



## AN ABSTRACT OF THE DISSERTATION OF

Ivo C.H.M. Pasman for the degree of Doctor of Philosophy in Ocean, Earth, and Atmospheric Sciences presented on December 7, 2018.

Title: Ensemble-variational Data Assimilation for Dynamically Challenging Coastal Ocean Circulation

Abstract approved: \_\_\_\_\_  
Alexander L. Kurapov

Currently, forecasts produced by the Oregon-Washington (OR-WA) Coastal Ocean Forecast System are constrained by assimilation of only surface observations. The 4-dimensional variational (4DVAR) data assimilation (DA) algorithm is utilized to combine the model and the data, with the time-independent forecast (“background”) error covariance **B**. In this study, two possible improvements are explored: (i) assimilation of subsurface observations from autonomous underwater gliders, using the present 4DVAR DA system, and (ii) estimation of **B** from an ensemble of 4DVAR runs (En4DVar).

We find that assimilation of glider observations alone creates erroneous energetic eddies in the vicinity of the glider transect. Assimilation of surface and subsurface observations in combination prevents these features from forming and reduces the forecast error.

The ocean state in the OR-WA region is highly variable in space and time. Therefore, it is unlikely that the current DA system, using a static **B** with a horizontally uniform covariance, can fully detect and correct the background errors. The En4DVar necessitated the development of: (a) the new, computationally-efficient Monte-Carlo localization method; (b) a Bayesian Hierarchical Model to obtain realistic estimates

for the ensemble wind perturbations from scatterometer observations; and (c) the cluster search method which can cut the wall time needed to calculate the DA correction by 30% compared to conventional 4DVAR.

Compared to the current static  $\mathbf{B}$ , the ensemble  $\mathbf{B}$  exhibits stronger covariances in frontal areas that change their location from one assimilation window to another. The En4DVar yields a solution with improved temperature and salinity properties on the shelf compared to 4DVAR with the static  $\mathbf{B}$ . However, assimilation of biased SST data amplified by the strong ensemble covariance in the area of the Columbia River plume yields strong unphysical changes in the surface salinity, which is not constrained by observations. 4DVAR was modified to reduce this effect.

©Copyright by Ivo C.H.M. Pasman  
December 7, 2018  
All Rights Reserved

Ensemble-variational Data Assimilation for Dynamically Challenging  
Coastal Ocean Circulation

by

Ivo C.H.M. Pasman

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of  
the requirements for the  
degree of

Doctor of Philosophy

Presented December 7, 2018  
Commencement June 2019

Doctor of Philosophy dissertation of Ivo C.H.M. Pasmans presented on  
December 7, 2018.

APPROVED:

---

Major Professor, representing Ocean, Earth, and Atmospheric Sciences

---

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.

---

Ivo C.H.M. Pasmans, Author

## ACKNOWLEDGEMENTS

This research would be impossible without the unwavering support and commitment from my advisor, Alexander Kurapov, to see this through to the end from the other side of the country. My thanks to Mike Kosro for taking over some administrative tasks in his absence as well as helping out with the comparison of the velocity observations. I also would like to thank Svetlana Erofeeva for her help with locating files from the Oregon-Washington Coastal Ocean Forecast System. My gratitude also extends to my committee: Gary Egbert, Xiaoli Fern, Mike Kosro, Alexander Kurapov and Tuba Ozkan-Haller. I am grateful to the college administrative staff for helping out with the neverending stream of forms and Tom Leach for the IT support provided. I am also thankful to my cohort, Emily Lemagie and Allison Einolf, and other students for the time spent and my parents for making living abroad bearable. Finally, I would like to thank all fellow PhDs, postdocs and other conference participants for the many enjoyable evenings spent talking science and non-science.

This study was made possible thanks to financial support from the National Oceanic and Atmospheric Administration (NOAA) Coastal Ocean Modeling Testbed (COMT) grant NA13NOS0120139, the NOAA Quantitative Observing System Assessment Program (QOSAP), National Science Foundation (NSF) grants OCE-0527168 and OCE-0961999, Integrated Ocean Observing System / Northwest Association of Networked Ocean Observing Systems (IOOS/NANOOS) grant NA16NOS0120019, the National Aeronautics and Space Administration (NASA) SWOT Science Definition Team project grant NNX13AD89G. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) under allocation TG-OCE160001, which is supported by NSF grant number ACI-1548562.

## CONTRIBUTION OF AUTHORS

A.L. Kurapov contributed to the writing and analysis in chapters 1-6. The high-frequency radar sea-surface velocity observations used in chapters 2,4 and 5 were obtained from P.M. Kosro. J. A. Barth, R.K. Shearman, and A. Erofeev collected the glider observations which are used in chapters 2 and 5. The NH10 buoy observations used in these two chapters were obtained from M. Levine and C. Risien. E. Maturi, A. Harris, X. Zhou and A. Ignatov helped with obtaining the satellite sea-surface temperature observations for chapters 2,4 and 5. Discharges for the rivers around the Salish Sea used in chapters 2, 4 and 5 were based on data provided by P. MacCready and S. Giddings. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official NOAA or U.S. government position, policy, or decision.

## TABLE OF CONTENTS

	<u>Page</u>
1 General Introduction . . . . .	1
2 Why Gliders Appreciate Good Company: Glider Assimilation in the Oregon-Washington Coastal Ocean 4DVAR System with and without Surface Observations . . . . .	8
2.1 Abstract . . . . .	9
2.2 Introduction . . . . .	9
2.3 The Dynamical Regime and the Model Setup . . . . .	13
2.3.1 Dynamical Regime . . . . .	13
2.3.2 Model . . . . .	14
2.3.3 Surface Observations . . . . .	16
2.3.4 Subsurface Observations . . . . .	17
2.3.5 Data Assimilation (DA) . . . . .	18
2.3.6 Experiments . . . . .	21
2.4 Results . . . . .	24
2.4.1 DA Impact on Surface Fields . . . . .	25
2.4.2 DA Impact on Subsurface Fields . . . . .	31
2.4.3 Post DA Persistence . . . . .	38
2.4.4 Contributions from Different Observation Types . . . . .	40
2.4.5 DA Impact on the Topology of the 26.5 Isopycnal Surface . . . . .	42
2.5 Conclusions and Discussion . . . . .	44
3 A Monte-Carlo Background Covariance Localization Method for an Ensemble-variational Assimilation System . . . . .	46
3.1 Abstract . . . . .	47
3.2 Introduction . . . . .	47
3.3 Method . . . . .	50
3.4 Computational Cost . . . . .	53
3.5 MC Localization in a 1D Model . . . . .	56
3.5.1 Convergence (Dependence on $M_{MC}$ ) . . . . .	59
3.5.2 Localization Distance ( $L_{MC}$ ) . . . . .	63
3.5.3 Shape of Covariance . . . . .	64
3.5.4 Spectrum . . . . .	66
3.6 Example from a 3D-model . . . . .	68

## TABLE OF CONTENTS (Continued)

	<u>Page</u>
3.7 Summary and Concluding Discussion . . . . .	70
4 Ensemble 4DVAR (En4DVar) Data Assimilation in a Coastal Ocean Circulation Model. Part I: System Design and Performance . . . . .	73
4.1 Abstract . . . . .	73
4.2 Introduction . . . . .	73
4.3 En4DVar System . . . . .	77
4.4 Cost Function Minimization . . . . .	82
4.4.1 RBCG . . . . .	84
4.4.2 Full Parallelization: Block Diagonal Conjugate Gradient Method	85
4.4.3 Partial Parallelization: Cluster Search Method . . . . .	85
4.5 Wind Perturbations . . . . .	88
4.6 Results . . . . .	94
4.6.1 Convergence . . . . .	94
4.6.2 Error Reduction . . . . .	99
4.6.3 Ensemble Reliability . . . . .	102
4.7 Conclusions and Discussion . . . . .	108
5 Ensemble 4DVAR (En4DVar) Data Assimilation in a Coastal Ocean Circulation Model. Part II: Performance . . . . .	112
5.1 Abstract . . . . .	112
5.2 Introduction . . . . .	113
5.3 Model experiments . . . . .	115
5.4 Hierarchical Box Salinity Constraint (HBSC) . . . . .	117
5.5 Covariances . . . . .	119
5.6 Columbia River Plume . . . . .	127
5.7 Verification . . . . .	132
5.7.1 Surface . . . . .	132
5.7.2 Subsurface . . . . .	136
5.8 Discussion and Conclusions . . . . .	143

## TABLE OF CONTENTS (Continued)

	<u>Page</u>
6 Conclusions . . . . .	148
6.1 Conclusions . . . . .	148
6.2 Outlook . . . . .	150
Bibliography . . . . .	151
Appendices . . . . .	175

## LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
1.1	Schematic overview of an operational numerical weather prediction or ocean prediction system. . . . .	2
1.2	Schematic overview of two DA corrections using (a) the ensemble Kalman filter, (b) 3DVAR, (c) 3DVAR-FGAT, (d) 4DVAR. Observations (black dots) are compared with the forecasts (solid black lines) for the time indicated by the black arrows. Based on this comparison and the magnitude of the background error covariance (grey) a DA correction (blue) is calculated. This corrected state then serves as the initial condition for the next analysis (dashed black lines). . . . .	4
2.1	Overview of the model region together with the assimilated glider transects (orange), the 200 m, 1000 m, 2000 m isobaths (grey lines) and the coastal shelf region used in the surface observation comparison in Figures 2.14 and 2.15 (grey). Also shown are the 31.5 isohaline on 21 July 2011 (black solid line) and the extent of the Columbia River plume on 13 August 2011 (dotted black line) according to the model without DA. . . . .	13
2.2	(a) Meridional wind-stress applied to the models close to the glider transects at 124.308°W, 44.634°N. The assimilation period (21 July to 11 August), is marked by the vertical dashed lines. The negative wind-stress corresponds to upwelling favorable northerly winds. (b) The surface salinity measured at 2 m depth at the NH10 buoy (124.308°W, 44.634°N) (black) and obtained from the long free model run without DA (blue) both filtered using a double application of a 24h-running mean filter. . . . .	15
2.3	Background error temperature-temperature covariance between a point on the surface at 124.843°W, 43.005°N and the vertical (solid black), covariance between a point at 100 m depth and the vertical (dashed blue) and the time-averaged observational variance (i.e. part of the diagonal of $\mathbf{R}$ ) for the glider temperature observations as used in the 4DVAR algorithm (dotted black). Also shown is the time-averaged vertical temperature profile obtained from glider observations over the period 21 July to 14 August 2011 (solid gray). . . . .	18
2.4	The experiment time line. . . . .	22

## LIST OF FIGURES (Continued)

<u>Figure</u>		
2.5 Time-averaged vertical profiles of (a) salinity and (b) temperature between 1 May and 1 October 2011 based on (black) glider observations and (blue) the long free model run. Solid (dashed) lines are the averages of the profiles when the Columbia River plume was present (absent) in the observations, i.e., when surface $S < 31.5$ ( $> 31.5$ ). . . . .	24	
2.6 Daily-averaged surface velocity fields (arrows) and relative vorticity in the top model layer (color scale) on 4 August 2011 from (a) the experiment without data assimilation, (b) analysis from Surface Only, (c) analysis from Glider Only and (d) analysis from Combined. The location of the glider measurements up to 8 August 2011 24:00 is indicated by the orange lines. Grey lines mark the 200 m, 1000 m, 2000 m isobaths. . . . .	26	
2.7 Enstrophy for the area between $42^{\circ}\text{N}$ , $47^{\circ}\text{N}$ , $127^{\circ}\text{W}$ and the coast for the case No DA (blue) and the analysis from the cases Glider Only (red), Surface Only (purple) and Combined (green). The vertical grid lines mark the beginning of each DA window, when instantaneous correction is applied. . . . .	27	
2.8 Daily-averaged surface temperature (color scale) and the 31.5 (solid line) salinity contour on 4 August 2011 from (a) case No DA, (b) analysis from Surface Only, (c) analysis from Glider Only, (d) analysis from Combined and (e) from satellite observations. The location of the glider measurements up to 8 August 2011 24:00 is indicated by the orange lines. Gray lines mark the 200 m, 1000 m, 2000 m isobaths. . . . .	28	
2.9 (a) Observed glider $T$ as a function of time and depth; (b)-(e) the model-observation $T$ difference, shown as color and selected model $T$ contours, for cases (top to bottom) No DA, Surface, Glider Only, and Combined. DA analyses are used. Contours are $T = 7, 8, 10, 12^{\circ}\text{C}$ . . . . .	29	
2.10 (a) Observed glider $S$ as a function of time and depth; (b)-(e) the model-observation $S$ difference, shown as color and selected model $S$ contours, for cases (top to bottom) No DA, Surface, Glider Only, and Combined. DA analyses are used. Contours are $S = 30.5, 31.5, 32.5, 33.5$ . . . . .	30	

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
2.11 RMSEs from analyses (a-c) and forecasts (d-f) averaged over the period 21 July 2011 to 11 August 2011 compared to glider salinity observations (first row), glider temperature observations (second row) and potential density (third row) as functions of depth. . . . .	33
2.12 Daily-averaged density along the line 44.64 °N from the analyses on 4 August 2011 for (a) experiment Surface Only, (b) experiment Glider Only and (c) experiment Combined. The dashed box indicates the area in which glider observations are available during this window. The dashed black line marks the $26.5 \text{ kg m}^{-3}$ isopycnal. . . . .	36
2.13 Standard deviation in the depths of the $24.5 \text{ kg m}^{-3}$ (a), $25.5 \text{ kg m}^{-3}$ (b) and $26.5 \text{ kg m}^{-3}$ (c) isopycnal from the hourly output of the analyses between 21 July to 11 August along 44.64°N, shown as functions of longitude in the different experiments. The dashed black line marks the most westward location of the glider. . . . .	37
2.14 Time series of the area-averaged, daily-averaged forecast RMSE as function of time for (a) HFR velocity, (b) SST, (c) glider temperature and (d) glider salinity for the cases Surface Only (purple), Glider Only (red), Combined (green) and No DA (blue). The vertical lines mark the instant of the last DA correction (8 August 2011). . . . .	39
2.15 The time series of the daily-averaged, area-averaged RMSE in (a) SST, (b) HFR velocity, (c) glider temperature, (d) glider salinity observations for the DA cases, in which the subsets of the available data sets are assimilated. The values are shown relative to the RMSE for case Combined, i.e., values below (above) one mean that the case is better (worse) than case Combined. Shown are DA analyses where only glider temperature and salinity (red), only glider temperature (blue), glider temperature, salinity and SST (green) and glider temperature, salinity and HFR (black) are assimilated. . . . .	41
2.16 $z_{26.5} \text{ kg m}^{-3}$ isopycnal on 21 July 2011 (top) and 10 August 2011 (bottom) from analysis for Glider Only (left), Surface Only (center) and Combined (right). The depth contours, every 25 m, are shown as dashed lines. The change in depth with respect to No DA is shown as color. Orange dots mark the location of the glider during up to and including the assimilation window. . . . .	43

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
3.1 1D example showing the construction of the modulated ensemble members for the case $L_{MC} = 51$ on a 501 point grid. a) Two raw ensemble members $\mathbf{x}^{(1)} - \bar{\mathbf{x}}$ (solid) and $\mathbf{x}^{(2)} - \bar{\mathbf{x}}$ (dashed). b) Vectors $\chi^{(85)}$ (dark blue), $\chi^{(253)}$ (green), $\chi^{(317)}$ (red), $\chi^{(432)}$ (light blue). c) The modulated ensemble members $(\mathbf{x}^{(1)} - \bar{\mathbf{x}}) \circ \chi^{(85)} \circ \gamma$ (dark blue), $(\mathbf{x}^{(2)} - \bar{\mathbf{x}}) \circ \chi^{(253)} \circ \gamma$ (green), $(\mathbf{x}^{(1)} - \bar{\mathbf{x}}) \circ \chi^{(317)} \circ \gamma$ (red), $(\mathbf{x}^{(2)} - \bar{\mathbf{x}}) \circ \chi^{(432)} \circ \gamma$ (light blue). . . . .	52
3.2 The relative error $\ \mathbf{B}_{full}\mathbf{v} - \mathbf{B}_{MC}\mathbf{v}\  / \ \mathbf{B}_{full}\mathbf{v}\ ^{-1}$ as function of $M_{MC}$ (the size of the sets $\mathbb{M}_k$ in the MC approximation) and different choices of $K$ (the number of raw ensemble members); note that $KM_{MC}$ is the total number of modulated ensemble members used. The solid line shows the mean relative error averaged over 1000 different realizations of the raw ensemble, the modulated ensemble and the vector $\mathbf{v}$ . Error bars indicate the interval around the mean in which 90% of the relative errors fall. . . . .	60
3.3 Each dot represents the mean relative error $\ \mathbf{B}_{full}\mathbf{v} - \mathbf{B}_{MC}\mathbf{v}\  / \ \mathbf{B}_{full}\mathbf{v}\ ^{-1}$ over 1000 realizations of $\mathbf{B}_{full}$ , $\mathbf{B}_{MC}$ and $\mathbf{v}$ for a different value of $K$ and $M_{MC}$ . Colors show the size of the raw ensemble ( $K$ ). The dashed line follows the relation $5.8(KM_{MC})^{-1/2}$ . . . . .	61
3.4 The comparison of the true (gray) and ensemble-generated covariances (black) computed with respect to the domain center point: a) mean (solid black line) +/- standard deviation (dashed black lines) of 50 raw ensemble covariances. b) mean (solid black line) +/- standard deviation (dashed black lines) of 50 modulated ensemble covariances with compact support length $L_{MC} = 61$ . c) as for b) but now $L_{MC} = 101$ . d) as for b) but with $L_{MC} = 141$ . . . . .	62
3.5 The relation between $\sigma$ as appearing in (3.16) and the value for $L_{MC}$ for which the RMS error between the true covariance and the localized ensemble covariance is minimal. Dashed line shows the linear fit to the data. . . . .	63
3.6 As figure 3.4, but now using ensemble members having (3.18) as true covariance function. . . . .	65

