages, remote sensing, and passive drifters). This work appeared as a publication in the *Journal of Atmospheric and Oceanic Technology* (Wilson & Özkan-Haller, 2012). The results have since been extended by Landon (2012), using the same methodology but with actual observations of currents from passive drifters deployed by Swick (2011). Landon found that the method was capable of generating accurate estimates of bathymetry in two different river reaches, was robust against errors in input parameters such as the expected length scale of unknown bathymetric features, and did not require unreasonable observational effort (ca. 10 drifter trajectories were sufficient for an accurate estimate).

In Chapter Four, we return to the surf zone environment, with the goal of refining the data assimilation method for potential use in a more operational-oriented application. Whereas Chapters Two and Three focused on how and why the method works in theory (taking advantage of a relatively controlled observational setting), here we shift to how it can be used in practice. To that end, we incorporate new methods for observing the surf

zone remotely, using shore-based optical and infrared video cameras, as well as marine radar. An obvious advantage of such methods is that the surf zone is an unforgiving environment for in-situ observation (due to breaking waves, strong currents, and often an unstable sandy bottom). Using remote sensing, one is able to observe for a longer period of time, and over a wider area. Surf zone bathymetry, however, cannot be observed remotely, hence the present method is a natural fit. We show that remote sensing data can be successfully used to control bathymetric errors, as in the previous two chapters. We also show how assimilation of data can permit model predictions of a rip current (whose presence is strongly tied to a bathymetric feature), even in the absence of any in-situ bathymetric observations.

It is worth mentioning that the main three chapters (2–4) of this dissertation also take three complimentary viewpoints on the problem of bathymetric uncertainty in predictive models. In Chapter Two, we show that bathymetric uncertainty due to rapid bathymetric change (time scale of days) has a strong impact on model error, which can be modeled and corrected using data assimilation. Chapter Three takes a more pragmatic approach, where one seeks to estimate bathymetry using observations of currents (the “inverse problem”), without giving much importance to the resulting effects on model predictions. Chapter Four takes the viewpoint of a forecaster, who is interested in “controlling” bathymetry error for the purposes of improving model predictions over time. Throughout, a unifying theme is the use of data assimilation to incorporate and manipulate uncertainty in the model and its inputs.

# DATA ASSIMILATION AND BATHYMETRIC INVERSION IN A 2DH SURF ZONE MODEL

G.W. Wilson, H.T. Özkan-Haller, and R.A. Holman

Journal of Geophysical Research: Oceans  
2000 Florida Avenue N.W., Washington DC, 20009-1277, U.S.A.  
vol. 115, C12057, doi:10.1029/2010JC006286, 2010

## 2.1 Introduction

Many nearshore circulation models utilize the depth- and wave-averaged equations of motion, coupled with a “wave driver” for transfer of momentum from incident gravity waves to surf zone currents. When validating these models (i.e. assessing their ability to match observations), one must consider two sources of error: mis-specification of physical processes in the model, and errors in model inputs, namely the underlying bathymetry and boundary conditions. Here, we will refer to these as “process error” and “input error”, respectively.

The majority of previous studies (e.g. Longuet-Higgins (1970b); Thornton & Guza (1986); Reniers & Battjes (1997)) have focused on minimizing process error, leading to improved parameterizations and empirical calibrations now standard in nearshore models. Meanwhile, the potential role of input error is often acknowledged but tends to be difficult to quantify, let alone to correct. An important example, which is the focus of this paper, is the presence of bathymetric uncertainty when modeling surf zone currents.

Bathymetric input error may appear in various forms. In the extreme case, where the bathymetry has not been measured, one is forced to assume some reasonable beach shape for the model. Even when measurements are available, they are subject to instrument error. Spatial undersampling may not resolve high-wavenumber bathymetric features (Plant *et al.*, 2002), and the resulting spatial smoothing may affect model outputs (Plant *et al.*, 2009). Temporal undersampling may also occur, as beach changes occur on daily or even hourly time scales. In field situations, these various sources of bathymetric input error may be as important as process error as constraints on model accuracy.

In this study, we address the issue of bathymetric input error from two perspectives. First, we seek to quantify the sensitivity of the model to errors in bathymetry. Second, at the same time, we evaluate a method for indirectly correcting bathymetric errors, by incorporating in-situ measurements of waves and currents. These two perspectives encompass data assimilation and bathymetric inversion.

In previous data assimilation efforts, Feddersen *et al.* (2004) used a variational approach (Bennett, 2002), deriving adjoint equations for a surf zone model involving linearized alongshore-uniform dynamics, to assimilate pressure and bi-directional current (PUV) measurements on a natural beach. Kurapov *et al.* (2007) extended this approach to the nonlinear 2DH time-dependent equations, and used the process of nonlinear shear instability in alongshore currents (Slinn *et al.*, 1998) as a test-bed for variational data assimilation. Both these studies focused on model sensitivity in the form of forcing errors,

which we would characterize as process error. Feddersen *et al.* (2004) also considered sensitivity to the bottom friction coefficient, i.e. input error.

Regarding the topic of surf zone bathymetric inversion, the majority of previous studies have focused on the technical challenge of observing surface wave properties, which are often related to water depth using simple physical models. For instance, previous studies have estimated bathymetry using the linear wave dispersion relationship (Stockdon & Holman, 2000), the nonlinear wave dispersion relationship (Catalán & Haller, 2007), or wave refraction (Splinter & Holman, 2009). Wave breaking dissipation proxies have also been used in combination with empirical models (Aarninkhof *et al.*, 2005) to infer bathymetric changes.

Recently, van Dongeren *et al.* (2008) have applied data assimilation techniques to the problem of bathymetric inversion, providing a fresh perspective on this long-standing problem. Their method employs a sequential least-squares estimator, which assimilates multiple remote sensing (video and radar) wave observations. While not as sophisticated as the variational schemes of Feddersen *et al.* (2004) or Kurapov *et al.* (2007), their technique stems from a comparable approach. Adjoint equations are derived, in this case, from simple localized models for the observed physical processes. Spatial covariance is neglected (although this may be unimportant for the spatially-dense observations being considered), and temporal covariance is approximated empirically. Despite these simplifications, their results show the most robust and reliable bathymetric inverse to date. This reliability stems from the important step of acknowledging data errors as well as model errors, and covariances thereof, in order to form a statistically optimal estimate.

The present work is conceptually similar to that of van Dongeren *et al.* (2008), but with some key differences. First, our method quantifies model sensitivity using statistical, rather than analytical, means. This facilitates the assimilation of arbitrary geophysical variables, without having to compute their derivatives with respect to depth (i.e. adjoint equations). Second, we incorporate spatial covariance, and hence can compute non-local corrections based on local in-situ measurements. Finally, we put special emphasis on bathymetric sensitivity and its role in the inversion problem.

The paper is organized as follows: section 2.2 introduces the in-situ data set. Section 2.3 presents the data assimilation methodology: the parameter estimation scheme, the hydrodynamic model, and the technique used to represent bathymetric uncertainty. Section 2.4 gives examples of the application of that methodology to field data. Sections 2.5 and 2.6 summarize and discuss the results, and give conclusions.

## 2.2 Observations

In this study, we will use a subset of the data collected during the SandyDuck '97 (SD97) experiment (Duck, NC); Figure 2.1 shows the experimental layout. Between 22 September and 31 October, the U.S. Army Corps of Engineers Field Research Facility (FRF) conducted 38 daily bathymetric surveys in the region shown (except for 19 October), using the CRAB survey vehicle (Birkemeier, 1984) to collect multiple across-shore transects of bathymetry spaced 25-50 m apart. These surveys were interpolated onto a regular grid with 5 m and 10 m spacing in the across-shore and alongshore directions, respectively, using a quadratic loess filter (Plant *et al.*, 2002) with interpolation filter length scales of 200 m in the alongshore direction, and 5 m in the across-shore direction. Additional bathymetry was incorporated from larger-scale surveys conducted on 16 September and 23 October, such that the total model domain was  $0 \leq x \leq 900$  m and  $0 \leq y \leq 1000$  m; the detailed daily surveys were stitched on top of the larger-scale bathymetry using weighted interpolation, i.e.

$$h = wh_0 + (1 - w)h_1, \quad (2.1)$$

where  $h_0$  is the larger-scale bathymetry,  $h_1$  is the detailed minigrid bathymetry, and  $w$  is a weighting function which ramps from 0 to 1 over 50 m at each edge (using a tanh shape). Herein, plots and figures will present bathymetric data in terms of distance from the National Geodetic Vertical Datum (NGVD) to the sea floor, or  $z_b$ ; that is,  $h = -z_b + z_t$ , where  $z_t$  is the still water level (which changes with time due to tide and large-scale surge).

An offshore array of 15 bottom-mounted pressure gages (labeled “8m-array” in Figure 2.1; Long (1996)) provided frequency-directional wave spectra at 3-hour intervals, which are used to specify the offshore boundary for the wave model (section 2.3.3). When the model time is not centered on a 3-hour collection time, a time-interpolation is applied using the scheme described in the SWAN wave model user manual ([www.wldelft.nl](http://www.wldelft.nl)).

In-situ measurements from sonar altimeters, pressure gages, and bi-directional current meters (SPUV) were provided at the locations shown in Figure 2.1 (Elgar *et al.*, 2001). These were processed to obtain the significant wave height  $H_{mo}$  (17 minute intervals), time-averaged currents (17 minute intervals), and estimated depths (three-hour intervals). The depth measurements were further processed by comparing to the daily bathymetric surveys; offsets which persisted for five or more days were removed from the sonar altimeter data, in cases where the offset was significant (using Welch’s t-test for significance of bias, with  $n \geq 5$  and  $p < 0.05$ ) and the correlation was significant

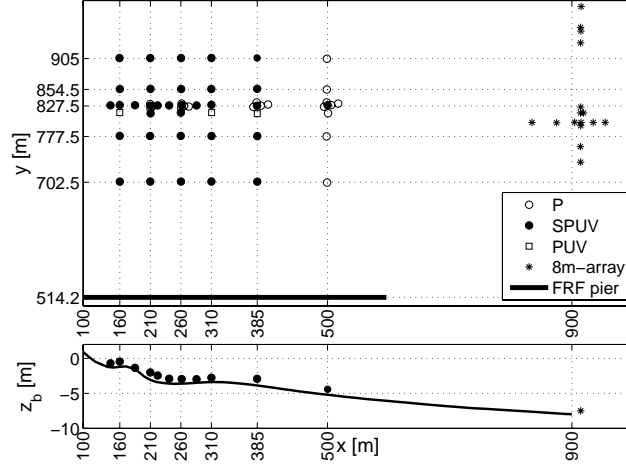


Figure 2.1: Top: plan-view map of observational array. Bottom: side-view, with depth transect from 22 September survey, showing typical bathymetric profile and sensor positions in water column. Still water level ranged from -0.5 m to 1.5 m NGVD during the experiment.

and positive ( $p < 0.05$ ,  $r > \sqrt{0.1}$ ). These corrections were typically small (less than 20 cm) and in deep water (depths greater than 3 m). Agreement with the CRAB surveyed depths was otherwise excellent (root-mean-square difference was 7.3 cm).

## 2.3 Methodology

In this section, we outline a methodology for assimilating data and generating bathymetric inversions based on point observations of wave height and alongshore current. As a general overview, the method involves the following steps:

1. Generate an ensemble  $h^f$  ( $f$  for “forecast”, following the standard notation) consisting of  $N$  realizations of bathymetry (in our application,  $N = 150$ ). The distribution of the ensemble should be representative of prior knowledge and uncertainty (section 2.3.2).
2. Apply the hydrodynamic numerical model (section 2.3.3) to each of the  $N$  bathymetric ensemble members, assuming other inputs are perfect (e.g. wave spectrum at offshore boundary), and store the output.
3. Compute the sample mean and covariance from the ensemble of modeled fields.



4. Generate an updated (posterior) state  $\psi^a$  ( $a$  for “analysis”, equation (2.3)), which includes bathymetry, wave height, currents, and calculate the posterior uncertainty  $C_{\psi\psi}^a$  (equation (2.5)).

We note this methodology is not particularly new or novel: Mourre *et al.* (2004) have previously applied ensemble-based methods (steps 1–3) to examine bathymetric sensitivity in a regional ocean model. The equations for optimally updating the model (step 4) are following a vast and ongoing literature on data assimilation using ensemble-based methods, for example the ensemble Kalman filter (EnKF) (see Evensen (2006), on which the present method is largely based, as well as references therein). The following section reviews the existing methodology, as it applies to the unique problem studied here: nearshore bathymetric sensitivity and inversion.

### 2.3.1 Theory for Bathymetric Inversion

To begin, we define some notation which will be useful in what follows. Suppose, for simplicity, we are dealing with a single observable  $v$ . In later sections we will in fact observe and assimilate two variables, alongshore current and wave height, but the extension is straightforward: simply augment  $v$  with  $H_{\text{mo}}$ . We are also given a model,  $v = G(h)$ , which makes predictions of  $v$  on a discrete spatial grid, given the water depth  $h$ . Here, we have assumed the model (including all boundary conditions and inputs other than  $h$ ) is “perfect”, so the error of the model prediction is due only to errors in  $h$ . To that point, we define the model input bathymetry  $h = h^t + p$ , where  $h^t$  is the true bathymetry. When the error  $p$  is included in the model input, the resulting prediction is  $v = G(h^t + p) = v^t + q$ .

Our goal is to obtain an optimal estimate of the true field on the model grid,  $v^t = G(h^t)$ , given a set of  $K$  observations which are themselves subject to some error  $\epsilon$ ,  $d = Lv^t + \epsilon$ . Here,  $L$  is a measurement operator, in our case simply a matrix which linearly interpolates from the gridded field to the measurement locations. We define the optimal posterior estimate  $\psi^a$  as the one which minimizes the following cost function:

$$\begin{aligned} \mathcal{J}[\psi] = & \left( \psi - \psi^f \right)^T W_{\psi\psi} \left( \psi - \psi^f \right) \\ & + (d - Lv)^T W_{dd} (d - Lv), \end{aligned} \quad (2.2)$$

where  $\psi = [v^T, h^T]^T$  is a state variable in which  $v$  is augmented with  $h$  (note  $v$  and  $h$  should each be treated as  $M \times 1$  vectors, where  $M$  is the total number of model

gridpoints),  $\psi^f$  corresponds to a prior estimate for  $v$  and  $h$ , and  $W_{\psi\psi}$  and  $W_{dd}$  are positive-definite weighting matrices.

Note that  $\mathcal{J}$  contains a “model” part and a “data” part: the model part says the posterior state should not stray too far from the prior (hence, it retains physics from the model solution), and the data part says the posterior should match closely with observations. If  $W_{\psi\psi} \gg W_{dd}$  (“perfect prior”) the posterior solution is just the prior  $\psi^f$ , and if  $W_{dd} \gg W_{\psi\psi}$  (“perfect data”) the posterior solution is an exact interpolation of the data  $d$ . Clearly, the perfect-prior assumption ignores the information contained in the observations; the perfect-data assumption, on the other hand, can lead to interpolation of observation noise. Hence, a central challenge of data assimilation is to find the correct balance between these two extremes by correctly choosing the weights  $W$ . In some cases, it is useful to “hedge” the estimate towards the prior, for instance if there is a possibility of instrument malfunction. In other cases, the observations are known to be very accurate and a perfect-data assumption is valid.

It can be shown (Evensen, 2006; Bennett, 2002) that the solution  $\psi^a$  which minimizes  $\mathcal{J}$  is given by

$$\psi^a = \psi^f + C_{\psi\psi} L_a^T (L C_{vv} L^T + C_{dd})^{-1} (d - L \psi^f), \quad (2.3)$$

where  $L_a = [L, 0^{K \times M}]$  is an augmented measurement operator for extracting  $v$  from  $\psi$ , and  $C_{\psi\psi}$  and  $C_{dd}$  are the inverse of the weights  $W_{\psi\psi}$  and  $W_{dd}$ . Specifically,  $C_{\psi\psi}$  has the following structure:

$$C_{\psi\psi} = \begin{pmatrix} C_{vv} & C_{vh} \\ C_{hv} & C_{hh} \end{pmatrix} \quad (2.4)$$

By choosing  $C_{\psi\psi}$  as the covariance, equation (2.3) gives the maximum likelihood estimator for Gaussian statistics.

At this point equation (2.3) can be viewed as a general solution, and the problem is reduced to specifying  $C_{\psi\psi}$  based on properties of the model ( $C_{dd}$  is typically specified as a diagonal matrix whose elements are the observation error variances). The simplest approach is to define  $C_{\psi\psi}$  a-priori without reference to the model itself. A more attractive approach, which recognizes the intrinsic properties of the model, is the method of representer expansions (Bennett, 2002). That method requires the specification of  $C_{hh}$ , but uses the model (via adjoint equations, which must be derived) to obtain the corresponding  $C_{vh}$  and  $C_{vv}$ . Feddersen *et al.* (2004) and Kurapov *et al.* (2007) used representer expansions, except their goal was to correct forcing and/or bottom friction, not bathymetry. van Dongeren *et al.* (2008) used a hybrid approach, where  $C_{vh}$ ,  $C_{vv}$ ,

and  $C_{hh}$  were assumed a-priori to be diagonal matrices (hence the correction is localized), but are related to one another by a physical model. In our application, we use an ensemble-based methodology, described next.

The crux of ensemble-based methods (e.g. the ensemble Kalman filter, Evensen (2006)) is that  $C_{\psi\psi}$  is approximated by the sample covariance of a representative ensemble  $\psi_i^f$ ,  $i = 1, 2, \dots, N$ . This ensemble is generated by applying the forward model  $G$  to an ensemble of inputs  $h_i^f$ , drawn from a statistical distribution specified by some reasonable  $C_{hh}$  (see section 2.3.2). In the update step, one applies equation (2.3) to each member of the ensemble (each time treating  $\psi^f = \psi_i^f$ ) to obtain the posterior members  $\psi_i^a$ . The sample mean of  $\psi_i^a$  is interpreted as the posterior state estimate, and the sample covariance provides a posterior estimate of uncertainty (under a Bayesian interpretation), given by

$$C_{\psi\psi}^a = C_{\psi\psi} - C_{\psi\psi} L_a^T (L C_{vv} L^T + C_{dd})^{-1} L_a C_{\psi\psi}. \quad (2.5)$$

### 2.3.2 Prior Bathymetric Ensemble

The previous section showed that the problem estimating  $v$  and  $h$ , based on observations  $d$ , hinges on the specification of the bathymetric covariance matrix  $C_{hh}$ . In our application, we do not explicitly define  $C_{hh}$ , but instead we construct an ensemble of bathymetric realizations  $h_i$ , in such a way as to represent the spread of potential bathymetric error to the best of our knowledge.  $C_{hh}$  is then approximated by the sample covariance of that ensemble.

We assume the dominant bathymetric error, in this context, is due to integrated sediment transport between bathymetric surveys (instrument error is also present, but we have attempted to minimize its impact using loess interpolation, see section 2.2). Hence, the reasonable spread of bathymetric realizations should be constrained by measurements (survey data from the recent past/future). We must also limit the ensemble to realistic bathymetries: perturbations around the prior mean  $h$  must not include physically-unrealistic shapes or features. To that end, we seek realizations on the state space where bathymetric change naturally occurred throughout the experiment. We approximate this space by applying an empirical orthogonal functions (EOF) decomposition to the complete set of interpolated bathymetric surveys over the entire experiment. For a particular bathymetric ensemble, the prior mean loadings are set to the time-mean observed loadings from a 72-hour period surrounding the target time. The prior stan-

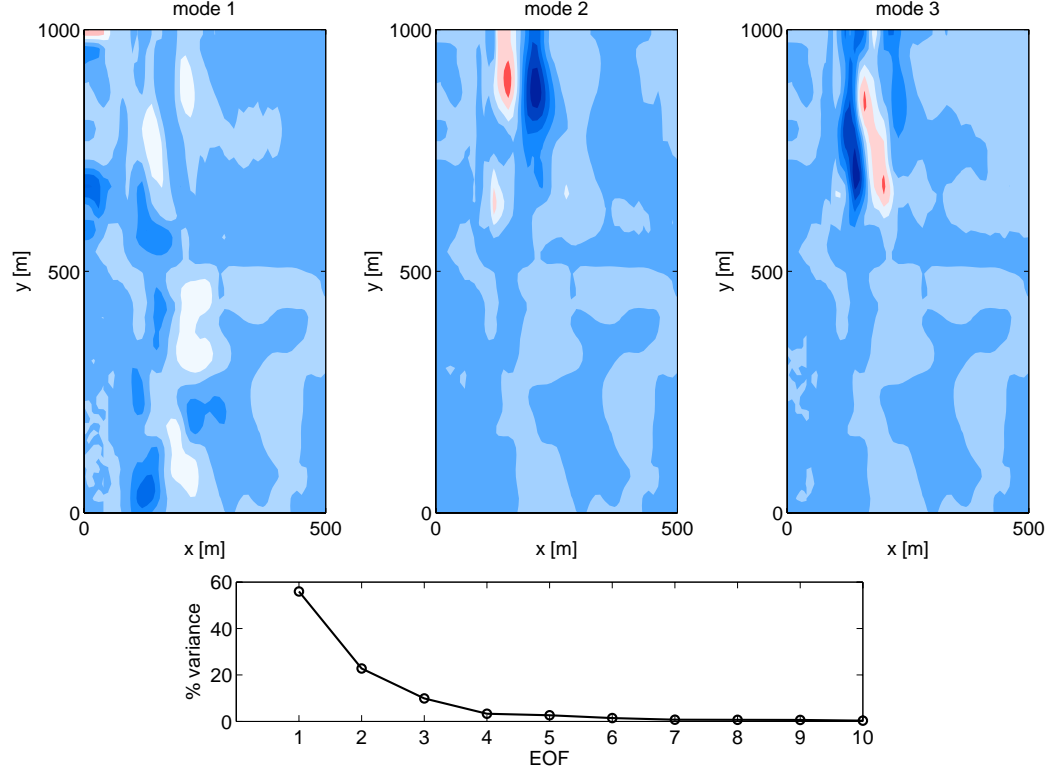


Figure 2.2: Top: leading modes in bathymetric EOF decomposition (normalized to unit variance). Bottom: percent of total variance for leading ten modes of bathymetric EOF decomposition.

dard deviation of loadings is set equal to the range of loadings observed over the same time period.

Figure 2.2 shows the leading modes of the EOF decomposition, and their corresponding percent of variance. The first mode represents full-domain surveyed change (only two full-domain surveys were conducted); subsequent modes show increasing detail mostly focused on the dynamics of the inner bar at  $x \approx 150$  m in the minigrid domain (defined as  $550 \text{ m} \leq y \leq 1000 \text{ m}$ ). For instance, the across-shore position and width of the inner bar is mainly determined by EOF modes 2 and 3.

### 2.3.3 Forward Model

Once the bathymetric ensemble is specified, we must generate the corresponding ensemble of observables  $v_i^f$ , which involves applying a forward model to each ensemble

member  $h_i^f$ . Here, we use the freely-available code **shoreCirc** (version 2.0, Svendsen *et al.* (2002)) to solve the depth-integrated and wave-averaged equations of motion for arbitrary bathymetry. These comprise the momentum equation,

$$\begin{aligned} \frac{\partial Q_\beta}{\partial t} + \frac{\partial}{\partial x_\alpha} \left( \frac{Q_\alpha Q_\beta}{(h + \eta)} \right) = & -g(h + \eta) \frac{\partial \eta}{\partial x_\beta} \\ & - \frac{1}{\rho} \left( \tau_\beta^s - \tau_\beta^b + \tau_\beta^M + \tau_\beta^{\text{Q3D}} \right) \\ & - \frac{1}{\rho} \frac{\partial S_{\alpha\beta}}{\partial x_\alpha}. \end{aligned} \quad (2.6)$$

and the conservation of mass equation,

$$\frac{\partial \eta}{\partial t} + \frac{\partial Q_\alpha}{\partial x_\alpha} = 0, \quad (2.7)$$

where  $\alpha$  and  $\beta$  are dummy indices for horizontal coordinates (summation is implied over repeated indices). In these equations,  $Q$  is the depth-integrated volume flux,  $h$  is the still water depth, and  $\eta$  is the wave-averaged water surface elevation. We define the depth-averaged across-shore and alongshore current as  $u = (Q_x - Q_{wx})/(h + \eta)$  and  $v = (Q_y - Q_{wy})/(h + \eta)$ , respectively, where  $Q_w$  is the contribution to volume flux from waves (approximated using linear wave theory).  $\tau^s$  and  $\tau^b$  are surface and bottom shear stresses,  $\tau^M$  is a non-dissipative momentum mixing,  $\tau^{\text{Q3D}}$  represents “quasi-3D” mixing (Svendsen & Putrevu, 1994), and  $S$  is the radiation stress tensor (Longuet-Higgins & Stewart, 1964). Details of the parameterizations of the various terms can be found in the **shoreCirc** manual (<http://chinacat.coastal.udel.edu/programs/nearcom>). Default values for physical constants were used throughout, except for the bottom friction coefficient  $f_w$  which was specified as 0.0053 (cf. Feddersen & Guza (2003)). We employ no-flux shoreline boundary conditions (the shoreline is defined as  $h = 0.05$  m), and radiation offshore boundary conditions. The lateral boundaries are treated as periodic, where a 300 m artificial buffer zone is added to enforce periodicity in the model inputs. The equations are solved on a mesh with  $(\Delta x, \Delta y) = (5, 10)$  m, and time-step  $\Delta t = 0.18$  s.

To define the time-independent model operator  $G$ , equations (2.6) and (2.7) are integrated from rest to steady-state. Shear instability of the alongshore current (Bowen & Holman, 1989) did not occur for the conditions tested here, except if the quasi-3D terms were turned off, similar to the results of Zhao *et al.* (2003). Hence, the steady-state flow corresponds to a single snapshot of the final model state.

To compute the radiation stress gradients due to wave motion, as well as other wave-related quantities which appear in equation (2.6) via parameterizations, we use another freely-available code, **SWAN** (Booij *et al.*, 1999). **SWAN** solves the spectral wave action-balance equation (Mei, 1983), and thus predicts the full-field wave spectral transformation. The model is initialized with measured wave frequency-directional spectra at the offshore boundary (see section 2.2). We include the effect of the wave roller, a mass of aerated water which travels on the face of breaking waves, using the formulation of Reniers & Battjes (1997) (also Reniers *et al.* (2004) and Ruessink *et al.* (2001)). We neglect interaction between the waves and the wave-averaged velocities.

The accuracy of the above model, although assumed perfect for the purposes of developing the data assimilation methodology, is in fact limited by many underlying assumptions about physical processes. In practice, it is very difficult to quantify the process model accuracy, except in very controlled laboratory conditions, because of the simultaneous presence of model input errors (the focus of the present work). An example of such a controlled validation is provided by Haas *et al.* (2003), who applied **shoreCirc** to simulate a laboratory rip current flow. In that study, **shoreCirc** was shown to reproduce the broad features of the 2DH flow, while smaller-scale flow details were shown to be influenced by errors in the bathymetric input. They also found the accuracy of **shoreCirc** to be comparable to that of a wave-resolving Boussinesq model.

The present model is also known to be as accurate as other available 2DH numerical codes under comparable field conditions. Wilson (2009) performed a validation of the present model for 455 hours of the SD97 field experiment, and found rms errors of order 5–15 cm for  $H_{mo}$  and 10–20 cm/s for  $v$  (larger errors occurring closer to shore). Similar values have been reported by Ruessink *et al.* (2001) (for a 1DH model), and Morris (2001) and Hsu *et al.* (2006) (for 2DH models), among others. Hence, the present model setup is considered representative of the state-of-the-art for depth-integrated wave-averaged nearshore prediction.

### 2.3.4 Observational Error Estimates

We assume observation error standard deviations of 6.7 cm/s in alongshore current, and 7.0 cm in significant wave height. These values encompass errors of measurement (instrument noise), as well as so-called representation errors (the two are added in quadrature). The former are due to practical issues of data collection and quality, while the latter are due to the fact that what the model predicts is not strictly comparable to what is

measured.

Measurement error standard deviation for  $v$  has been estimated using laboratory and field calibration (Feddersen & Guza, 2003) as  $(\sigma_0^2 + (\alpha|v|)^2)^{1/2}$ , where  $\sigma_0 = 5$  cm/s, and  $\alpha = 0.05$ . For the present case,  $|v| \sim 0.75$  m/s, and we therefore assume a measurement error standard deviation of 6.25 cm/s for  $v$ .

To obtain an estimate of measurement error for  $H_{\text{mo}}$  during SD97, we have compared measurement differences for sensors placed less than 4.5 m apart in the alongshore direction and less than 0.55 m apart in the across-shore direction (four sensor pairs passed this criteria, located from  $x = 210$  m to  $x = 261$  m). The standard deviation of measurement differences, based on over 2500 hours of data, ranged from 3.6 cm to 6.3 cm. Values were increased for sensors closer to shore, and for increasing offshore wave height. Hence we assume a (conservative) measurement error standard deviation of 6.5 cm for  $H_{\text{mo}}$ .

Several potential sources of representation error exist in the present model. One example is the fact that the measurements were collected at a particular water depth, whereas the model predicts depth-averaged flow. Further, the measurements may have been sampled during slowly-varying conditions (such that time averaging of observational data does not suffice to remove the variability), or may even be unsteady (Bowen & Holman, 1989), whereas the model predicts the steady-state waves and flow which would occur under static conditions. The treatment of representation error is not trivial, and is the subject of ongoing research (e.g. Oke & Sakov (2007), and references therein). Here, we simply assume a constant, spatially-uniform contribution to the observational error, of 2.5 cm/s for  $v$  and 2.5 cm for  $H_{\text{mo}}$ . We have tested different values of total observational error, and find no qualitative change in the posterior solution.

### 2.3.5 Underlying Assumptions and Optimality

Several assumptions underlie the derivation of equation (2.3), which should be kept in mind when applying the method. Importantly, we have assumed that the model is “perfect”, in the sense that if the true inputs  $h^t$  were known, the output  $v^t = G(h^t)$  would be exact. This assumption pertains to physical processes in the model, as well as boundary conditions: for the present application, the latter is important because the offshore boundary condition (an input wave spectrum) is itself derived from measurements. While this effect may be reduced at locations far from the boundary (i.e. the inner surf zone), it is likely not negligible.

Another factor in the interpretation of equation (2.3) as an “optimal” solution is the

assumption of Gaussian statistics. In our case, the model operator  $G$  is nonlinear, hence the statistics are not likely to be Gaussian. Therefore, we will avoid the use of the term “optimal” in describing the posterior estimates. Instead, we interpret the results as a least-squares estimate, based on approximate model statistics.

Finally, the quality of the posterior estimate is conditioned by the quality of the prior statistics. Specifically, one must define an appropriate prior mean/covariance for the bathymetry, and a reasonable error model for the observations. The present results are based on rational and well-defined estimates of those statistics, as described above, but these estimates are still subjective, to some degree. In practice, we have found the quality of the posterior to be degraded if the prior statistics are not carefully defined, and this may be unavoidable in the absence of extensive observational data. An attractive extension of the present method would be to include time-evolution in the ensemble statistics, as in the sequential method of van Dongeren *et al.* (2008). Using that approach, the prior statistics are only specified once, and are continually updated whenever measurements become available (using equations (2.3) and (2.5); this is the ensemble Kalman filter, Evensen (2006)). Such an extension is highly recommended for future application of the present method.

The above caveats underscore the need for cross-validation when applying the present method. Therefore, in the following, we first show the applicability and skill of the method before using it to assess the sensitivity of modeled circulation to bathymetric uncertainty.

## 2.4 Results

In this section, we demonstrate the ability of the statistical inverse method to estimate bathymetry, in a situation where it was not possible to conduct a bathymetric survey. Our primary example case is for 1530 EST on 20 October, for which the dynamics are 2DH. This time was selected due to its interesting morphodynamic setting, the presence of strongly-2DH flow features, and a low rate of instrument malfunctions in shallow water. Before moving to this more-complex 2DH case, however, we will present a 1DH case study from the same day, at 1130 EST. The 1DH case will serve to introduce important conceptual topics related to the assimilation methodology.



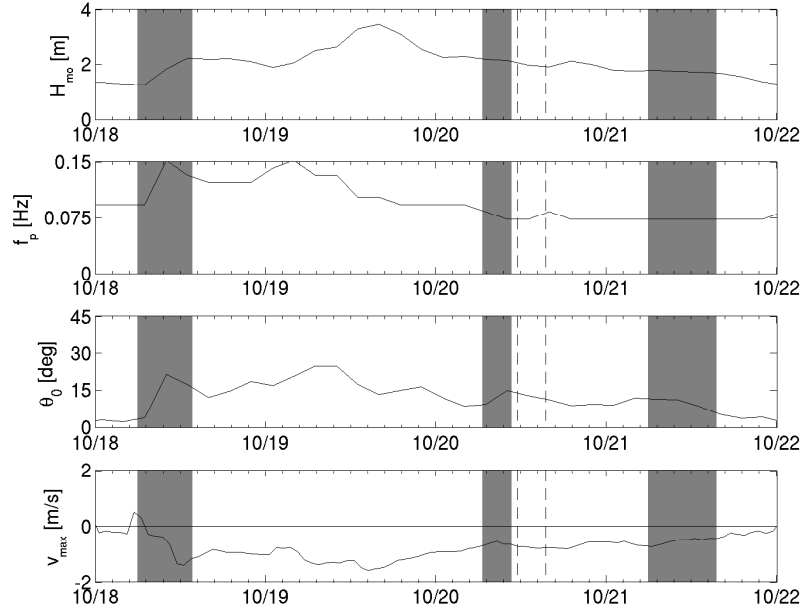


Figure 2.3: Conditions observed for the period from 18 October to 21 October during the SD97 experiment: offshore significant wave height  $H_{mo}$ , peak frequency  $f_p$ , mean wave angle  $\theta_0$  (Kuik *et al.*, 1988) (positive is from the north), and maximum alongshore current in SPUV array  $v_{max}$  (positive is towards the north). Shaded regions represent times when bathymetric surveys were being performed (see Figure 2.4). Dashed lines correspond to 20 October, 1130 EST and 1530 EST (sections 2.4.2 and 2.4.3).

### 2.4.1 Physical Setting

The conditions surrounding 20 October were strongly influenced by the passage of a Nor’easter storm, which peaked during the hours 1600–1900 EST on 19 October. The measured significant wave height (at 8 m depth) during the storm was 3.4 m. Somewhat less-energetic conditions continued throughout the day on 20 October. Figure 2.3 summarizes the observed conditions.