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
3.7 Power spectral density as function of the wave number $\kappa$ for raw ensemble members (black) and for the modulated ensemble members with $L_{MC} = 101$ and $M_{MC} = 400$ (grey). Solid lines show the power spectral density averaged over 500 ensemble members. Dashed lines show the 5% and 95% percentiles. The power spectrum of raw ensemble members as given by (3.15) is shown as a line of stars. . . . .	67
3.8 Ensemble covariance (in $10^{-3}\text{C}$ ) on 0000 UTC 4 May 2011 between the sea surface temperature at the location marked by the dot and the sea surface salinity field in the Columbia River plume (black line). Left: for a nonlocalized 50-member ensemble. Right: for the same 50-member ensemble but now using MC localization with $M_{MC} = 15000$ and $L_{MC} = 51$ in both zonal and meridional direction corresponding to a localization distance of 100 km. In the area enclosed by the dashed lines the nonlocalized sample correlation is significantly different from zero at the 90%-level. The 200m, 1000m and 2000m isobaths are shown as grey lines. . . . .	69
4.1 Location of (a) HFR, (b) SST and (c) SSH observations in the model for the window starting on 26 August 2011. . . . .	78
4.2 Overview of the En4DVar system. Panels a, b, and c show the progression of tasks. In (a), the main run (task 1, solid line) and ensemble members (with the envelope shown as dashed lines, task 2) are run for six days. The first three days are the analyses (blue), the last three days are the forecasts (red). In (b), task 3 is <b>B</b> calculation from the ensemble, and task 4 is the calculation of the DA corrections for the main run and the ensemble members using information from the observations (black circles). In (c), task 5 are the new six day model runs started from the corrected ocean states. . . . .	81
4.3 (a) $\hat{\sigma}_{EOF,1}\mathbf{w}_{EOF,1}$ , first EOF of the wind field scaled by its standard deviation in the wind perturbations, (b) $\hat{\sigma}_{EOF,2}\mathbf{w}_{EOF,2}$ the second EOF of the wind field scaled by its standard deviation in the wind perturbations, (c) $\mathbf{w}_S$ , example of a small-scale wind error field for one time. Color scale shows the wind error speed of the different wind fields in $\text{m s}^{-1}$ . . . . .	90

## LIST OF FIGURES (Continued)

<u>Figure</u>		<u>Page</u>
4.4 A priori (red) and posteriori (blue) probability density distributions for the variance of the EOF coefficients $\sigma_{EOF,i}^2$ . The variance for which the blue distribution attains its maximum is used as estimate $\hat{\sigma}_{EOF,i}^2$ . For reference, the time-variance explained by each mode is shown as percentage. Indicated by the dash line is 9% value of the time-variance.	. . . . .	91
4.5 Power spectral density of the meridional wind field as a function of the meridional wave number from the interpolated NAM wind field (blue) and the perturbed wind fields of the ensemble members (grey) on 8 August 2011 00:00. Also shown are the fits to the mean of the ensemble power spectral densities (dashed black line) and the NAM power spectral density (dashed blue line) and the 95% confidence interval for the PSD estimates.	. . . . .	93
4.6 Value of the cost function as function of the number of inner loop iterations using different levels of parallelization $N_s$ for the window starting (a) on 31 May 2011 and (b) on 26 August 2011. For reference, the value of the cost function after 12 inner loop iterations with $N_s = 4$ is marked by a dashed black line. Panel c shows how many iterations RBCG need to reach this level compared to cluster search with $N_s$ clusters.	. . . . .	95
4.7 Expected normalized RMSE in (a) SST, (b) HFR daily-averaged surface velocity and (c) SSH observations as function of the number of inner loop iterations for the window starting on 26 August 2011.	. . . . .	96
4.8 DA correction to the surface temperature field (top row), surface velocity field (center row) and sea-surface height (bottom column) based on 12 iterations of the RBCG method (left column) and the differences between this and the DA correction using cluster search with $N_s = 4$ (right column).	. . . . .	98

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
4.9 The RMS of the normalized forecast errors, as contained in vector $\hat{\mathbf{d}}$ , after $\hat{\mathbf{B}}$ -projection expressed as ratio versus the RMS of the total forecast error for (a) SST, (b) HFR and (c) SSH observations for each DA window. Grey area shows the range of this ratio over all ensemble members, while the black solid line marks the ensemble mean of the ratio. Dashed black line marks the value of 1 that would be obtained if the innovation vector $\hat{\mathbf{d}}$ of the ensemble lies completely in the search space. . . . .	99
4.10 (a) Value of the cost function for the main run (green) and ensemble members (grey) prior to cost function minimization. (b) value of the cost function after minimization ( $N_s = 12$ , $I = 12$ ). The 95%-confidence interval for the expected value of the cost function is shown as a pair of dashed black lines. . . . .	101
4.11 RMSE per 3-day window for (a) SST, (b) HFR daily-averaged velocity and (c) SSH observations from the model without DA (blue), the Main run (green), the different ensemble members (grey) and the ensemble average (black). The left side of each line piece marks the RMSE in the analysis, the right side the RMSE in the forecast. . . . .	102
4.12 The 3-day model bias (model-observations) per window for (a) SST and (b) HFR daily-averaged velocity observations from the model without DA (blue), the main run (green), the different ensemble members (grey) and the ensemble average (black). The left side of each line piece marks the bias in the analysis, the right side the bias in the forecast. . . . .	104
4.13 Rank of the perturbed (a) SST, (b) HFR daily-averaged surface velocity and (c) SSH observations within the ensemble. . . . .	105
4.14 Estimates for total error variance (first row), observational error variance (center row) and background error variance (bottom row) based on SST observations (left column), HFR observations (center column) and SSH observations (right column). Blue lines are estimates based on (4.24)-(4.26), while the black lines indicate estimates based on the specified $\mathbf{R}$ and the ensemble variance. . . . .	107
5.1 Overview of the model domain and the boxes used in HBSC. Boxes at the different levels of the hierarchy are shown with different colors. . . . .	118

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
5.2 Surface maps of the background error correlations in the interior ocean, away from strong fronts. Shown are correlations of $T$ at the grey dot location and (left) surface $T$ and (right) surface $S$ . (Top) balanced <b>B</b> , (bottom) ensemble <b>B</b> , where correlations are computed using ensemble member perturbations from all the windows. $T(\mathbf{r}_0)$ -surface velocity correlations are shown as vectors in each panel. . . . .	121
5.3 Surface maps of ensemble background error correlations for the reference point at the inshore edge of the Columbia River front (black line is $S = 31.5$ ). Shown are surface maps of the background error correlations of $T$ at the grey dot location and (a) surface $T$ and (b) surface $S$ . The correlations are computed in a single assimilation window, starting on 9 July 2011. $T(\mathbf{r}_0)$ -surface velocity correlations are shown as vectors in each panel. . . . .	122
5.4 Surface (top) $T-T$ and (bottom) $T-S$ background error correlations from experiment Ens for randomly selected 200-km meridional sections (left) outside and (right) inside the Columbia River plume. Error ensemble correlations are computed for surface $T$ at the center point (distance= 0 km) and SST and SSS along the sections. Grey lines: localized ensemble correlation in each window; dashed: averaged correlations, dotted: average correlations for raw (not localized) ensembles; solid black: the balanced <b>B</b> correlations. . . . .	124
5.5 Point-by-point surface (a-c) temperature-temperature and (d-f) temperature-salinity ensemble covariance for 1 May (1st column), 30 July (2nd column), 28 September 2011 (3rd column). Solid black lines mark the (a-c) $15^{\circ}\text{C}$ isotherm and (d-f) $31.5$ isohaline. . . . .	126
5.6 (a-c) Maximum point-by-point SST-surface velocity ensemble covariance for 1 May (1st column), 30 July (2nd column), 28 September 2011 (3rd column). (d-f) as (a-c) but now for the ensemble surface velocity variance. Solid black lines mark the (a-c) $15^{\circ}\text{C}$ isotherm and (d-f) $31.5$ isohaline. . . . .	128

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
5.7 Change in fresh water volume in the river plume ( $S < 31.5$ ) since 19 April 2011 in experiment No DA (dark blue) and the analyses from experiments Ens (light blue), Ens-HBSC (green), Bal-HBSC (red). Also shown is the cumulative discharge of the Columbia River since 19 April 2011 (black). . . . .	129
5.8 The study of DA-induced SSS changes, 21 July 2011: (a) forecast SSS in experiment Ens, prior to DA correction; the black line is $S = 31.5$ ; (b) analysis SSS, after DA; (c) difference in SST between the Ens forecast and observed SST; and (d) the DA correction to SSS field. In (c) and (d), contours show the point-by-point SST-SSS covariance $-0.39$ , $-0.26$ and $-0.13$ °C. . . . .	131
5.9 Time series of the area-averaged RMSE for (a) SST, (b) daily-averaged zonal velocity, (c) daily-averaged meridional velocity and (d) SSH observations with along-track mean removed. Results from experiment No DA (dark blue), experiment Ens (light blue), experiment Ens-HBSC (green) and Bal-HBSC (red) are shown. Dots at the left (right) side of the DA line segments correspond to the analysis (forecast). . . . .	133
5.10 Surface maps of time-averaged RMSE. (Top to bottom): SST, SSH, and velocity. (Left) Case no DA RMSE; (center) Bal-HBSC forecast RMSE minus No DA RMSE; (right) Ens-HBSC forecast RMSE minus Bal-HBSC forecast RMSE. Ellipses in the bottom row indicate the second moment (RMSE squared) of the errors in the surface velocities. The scale circle indicates a RMSE squared of $0.1 \text{ m}^2 \text{s}^{-2}$ . The 200, 1000 and 2000 m isobath contours are shown as grey lines. . . . .	135
5.11 Average surface currents from (a) HFR observations, (b) experiment No DA and from the forecasts produced by experiments (c) Bal-HBSC and (d) Ens-HBSC over the period 19 April to 1 October 2011. The 200, 1000 and 2000 m isobaths are shown as grey lines. . . . .	136
5.12 Location of in-situ observations. Shown are glider positions prior to 30 June 2011 (red), after 30 June 2011 (purple), Argo floats outside the plume (green) and inside the river plume (blue) as well as the location of the NH10 buoy (black). The 200, 1000 and 2000 m isobath contours are shown as grey. . . . .	138

## LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
5.13 Glider-observed and model forecast $T$ - $S$ diagrams along the glider transect near $44.65^{\circ}\text{N}$ . (Top) 19 April 00:00 to 30 June 00:00 2011, (bottom) 30 June 00:00 through 1 October 2011 00:00. (Left to right): observed, No DA, Bal-HMSC, Ens-HMSC. Colors indicate the depth at which the observations are taken. Black solid line is $T = 13.5^{\circ}\text{C}$ and the black dashed line corresponds to the assumed $\delta T = -\alpha^{-1}\delta S$ relation in the balance operator $\mathbf{B}$ (E.2). . . . .	139
5.14 Vertical profiles of the time-averaged, 19 April 2011 to 1 October 2011 (Top) RMSE and (bottom) bias in NH10 temperature. (left) model analyses (right) forecasts. Experiments: No DA (dark blue), Ens (light blue), Ens-HBSC (green) and Bal-HBSC (red) over the period 19 April 2011 till 1 October 2011. The “+” symbol marks the RMSE and bias between the GOES SST observations in the 6 km radius around the mooring and the mooring temperature at 2 m depth (i.e. GOES minus mooring temperature). . . . .	141
5.15 As Figure 5.14, but now for salinity. . . . .	142
5.16 Time-average temperature-depth profiles (top row) and salinity-depth profiles (bottom row) from ARGO observations (black) and forecasts from experiment No DA (blue), Ens (light blue), Ens-HBSC (green) and Bal-HBSC (red). Separate average profiles are shown based on profiles taken while the float was outside the river plume (a,c) and while the float was in or beneath the river plume (b,d). See Figure 5.12 for the location of the ARGO floats. . . . .	144

## LIST OF TABLES

<u>Table</u>	<u>Page</u>
2.1 RMSE for different types of observations and experiments. Top block: Depth-averaged subsurface (only including observations below 25 m depth) analysis RMSE for the period 21 July to 10 August 2011 (see Figure 2.11a-c). Center block: As top block, but now for the forecast RMSE (see Figure 2.11d-f). Bottom block: Forecast RMSE error over the period 8 August to 1 September (see Figure 2.14). In each row the case with the lowest RMSE is emphasized. . . . .	32
3.1 Proportionality constant $\alpha$ ( $\pm$ standard deviation) between $\hat{L}_{MC}$ and $\sigma$ as function of the number of raw ensemble members $K$ . . . . .	63
4.1 RMS and the mean of the difference experiment minus observations as shown in Figure 4.11 and 4.12 for the different models and over the period 22 April to 28 September 2011. . . . .	103
5.1 Time-averaged RMSE, 19 April to 1 October 2011. Smallest analysis and forecast RMSE for each observation type are highlighted. . . . .	134

## LIST OF APPENDICES

	<u>Page</u>
A Mixing Algorithm . . . . .	176
B Fourier Transform of the Modulated Ensemble Members . . . . .	177
C Pseudocode Minimization Algorithms . . . . .	180
D Estimation Conditional Distribution $\sigma_{EOF,i}$ . . . . .	184
E Balance Operator Covariance . . . . .	188
F Ensemble for the Balance Operator . . . . .	190

## LIST OF APPENDIX TABLES

<u>Table</u>	<u>Page</u>
C.1 Pseudocode for RBCG using (4.6) for ensemble member $m$ . . . . .	180
C.2 Pseudocode for block diagonal CG with B-preconditioning using (4.6). .	181
C.3 Pseudocode for the cluster search method. $\mathbf{e}_j$ is a unit vector in direction $j$ . . . . .	181

“Zoo’n waterplas heeft ’t maar goed, die golft maar en weerspiegelt de wolken, is  
aldoor anders en blijft toch gelijk. Heeft nergens last van.”

*Nescio (1911, p. 48)*

## Chapter 1: General Introduction

In 1922 Lewis Fry Richardson published on his first, failed, attempt to predict the weather near Munich, Germany (*Richardson*, 1922). Since then numerical weather prediction (NWP) has evolved into an industry that is indispensable for modern society. In the second half of the 20th century, the need for similar products for the ocean arose. Initially, this was driven by the Navy that needed reliable estimates of the thermal structure to support anti-submarine warfare (*Burnett et al.*, 2014). Nowadays, numerical ocean forecasts are also used by fishermen to locate the best fishing grounds, shipping companies for optimal ship routing, and government agencies to track marine debris, provide guidance in oil spill clean up efforts, e.g. those carried out after the Deepwater Horizon disaster (*Liu et al.*, 2011), and to predict harmful algae blooms that would necessitate the closure of shellfish farms and beaches (*Jochens et al.*, 2010).

Predictions in both operational NWP systems as well as numerical oceanographic prediction systems are produced following the same basic procedure: at fixed time intervals, or time windows, a forecast is issued; observations coming in during the next time window are compared with the forecast; based on this comparison a correction to the forecast is calculated. The corrected ocean/atmosphere state is then used as starting point for a, possibly nonlinear, numerical model in the next window (see Figure 1.1). The process in which the forecast is combined with the observations in order to obtain a more accurate estimate of the true ocean/atmosphere is called data assimilation (DA). In this framework, the forecast prior to DA is referred to as the background, while the corrected ocean/atmosphere state is referred to as the analysis.

One could construct a naive DA system by replacing, at the observation location and time, the model result with the observed values at the observation locations and times. This concept suffers from different flaws when applied to the ocean. First, it creates unphysical, discontinuous ocean states. The ocean will start to restore

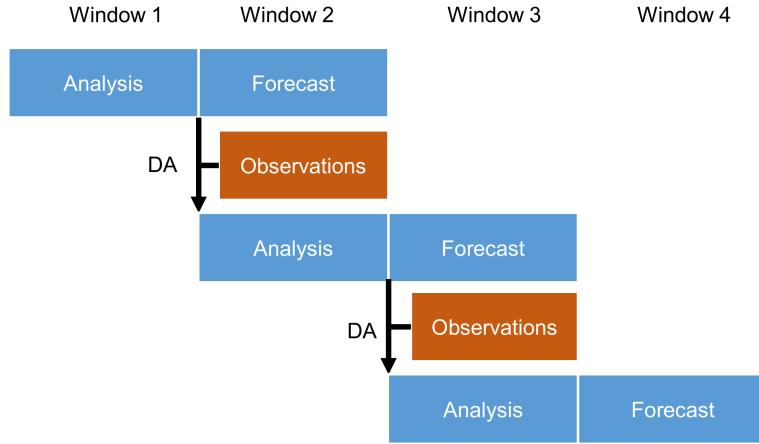


Figure 1.1: Schematic overview of an operational numerical weather prediction or ocean prediction system.

geostrophic balance. The waves generated during this process can actually deteriorate the analysis (*Daley, 1981; Lynch and Huang, 1992*). Second, it neglects the presence of observational error in the measurements. Third, if the model is erroneous at the observation location, it can reasonably be expected that it is erroneous at nearby points too. A more rigorous way to deal with the errors is to determine the most likely value of the background error given the observations and subtract this estimate from the background. More formally, given observations  $\mathbf{y}$  the most-likely estimate for the background error  $\hat{\epsilon}_B$  is found by maximizing the probability density (see e.g. *Lorenc, 1986; Purser, 1984*)

$$p(\epsilon_B | \mathbf{y}) \sim p_R(\mathbf{y} | \epsilon_B) p_B(\epsilon_B). \quad (1.1)$$

Here Bayes' theorem has been used with  $\sim$  indicating proportionality;  $p_R$  and  $p_B$  are the observation error and background error probability density distributions, respectively. Though the selection of suitable probability distributions is still an area of active research (see *Bocquet et al., 2010*, and references therein), the current practice in weather and ocean forecasting systems is to assume that the observational and

background errors follow a normal distribution with zero mean and covariances  $\mathbf{R}$  and  $\mathbf{B}$  respectively. It can be shown that under these assumptions maximization of  $p(\epsilon_B|\mathbf{y})$  is equal to minimization of a quadratic cost function  $J(\epsilon_B)$  (*Lorenc*, 1986).

The ocean state in the numerical models is stored as a high-dimensional ( $O(10^7) - O(10^9)$ ) vector. As a consequence, the optimization of (1.1) is computationally intensive and explicit storage of  $\mathbf{B}$  is impossible, even at the current levels of memory available in supercomputer clusters. Over the last two and half decades two families of DA methods have been developed to deal with these limitations: ensemble Kalman filters/smoothers (*Evensen*, 1994; *Evensen and van Leeuwen*, 2000; *Houtekamer and Mitchell*, 1998; *Oke et al.*, 2002a) and variational methods like 3DVAR (*Sasaki*, 1970a,b), 3DVAR-FGAT (*Fisher and Andersson*, 2001), and 4DVAR (*Bennett*, 1992; *Chua and Bennett*, 2001; *Courtier et al.*, 1994; *Dimet and Talagrand*, 1986; *Lewis and Derber*, 1985; *Navon and de Villiers*, 1983). As Figure 1.2 shows, the methodologies differ in whether and how they evolve  $\mathbf{B}$  in time and when the observations are used. In Kalman filters and 3DVAR it is assumed during the maximization of (1.1) that all observations are taken at the same time (see Figure 1.2a,b). Both in 3DVAR-FGAT and 4DVAR observations can be compared to the background state at the actual observation time (see Figure 1.2c,d). But whereas in 3DVAR-FGAT it is assumed that the background error is time-independent ( $\epsilon(t) = \epsilon(t_0)$ ) throughout the window,  $\epsilon(t_0)$  is explicitly propagated forward in time in 4DVAR. Consequently, 4DVAR is capable of detecting the fastest growing error modes (*Thépaut et al.*, 1996) and correct for these in the DA correction. Due to its abilities to capture error dynamics and to assimilate observations at the correct time, 4DVAR outperforms other variational methods tested in the NWP framework (*Lorenc and Jardak*, 2018; *Lorenc and Rawlins*, 2005). However, traditionally, in operational implementations of 4DVAR in NWP and ocean prediction (e.g. *Kurapov et al.*, 2011; *Moore et al.*, 2011),  $\mathbf{B}$  is assumed to be the same in each assimilation window. While at the same time in the Kalman filters/smoothers  $\hat{\epsilon}_B(t_0)$  is estimated from an ensemble and the growth of  $\mathbf{B}$  after  $t_0$  is explicitly taken into account by propagating the ensemble members forward in time using the nonlinear numerical model.

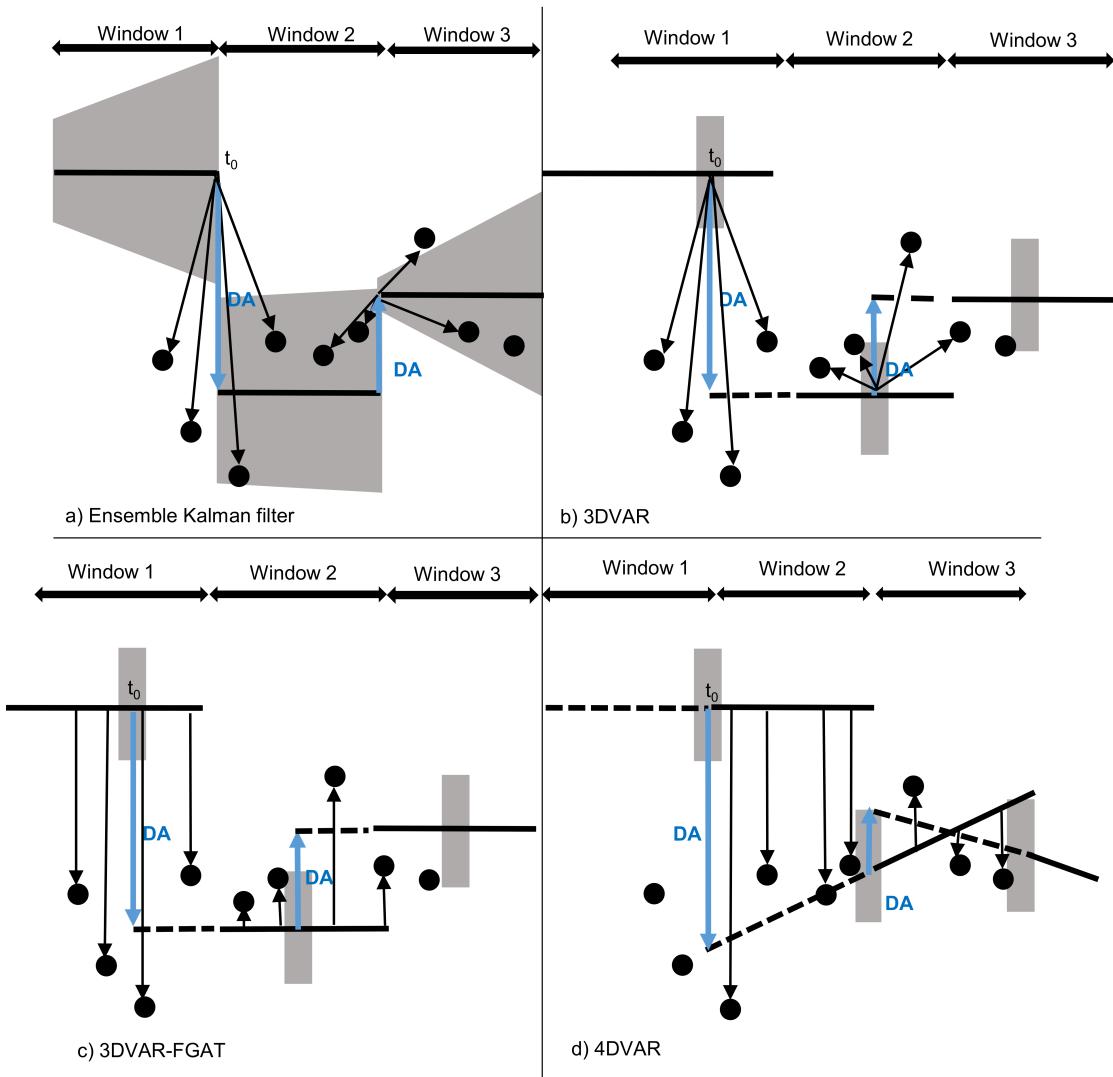


Figure 1.2: Schematic overview of two DA corrections using (a) the ensemble Kalman filter, (b) 3DVAR, (c) 3DVAR-FGAT, (d) 4DVAR. Observations (black dots) are compared with the forecasts (solid black lines) for the time indicated by the black arrows. Based on this comparison and the magnitude of the background error covariance (grey) a DA correction (blue) is calculated. This corrected state then serves as the initial condition for the next analysis (dashed black lines).

The central question of this study is whether combining the 4DVAR and the ensemble method in the En4DVar hybrid can lead to better coastal ocean prediction. Many aspects of this study were motivated by our aspiration to improve the performance of the pilot Oregon-Washington (OR-WA) Coastal Ocean Forecast System (*Kurapov et al.*, 2011; *Yu et al.*, 2012) that has been run operationally at OSU. This system uses 4DVAR in 3-day windows and produces daily updates of 3-day forecasts of the temperature, salinity, velocities and sea-surface height for the Pacific Ocean offshore of Oregon and Washington, at the U.S. West coast.

Currently, only remote observations of the ocean surface are assimilated into the OR-WA system. In absence of observations, the subsurface structure of the DA corrections derives in large part from **B**. Assimilation of in-situ subsurface observations of temperature and salinity by low-power autonomous underwater vehicles, gliders, has the potential to constrain the subsurface model error. While efforts are underway to establish a denser network of gliders as part of the Ocean Observing Initiative (OOI) project, supported by the National Science Foundation (NSF), the glider coverage in the area is currently very limited. Our concern has been that in the absence of surface observations, the assimilation of glider observations can create local DA corrections to subsurface stratification with the horizontal scale set by **B**. For coastal problems this scale is usually chosen to be close to the baroclinic Rossby radius of deformation. Then by virtue of the error covariance and/or dynamical adjustment one could expect that DA will generate baroclinic eddies around the assimilation sites. In chapter 2, currently under review with the Journal of Geophysical Research: Oceans, we investigate whether this is a serious problem. In the same study we test whether combined assimilation of subsurface observations with area-covering surface observations can prevent unphysical corrections from forming and direct the DA system to make subsurface corrections with accurate horizontal length scales.

The Columbia River plume (*Berdeal et al.*, 2002; *Hickey et al.*, 2005; *Liu et al.*, 2009) and seasonal coastal upwelling (*Castelao and Barth*, 2005; *Halpern*, 1976; *Huyer*, 1983) create strong temperature and salinity fronts in this region. These fronts move under the influence of changing winds creating an ocean state that is

highly variable both in space and in time. Considering this, it is improbable that the current assumption in the OR-WA system that  $\mathbf{B}(t_0)$  is time-independent holds. To capture the errors of the day in  $\mathbf{B}(t_0)$ , we have come up with a new DA setup for the OR-WA system. In this new setup  $\mathbf{B}(t_0)$  used in the 4DVAR method will be obtained from an ensemble of 4DVAR runs (En4DVar). This combines the ability of the ensemble Kalman filter to adapt  $\mathbf{B}(t_0)$  to the changing ocean conditions with all the advantages of 4DVAR. Obtaining a good estimate for  $\mathbf{B}(t_0)$  from an ensemble of limited ( $O(10^1) - O(10^2)$ ) size is nontrivial. One of the problems is the occurrence of sampling errors in the ensemble covariance. Localization is commonly used to remove the spurious long-distance correlations created by the sampling errors (e.g. *Bishop and Hodyss*, 2007; *Buehner et al.*, 2009; *Clayton et al.*, 2013; *Houtekamer and Mitchell*, 2001). It has the additional benefit that it increases the rank of  $\mathbf{B}$ . However, localization can be time-consuming. In chapter 3 a computationally efficient localization method, called Monte Carlo localization, is presented, to overcome this problem. This work also appeared as a publication in Monthly Weather Review (*Pasmans and Kurapov*, 2017).

Not only ensemble localization, but also the optimization of (1.1) in 4DVAR is computationally demanding. In 4DVAR, the cost function minimization requires repeated, sequential application of the adjoint (ADJ) model,  $\mathbf{B}(t_0)$  and the tangent linear (TL) model in iteration after iteration. The need in En4DVar to do this for the entire ensemble, in order to provide a correction to every member, only compounds these computational demands. In chapter 4 we show that by making some approximations it is possible to simultaneously apply 4DVAR to all ensemble members by running  $N_s$  ADJ-TL computations in parallel, where  $N_s$  is relatively small (e.g., in our applications  $N_s = 4$ ). This new cluster search method, allows us to reduce the number of iterations, and thus the wall time needed to reach a certain level of cost function reduction. This comes at a price of running  $N_s$  ADJ and TL model computations simultaneously. This is certainly feasible for our relatively small OR-WA grid and can be scaled to larger prediction systems run on today's supercomputers (*Kurapov et al.*, 2017). In chapter 5, the performance of the En4DVar system is evaluated

and compared to that using the “traditional” 4DVAR with the static **B**.

Finally, summary and closing remarks are provided in chapter 6.

Chapter 2: Why Gliders Appreciate Good Company: Glider  
Assimilation in the Oregon-Washington Coastal Ocean 4DVAR  
System with and without Surface Observations

I. Pasmans, A.L. Kurapov, J.A. Barth, A. Ignatov, P.M. Kosro, R.K. Shearman

Journal of Geophysical Research: Oceans  
2018 (*Submitted*).

## 2.1 Abstract

Giders are low-power autonomous underwater vehicles used to obtain oceanic measurements in vertical sections. Assimilation of glider temperature and salinity into coastal ocean circulation models holds the potential to improve the ocean subsurface structure estimate. In this study, the impact of assimilation of glider observations is studied using a 4DVAR data assimilation and forecast system set offshore of Oregon and Washington on the U.S. West Coast. Four test cases are compared: (1) no assimilation, (2) assimilation of glider temperature and salinity data alone, (3) assimilation of the glider data in combination with the surface observations including satellite SST, SSH and high-frequency radar (HFR) surface velocities, and (4) assimilation of the surface data alone. It is found that the assimilation of glider observations alone creates unphysical eddies in the vicinity of the glider transect. As a consequence, the forecast errors in the surface velocity and temperature increase compared to the case without data assimilation. Assimilation of surface and subsurface observations in combination prevents these features from forming and reduces the errors in the forecasts for the subsurface fields compared to the other three experiments. These improvements persisted in 21-day forecasts run after the last data assimilation cycle.

## 2.2 Introduction

The Oregon State University (OSU) coastal ocean forecast system provides daily updates of 3-day forecasts of shelf currents, temperature and other physical variables of interest along the Oregon (OR) and Washington (WA) coasts (*Erofeeva, 2018; NANOOS, 2018*). The information provided by the system has been used as guidance by local fishermen, a government agency in charge of environmental hazard response, and other users. The system uses data assimilation (DA) to improve the accuracy of initial conditions for the forecasts. Currently, only surface observations are assimilated, including satellite sea surface temperature (SST), satellite sea surface height (SSH) (*Kurapov et al., 2011*), and surface currents from a network of land-based high-frequency radars (HFR) (*Yu et al., 2012*). Subsurface temperature and salinity

observations by gliders, autonomous platforms capable of making subsurface in-situ measurements, have been available since 2006 (*COAS*, 2016), but have until recently been limited to single transects like the Newport transect. Because of their limited availability, subsurface in-situ observations have so far not been included. The ongoing Ocean Observatories Initiative project (*OOI*, 2018) includes an effort for regular deployment of a fleet of gliders along several lines in this region and thus can provide a more extensive source of subsurface information in the future. The possibility of using these measurements in our forecast system has encouraged us to investigate the impact of glider DA using the 2011 archive data from just one section in Oregon.

The subsurface salinity ( $S$ ) and temperature ( $T$ ) measurements taken by the gliders can potentially provide a valuable constraint on the subsurface stratification and location of fronts and eddies along the glider path. However, assimilation in short time intervals will use data only along a limited segment of the track and will most probably result in local corrections to the vertical stratification, in particular, if a relatively short (25 – 50 km) horizontal decorrelation scale is assumed for the model background errors (*Kurapov et al.*, 2011). If the background error covariance implies dynamical balances in the correction, including geostrophy and thermal wind balance (*Weaver et al.*, 2005), this local correction in the vertical stratification will yield horizontal density gradients that will be balanced by a correction in the baroclinic currents. These corrections will take the form of eddies that could potentially degrade the forecasts. Even if the balanced covariance is not utilized, the local correction in initial conditions for the density will eventually be balanced by changes in the velocity field by dynamical adjustment. So we hypothesize that in-situ hydrographic profile data assimilation can result in spurious baroclinic eddy generation that will deteriorate forecasts. Since surface observations including satellite SST, SSH and HFR surface currents all contain information about the eddies, their assimilation in combination with the in-situ glider data can potentially allow the fit to the subsurface data without erroneous eddy generation. In that case, the correction to the vertical stratification should extend to distances larger than the Rossby radius of deformation.

Reports of earlier test studies involving glider DA are a mixture of success stories

and signs of complications. *Pan et al.* (2014) compared results of glider DA to mooring data west of Florida and found a reduction of the root-mean-square error (RMSE) with respect to mooring  $T$  and  $S$  in the forecasts compared to the control run (model without DA). *Jones et al.* (2012) compared the DA analyses and the no-DA model results southeast of Tasmania and found reduction in the nearshore mooring temperature RMSE as well as a 40% reduction in the RMSE in satellite SST observations that were not assimilated. *Shulman et al.* (2009) found that after assimilating data from 10-15 gliders in an approximately  $100 \times 200$  km area near Monterey Bay (Central California) the RMSE in mooring temperature and salinity decreased and the correlation between mooring temperature, salinity and velocities increased compared to a model without DA. These studies do not report on adverse eddy variability introduced by the glider assimilation. In both *Shulman et al.* (2009) and *Jones et al.* (2012), the data assimilation correction was added to the model in small increments to avoid shocks to the system. This potentially helped to alleviate the problem of bogus eddy variability. Combined assimilation of surface and subsurface data has also been carried out successfully. E.g., *Ngodock and Carrier* (2014b) assimilated several glider transects in Monterey Bay in combination with SST and SSH and found a lower number of large discrepancies between observations and forecast predictions than in the control run. *Zhang et al.* (2010) assimilated glider observations in combination with HFR currents and satellite SST in the New York Bight. They reported 25%-60% reduction in the analysis RMSE for these observations (i.e. data fits) compared to forecasts. Reduction of the subsurface temperature was attributed in part to SST assimilation. *Matthews et al.* (2012) assimilated SST, SSH, and subsurface glider  $T$  and  $S$  and also found reductions in the RMSE for these observations in the analyses (up to 40%) compared to a control (no-DA) run. At the same time, however, there are indications that the assimilation of solely glider observations can be problematic. E.g., *Melet et al.* (2012) disclose that assimilation of  $T$  and  $S$  from 10 gliders off Samoa actually produced analyses with higher RMSEs in the model  $T$  and  $S$  fields than in the control run serving as truth. *Li et al.* (2013) found that assimilation of glider observations alone created spurious small-scale structures in the surface velocity field

in Prince William Sound, Alaska. *Dobricic et al.* (2010) found that assimilation of glider observations alone in the eastern Ionic Sea reduced the RMSE in the surface and subsurface temperature and salinity observations, but increased RMSEs in SSH and current observations in the top 200m of the water column by as much as 23% compared to the control run.

In this study, we want to find ways in which glider observations can be assimilated usefully in our OR-WA coastal ocean forecast system. In particular, we will test the following hypotheses: (1) assimilation of glider observations alone in an area of the continental shelf and slope can create unobserved eddy variability and degrade the forecast accuracy; (2) assimilation of the glider observations in combination with surface observations yields better forecast error statistics than those obtained by assimilating only glider observations or only surface observations; (3) combined surface-glider data assimilation can help extend the zone of impact of the glider data beyond the direct vicinity of the glider transect. This study looks at a combination of conditions that have not been considered in previous studies: it uses a 4DVAR DA system in a series of time intervals, the glider is assimilated along only one transect and is located in the immediate vicinity of the seasonal upwelling front with a large river plume being present on the offshore side of upwelling front. Furthermore, we make a comparison of the non-DA model, the glider-only, combined glider-surface and surface-only DA cases and focus not only on the data fits (analyses) but also on 3-day and longer-term forecasts.

This paper is arranged as follows: section 2.3 describes the model setup, the assimilated observations and our DA system. Section 2.4 looks at the corrections in these experiments and the accuracy of the forecasts in the different cases. Finally, section 2.5 will conclude to what extent the results agree or disagree with our three hypotheses.

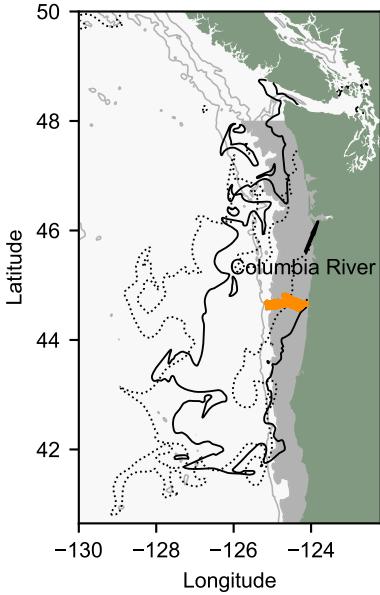


Figure 2.1: Overview of the model region together with the assimilated glider transects (orange), the 200 m, 1000 m, 2000 m isobaths (grey lines) and the coastal shelf region used in the surface observation comparison in Figures 2.14 and 2.15 (grey). Also shown are the 31.5 isohaline on 21 July 2011 (black solid line) and the extent of the Columbia River plume on 13 August 2011 (dotted black line) according to the model without DA.