No bathymetric survey was conducted on 19 October due to dangerous conditions. Complete minigrid surveys were conducted, however, on 18 October, 0600–1340 EST, and 21 October, 0550–1530 EST. Also, a limited survey was conducted on 20 October, 0630–1040 EST. Figure 2.4 summarizes these bathymetric observations. The sequence of surveyed bathymetries illustrates the speed with which bathymetric change occurred in the days surrounding the storm. The surveyed transects suggest changes in across-

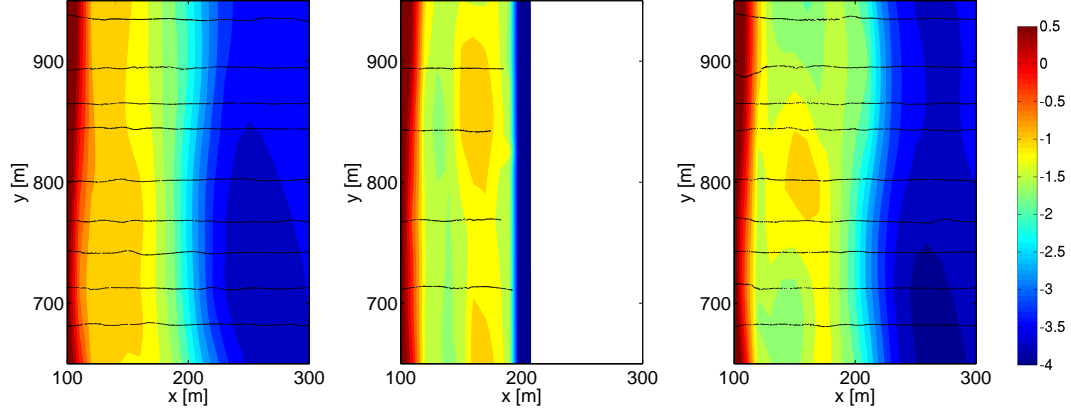


Figure 2.4: Black: bathymetric transects collected by CRAB on (from left to right) 18, 20, and 21 October. Colors: interpolated surveyed bathymetry  $z_b$  (hotter colors represent shallower water).

shore bar profile, alongshore variability, or both (the exact morphodynamics may not be resolved by the surveys). This rapid bathymetric change, combined with the paucity of survey data on 19 and 20 October, makes specification of model bathymetry quite difficult for the target model times, which can be between survey times. Temporal interpolation from surveys would be a questionable approach, as the surveys, as well as the sonar altimeters, suggest a non-monotonic change through time. Hence the present method has practical relevance, because it makes use of additional time-resolved measurements (wave height, velocity) to improve the bathymetric estimate.

The observed flow during the 19 October storm was alongshore uniform (1DH) and reached speeds of up to 1.6 m/s. Over the course of the day of 20 October, conditions changed such that the observed flow was weaker, and exhibited alongshore-nonuniformity (2DH). Indeed, as we will see in later sections, assimilation of data on 20 October, 1530 EST, leads to a 2DH posterior model state. However, at earlier times on 20 October, particularly at high tides, the observed (and posterior estimated) flow was closer to 1DH. Next, we study such a case, 1130 EST, as a simple dynamical setting in which to introduce the present method.

#### 2.4.2 Conceptual Interpretation: 1DH Case Study

In this section, we run the forward model assuming  $\partial/\partial y = 0$  in the governing equations, using as the bathymetry a single transect  $y = 828$  m from the 2DH bathymetric ensemble

(observational data are taken from the same transect). The assumption of 1DH dynamics is only approximately valid, here. For instance,  $v$  measured on the transect  $x = 160$  m varied from -45 cm/s ( $y = 704$  m) to -70 cm/s ( $y = 816$  m) (no other sensors were functioning at the time on  $x = 160$  m). On the transect  $x = 210$  m (five sensors),  $v$  varied from -68 cm/s ( $y = 906$  m) to -50 cm/s ( $y = 816$  m), with mean -59 cm/s and standard deviation 7.1 cm/s. However, neither the measurements nor a 2DH assimilation indicated any strongly-2DH features such as rip currents. Moreover, our purpose in this section is to elucidate the mechanics of the assimilation in the context of simple 1DH model dynamics. Cross-validation using more-accurate 2DH dynamics will be taken up in later sections.

Figure 2.5 shows the prior and posterior predictions of bathymetry, velocity, and wave height, compared to measurements. The prior bathymetry does not include a sharp nearshore bar, as was measured by the sonar altimeters, and confirmed by a nearby CRAB survey transect. Hence the prior alongshore current jet is too broad, causing  $v$  to be overpredicted at the innermost sensor. Similarly, the offshore face of the inner bar is too shallow in the prior, causing increased wave breaking and hence underprediction of  $H_{mo}$  at nearby locations (e.g. compare wave transformation from  $x = 210$  m to  $x = 185$  m). After assimilating data, the above errors are reduced and the overall fit is improved, including the fit to  $h$  (which was not assimilated).

In order to understand how equation (2.3) used the observed model-data misfit ( $d - Lv^f$ ) to update the full model state  $\psi$ , it is useful to examine the coefficient matrix  $r = C_{\psi\psi}L_a^T$ . In the language of data assimilation,  $r$  is usually referred to as the matrix of representers. Each column  $r_k$  quantifies the sensitivity of the model to a particular observation (the  $k$ 'th observation). Hence by analyzing these columns, suitably normalized, we can better understand how the overall model corrections are assembled (Kurapov *et al.*, 2009). Here, we will normalize as follows,

$$\hat{r} = r \frac{\sigma_m + \sigma_d}{\sigma_m^2 + \sigma_d^2} . \quad (2.8)$$

This normalization is obtained by taking model and data errors to be equal to their standard deviations (denoted  $\sigma_m$  and  $\sigma_d$ , respectively), and assuming an isolated observation. We will interpret maps of  $\hat{r}$  as the “potential correction” which can be obtained by assimilating a particular measurement. For instance, we will write  $\hat{r}_{hv}$  as the potential correction to bathymetry  $h$  from assimilating velocity  $v$ . In this way, we may compare the magnitude and spatial pattern of the correction derived from each element

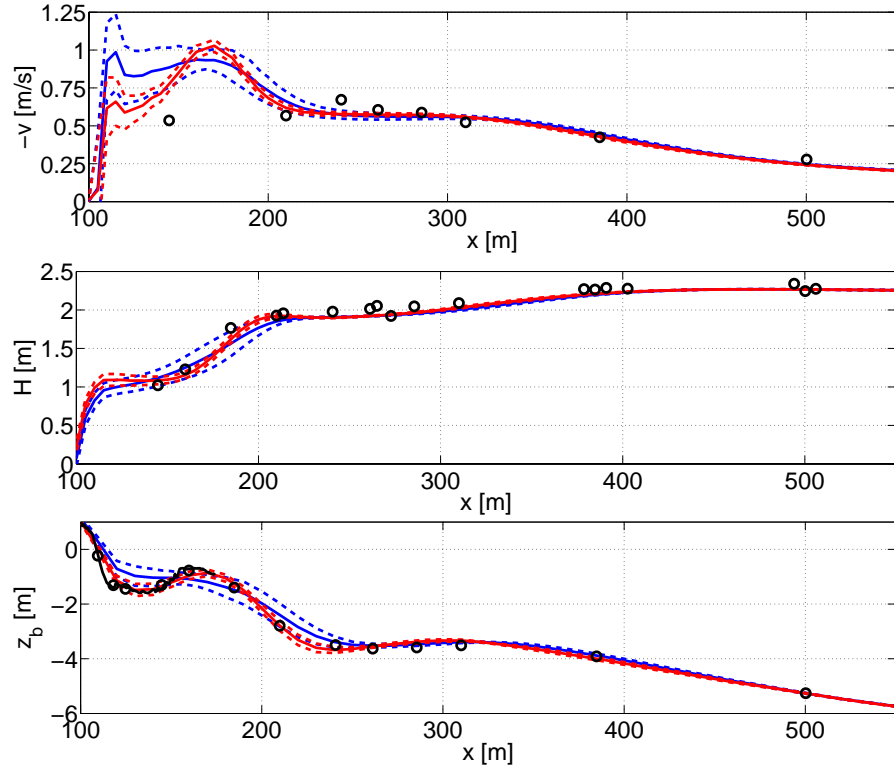


Figure 2.5: Prior (blue), posterior (red), and measured (black) data for 1DH case study, 20 October 1130 EST. Dashed lines represent  $\pm$  one standard deviation. Black circles correspond to data from fixed instruments (SPUV); black line is CRAB transect collected at  $y = 843$  m, on 20 October, 1036–1040 EST. Still water level was 1.16 m NGVD.

in the measurement array. Note this analysis is performed without reference to the actual measurements;  $r$  is a property of the model (and prior statistics) only. We have also compared equation (2.8) to the contributions of actual measurements to the overall model update in equation (2.3), and generally find good agreement.

Figure 2.6 shows scaled representers for each instrument on the observational transect (including instruments which were not functioning at 1130 EST). Both observation types show the largest potential corrections coming from sensors in the inner surf zone ( $x < 250$  m). Potential corrections from outer surf zone sensors are small, because of small prior uncertainty (cf. Figure 2.5) and/or lack of bathymetric sensitivity at those locations. Hereafter, we will concentrate on the inner surf zone.

Focusing first on  $\hat{r}_{hH}$ , underprediction of  $H_{\text{mo}}$  typically resulted in a local increase (deepening) of  $h$ ; that is,  $\hat{r}_{hH}$  is locally-positive. This agrees with the expectation based on saturated depth-limited breaking,  $H_{\text{mo}} = \gamma h$ , with  $\gamma > 0$  (i.e.  $\partial H_{\text{mo}}/\partial h > 0$ ). To be more precise, consider the following local approximation of  $C_{Hh}$  near the point  $x_0$ :

$$\begin{aligned} C_{Hh} &= E[\delta H_{\text{mo}} \delta h] \\ &\approx E \left[ \left. \frac{\delta H_{\text{mo}}}{\delta h} \right|_{x_0} \delta h \delta h \right] \\ &= \left. \frac{\delta H_{\text{mo}}}{\delta h} \right|_{x_0} C_{hh}, \end{aligned} \tag{2.9}$$

where  $E$  is expected value, and  $\delta H_{\text{mo}}/\delta h$  denotes the relative increment of  $H_{\text{mo}}$  for a given increment of  $h$ , evaluated based on the prior statistics at a given point. As suggested above, the prior statistics for the present case indeed gave  $\delta H_{\text{mo}}/\delta h > 0$  for locations where waves were breaking; in fact,  $\delta H_{\text{mo}}/\delta h$  was significantly correlated with the prior wave dissipation ( $r = 0.58$ ,  $p = 10^{-10}$ ). Positive  $\delta H_{\text{mo}}/\delta h$  occurred in the inner surf zone where dissipation was large, while small (or even negative)  $\delta H_{\text{mo}}/\delta h$  occurred outside the surf zone and in reshoring regions. At locations near the maxima of wave dissipation,  $\delta H_{\text{mo}}/\delta h$  had a value of approximately 0.5.

Turning next to  $\hat{r}_{hv}$ , we note that an underprediction of  $-v$  in the prior (i.e. prior predicted current not as large as observed current towards the south, or negative  $y$ , direction) always produces a local decrease (shoaling) of  $h$ ; that is,  $-\hat{r}_{hv}$  is locally-negative. Conceptually, this behavior is due to the fact that local maxima in  $|v|$  tend to be associated with local minima in depth (e.g. a sand bar). Indeed, the prior statistics for the present case gave  $\delta(-v)/\delta h < 0$  for all  $x > 105$  m.

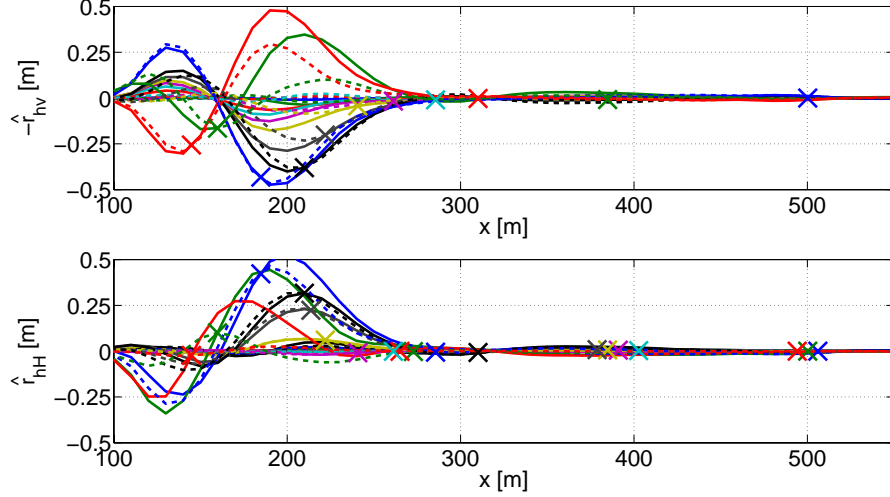


Figure 2.6: Solid lines: scaled representers. Dashed lines: corresponding approximation based on extrapolation with  $C_{hh}$  (see text). Crosses indicate location of measurement for each representer.  $\hat{r}_{hv}$  has been negated for comparison with Figure 2.5.

The above interpretation highlights the role of the model dynamics for determining the local values of  $\delta H_{mo}/\delta h$  and  $\delta v/\delta h$ , and hence the magnitudes of the representers themselves. Non-local corrections to bathymetry, on the other hand, are derived from a combination of model dynamics and the assumed prior covariance  $C_{hh}$ . In order to judge the balance between these two contributions, we may compare  $\hat{r}$  to the approximation based on (2.9) (dashed lines in Figure 2.6), which is representative of the contribution of  $C_{hh}$  to the non-local correction. Clearly,  $C_{hh}$  plays an important role in determining the basic structure of  $\hat{r}$ , and hence the corrections themselves, while model dynamics mainly act to amplify and/or shift that structure. This highlights the importance of choosing appropriate prior statistics, if non-local corrections are to be trusted.

Finally, we note there are qualitative differences between the shape and magnitude of  $\hat{r}_{hv}$  and  $\hat{r}_{hH}$ , suggesting  $v$  and  $H_{mo}$  play different roles in the overall correction. To further illustrate this fact, Figure 2.7 shows the posterior model state when each observation type (either  $v$  or  $H_{mo}$ ) is assimilated individually. The results are best understood by considering differences in the posterior  $H_{mo}$ . When assimilating  $v$  only, the spatial gradient of  $H_{mo}$  is altered in the inner surf zone, resulting in a correction to the wave-induced forcing (not plotted); however, the resulting  $H_{mo}$  is not in good agreement with observations (in particular, the posterior breakpoint is too far offshore). Conversely, when assimilating  $H_{mo}$  only, the magnitude of  $H_{mo}$  is improved but the

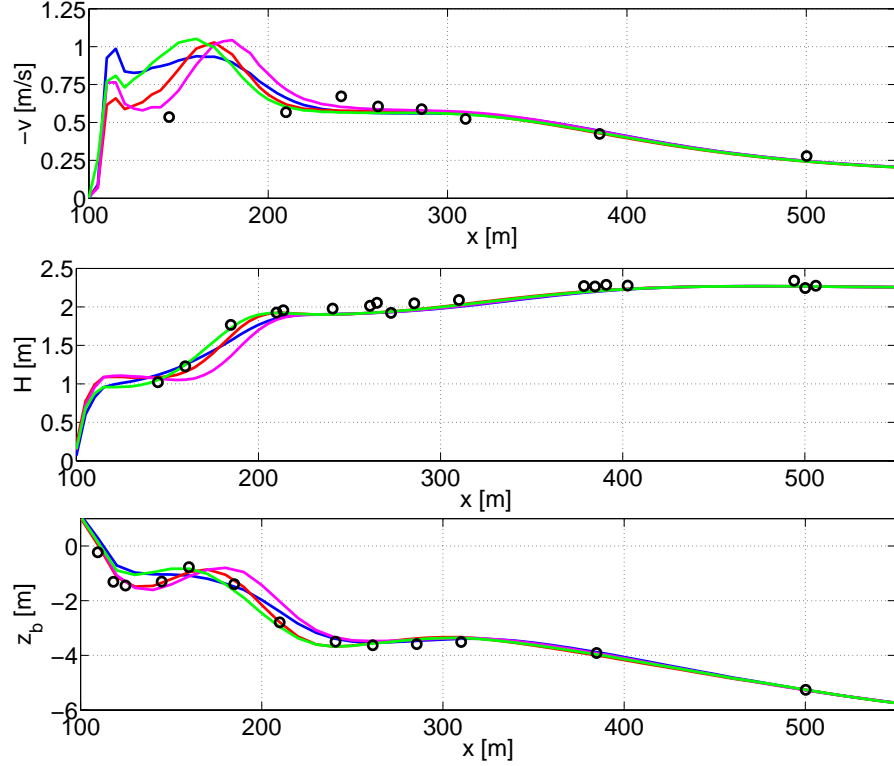


Figure 2.7: Prior (blue), measured (black) and posterior data when assimilating  $H_{mo}$  only (green),  $v$  only (magenta), or both (red).

resulting change in wave-induced forcing does not lead to an improved  $v$  (particularly at the innermost sensor). Assimilating  $v$  and  $H_{mo}$  together allows the forcing to be corrected without severely affecting the accuracy of  $H_{mo}$ , resulting in an improved overall agreement for all variables (also see section 2.4.4). The representers (Figure 2.8) confirm the above interpretation: for the most-shoreward sensors,  $\hat{r}_{HH}$  and  $\hat{r}_{Hv}$  indicate controls on magnitude and gradient, respectively. That is, the most-shoreward observation points correspond to anti-nodes of  $\hat{r}_{HH}$ , and nodes of  $\hat{r}_{Hv}$ . These sensors dominate the overall correction, producing the above behavior. Further-offshore sensors do not show such a clear contrast between  $\hat{r}_{HH}$  and  $\hat{r}_{Hv}$ , likely due to differences in the qualitative dynamics (e.g. the influence of momentum mixing).

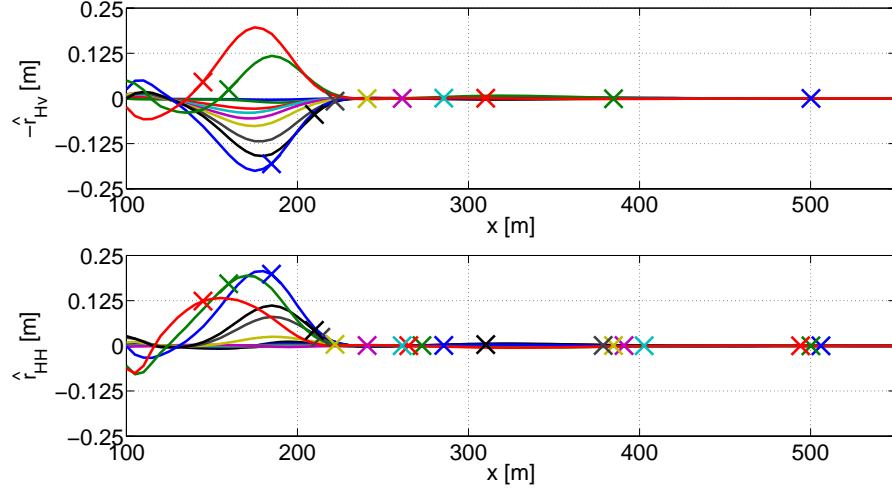


Figure 2.8: Scaled representers for correction of  $H_{\text{mo}}$ . Crosses indicate location of measurement for each representer.  $\hat{r}_{Hv}$  has been negated for comparison with Figure 2.5.

### 2.4.3 Assimilation During 2DH Flow

Having established the conceptual framework for assimilation under simple 1DH dynamics, we now move to a 2DH case: 20 October, 1530 EST.

Figure 2.9 shows the prior prediction of  $v$  and  $H_{\text{mo}}$ . Recall this corresponds essentially to a forward model run, with bathymetry derived from a smoothed interpolation in EOF space (section 2.3.2). The predictions in the outer surf zone are fairly consistent with observations (overprediction of offshore wave height is likely due to error in the offshore boundary conditions). For sensors in the inner surf zone,  $x < 250$  m, however, the flow becomes alongshore-nonuniform, and the velocity predictions are highly inaccurate in magnitude and even wrong in direction (see sensors at  $y \approx 830$  m and  $y \approx 700$  m). Given the known bathymetric sampling issues (see section 2.4.1), we will now explore the possibility that the model error is due to mis-specification of  $h$ .

Figure 2.10 compares the posterior velocity field to observations, after assimilating  $H_{\text{mo}}$  (46 observations) and  $v$  (29 observations). Model-data agreement in alongshore current is improved, particularly in the inner surf zone, which is to be expected because that data was assimilated. Importantly, the modeled across-shore currents  $u$  are also improved, which can be interpreted as a cross-validation for the update step ( $u$  was not assimilated).

The posterior bathymetry is also an improvement over the prior. Figure 2.11 shows an across-shore transect comparing the prior and posterior  $h$  to that measured by the



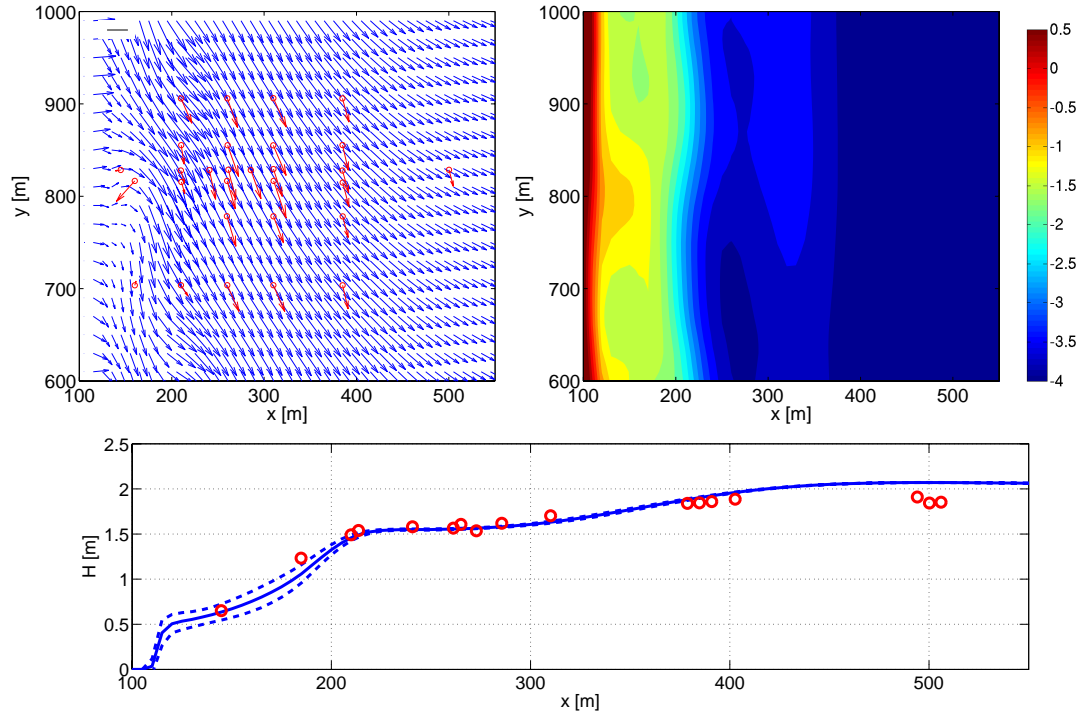


Figure 2.9: Prior mean velocity, bathymetry, and wave height on transect  $y = 828$  m, for 20 October, 1530 EST. Red arrows in left-hand plot are observed velocity; blue arrows are modeled velocity, plotted at even gridpoints; scale arrow in upper left is 50 cm/s. Colors in right-hand plot are model  $z_b$  (still water level was 0.17 m NGVD). Solid and dashed lines in the lowermost plot are prior mean and standard deviation of  $H_{mo}$ , and red circles are measured  $H_{mo}$ .

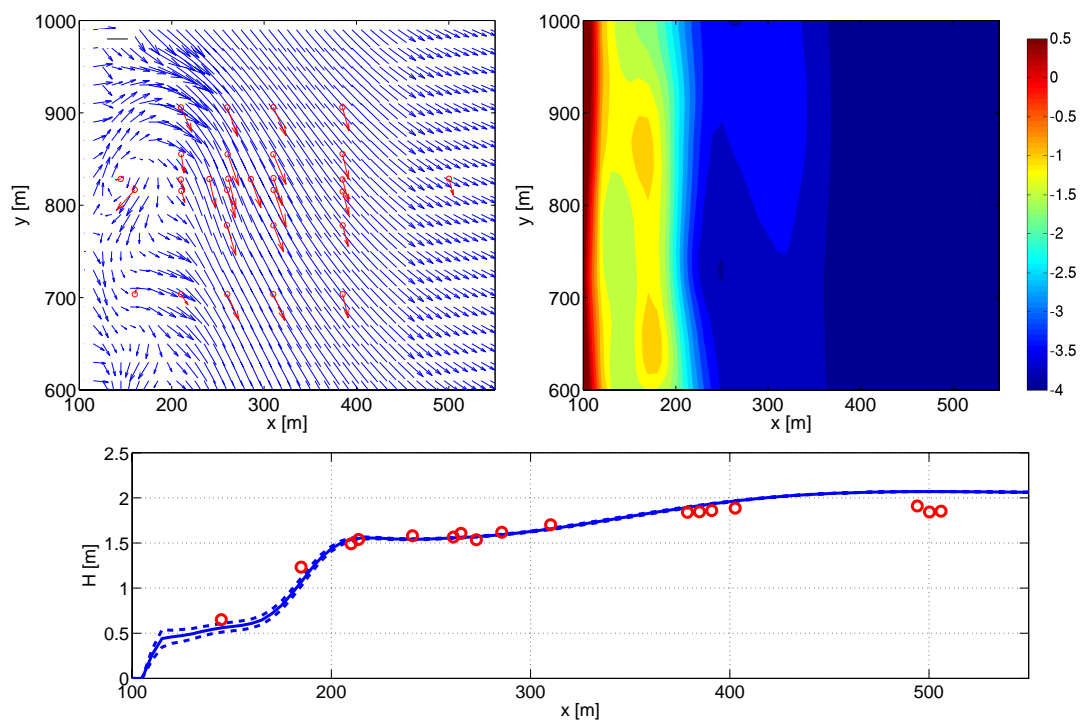


Figure 2.10: As in Figure 2.9, but for posterior (updated) fields.

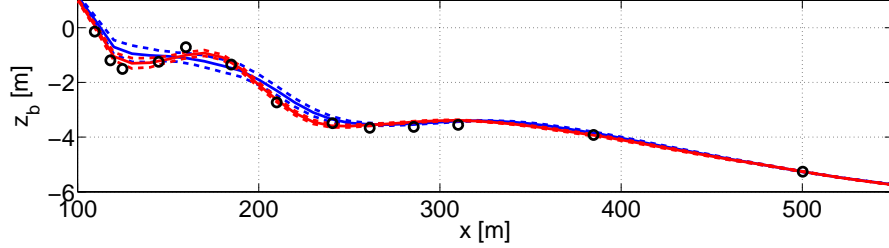


Figure 2.11: Across-shore transect ( $y = 828$  m) of prior (blue), posterior (red), and measured (black)  $z_b$  (still water level was 0.17 m NGVD). Dashed lines represent  $\pm$  one standard deviation.

sonar altimeters (again, these measurements were not assimilated). The comparison shows that the update step correctly adjusted the prior in the direction of the actual (measured) bathymetry. As in the 1DH case study, major corrections occurred in the inner surf zone, whereas outer surf zone ( $x > 250$  m) corrections were relatively small.

Next, as in the 1DH case study, we examine the scaled representers  $\hat{r}$  for 2DH flow. Figure 2.12 (top plots) shows maps from  $\hat{r}_{hH}$ , the potential correction to  $h$  from assimilating measurements of  $H_{\text{mo}}$ , at four different locations ranging from the inner to outer surf zone. Clearly,  $H_{\text{mo}}$  is effective for constraining local bathymetry in the inner surf zone, where wave height is strongly controlled by water depth because of depth-limited wave breaking. For measurements in the outer surf zone,  $\hat{r}_{hH}$  is much smaller in magnitude, indicating a smaller potential for correction in that region.

Figure 2.12 (bottom plots) shows maps from  $\hat{r}_{hv}$ , the potential corrections to  $h$  from observing  $v$ . A distinguishing feature of  $\hat{r}_{hv}$  in this case is that sensors offshore of the inner bar contribute information about bathymetry onshore of the inner bar. This did not occur under 1DH dynamics, and hence is attributed to advection by 2DH currents. This also means a greater number of sensors for  $v$  could provide significant (potential) corrections to  $h$ : 19 columns of  $\hat{r}_{hv}$  had maximum magnitude exceeding 10 cm, as opposed to only six for  $\hat{r}_{hH}$ . The actual model corrections when assimilating data reflect the same trend: five observations of  $v$  contributed corrections of more than 10 cm to  $h$ , compared to two for  $H_{\text{mo}}$ . Thus, the  $v$  array included a greater number of “useful” sensors. The corresponding correction to  $h$  should be considered more stable, in the sense that it is more robust against isolated observation errors dominating the overall correction. A more rigorous way to examine the stability of the measurement array is to compute the singular value decomposition of the matrices  $Lr_{hv}$  and  $Lr_{hH}$  (so-called “array-mode analysis”, Bennett (2002); Kurapov *et al.* (2009)). Five of the singular values for  $Lr_{hv}$

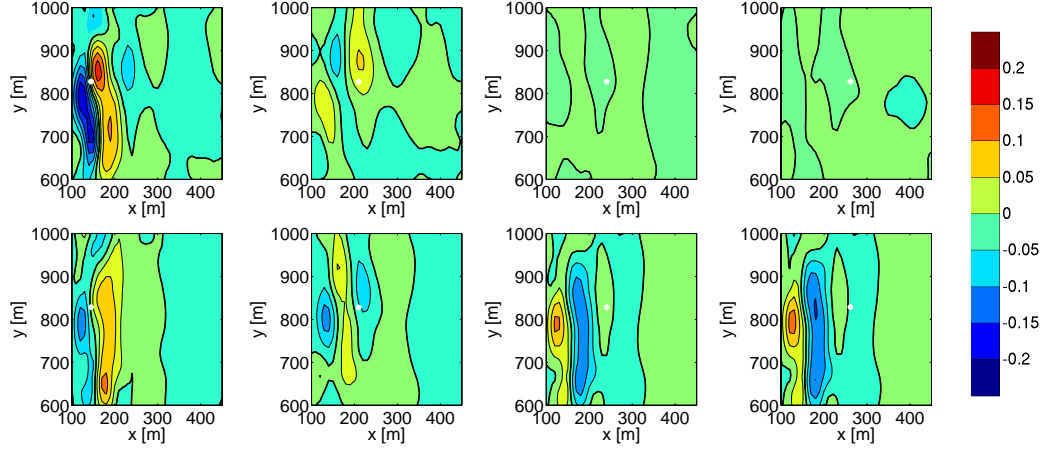


Figure 2.12: Maps from scaled representer sub-matrix  $\hat{r}_{hH}$  (top) and  $\hat{r}_{hv}$  (bottom), for select measurement locations (white dots). Thick black contour is 0 cm, and subsequent contours are plotted at 5 cm intervals.

exceeded the observational noise level, indicating stable array-modes; two singular values for  $Lr_{hH}$  passed the same criteria. Thus, again,  $v$  was the more stable observation type for the present case.

#### 2.4.4 Skill Statistics

Next, we evaluate the skill of the posterior model state, compared to that of the prior. Our aim is to quantify the improvement in the model state when assimilating observations of  $H_{mo}$  and  $v$ , together as well as individually. Here, we will limit our discussion to the inner surf zone region  $x < 250$  m, for two reasons: first, as noted earlier, corrections were broadly confined to this region, whereas the outer surf zone was constrained by low prior uncertainty; second, the point  $x = 250$  m corresponds to a minimum in modeled wave dissipation, separating two distinct wave breaking regions and flow regimes. For completeness, skill statistics for  $x > 250$  m are listed in Appendix B, Table B.1. Indeed, the model updates in that region were small, for the reasons stated above, and often did not result in an improved fit to cross-validation variables.

The skill of the prior and posterior model states will be assessed in a probabilistic validation framework (Casati *et al.*, 2008), taking into account the predicted model state  $\psi$  as well as the predicted uncertainty  $C_{\psi\psi}$ . Specifically, we adopt the Continuous Ranked Probability Score (CRPS; appendix A) to assess skill, which measures the overall

			1DH (1130 EST)			2DH (1530 EST)		
variable(s) assimilated	variable updated	units	$\epsilon$	$CRPS$	$S$	$\epsilon$	$CRPS$	$S$
none (prior)	$u$	m/s	0.13	0.13	-	0.33	0.68	-
	$v$	m/s	0.22	0.24	-	0.29	0.56	-
	$H_{\text{mo}}$	m	0.11	0.11	-	0.11	0.18	-
	$h$	m	0.45	0.75	-	0.41	0.88	-
$H_{\text{mo}}, v$	$u$	m/s	0.13	0.16	-0.17	0.19	0.36	0.48
	$v$	m/s	0.15	0.18	0.27	0.054	0.055	0.90
	$H_{\text{mo}}$	m	0.080	0.092	0.16	0.084	0.15	0.18
	$h$	m	0.20	0.28	0.63	0.18	0.30	0.66
$H_{\text{mo}}$	$u$	m/s	0.11	0.12	0.11	0.30	0.62	0.093
	$v$	m/s	0.26	0.35	-0.47	0.28	0.55	0.013
	$H_{\text{mo}}$	m	0.041	0.033	0.70	0.064	0.10	0.46
	$h$	m	0.36	0.61	0.18	0.37	0.83	0.061
$v$	$u$	m/s	0.17	0.21	-0.54	0.20	0.36	0.48
	$v$	m/s	0.12	0.11	0.52	0.054	0.053	0.90
	$H_{\text{mo}}$	m	0.19	0.30	-1.7	0.13	0.25	-0.34
	$h$	m	0.42	0.74	0.016	0.20	0.32	0.64

Table 2.1: Model accuracy statistics before and after assimilation, for sensors in  $x < 250$  m.  $\epsilon$  is rms difference between model and observations, CRPS is the Continuous Ranked Probability score (see text, and appendix A), and  $S$  is a skill score (equation (2.10)). The calculation of CRPS for  $u$  and  $h$  assume observational error standard deviations of 6.7 cm/s (same as for  $v$ ) and 10 cm, respectively.

difference between prediction and observation probability density functions. In the limit of deterministic predictions and observations, CRPS is equal to the root-mean-square (rms) error. We also compute a skill score,

$$S = 1 - (CRPS)_a / (CRPS)_f, \quad (2.10)$$

which indicates whether the posterior state (subscript  $a$ ) has improved skill relative to the prior (subscript  $f$ ). Finally, for completeness, we also report the rms error, as a simple and easy to understand measure of accuracy which does not take into account the predicted uncertainty. These statistics are given in Table 2.1, and the results are discussed next.

## 2.5 Discussion

### 2.5.1 Assimilation Skill

Table 2.1 reports statistics which assess the improvement in model skill when assimilating different combinations of data. In general, if the present methodology is skillful, we should find a decrease in rms error and CRPS (i.e. positive skill score  $S$ ) as a result of assimilating data. When this is not the case, we will generally assume the inversion is converging (with respect to increasing number of observations) on an incorrect posterior state  $\psi$ . In this section, we ask: what data were required to obtain a skillful inverse, in the above sense?

First, we consider the case where both  $v$  and  $H_{\text{mo}}$  are assimilated, under 2DH dynamics (20 October, 1530 EST). The resulting posterior state is improved in all variables, including the cross-validation variables  $u$  and  $h$ . This indicates the assimilation of data has introduced a realistic correction to the overall model state. The error that remains in the posterior estimate represents a combination of still-unresolved uncertainty in the input  $h$ , observational/representation errors, and (importantly) errors due to model physics.

When only  $v$  is assimilated (2DH case), the situation is quite different. Both  $u$  and  $v$  are brought into good agreement with the observations, and CRPS indicates positive skill. The prediction for  $h$  is also improved, although not to the extent as when  $H_{\text{mo}}$  and  $v$  were assimilated together. However, the posterior  $H_{\text{mo}}$  is actually less accurate than the prior, and has larger CRPS, indicating the assimilation is not converging towards the true  $H_{\text{mo}}$ . Overall, cross-validation suggests the assimilation is overfitting the velocity data, at the expense of  $H_{\text{mo}}$  (we define overfitting, here, as achieving improved skill in one variable, at the expense of any other variable).

When only  $H_{\text{mo}}$  is assimilated (2DH case), we find a similar result to when only  $v$  was assimilated. The posterior state is an improved fit to the assimilated variable, but not to the unassimilated variables (in this case, the skill for  $u$ ,  $v$  and  $h$  is essentially unchanged). Again, the result may be converging on an incorrect posterior model state.

Statistics from the 1DH case (20 October, 1130 EST) indicate similar results, with overfitting occurring unless  $H_{\text{mo}}$  and  $v$  are assimilated together. One distinction between the 1DH and 2DH cases lies in their ability to correct the across-shore current  $u$ . In the 2DH case,  $u$  was improved when  $v$  was assimilated, but made worse when  $H_{\text{mo}}$  was assimilated; the opposite was true for the 1DH case. This is perhaps not surprising: under 1DH dynamics,  $u$  is entirely due to the below-trough return flow of wave volume flux, which is in turn directly related to  $H_{\text{mo}}$ . In contrast, in the 2DH case, the cross-

shore current is likely driven as a result of non-local alongshore-nonuniform dynamics. Another distinction between the 1DH and 2DH results is that, in the 1DH case,  $h$  could be better corrected by assimilation of  $H_{\text{mo}}$  than by assimilation of  $v$ . We note, however, that the 1DH case had only three active sensors for  $v$  in the inner surf zone, compared to six active sensors for  $H_{\text{mo}}$ .

A result common to both 1DH and 2DH cases is that the true ocean state  $\psi$  can only be recovered by assimilating both variables  $v$  and  $H_{\text{mo}}$ . This may be partly attributed to the fact  $v$  and  $H_{\text{mo}}$  provide different (complimentary) information with regard to the dynamics, as demonstrated using the 1DH model in section 2.4.2. However, other factors may serve to exacerbate the overfitting behavior. For instance, we have already noted that the specified prior  $C_{hh}$  influences the shape of representers, and hence the correction itself; errors in this specification could lead to unexpected results. Errors may also exist in the forward model, causing the true  $v$  and  $H_{\text{mo}}$  to be incompatible under the “perfect-model” assumption; Plant *et al.* (2009) find an analogous result, where artificial smoothing of bathymetry leads to decreased error in  $H_{\text{mo}}$  but increased error in  $v$ . One way to exclude the influence of the above effects is to extract synthetic observations from a forward model run with idealized bathymetry. From that experiment, we find a similar result as above: assimilating  $v$  appears to correct the gradient of  $H_{\text{mo}}$ , and hence the wave-induced forcing, while the magnitude of  $H_{\text{mo}}$  is not improved (and vice-versa). However, the synthetic tests do not show strongly-negative skill in the unassimilated variable as was the case with real observations. Hence, we cannot not rule out the possibility of model error being present.

To summarize, the ensemble-based method was successful in assimilating observations and correcting bathymetry, when using all of the available data for  $v$  and  $H_{\text{mo}}$ . The resulting prediction is an improvement over the prior (rms error and CRPS are both decreased). When one observation type (either  $v$  or  $H_{\text{mo}}$ ) was withheld, bathymetry could still be improved relative to the prior, but only at the expense of a poor posterior prediction of the unassimilated variable (either  $H_{\text{mo}}$  or  $v$ ). This is explained by considering the complimentary information carried by each variable, although other factors are considered. In any case, there is an inherent benefit of assimilating (semi)independent data types.

### 2.5.2 Bathymetric Input Sensitivity

In sections 2.4.2 and 2.4.3, representers were used to clarify how prior error/sensitivity is utilized for the assimilation of data: by combining all of the representers (with appropriate normalization), one obtains the posterior model state. The magnitude of the potential correction fields  $\hat{r}$  indicate strong model sensitivity between the observed variables  $v$  and  $H_{\text{mo}}$  and the target variable  $h$ .

A closely-related problem is the extent to which  $v$  and  $H_{\text{mo}}$  are sensitive to errors in  $h$ . A direct quantification of this sensitivity is given by the prior standard deviation of modeled  $v$  and  $H_{\text{mo}}$ , which, for October 1530 EST, had maximum values of 31 cm/s and 19 cm, respectively. At the observation locations, prior standard deviations ranged from 0–29 cm/s (for  $v$ ), and 0–10 cm (for  $H_{\text{mo}}$ ), with larger values occurring closer to shore. The 1DH case (20 October, 1130 EST) gave similar values, except for  $v$  at the observation locations which ranged from 0–13 cm/s. Given that model validation studies have reported errors on these same orders of magnitude (e.g. Ruessink *et al.* (2001)), this suggests bathymetric input error may be equally as important as process error, for cases like the ones we consider here.

It should be noted, however, that the above results are influenced by the specified prior statistics for  $h$ . In particular, the uncertainty in  $h$  is constructed, here, to reflect unresolved changes in bathymetry between surveys. Other sources of uncertainty, such as instrument error or spatial over-smoothing (Plant *et al.*, 2009), could be treated by a similar analysis, with  $C_{hh}$  redefined appropriately.

### 2.5.3 Effect of Sampling Scheme

In the preceding sections, we have performed the model inverse using all available measurements. However, the present data set, from the SD97 experiment, had an unusually large observational array which sampled both alongshore and across-shore variability. Previous experiments such as SuperDuck (Oltman-Shay & Howd, 1989) and DUCK94 (Elgar *et al.*, 1997; Feddersen & Guza, 1998; Gallagher *et al.*, 1998) have focused on only alongshore or across-shore variability, respectively. It is natural to ask whether the present method can be applied under a more limited experimental layout.

Figures 2.13 and 2.14 show the posterior wave and current fields for October 20, 1530 EST (cf. Figure 2.10), obtained by assimilating  $v$  and  $H_{\text{mo}}$  from a single alongshore or across-shore transect. The sampling schemes are similar to SuperDuck and DUCK94, respectively. We find that, in the present case, either sampling scheme is sufficient to im-



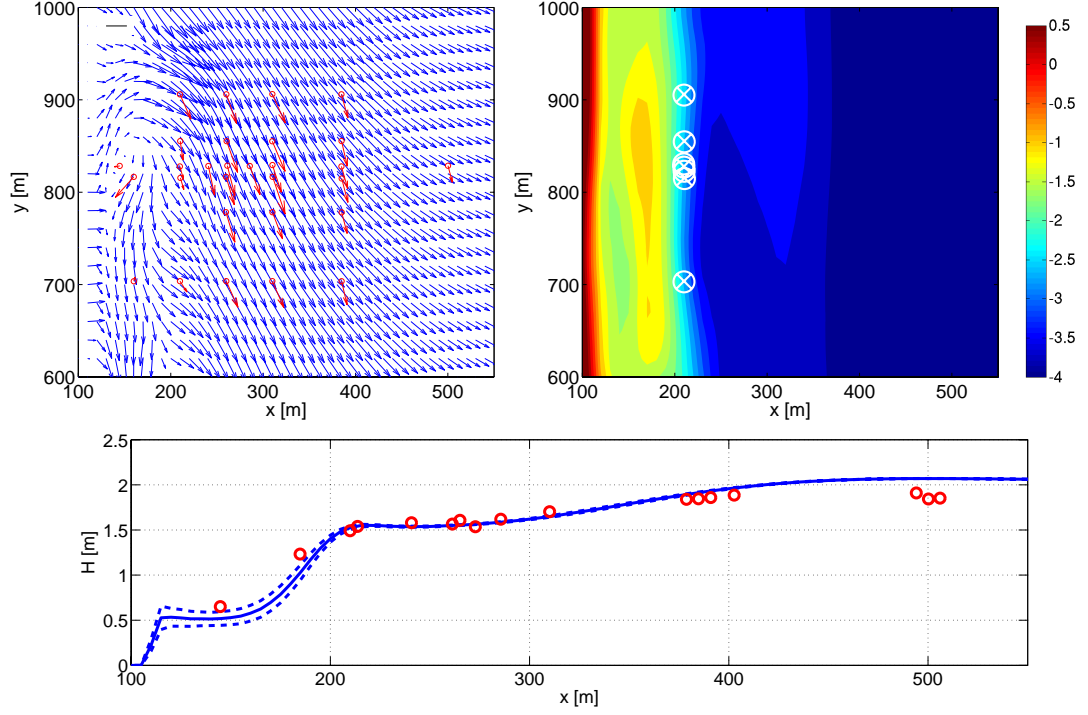


Figure 2.13: As in Figure 2.10, but for posterior (updated) fields using alongshore transect sampling scheme. Assimilated observations of  $v$  are marked by white crosses, and observations of  $H_{mo}$  are marked by white circles.

prove the prediction of the model state (a positive skill score  $S$  is found for all variables). However, the posterior bathymetry is more accurate when using the alongshore array (skill  $S = 0.77$ , rms error  $\epsilon = 15$  cm, taking measurements from  $x < 250$  m), compared to the across-shore array ( $S = 0.45$ ,  $\epsilon = 26$  cm). On the other hand, the alongshore array was less able to constrain wave height ( $S = 0.051$ ,  $\epsilon = 10$  cm), compared to the across-shore array ( $S = 0.27$ ,  $\epsilon = 8.0$  cm).

#### 2.5.4 Assimilation of Other Observational Data Types

So far, we have presented results for assimilation of  $v$  and  $H_{mo}$ , two commonly-measured observational data types. However, standard surf zone instruments are also capable of recovering additional information about the wave field, including wave directional information. As an example, here we test the assimilation of the radiation stress tensor component  $S_{xy}$ .

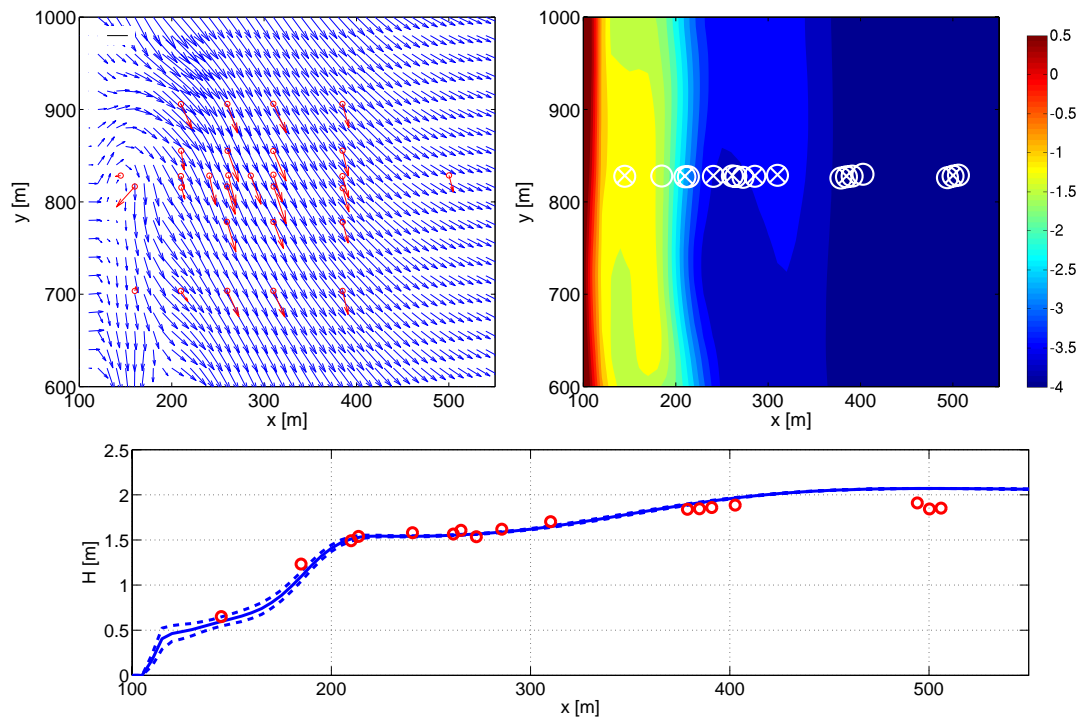


Figure 2.14: As in Figure 2.13, but for across-shore sampling scheme.

An observational estimate of  $S_{xy}$  can be computed from a cross-spectral analysis of timeseries of  $u$  and  $v$  (Higgins *et al.*, 1981). Here, we computed the cross-spectrum from 17 minute records, using a Bartlett taper. Depth-dependence is accounted for using linear wave theory.

Guza & Thornton (1978) have noted that  $S_{xy}$  is generally a statistically unstable observation which can be very difficult to measure, particularly due to instrument alignment issues. Hence, the choice of observational error for this derived quantity was not obvious, and we chose an error of 20 N/m (in the present test case,  $S_{xy}$  varied from -87–230 N/m). Values of 15 N/m and 30 N/m were also tested, and did not change the qualitative conclusions that follow.

When  $S_{xy}$  is assimilated (alone) for the 20 October 1530 EST case presented in section 2.4, the effect on the posterior bathymetry is qualitatively similar to that found when assimilating  $v$ . Specifically, the areas onshore and offshore of the inner bar at  $x \approx 160$  m were made deeper, a correction which could not be attained by assimilating  $H_{mo}$  alone. Hence,  $S_{xy}$  could be used in conjunction with  $H_{mo}$  to generate an improved posterior bathymetry ( $S = 0.65$ ,  $\epsilon = 18$  cm, for  $x < 250$  m). On the other hand, assimilation of  $S_{xy}$  was not found to be a substitute for the information provided by  $v$ . The skill of the posterior  $v$  was not much improved by assimilation of  $S_{xy}$  ( $S = 0.31$ ,  $\epsilon = 22$  cm/s), compared to when  $v$  itself was assimilated. Results were similar for the 1DH case. In summary, then,  $S_{xy}$  appears to provide information about bathymetry, but further work would be required to incorporate this data type into an accurate assimilation.

## 2.6 Conclusions

In this study, we have applied standard methods from data assimilation to examine the sensitivity of surf zone models to bathymetric uncertainty. Our purpose was twofold: to directly analyze the impact of bathymetric uncertainty on a surf zone model using field data, and to demonstrate the potential of ensemble-based data assimilation for 2DH nearshore prediction.

The results presented here show that, even in an extensively-sampled experimental setting (SD97, possibly the most detailed short-term bathymetric data set available to date), bathymetric uncertainty can play a leading role in determining the error of hindcast model circulation. This was demonstrated in several ways, as described below.

Figures 2.5 and 2.9 illustrate that the best prior estimate of bathymetry can lead to poor model results in the inner surf zone for a particular field case. On its face, this

could indicate a problem with model physics, or a problem with model inputs. However, by assimilating data under the assumption of perfect model physics and uncertain bathymetry, we were able to derive a consistent model state. Thus, we conclude that the standard approach of estimating bathymetry from recent bathymetric surveys and running the forward model did not apply here, as the bathymetry was changing rapidly (time scales of hours–days) and was therefore very uncertain.

The underlying details of the assimilation step were investigated using an analysis of the matrix of representers (Kurapov *et al.*, 2009), i.e. the interrelationship (covariance) between the modeled  $H_{\text{mo}}$ ,  $v$ , and  $h$ . We found that surf zone  $H_{\text{mo}}$  tended to provide slightly-larger magnitudes of correction to  $h$ , but corrections were relatively localized in space. Under 2DH flow, observations of  $v$  could provide non-local corrections to  $h$ , which meant a larger portion of the observational array for  $v$  was useful for the assimilation. The pattern and magnitude of the corrections are determined by a combination of  $C_{hh}$  (which must be specified) and the model dynamics. In this case,  $C_{hh}$  represents the estimated variability in  $h$  due to unresolved bathymetric change through time.

Table 2.1 reports the model-data misfit before and after assimilating data. For 2DH (1DH) flow, root-mean-square errors in  $v$  and  $H_{\text{mo}}$  in the inner surf zone were reduced by 81% (27%) and 24% (27%), respectively. Errors in  $h$  and  $u$ , variables which were not assimilated, were also reduced, by 56% (56%) and 42% (0%) respectively. Thus a significant portion of model output errors were linked to input (bathymetric) errors in a self-consistent way. However, we also found the assimilation to be sensitive to the type of observations used: only by assimilating multiple observation types ( $v$  and  $H_{\text{mo}}$  together) were we able to avoid overfitting the data. This was explained by considering the different types of information carried by each variable: observations of  $v$  were useful for constraining gradients of  $H_{\text{mo}}$ , but not magnitudes, and vice-versa for  $H_{\text{mo}}$ .

The above results all point to a strong model sensitivity to the input  $h$ ; this sensitivity was quantified directly using the prior ensemble variance. The estimated uncertainty in  $v$  and  $H_{\text{mo}}$  due to uncertainty in  $h$  was found to be up to 29 cm/s and 10 cm, respectively, at the measurement locations. These values are comparable to what is reported in typical field validation studies (e.g. Ruessink *et al.* (2001)). We stress, however, the modeled uncertainty is conditioned by the (specified) uncertainty in  $h$ .

Finally, we have tested the above results when using a subset of the available measurements. It was shown that an accurate posterior bathymetry and velocity field can be obtained when using only a single alongshore array of sensors, noting the present situation had strongly 2DH flow and therefore this configuration provides non-redundant

information. An across-shore array gave a less skillful posterior estimate of bathymetry and velocity, but was necessary for an accurate estimate of wave height.

Based on the above results, we conclude 2DH velocity and wave height observations do provide information about surf zone bathymetry, which can be exploited using statistical methods. Conversely, uncertainty in bathymetry (which is often large due to sampling constraints) can have a strong impact on model skill, a fact which should be considered when validating models.

# ENSEMBLE-BASED DATA ASSIMILATION FOR ESTIMATION OF RIVER DEPTHS

G.W. Wilson and H.T. Özkan-Haller

Journal of Atmospheric and Oceanic Technology  
45 Beacon Street Boston MA, 02108-3693, U.S.A.  
vol. 29, 1558–1568, doi:10.1175/JTECH-D-12-00014.1, 2012

### 3.1 Introduction

Flow in freshwater channels (rivers and upper estuaries) is governed by a number of factors, including the rate of discharge and the channel geometry. The latter may include effects due to large scale bedforms, such as bars and holes. Knowledge of bathymetry is therefore essential for modeling the channel hydrodynamics, particularly at the reach scale. In many natural settings, however, it is not practical to obtain a detailed field survey of bathymetry. Various methods have been developed using remote sensing technologies to replace or supplement field survey data (e.g. lidar (Hilldale & Raff, 2008), hyperspectral imaging (Legleiter *et al.*, 2009)); however, these methods are typically limited to clear shallow water. Hence, other investigators have proposed indirect methods for estimating bathymetry based on non-bathymetric variables that can be more easily measured, such as Lagrangian drifter trajectories (Honnorat *et al.*, 2010), or water surface elevation maps (typically for larger-scale applications, Andreadis *et al.* (2007); Durand *et al.* (2008)). The present work is another such method: we seek to estimate bathymetry indirectly using measurements of Eulerian velocity.

The use of measured river velocity to infer bathymetry relies on the strong sensitivity between those two variables (Smith & McLean, 1984). The main obstacles in exploiting this sensitivity involve (a) how the measurements are to be collected, and (b) how the sensitivity can be represented as an inverse model. Our focus here is on issue (b), the development and verification of the inverse model, using twin tests. Issue (a) will also be addressed to some extent by designing the twin tests based on the capabilities of existing instrument technology, including limitations in observational accuracy and spatial resolution.

In developing the inverse model, we will use tools from data assimilation, a methodological approach which combines known uncertainty in models and observations to produce a statistically-optimal estimate of the true state of a system. Data assimilation for estimation of model parameters (such as bathymetry) has been a topic of recent interest in river modeling (Andreadis *et al.*, 2007; Durand *et al.*, 2008; Tossavainen *et al.*, 2008; Honnorat *et al.*, 2010; Rafiee *et al.*, 2011). Most recently, Zaron *et al.* (2011) demonstrated the success of this approach to estimate bathymetry using remotely sensed velocity in an estuarine setting. Our work builds on this existing literature, continuing a trend towards field applications on natural channels at reach scales.

The paper is organized as follows. Section 3.2 describes the parameter estimation methodology. Section 3.3 then applies this methodology to three test cases: the first

case involves a one-dimensional channel with linearized dynamics, presented as a simple case which illustrates the use of the method; the second and third test cases involve two natural channels, to demonstrate the real-world applicability of the method. Sections 3.4 and 3.5 summarize and discuss the results.

## 3.2 Methods

### 3.2.1 Bathymetric Inversion Method

The method used here employs state-augmentation in an ensemble-based statistical estimator, treating bathymetry as a fixed model parameter. This approach is largely following Wilson *et al.* (2010), who applied similar methodology but in a nearshore ocean environment.

The goal of the method is to combine a prior estimate of bathymetry, a known discharge, and point measurements of velocity (including measurement uncertainty or noise), to produce a posterior estimate of bathymetry. Specifically, we seek to minimize the following cost function:

$$\begin{aligned} \mathcal{J}[\psi] = & \left( \psi - \psi^f \right)^T C_{\psi\psi}^{-1} \left( \psi - \psi^f \right) \\ & + (d - Lu)^T C_{dd}^{-1} (d - Lu). \end{aligned} \quad (3.1)$$

Here,  $\psi = [u^T, h^T]^T$  is a model state variable, consisting of a  $M \times 1$  vector  $u$  containing the velocity at each of the  $M$  model gridpoints, augmented with a  $M \times 1$  vector  $h$  containing the corresponding water depths. The  $K \times 1$  vector  $d$  contains the measurements, and  $L$  is a  $K \times M$  matrix which serves to extract the corresponding modeled values at the measurement locations. The matrices  $C_{\psi\psi}$  and  $C_{dd}$  are covariances of model and measurement uncertainties, respectively. The superscript  $f$  denotes the prior, or “forecast” model state, considered the most-likely state if no measurements were available.

The posterior, or “analysis” model state  $\psi^a$  is the one which minimizes  $\mathcal{J}$  above, and is given by (e.g., Evensen (2006); Bennett (2002))

$$\psi^a = \psi^f + C_{\psi\psi} L_a^T (L C_{uu} L^T + C_{dd})^{-1} (d - L \psi^f), \quad (3.2)$$

and the posterior covariance (interpreted as an estimate of uncertainty) is given by

$$C_{\psi\psi}^a = C_{\psi\psi} - C_{\psi\psi} L_a^T (L C_{uu} L^T + C_{dd})^{-1} L_a C_{\psi\psi}, \quad (3.3)$$



where  $C_{uu}$  is the upper-left submatrix of  $C_{\psi\psi}$ , and  $L_a = [L, 0^{K \times M}]$ . For Gaussian statistics,  $\psi^a$  can be interpreted as the maximum likelihood estimate of the true state, given all available information. For nonlinear dynamics this is not the case; instead we interpret  $\psi^a$  as a least-squares estimate, whose usefulness must be evaluated based on experiment, as will be done in subsequent sections.

### 3.2.2 Assumed Known Information

For the present application, we will assume the only unknown information is the deviation of channel depth from a nominal along-channel-uniform shape (e.g. a parabolic or piecewise-linear channel cross-section). While this nominal shape would need to be estimated somehow (for instance from a survey transect, or by using lower-order (channel-averaged) model equations and a separate statistical estimator) we consider this to be a separate problem from the one addressed here. Additionally, we will assume that the discharge is known during the time when measurements are collected; estimating discharge accurately would require a consideration of the larger-scale hydrology, and is not in the scope of the present work.

In summary, the assumed known quantities (or “inputs”) for the present method are (a) the nominal channel cross-section; (b) the discharge; and (c) a statistical characterization of the unknown deviations in depth, to be discussed next.

### 3.2.3 Specification of Prior

The solution (3.2) is so far incomplete as we have not yet specified a statistical model for the prior (mean and covariance). These appear as submatrices of  $C_{\psi\psi}$  in equation (3.2),

$$C_{\psi\psi} = \begin{bmatrix} C_{uu} & C_{uh} \\ C_{hu} & C_{hh} \end{bmatrix}. \quad (3.4)$$

In general, we would like to define a covariance  $C_{hh}$  representing uncertainty in the bathymetry, and from it derive (using a model) the corresponding velocity covariance,  $C_{uu}$ , as well as the covariance between velocity and bathymetry,  $C_{hu}$ . This can be done, for example, using adjoint model equations (Bennett, 2002; Zaron *et al.*, 2011). Another approach, used here, is to approximate the covariances using ensembles. An ensemble of  $h$  is drawn from a specified distribution (defined below), a numerical model is used to compute the corresponding  $u$  for each member of the ensemble, and then  $C_{\psi\psi}$  is

estimated using the sample covariance.

In the present application, the prior mean bathymetry is chosen on a case-by-case basis, but generally consists of a nominal channel cross-section and bank geometry. Perturbations around this prior mean are based on a bell-shaped covariance,

$$C_{hh}(\Delta x, \Delta y) = \sigma_h^2 \exp \left[ -3 \frac{\Delta x^2}{L_x^2} - 3 \frac{\Delta y^2}{L_y^2} \right], \quad (3.5)$$

where  $\Delta x$  and  $\Delta y$  are separation distances;  $L_x$ ,  $L_y$ , and  $\sigma_h$  are parameters representing the expected length and amplitude scales for bathymetric perturbations. We generate realizations from (3.5) using the Fourier Transform method described in Evensen (2006) (Fortran code available from [enkf.nersc.no](http://enkf.nersc.no)).

Note that the distribution  $C_{hh}$  as defined above includes the potential for small or even negative water depths in individual realizations of  $h$ . In practice, this can be problematic for the numerical model. Hence, in the cases discussed below we define a truncated distribution such that  $h > 0.5$  m. Similarly, some bathymetric perturbations can generate flows for which the Froude Number is large; in our model setup, such flows produce sharp steplike features in the free surface, ultimately leading to poor representation of the upstream open boundary condition. Hence, we also reject/replace any realizations for which the maximum Froude Number is greater than 0.5.

Finally, we note that ensemble estimates of  $C_{\psi\psi}$  may introduce spurious long-range correlations, which can contaminate the result of equation (3.2). To combat this, we follow the approach used by Hamill *et al.* (2001), wherein the estimated covariance is localized using element-by-element multiplication with a bell-shaped correlation function having compact support. Specifically, they define

$$(C_{\psi\psi})_{ij} = S_{ij} C_{\psi\psi}^{\text{ens}}{}_{ij}, \quad (3.6)$$

where  $S_{ij}$  is the desired correlation function. They suggest

$$S_{ij} = \Omega \left( \sqrt{10/3} L_{\text{loc}}, |x_i - x_j| \right) \quad (3.7)$$

where  $L_{\text{loc}}$  is the desired localization length scale, and  $\Omega(a, b)$  is given by

$$\Omega(a, b) = \begin{cases} -\frac{1}{4} \left(\frac{b}{a}\right)^5 + \frac{1}{2} \left(\frac{b}{a}\right)^4 + \frac{5}{8} \left(\frac{b}{a}\right)^3 - \frac{5}{3} \left(\frac{b}{a}\right)^2 + 1, & 0 \leq b \leq a \\ \frac{1}{12} \left(\frac{b}{a}\right)^5 - \frac{1}{2} \left(\frac{b}{a}\right)^4 + \frac{5}{8} \left(\frac{b}{a}\right)^3 + \frac{5}{3} \left(\frac{b}{a}\right)^2 - 5 \left(\frac{b}{a}\right) + 4 - \frac{2}{3} \left(\frac{b}{a}\right)^{-1}, & a \leq b \leq 2a \\ 0, & b > 2a \end{cases} \quad (3.8)$$

Using this method, any spurious non-local correlations are eliminated from the estimate of  $C_{\psi\psi}$ , and the resulting posterior estimate is improved.

### 3.3 Verification

#### 3.3.1 Idealized Test Case: Straight-Channel With a Bump

In this section we apply the present method in a simplified setting: flow with small Froude Number, with no variability in the across-stream direction (i.e. no effect of side walls). In this idealized case, the relationship between  $u$  and  $h$  can be written explicitly using a rigid-lid approximation  $hu = Q$ , where  $Q$  is a constant (known) discharge. This will allow us to illustrate (and test) some potential sources of error in equation (3.2).

As a specific test case, we consider the bathymetry shown schematically in Figure 3.1, consisting of a flat bottom interrupted by a Gaussian-shaped bump located at the center of the model domain:

$$h^t(x) = h_0 - h_b \exp \left[ -3 \frac{(x - l/2)^2}{l_b^2} \right], \quad (3.9)$$

where  $l = 500$  m is the domain length,  $l_b = 50$  m is the width scale of the bump,  $h_0 = 5$  m is the depth at  $x = 0$ , and  $h_b$  is the height of the bump. The prior bathymetry is defined as  $h^f = h_0$ , i.e. the position and height of the bump is unknown. The goal of the depth inversion is to detect the bump width, height, and location, based on measurements of velocity.

Measurements are defined by sampling at 12.5 m spacing from the “true” velocity  $u^t$ , defined as the velocity given by the rigid-lid dynamics with true bathymetry:  $u^t(x) = u_0 h_0 / h^t(x)$ , where  $u_0 = 0.5$  m/s is the velocity at  $x = 0$ . The measurement error covariance is taken to be diagonal,  $C_{dd} = (0.01 \text{ m})^2 I$  (where  $I$  is the identity matrix), although noise is not explicitly added to the data for these experiments. Unless otherwise stated, the prior ensemble uses parameters  $L_x = 50$  m,  $L_{\text{loc}} = 100$  m, and  $\sigma_h = 1$  m

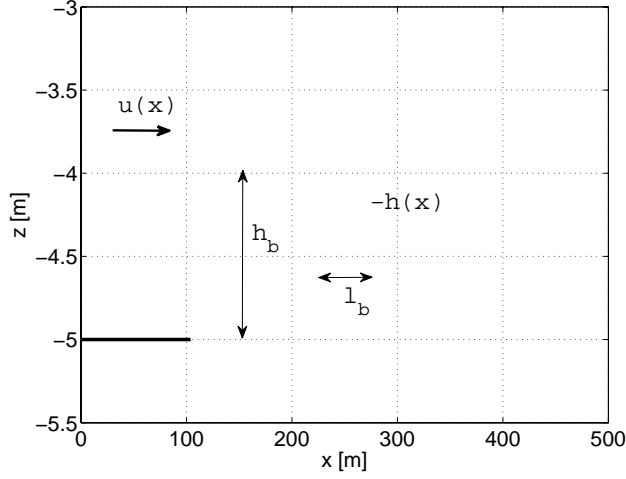


Figure 3.1: Schematic of 1D test bathymetry  $h(x)$ , equation (3.9), showing the location of the river bottom (or negative-depth,  $-h(x)$ ) as a function of along-channel distance ( $x$ ). The flow velocity  $u(x)$  is assumed depth-uniform, and is in the  $+x$  direction.

(see equations (3.5) and (3.6)). Also, unless otherwise stated, the same ensemble of 500 members is used for each test below.

Figure 3.2 shows examples of inverse bathymetry estimates, for several values of bump size  $h_b$ . In general, the estimator overpredicts the height of small bumps, and underpredicts that of large bumps. This can be viewed as a consequence of the implicit linearization in equation (3.2), when relating  $u$  and  $h$ . To see this, consider the correction  $\Delta h = h^a - h^f$  induced by an observed velocity error  $\Delta u = u_{\text{meas}} - u^f$ ; neglecting spatial correlation ( $C_{hu}$ ,  $C_{uu}$  scalar) and also neglecting measurement error ( $C_{dd} = 0$ ), the correction can be written as

$$\Delta h = \frac{C_{hu}}{C_{uu}} \Delta u. \quad (3.10)$$

Assuming a relationship  $u = u(h)$ , we can write, formally,

$$\begin{aligned} \Delta h &= \frac{E[\delta h \delta u]}{E[\delta u \delta u]} \Delta u \\ &= \frac{E\left[\left(\frac{dh}{du}\bigg|_{u^f} \delta u + \frac{1}{2} \frac{d^2 h}{du^2}\bigg|_{u^f} \delta u^2 + \dots\right) \delta u\right]}{E[\delta u \delta u]} \Delta u \\ &= \frac{dh}{du}\bigg|_{u^f} \Delta u + \frac{1}{2} \frac{d^2 h}{du^2}\bigg|_{u^f} \frac{E[\delta u^3]}{E[\delta u \delta u]} \Delta u + \dots \end{aligned} \quad (3.11)$$

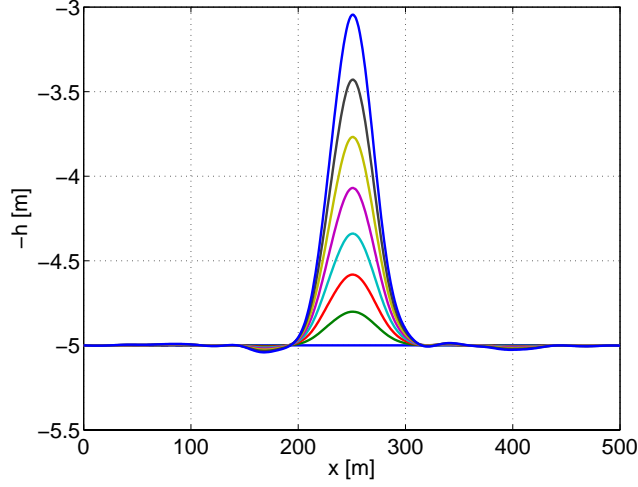


Figure 3.2: Example estimates of  $h(x)$  using equation (3.2), for various values of bump size  $h_b$ ; corresponding true  $h(x)$  are plotted as dashed lines.