## 2.3 The Dynamical Regime and the Model Setup

### 2.3.1 Dynamical Regime

The model domain is shown in Figure 2.1. The summer conditions in this region are characterized by predominantly southward winds that force offshore Ekman transport in the surface layer, upwelling, and a baroclinic equatorward coastal jet (*Huyer, 1977, 1983; Kosro, 2005; Mooers et al., 1976*). As the upwelling develops, the coastal current exhibits instabilities and jets separate off the shelf carrying cold, nutrient-rich coastal waters offshore (*Barth et al., 2000; Koch et al., 2010; Oke et al., 2002a,b,c*). The near-surface salinity variability in our study domain is influenced by the outflow of the Columbia River at 46.25°N (*Hickey et al., 1998*). Here we use the 31.5

isohaline contour from our model solution to show the edge of the plume. During winter the river plume typically flows northward along the coast. During summer the wind and coastal upwelling displace the plume south and, offshore (dashed line in Figure 2.1). During periods of relaxation from upwelling conditions and episodic downwelling events, the plume is pressed against the Oregon coast south of the river mouth (solid line) (*Hickey et al.*, 2005; *Liu et al.*, 2009).

### 2.3.2 Model

The ocean state is simulated using the Region Ocean Modeling System (ROMS) version 3.6 ([www.myroms.org](http://www.myroms.org)). This is a nonlinear, finite-volume, hydrostatic, Boussinesq model featuring advanced numerics (*Shchepetkin and McWilliams*, 2003, 2005). The model grid resolution is approximately  $2 \times 2$  km in the horizontal and 40 terrain-following layers in the vertical direction. Baroclinic modes are resolved using a time step of 90 s, while the barotropic time-step is 3 s. The bathymetry is obtained by averaging depths from the U.S. Coastal Relief Model (*NCEP*, 2003) and ETOPO2 (*NCEP*, 2006) and smoothed using *Sikirić et al.* (2009). Boundary forcing for all non-tidal fields is taken from the Hycom-NCODA  $1/12^\circ$  analysis (*COAPS*, 2015), while barotropic tidal components at the boundary are added from the TPXO tidal model (*Egbert and Erofeeva*, 2002, 2010). The boundary conditions are imposed using *Chapman* (1985) for the free surface, *Flather* (1976) for the barotropic velocities and mixed radiation-nudging boundary condition scheme for the baroclinic velocities, temperature, and salinity (*Marchesiello et al.*, 2001). Surface momentum and heat fluxes are calculated using the ROMS bulk flux parametrization. The wind, surface air pressure, net short-wave radiation, downward long-wave radiation, air temperature and relative humidity fields required by the bulk flux parametrization (*Fairall et al.*, 2003) are taken from the NAM analysis (*NCEP*, 2011). The model does not include evaporation or precipitation. Mellor-Yamada 2.5 is used as the turbulent closure scheme (*Mellor and Yamada*, 1982). The Columbia River estuary is represented as an idealized 50 km long channel. The Columbia River together with 15 rivers in the Salish Sea (*Banas et al.*, 2015) are implemented as point sources with a uniform

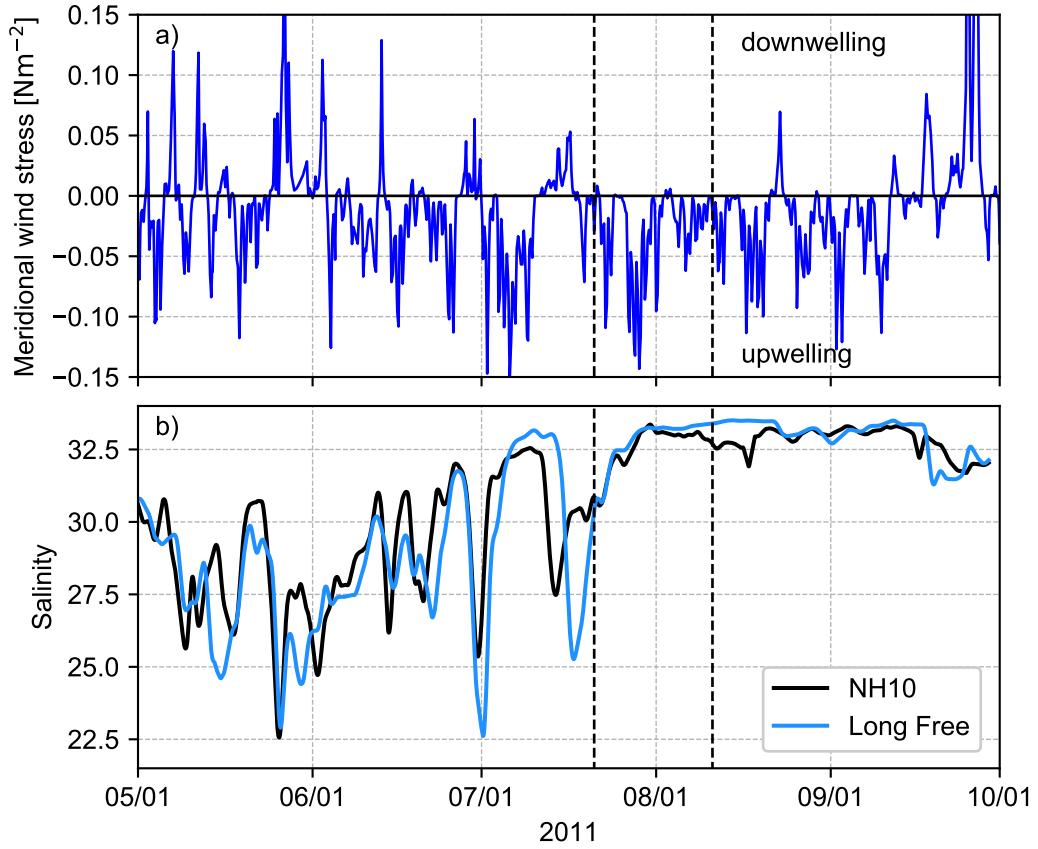


Figure 2.2: (a) Meridional wind-stress applied to the models close to the glider transects at  $124.308^{\circ}\text{W}$ ,  $44.634^{\circ}\text{N}$ . The assimilation period (21 July to 11 August), is marked by the vertical dashed lines. The negative wind-stress corresponds to upwelling favorable northerly winds. (b) The surface salinity measured at 2 m depth at the NH10 buoy ( $124.308^{\circ}\text{W}$ ,  $44.634^{\circ}\text{N}$ ) (black) and obtained from the long free model run without DA (blue) both filtered using a double application of a 24h-running mean filter.

discharge throughout the water column. The Columbia River discharge is based on measurements at the U.S. Geological Survey Beaver Army Terminal (*USGS*, 2015). The river temperature is provided as climatology based on several years of observations at the same station. The discharge and river temperature of the Fraser River and the small rivers in Puget Sound are based on the climatology obtained from

data measured by *Environment Canada* (2015); *USGS* (2015) and provided by P. MacCready and S. Giddings (*Giddings et al.*, 2014; *MacCready and Giddings*, 2016). During the analysis period (21 July-11 August), the river discharge of the Columbia River varies between  $6.3$  and  $9.6 \times 10^3 \text{ m}^3\text{s}^{-1}$  with a mean of  $7.7 \times 10^3 \text{ m}^3\text{s}^{-1}$ , which is anomalously high compared to the average discharge of  $6.0 \times 10^3 \text{ m}^3\text{s}^{-1}$  (*USGS*, 2015). The cumulative outflow of the other rivers varies between  $3.8$  and  $5.3 \times 10^3 \text{ m}^3\text{s}^{-1}$  with a mean of  $4.6 \times 10^3 \text{ m}^3\text{s}^{-1}$ . Model runs were carried out on a local cluster at OSU as well as the COMET supercluster via the XSEDE framework (*Towns et al.*, 2014).

### 2.3.3 Surface Observations

Along the Oregon coast, a network of HFR stations was in place in 2011 primarily covering the Oregon shelf and shelf break between  $41 - 46^\circ\text{N}$ . Hourly radial surface velocity components are available from each site (*Kosro*, 2017). Observations are binned into super-observations by vector-averaging in approximately  $5 \times 5 \text{ km}$  boxes. If the directional spread of the velocities within a box is more than  $9^\circ$  the observations in the box are rejected. Hourly time series in a single day are verified at each point and incomplete series are also rejected. The full-time series at a point is averaged over a day, and the daily-averaged data are matched to the daily-averaged model radial component estimate at the same point.

SST is obtained from NOAA19 and Metopa-A satellites (*Ignatov et al.*, 2016). For each hour the observations, which are available on a  $0.02^\circ \times 0.02^\circ$  grid, are binned into the cells of a  $4 \times 4 \text{ km}$  grid and the observations within each grid cell are averaged to create super-observations. Starting from the end of a 3-day data assimilation window and working towards the beginning, the resulting observations are then thinned such that the time between two observations in the same grid cell is at least three hours. These observations are assimilated as instantaneous point measurements.

Alongtrack sea surface anomaly plus the mean dynamic topography from the Jason-1, Jason-2, Cryosat and Envisat satellites are obtained from *Scharroo et al.* (2013). The slope in these observations provides information about relatively slowly changing, subinertial, surface geostrophic currents and these data are used to con-

strain the non-tidal, 24-hour averaged model SSH slope. To match these data and the tide-resolving model, the following procedure is used. Model tidal harmonic constants are obtained from the analysis of a long no-DA run with T-TIDE (*Pawlowicz et al.*, 2002). The model tidal SSH is computed using eight dominant harmonic constants at the points along the altimetry track for a 24h-period centered around the observation time. This signal is then time averaged and added to the satellite-derived absolute dynamic topography. Geographically correlated errors are known to contribute significantly to the total altimetry error (*Dettmering and Bosch*, 2010; *Labroue et al.*, 2012). To remove these large-scale errors as much as possible, the mean of each track through the model domain is calculated and removed. In the DA scheme, the resulting observation is matched to the 24h-averaged model SSH minus the along-track mean.

### 2.3.4 Subsurface Observations

Glider temperature and salinity observations along the Newport line ( $44.65^{\circ}\text{N}$ , see Figure 2.1) are binned on a grid with a 1 km resolution in the horizontal direction, along the glider track, and 4 m in the vertical direction. The profiles are provided together with the variance of the raw observations for each point (*Erofeev*, 2015; *Mazzini et al.*, 2014; *Saldías et al.*, 2016). For each vertical profile the observations are boxed and averaged within each grid cell. Observational variances for the boxed glider observations are calculated as follows:

$$\sigma_{\xi}^2 = \frac{1}{M} \sum_{j=1}^M \sigma_{\xi,j}^2 + \frac{1}{M} \sum_{j=1}^M (\xi_j - \bar{\xi})^2, \quad (2.1)$$

where  $\xi$  is either glider  $T$  or  $S$ ,  $M$  is the number of profile observations in the box,  $\sigma_{\xi,j}$  is the standard deviation of the  $j$ -th observation as specified in the data, and  $\bar{\xi}$  is the average of the observations in the box. An additional  $(5.7 \cdot 10^{-2} \text{ }^{\circ}\text{C})^2$  or  $(1.30 \times 10^{-2})^2$  is added to this to account for the measurement error variance in  $T$  or  $S$ , correspondingly. These boxed and averaged observations are assimilated as

instantaneous point measurements.

### 2.3.5 Data Assimilation (DA)

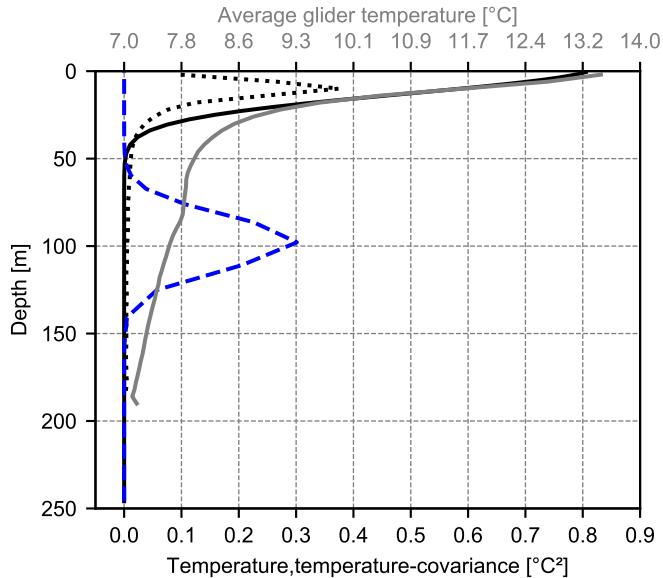


Figure 2.3: Background error temperature-temperature covariance between a point on the surface at  $124.843^{\circ}\text{W}$ ,  $43.005^{\circ}\text{N}$  and the vertical (solid black), covariance between a point at 100 m depth and the vertical (dashed blue) and the time-averaged observational variance (i.e. part of the diagonal of  $\mathbf{R}$ ) for the glider temperature observations as used in the 4DVAR algorithm (dotted black). Also shown is the time-averaged vertical temperature profile obtained from glider observations over the period 21 July to 14 August 2011 (solid gray).

DA proceeds in a series of 3-day windows. In each window, the correction to the initial conditions is found using the Advanced Variational Regional Ocean Representer Analyzer (AVRORA) 4DVAR system developed in-house (*Kurapov et al.*, 2009, 2011; *Yu et al.*, 2012). The AVRORA tangent linear and adjoint codes are algorithmically and dynamically consistent with the nonlinear ROMS and their modular structure is very suitable to experimentation with different background error covariances (see e.g. chapter 3) and data functionals. In particular, we can easily implement assimilation

of the SSH slope (see *Kurapov et al.*, 2011) or daily-averaged observations.

After the corrected initial conditions are obtained using AVRORA 4DVAR, the nonlinear ROMS is run for the three days, to yield the analysis, and is continued for another three days providing the forecast. The forecast, saved hourly, provides the background solution for linearization in the next assimilation window. The correction  $\delta\mathbf{x}$  to the ocean state at the beginning of the forecast is found by minimizing the penalty function (*Courtier et al.*, 1994):

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{d} - \mathbf{H}\mathbf{M}\delta\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\mathbf{M}\delta\mathbf{x}). \quad (2.2)$$

Here,  $\mathbf{M}$  is the tangent linear model, i.e. the nonlinear model linearized around the forecast on a  $4 \times 4$  km grid,  $\mathbf{H}$  is the linear operator that maps the ocean state to the space of observations,  $\mathbf{d}$  is the innovation vector, i.e. the difference between the observations and the predictions for those observations obtained by applying  $\mathbf{H}$  to the forecast,  $\mathbf{R}$  is the observation error covariance and  $\mathbf{B}$  is the background error covariance, i.e. the covariance of the errors present in the ocean state at the beginning of the forecast.  $T$  denotes the matrix transpose.

As  $\mathbf{B}$ , the balanced operator covariance (*Weaver et al.*, 2005) is used with a modification to account for the shallow shelf depth (see *Kurapov et al.*, 2011). This covariance assumes that errors in all fields are related to the background temperature errors via a simple linear temperature-salinity relation, the linear equation of state, and thermal wind balance. To speed up computations, the background temperature error covariance is assumed to be Gaussian and separable into two horizontal directions and along the vertical  $s$ -coordinate. The horizontal correlation length scale in each direction is 25 km, corresponding to the open ocean Rossby radius of the deformation for the first baroclinic mode (*Chelton et al.*, 1998). With this choice, the DA system is expected to be able to make corrections for eddies that are absent, or wrongly located, in the forecasts. The validity of this assumption for the horizontal error scale is studied in more detail in section 5.5. The vertical correlation scale is the same at each vertical profile in the terrain following  $s$ -coordinates and is thus proportional to the total depth; it is chosen to be 15 m in 250 m deep water. The

temperature error standard deviation  $\sigma_T$  is reduced exponentially with depth, with a vertical decay scale of  $l_z = 100$  m. To estimate  $\sigma_T(z)$  at the surface ( $z = 0$ ), the daily-averaged temperatures from the long free run are compared with time-series data from National Data Buoy Center buoys 46015, 46022, 46027, 46029, 46041, 46050, 46087, 46088, 46089, 46094, 46211, 46229, 46243, 46244, 46248 (NDBC, 2016) and the standard deviation of the difference is calculated for each buoy. The median of these standard deviations yields  $\sigma_T(0) = 0.9$  °C. It is expected that DA will remove any bias that is present in the model SST field for 19 April 2011 within the first few windows. No further correction for model SST bias is applied.

The observational error covariance matrix  $\mathbf{R}$  is assumed to be diagonal. The initial estimates of the variances for the surface observations are obtained using equations similar to (2.1). The observations are then rescaled based on the following expression (Bennett, 1992; Desroziers *et al.*, 2005):

$$\mathbf{HMBM}^T \mathbf{H}^T + \mathbf{R} = \langle \mathbf{d}\mathbf{d}^T \rangle \quad (2.3)$$

where  $\langle \cdot \rangle$  is the statistical average. Here we approximate the statistical average by the spatial average. Or more specifically, for the window 9–11 July we estimate the background surface error variance using different types of observations in the top 6 m of the water column by  $\langle d_i^2 - \mathbf{R}_{ii} \rangle_\xi$ , where  $\xi$  denotes the observation type,  $i$  runs over the observations of this type and  $\mathbf{R}_{ii} = \sigma_\xi^2$  is the initial estimate of the error variance of the  $i$ -th observation. Based on this comparison we find that the background error standard deviations for the different fields are too small. More specifically, we find that surface temperature should be rescaled by a factor  $\alpha_{glider,T} = 1.3$  (glider) or  $\alpha_{SST} = 1.1$  (satellite). Similarly, surface salinity should be rescaled by  $\alpha_{glider,S} = 12.5$ , surface height by  $\alpha_{SSH} = 8.4$  (SSH) and surface velocities by  $\alpha_{HFR} = 3.8$  (HFR). The balanced operator does not allow to change the background error variance for different fields separately by multiplying each of them with  $\alpha_\xi^2$ . Instead, we multiply the diagonal elements of  $\mathbf{R}$  with  $\alpha_\xi^{-2}$  for all observations of type  $\xi$ . For the surface observations we eventually fix  $\sigma_{SST} = 0.39$  °C,  $\sigma_{HFR} = 4.86$  cm s<sup>-1</sup>,  $\sigma_{SSH} = 6.9$  mm. As the glider observational error variance varies strongly with depth,

$\sigma_{\text{glider},T}$  and  $\sigma_{\text{glider},S}$  are not adjusted using the same technique. Instead, they are found by applying the scaling to observational error variances found using (2.1). To prevent overfitting the glider data, we limit  $\sigma_\xi(z) \geq \hat{\sigma}_\xi \exp(z/l_z)$  where  $\xi$  is either glider temperature or salinity and  $\hat{\sigma}_\xi$  is the median of the rescaled observational error standard deviation in the top 6 m. As Figure 2.3 shows, the resulting values of  $\sigma_{\xi,T}^2$  decay rapidly with depth. The impact of outliers is reduced by rescaling  $\mathbf{R}_{ii}$  to  $(0.1|\mathbf{d}_i|)^2$  if  $|\mathbf{d}_i| > 10\sqrt{\mathbf{R}_{ii}}$ .