Whereas the desired “true” correction would be

$$\Delta h = \frac{dh}{du}_{u^f} \Delta u + \frac{1}{2} \frac{d^2 h}{du^2}_{u^f} \Delta u^2 + \dots \quad (3.12)$$

If the relationship between  $u$  and  $h$  is linear, then these two expressions agree, and (3.2) is valid. A nonlinear relationship, on the other hand, as in the present case  $u = Q/h$ , will result in error, especially if  $\Delta u$  is large. This is illustrated in Figure 3.3, which plots the bathymetry correction,  $\Delta h = h^f - h$ , as a function of the prior velocity error,  $\Delta u = u^f - u_{\text{meas}}$ . Equation (3.2) predicts this relationship is approximately linear (solid line), whereas the true relationship (dashed line) is not. Referring back to Figure 3.2, this same linearization is responsible for the underprediction of the height of small bumps, and overprediction of that of large bumps. More generally, we should expect the bathymetry estimate to have larger errors if the prior  $h^f$  is further from the truth.

Next, we consider the effect of ensemble size when computing covariance estimates for equation (3.2). To test this, we generated different sized ensembles ranging from 10 to 1000 members (in increments of 10). Ten independent ensembles were generated for each ensemble size, in order to account for random variability. Then, for each ensemble, we applied equation (3.2), and computed the root-mean-square error of the resulting bathymetry estimate. This is plotted as a function of ensemble size, in Figure 3.4. As expected, error is reduced with increasing ensemble size; the estimate appears to be

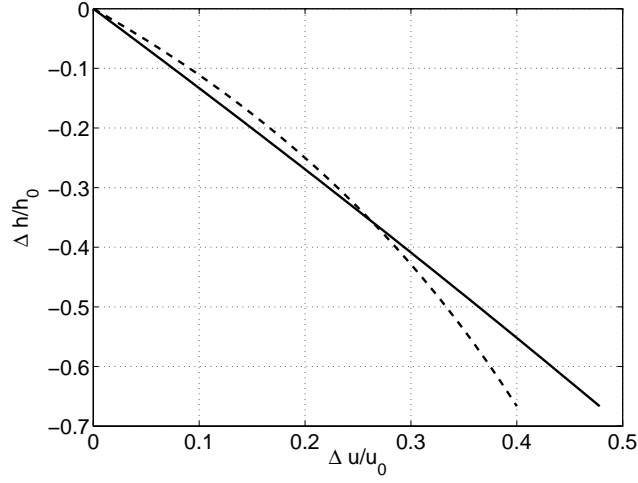


Figure 3.3: Estimated depth correction  $\Delta h = h^f - h$  from equation (3.2), plotted as a function of prior velocity error  $\Delta u = u^f - u_{\text{meas}}$ , at the location  $x = l/2$ . True relationship (using  $u = Q/h$ ) is plotted as dashed line; differences between true and estimated relationship are attributed to nonlinearity (see text).

converged for ensemble sizes greater than about 500 members.

Lastly, Figure 3.5 (left panel) shows the effect of covariance decorrelation length  $L_x$  on the depth estimate. Assuming a decorrelation length less than the observation sampling rate (in this case 12.5 m) results in overly-localized corrections; assuming an unrealistically-large decorrelation length, on the other hand, results in overly-smoothed corrections. Hence, the choice of  $L_x$  should, ideally, take into account the expected bathymetric scales as well as the sampling scale. Figure 3.5 (right panel) quantifies this dependence for various measurement sample spacings  $dx_{\text{obs}}$  (the point  $x = 250$  m is always included in the measurements, and other measurements are spaced a distance  $dx_{\text{obs}}$  apart). In all cases, error in the bathymetry estimate is minimized by choosing  $L_x$  to be close to the true bump length scale  $l_b = 50$  m. However, when the measurement sample spacing is small, there is less penalty for choosing a smaller value of  $L_x$ . Note in realistic applications it would be advantageous to choose  $L_x$  as small as possible so that smaller scales of bathymetry, if they exist, would also be well-estimated. Hence as a rule of thumb based on the present tests, we recommend choosing  $L_x$  as the lesser of (a) three times the measurement sample spacing, and (b) the maximum expected bathymetric length scale.

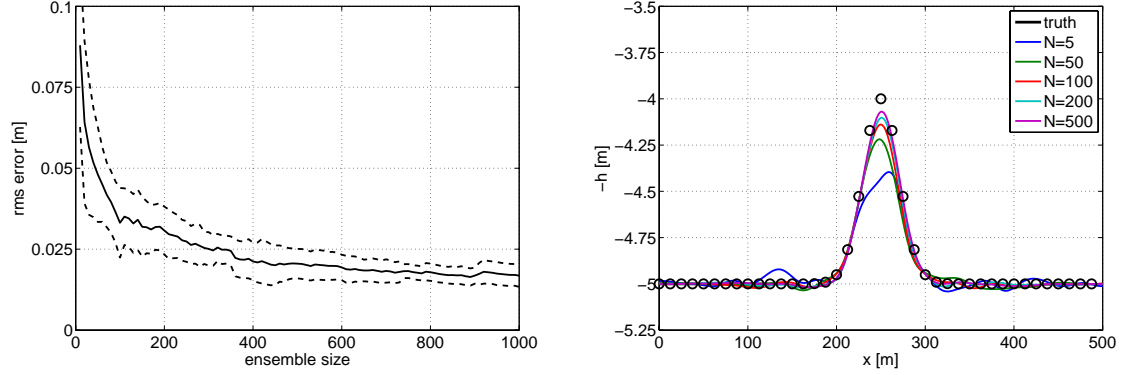


Figure 3.4: Left: convergence of root-mean-square error (rmse) for increasing ensemble size; average (solid) and  $\pm$  one standard deviation (dashed), computed over 10 realizations of each ensemble size. Right: example estimates using various ensemble sizes  $N$ .

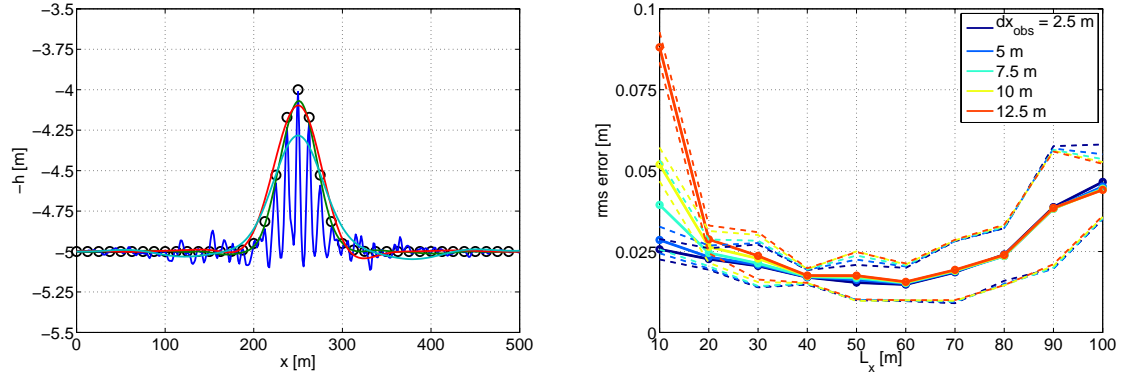


Figure 3.5: Left: depth estimates of a bump with  $h_b = 1$  m, for  $L_x = 5$  m (blue), 50 m (green), 100 m (red), and 200 m (cyan). True depth is plotted in black, and measurement locations are marked by circles (measurement sample spacing is  $dx_{\text{obs}} = 12.5$  m). Right: root-mean-square (rms) difference between true and estimated bathymetry, using same bump geometry, for various decorrelation lengths  $L_x$ , and various measurement sample spacings  $dx_{\text{obs}}$ . Ten realizations of 500-member ensembles are computed for each combination of  $L_x$  and  $dx_{\text{obs}}$ ; solid line represents average rms error, and dashed lines are  $\pm$  one standard deviation.

### 3.3.2 Test Cases With Realistic Bathymetry

Thus far, we have considered several specific aspects of the depth inversion process which can affect the result. These include the nonlinearity in the relationship between  $u$  and  $h$ , the specification of the prior covariance, and the impact of ensemble-based estimates. These effects can be isolated when assuming highly-idealized dynamics and geometry; in a realistic case, however, they act simultaneously. We next test the method using two real-world reaches with known bathymetry.

#### 3.3.2.1 Test Case A: Snohomish River, WA

The first case we consider is based on a 2.3 km reach of the Snohomish River, WA (TerraSond, 2009), which was the site of the COHerent STructures in Rivers EXperiment (or, COHSTREX) (Chickadel *et al.*, 2009; Giddings *et al.*, 2011). As part of the COHSTREX experiment, a high-resolution bathymetric survey was conducted at this site during September, 2009 (TerraSond, 2009).

The river is nominally 100 m wide and its depth varies with the tide, having channel depths in the nominal range 3–6 m. Discharge is also tidally-influenced, spanning a range of about  $\pm 300 \text{ m}^3/\text{s}$ . Salinity intrusions occur with each tide cycle, but here we will restrict our attention to tidal phases when measured salinity was negligible. Specifically, we assume an ebb-tide phase with discharge  $180 \text{ m}^3/\text{s}$  and nominal depth 3.5 m, and we neglect buoyancy effects in the equations of motion.

#### 3.3.2.2 Test Case B: Kootenai River, ID

As a second test case, we consider a reach of the Kootenai River, ID. This channel is deeper than that of the Snohomish River (nominal depth 6.5 m), and has larger spatial scales in its bathymetry and bank geometry.

Bathymetric data for this test case were collated from existing U.S. Geological Survey measurements (Barton *et al.*, 2004), as well as supplementary survey data collected in 2010, as described by Swick (2011). The same 2010 field program (Swick, 2011) provided measurements of river stage and discharge on which we have based our tests. Adjoining streams and side-channels were excluded from the model domain, hence their influence on discharge is not accounted for in the present model.



### 3.3.2.3 Governing Equations and Numerical Model

The river dynamics (necessary to provide forcing to the inverse model) are simulated using depth-averaged hydrostatic equations of motion. To solve the equations numerically, we employ the community code ROMS (Regional Ocean Modeling System) (Shchepetkin & McWilliams, 2005). The ROMS code integrates the equations of motion on a curvilinear grid, which we generate using the “gridgen” software of Pavel Sakov (<http://www.marine.csiro.au/~sakov>).

The model grid for the Snohomish River includes 651 points along-channel (nominal spacings 2–6 m) and 60 points across-channel (nominal spacings 1–3 m). For the Kootenai River, we use 739 points along-channel (nominal spacings 6–10 m) and 35 points across-channel (nominal spacings 3–5 m). In both cases, the model time step is 0.025 seconds.

The ROMS code includes a variety of options for model physics and parameterizations; some essential details on our particular setup are as follows. The model includes bottom stress in the form of a quadratic drag law, with drag coefficient  $C_f = 0.01$ , and harmonic mixing with eddy viscosity  $\nu_t = 0.01 \text{ m}^2/\text{s}$ . Boundary conditions at the river banks (defined as the location where depth is less than 0.1 m) are no-slip; upstream and downstream boundary conditions are ramped up from rest to a steady discharge over a period of 30 minutes. Discharge is then held fixed at the downstream boundary until the flow reaches a steady state. The ramping-up phase excites waves which must be allowed to propagate out of the model domain, hence we employ Flather and Chapman boundary conditions (Palma & Matano, 1998) for velocity and free surface at the upstream boundary. For the incoming velocity and free surface height, which must be pre-specified, we assume a channel-averaged balance between downstream pressure gradient and bottom stress; some tuning is required in this calculation, to account for the effective drag due to nonuniform channel geometry.

### 3.3.2.4 Prior Statistics

We use a 500-member ensemble of bathymetric perturbations, drawing from the distribution described in section 3.2.3. For the Snohomish River, we choose parameters  $L_x = 100 \text{ m}$  (along-channel decorrelation length),  $L_y = 50 \text{ m}$  (across-channel),  $\sigma_h = 0.75 \text{ m}$ , and  $L_{\text{loc}} = 100 \text{ m}$ . The Kootenai River has naturally larger spatial scales, hence we increase the prior decorrelation length scales to  $L_x = 200 \text{ m}$ ,  $L_y = 100 \text{ m}$  (other parameters are unchanged). Note we do not consider the measurement sample spacing in the choice of

$L_x$  and  $L_y$ ; this is to ensure an objective comparison among various sampling schemes (defined below).

As discussed in section 3.3.1, the present method is, in a sense, linearized about the prior mean state, and hence its accuracy depends partly on the choice of the prior mean. For both the Snohomish and Kootenai Rivers, we assume the prior mean bathymetry is uniform in the along-channel direction, i.e.  $h^f(x, y) = h^f(y)$ , where  $y$  is the across-channel coordinate. For the Snohomish River, we assume  $h^f(y)$  is piecewise-linear, having a bank slope of 1:5 and a horizontal bottom at 3.5 meters depth (Figure 3.6a). For the Kootenai River, we use a parabolic shape for  $h^f(y)$  (Figure 3.9a). In both cases,  $h^f$  approximately matches the measured along-channel-averaged depth profile, hence is partly based on known information. In practice,  $h^f$  could perhaps be estimated by assuming highly simplified and/or channel-averaged dynamics, as in Zaron *et al.* (2011).

### 3.3.2.5 Simulated Measurement Schemes

Next, we employ a similar test methodology as above (for idealized 1D dynamics), where synthetic measurements are extracted from a forward model run with true bathymetry, and are then used to generate bathymetry estimates. Specifically, we generate a velocity field for the true (measured) bathymetry using the model described in section 3.3.2.3 — this is assumed to be equivalent to the true velocity, i.e. errors in model physics are not considered. The velocity is then “measured” by interpolating to a set of measurement points, and adding normally-distributed random noise to simulate instrument noise. Finally, these measurements are used in equation 3.2 to estimate bathymetry, and the result is compared to the truth (a “twin test”). We will test several possible measurement sampling schemes, representing realistic observational capabilities.

First, we test isolated point measurements of 2D velocity. In this case, we assume high accuracy is obtained at the expense of spatial resolution, for example using an in-situ gage. Measurements are taken from gridpoints along the channel centerline, with a nominal spacing of 200 m along-channel. Measurement error standard deviation is taken to be 0.01 m/s (e.g. the typical upper limit for bias error in acoustic Doppler current profiler instruments, RDInstruments (1996)).

A second test involves assimilation of 2D velocities, measured along a simulated drifter track, e.g. Swick (2011). The drifter is released near the center of the channel at the inlet, and 2D model velocities are interpolated to the drifter track at 5 m along-track spacing. Measurement error standard deviation is taken to be 0.05 m/s.

Case	Full-Domain		Obs.-Points	
	rmse [m]	$r^2$	rmse [m]	$r^2$
Prior	1.4	0.18	1.4	0.014
Pts	1.2	0.37	1.1	0.61
$u$	1.0	0.57	0.91	0.64
$v$	1.2	0.37	1.3	0.27
$u, v$	0.84	0.70	0.76	0.77
Drifter	1.1	0.52	1.4	0.48

Table 3.1: Accuracy statistics for Snohomish River bathymetry estimates, comparing several measurement sampling schemes. Labels in first column are: Prior, no assimilation; Pts, assimilating isolated point observations of 2D velocity;  $u$ , assimilating observations of along-channel (but not across-channel) velocity on a  $\sim 10$  m grid;  $v$ , assimilating observations of across-channel (but not along-channel) velocity on a  $\sim 10$  m grid;  $u, v$ , assimilating observations of 2D velocity on  $\sim 10$  m grid; and Drifter, assimilating 2D velocity sampled along a simulated drifter track. The statistics are root-mean-square error (rmse) and squared-correlation ( $r^2$ ); separate columns show statistics computed using all model gridpoints as the sample (“Full-Domain”), and using only the measurement locations as the sample (“Obs.-Points”).

The final test case involves assimilation of 2D velocities on a coarse grid, similar to what could be obtained from remote sensing, e.g. Plant *et al.* (2005) using Doppler radar. Data are subsampled from model gridpoints, at nominally  $10 \times 10$  m spatial resolution, excluding locations within 15 m of the river banks. Measurement error standard deviation is taken to be 0.1 m/s.

### 3.3.2.6 Results, Snohomish River

Figure 3.6 shows the prior bathymetry, and the true bathymetry for the Snohomish River. To reiterate, the goal of the assimilation is to reduce error in the prior, i.e. obtain a better match to the truth. Results for the various sampling schemes outlined above are shown in Figure 3.7. Accuracy statistics are given in Table 3.1. In all cases, the posterior bathymetry is an improvement over the prior, indicating positive skill for the depth-inversion routine. As expected, higher skill can be obtained by reducing the observational error and/or increasing the observational resolution. The various sampling schemes tested here reflect inherent trade-offs between those two factors.

The most bathymetric information is recovered when spatially-dense observations of both across-channel and along-channel velocities are assimilated (Figure 3.7a). Assimi-

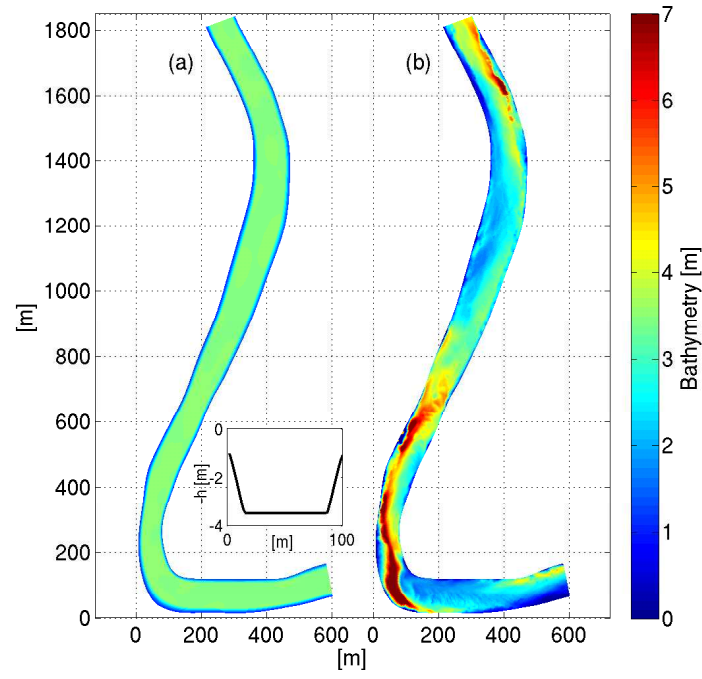


Figure 3.6: Prior bathymetry (a) and measured bathymetry (b) for Snohomish River test case. Inset shows along-channel-average of prior bathymetry, plotted as a function of across-channel position.

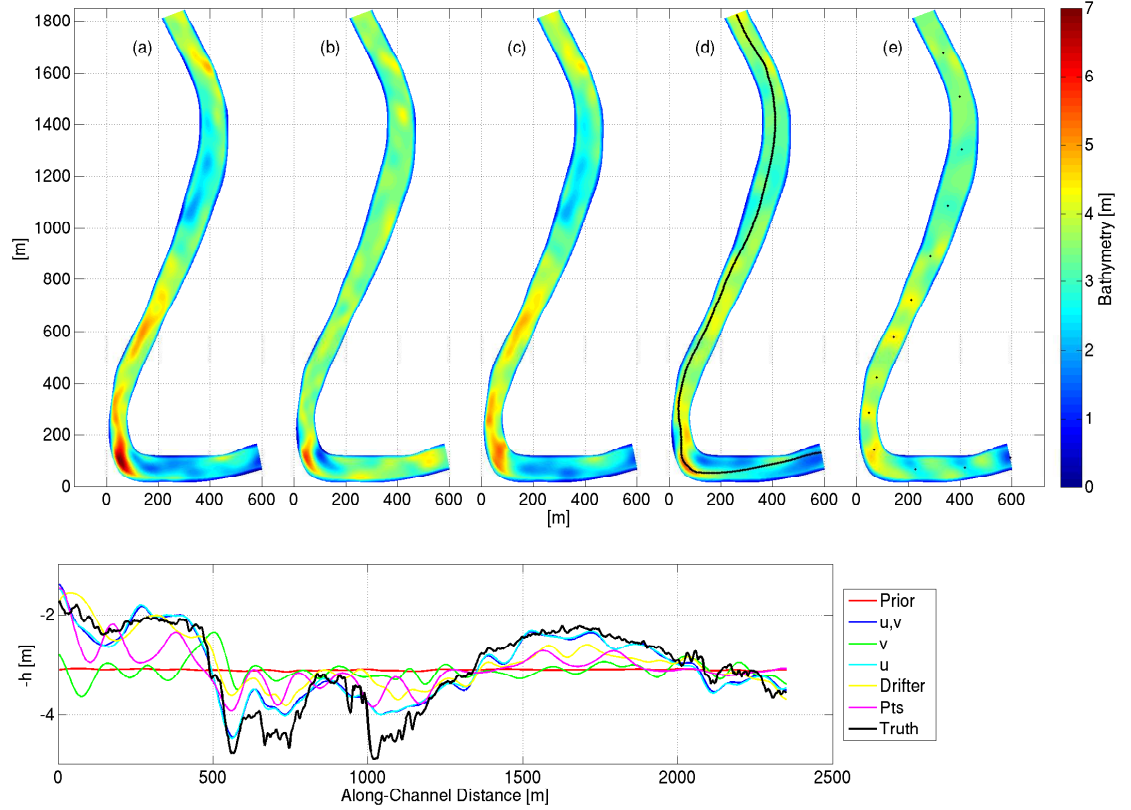


Figure 3.7: Posterior bathymetry for Snohomish River test case, with various sampling schemes: (a), assimilating 2D velocities,  $\sim 10$  m grid; (b), assimilating across-channel velocity only,  $\sim 10$  m grid; (c), assimilating along-channel velocity only,  $\sim 10$  m grid; (d), assimilating across- and along-channel velocities at 5 m spacing on a simulated drifter track; and (e), assimilating isolated point measurements of across- and along-channel velocities. For (d) and (e), the locations of measurements are indicated by black dots. Bottom plot shows across-channel-averaged depths, compared to truth.

lating along-channel velocity (Figure 3.7c) produces a larger and more accurate correction than across-channel velocity (Figure 3.7b). However, we note that the assimilation of across-channel velocity appears to provide/reinforce the across-channel structure of the bathymetry, for example the location of the thalweg in the southern river bend.

Assimilation of 2D velocities along a simulated drifter track (Figure 3.7d) shows promising results, at least in terms of resolving the along-channel variability of bathymetry. Similarly, estimates from assimilating point observations are somewhat accurate near the observations themselves, but little to no information is gained in terms of non-local features.

Finally, although thus far our focus has been on the estimation of bathymetry itself, we note that equation (3.3) also provides an estimate of the posterior uncertainty in bathymetry. Figure 3.8 shows the maps of posterior standard deviation (square root of diagonal of  $C_{hh}^a$ ), for each of the depth estimates in Figure 3.7. Recall the prior standard deviation was  $\sigma_h = 0.75$  m; by definition, assimilation of data causes a reduction of standard deviation below this value. Cases where the posterior bathymetry estimate is skillful (e.g. assimilation of spatially-dense 2D velocities, Figures 3.7a and 3.8a) correspond to smaller posterior standard deviation, as expected. Likewise, locations where the bathymetry was not significantly changed due to assimilation of data (e.g. far from observation points, Figures 3.7d and 3.8d) have almost no reduction in standard deviation. This demonstrates that the posterior standard deviation is providing meaningful information about bathymetric uncertainty, which could be used to aid the interpretation of the result, or even to guide the further collection of observations.

### 3.3.2.7 Results, Kootenai River

The results for the Kootenai River are similar to those for Snohomish River. Accuracy statistics are summarized in Table 3.2. As an example, Figure 3.9 shows the posterior depth estimate from assimilating  $\sim 10$  m-gridded 2D velocities; results from other tests are listed in Table 3.2 but are not plotted, for brevity (conclusions are similar to Snohomish River, Figure 3.7).

## 3.4 Discussion: Applicability to Real Observations

The present results have shown the potential for estimating river bathymetry using spatially-dense measurements of velocity. This has been done using synthetic observa-

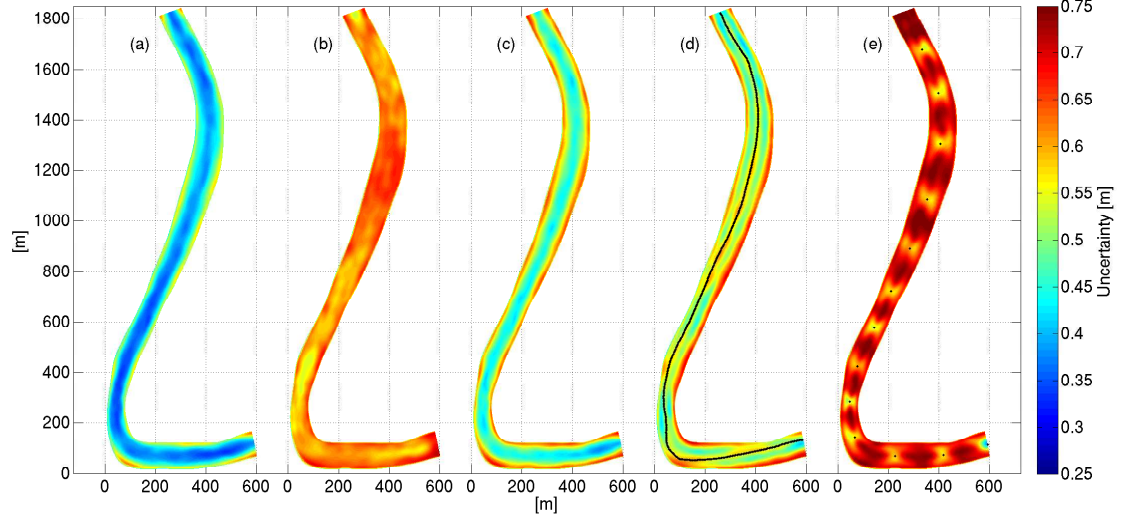


Figure 3.8: Posterior standard deviation (i.e. uncertainty) of bathymetry for Snohomish River test case, after assimilation of measurements using various sampling schemes. Labels (a)–(d) are as in Figure 3.7.

Case	Full-Domain		Obs.-Points	
	rmse [m]	$r^2$	rmse [m]	$r^2$
Prior	1.6	0.52	1.8	0.19
Pts	1.4	0.63	1.2	0.62
$u$	1.1	0.78	1.2	0.64
$v$	1.6	0.56	1.7	0.24
$u, v$	1.1	0.77	1.1	0.67
Drifter	1.3	0.71	1.5	0.45

Table 3.2: Accuracy statistics for Kootenai River test cases (cf. Table 3.1 for labeling conventions).

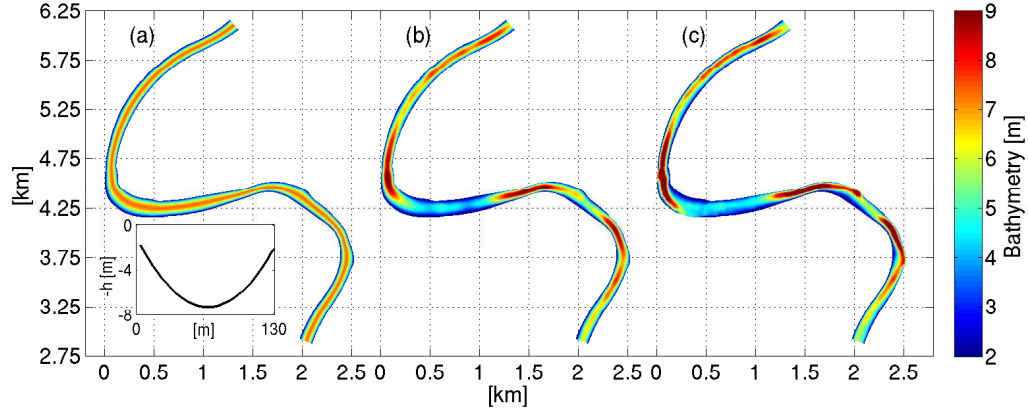


Figure 3.9: Prior (a), posterior (b), and measured (c) bathymetry, for Kootenai River, assimilating 2D velocities on  $\sim 10$  m grid. Inset shows along-channel-average of prior bathymetry, plotted as a function of across-channel position.

tions derived from the same model (i.e., twin tests) in order to demonstrate the method in an idealized setting. Extension to real observations is a logical next step, however several important hurdles remain.

Model error is an obvious concern for real-world application of this method, and has not been considered here. One question is whether the present method would be significantly contaminated by spatial variability in bottom stress, which is parameterized using a constant bottom roughness in the numerical model. Similarly, the model used here is only capable of simulating depth-averaged hydrostatic flow; in cases where the flow is non-hydrostatic, or strongly depth-dependent, the resulting estimate of bathymetry would be contaminated, as shown by Honnorat *et al.* (2010). And finally, we have also assumed prior knowledge of river discharge and channel-averaged depth; error in those parameters would correspond to error in model boundary conditions, and would affect the posterior estimate.

Another assumption made in the present methodology is that of a linear relationship between velocity and bathymetry. Because this assumption is violated, the update equation (3.2) is not truly optimal, but is in a sense a linearization around the prior state (this was demonstrated for a simplified case in equation (3.11)). Figure 3.3 illustrates the effect of this linearization on the posterior prediction. Based on this interpretation, the method may produce large errors in cases where the prior model state is very far from the truth. Possibly this error could be reduced by introducing an ad-hoc iterative



scheme. We suggest experiments with such a procedure (and its statistical justification) as a topic for future work.

Finally, it is important to emphasize the role of observational uncertainty. Without an accurate model for observational error, one risks overfitting to unphysical noise. Hence, data must be properly vetted for quality and given appropriate error bars. This seems particularly relevant to the present method, which benefits greatly from spatially dense observations as in remote sensing data; often such data require careful quality control.

Despite the above caveats, we believe the present method is a useful tool for river depth estimation. It has the advantage of being easily extensible to new observation types, and is able to handle observational uncertainty/error. Continuing developments in remote sensing techniques for river flow (e.g. Plant *et al.* (2005); Chickadel *et al.* (2011); Puleo *et al.* (2012)) make this a promising possibility. Application of the present method to such data will be an interesting challenge and may reveal new technical or physical insights.

### 3.5 Summary

We find that the sensitivity of river velocity to variations in river bathymetry is strong enough to be exploited in an inverse model for realistic observational data quality. In other words, measurements of river velocity can potentially be used to estimate bathymetry. We have presented a methodology for doing so, using a least-squares estimator, which takes into account the prior bathymetric uncertainty as well as the measurement uncertainty.

The inverse method was verified using synthetic twin tests. Under highly-simplified 1D channel dynamics, the method is capable of detecting bathymetric perturbations with high skill (section 3.3.1). In more-complex applications with real-world channel geometry (section 3.3.2), the method still produces skillful corrections to bathymetry, quantified in Tables 3.1 and 3.2. Moreover, the method also quantifies the posterior uncertainty of the bathymetry estimate, which can aid the interpretation of the result.

The effect of various observational sampling schemes on the accuracy of the posterior bathymetry was also investigated using synthetic tests. These schemes represent inherent practical trade-offs between spatial resolution and observational accuracy. We find that higher spatial resolution of observations, and observation of both across-channel and along-channel velocity, can help to resolve more detailed features such as the location of the river thalweg, or individual bumps/holes. The most highly-resolved observational

scheme produced the most skillful estimate, despite having larger observational error.

The application of this method using real measurements, especially remote-sensing data, is promising based on the present results. We note, here, that detailed estimates of measurement error are also valuable and are crucial for the accuracy of the present method. Also, a prior estimate of mean river depth and discharge is required — for example, from a coarser-scale model. Despite these caveats, the present results suggest that river bathymetric inversion is within reach of current observational and modeling capabilities.

## 4. Surf Zone Bathymetry and Circulation Predictions via Data Assimilation of Remote Sensing Observations

### 4.1 Introduction

The surf zone is defined as the coastal region where the effects of wave breaking dominate the hydrodynamics. It is characterized by large waves and strong currents, and often has a sandy/unstable bottom, and hence can be a challenging or even hazardous environment for in-situ observation. Because of this, remote sensing has played an important role in surf zone research; for a recent review, see Holman & Haller (2013). The most common implementation has been optical imagery collected from shore-based platforms such as Argus cameras (Holman & Stanley, 2007). Recently, efforts have also been made to exploit new imaging mechanisms such as radar (Catalán *et al.*, 2011; Puleo *et al.*, 2003; Haller *et al.*, 2013), infrared (Chickadel *et al.* (2009), and as discussed herein), and LIDAR (Blenkinsopp *et al.*, 2012), as well as new platforms such as stereo-video (de Vries *et al.*, 2011; Palmsten & Holman, 2011) and airborne. Furthermore, image processing techniques are increasingly being used to translate surf zone imagery into quantitative data products (e.g., using particle image velocimetry, Holland *et al.* (2001); Puleo *et al.* (2003); as well as other techniques discussed herein). And beyond remote sensing, other non-traditional measurement techniques such as passive GPS-equipped drifters (MacMahan *et al.*, 2010) and fluorescent dye tracers (Clark *et al.*, 2009) are now becoming standard.