Minimization of (2.2) is carried out using the Restricted B-preconditioned Conjugate Gradient method (RBCG) (*Gürol et al.*, 2014). Out of time considerations minimization of (2.2) is terminated after 14 inner loop iterations. Also out of time considerations, no outer loop iterations are performed. Limited parallelization is implemented by using the Orthogonal Multiple Search Direction Conjugate Gradient method (MSDO-CG) (*Grigori et al.*, 2016). In this method 4 new search directions are generated per iteration eventually creating a correction that lies in a subspace of dimension 56. In this respect the current AVRORA implementation differs from the one in *Kurapov et al.* (2009, 2011); *Yu et al.* (2012).

### 2.3.6 Experiments

The model was started on 2 January 2011 from the global Navy Hycom analysis results. After making local corrections to the bathymetry to compensate for insufficient tidal mixing thus reducing the model SST warm bias in the Strait of Juan de Fuca, the model was restarted from 27 February 2011 and ran till 1 October 2011. This case is referred to as the long free model run. To make sure glider DA results are not overshadowed by the shock of DA initialization, we first spin up DA starting on 7 July assimilating SSH, SST and HFR velocities (see the experiment time line in Figure 2.4). The output of this DA model (case *Surface Only*) on 21 July is then used as the initial condition for the comparison of this case and the other three base test cases. In the case *No DA*, data are not assimilated after 21 July. In the *Glider Only* DA case, glider *T* and *S* are only assimilated. In the *Combined* DA case, all aforementioned observations are used. The last DA correction (cases Surface Only,

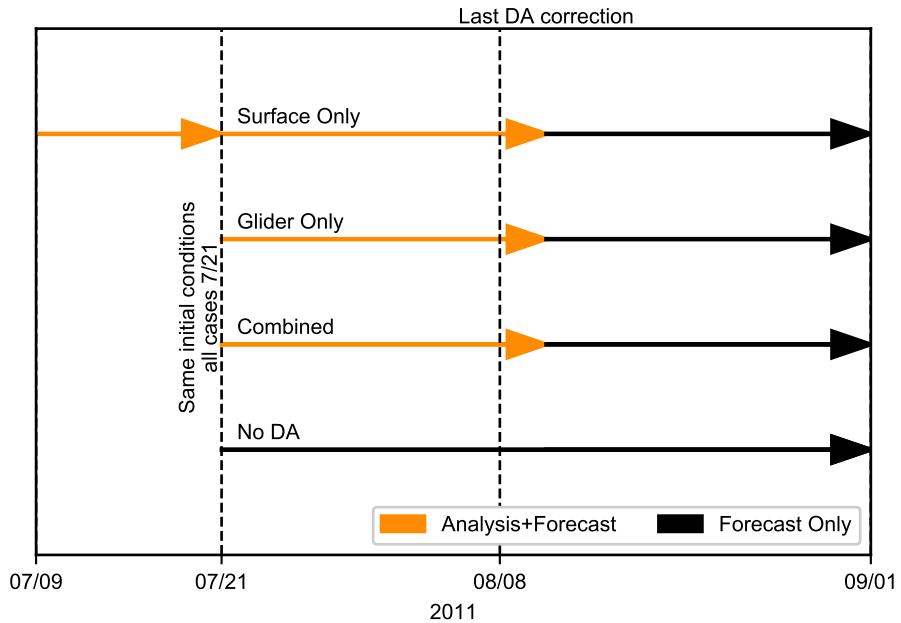


Figure 2.4: The experiment time line.

Glider Only, and Combined) takes place on 8 August 2011. After this, all the experiments are continued as forecasts without further assimilation until 1 September 2011. In spring 2011 the freshwater transport from the Columbia River was anomalously large, creating a large river plume. Upwelling favorable winds preceding the DA period (see Figure 2.2a) turned the plume toward Oregon and offshore (*Hickey et al.*, 2005; *Liu et al.*, 2009). Additional upwelling events take place during the DA period between 21-30 July and between 6-8, 11-14 and 16-22 August (see Figure 2.2a). The assimilation experiments are set purposely during this period, when the salinity signal sampled by the glider is particularly strong.

The model without assimilation reproduces the plume dynamics qualitatively correctly. For instance, the extent and geometry of the plume is similar to that revealed in satellite ocean color data (*Saldías et al.*, 2016). Events of surface water freshening over the Oregon shelf following upwelling wind relaxation and reversal are observed at the NH10/NDBC46094 mooring (124.308°W, 44.634°N, see *NDBC* (2016)). The

timing and magnitude of these events is predicted well by the long free model without DA (Figure 2.2b). A further comparison is provided in Figure 2.5. This figure shows the average  $T$  and  $S$  profiles from glider observations and from the long free run without DA for the period 1 May 2011 to 1 October 2011. Different profiles are compiled for times during which the plume was present at the surface (glider surface salinity  $< 31.5$ ) and during times it was not. In general, the model, even without DA, is able to reproduce the time-averaged temperature and salinity profiles in the top 150 m along the glider transect. On smaller time and space scales, i.e. the scales we are hoping to correct with DA, incorrect prediction of the location of the plume front by the model might still lead to significant errors in the glider salinity observations, as presence or absence of the plume can create large variations in surface salinity (see Figure 2.5a).

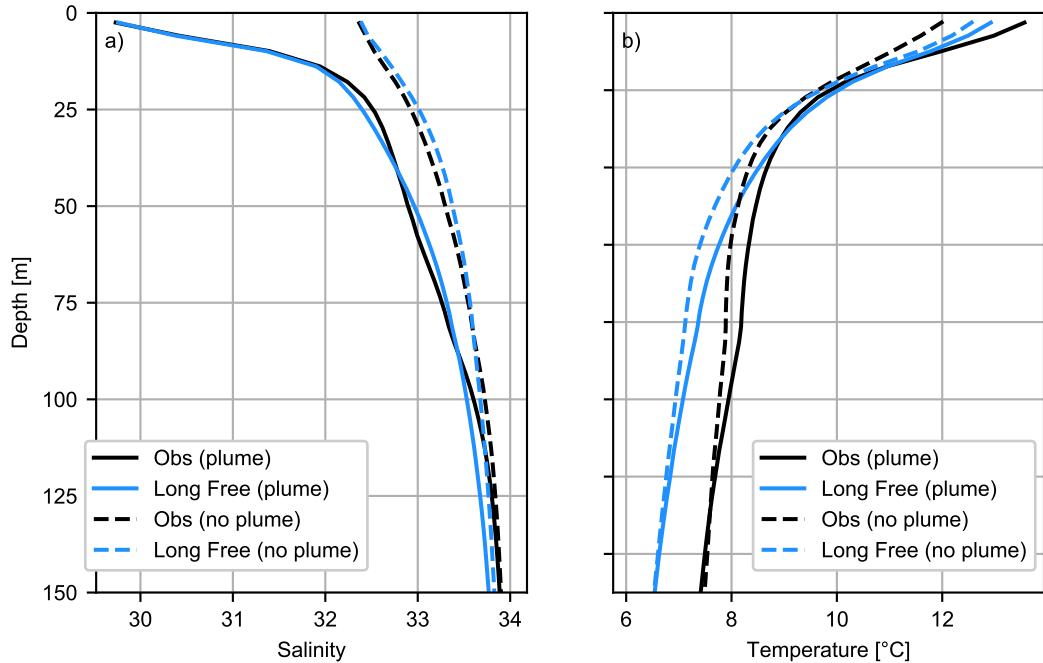


Figure 2.5: Time-averaged vertical profiles of (a) salinity and (b) temperature between 1 May and 1 October 2011 based on (black) glider observations and (blue) the long free model run. Solid (dashed) lines are the averages of the profiles when the Columbia River plume was present (absent) in the observations, i.e., when surface  $S < 31.5$  ( $> 31.5$ ).

## 2.4 Results

In this section we will first demonstrate that assimilation of glider observations alone using our DA system creates erroneous eddies and that DA case Combined prevents these eddies from forming while still fitting the glider data. Then we compare forecasts from the different experiments. Additional sensitivity tests are performed by assimilating subsets of the surface data to learn which surface data types contribute most to the improved performance. Finally, the subsurface differences between the DA cases Surface Only and Combined will be compared to determine if glider assimilation impacts the topology of subsurface isopycnal layers on a regional scale.

### 2.4.1 DA Impact on Surface Fields

The negative impact of glider only DA on the geostrophic eddy field can be seen in plots of the surface relative vorticity, which is calculated by taking the curl of the daily-averaged horizontal velocities in the top model layer. In Figure 2.6, it is shown together with the surface daily-averaged currents on 4 August 2011 for the four experiments. Cases No DA, Surface Only, and Combined show similar structures except that in the DA cases separation at Cape Blanco ( $43^{\circ}\text{N}$ ) is qualitatively more vigorous and farther towards the west than in the case No DA. Assimilation of glider observations alone (Figure 2.6c) in the four time windows preceding this analysis resulted in generation of a strong eddy field that surrounds the glider transect and which is absent in the other three cases. The eddies are persistent: after creation at the beginning of the window they move and deform under influence of the current, but they remain present during the forecast and can be tracked in the future assimilation windows.

Are the large erroneous eddies implied mostly by the correction in the initial conditions, or can they result from the growth of baroclinic instabilities of the coastal jet, caused by the (relatively small) initial DA perturbation? To answer this question, we use hourly DA analysis outputs and calculate enstrophy  $\frac{1}{2} \int |\nabla \times \mathbf{u}_s|^2 dA$  for the area most impacted by the eddies, between  $42^{\circ}\text{N}$ ,  $47^{\circ}\text{N}$ ,  $127^{\circ}\text{W}$  and the coast; here  $\mathbf{u}_s$  is the horizontal surface velocity (Figure 2.7). We find that in the three cases, No DA, Surface Only, and Combined, the area-integrated enstrophy is remarkably similar. In case Glider Only, eddies emerging from DA increase the enstrophy by a factor of three. The largest changes are associated with instantaneous DA corrections at the beginning of each DA window (vertical dashed lines in Figure 2.7). These changes are generally followed by a slow decrease of enstrophy toward the end of the 3-day window. We conclude that the enstrophy growth in the Glider Only case is mostly due to the instantaneous DA correction in the density field.

Figure 2.8 shows the daily-averaged SST obtained from the different cases (all analysis) and satellite SST observations available on 4 August 2011, the same day as that in Figure 2.6. The results for experiment Surface Only and Combined are

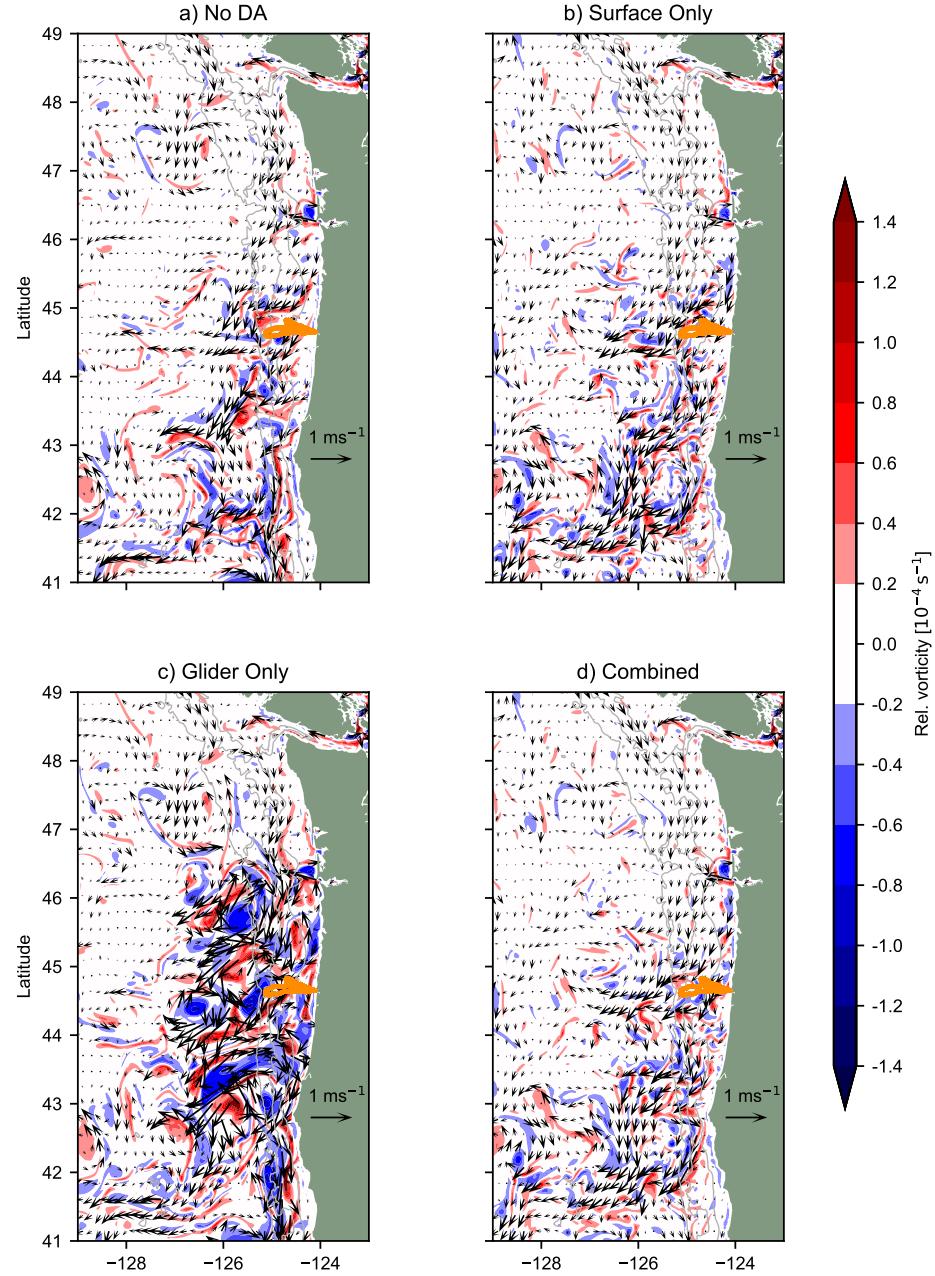


Figure 2.6: Daily-averaged surface velocity fields (arrows) and relative vorticity in the top model layer (color scale) on 4 August 2011 from (a) the experiment without data assimilation, (b) analysis from Surface Only, (c) analysis from Glider Only and (d) analysis from Combined. The location of the glider measurements up to 8 August 2011 24:00 is indicated by the orange lines. Grey lines mark the 200 m, 1000 m, 2000 m isobaths.

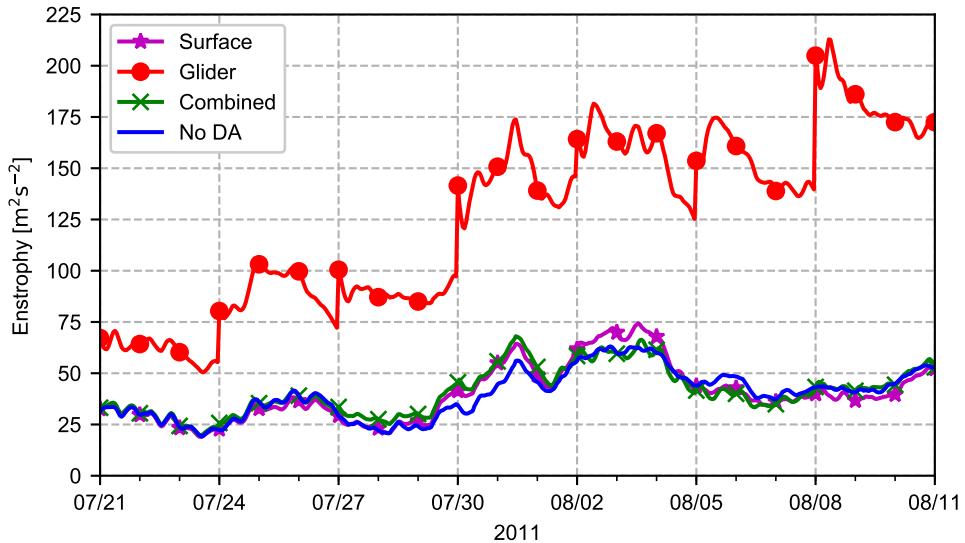


Figure 2.7: Enstrophy for the area between  $42^{\circ}\text{N}$ ,  $47^{\circ}\text{N}$ ,  $127^{\circ}\text{W}$  and the coast for the case No DA (blue) and the analysis from the cases Glider Only (red), Surface Only (purple) and Combined (green). The vertical grid lines mark the beginning of each DA window, when instantaneous correction is applied.

nearly identical. Once again Glider Only differs significantly. Eddies are clearly visible within the Columbia river plume (solid black contours), some with a high-temperature core ( $> 18^{\circ}\text{C}$ ). While in the cases Surface Only and Combined the upwelling zone is continuous between  $45^{\circ}\text{N}$  and the southern boundary and is widening toward the south (in agreement with the assimilated SST), in the case Glider Only the upwelling separation zone is split into two, one over the mid-Oregon shelf and another south of Cape Blanco ( $43^{\circ}\text{N}$ ). The eddies introduced by glider DA are not only unphysical, judged by their high enstrophy, but also damaging to the prediction of the front geometry and associated alongshore transport.

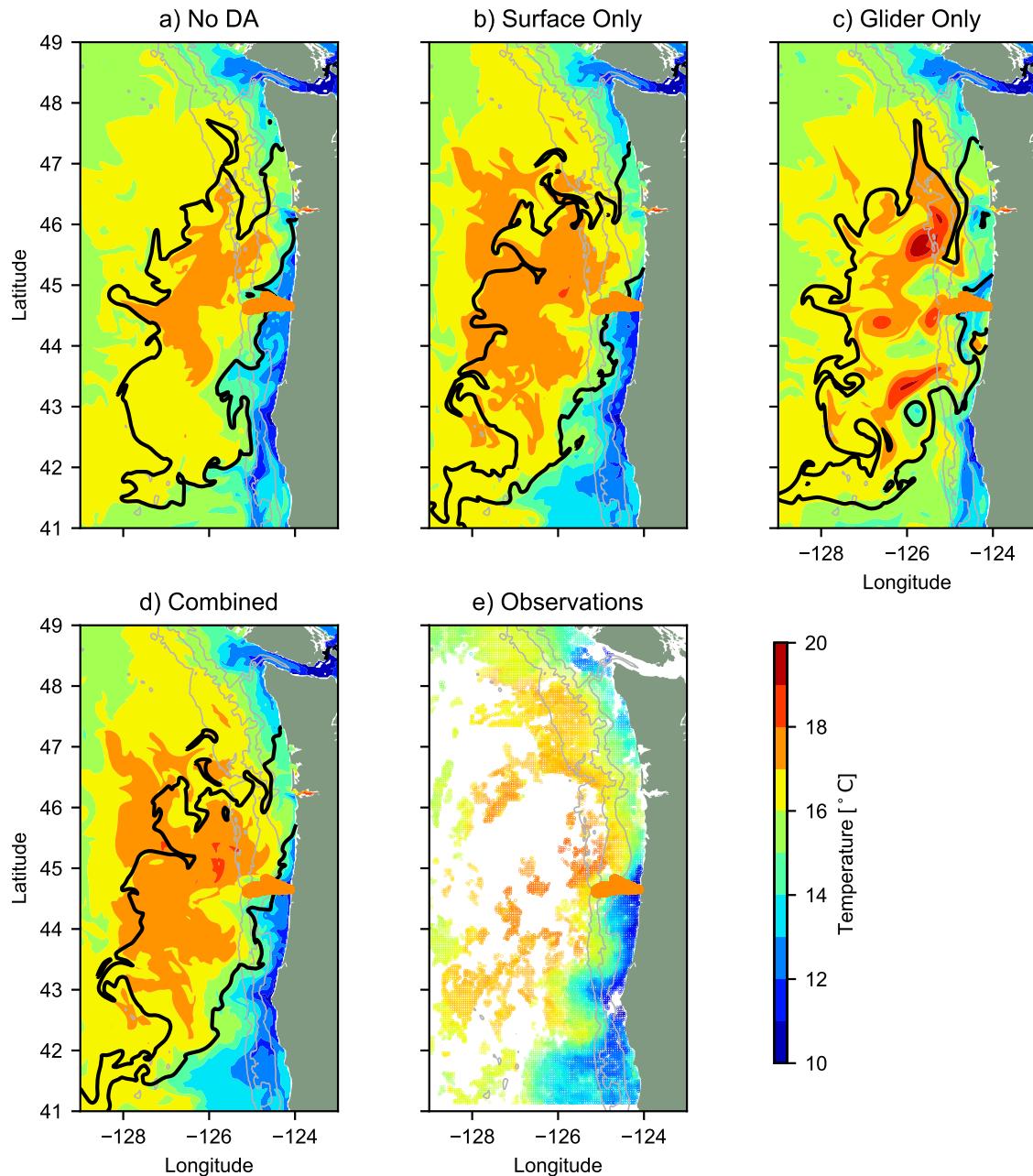


Figure 2.8: Daily-averaged surface temperature (color scale) and the 31.5 (solid line) salinity contour on 4 August 2011 from (a) case No DA, (b) analysis from Surface Only, (c) analysis from Glider Only, (d) analysis from Combined and (e) from satellite observations. The location of the glider measurements up to 8 August 2011 24:00 is indicated by the orange lines. Gray lines mark the 200 m, 1000 m, 2000 m isobaths.

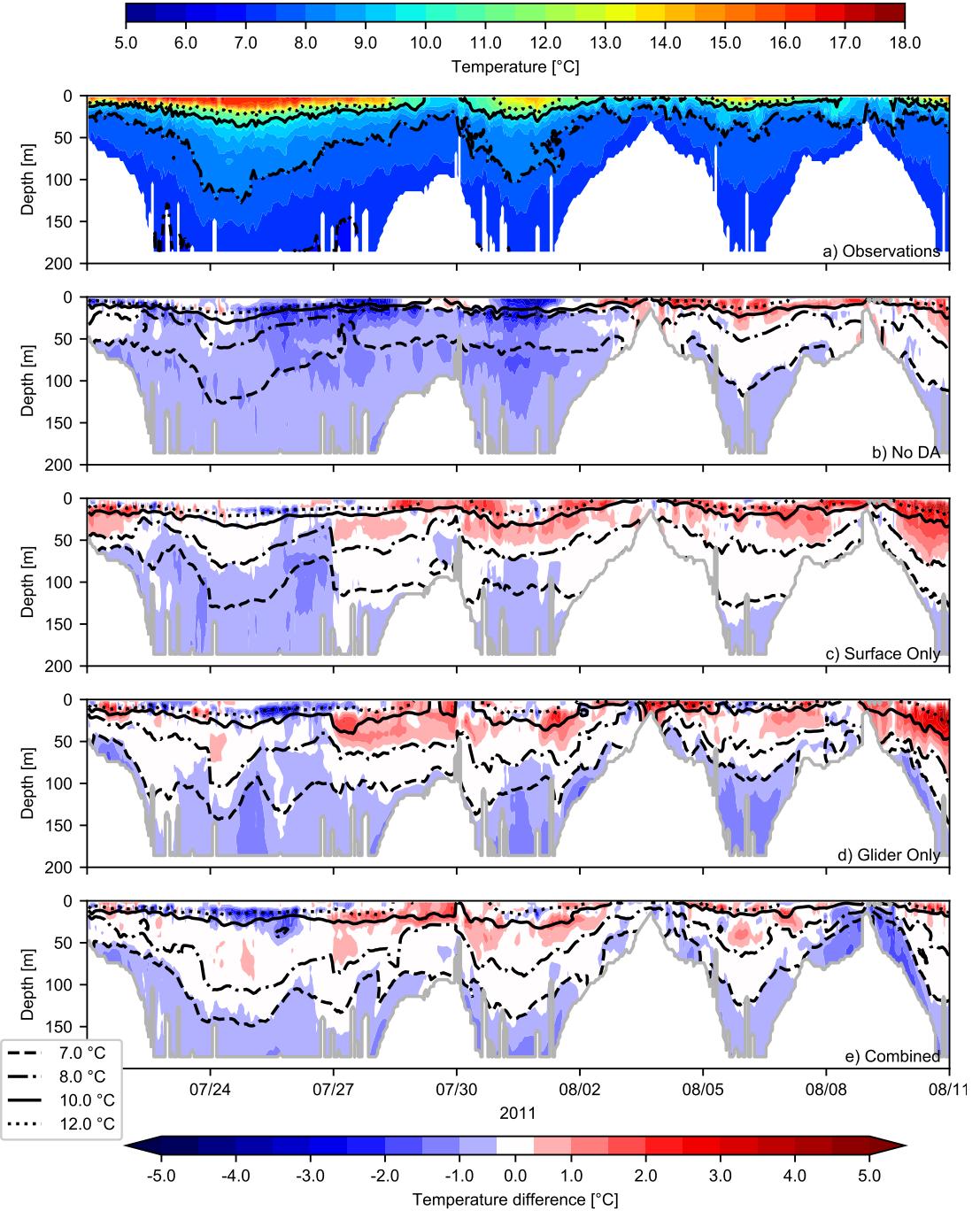


Figure 2.9: (a) Observed glider  $T$  as a function of time and depth; (b)-(e) the model-observation  $T$  difference, shown as color and selected model  $T$  contours, for cases (top to bottom) No DA, Surface, Glider Only, and Combined. DA analyses are used. Contours are  $T = 7, 8, 10, 12^{\circ}\text{C}$ .

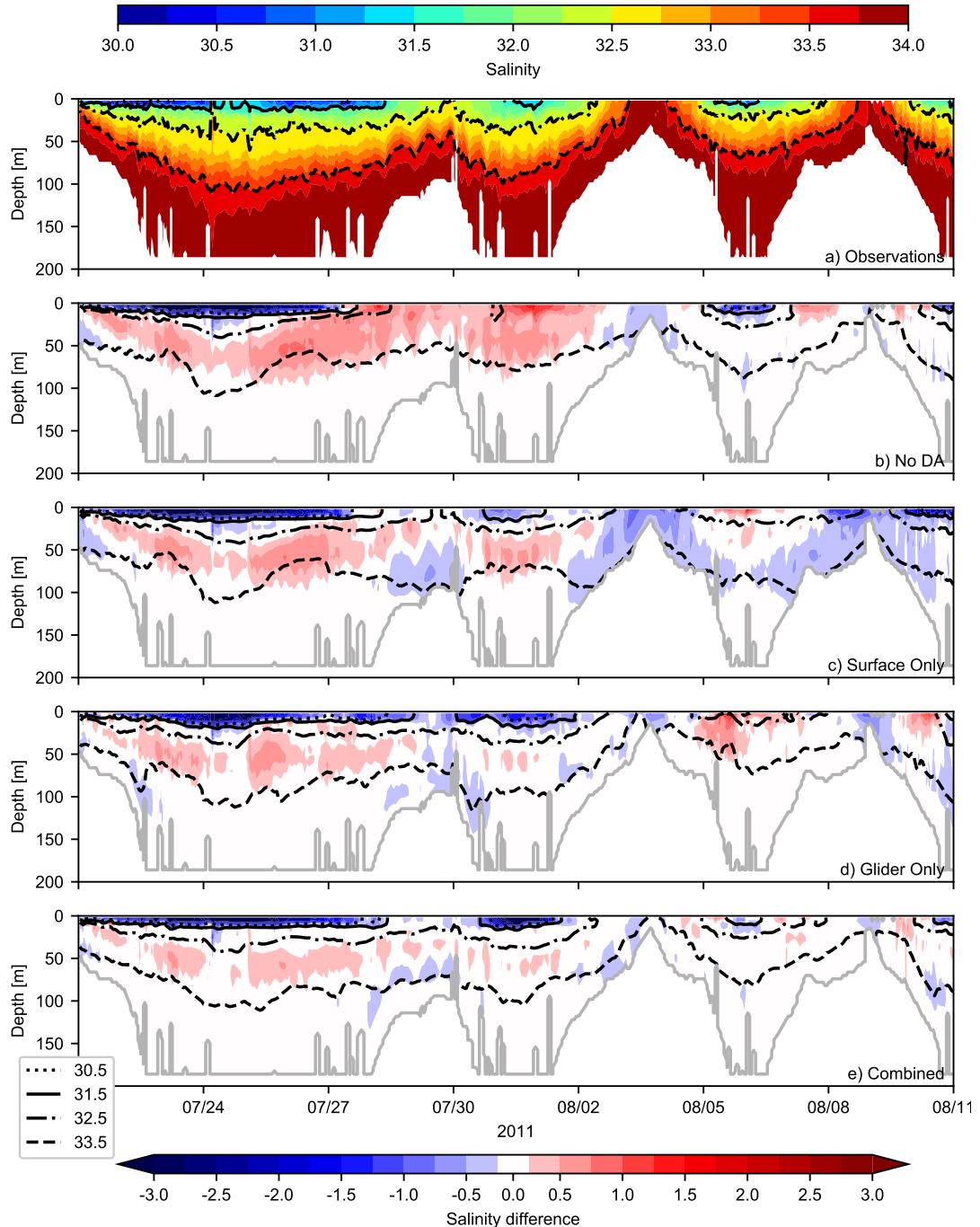


Figure 2.10: (a) Observed glider  $S$  as a function of time and depth; (b)-(e) the model-observation  $S$  difference, shown as color and selected model  $S$  contours, for cases (top to bottom) No DA, Surface, Glider Only, and Combined. DA analyses are used. Contours are  $S = 30.5, 31.5, 32.5, 33.5$ .

## 2.4.2 DA Impact on Subsurface Fields

Figure 2.9a shows glider temperature sections as a function of time and the vertical coordinate, for the period July 21 through August 11. Below it, in Figure 2.9b-e, the difference between the four base model cases and the glider temperature is shown as color. Selected temperature contours for each model case are also shown. For the DA cases (Figure 2.9c-e), analysis fields are sampled along the glider path. Figure 2.10 exhibits similar plots for salinity and model-data salinity differences.

As seen in the observations (Figure 2.9a and 2.10a), the Columbia River plume is characterized by a layer of relatively warmer and fresher ( $S < 31.5$ ) water near the surface. In the beginning of the series shown, before the start of the major upwelling event of 21-30 July, the glider finds the river plume waters over the shelf, close to coast. After the upwelling pushes the river plume waters offshore, the glider crosses through the river plume front on every lap between the offshore and shelf waters. Over the shelf, the glider samples through the colder and saltier upwelled waters. In particular, as seen in the data, the  $S = 33.5$  isohaline is found at depths near 100 m in offshore waters and it is outcropping at the ocean surface over the shelf on 3-4 August.

Every model case (Figure 2.9b-e and 2.10b-e) shows these patterns qualitatively correctly, although the extent of the river plume reaching the glider section line can differ from case to case. The model-data difference plot for case No DA shows the model negative bias throughout the water column in the first part of the time series (see Figure 2.9b). During the same time period, salinity bias is positive in the range of depths between 0 – 100 m (more precisely, between the 32.5 and 33.5 contours, Figure 2.10b). The DA Surface case (Figure 2.9c and 2.10c) reduces the subsurface bias in temperature between 100 m and the surface. Consistent with the prescribed standard deviation in the model temperature error covariance (see section 2.3.5), the SST data can influence the subsurface temperature down to these depths. At the same time, the positive bias in subsurface salinity is not removed. In the latter part of the time series surface data assimilation yields increased positive temperature bias in the shallow part of the water column ( $z > -50$  m). Since the DA change in salinity is not

directly constrained by the salinity data and is mostly influenced by the choice of the balanced model error covariance, in which the temperature correction is negatively correlated with the salinity correction, warming near the surface can be associated with a pattern of freshening of the water column for  $S < 33.5$  (see Figure 2.10c). The cases involving the glider data assimilation (i.e., Glider Only, Figure 2.9d and 2.10d, and Combined, Figure 2.9e and 2.10e) show that the DA system is able to fit both the temperature and salinity data and remove major biases in analyses compared to the No DA case. Qualitatively, looking at the finer details, one could favor case Combined.

Table 2.1: RMSE for different types of observations and experiments. Top block: Depth-averaged subsurface (only including observations below 25 m depth) analysis RMSE for the period 21 July to 10 August 2011 (see Figure 2.11a-c). Center block: As top block, but now for the forecast RMSE (see Figure 2.11d-f). Bottom block: Forecast RMSE error over the period 8 August to 1 September (see Figure 2.14). In each row the case with the lowest RMSE is emphasized.

	Surface only	Glider only	Combined	No DA
Analysis glider $S$	0.27	0.21	<b>0.17</b>	0.25
Analysis glider $T$ [°C]	0.65	0.76	<b>0.64</b>	0.84
Analysis glider $\rho$ [kg m <sup>-3</sup> ]	0.24	0.18	<b>0.13</b>	0.27
Forecast glider $S$	0.29	0.25	<b>0.21</b>	0.25
Forecast glider $T$ [°C]	<b>0.74</b>	0.83	<b>0.74</b>	0.84
Forecast glider $\rho$ [kg m <sup>-3</sup> ]	0.27	0.25	<b>0.20</b>	0.27
Extended SST [°C]	1.09	1.77	<b>1.05</b>	1.19
Extended HFR [m s <sup>-1</sup> ]	<b>0.14</b>	0.27	0.16	0.17
Extended glider $T$ [°C]	1.78	2.93	<b>1.20</b>	1.39
Extended glider $S$	0.61	0.62	<b>0.33</b>	0.59

Quantitative assessment is provided by calculating profiles of the model-data root-mean-square difference (or “error”, RMSE). These are obtained by time-averaging, at each horizontal level, squares of differences as shown in Figure 2.9 and 2.10. Additionally, the RMSE for the potential density is calculated, all using the hourly model outputs. The results are presented separately for the DA analyses (Figure 2.11abc)

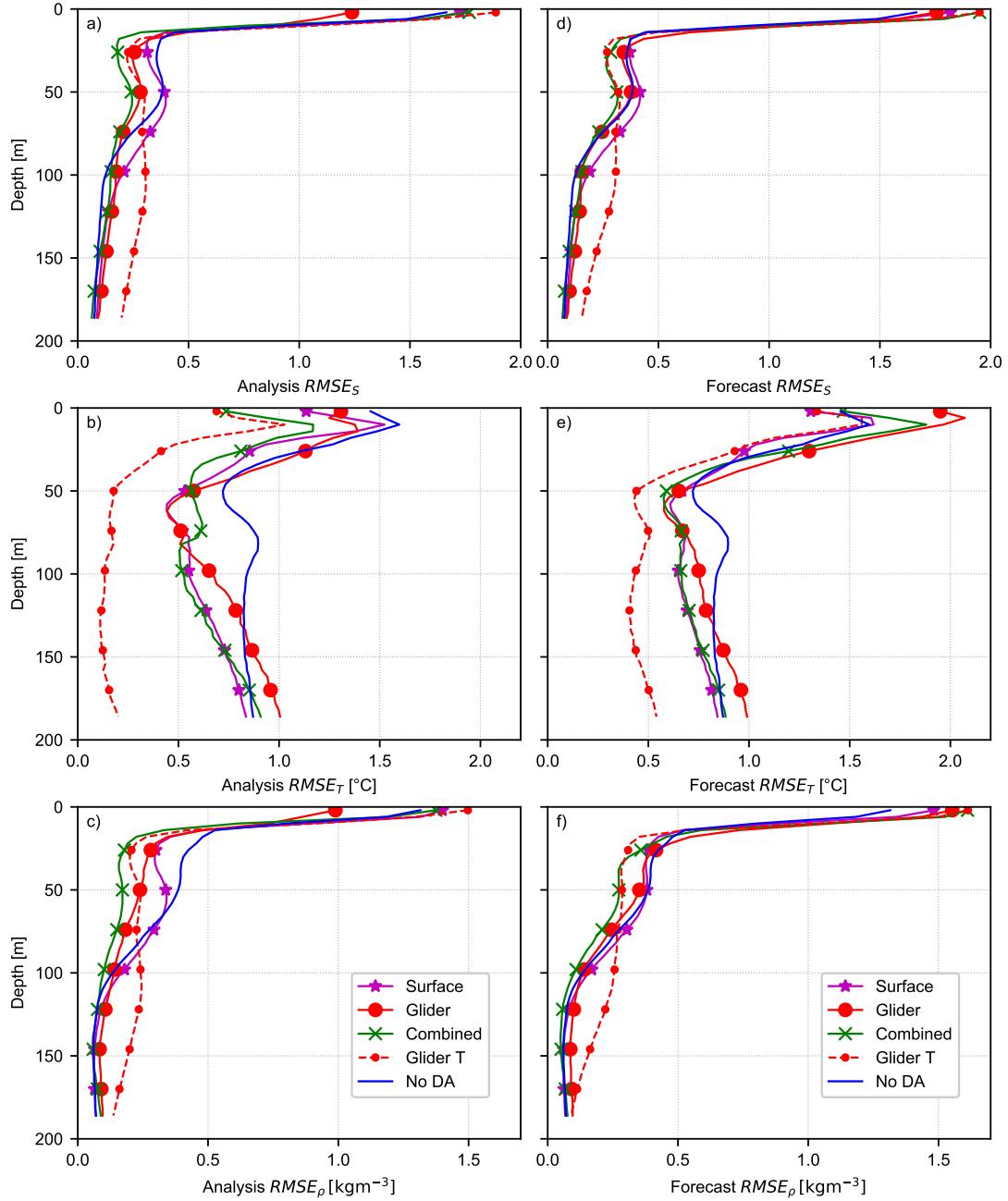


Figure 2.11: RMSEs from analyses (a-c) and forecasts (d-f) averaged over the period 21 July 2011 to 11 August 2011 compared to glider salinity observations (first row), glider temperature observations (second row) and potential density (third row) as functions of depth.

and day 1-3 forecasts (Figure 2.11def). Case Surface Only produces analyses with smaller RMSE in  $T$  than case No DA throughout the water column, but with slightly larger RMSEs in  $S$  and  $\rho$ . Case Glider Only (red) yields improvement over case No DA throughout the water column in terms of  $T$ ,  $S$ , and density. Case Combined has a smaller RMSE than case Glider Only in most of the water column. The notable exception is at the surface, where Case Glider Only yields the best improvement in terms of salinity (see Figure 2.11a) and density (see Figure 2.11c), but not in terms of temperature (see Figure 2.11b). In other words, using the surface and glider data in combination allows for a better fit to the glider data than in case Glider Only. Compared to case Surface Only case Combined shows improved fit to glider  $T$ ,  $S$  and  $\rho$  in the top 100 m of the analysis. This shows that the ratio of the observational error covariance of the glider  $T$  observations (see Figure 2.3) versus the observational error covariance for the SST observations ( $(0.39^{\circ}\text{C})^2$ ) is small enough that, notwithstanding that there are more SST than glider  $T$  observations, glider  $T$  observations can still contribute to the DA correction of the temperature in the vicinity of the glider transect.