A common factor among all the above measurement techniques and data products is the ability to sample a broad spatial and temporal range, compared to traditional in-situ instruments (i.e., bottom-mounted gages and profilers). The trade-off, however, is usually in terms of measurement uncertainty (“noise”) and/or the inability to sample continuous high-quality data at a fixed location (“sparseness”). Hence the use of such data requires an ability to filter through sparse and noisy observations. Data assimilation is a powerful approach to this problem where one seeks to utilize the full information contained in observations (including their uncertainty), combined with knowledge of physical processes, to generate a statistically optimal estimate of an unknown variable. One of the major triumphs of data assimilation, for example, has been the use of satellite

remote sensing data to reduce errors in numerical weather prediction. It is estimated that three quarters of the effective information used by such models now comes from satellite data (Cardinali *et al.*, 2004), and this has resulted in marked improvement in forecast skill. The challenges faced in that case (noisy/sparse imagery from non-traditional platforms) are completely analogous to the situation described above. Hence, data assimilation methods are a natural fit for surf zone prediction using remote sensing.

A necessary prerequisite for applying data assimilation to the surf zone is the specification of uncertainty in the data, and, importantly, in the model itself. Here, we argue that the dominant source of surf zone model uncertainty is often due to uncertain bathymetry. This is explained next.

Bathymetry plays a central role in wave transformation and breaking, which is in turn responsible for the transfer of momentum from waves to time-averaged surf zone currents. This leads to the suspension and transport of sediment, which ultimately causes changes in the bathymetry itself. Hence, not only is bathymetry difficult to measure directly due to operational considerations (breaking waves, strong currents), it is also time-varying. Indeed, significant bathymetric change can occur even on daily time scales (Lippmann & Holman, 1990), which can be responsible for significant model error if not corrected (Wilson *et al.*, 2010). As a result, errors in bathymetry are often cited as a fundamental barrier to operational nearshore modeling (Allard *et al.*, 2008; Austin *et al.*, 2012).

One approach to minimizing bathymetric error is to incorporate the physics of sediment transport into the modeling system, thus reducing the need for continual measurements of bathymetric change. Here, however, the modeler is faced with several challenges. First, the bathymetry and sediment properties (e.g., grain size) still must be initialized in the model. Second, present knowledge of sediment transport processes is far from mature and is the subject of ongoing research (e.g., Hoefel & Elgar (2003); Henderson *et al.* (2004)). Finally, the spatial scales associated with sediment transport processes (of order centimeters to millimeters in the turbulent bottom boundary layer) are much smaller than the typical resolution of wave and circulation models, hence one must rely on sub-grid-scale parameterization. The accuracy of such parameterizations is often suspect, to the point that fundamental predictability of medium-to-large scale bathymetric change has been questioned by some (Coco & Murray, 2007; Plant *et al.*, 2006).

To bring the various threads of discussion, above, together: in the present work we propose to use remote sensing data to control errors in surf zone bathymetry via data assimilation. This is a departure from the traditional approach of driving a nearshore model

with direct measurements of bathymetry, and/or deterministically modeling bathymetric change through time. Instead, the problem is cast in terms of a non-deterministic modeling system which incorporates bathymetric uncertainty (due to either unknown initial bathymetry or unknown bathymetric change, or both) as a random time-varying model parameter. The time-evolution of this parameter is treated very simply, and no attempt is made to incorporate physical models for sediment transport processes. Instead, bathymetry is estimated/controlled via the assimilation of data alone.

We approach this problem with two goals in mind. First (a), assimilation of data results in an improved estimate of bathymetry, which can be used to improve model predictions of other variables (e.g., currents) and reduce forecast error. Second (b), as more data are assimilated the estimated bathymetry will become increasingly accurate and hence could be used for other applications (e.g., monitoring bathymetric change through time). The latter goal (b) has been the focus of several recent applications of data assimilation in nearshore and shallow water environments (van Dongeren *et al.*, 2008; Holman *et al.*, 2013; Wilson & Özkan-Haller, 2012; Zaron *et al.*, 2011). In that case, the focus is on inversion of a physical model to assimilate potentially-noisy observations and predict bathymetry. The work of van Dongeren *et al.* (2008) and Holman *et al.* (2013) in particular highlight the benefits of using surf zone remote sensing data in such an application. The other goal (a), control of bathymetric error in a predictive surf zone model, was investigated by Wilson *et al.* (2010), who demonstrated model skill was strongly influenced by bathymetric error in a field setting, which could be corrected by assimilation of data.

The work of Wilson *et al.* (2010) can be viewed as a precursor to the present study. They found that bathymetric errors could be corrected using single-time observations of wave heights and alongshore currents from an in-situ instrument array. Here, we extend their methods to assimilate time-dependent remote sensing observations of multiple geophysical variables, a more realistic scenario for operational use. Our modeling system, described in section 4.3, assimilates remote sensing data as they become available (in this case once every half hour), and continually updates the estimated bathymetry and its uncertainty without the need for direct in-situ observation. In section 4.2, we outline a series of remote sensing data products, as well as in-situ data, collected during a 2010 field experiment, which we use to test the system. Results are shown in sections 4.4 and 4.5. We find that bathymetric errors can be successfully controlled using remote sensing data alone, and this leads to significant improvement in the ability of the model to predict surf zone currents.

## 4.2 Observations

### 4.2.1 Experiment

Observations were collected during a field experiment in September 2010 at the U.S. Army Corps of Engineers Field Research Facility (FRF) in Duck, NC. The study domain encompassed approximately 1 km of a sandy barrier island beach, from the shore to approximately 8 m water depth (roughly 800 m offshore). Figure 4.1 shows the experimental layout, including the locations of in-situ instruments and the footprints of remote sensing observations. These instruments were deployed by several in-situ and remote sensing research groups, who provided pre-processed data sets for use in our data assimilation system. Their methods are reviewed next.

### 4.2.2 In-Situ Data

Bathymetry at the FRF has been regularly surveyed (approximately fortnightly) since 1981 using the CRAB amphibious survey vehicle (Birkemeier & Holland, 2001), which uses RTK-GPS to obtain vertical accuracy of approximately 5 cm. A standard bathymetric survey consists of across-shore transects at roughly 50 m spacing in the alongshore direction. Each such transect extends from the subaerial beach to beyond the offshore extent of our study domain. Two surveys were performed around the time of the present experiment, on September 6 and 15. This data will be used for cross-validation of our bathymetry estimation routine in section 4.4. Additionally, we calculated a climatological bathymetry from 253 archived surveys collected between 1981 and July 2010, which we used to initialize our data assimilation system (again, see section 4.4).

Measurements of the incoming waves at the offshore boundary of our study domain ( $x = 900$  m, or about 8 m depth) are also provided by the FRF, using an array of 15 bottom mounted pressure sensors, referred to as the 8m-array (Long, 1996). These data are processed to form estimates of frequency-directional wave spectra, reported every three hours, which we use as alongshore-uniform offshore boundary conditions in our numerical model.

Additional in-situ data collected within the model domain came from three bottom mounted co-located pressure and acoustic-Doppler current profiler sensors on the transect  $y = 940$  m (Mulligan *et al.*, 2010), and a bottom mounted acoustic-Doppler current profiling instrument at  $(x, y) = (191, 714)$  m. These are used for verification of remote sensing observations.

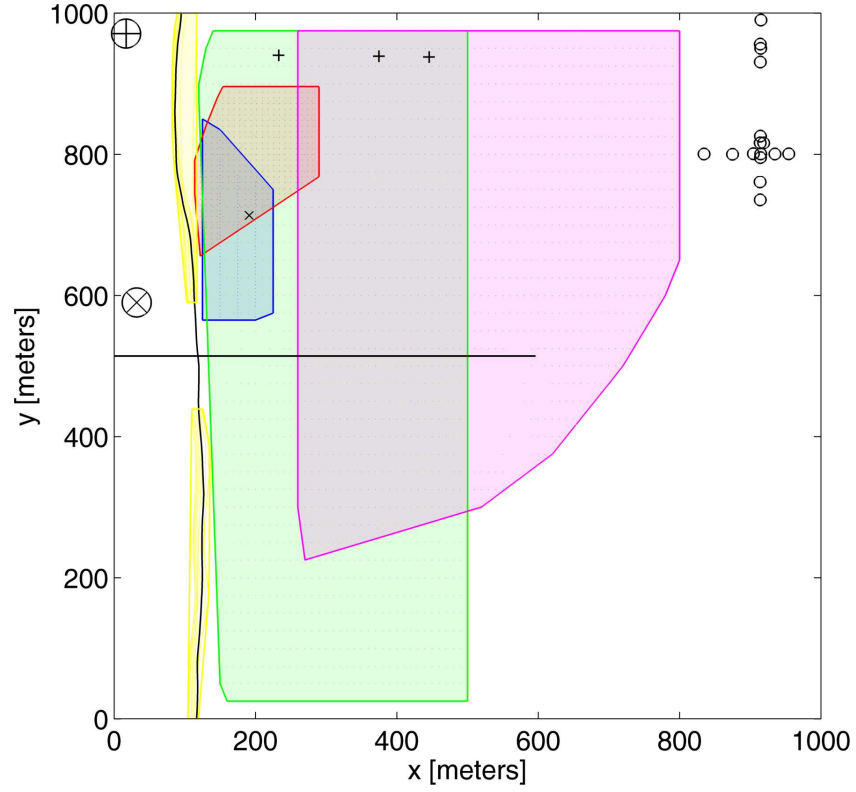


Figure 4.1: Map of 2010 field experiment, showing locations of in-situ and remote sensing data. Location of shoreline and FRF pier are plotted as black lines for reference (off-shore is in the positive- $x$  direction). In-situ data consist of 8m-array pressure gages ( $\circ$ ), co-located FRF pressure and current-profiling instruments (+), and acoustic Doppler current profiler ( $\times$ ). Remote sensing data, from Argus tower ( $\otimes$ ) and radar tower ( $\oplus$ ), consist of Optical Current Meter (blue), Infrared PIV (red), CB1-optical (green), CB1-radar (magenta), and shoreline (yellow).

### 4.2.3 Optical Remote Sensing

Optical remote sensing data was collected from an Argus station (Holman & Stanley, 2007), consisting of five video cameras mounted on a 43 m tower located at  $(x, y) = (32, 590)$  m. The combined field of view covers the full alongshore extent of the study site, and extends out to  $x = 500$  m. Pixel resolution degrades with distance from the tower, roughly 0.25–10 m. The cameras recorded data in 17 minute bursts, sampling at 2 Hz, starting every 1/2 hour during daylight hours.

#### 4.2.3.1 Shoreline Identification

The most-basic data products from the Argus system are the time-mean and variance of pixel intensity, which were saved at the top of every hour in this case. These are often used to estimate the location of submerged bathymetric features which affect wave breaking (a strong optical signal). Another use of such images is for shoreline detection, i.e., estimating the location of the zero contour in mean water level (e.g., Plant & Holman (1997)). Here, we estimate the shoreline as a function of  $y$  using the most-shoreward local maximum of image variance. An example is shown in Figure 4.2. These observations were recorded at one meter spacing in  $y$ , excluding locations within 75 m of the FRF pier where the shorebreak was often masked by waves breaking on the FRF pier pilings. Initially, estimates were extracted automatically, with fair success; however manual corrections were also used, e.g., to deal with uneven lighting or irregular wave breaking patterns. The data were also smoothed using a quadratic Loess interpolation having a alongshore window span of 100 m. Finally, a conservative error estimate of 10 m standard deviation in  $x$  was assigned for all data.

#### 4.2.3.2 Optical Current Meter

A second data product from the Argus system uses the Optical Current Meter technique (hereafter, OCM) for measuring alongshore current, as originally developed by Chickadel *et al.* (2003) (hereafter, CHF). In the present experiment, we defined five alongshore transects of pixels, spaced 25 m in the across-shore, starting at  $x = 125$  m. 17-minute timeseries from these transects were processed every 1/2 hour using a moving analysis window of width 30 m, to obtain estimates of alongshore current at 5 m alongshore resolution. To calculate the estimates, data within the analysis window were first bin-averaged to a uniform sampling resolution equal to the maximum pixel spacing



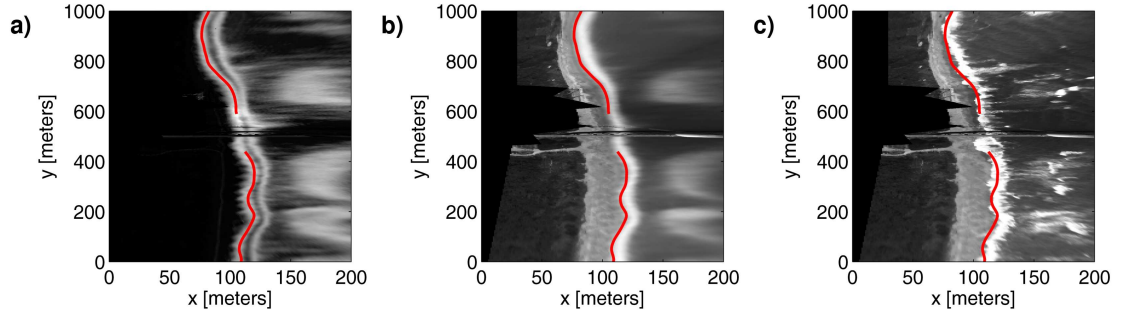


Figure 4.2: Example shoreline detected from optical imagery on September 13, 1200 EST. Shoreline (red line) is defined as shoremost maximum of variance image (a); time-exposure (b) and snapshot (c) images are also shown for reference.

in the window; the CHF algorithm then applies a 2D Fourier decomposition, and uses a parametric spectral fit to identify slowly moving features. Such features are usually associated with the alongshore drift of foam (initially generated during wave breaking), which is a proxy for alongshore current.

Uncertainty estimates for the OCM product are provided based on the spectral fitting procedure, which includes a noise model as described by CHF. Automated quality control of data also follows CHF. Additionally, periods where raindrops were present on the windows in front of the cameras occurred for 6 hours during the experiment, and those times were manually excluded (was not flagged by automated quality control). No rain occurred during the specific time period analyzed in section 4.4. Finally, we also excluded estimates for which the raw-data alongshore pixel spacing was greater than one meter.

Figure 4.3a shows an example OCM data product (red arrows) at a time with dense data coverage, overlaid on a time-exposure optical image to indicate the locations of wave breaking. For visualization purposes, this figure uses spatial smoothing of the OCM estimates to reduce the effects of noise, such that the across-shore component of current could be estimated from the continuity equation using measured water depths (this processing was not applied when assimilating the data, or when comparing to in-situ observations below). Figure 4.3a also shows time-averaged predictions from a forward numerical model (yellow arrows; model is described in section 4.3.1), using measured bathymetry from September 15. This demonstrates that the model could agree well with the remote sensing data, given accurate bathymetry. In this case the model is correctly predicting a gyre-like flow caused by a gap in the nearshore sandbar (see Figures 4.9c and 4.9d). This type of circulation was common during the experiment, and mainly

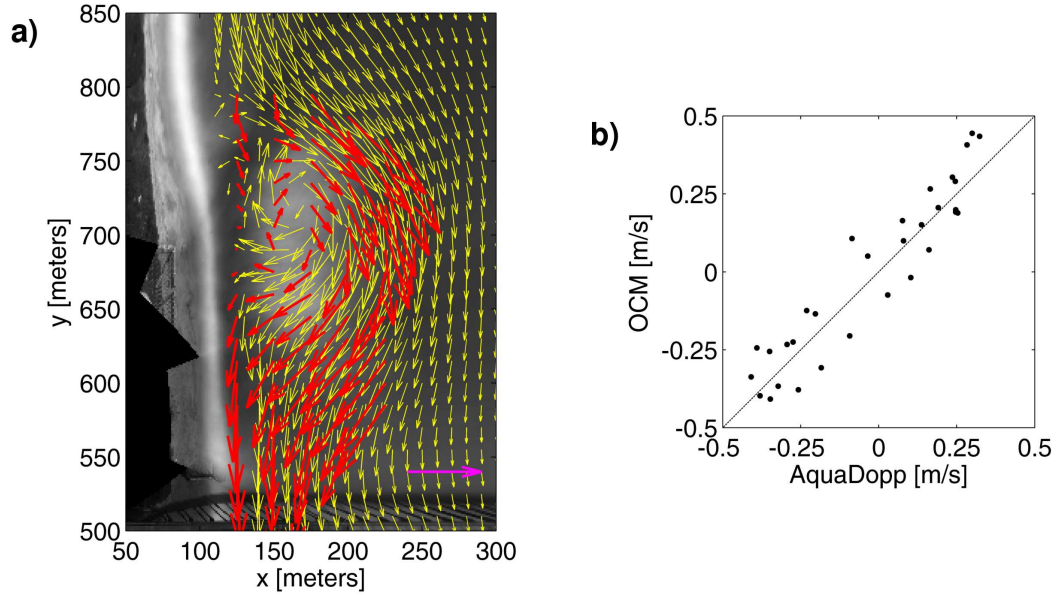


Figure 4.3: Left (a): example data product for Optical Current Meter (OCM): September 13, 0830 EST. Red arrows represent spatially-smoothed measured alongshore currents, with across-shore currents derived using continuity equation; yellow arrows are a numerical model prediction using measured bathymetry; magenta scale arrow in lower right represents 0.5 m/s; background is an Argus time-exposed image from the same time period. Right (b): comparison between OCM alongshore current, and in-situ (AquaDopp) measurements of depth-averaged current.

occurred during low tide (Haller *et al.*, 2013).

Figure 4.3b shows a comparison between OCM data and depth-averaged currents measured by an AquaDopp instrument at  $(x, y) = (191, 714)$  m (nominal depth 1.5 m). The data shown cover the period September 11–14 (OCM collections began Sep. 11, and in-situ collection ended Sep. 14). To make this comparison, quality-controlled OCM observations were averaged within a radius of 20 m around the in-situ gage, excluding cases with fewer than five such observations; 31 OCM collections passed this criteria. The in-situ data were then time-averaged over the OCM collection window. Although the data set is small, the results indicate good skill for the OCM measurements: root-mean-square (rms) error was 9.3 cm/s, and  $r^2 = 0.88$ , comparable to the verification results of CHF.

### 4.2.3.3 Wave Celerity and Direction

A third optical data product comes from the so-called “cBathy phase-one” routine (hereafter CB1) developed by Holman *et al.* (2013). The CB1 algorithm analyzes imagery downsampled to 5 m by 10 m resolution, within a movable spatial window of width 40 m (across-shore) by 100 m (alongshore). For each such window, CB1 calculates a cross-spectral matrix, retaining only the first singular vector thereof, then fits the resulting phase maps using sinusoidal waveforms at pre-specified frequency bands. This gives estimates of wavenumber (scalar) and wave angle for each band, as well as estimated uncertainty based on the fitting routine. Data are then reported for the four most coherent frequency bands. We will refer to the wavenumber product as CB1k, and the wave angle product as CB1a. The above processes is repeated at different locations by shifting the analysis window to obtain an output resolution of 10 m by 25 m, and collections occur every 1/2 hour. Automated quality control of the data follows Holman *et al.* (2013); we also increased the threshold for phase-map fit skill from 0.5 (default) to 0.75, and we excluded data for which the analysis window included subaerial (dry) points (based on the identified shoreline, section 4.2.3.1).

Holman *et al.* (2013) showed that by fitting CB1k frequency-wavenumber pairs to the linear wave dispersion relationship they could extract accurate estimates of water depth. When combined using a Kalman filter (not unlike the one used here), this information produced accurate spatial maps of bathymetry, which they verified using in-situ surveys. One minor shortcoming of their results was the Kalman filter predicted much lower bathymetric error than was observed. This may be partly due to unrealistically-small observational uncertainty in the assimilation system. The same issue was present in our system when assimilating CB1k data, and to avoid it we chose to increase the CB1k observational uncertainty by a factor of two. This resulted in similar estimates of bathymetry, but more-realistic error estimates.

The bathymetry verification by Holman *et al.* (2013) suggests the CB1k wavenumber data are accurate, in the sense they can be used to predict water depth. For additional verification, we also compared the CB1k data to wavenumber data from in-situ pressure gage measurements. The in-situ data consist of 34-minute 2 Hz timeseries, collected hourly, at locations  $(x, y) = (233, 940)$ ,  $(375, 939)$ , and  $(446, 938)$  m (see Figure 4.1). For each collection period, we performed a cross-spectral analysis between pairs of adjacent in-situ gages to extract estimates of the across-shore component of wavenumber at each CB1k frequency band. Corresponding estimates from CB1k (taking wave angle into ac-

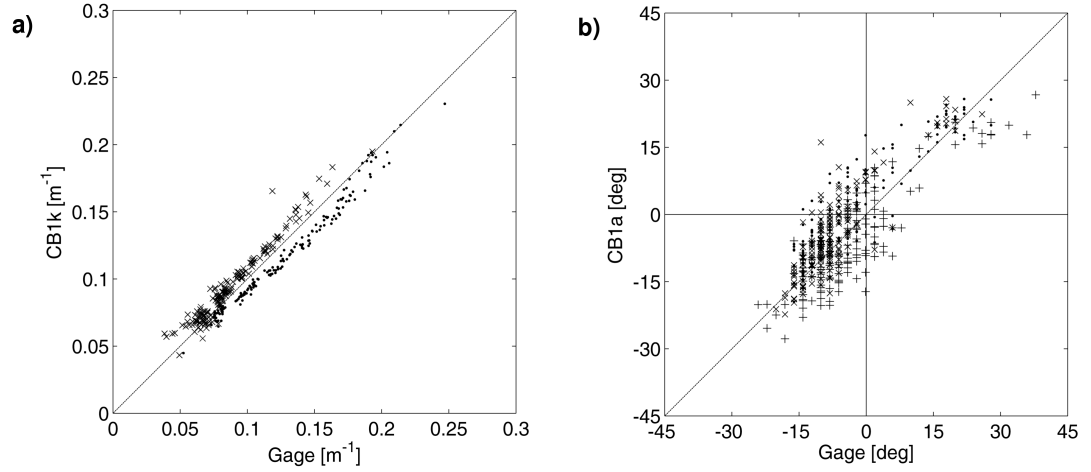


Figure 4.4: Left (a): comparison between wavenumber from optical CB1k product, and an estimate from cross-spectral analysis of in-situ pressure gage pairs. Dots represent in-situ gage pair  $(x, y) = (233, 940), (375, 939)$  m; crosses represent  $(375, 939), (446, 938)$  m. Right (b): comparison between wave angle CB1a and peak wave angle from analysis of in-situ gage data. Gages are located at  $(x, y) = (233, 940)$  m (dots)  $(375, 939)$  m (crosses), and  $(446, 938)$  m (plusses).

count using CB1a) were extracted by interpolating to the midpoints between gage pairs. Figure 4.4a shows the resulting comparison, for 75 CB1k data collections covering the time period September 9–17. The agreement is good, with overall rms-error  $0.011 \text{ m}^{-1}$  and  $r^2 = 0.93$ . Much of the error is due to an apparent bias between the in-situ and remote sensing estimates, which could be explained by reasonable synchronization offsets among the in-situ instrument clocks (1–2 seconds; K. Hathaway, personal communication). If this bias was removed, skill increased to rms-error  $0.0060 \text{ m}^{-1}$  and  $r^2 = 0.97$ .

Similarly, Figure 4.4b shows a comparison between CB1a wave angle data and observations from the same three in-situ gages as above, for the same time period (in this case 83 collections; the number of comparable data is different than for CB1k because the data locations differ). To make this comparison, wave spectra were estimated from the in-situ data as described by Hathaway & Hanson (2011), and peak wave directions then were extracted for each of the CB1a frequency bands. The CB1a data were then spatially interpolated to the gage locations. The results show CB1a represents the peak wave angle with fair accuracy: rms-error is  $6.4^\circ$  and  $r^2 = 0.71$  (note, part of the differences may also be due to uncertainty in the in-situ data). A comparison was also made to the in-situ mean wave angle using the definition of Kuik *et al.* (1988), which

resulted in worse agreement: rms-error  $8.1^\circ$  and  $r^2 = 0.67$ . The fact CB1a agrees better with peak angle rather than mean angle is likely due to the CB1 processing technique, which filters data by extracting only the first EOF mode of the cross-spectral matrix. For broad or multi-modal wave spectra, additional EOF modes may include significant information (other wave trains), but are not considered. The problem of making best use of multi-directional wave information in cBathy is a subject of ongoing research.

#### 4.2.4 Infrared Remote Sensing: Particle Image Velocimetry

An infrared video camera, similar to the system used by Chickadel *et al.* (2009), was also deployed on the Argus station tower during this experiment. This imagery was analyzed in 30 minute bursts using particle image velocimetry (PIV) to extract velocity and its uncertainty, a product we will refer to as IR-PIV. The output resolution of this product is  $8 \times 8$  m (raw pixel data has resolution of order 1 m), using analysis windows of  $16 \times 16$  m (i.e., 50% overlap), and data are reported every 1/2 hour. The incident wave signal was removed from the imagery prior to processing by extracting the minimum pixel intensity over a moving 10-second window. It is assumed the tracked features correspond to remnant foam (which is typically cooler than surface water and/or recently-generated foam), and remnant/active coherent structures, both of which are generated by wave breaking and are passively advected by mean currents.

Quality control was defined by excluding measurements for which the PIV algorithm used fewer than 30 samples within its 30 minute analysis window. Also, we excluded measurements which were within 20 m of the shoreline. Finally, although this method provides estimates of both  $x$  and  $y$  components of current, we only consider the alongshore current when assimilating data. This is because our numerical model is not designed to reproduce the stronger depth-variability expected in the across-shore current.

An example IR-PIV result is shown in Figure 4.5a, similar to Figure 4.3a. Again, we find that the numerical model agrees qualitatively well with the remote sensing data, in a case where bathymetry is accurately known. The remote sensing data also compare well with in-situ observations, as shown in Figure 4.5b (using the same in-situ data set and method of comparison as described above for the OCM data, Figure 4.3b). In this case, the comparison spanned September 9–14, and included 63 IR-PIV collections. Root-mean-square error for the IR-PIV data was 8.9 cm/s, and  $r^2 = 0.71$ .

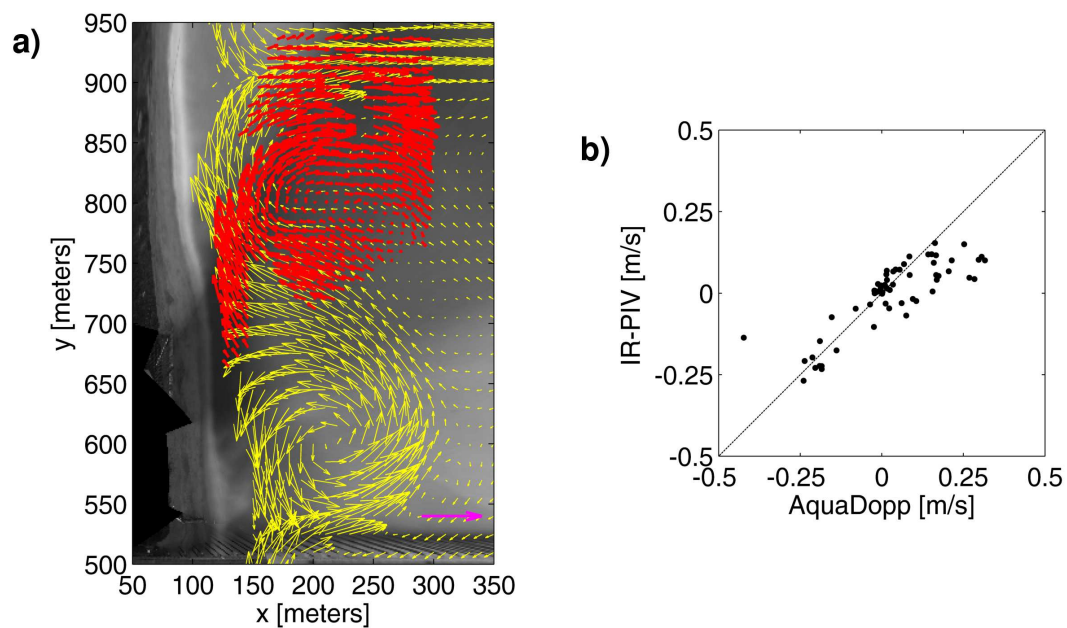


Figure 4.5: Left (a): example data product for Infrared Particle Image Velocimetry: September 13, 1800 EST. Red arrows represent measured currents; yellow arrows are a numerical model prediction using surveyed bathymetry; magenta scale arrow in lower right represents 0.5 m/s; background is a time-exposed image during the same period (unfortunately not very informative due to low light). Right (b): comparison of IR-PIV estimates of alongshore current, and in-situ (AquaDopp) measurements of depth-averaged current.

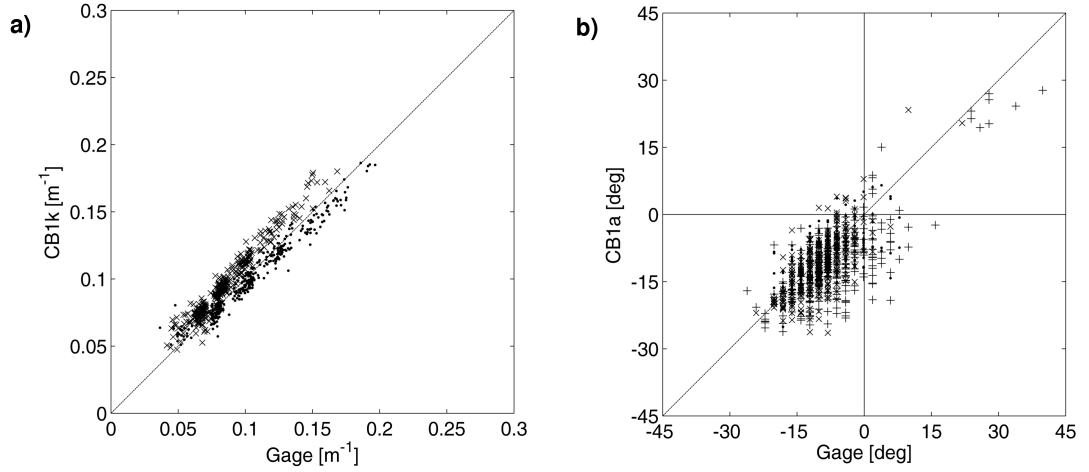


Figure 4.6: As in Figure 4.4, but for CB1-radar.

#### 4.2.5 Radar Remote Sensing: Wave Celerity and Direction

An X-band marine radar, described by Haller *et al.* (2013), was mounted at a height of 14 m during the experiment, at the location  $(x, y) = (17, 971)$  m. For this experiment, the system measured backscatter intensity in 330 range bins (total range 1 km), and 270 azimuthal bins. Data were collected at the top of each hour, and each collection consisted of 760 antenna rotations at a rate of 46 rotations per minute (17 minutes total).

Haller *et al.* (2013) have demonstrated this instrument's effectiveness for imaging waves and wave-averaged properties (e.g., locations of wave breaking, presence of rip currents) during the present experiment. The imaging mechanism is scattering from centimeter-scale sea surface roughness, for example due to wind or wave breaking; the scattering is then modulated by incident waves, resulting in a strong wave signal in the imagery. Wavefield information could therefore be extracted using the CB1 routine, as in section 4.2.3.3 (the same analysis window and output resolution was used). We will refer to this as CB1k-radar and CB1a-radar, to distinguish it from the optical products.

Figure 4.6 shows a comparison between in-situ data and CB1-radar products, where the in-situ data is calculated as in Figure 4.4. This comparison used 121 CB1k-radar collections and 128 CB1a-radar collections, spanning September 10–17. The CB1-radar accuracy appears to be comparable to the optical-based product: rms-error and  $r^2$  were  $0.010 \text{ m}^{-1}$  and 0.89, respectively, for wavenumber ( $0.0065 \text{ m}^{-1}$  and 0.95 after removing per-instrument-pair bias), and  $5.7^\circ$  and 0.47 for wave angle.

Direct comparison of the radar- and optical-based CB1 products also showed good agreement for the most part. However, we did identify a systematic bias towards smaller predictions of wavenumber by CB1k-radar, in a nearshore region roughly the equal to the extent of the surf zone. Because the radar observations are known to be of poor quality in that region compared to optical observations (M. Haller, personal communication), it was decided to discard radar-based data for  $x < 250$  m. Conversely, we noted that the radar-based product, unlike the optical product, did not require inflation of the observational uncertainty estimate in order to obtain reasonable assimilation results. Indeed, if the radar-based observation uncertainty was inflated, the result was an over-reliance on the optical data (and hence almost no added benefit from radar). It is not known why the two products required differing calibration for observational uncertainty, but we note this is a notoriously difficult quantity to estimate for remote sensing data, and is the subject of ongoing research in the data assimilation community (e.g., Desroziers *et al.* (2006); Li *et al.* (2009)).

### 4.3 Modeling and Bathymetry Inversion System

To reiterate, our goal is to predict surf zone waves and currents in an setting where bathymetry is uncertain. To that end, we utilize a combination of a forward numerical model and a data-assimilating inverse model. The overall modeling system thus tracks an uncertain estimate of bathymetry, and controls that estimate by assimilation of remote sensing data.

Our methodology follows Wilson *et al.* (2010), but includes extensions for time-varying sequential estimation and hence implements the ensemble Kalman filter (EnKF, Evensen (2006)). It also makes use of different observational data, and a slightly different physical model. In overview, the basic procedure is as follows:

1. Define an initial background ensemble consisting of 200 realizations of bathymetry (section 4.3.3).
2. For each member of the ensemble, apply the hydrodynamic numerical model (section 4.3.1) with fixed boundary conditions for the target observation time.
3. Define the observational data set for the target time (section 4.2) and its uncertainty (section 4.3.6), and extract corresponding predictions from the ensemble (section 4.3.5).



4. Apply the EnKF update equations with state augmentation (section 4.3.4), to assimilate the observations and thereby obtain an updated ensemble of bathymetry.
5. Adjust the ensemble spread to account for unresolved sediment transport and potential shortcomings of the filter, and resample to replace any failed ensemble members (section 4.3.7).
6. Move to the next observation time, and repeat from step 2.

After a sufficiently long period of time, multiple assimilation cycles should refine the ensemble of bathymetry such that its distribution represents an improved state of knowledge given the observations. The mean of the ensemble will represent the maximum likelihood estimate of the true bathymetry, and the covariance will represent the expected uncertainty of that estimate. This will be tested in section 4.4.