For the short-term forecasts (see Figure 2.11def), the spread between the curves is smaller than for the analyses. In particular, the improvement in  $T$  in case Combined versus case Surface Only does not persist above 25 m depth. The result is that, at least in the vertical profiles, case Combined acts like a compromise between Glider Only and Surface Only with the glider observations providing improved forecast below the surface layer and the surface observations partially suppressing the deterioration of the forecast for the surface layer caused by the assimilation of the glider observations.

A short summary of the results from Figure 2.11 can be found in Table 2.1, which shows the time- and depth-averaged RMSE below the surface layer for analyses and forecasts over the same period. Here, each row presents the estimate for a particular field ( $T$ ,  $S$ , or  $\rho$ ), either the analysis or the forecast. Each column corresponds to one of the base model cases. The smallest RMSE value in each row is shown as bold. Case Combined is a winner for the region below the surface layer.

A possible explanation for the fact that RMSEs for surface salinity and den-

sity are larger in case Combined compared to the analysis of case Glider Only (see Figure 2.11a,c) might lie in the structure of the background covariance error used. Figure 2.9b and Figure 2.10b show that the plume is too fresh and too cold during the period 21-27 July. We hypothesize that DA cannot correct for the observed plume temperature and salinity errors simultaneously, as in our **B** (see section 2.3.5) temperature and salinity background errors are assumed to have a negative correlation. In case Glider Only, DA partially reduces the salinity error at the expense of a larger temperature error, while for case Combined, in which the more numerous SST observations are present, the largest reduction of  $J$  (see (2.2)) is achieved by fitting the SST. To test this hypothesis, a DA experiment for the period 21 July to 11 August is performed in which glider  $T$  is only assimilated (case Glider T). The resulting RMSEs in the analysis of case Glider T show that lower temperature RMSE can be obtained by not assimilating salinity observations (Figure 2.11b) and hence that it is the salinity that constrains the fit to the glider  $T$  observations. However, the reduction in  $T$  RMSE comes at the expense of an increase in glider  $S$  RMSE (Figure 2.11a) proving that the static **B** used is not always able to correctly represent the relationship between  $T$  and  $S$  errors in the model.

In Glider Only DA analyses, the subsurface  $T$  and  $S$  errors are reduced along the glider path. However, when we extend the cross-shore transect to the west of the glider range, problems emerge in case Glider Only. To illustrate, the potential density is calculated in a cross-section at 46.64°N using daily-averaged DA analysis  $T$  and  $S$ . The example of the density field on 4 August 2011 is shown in Figure 2.12. Also shown in the figure are the 31.5 isohaline (solid black line) and the  $26.5 \text{ kg m}^{-3}$  isopycnal contour (dashed black). The former is important because it marks the edge of the Columbia River water plume. The latter is important because it marks the upper boundary of the cold, oxygen-poor, high-nutrient source water (Adams *et al.*, 2013). Its arrival on the shelf contributes to the frequent hypoxic conditions observed on the Oregon shelf during the summer (Connolly *et al.*, 2010; Siedlecki *et al.*, 2015). Figure 2.12a,b exhibits two noticeable differences between the cases Surface Only and Glider Only. First, glider DA affects the amount of source water

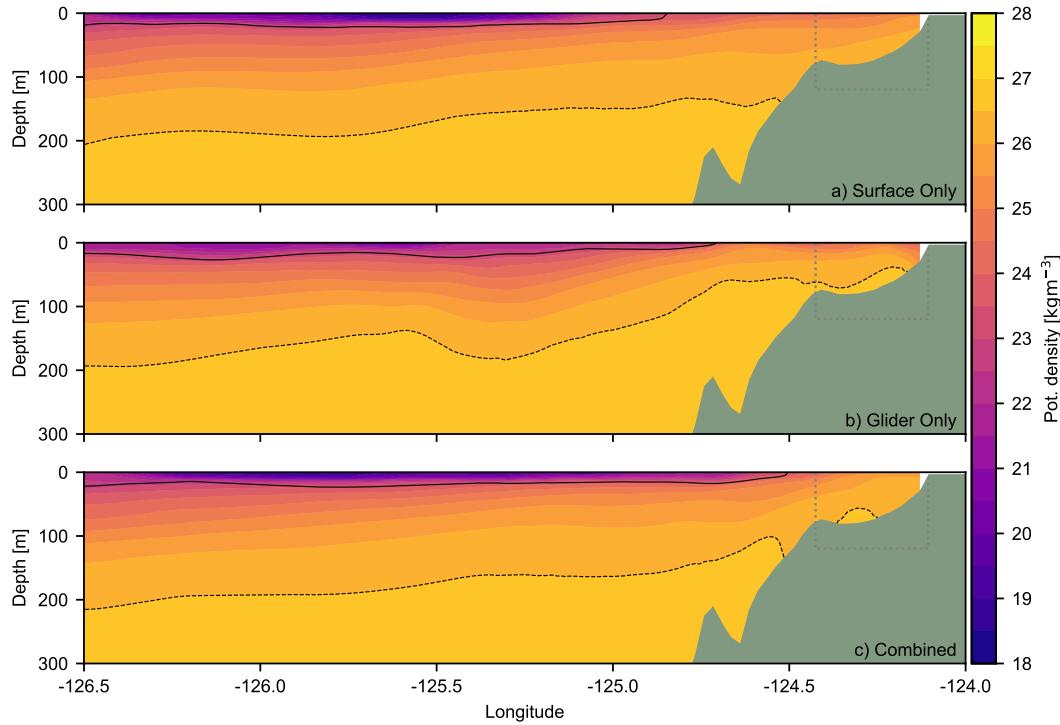


Figure 2.12: Daily-averaged density along the line  $44.64^{\circ}\text{N}$  from the analyses on 4 August 2011 for (a) experiment Surface Only, (b) experiment Glider Only and (c) experiment Combined. The dashed box indicates the area in which glider observations are available during this window. The dashed black line marks the  $26.5 \text{ kg m}^{-3}$  isopycnal.

(potential density  $> 26.5 \text{ kg m}^{-3}$ ) predicted over the shelf. Second, west of the glider transect, at  $125.3^{\circ}\text{W}$ , the assimilation of glider data alone leaves a sizable impact on the subsurface structure. The 40 m depression of the  $26.5 \text{ kg m}^{-3}$  isopycnal at  $125.3^{\circ}\text{W}$  is dynamically consistent with the erroneous anticyclonic eddy at this location in Figure 2.6c. Case Combined (Figure 2.12c) predicts denser waters over the shelf compared to case Surface Only and at the same time defies strong, possibly erroneous, corrugations of the  $26.5 \text{ kg m}^{-3}$  density surface offshore of the shelf break.

To provide a more quantitative assessment of the DA impact on the depth of

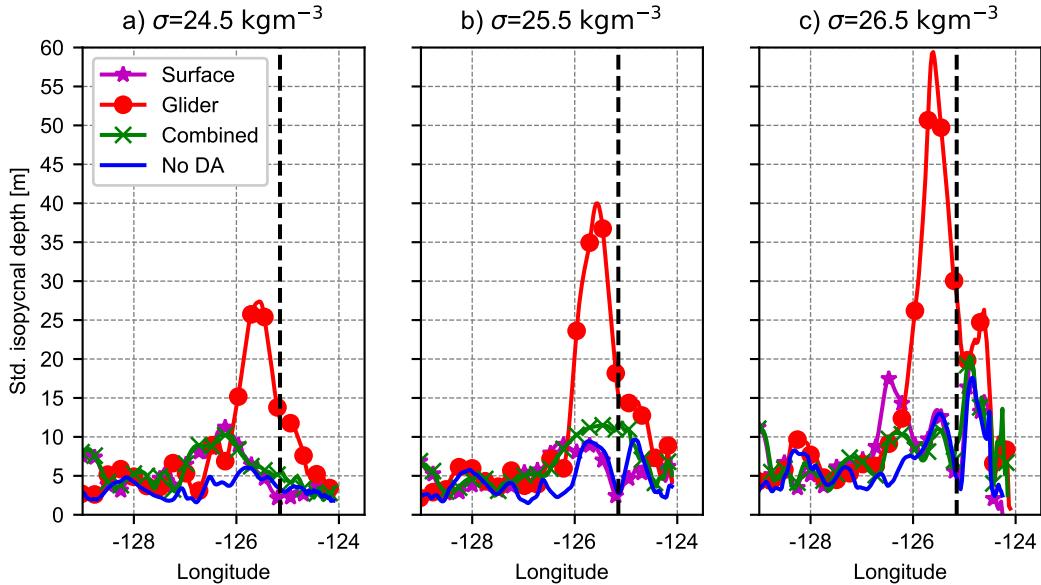


Figure 2.13: Standard deviation in the depths of the  $24.5 \text{ kg m}^{-3}$  (a),  $25.5 \text{ kg m}^{-3}$  (b) and  $26.5 \text{ kg m}^{-3}$  (c) isopycnal from the hourly output of the analyses between 21 July to 11 August along  $44.64^\circ\text{N}$ , shown as functions of longitude in the different experiments. The dashed black line marks the most westward location of the glider.

the subsurface isopycnal surfaces in this section, the depth of the 24.5, 25.5, and  $26.5 \text{ kg m}^{-3}$  surfaces are calculated from the hourly snapshots over the entire period for which DA analyses are available (21 July to 11 August). The DA corrections do occasionally create an unstable profile in the surface layer in the first few hours after the correction. This makes the definition of the isopycnal depth ambiguous, in particular, for the shallower isopycnal surfaces. To overcome this problem the model profiles are mixed into stable profiles during post-processing before the isopycnal depth is determined (see appendix A). For each longitude along this zonal section the standard deviation of the hourly depths is determined (Figure 2.13). The standard deviation is higher for the cases Surface Only and Combined than for No DA (on average +28% and +17%, respectively, for the  $26.5 \text{ kg m}^{-3}$  isopycnal surface). 5–10 m deviations can be partly associated with the internal tides resolved by the 2-km model (Kurapov *et al.*, 2003; Osborne *et al.*, 2011). Assimilation of glider observations alone

causes much greater variability along the glider transect (on average +73% increase in the standard deviation for  $26.5 \text{ kg m}^{-3}$ ) . The strongest variability is found just west of the glider transect where standard deviations can be 5-7 times as large as in the other experiments. Variability drops to levels comparable to other experiments at the distance of 80-100 km west of the glider transect, which is 3-4 times the horizontal correlation scale in the background error covariance (see section 2.3.5).

### 2.4.3 Post DA Persistence

After the last DA correction on August 8, the No DA and the three DA cases are continued as forecasts until 1 September 2011. The day 1-3 forecasts (21 July 21 to 11 August) and extended forecasts (12 August to 1 September) are compared to the HFR velocity and SST as well as glider  $T$  and  $S$  observations. The RMSE for each observation type and each day is calculated using only observations close enough to the coast and the glider section, specifically, in the shaded area in Figure 2.1. The area-averaged RMSE time series are presented in Figure 2.14. Forecast quality in the case Glider Only deteriorates quickly when compared against HFR and SST data (Figure 2.14a and b). On average over the entire period, case Combined SST RMSE is larger than the case Surface SST RMSE by only 1%; case Combined HFR RMSE is larger by only 9%. Based on these we conclude that the effect of adding the glider data to these surface fields is limited. Compared to glider  $T$  (Figure 2.14c), the accuracy of short-term forecasts for the Glider Only case (21 July to 8 August) is acceptable and comparable to the that from the other three cases (with RMSEs near  $1^\circ\text{C}$ ). However, due to DA-induced eddy transport, long-term forecasts from this case are of poor quality with RMSEs substantially higher than in the other cases (bottom block Table 2.1). The long-range forecast RMSE computed against the glider  $S$  data (Figure 2.14d), is large and comparable to cases No DA and Surface Only (see Table 2.1). The salinity RMSE for the long-range forecast following the Combined DA series of corrections is a major improvement over the other three cases. Similarly, the  $T$  RMSE over this period is lower than in case Combined. These results indicate that case Combined holds an edge as both short and longer range predictor.

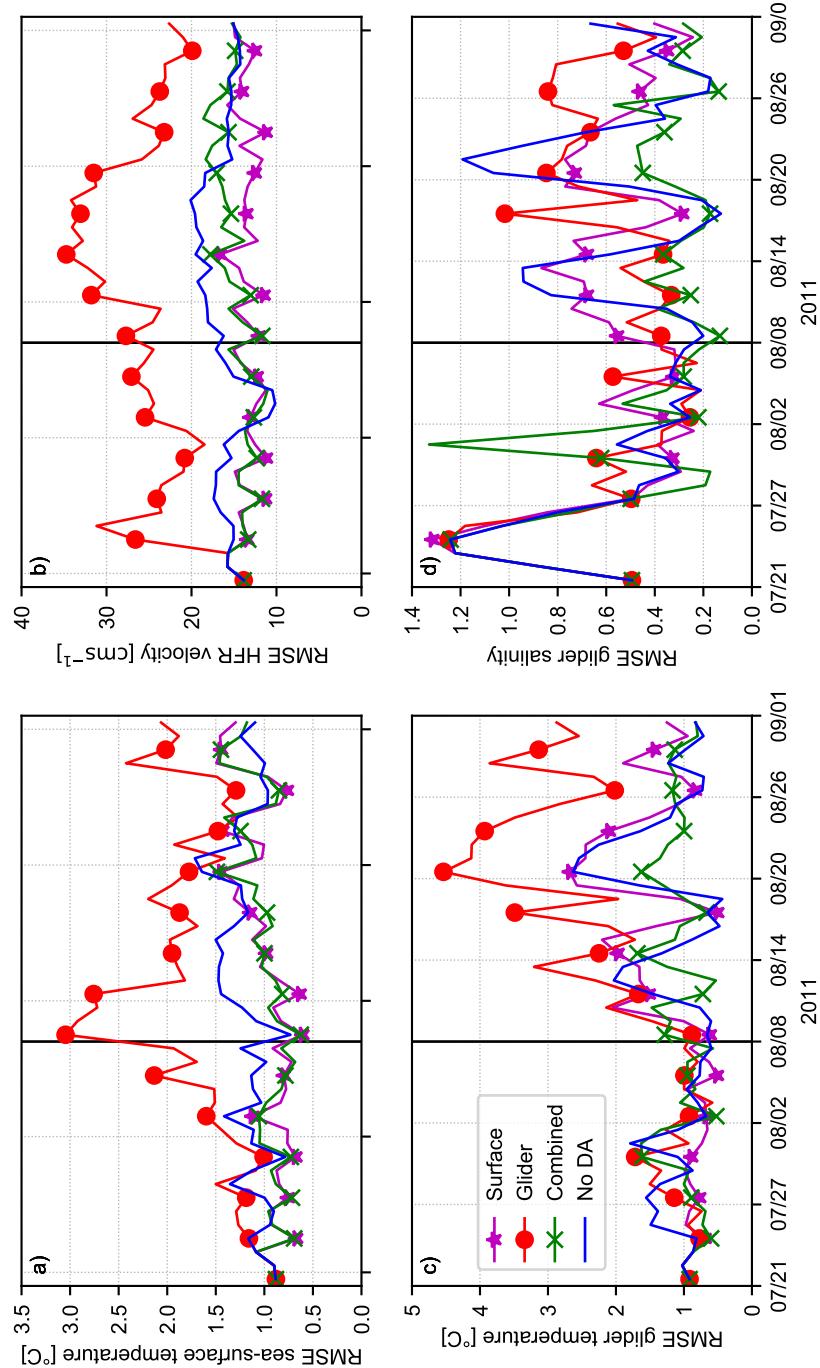


Figure 2.14: Time series of the area-averaged, daily-averaged forecast RMSE as function of time for (a) SST, (b) HFR velocity, (c) glider temperature and (d) glider salinity for the cases Surface Only (purple), Glider Only (red), Combined (green) and No DA (blue). The vertical lines mark the instant of the last DA correction (8 August 2011).