### 4.3.1 Forward Model

We begin by defining the forward model, which represents transformation and breaking of waves as they enter shallow water (using linear wave theory), and the subsequent transfer of momentum from waves to depth- and time-averaged currents, as represented by the radiation stress (Longuet-Higgins & Stewart, 1964). This basic description of surf zone dynamics was first introduced by Longuet-Higgins (1970b) and Bowen (1969). Physical parameterizations for unresolved processes such as turbulence and wave dissipation have since reached a fairly high level of maturity, and the resulting predictive models have been validated extensively in field and laboratory settings (e.g., Ruessink *et al.* (2001); Haas *et al.* (2003)). We describe our specific implementation next.

In our application, we use SWAN (Booij *et al.*, 1999) to simulate incident waves, and the Regional Ocean Modeling System (ROMS, Shchepetkin & McWilliams (2005)) to simulate time-averaged currents. Both models will be described in detail below. The model domain extends from the 10 cm depth contour (i.e., the shoreline,  $x \approx 100$  m) to an offshore boundary  $x = 900$  m, and for an alongshore span  $-105 < y < 1410$  m. Model grid spacing is 10 m (across-shore) by 15 m (alongshore). The domain is assumed to be periodic in the  $y$ -direction, and the area  $1110 < y < 1410$  m is used as a buffer zone over which the bathymetry is smoothly ramped to satisfy periodicity (the buffer zone is reapplied each time bathymetry is updated via data assimilation). The model is re-initialized from rest for each observation period; continuous simulation

would be problematic, because the model bathymetry changes abruptly each time data are assimilated.

The wave part of the model, SWAN, solves the stationary conservation of wave action equation (Mei, 1983), which governs the transformation of wave energy (i.e., the frequency-directional wave spectrum) from the offshore boundary to the shoreline. Offshore boundary conditions are specified for a given observation time using the FRF 8m-array measurements, which are assumed alongshore-uniform and are interpolated in time using the internal SWAN routine. Energy dissipation due to wave breaking is included using the parameterization of Battjes & Janssen (1978), with default physical constants in SWAN. The effect of currents on waves (i.e., wave-current interaction) is not included, so that SWAN runs as a standalone model.

The wave spectral predictions from SWAN are then used to compute radiation stress gradients (Longuet-Higgins & Stewart, 1964), which are passed as a static input to ROMS. The effect of wave rollers, the aerated mass of water riding on top of breaking waves which acts as a reservoir for momentum (Svendsen, 1984), is included following Reniers *et al.* (2004).

ROMS, in turn, solves the Reynolds-averaged hydrostatic Navier Stokes equations, which are also averaged in depth and in time, over the time scale of waves (this time averaging produces the radiation stress gradient terms noted above). The model is allowed to spin-up for seven hours, and then model outputs are averaged over 30 minutes to simulate an observational data collection period. Bottom stress is parameterized following Svendsen & Putrevu (1990), with a drag coefficient  $f_w = 0.0053$  chosen based on an analysis of field data on this beach by Feddersen & Guza (2003) for a similar, though not identical, bottom stress formulation. Surface stress is assumed to be due to wind only, and is modeled using the parameterization of Smith (1988), using wind measurements from the offshore end of the FRF pier and assuming nominal values for air temperature,  $10^\circ \text{ C}$ , and density,  $1.22 \text{ kg/m}^3$ . Horizontal momentum mixing is modeled using an eddy viscosity, parameterized for the surf zone following Haas *et al.* (2003). Tides are included as a static adjustment to the water level, as measured by a tide gage at the end of the FRF pier (note, this implies the sub-tidal depth is what is estimated by assimilation of data, rather than the total water depth). The shoreline boundary condition is no-slip, applied at the 10 cm depth contour, and a radiation condition is applied at the offshore model boundary.

### 4.3.2 Mathematical Statement of Inverse Problem

Next we introduce the inverse model, which incorporates bathymetric uncertainty into the forward model, and then attempts to control bathymetric error (hence estimate bathymetry) using observations.

In order to formally define the inverse problem, we begin by introducing some notation and nomenclature. First, we define a state vector  $\psi$ , consisting of a concatenation of all the relevant variables in the model (i.e.,  $h, u, v, k$ , etc.), including bathymetry, for all model gridpoints. The dimensions of  $\psi$  are therefore  $MV \times 1$ , where  $M$  is the number of model gridpoints and  $V$  is the number of model variables. We also define a vector of observations,  $d$ , having dimensions  $K \times 1$ , and a  $K \times MV$  matrix  $L$  which serves to map  $\psi$  to the observation space. For example, if  $d$  comprises a list of observations of the velocity  $u$  at specific locations, then  $L\psi$  represents interpolation of the model  $u$  to those locations.

In general, the number of degrees of freedom contained in the observations is far less than that of the model, so that estimation of  $\psi$  (or even a subset of  $\psi$  such as bathymetry) using  $d$  alone is a mathematically under-determined problem. For that reason, any method which seeks to estimate bathymetry from sparse observations must introduce regularizing assumptions (extra information) to constrain the inverse problem. In the usual formulation of data assimilation, this information is given explicitly in the form of a prior or “background” model state  $\psi^b$  and its covariance  $C_b$ . These represent the best possible estimate of the model state, and the uncertainty of that estimate, before observations are taken into account. Given this information, the problem becomes overdetermined, and one can seek a generalized inverse by minimizing a least-squares cost function:

$$J[\psi] = \left(\psi - \psi^b\right)^T C_b^{-1} \left(\psi - \psi^b\right) + (d - L\psi)^T C_d^{-1} (d - L\psi). \quad (4.1)$$

In this cost function, the first term penalizes departure of the estimate of  $\psi$  from the background  $\psi^b$ , and the second term penalizes misfits to the observations. These are weighted using the matrices  $C_b$  and  $C_d$ , which represent the expected covariance of errors (i.e., uncertainty) in the background and the observations, respectively.

The solution to this minimization problem gives an updated estimate of  $\psi$  and its covariance, for a given set of observations. In our application, we then extract the bathymetry from this updated state, and use it to form the background for the next observation time. This is described in more detail in the following sections.

### 4.3.3 Covariance Modeling Using Ensembles

A central aspect of any data assimilation system is the method used to construct the background  $\psi^b$  and  $C_b$  given a specified distribution of uncertain parameters, in this case bathymetry. It is easy to see that given a background bathymetry  $h^b$ , plus boundary conditions (which we assume are known), one could run a forward numerical model to produce  $\psi^b$ . Calculation of  $C_b$ , on the other hand, is not straightforward. The approach we use here follows Evensen (2006), as implemented by Wilson *et al.* (2010) for surf zone bathymetric inversion.

For the first assimilation step, we generate  $N = 200$  realizations of bathymetric perturbations, using the Fourier Transform method described in Evensen (2006) (Fortran code available from [enkf.nersc.no](http://enkf.nersc.no)), which draws from the covariance

$$C_h(\Delta x, \Delta y) = \sigma_h^2 \exp \left[ -3 \left( \frac{\Delta x^2}{L_x^2} + \frac{\Delta y^2}{L_y^2} \right) \right]. \quad (4.2)$$

In the present experiments we choose  $L_x = L_y = 100$  m, and  $\sigma_h = 0.5$  m, representative of the presumed typical length scales of unknown bathymetric features. These perturbations are then added to a prescribed initial background estimate of bathymetry  $h^b$  (e.g., see section 4.4) to form a bathymetric ensemble. We then execute the forward model (section 4.3.1) for each member of the ensemble, resulting in an ensemble of full model state vectors which we will denote  $\psi_i^b$ . The sample mean of this latter ensemble is used for  $\psi^b$ , and the sample covariance is used for  $C_b$ . After the first assimilation step, the ensemble and its covariance evolves via assimilation of data (section 4.3.4), and equation (4.2) is no longer used.

A common issue with ensemble-based covariance approximations is the potential for spurious long-range spatial correlations in the estimated  $C_b$ . Hence, following Hamill *et al.* (2001), we localize all sample covariances using a Schur product with a compactly-supported correlation function. The correlation function used here is the same as used by Hamill *et al.* (2001), with a length scale of 75 m. This yields a cutoff separation distance of roughly 150 m beyond which all covariances are effectively set to zero.

We also note the present approach does not constrain realizations of  $h$  to be strictly positive. This is physically acceptable, as negative water depth can simply be interpreted as dry land. However, problems arise if negative depths occur in locations where waves or currents were measured. In that case, it is not possible to “measure” the ensemble as required in the assimilation process (section 4.3.5). To circumvent this issue, we

define a rule that only one zero-crossing of still-water-depth may occur for any given  $y$  location, meaning the shoreline is a single-valued function of  $y$ , and no “islands” are allowed. To enforce this rule, depths are truncated to a minimum of 0.25 m at all points offshore of the first zero-crossing. In cases where truncation would cause a change in depth of more than 0.5 m, the realization is completely removed from the ensemble. If after these changes a given observation is still not measurable across all of the ensemble members (which occurred for some CB1 data near the shoreline), that observation is removed from the assimilation process. For the experiment of section 4.4.2, the above rules caused depths to be truncated in 4.1% of realizations, 0.5% of realizations had to be discarded, and 3.5% of the available CB1 data was discarded.

#### 4.3.4 Update Step

With the above definitions, we are now prepared to assimilate data and update the model state by minimizing the cost function (4.1). A formal minimization (e.g., Evensen (2006); Bennett (2002)) results in the following equation for the updated (or “analysis”) state:

$$\psi^a = \psi^b + C_b L^T (L C_b L^T + C_d)^{-1} (d - L \psi^b + e), \quad (4.3)$$

where the superscript “ $\dagger$ ” indicates a Moore-Penrose matrix inverse, which accounts for possible conditioning problems when the number of observations is larger than the ensemble size (Evensen & van Leeuwen, 1996). Equation (4.3) is applied to each member of the prior ensemble, producing an updated ensemble of state vectors (with updated sample covariance), from which we extract the updated bathymetry ensemble,  $h_i^a$ . Note we must introduce random measurement perturbations  $e_i$  for each member, with mean zero and covariance  $C_d$ , in order for updated ensemble to have the correct covariance (Houtekamer & Mitchell, 1998). Also note in practice we usually forego the computation of the entire state update and instead compute only the rows of  $\psi^a$  which correspond to the bathymetry  $h^a$ .

Equation (4.3), the EnKF update equation, is the core of the assimilation system. At this point, however, there are a few remaining details regarding our particular implementation, which we describe next.

### 4.3.5 Observation Operator

Recall the observation operator  $L$  serves to “measure” the model state  $\psi$ , and hence produce model predictions corresponding to the observations  $d$ . In practice,  $L$  is not specified explicitly, rather each observed variable is extracted from the model (i.e., measured) using a set algorithm, described below. This reveals a minor abuse of notation in equation (4.3) in cases where the measurement process is nonlinear. In actual fact, in our system  $L\psi_i^b$  is defined as  $L(\psi_i^b)$ , where  $L$  is a function (possibly nonlinear) which maps to the observation space. Similarly,  $C_b L^T$  is defined as the sample covariance (including localization) between  $\psi_i^b$  and  $L(\psi_i^b)$  (and similarly for  $LC_b L^T$ ). The function  $L$  is described next for each observation type.

Measurements of currents are treated simply by linearly interpolating the predicted  $u$  and  $v$  from the model grid to the observation locations. This is the simplest of observation operators, because the forward model already outputs  $u$  and  $v$  explicitly.

Wavenumber measurements are defined using a sub-model for wave dispersion, applied as a function of depth, waves and currents. Following the recommendation of Catalán & Haller (2007), we use the dispersion relationship of Kirby & Dalrymple (1986) (hereafter KD86),

$$\left(\sigma - \vec{k} \cdot \vec{u}\right)^2 = gk \left[1 + f_1 \epsilon^2 E \tanh(kD + f_2 \epsilon)\right], \quad (4.4)$$

where

$$\epsilon = \frac{kH}{2}, \quad E = \frac{8 + \cosh 4kD - 2 \tanh^2 kD}{8 \sinh^4 kD},$$

$$f_1(kD) = \tanh^5 kD, \quad f_2(kD) = \frac{kD}{\sinh kD}^4.$$

In these equations,  $\sigma$  is the radial wave frequency,  $\vec{k}$  is the wavenumber,  $H$  is the wave height, and  $D$  is the mean water depth. The KD86 model includes the effect of currents on waves, as well as the effect of finite wave amplitude. Note these effects are not included in our implementation of SWAN, hence the use of KD86 in the inversion step is somewhat ad-hoc; however, we found that if they were not included, the resulting estimate of bathymetry was biased towards being too deep. Another aspect of the wavenumber measurement is the fact the remote sensing measurements are computed over a large spatial footprint. To account for this, all variables in KD86 are averaged over the same footprint before applying the equations.

Measurements of the location of the shoreline are defined by interpolating the  $x$ -locations of the modeled shoreline to the  $y$ -locations of the observations. An alternative would be to treat the measurement as an observation of zero depth at the measured  $(x, y)$ ; however, in that case the observation uncertainty would be difficult to define, as the sensor does not actually “observe” depth. A related point is that the shoreline measurements in  $L(\psi^b)$  do not have a well-defined  $(x, y)$  location (because  $x$  is treated as an uncertain observable). This leads to ambiguity when applying localization to the sample covariances; we chose to localize based on the modeled mean  $x$ -location of the shoreline (an alternative would be to use the observed  $x$ -location).

Measurements of wave angle were the most difficult to represent. Recall (section 4.2.3.3), the CB1 algorithm filters its observational data by extracting the leading EOF, which is likely (though not guaranteed) to be associated with the most-energetic wave train. With that in mind, we extract frequency-directional spectra from the wave model, integrate with respect to frequency over each CB1a frequency band, then extract the peak wave angle in each band. This procedure was repeated for each of the CB1a output locations, and results were then averaged over the CB1a analysis windows.

#### 4.3.6 Observation Error Covariance

To define  $C_d$ , we begin by noting that its main diagonal corresponds to the estimated error variances for the observations, which are already given by the remote sensing data analysis (section 4.2). In addition, however, we must take into account the fact the remote sensing data are derived using analysis windows, and those analysis windows can overlap, which would imply spatial correlation of the observation error covariance (i.e.,  $C_d$  should also include off-diagonal terms). If this fact was not accounted for we would effectively assimilate the same information multiple times.

We therefore model the full observation error covariance  $C_d$  as block-diagonal, with each block representing a particular data product (that is, errors between different data products are assumed uncorrelated). The  $i$ 'th block of  $C_d$  (representing one data product) is modeled by

$$C_{di}(\Delta x, \Delta y) = w \Sigma \exp \left[ -3 \left( \frac{\Delta x^2}{L_{xi}^2} + \frac{\Delta y^2}{L_{yi}^2} \right) \right] \Sigma^T + (1 - w) \Sigma \Sigma^T, \quad (4.5)$$

where  $\Delta x$  and  $\Delta y$  are the separation between observations,  $L_{xi}$  and  $L_{yi}$  are the analysis window half-widths in the  $x$  and  $y$  directions,  $\Sigma$  is a diagonal matrix containing the

observation error standard deviations, and  $w = 0.9$  is a weighting factor. In the case of the shoreline observations (the only data product which does not have a clearly-defined analysis window) a representative analysis window half-width of 30 m was used, based on the alongshore smoothing function that was applied to the measurements. Additionally, CB1 observation errors from different frequency bands are assumed to be uncorrelated (likewise for CB1-radar), as are OCM observation errors from different alongshore transects.

The weighting factor  $w$  allows us to hedge our estimate of  $C_d$  towards a more-traditional diagonal matrix, which helps to ensure  $C_d$  is well-conditioned — note the more the measurement analysis windows overlap, the less independent are the rows of  $C_d$ , which influences the condition of the matrix. The choice of  $w$  was also used as a rudimentary calibration parameter for  $C_d$ . With  $w = 0$  (diagonal  $C_d$ ), we found the ensemble spread was strongly underestimated, and corrections to bathymetry were amplified, sometimes at the expense of accuracy; we interpret this as overfitting the observations. With  $w = 1$ , we found  $C_d$  was not as well conditioned, and estimates of some bathymetric features were overly smeared-out. The choice  $w = 0.9$  was a compromise between those two extremes.

#### 4.3.7 Ensemble Resampling and Covariance Inflation

The bathymetric ensemble obtained from the update equation (4.3) forms the basis for a new background ensemble for the next assimilation time. However, recall our methodology includes the possibility of excluding ensemble members with unacceptable depth variations (e.g., large islands); similarly, in rare cases certain realizations induce numerical instability in the forward model and therefore must be excluded (this occurred for 0.13% of all realizations in the experiment of section 4.4.2). Over time, these restrictions could lead to unacceptable shrinking of the ensemble size. Hence, we choose to resample the ensemble after each update, producing a new ensemble of 200 members with conserved sample mean and covariance. This resampling is calculated using code from the EnKF-Matlab package by P. Sakov (available at <http://enkf.nersc.no/Code/EnKF-Matlab/enkf-matlab-0.30.tar.gz>).

Lastly, before proceeding to the next observation time, the spread (uncertainty) of the ensemble is inflated using additive random noise. This is done for two reasons. First, our forward model does not account for time-evolution of bathymetry by sediment transport, hence this must be incorporated as additional uncertainty (similar methods



were used by van Dongeren *et al.* (2008) and Holman *et al.* (2013)). Effectively, this is a simple statistical “forward model” representing bathymetric change. Second, the assimilation method used here is known to be sub-optimal (e.g., due to nonlinearity in the forward model and observation operator, and finite ensemble size), which can lead to underprediction of ensemble spread. In either case, if inflation is not used one risks developing unrealistically-small ensemble spread, such that the true bathymetry is no longer a viable realization. This can cause a situation known as filter divergence, where the filter is so “certain” of an inaccurate model state that it effectively ignores new observations. The problem of filter divergence is well-known in geophysical data assimilation, and inflation is used routinely to mitigate it (e.g., Hamill *et al.* (2001)).

Construction of the additive inflation noise is primarily based on accounting for unknown bathymetric change due to sediment transport. This was quantified for our field site by Holman *et al.* (2013), who suggest the following empirical formula for the growth rate of variance, based on in-situ field measurements of bathymetric change:

$$Q(x, H_{m0}) = C_Q H_{m0}^2 \exp -\frac{(x - x_0)^2}{\sigma_x^2} . \quad (4.6)$$

Here,  $x$  is the across-shore coordinate,  $C_Q = 0.067 \text{ days}^{-1}$ ,  $H_{m0}$  is significant wave height measured at the 8m-array,  $x_0 = 150 \text{ m}$ , and  $\sigma_x = 100 \text{ m}$  (the values of  $x_0$  and  $\sigma_x$  reflect the typical location of breaking waves at this particular beach). By integrating  $Q$  between successive assimilation cycles (see section 4.2) we obtain a baseline envelope for the inflation noise variance. Next, we set thresholds on this envelope such that, once the inflative noise is added to the ensemble, the ensemble spread (standard deviation) will be between 0.25 m and 0.75 m. The use of a lower bound on variance follows the “conditional covariance inflation” methodology used for parameter estimation by Aksoy *et al.* (2006), and is intended to avoid filter divergence. The use of an upper bound is to avoid unbounded growth of ensemble spread in sparsely-observed regions (Hamill & Whitaker, 2005). After applying these thresholds, the square-root of the variance envelope is multiplied by unit-variance random noise, using covariance as in equation (4.2) with  $L_x = L_y = 100 \text{ m}$ . This noise is then added to the bathymetry ensemble before moving to the next observation time.

## 4.4 Results

Next, we apply the modeling/assimilation system described in section 4.3 to the observational data set described in section 4.2.

### 4.4.1 Experiment Setup

Figure 4.7 shows the conditions observed during the experiment, and Figure 4.8 shows the number of observations reported by each remote sensing data product. Based on this data set, we chose to test our modeling system by assimilating data on September 13, for an 11.5 hour period during daylight hours (0700–1830 EST). During the development of our system we also tested other times and longer assimilation windows, and generally found similar results (although, see section 4.5.1). The main criteria for selecting a targeted test period was that all data products were consistently reporting with good spatial coverage, allowing for a fair comparison between the assimilation of different observation types (section 4.4.4). On other days, unavoidable factors caused one or more data products to be poorly represented, such as mismatched data start/end times (e.g., OCM data not available before September 11) or weather conditions (e.g., rain, which occurred on September 12 during otherwise favorable conditions). Another consideration was that wave heights on September 13 were sufficient to cause consistent breaking over the nearshore bar/terrace throughout the day, which drove significant surf zone currents. On days with smaller wave heights, currents were sometimes driven by other forces not represented by the forward model, resulting in errors which will be discussed in section 4.5.1.

As shown in Figure 4.7, combined significant wave height during the test case varied from 0.7–1.1 m. Wave spectra were somewhat complex, consisting of at least two sea components and a weaker swell component. The observed dynamics were spatially nonuniform, primarily due to the presence of nonuniform bathymetric features (see surveyed bathymetry in Figures 4.9c and 4.9d). Most notably, a rip current was observed during low tide at  $y \approx 800$ –900 m, apparently due to a gap in the nearshore bar/terrace which caused a nonuniform wave breaking pattern. This rip current is of particular interest, because it was in the field of view of all of the remote sensing instruments. For instance, it can be seen in the example data shown in Figures 4.3a and 4.5a.

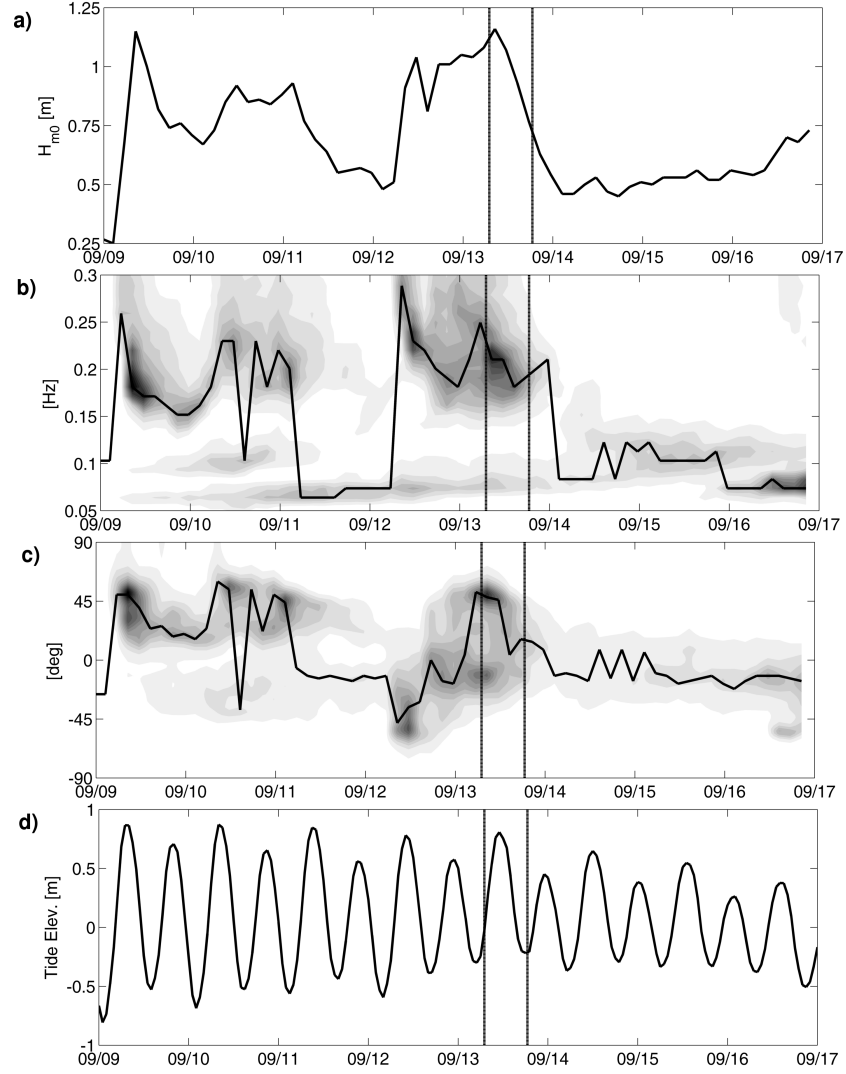


Figure 4.7: Conditions observed during experiment, vs. time in EST. The start/end time of the September 13 test case presented in section 4.4 is marked by vertical lines. Plots (a)–(c) show observed wave conditions in 8 meters depth: significant wave height  $H_{m0}$ , and integrated frequency and directional wave energy density (shown as shading in (b) and (c), normalized to unit energy for each time), with peak frequency and direction marked by solid lines. Wave directions are measured counterclockwise from the positive- $x$  axis to the direction waves are coming from. Plot (d) shows observed tidal elevations relative to the NAVD88 vertical datum.

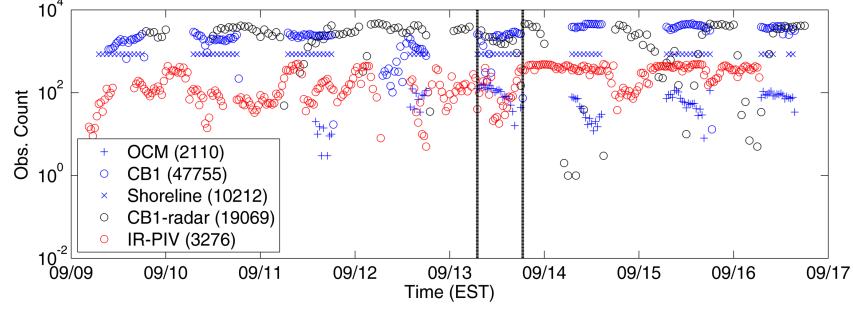


Figure 4.8: Number of remote sensing observation data points, vs. time in EST. The start/end time of the September 13 test case presented in section 4.4 is marked by vertical lines. Legend indicates the type of observation; the total number of observations during the September 13 test period is listed in parentheses.

#### 4.4.2 Estimated Bathymetry

To initialize the data assimilation system, we defined a highly simplified background bathymetry based on 29 years of bathymetric surveys at the field site (253 surveys in total). This data was merged/interpolated to the model grid using a linear Loess interpolator (Plant *et al.*, 2002) having length scales of  $l_x = 20$  m and  $l_y = 200$  m. The resulting background bathymetry is shown in Figure 4.9a. Uncertainty for this background bathymetry,  $C_h$ , was initialized as described in section 4.3.3. Note this background estimate includes almost no information pertaining to the actual bathymetry during the experiment, except for the presence of mild scour under the FRF pier, and the approximate beach slope; this ensures that any subsequent corrections to bathymetry can be clearly attributed to information in the assimilated data, rather than user-specified prior knowledge.

Figure 4.9b shows the estimated bathymetry after assimilating the 11.5 hour test data set (24 observational data cycles). For this test, we assimilated all available remote sensing data products (see section 4.2 and Figure 4.8), with the exception of wave angle observations. The reason for excluding wave angle observations is they were found to degrade the bathymetry estimate; this will be discussed in section 4.4.4.

The accuracy of the estimated bathymetry can be assessed qualitatively by comparing to the survey data in Figures 4.9c and 4.9d. These show a nearshore bar at  $x \approx 200$  m, which migrated onshore over time to form a more terrace-like nearshore feature. The bar/terrace was also incised with several channels ( $y \approx 100, 250, 900$  m) which caused alongshore nonuniform wave breaking patterns and rip currents. There was also

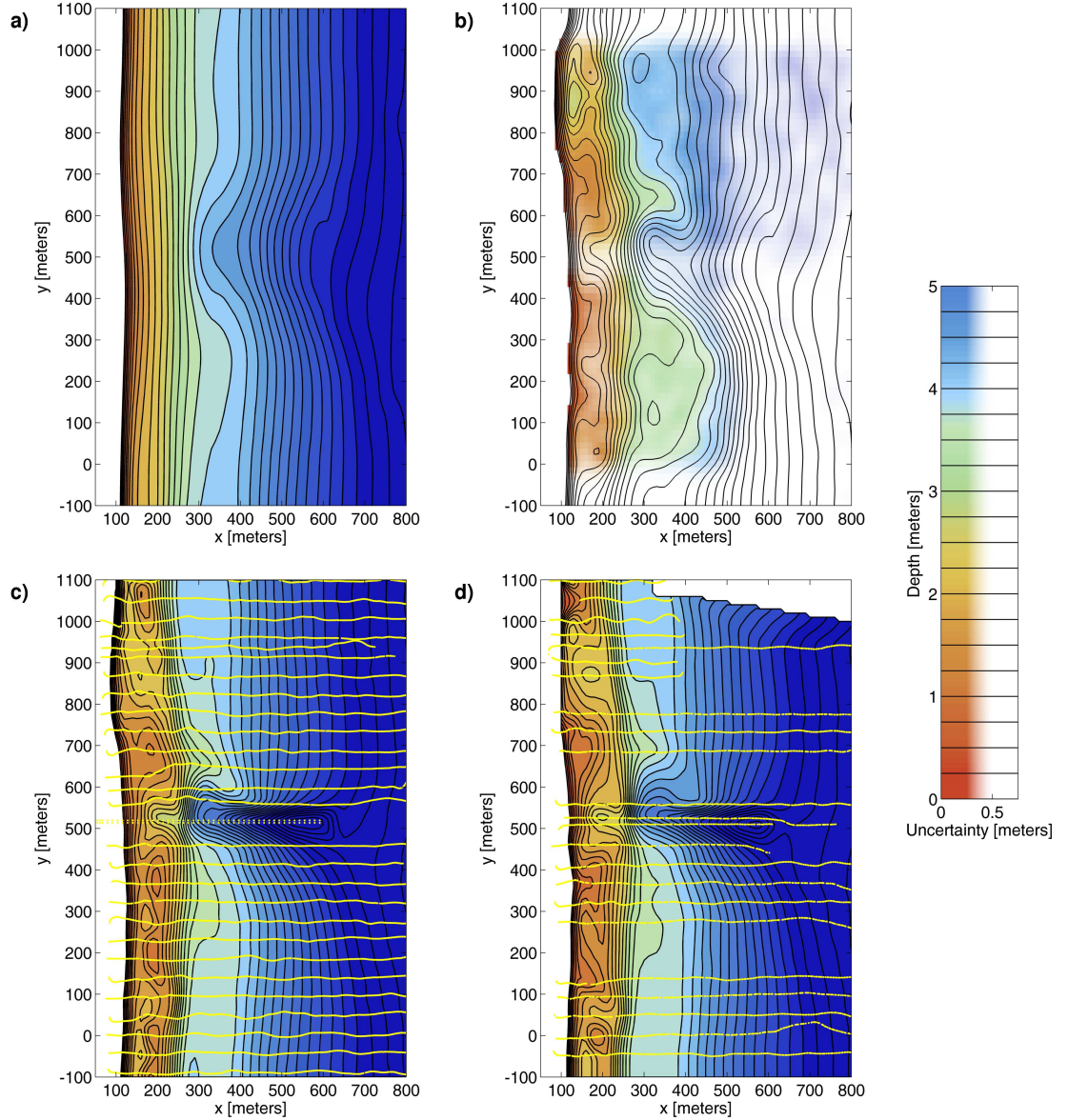


Figure 4.9: Plot (a): background bathymetry  $h^b$  used to initialize the ensemble assimilation system: a climatological average of bathymetry observed at the field site. Plot (b): updated bathymetry after assimilating all observations except wave angle, over 24 assimilation cycles (ending 1830 EST). Plots (c,d): surveyed bathymetry from September 6 (c) and September 15 (d), where individual survey data points are plotted as yellow dots, and interpolated depths are shown as colors (white area in upper-right of Sep. 15 bathymetry is due to a gap in the data). Colorbar with contour marks at far right applies for all plots, and refers to depth relative to the NAVD88 vertical datum; in the case of the updated bathymetry (b), the same colors and contours are used, but color transparency is scaled to represent the posterior estimate of bathymetric uncertainty.

a prominent trench at  $y \approx 500$  m, which is a persistent feature at this site due to scour around the FRF pier pilings. While none of these qualitative features existed in the initial background bathymetry (Figure 4.9a), all are fairly well represented in the final estimated bathymetry (Figure 4.9b). This is further illustrated using individual across-shore and alongshore transects in Figure 4.10. A transect of the estimated bathymetry over the nearshore bar/terrace (red line in Figure 4.10a) shows that assimilation of data correctly captures the location and approximate amplitude of rip channels, in good agreement with the survey data from September 15 (black line). An across-shore transect at  $y = 690$  m shows that the across-shore profile of the terrace is also fairly well captured, as is a secondary bar at  $x \approx 350$  m; another across-shore profile at  $y = 870$  m also shows good agreement, including a corrected shoreline location, although in that case the inner bar location is mis-predicted.

Figure 4.11 shows differences between the raw survey data and the estimates of bathymetry before and after data assimilation. Positive values in this plot represent overestimates of depth. Despite ambiguity as to which survey should represent the “truth” for September 13, both surveys indicate the bathymetry estimate is generally improved by the assimilation of data (also cf. Table 4.1 in section 4.4.4). For example, the initial estimate did not include a bar/terrace, resulting in the initial bathymetry being overly deep for approximately  $150 < x < 250$  m, an error which was largely corrected by data assimilation.

A region of low skill, on the other hand, occurred in the south part of the domain offshore of the surf zone, roughly  $0 < y < 400$  m and  $250 < x < 500$  m, where the system estimated overly-shallow depths. Note only one observation type, wavenumber, was assimilated in this region (see Figure 4.1). Inspection of the data showed CB1k measured wave celerities were indeed consistent with such shallow depths (based on linear wave dispersion), and the alternative data assimilation method of cBathy Phases 2–3 (Holman *et al.*, 2013) produced similar results. CB1k-radar data were sparse in this region, but the few data points that were available showed larger wave celerity than measured by CB1k (i.e., consistent with larger depths). Hence, we suspect low skill in this region was due to an isolated problem with observational (CB1k) data quality, not with the data assimilation method.

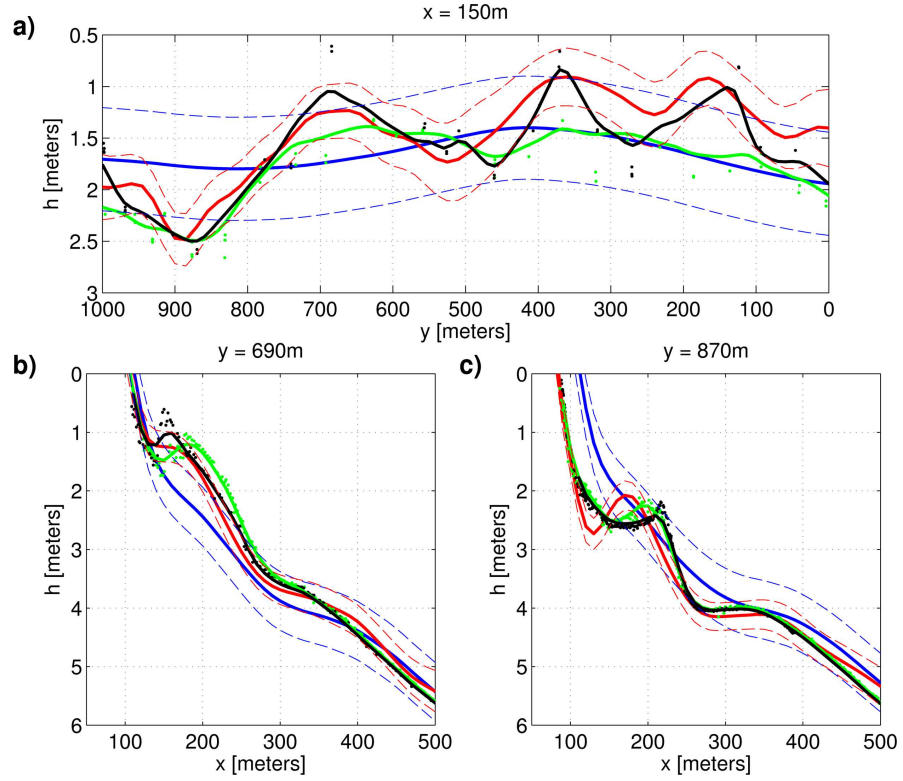


Figure 4.10: Transects from initial (blue) and final (red) estimated bathymetry (as in Figures 4.9a and 4.9b, respectively), and  $\pm$  one standard deviation (dashed lines), compared to measured bathymetry on September 6 (green) and September 15 (black). Alongshore transect (a) is from  $x = 150$  m, located over the nearshore terrace; across-shore transects (b,c) are from  $y = 690, 870$  m. Dots represent measured data from within 2 meters of the alongshore transect in plot (a), and within 10 meters of the across-shore transects in plots (b,c); green and black lines represent smoothed 2D interpolation of raw survey data, as in Figures 4.9c and 4.9d.



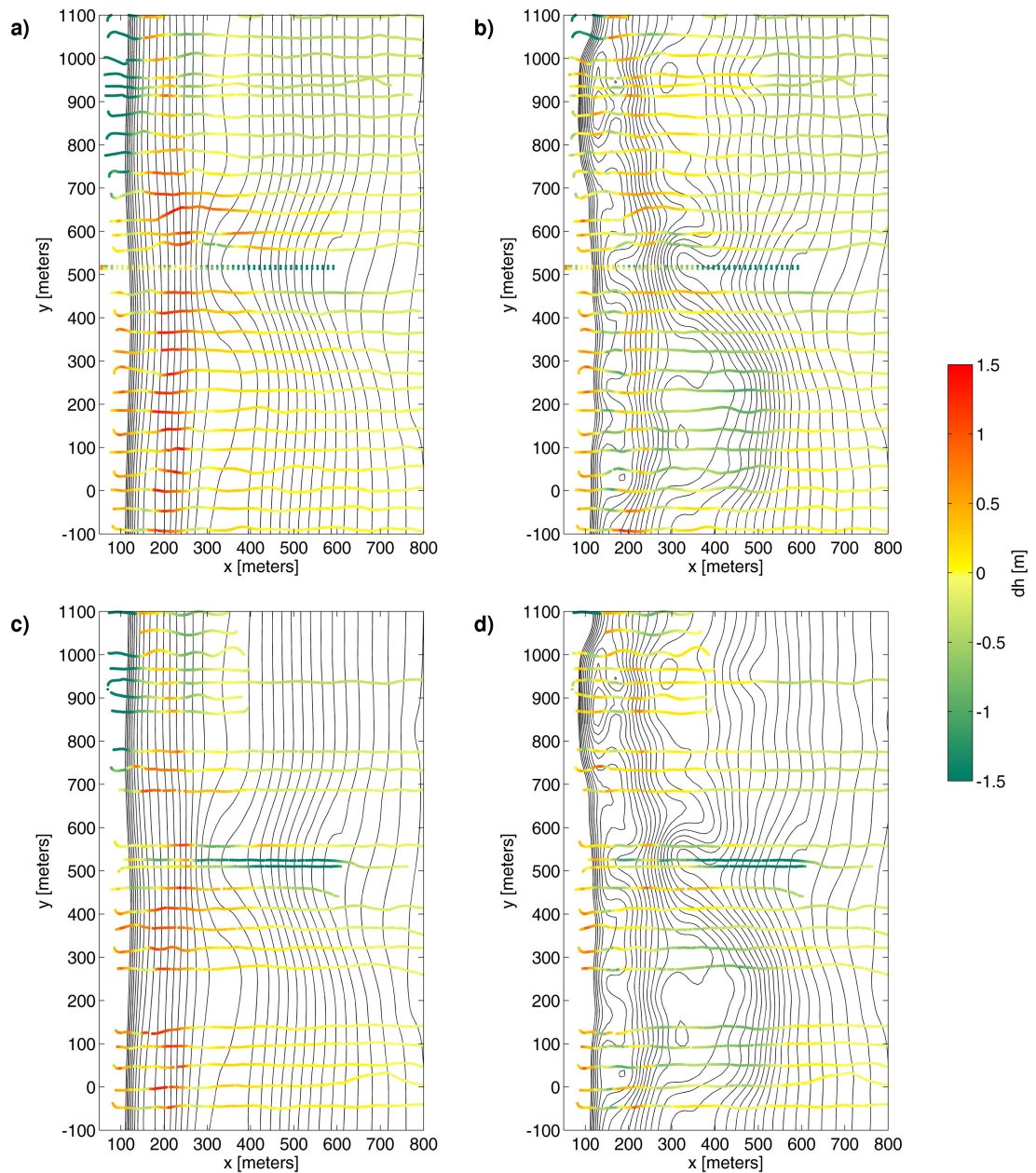


Figure 4.11: Difference between estimated bathymetry and raw bathymetric survey data. Top plots (a,c) use data from the September 6 survey, showing differences before (a) and after (b) data assimilation; bottom plots (c,d) are the same, but using data from the September 15 survey. Positive differences indicate an overprediction of depth. Black contours are the estimated depths as in Figure 4.9.



### 4.4.3 Improved Prediction of Currents

An important aspect of the assimilation system is its ability to improve overall model skill as a result of assimilating observations and correcting bathymetry. To demonstrate this, we test the ability of the system to predict a persistent rip current which appeared throughout the experiment during low tides, at  $y \approx 900$  m, coincident with a gap in the nearshore bar/terrace. The rip was well-imaged by IR-PIV measurements, and was also visible in time-averaged radar backscatter imagery (Haller *et al.*, 2013).

On the day of our test case, observations did not indicate the presence of the rip current at  $y \approx 900$  m until 1500 EST, at which point it appeared and persisted until the end of the day. Hence, to test our system we consider assimilating data within the time period 0700–1200 EST (i.e., before the rip was observed), then running the model forward, using the same boundary conditions but without assimilating data, to predict currents at low tide, 1800 EST. The reason 1200 EST was chosen as a cutoff was that near that time a reversal of currents was observed at  $y \approx 700$  m, perhaps a precursor to the formation of the rip current. It also should be noted that the 1200 EST bathymetry estimate (not shown) already contained the qualitative bathymetric features described above for the final estimate shown in Figure 4.9b, and estimates were not significantly different for other nearby times.

Figure 4.12 shows the resulting prediction of currents at 1800 EST. If no data were assimilated, the bathymetry remained nearly alongshore uniform (that is, as in Figure 4.9a), and no rip current was predicted (not shown). After assimilating the 0700–1200 EST data, the model was capable of predicting the rip current in roughly the correct location (Figure 4.12a). Hence, assimilation of data resulted in the prediction of a bathymetry-controlled rip current, without the use of any direct bathymetry observations. An even more accurate prediction was obtained if data were also assimilated during 1200–1800 EST (Figure 4.12b).

As an aside, note we do not expect the forward model to be capable of predicting the trajectory of the rip current once it exits the surf zone (roughly  $x > 200$  m in this case), regardless of the accuracy of bathymetry. In that region, there is likely to be a strong influence from wave-current interaction (Haas *et al.*, 1998; Yu & Slinn, 2003) as well as 3D aspects of circulation (Haas & Svendsen, 2002), neither of which is included in our forward model. The prediction of the trajectory in Figure 4.12b appears accurate, but was found to be sensitive to small details in the bathymetry and so may simply have been a coincidence. On the other hand, the ability to predict the location of the rip

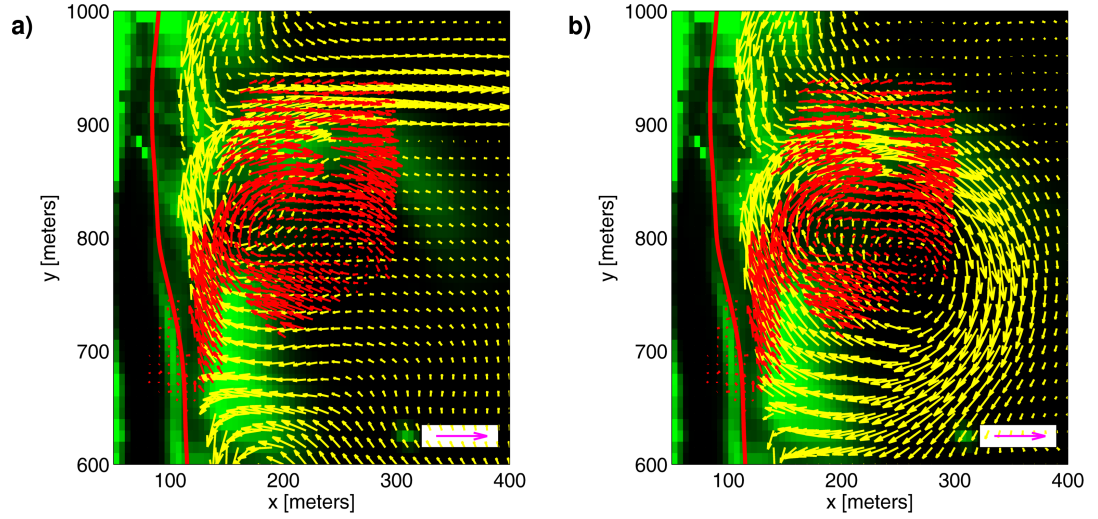


Figure 4.12: Measurements and forward model predictions for September 13, 1800 EST (low tide). Background shading represents X-band radar backscatter averaged over a 17 minute collection period; high backscatter (green) corresponds to breaking waves, and/or surface roughness due to the presence of a rip current (Haller *et al.*, 2013). Red arrows are IR-PIV measurements of currents (scale arrow in lower-right is 50 cm/s), which confirm the presence of a rip current at  $y \approx 900$  m. Red line is optical-based shoreline position. Yellow arrows are predicted time-averaged currents from a forward model run using (a) the estimated bathymetry after assimilation of data from 0700–1200 EST, before any rip current was observed, and (b) the estimated bathymetry after assimilation of data up to and including 1800 EST.

inside the surf zone (after assimilating data) is within the expected capabilities of the model, and was a robust result.

#### 4.4.4 Observation Impact and Quantitative Skill Assessment

To judge the impact of each individual observation type on the estimated bathymetry, we conducted a series of data-denial experiments, in which only one observation type was used in addition to shoreline observations. The reason shoreline observations were always included was that the initialized shoreline (i.e., from the climatological average bathymetry, Figure 4.9a) was further offshore than the true shoreline; if the shoreline location was not corrected, many observations fell on “dry land” in the model and hence could not be assimilated.

Figure 4.13 shows the final bathymetry estimate (after 24 assimilation cycles) for each observation type, including a case where only shoreline data were assimilated. With the exception of wave angle observations, discussed in more detail below, assimilation of data generally produced qualitative improvement in the bathymetry estimate in the region where observations were available (see Figure 4.1). In areas where there were no observations, assimilation has less of an effect and the estimated uncertainty is correspondingly large.

Table 4.1 presents skill statistics for the bathymetry estimates from the various data-denial experiments presented in Figure 4.13, as well as for the full assimilation test, Figure 4.9b, and the case with no assimilation, Figure 4.9a. Skill is assessed by comparison to raw data from each of the bathymetric surveys (Figures 4.9c and 4.9d), for two different sub-regions: (A) the region where both wave and current observations were available (i.e., union of red and blue polygons in Figure 4.1), and (B) the combined region spanned by all of the observations (nearly the entire model domain). The statistics consistently show that assimilation of either wavenumber or alongshore current produced a quantitative improvement in bathymetric accuracy, relative to the initial estimate with no assimilation. Wavenumber observations produced a more accurate estimate than did alongshore current observations, which we attribute to two factors. First, the relationship between wavenumber and bathymetry is more clear-cut, via the wave dispersion relationship (equation (4.4)). Second, the density of wavenumber observations far exceeded that of currents (see Figure 4.8); this is because waves are nearly always visible in the remote sensing imagery, and can be analyzed at multiple frequencies for each location, whereas observations of currents rely on tracking of ephemeral image features. Our

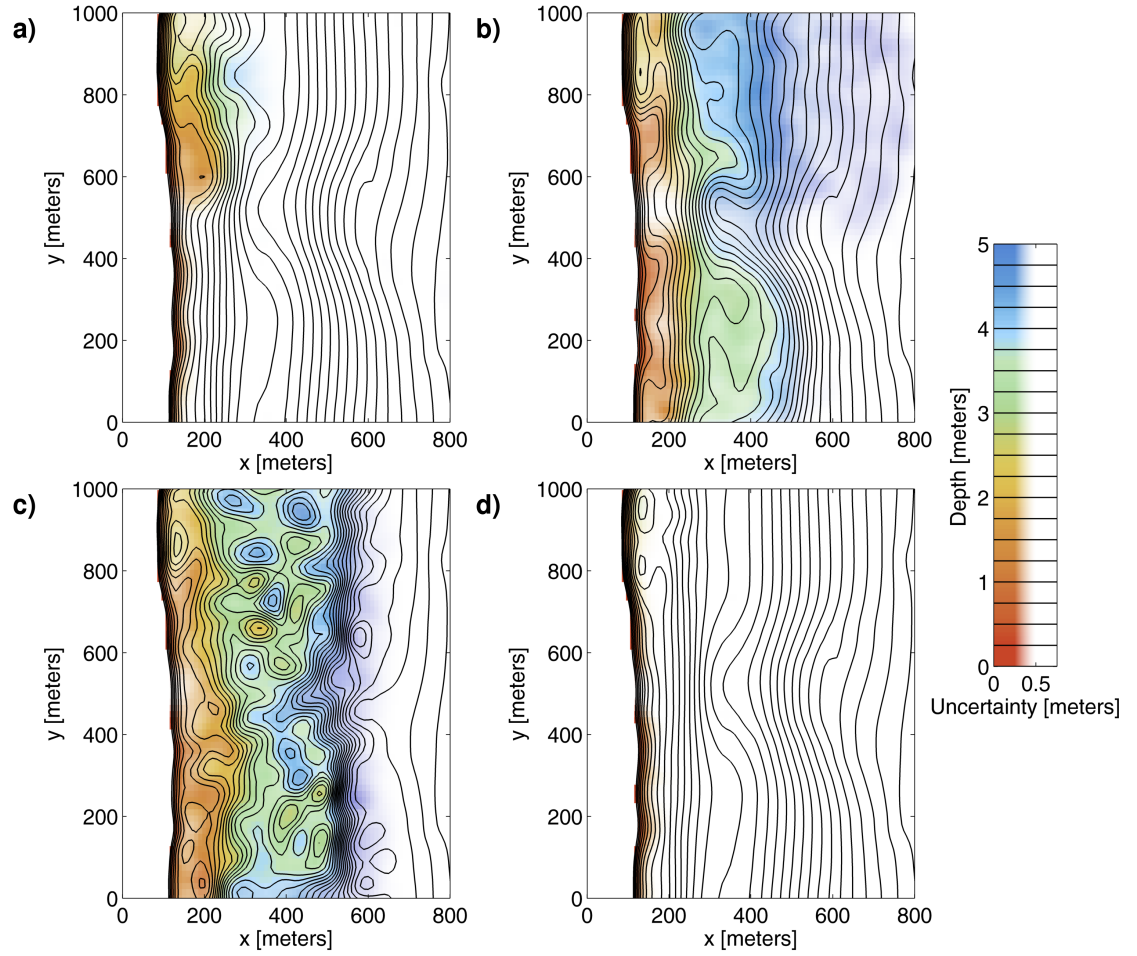


Figure 4.13: Estimated bathymetry from data-denial experiments (see text), plotted as in Figure 4.9b. Each result is from assimilating shoreline observations and one other observation type: (a), alongshore currents (OCM and IR-PIV); (b), wavenumber (CB1k and CB1k-radar); (c), wave angle (CB1a and CB1a-radar); (d), shoreline only.

Obs. Type	Sep. 6 Survey				Sep. 15 Survey			
	Region A		Region B		Region A		Region B	
	$\epsilon$ (cm)	$r^2$	$\epsilon$ (cm)	$r^2$	$\epsilon$ (cm)	$r^2$	$\epsilon$ (cm)	$r^2$
Initial	57	0.49	51	0.93	55	0.68	65	0.88
(a) Current	35	0.81	41	0.96	43	0.85	52	0.93
(b) Wavenumber	33	0.81	39	0.96	36	0.86	48	0.94
(c) Wave Angle	46	0.58	79	0.88	48	0.75	86	0.85
(d) Shoreline-Only	53	0.74	45	0.95	59	0.81	53	0.93
All	27	0.86	38	0.97	36	0.86	49	0.94

Table 4.1: Skill statistics, comparing September 6 and September 15 raw bathymetric survey data (Figures 4.9c and 4.9d, resp.) to estimates from data-denial experiments presented in Figure 4.13 (labels (a)–(d) are following Figure 4.13), as well as the initial estimate before data assimilation (“Initial”, Figure 4.9a) and the estimate when assimilating all observational data except wave angles (“All”, Figure 4.9b). The statistic  $\epsilon$  is defined as the root-mean-square difference between surveyed and estimated bathymetry, and  $r^2$  is the squared correlation. Both statistics are computed over two regions: Region “A” is the area where both alongshore current and wavenumber were observed (union of red and blue polygons in Figure 4.1); Region “B” is the union of all observed areas (union of all polygons in Figure 4.1).

final comment on Table 4.1 is that assimilation of wavenumber, alongshore current, and shoreline observations together (row labeled “All”) produced accuracy similar to that of assimilating wavenumber observations; this may be again partly due to the sheer number of wavenumber observations used, which would cause the estimate to be dominated by that observation type.

Another interesting data-denial experiment involves the effect of spatial coverage on the estimation of bathymetry from alongshore currents: Figure 4.14 compares the results when assimilating OCM currents vs. IR-PIV currents (again, in addition to shoreline data). In this comparison, we note the observed region coincides with the region presented in Figure 4.12, where a gap in the nearshore bar/terrace caused a rip current to occur. It happened that the IR-PIV data coverage was largely over the gap itself, whereas the OCM data was concentrated slightly to the south, over the bar/terrace. Hence in the case where only the IR-PIV data were assimilated (Figure 4.14b) the assimilation system would have had fewer observations over the bar/terrace. In that situation, the system apparently obtained a fit to the observations by creating a deep channel at  $y \approx 900$  m. The OCM assimilation run, on the other hand, obtained a fit to observations by creating a “bump” at  $y \approx 700$  m (Figure 4.14a), where data were readily available. If both OCM

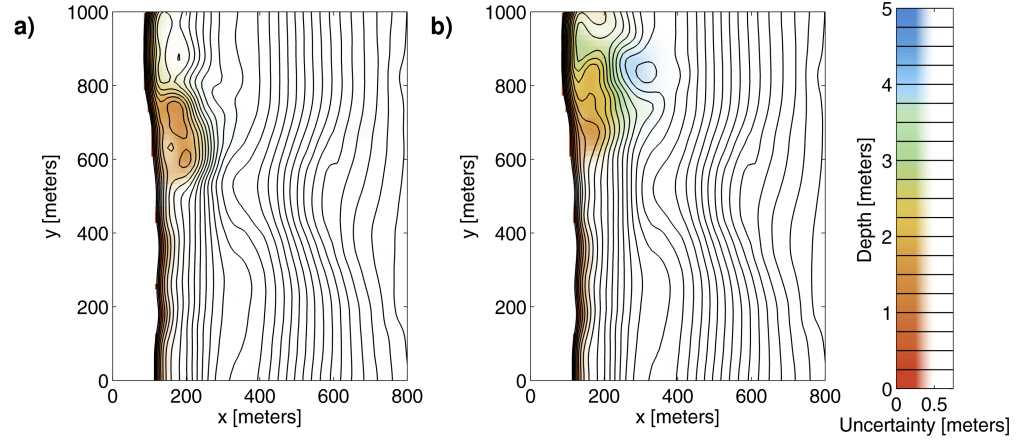


Figure 4.14: As in Figure 4.13, but assimilating shoreline plus (a) OCM currents, and (b) IR-PIV currents. Note the different observational footprint of the two products, and the resulting differences in estimated bathymetry.

and IR-PIV were assimilated together (Figure 4.13a) the estimate included both the bump and a more-realistic channel, essentially an average of the two individual results. Surprisingly, despite their differences all three of these bathymetry estimates resulted in the prediction of a low-tide rip current at  $y \approx 900$  m, similar to Figure 4.12. This illustrates the complex and indirect relationship between bathymetry and currents, and underscores the importance of data coverage when attempting to invert that relationship.

Lastly, we turn to the least skillful estimate of bathymetry, which was from wave angle observations, Figure 4.13c. In that case, we found that for assimilation cycles early on in the test period (not shown) some aspects of the estimated bathymetry appeared to be related to actual bathymetric features. However, as more data were assimilated gross perturbations began to appear, which removed nearly all skill. We attribute this to the fact the incident wavefield contained multiple directional components (see observations in Figure 4.7c), resulting in an unstable observation as discussed in section 4.3.5. That is, we found SWAN would often report the angle of one wave train, whereas CB1a and/or CB1a-radar would observe the angle of a different wave train at the same frequency. This caused severe unexpected differences between model and data, and hence resulted in severe errors. Moreover, error in the forward model may have also played a role here, as the assimilation of CB1a places a high demand on the detailed wave spectral transformation predicted by SWAN. To assess both of the above issues, we tried replacing the CB1a data with predictions of peak wave angle from a forward model run with measured bathymetry, using the same frequencies, locations, and estimated observation

uncertainty as in the real CB1a data set. The estimate of bathymetry had more skill in this twin-test environment, which indicates model error (the inability of the model to predict the CB1a representative wave direction) may have been important.

## 4.5 Discussion

The above results show how an EnKF-based data assimilation system can be used to estimate bathymetry as an uncertain parameter in a nearshore model for waves and currents, by assimilating remote sensing data alone. Next, we discuss some potential shortcomings and pitfalls of the system.

### 4.5.1 Effects of Model Error

An important consideration in the application of the assimilation system is the potential for unaccounted-for errors in the forward model. In our system, we assume the model physics and boundary conditions are perfect, and model error is entirely due to errors in bathymetry. However, this is not always the case. One exception occurred in the 2010 experiment during high tide on September 11, at which time a 30–40 cm/s along-shore current was observed in the in-situ and remote sensing data, despite minimal wave breaking and wind (winds on this day were less than 5 m/s). This current was observed even in 8 m depth, and was presumably caused by larger-scale processes (Lentz *et al.*, 1999) which are not included in our model. When alongshore current observations were assimilated in this case, the system obtained a fit to the observations by producing a spurious nearshore bar. This error persisted until later that day, when waves heights increased and offshore currents weakened, such that nearshore currents were once again driven by wind- and wave-induced forcing as assumed by the model. With the model error thus reduced, the system gradually re-corrected the spurious nearshore bar and regained skill by September 12.

Another potentially important source of model error is due to offshore wave boundary conditions. In the present experiment, we have relied on in-situ observations of frequency-directional wave spectra from a highly-accurate (and unique) observational array. In many realistic applications, however, boundary conditions would be derived from larger scale wave forecasts, which are error-prone. We did attempt to assimilate data when using forecasted boundary conditions, but this was not successful. The forecast was produced specifically for this experiment, and is described in Appendix C. It is considered

representative of a typical regional wave forecasting system. Unfortunately, however, during our test case (September 13) the majority of wave energy was generated locally by a small storm, whose wind speeds were underpredicted by the atmospheric component of the forecasting model. As a result, waves and currents in our nearshore model were also underpredicted. Our system does not account for such biases, which resulted in errors in the predicted bathymetry when assimilating data. Specifically, the system predicted a deep channel extending out from  $y \approx 900$  m, which was apparently required to fit the observations of the rip current during low tide. We conclude that boundary condition error remains a barrier to “in-situ-free” nearshore prediction. An interesting avenue for future work would be the development of a data assimilation system which corrects for both bathymetric error (as in the present work) and boundary condition error. For an example of the latter problem, correcting boundary condition errors in a nearshore wave model using data assimilation, we refer the reader to the work of Veeramony *et al.* (2010).

#### 4.5.2 Representation of Posterior Uncertainty

An aspect of the system which we have not chosen to focus on, but is important to discuss, is its ability to track bathymetric uncertainty as data are assimilated. Unfortunately, we noted a tendency to underpredict this uncertainty (as compared to errors based on bathymetric survey data), similar to the results reported by Holman *et al.* (2013). In our case, we chose to combat this by using conditional additive inflation of the ensemble spread (section 4.3.7). This inflation was generally applied in regions near shore with dense observational data coverage. In other words, the estimated bathymetric uncertainty in such regions was held close or equal to our specified minimum value of 0.25 m (after about five observation cycles), and in that sense it was not truly “dynamically” updated over time.

We attribute this undesirable behavior to the following factors. First, we have not attempted to account for errors in the forward model (e.g., section 4.5.1), which would tend to cause underprediction of uncertainty (Houtekamer *et al.*, 2008). Second, there is the possibility that the observational error covariance is poorly specified (this is a difficult problem for remote sensing data). Our attempt to include spatial correlation in observation errors (section 4.3.6) did reduce the tendency to underpredict bathymetric uncertainty, but did not fix the problem completely.

To summarize: longer term applications of this method would likely benefit from further calibration of the filter. This would include refining the method used for covariance



inflation, adding a systematic representation of model error, and improving the specification of observational error covariance. The methods used here (see sections 4.3.6 and 4.3.7) are a first attempt, but more sophisticated methods do exist (e.g., Dee (1995); Houtekamer *et al.* (2008); Li *et al.* (2009)).

### 4.5.3 Computational Efficiency

A potential shortcoming of our system is the computational cost associated with a large ensemble (in our case 200 members) of forward model runs.

We ran our tests using an array of 2.67 GHz Intel Xeon processors, each of which could execute 12 simultaneous forward model runs in roughly 13 minutes. Assimilation updates were calculated using the Matlab Parallel Toolbox, and took roughly 5 minutes. In total, then, our system was capable of assimilating data within the timeframe of the 30 minute observation collection cycle. The majority of runtime was spent in spinning-up the circulation model, hence runtime could have been significantly reduced by using a more clever model initialization.

Another factor in computational efficiency is the number of ensemble members required to obtain an accurate result. To assess this in the full system would require repeated tests with increasingly large ensemble size, which would be impractical. Instead, we modified our system to assimilate only CB1k observations, and to use only a single SWAN model run to represent wave height (and assuming zero current) in equation (4.4). This eliminated the need for an ensemble of wave and circulation model runs, and hence greatly reduced runtime. Figure 4.15 shows the convergence of bathymetry estimated by this simplified system for the test case described in section 4.4.2, by comparing results with increasingly large ensemble sizes to a reference run having 300 members. Note even for the smallest ensemble size tested, 50 members, the effect on the bathymetry estimate is not excessive (order 10 cm). In fact, the estimated bathymetry was not qualitatively different for any of the ensemble sizes tested. With that in mind, we then tested the assimilation of all available data (except wave angle), in the full system using a 50-member ensemble; the result is shown in Figure 4.16, which should be compared to the 200-member result shown in Figure 4.9b. The 50-member estimate still includes basic bathymetric features, although it appears to be prone to error at short length scales.

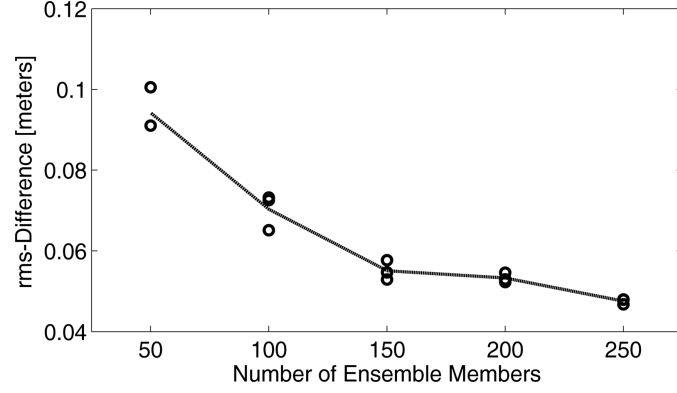


Figure 4.15: Root-mean-square difference in estimated bathymetry for various ensemble sizes in a simplified assimilation system (see text), compared to a reference run with 300 ensemble members. In each case, differences are computed over the domain  $50 < x < 500$  m and  $0 < y < 1000$  m, and over each of the 24 assimilation cycles. Three realizations (indicated by circles) were performed for each ensemble size, to account for differences due to the stochastic nature of the filter. Note, the rms-difference computed for two runs with 300 ensemble members each was 4.8 cm, which defines an approximate “noise floor” for the statistic.

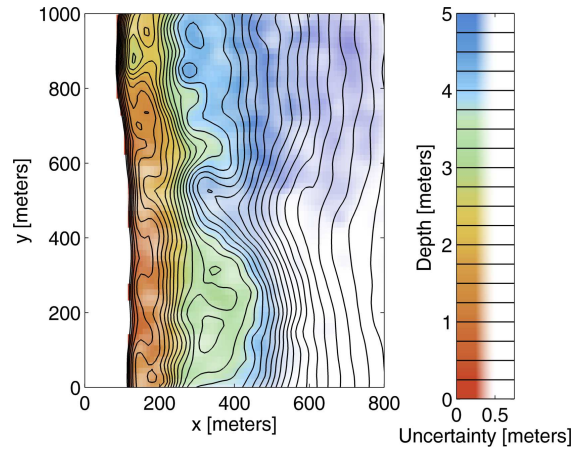


Figure 4.16: Bathymetry estimate as in Figure 4.9b, but using 1/4 ensemble size (50 members rather than 200).

## 4.6 Summary

The present work demonstrated a new application of the ensemble Kalman filter (EnKF) to the problem of surf zone bathymetric uncertainty. The method was applied to a test case spanning 11.5 hours of remote sensing observations collected at Duck, NC. We showed that assimilation of wavefield observations (frequency-wavenumber pairs, i.e., wave celerity), circulation observations (alongshore current), and shoreline observations led to an improved estimate of bathymetry. After assimilating data, the model became capable of predicting an observed surf zone rip current, without the use of any in-situ bathymetry observations.

An important feature of the EnKF method is the ease with which it can be extended for assimilation of new geophysical data types, and for new physical processes in the forward model, without the need to redefine the assimilation system itself. This contrasts with existing methods such as that of van Dongeren *et al.* (2008), who relied on explicit knowledge of the derivative of the observable with respect to depth (effectively, an adjoint model), or Holman *et al.* (2013), who converted the observations to depth estimates prior to assimilation. The EnKF has no such requirements, which is beneficial when incorporating/testing the assimilation of new and novel observation types such as remotely sensed time-averaged currents or wave angle. The ability of this method to assimilate currents (which, to our knowledge, is unique for this type of application) may be a particular advantage in environments where observation coverage varies between different observation types. An example would be a coastal inlet, where wavefield observations would dominate near shore, but observations of currents would dominate within the inlet itself.

Likewise, the ability to assimilate multiple geophysical variables allowed us to assess the relative usefulness of each variable with regard to bathymetry inversion. In this experiment, frequency-wavenumber observations were available in high density over a broad field of view, and those observations were most successful for estimating bathymetry. Assimilation of alongshore current observations was also successful, although there was evidence that the bathymetry was not uniquely determined unless the data had good spatial coverage. Wave angle observations were not successfully assimilated in this case, due to an inability to properly represent the observations using the numerical model.

The use of remote sensing data to help control errors in a surf zone model is attractive due to the difficult nature of in-situ observation. We have shown that using remote sensing data alone we may overcome significant model errors caused by uncertain bathymetry.

This suggests the possibility of an operational nearshore forecasting system which does not rely on in-situ data for its inputs. We found, however, that a remaining barrier to such an application would be the influence of errors in model boundary conditions, in particular the accurate specification of waves at the offshore boundary.

## 5. General Conclusion

Recalling the Introduction, our primary motivation for studying nearshore and fluvial hydrodynamics is its broad relevance to many important engineering and scientific applications. Taking inspiration from Saint-Venant’s work, we have not attempted to introduce new physics to this problem, but rather to attack a key gap in the existing approaches. In particular, we have focused on the role of uncertain bathymetry in the application of predictive models, an issue which is pervasive in real-world nearshore and fluvial environments, but has heretofore not been directly addressed. We introduce a new application of data assimilation for this problem, providing a way to incorporate bathymetric uncertainty in model predictions, and to control that uncertainty using available observations. Chapters 2–4 give three examples of this approach, which reflect three key conclusions, outlined next.

*Conclusion A: bathymetry can be a dominant source of error in surf zone models.* In Chapter Two, we showed a case where bathymetry was surveyed extensively on a daily basis, yet unresolved rapid bathymetric change led to significant model uncertainty and error. This was confirmed by the fact that, once data were assimilated, the model located an improved solution which was consistent with in-situ point observations of both bathymetry and waves/currents. That is, the errors were indeed largely due to bathymetry, not other deficiencies in the model.

*Conclusion B: bathymetry can be estimated by exploiting the relationship between bathymetric uncertainty and model error.* Chapter Two explored this concept in a controlled observational setting (using single-time data from an in-situ array); we described how our data assimilative model represents uncertainty among multiple state variables, and demonstrated how this information is used to produce an updated bathymetry estimate. Chapter Three also focused on this aspect of the problem, using measurements of currents to estimate bathymetry as an unknown parameter (i.e., the “inverse problem”).

*Conclusion C: data assimilation can be used in a surf zone forecasting model, to control errors due to uncertain bathymetry, without the requirement of direct in-situ observation.* In Chapter Four, we demonstrated an application of our method using surf zone measurements from shore-based remote sensing platforms. Assimilation of these measurements resulted in the prediction a rip current without the use of any in-

situ bathymetry observations. This suggests an interesting application of our method for operational prediction and forecasting, where the inability to specify bathymetry is often cited as a limiting factor.

Overall, data assimilation has been shown to be an effective tool for understanding and manipulating bathymetric uncertainty in a general shallow water setting. It is our hope that these results will improve the ability to predict natural flows, and will help elucidate the role of uncertainty in those predictions.

## Bibliography

- Aarninkhof, S.G.J., Ruessink, B.G. & Roelvink, J.A. (2005). Nearshore subtidal bathymetry from time-exposure video images. *Journal of Geophysical Research*, **110**(C06011).
- Aksoy, A., Zhang, F. & Nielsen-Gammon, J.W. (2006). Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Monthly Weather Review*, **134**, 2951–2970.
- Allard, R., Dykes, J., Hsu, Y.L., Kaihatu, J. & Conley, D. (2008). A real-time nearshore wave and current prediction system. *Journal of Marine Systems*, **69**, 37–58.
- Amante, C. & Eakins, B.W. (2009). *ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, National Geophysical Data Center, Marine Geology and Geophysics Division.
- Andreadis, K.M., Clark, E.A., Lettenmaier, D.P. & Alsdorf, D.E. (2007). Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model. *Geophysical Research Letters*, **34**, 10403.
- Austin, M.J., Scott, T.M., Russell, P.E. & Masselink, G. (2012). Rip current prediction: Development, validation, and evaluation of an operational tool. *Journal of Coastal Research*, **29**, 283–300.
- Baringhous, L. & Franz, C. (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis*, **88**, 190–206.
- Barré Saint-Venant, A. (1871). Théorie du mouvement non permanent des eaux, avec application aux crues des rivières et à l'introduction des marées dans leur lit. *Comptes Rendus des Séances de l'Académie des Sciences. Paris*, **73**, 147–154.
- Barton, G.J., Moran, E.H. & Berenbrock, C. (2004). Surveying cross sections of the Kootenai River between Libby Dam, Montana, and Kootenay Lake, British Columbia, Canada. U.S. Geological Survey Open-File Report 2004-1045. Tech. rep.
- Bathurst, J.C., Hey, R.D. & Thorne, C.R. (1979). Secondary flow and shear stress at river bends. *Journal of the Hydraulics Division*, **105**, 1277–1295.
- Battjes, J.A. & Janssen, J.P.F.M. (1978). Energy loss and set-up due to breaking of random waves. In *Proc. 16th Int. Coastal Eng. Conf.*, 570–587, Am. Soc. of Civ. Eng., New York.

- Bennett, A.F. (2002). *Inverse Modeling of the Ocean and Atmosphere*. Cambridge Univ. Press, New York.
- Birkemeier, W.A. (1984). The CRAB: a unique nearshore surveying vehicle. *Journal of Surveying Engineering*, **110**, 1–7.
- Birkemeier, W.A. & Holland, K.T. (2001). The Corps of Engineers' field research facility: more than two decades of coastal research. *Shore and Beach*, **69**, 3–12.
- Blanton, B., Madry, S., Galluppi, K., Gamiel, K., Lander, H., Reed, M., Stillwell, L., Blanchard-Montgomery, M., Luettich, R., Mattocks, C., Fulcher, C., Vickery, P., Hanson, J., Devaliere, E. & McCormick, J. (2008). Draft report for the state of North Carolina floodplain mapping project: coastal flood analysis system. Submittal N.1: Topographic/Bathymetric Data. Tech. rep., State of North Carolina.
- Blenkinsopp, C.E., Turner, I.L., Allis, M.J., Peirson, W.L. & Garden, L.E. (2012). Application of LIDAR technology for measurement of time-varying free-surface profiles in a laboratory wave flume. *Coastal Engineering*, **68**, 1–5.
- Booij, N., Ris, R.C. & Holthuijsen, L.H. (1999). A third-generation wave model for coastal regions 1. Model description and validation. *Journal of Geophysical Research*, **104**(C4), 7649–7666.
- Bowen, A.J. (1969). The generation of longshore currents on a plane beach. *Journal of Marine Research*, **27**, 206–215.
- Bowen, A.J. & Holman, R.A. (1989). Shear instabilities of the mean longshore current. *Journal of Geophysical Research*, **94**(C2), 18,023–18,030.
- Bowen, A.J. & Pinless, S.J. (1974). Effects of bank raising along the Thames. In *Proc. 14th Int. Coastal Eng. Conf.*, vol. 3, 2471–2482.
- Cardinali, C., Pezzulli, S. & Andersson, E. (2004). Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **130**, 2767–2786.
- Casati, B., Stephenson, D.B., Nurmi, P., Ghelli, A., Pocerich, M., Damrath, U., Ebert, E.E., Brown, B.G. & Mason, S. (2008). Forecast verification: current status and future directions. *Meteorological Applications*, **15**, 3–18.
- Catalán, P.A. & Haller, M.C. (2007). Remote sensing of breaking wave phase speeds with application to non-linear depth inversions. *Coastal Engineering*, **55**, 93–111.
- Catalán, P.A., Haller, M.C., Holman, R.A. & Plant, W.J. (2011). Optical and microwave detection of wave breaking in the surf zone. *IEEE Transactions on Geoscience and Remote Sensing*, **49**, 1879–1893.



- Center, E.M. (2003). NCEP Office Note 442: The GFS atmospheric model. Tech. rep., NOAA / NWS / NCEP, 14 pp.
- Chickadel, C.C., Holman, R.A. & Freilich, M.H. (2003). An optical technique for the measurement of longshore currents. *Journal of Geophysical Research*, **108**(C11), 3364.
- Chickadel, C.C., Horner-Devine, A.R., Talke, S.A. & Jessup, A.T. (2009). Vertical boil propagation from a submerged estuarine sill. *Geophysical Research Letters*, **36**, L10601.
- Chickadel, C.C., Talke, S.A., Horner-Devine, A.R. & Jessup, A.T. (2011). Infrared-based measurements of velocity, turbulent kinetic energy, and dissipation at the water surface in a tidal river. *Geoscience and Remote Sensing Letters, IEEE*, **8**, 849–853.
- Clark, D.B., Feddersen, F., Omand, M.M. & Guza, R.T. (2009). Measuring fluorescent dye in the bubbly and sediment-laden surfzone. *Water, Air, & Soil Pollution*, **204**, 103–115.
- Clark, D.B., Elgar, S. & Raubenheimer, B. (2012). Vorticity generation by short-crested wave breaking. *Geophysical Research Letters*, **39**, L24604.
- Coco, G. & Murray, A.B. (2007). Patterns in the sand: from forcing templates to self-organization. *Geomorphology*, **91**, 271–290.
- de Vries, S., Hill, D.F., de Schipper, M.A. & Stive, M.J.F. (2011). Remote sensing of surf zone waves using stereo imaging. *Coastal Engineering*, **58**, 239–250.
- Dee, D.P. (1995). On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, **123**, 1128–1145.
- Desroziers, G., Berre, L., Chapnik, B. & Poli, P. (2006). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3385–3396.
- Durand, M., Andreadis, K.M., Alsdorf, D.E., Lettenmaier, D.P., Moller, D. & Wilson, M. (2008). Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophysical Research Letters*, **35**, L20401.
- Elgar, S., Guza, R.T., Raubenheimer, B., Herbers, T.H.C. & Gallagher, E.L. (1997). Spectral evolution of shoaling and breaking waves on a barred beach. *Journal of Geophysical Research*, **102**(C7), 15,797–15,805.
- Elgar, S., Guza, R.T., O'Reilly, W.C., Raubenheimer, B. & Herbers, T.H.C. (2001). Wave energy and direction observed near a pier. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, **207**, 2–6.
- Evensen, G. (2006). *Data Assimilation: The Ensemble Kalman Filter*. Springer.

- Evensen, G. & van Leeuwen, P.J. (1996). Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Monthly Weather Review*, **124**, 85–96.
- Feddersen, F. & Guza, R.T. (1998). Alongshore momentum balances in the nearshore. *Journal of Geophysical Research*, **103**(C8), 15,667–15,676.
- Feddersen, F. & Guza, R.T. (2003). Observations of nearshore circulation: Alongshore uniformity. *Journal of Geophysical Research*, **108**(C1).
- Feddersen, F., Guza, R.T. & Elgar, S. (2004). Inverse modeling of one-dimensional setup and alongshore current in the nearshore. *Journal of Physical Oceanography*, **34**, 920–933.
- Feng, Z., Reniers, A., Haus, B.K. & Solo-Gabriele, H.M. (2013). Modeling sediment-related enterococci loading, transport, and inactivation at an embayed nonpoint source beach. *Water Resources Research*.
- Gallagher, E.L., Elgar, S. & Guza, R.T. (1998). Observations of sand bar evolution on a natural beach. *Journal of Geophysical Research*, **103**, 3203–3215.
- Galvin, C.J. (1967). Longshore current velocity: a review of theory and data. *Reviews of Geophysics*, **5**, 287–304.
- Garcez-Faria, A.F., Thornton, E.B., Lippmann, T.C. & Stanton, T.P. (2000). Undertow over a barred beach. *Journal of Geophysical Research*, **105**(C7), 16999–17010.
- García-Medina, G., Özkan-Haller, H.T., Ruggiero, P. & Oskamp, J. (2013). An inner-shelf wave forecasting system for the U.S. Pacific Northwest. *in press, Weather and Forecasting*.
- Giddings, S.N., Fong, D.A. & Monismith, S.G. (2011). Role of straining and advection in the intratidal evolution of stratification, vertical mixing, and longitudinal dispersion of a shallow, macrotidal, salt wedge estuary. *Journal of Geophysical Research*, **116**(C3), C03003.
- Gneiting, T., Stanberry, L.I., Grit, E.P., Held, L. & Johnson, N.A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *Test*, **17**, 211–235.
- Grant, W.D. & Madsen, O.S. (1979). Combined wave and current interaction with a rough bottom. *Journal of Geophysical Research*, **84**(C4), 1797–1808.
- Guza, R.T. & Thornton, E.B. (1978). Variability of longshore currents. In *Proc. 16th Int. Coastal Eng. Conf.*, 756–775, Am. Soc. of Civ. Eng., New York.
- Haas, K.A. & Svendsen, I.A. (2002). Laboratory measurements of the vertical structure of rip currents. *Journal of Geophysical Research*, **107**(C5), 3047.

- Haas, K.A., Svendsen, I.A. & Haller, M.C. (1998). Numerical modeling of nearshore circulation on a barred beach with rip channels. In *Proc. 26th Int. Coastal Eng. Conf.*, vol. 1, 801–814, Am. Soc. of Civ. Eng.
- Haas, K.A., Svendsen, I.A., Haller, M.C. & Zhao, Q. (2003). Quasi-three-dimensional modeling of rip current systems. *Journal of Geophysical Research*, **108**(C7), **3217**.
- Haller, M.C., Honegger, D. & Catalán, P.A. (2013). Rip current observations via marine radar. *submitted to Journal of Waterway, Port, Coastal, and Ocean Engineering*.
- Hamill, T.M. & Whitaker, J.S. (2005). Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Monthly Weather Review*, **133**, 3132–3147.
- Hamill, T.M., Whitaker, J.S. & Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, **129**, 2776–2790.
- Hathaway, K.K. & Hanson, J.L. (2011). Wave transformation and setup from a cross-shore array of acoustic Doppler profilers. 12th International Workshop on Wave Hindcasting and Forecasting & 3rd Coastal Hazard Symposium, <http://www.waveworkshop.org/12thWaves/>.
- Henderson, S.M., Allen, J.S. & Newberger, P.A. (2004). Nearshore sandbar migration predicted by an eddy-diffusive boundary layer model. *Journal of Geophysical Research*, **109**(C06024).
- Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15**, 559–570.
- Higgins, A.L., Seymour, R.J. & Pawka, S.S. (1981). A compact representation of ocean wave directionality. *Applied Ocean Research*, **3**, 105–112.
- Hilldale, R.C. & Raff, D. (2008). Assessing the ability of airborne LIDAR to map river bathymetry. *Earth Surface Processes and Landforms*, **33**, 773–783.
- Hoefel, F. & Elgar, S. (2003). Wave-induced sediment transport and sandbar migration. *Science*, **299**, 1885.
- Holland, K.T., Puleo, J.A. & Kooney, T.N. (2001). Quantification of swash flows using video-based particle image velocimetry. *Coastal Engineering*, **44**, 65–77.
- Holman, R.A. & Haller, M.C. (2013). Remote sensing of the nearshore. *Annual Reviews of Marine Science*, **5**, 95–113.
- Holman, R.A. & Stanley, J. (2007). The history and technical capabilities of Argus. *Coastal Engineering*, **54**, 477–491.

- Holman, R.A., Plant, N.G. & Holland, K.T. (2013). cBathy: A robust algorithm for estimating nearshore bathymetry. *in press, Journal of Geophysical Research*.
- Honnorat, M., Monnier, J., Rivière, N., Huot, É. & Le Dimet, F.X. (2010). Identification of equivalent topography in an open channel flow using Lagrangian data assimilation. *Computing and Visualization in Science*, **13**, 111–119.
- Houtekamer, P.K. & Mitchell, H.L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, **126**, 796–811.
- Houtekamer, P.L., Mitchell, H.L. & Xingxiu, D. (2008). Model error representation in an operational ensemble Kalman filter. *Monthly Weather Review*, **137**, 2126–2143.
- Hsu, Y.L., Dykes, J.D. & Allard, R.A. (2006). Evaluation of Delft3D performance in nearshore flows. Tech. rep., Naval Research Laboratory.
- Ikedda, S. & Parker, G., eds. (1989). *River Meandering*. No. 12 in Water Resources Monograph, American Geophysical Union.
- Kirby, J.T. & Dalrymple, R.A. (1986). An approximate model for nonlinear dispersion in monochromatic wave propagation models. *Coastal Engineering*, **9**, 545–561.
- Kuik, A.H., van Vledder, G.P. & Holthuijsen, L.H. (1988). A method for the routine analysis of pitch-and-roll buoy wave data. *Journal of Physical Oceanography*, **18**, 1020–1034.
- Kurapov, A.L., Egbert, G.D., Allen, J.S. & Miller, R.N. (2007). Representer-based variational data assimilation in a nonlinear model of nearshore circulation. *Journal of Geophysical Research*, **112**(C11019).
- Kurapov, A.L., Egbert, G.D., Allen, J.S. & Miller, R.N. (2009). Representer-based analyses in the coastal upwelling system. *Dynamics of Atmospheres and Oceans*, **48**, 198–218.
- Landon, K. (2012). *Ensemble-based data assimilation and depth inversion on the Kootenai River, ID, U. S. A.*. Master’s thesis, Oregon State University, Department of Civil Engineering.
- Legleiter, C.J., Roberts, D.A. & Lawrence, R.L. (2009). Spectrally based remote sensing of river bathymetry. *Earth Surface Processes and Landforms*, **34**, 1039–1059.
- Lentz, S., Guza, R.T., Elgar, S., Feddersen, F. & Herbers, T.H.C. (1999). Momentum balances on the North Carolina inner shelf. *Journal of Geophysical Research*, **104**(C8), 18205–18226.
- Li, H., Kalnay, E. & Miyoshi, T. (2009). Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135**, 523–533.

- Lippmann, T.C. & Holman, R.A. (1990). The spatial and temporal variability of sand bar morphology. *Journal of Geophysical Research*, **95**(C7), 11,575–11,590.
- Long, C.E. (1996). *Index and bulk parameters for frequency-directional spectra measured at CERC Field Research Facility, July 1994 to August 1995*. Misc. Pap. CERC-96-6, U. S. Army Eng. Waterw. Exp. Stn., Vicksburg, Miss.
- Longuet-Higgins, M.S. (1970a). Longshore currents generated by obliquely incident sea waves, 1. *Journal of Geophysical Research*, **75**, 6778–6789.
- Longuet-Higgins, M.S. (1970b). Longshore currents generated by obliquely incident sea waves, 2. *Journal of Geophysical Research*, **75**, 6790–6801.
- Longuet-Higgins, M.S. & Stewart, R.W. (1964). Radiation stress in water waves, a physical discussion with applications. *Deep-Sea Research*, **11**, 529–563.
- MacMahan, J., Brown, J., Thornton, E., Reniers, A., Stanton, T., Henriquez, M., Gallagher, E., Morrison, J., Austin, M.J., Scott, T.M. & Senechal, N. (2010). Mean Lagrangian flow behavior on an open coast rip-channeled beach: a new perspective. *Marine Geology*, **268**, 1–15.
- Mei, C.C. (1983). *The Applied Dynamics of Ocean Surface Waves*. Wiley, New York.
- Morris, B.J. (2001). *Nearshore wave and current dynamics*. Ph.D. thesis, Naval Postgraduate School.
- Mourre, B., Mey, P.D., Lyard, F. & Provost, C.L. (2004). Assimilation of sea level data over continental shelves: an ensemble method for the exploration of model errors due to uncertainties in bathymetry. *Dynamics of Atmospheres and Oceans*, **38**, 93–101.
- Mulligan, R.P., Hanson, J.L. & Hathaway, K.K. (2010). Observations of wave breaking and surf zone width from a real-time cross-shore array of wave and current sensors at Duck, NC. In *Proceedings of the IEEE/OES/CWTM 10th Working Conference on Current Measurement Technology*, 130–137.
- Nadaoka, K. & Yagi, H. (1998). Shallow-water turbulence modeling and horizontal large-eddy computation of river flow. *Journal of Hydraulic Engineering*, **124**, 493–500.
- Nielsen, P. (1992). *Coastal Bottom Boundary Layers and Sediment Transport*, vol. 4. World Scientific Publishing Company Incorporated.
- Oke, P.R. & Sakov, P. (2007). Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology*, **25**, 1004–1017.
- Oltman-Shay, J. & Howd, P.A. (1989). Shear instabilities of the mean longshore current 2. Field observations. *Journal of Geophysical Research*, **94**(C2), 18,031–18,042.

- Özkan Haller, H.T. & Kirby, J.T. (1999). Nonlinear evolution of shear instabilities of the longshore current: A comparison of observations and computations. *Journal of Geophysical Research*, **104**(C11), 25,953–25,984.
- Palma, E.D. & Matano, R.P. (1998). On the implementation of passive open boundary conditions for a general circulation model: the barotropic mode. *Journal of Geophysical Research*, **103**(C1), 1319–1341.
- Palmsten, M.L. & Holman, R.A. (2011). Laboratory investigation of dune erosion using stereo video. *Coastal Engineering*, **60**, 123–135.
- Plant, N.G. & Holman, R.A. (1997). Intertidal beach profile estimation using video images. *Marine Geology*, **140**, 1–24.
- Plant, N.G., Holland, K.T. & Puleo, J.A. (2002). Analysis of the scale of errors in nearshore bathymetric data. *Marine Geology*, **191**, 71–86.
- Plant, N.G., Holland, K.T. & Holman, R.A. (2006). A dynamical attractor governs beach response to storms. *Geophysical Research Letters*, **33**(L17607).
- Plant, N.G., Edwards, K.L., Kaihatu, J.M., Veeramony, J., Hsu, L. & Holland, K.T. (2009). The effect of bathymetric filtering on nearshore process model results. *Coastal Engineering*, **56**, 484–493.
- Plant, W.J., Keller, W.C. & Hayes, K. (2005). Measurement of river surface currents with coherent microwave systems. *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 1242–1257.
- Puleo, J.A., Farquharson, G., Frasier, S.J. & Holland, K.T. (2003). Comparison of optical and radar measurements of surf and swash zone velocity fields. *Journal of Geophysical Research*, **108**(C3), 3100.
- Puleo, J.A., McKenna, T.E., Holland, K.T. & Calantoni, J. (2012). Quantifying riverine surface currents from time sequences of thermal infrared imagery. *Water Resources Research*, **48**(W01527).
- Rafiee, M., Tinka, A., Thai, J. & Bayen, A.M. (2011). Combined state-parameter estimation for shallow water equations. In *American Control Conference*, 1333–1339, IEEE.
- RDInstruments (1996). *Acoustic Doppler Current Profiler Principles of Operation, a Practical Primer*. 9855 Businesspark Ave. , San Diego, California 92131 USA, 2nd edn.
- Reniers, A.H.J.M. & Battjes, J.A. (1997). A laboratory study of longshore currents over barred and non-barred beaches. *Coastal Engineering*, **30**, 1–22.

- Reniers, A.J.H.M., Roelvink, J.A. & Thornton, E.B. (2004). Morphodynamic modeling of an embayed beach under wave group forcing. *Journal of Geophysical Research*, **109**(C01030).
- Rilov, G., Dudas, S.E., Menge, B.A., Grantham, B.A., Lubchenco, J. & Schiel, D.R. (2008). The surf zone: a semi-permeable barrier to onshore recruitment of invertebrate larvae? *Journal of Experimental Marine Biology and Ecology*, **361**, 59–74.
- Rogers, E., DiMego, G.J., Black, T.L., Ek, M.B., Ferrier, B.S., Gayno, G.A., Janjic, Z., Lin, Y., Pyle, M.E., Wong, V.C. *et al.* (2009). The NCEP North American mesoscale modeling system: Recent changes and future plans. In *Preprints, 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction*, 2A. 4, American Meteorological Society, Omaha, NE, available online at <http://ams.confex.com/ams/pdfpapers/154114.pdf>.
- Ruessink, B.G., Miles, J.R., Feddersen, F., Guza, R.T. & Elgar, S. (2001). Modeling the alongshore current on barred beaches. *Journal of Geophysical Research*, **106**(C10), 22,451–22,463.
- Sela, J.G. (1980). Spectral modeling at the National Meteorological Center. *Monthly Weather Review*, **108**, 1279–1292.
- Shanks, A.L., Morgan, S.G., MacMahan, J. & Reniers, A.J.H.M. (2010). Surf zone physical and morphological regime as determinants of temporal and spatial variation in larval recruitment. *Journal of Experimental Marine Biology and Ecology*, **392**, 140–150.
- Shchepetkin, A.F. & McWilliams, J.C. (2005). The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, **9**, 347–404.
- Slinn, D.N., Newberger, P.A. & Holman, R.A. (1998). Nonlinear shear instabilities of alongshore currents. *Journal of Geophysical Research*, **103**(C9), 18,357–18,379.
- Slinn, D.N., Allen, J.S. & Holman, R.A. (2000). Alongshore currents over variable beach topography. *Journal of Geophysical Research*, **105**(C7), 16,971–16,998.
- Smith, J.D. & McLean, S. (1984). A model for flow in meandering streams. *Water Resources Research*, **20**, 1301–1315.
- Smith, J.D. & McLean, S.R. (1977). Spatially averaged flow over a wavy surface. *Journal of Geophysical research*, **82**(12), 1735–1746.
- Smith, S.D. (1988). Coefficients for sea surface wind stress, heat flux, and wind profiles as a function of wind speed and temperature. *Journal of Geophysical Research*, **93**(15), 467–15.

- Splinter, K.D. & Holman, R.A. (2009). Bathymetry estimation from single-frame images of nearshore waves. *IEEE Transactions on Geoscience and Remote Sensing*, **47**, 3151–3160.
- Stockdon, H.F. & Holman, R.A. (2000). Estimation of wave phase speed and nearshore bathymetry from video imagery. *Journal of Geophysical Research*, **105**(C9), 22,015–22,033.
- Svendsen, I.A. (1984). Wave heights and set-up in a surf zone. *Coastal Engineering*, **8**, 303–329.
- Svendsen, I.A. & Putrevu, U. (1990). Nearshore circulation with 3-D profiles. In *Proc. 22nd Int. Coastal Eng. Conf.*, 241–254, Am. Soc. of Civ. Eng., New York.
- Svendsen, I.A. & Putrevu, U. (1994). Nearshore mixing and dispersion. *Proceedings of the Royal Society of London A*, **445**, 561–576.
- Svendsen, I.A., Haas, K. & Zhao, Q. (2002). Quasi-3D nearshore circulation model, SHORECIRC, Report #2002-01. Tech. rep., Center for Applied Coastal Research. University of Delaware.
- Sverdrup, H.U. & Munk, W.H. (1946). Theoretical and empirical relations in forecasting breakers and surf. *Transactions, American Geophysical Union*, **27**, 828–836.
- Swick, W. (2011). *Field and numerical study on natural river mixing*. Ph.D. thesis, Naval Postgraduate School.
- TerraSond, L. (2009). Snohomish river bathymetric survey, Everett, WA, hydrographic survey report. Tech. rep., for Applied Physics Laboratory, University of Washington.
- Thornton, E.B. & Guza, R.T. (1986). Surf zone longshore currents and random waves: field data and models. *Journal of Physical Oceanography*, **16**, 1165–1178.
- Tolman, H.L. (2002). Validation of WAVEWATCH III version 1.15 for a global domain. Tech. rep., NOAA / NWS / NCEP / OMB, 33 pp.
- Tolman, H.L. (2006). Development of a multi-grid version of WAVEWATCH III. Tech. rep., NOAA / NWS / NCEP / MMAB, 88 pp.
- Tossavainen, O.P., Percelay, J., Tinka, A., Wu, Q. & Bayen, A.M. (2008). Ensemble Kalman filter based state estimation in 2D shallow water equations using Lagrangian sensing and state augmentation. In *47th IEEE Conference on Decision and Control*, 1783–1790.
- van Dongeren, A., Plant, N., Cohen, A., Roelvink, D., Haller, M.C. & Catalán, P. (2008). Beach Wizard: nearshore bathymetry estimation through assimilation of model computations and remote observations. *Coastal Engineering*, **55**, 1016–1027.



- Veeramony, J., Walker, D. & Hsu, L. (2010). A variational data assimilation system for nearshore applications of SWAN. *Ocean Modelling*, **35**, 206–214.
- Wilson, G.W. (2009). *Field validation of a nearshore circulation model for alongshore-nonuniform flows*. Master's thesis, Oregon State University.
- Wilson, G.W. & Özkan-Haller, H.T. (2012). Ensemble-based data assimilation for estimation of river depths. *Journal of Atmospheric and Oceanic Technology*, **29**, 1558–1568.
- Wilson, G.W., Özkan-Haller, H.T. & Holman, R.A. (2010). Data assimilation and bathymetric inversion in a 2DH surf zone model. *Journal of Geophysical Research*, **115**(C12057).
- Wilson, G.W., Özkan-Haller, H.T. & Holman, R.A. (2013). Quantifying the length-scale dependence of surf zone advection. *in press, Journal of Geophysical Research*.
- Wright, L.D. & Short, A.D. (1984). Morphodynamic variability of surf zones and beaches: A synthesis. *Marine Geology*, **56**, 93–118.
- Yu, J. & Slinn, D.N. (2003). Effects of wave-current interaction on rip currents. *Journal of Geophysical Research*, **108**(C3).
- Zaron, E.D., Pradal, M.A., Miller, P.D., Blumberg, A.F., Georgas, N., Li, W. & Muccino Cornuelle, J. (2011). Bottom topography mapping via nonlinear data assimilation. *Journal of Atmospheric and Oceanic Technology*, **28**, 1606–1623.
- Zhao, Q., Svendsen, I.A. & Haas, K. (2003). Three-dimensional effects in shear waves. *Journal of Geophysical Research*, **108**(C8).

## APPENDICES

## A. Continuous Ranked Probability Score

In section 2.4.4, we test the skill of the prior and posterior estimates of the ocean state  $\psi$ , by cross-validation with observations. This involves testing the accuracy of the ocean state prediction, as well as the predicted uncertainty. Both must be assessed together in order to fully characterize the skill of the assimilation methodology. The validation is carried out using the Continuous Ranked Probability Score (CRPS), defined below.

For a given probabilistic forecast of a scalar random variable  $x$  (e.g. the prior or posterior model state  $\psi$  and its uncertainty  $C_{\psi\psi}$ ), define the cumulative distribution function (cdf)  $f_X(x)$ . Also define the cdf of the same random variable  $x$  as determined from an observation of the same variable,  $f_Y(x)$ . Then the CRPS is defined by (Hersbach, 2000; Gneiting *et al.*, 2008; Casati *et al.*, 2008)

$$CRPS = \int_{-\infty}^{\infty} (f_X(x) - f_Y(x))^2 dx. \quad (\text{A.1})$$

Note this is a generalization of the standard definition, allowing for observational uncertainty (e.g. instrument error).

Figure A.1 shows a graphical interpretation of the CRPS, as the squared area of the regions pointed to by the arrows. When the probability density functions (pdf's) of the prediction and the observation coincide, the CRPS goes to zero; large values of CRPS indicate an unskilled prediction. Note that CRPS takes into account both calibration (the agreement of the predicted and observed expected value) and sharpness (the agreement of the predicted and observed uncertainty). The prediction depicted in Figure A.1 is fairly well calibrated, but is not particularly sharp.

The extension to multiple observations (vector-valued random variables  $x$ ) follows Gneiting *et al.* (2008), who point out the following identity (from Baringhous & Franz (2004), Lemma 2.2):

$$\begin{aligned} CRPS &= E[\|X - Y\|] - \frac{1}{2}E[\|X - X'\|] \\ &\quad - \frac{1}{2}E[\|Y - Y'\|] \geq 0, \end{aligned} \quad (\text{A.2})$$

where  $X$  and  $X'$  are independent realizations following the cdf  $f_X$  (similar for  $Y$  and

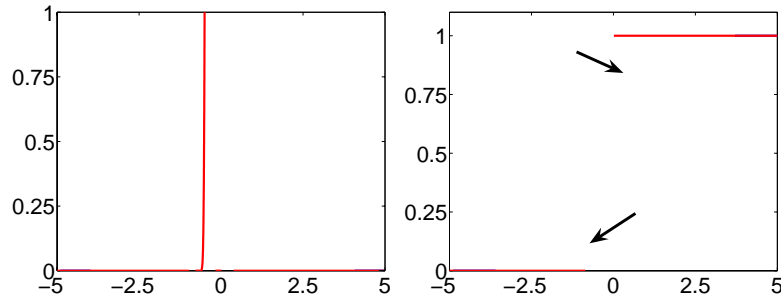


Figure A.1: pdf (left) and cdf (right) representations of the prediction (blue) and observation (red).

$Y'$ ),  $\|\cdot\|$  is the Euclidean norm, and  $E$  denotes expected value (computed herein using Monte-Carlo methods). With the norm thus-defined, CRPS is naturally extended for any number of observations. This definition also means CRPS reduces to the rms error when the variables are deterministic.

## B. Skill Statistics for $x > 250$ m

For completeness, the following table lists model skill for the offshore sensors  $x > 250$  m (discussed briefly in section 2.4.4).

			1DH (1130 EST)			2DH (1530 EST)		
variable(s) assimilated	variable updated	units	$\epsilon$	$CRPS$	$S$	$\epsilon$	$CRPS$	$S$
none (prior)	$u$	m/s	0.19	0.34	-	0.15	0.48	-
	$v$	m/s	0.032	0.042	-	0.14	0.42	-
	$H_{\text{mo}}$	m	0.075	0.17	-	0.11	0.44	-
	$h$	m	0.091	0.085	-	0.12	0.23	-
$H_{\text{mo}}, v$	$u$	m/s	0.20	0.35	-0.029	0.16	0.53	-0.11
	$v$	m/s	0.028	0.044	-0.049	0.091	0.26	0.37
	$H_{\text{mo}}$	m	0.068	0.15	0.10	0.11	0.44	-0.0077
	$h$	m	0.13	0.16	-0.86	0.14	0.32	-0.38
$H_{\text{mo}}$	$u$	m/s	0.19	0.35	-0.017	0.15	0.50	-0.048
	$v$	m/s	0.030	0.046	-0.098	0.16	0.50	-0.20
	$H_{\text{mo}}$	m	0.070	0.16	0.077	0.11	0.44	-0.0015
	$h$	m	0.11	0.12	-0.37	0.11	0.22	0.065
$v$	$u$	m/s	0.19	0.34	-0.014	0.16	0.51	-0.079
	$v$	m/s	0.028	0.042	0.0044	0.084	0.24	0.43
	$H_{\text{mo}}$	m	0.073	0.16	0.032	0.11	0.44	-0.0098
	$h$	m	0.12	0.13	-0.58	0.15	0.36	-0.53

Table B.1: As in Table 1, but for  $x > 250$  m.

## C. Global Wave Forecasting Model

In section 4.5.1, we discuss the effect of using forecasted (error-prone) wave boundary conditions in our assimilation system, rather than boundary conditions derived from in-situ measurements. Here, we describe the model used to generate those forecasts.

The model uses a series of nested WaveWatch III (Tolman, 2002, 2006) and SWAN (Booij *et al.*, 1999) grids, summarized in Table C.1. At the global and basin (Western North Atlantic) scale, the grids are based on the National Centers for Environmental Prediction operational wave model, which was re-implemented as described by García-Medina *et al.* (2013). Regional predictions are then calculated over the continental shelf using a 6 arc-minute resolution WaveWatch III model, and a 1 arc-minute resolution SWAN model extending to roughly 50 km offshore of our study site. In turn, the latter (SWAN) model is used to generate boundary conditions for the nearshore model described in section 4.3.1.

“Forecast” winds were derived by blending hindcast products from the Global Forecasting System (Center, 2003; Sela, 1980) and the 12 km resolution North Atlantic Model (Rogers *et al.*, 2009). True forecast products were not available, but it is assumed that these hindcasts would be comparable to forecasts at short lead times. Bathymetry for the shelf and FRF grids use data from ETOPO 1 (Amante & Eakins, 2009) and the U.S. Army Corps of Engineers (Blanton *et al.*, 2008), respectively.

Grid	Resolution arc-degrees	Lower Left Lat,Lon	Upper Right Lat,Lon
Global	$1.25^\circ \times 1^\circ$	$-78^\circ, 0^\circ$	$78^\circ, 359.5^\circ$
Basin	$0.25^\circ \times 0.25^\circ$	$0^\circ, -98^\circ$	$50^\circ, -30^\circ$
Shelf	$6' \times 6'$	$30^\circ, -82^\circ$	$40^\circ, -70^\circ$
FRF	$1' \times 1'$	$36.10^\circ, -75.80^\circ$	$36.30^\circ, -75.60^\circ$

Table C.1: Description of grids used by wave forecasting model.