#### 2.4.4 Contributions from Different Observation Types

Here we answer two questions: (1) what type of the subsurface observation,  $T$  or  $S$ , is the main contributor to the deterioration of the DA results and (2) which type of surface observations provides is most strongly constraining the eddy variability over the shelf and slope. Three additional DA cases are run using subsets of the assimilation sets: Glider  $T$ , Glider+SST and Glider+HFR. In the last two cases, both glider  $T$  and  $S$  plus one of the surface sets are assimilated. An experiment that assimilates only glider and SSH observations is not executed as the number of SSH observations on the shelf is small and therefore it is deemed unlikely that the SSH observations are a major contributor to the improvements observed in the Combined case. The daily-averaged analysis RMSEs are computed for all the cases with respect to the assimilated SST, HFR surface velocity, glider  $T$  and  $S$ . The ratio between these RMSE and their equivalents in the case Combined is shown in Figure 2.15. Only observations located in the gray shelf area shown in Figure 2.1 are included in the calculation. The curve for the Glider Only case is added for comparison.

Figure 2.14c and 2.14d show that Case Glider  $T$  provides the best fit to glider  $T$  observations in the analysis, but at the expense of a decreasing fit to the salinity, as was already inferred from Figure 2.11b. Erroneous eddies are still found in case Glider  $T$ , but they are weaker than in case Glider Only and not as numerous (not shown). Consequently, Glider  $T$  provides better fit to HFR and SST observations than case Glider Only. Case Glider+SST (green line in Figure 2.15b) has the SST RMSE close to that in case Combined, while case Glider+HFR (black line in Figure 2.15b) has a velocity RMSE close to that in case Combined. These two facts suggest that corrections induced by the SST and HFR observations do not conflict with each other. SST observations contribute the most to constraining eddy variability: HFR RMSE increases only by 36% if only glider and SST are assimilated (green line in Figure 2.15b), while SST RSME increases by a factor of 3 in the case Glider+HFR (black line in Figure 2.15a). Part of this difference will stem from the fact that in case Glider+SST the number of observations assimilated is an order of magnitude larger than in Glider+HFR.

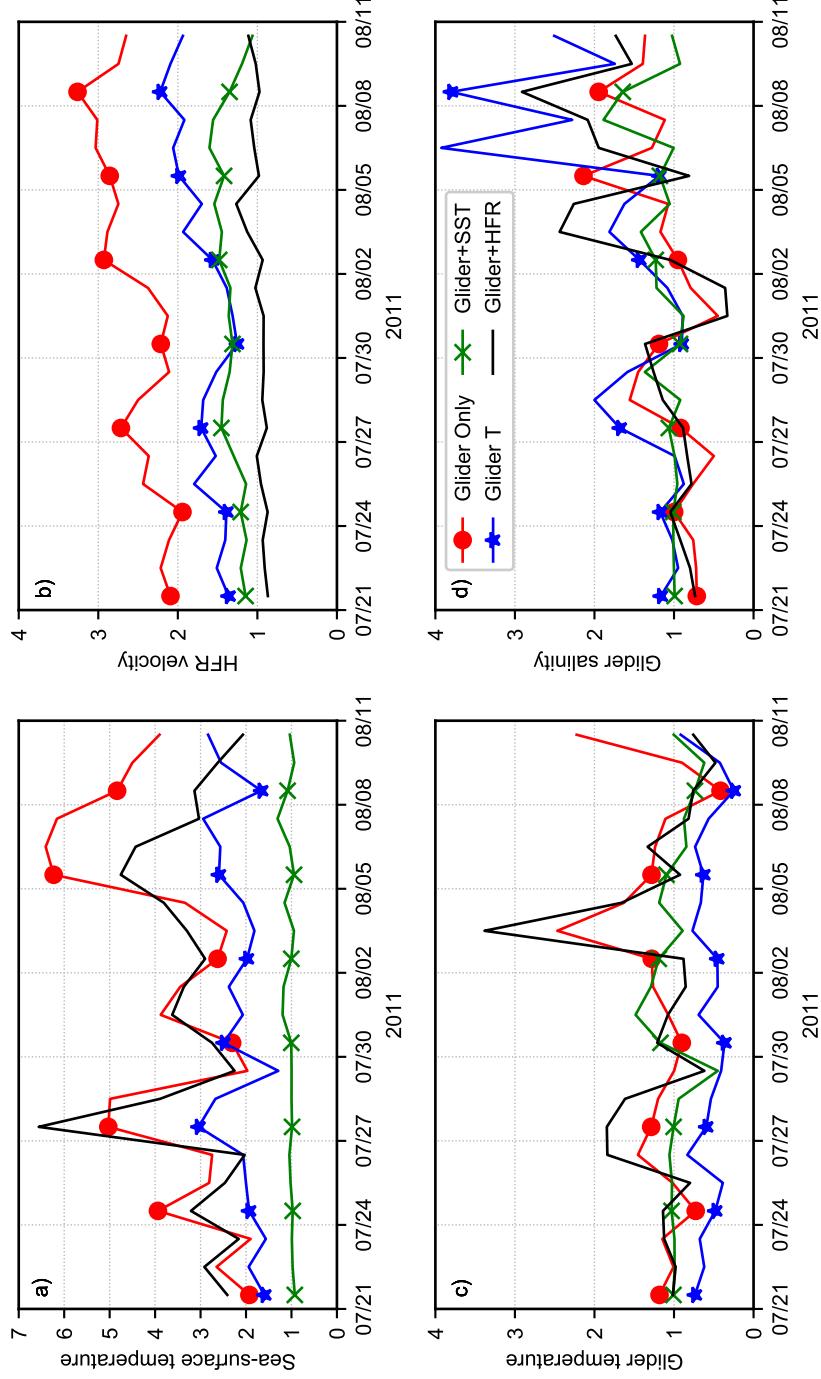


Figure 2.15: The time series of the daily-averaged, area-averaged RMSE in (a) SST, (b) HFR velocity, (c) glider temperature, (d) glider salinity observations for the DA cases, in which the subsets of the available data sets are assimilated. The values are shown relative to the RMSE for case Combined, i.e., values below (above) one mean that the case is better (worse) than case Combined. Shown are DA analyses where only glider temperature and salinity (red), only glider temperature (blue), glider temperature, salinity and SST (green) and glider temperature, salinity and HFR (black) are assimilated.

#### 2.4.5 DA Impact on the Topology of the 26.5 Isopycnal Surface

Using the daily-averaged temperature and salinity fields on the first day in the first data assimilation window (21 July 2011) and the last day in the last data assimilation window (10 August 2011), the depth of the  $26.5 \text{ kg m}^{-3}$  isopycnal surface,  $z_{26.5}$ , is determined. The difference in depth of this isopycnal surface between each of the three base DA experiments and case No DA is shown in Figure 2.16.

In the first window, in case Glider Only (Figure 2.16a) the impact of the glider observations on  $z_{26.5}$  is largest in the area near the glider and southward along the shelf slope where information can be propagated backward in time by the 4DVAR algorithm with coastally trapped waves (*Kurapov et al.*, 1999, 2002, 2011). The impact in the case Combined (Figure 2.16b) is similar to Surface Only (Figure 2.16c) and differences in depth are more than a factor two as small as those in case Glider Only. Far away from the glider transect, corrections to  $z_{26.5}$  are similar in cases Combined and Surface Only.

In the last assimilation window (Figure 2.16d-f), the differences in  $z_{26.5}$  between the assimilation cases and No DA are a result of the accumulated DA changes in a series of windows and their dynamical evolution over 21 days. Changes in case Glider Only reach further north and south than in Figure 2.16a as an assemblage of eddies. Structures in cases Surface Only and Combined are very similar. The large-scale uplifting of  $z_{26.5}$  south of Cape Blanco ( $43^\circ\text{N}$ ) is associated with the correction of the location, strength, and direction of the separated coastal jet (cf. Figure 2.6b,d). The glider impact in case Combined can be seen southwest of the glider transect ( $44 - 44.6^\circ\text{N}$ ) where the isopycnal surface is uplifted compared to case No DA. At the beginning of our study we hypothesized that Glider Only assimilation will have a very local effect on the depth of the isopycnal surfaces and create eddies. We also hypothesized that combined assimilation of the in-situ and surface data would inhibit this eddy activity. This can happen since SST (and HFR data as well) carry information about the eddy field. E.g., cyclonic (anticyclonic) eddies are cold (warm) cored. If SST and HF radar do not indicate an eddy at the location near the glider and if the measured stratification is significantly different from the forecast then both

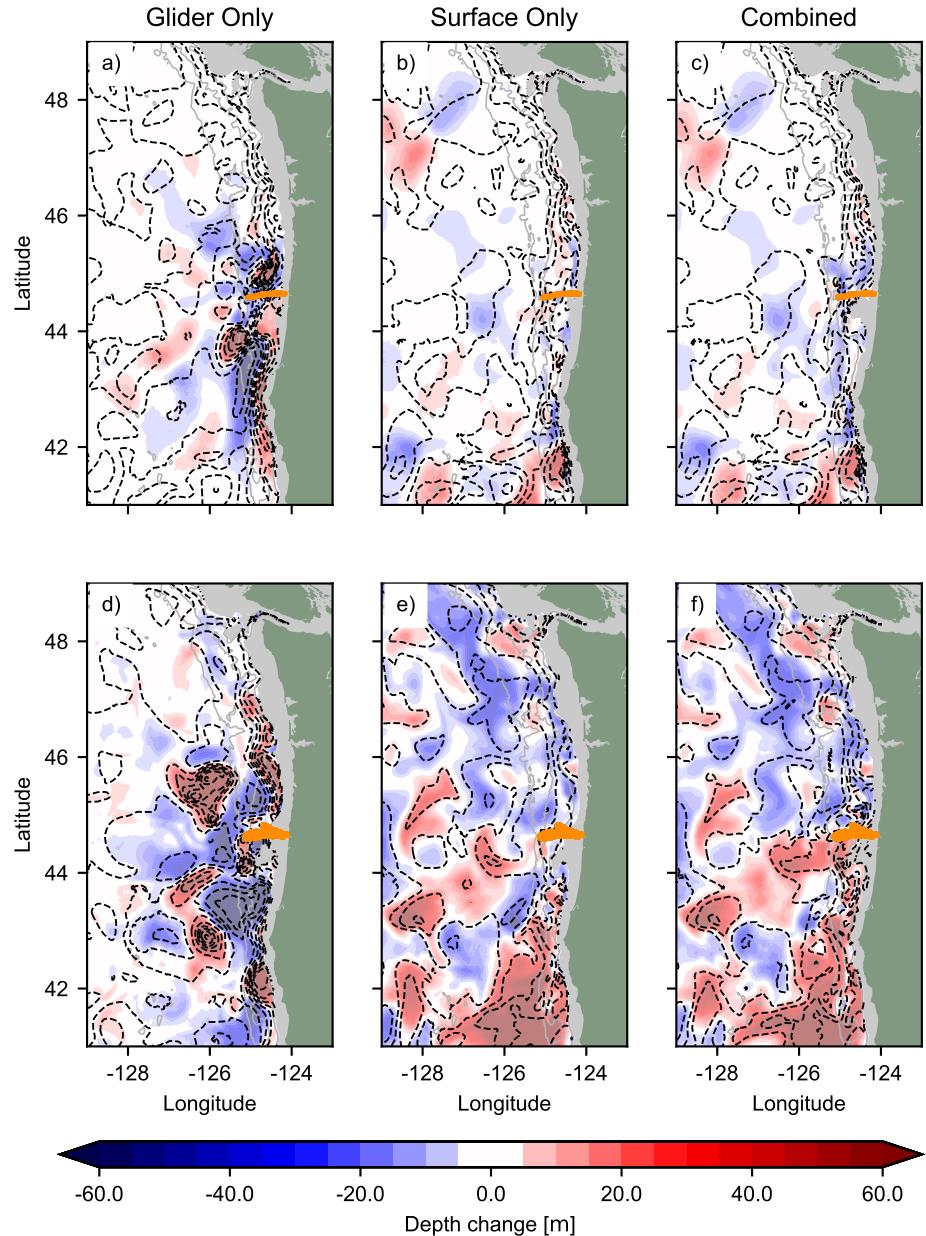


Figure 2.16:  $z_{26.5} \text{ kg m}^{-3}$  isopycnal on 21 July 2011 (top) and 10 August 2011 (bottom) from analysis for Glider Only (left), Surface Only (center) and Combined (right). The depth contours, every 25 m, are shown as dashed lines. The change in depth with respect to No DA is shown as color. Orange dots mark the location of the glider during up to and including the assimilation window.

the surface and subsurface data can be fit by adjusting the isopycnal surfaces at distances larger than the eddy scale. In our study we tried to demonstrate this, but could not find a clear illustration, partly since in our prior model the stratification is already close to observed (see Figure 2.5).

## 2.5 Conclusions and Discussion

This study shows it can be dangerous to assimilate glider vertical profile data in a high-resolution coastal ocean model if these data are not supported by surface observations. Without the surface observations, strong eddy variability can emerge. In a high-resolution DA system, gliders will appreciate the good company of the surface data because surface maps (mostly satellite SST in our case) constrain the eddy variability and help fit both the surface and subsurface data. Combined assimilation also helps to improve the subsurface in the analyses and forecasts. However, compared to case Surface Only the improvement mainly takes place in the salinity field, while RMSEs in temperature are similar to those in Surface Only and forecast accuracy for the surface is less than in case Surface Only. The improvement in long-term subsurface forecasts can last up to 3 weeks. The enstrophy (see Figure 2.7) can be used as a criterion for the monitoring of bogus eddy activity especially since this parameter remained very similar between cases No DA, Surface Only, and Combined, while it stood out as too large in the case Glider Only.

We find a large difference in the fit to the near-surface salinity data between the case in which unphysical eddies are formed (Glider Only) and the cases in which the erroneous eddies are not exhibited (Surface Only, Combined) or less pronounced (Glider T). This suggests that the formation of the erroneous eddy variability can in part be attributed to deficiencies in the background model  $T$  and  $S$  and a limited ability of the system to fit these data together. This can be exaggerated by the use of the static, balanced background error covariance, in which a simple negative correlation between errors in  $T$  and  $S$  is assumed, while  $T$  and  $S$  errors in the area of the river plume were sometimes positively correlated. It would be interesting to see (in future studies) if a different choice of the error covariance, e.g., an ensemble-based

covariance (*Pasmans and Kurapov*, 2017), helps to improve the fit and reduce the forecast error.

While our series of experiments concerns the glider data, we anticipate that assimilation of other in-situ assets, e.g., Argo profiles, can cause similar problems. We note that a useful signal in a given observed density profile can either say something about large (basin) scale model errors in the stratification or errors at the scale of geostrophic eddies. *Li et al.* (2015a,b) proposed to explicitly separate the large- and small-scale signal by spatial filtering into two subsets and fit those using penalty functions with substantially different horizontal correlation scales. We had hoped that combined surface and in-situ assimilation would help the system distinguish between these two scales and that DA would create subsurface corrections that reflect both the long scales and short scales. However, Figure 2.16b,c showed us that away from the glider transect corrections to the  $26.5 \text{ kg m}^{-3}$  isopycnal are the same as those in case Surface Only and hence our study could not produce an example of large-scale subsurface corrections that can be attributed to the assimilation of glider observations.

In conclusion, the result of this study provides useful guidance for the inclusion of glider observations into the Oregon-Washington near-real-time DA system: to utilize glider observations only when SST and/or HFR coverage is good. In a near future, the SSH obtained by a wide-swath altimeter (*Fu and Rodriguez*, 2004) can possibly provide additional constraint limiting the erroneous eddy variability caused by in-situ DA.

Chapter 3: A Monte-Carlo Background Covariance Localization  
Method for an Ensemble-variational Assimilation System

I. Pasmans and A.L. Kurapov

Monthly Weather Review  
vol. 145, 4543-4557, doi.org/10.1175/MWR-D-16-0424.1, 2017  
©American Meteorological Society. Used with permission.

### 3.1 Abstract

Spurious long distance correlations in estimates of the background error covariance can deteriorate the performance of ensemble-based data assimilation methods. In this study a localization method, called Monte-Carlo (MC) localization, is presented to remove these correlations. It is particularly useful for use in high-dimensional ensemble-variational data assimilation systems. In this method, raw ensemble members are truncated by multiplying them with functions having compact support. This creates a larger ensemble, in which points spaced further apart than the size of the compact support have zero correlation. The localized background error covariance is then estimated as the sample covariance of this larger ensemble. It is hypothesized that this localized background error covariance can be approximated by the MC approximation method using a limited set of the truncated ensemble members. This hypothesis is tested here on a grid with 1001 grid points and assuming a Gaussian true background error covariance. It is found that the mean relative error has an upperbound that scales with the inverse square root of the number of truncated ensemble members. In the case studied the size of the support for which the localized background covariance best approximates the true background covariance increases with increasing number of raw ensemble members and is close to 4 times the standard deviation of the Gaussian when 20 raw ensemble members are used. In the Fourier space the localization manifests itself as a convolution resulting in smoothing of the power spectral density of the ensemble members.

### 3.2 Introduction

Operational implementations of 4DVAR data assimilation in ocean models (*Kurapov et al., 2011; Moore et al., 2011; Ngodock and Carrier, 2014a; Yu et al., 2012*) require estimation of the covariance of the errors in the initial conditions prior to assimilation, or the background error covariance. The operational model that motivated our study is the Oregon-Washington coastal forecasting system (*Erofeeva, 2018*). The configuration of this model is similar to that described in *Kim et al. (2014)*. Its

domain is approximately 400 by 800 km with a 2-km resolution in the horizontal and 40 terrain-following layers in the vertical. Its total dimensionality is of  $O(10^7)$ . The model assimilates sea surface temperature, altimeter sea surface height, and daily-averaged surface current observations from high-frequency radars and uses a static background covariance (*Kurapov et al.*, 2011). However, the background error statistics are expected to vary in space and time as result of changing environmental conditions and in particular due to the presence the Columbia River plume (*Liu et al.*, 2009). To capture this variability, the background error covariance can be estimated from an ensemble of model runs.

A well-known problem with the estimation of the covariance from a limited number of ensemble members is the presence of spurious long-tail correlations that are detrimental to data assimilation performance (*Hamill et al.*, 2001; *Houtekamer and Mitchell*, 2001). To remove these correlations, various localization schemes have been proposed.

Many known localization schemes multiply the sample covariance on an element-by-element basis with a localized correlation matrix  $\mathbf{C}$ , i.e. they use the Schur product. It is often assumed that correlations in  $\mathbf{C}$  decrease with physical distance. In particular, piecewise rational functions with compact support (*Gaspari and Cohn*, 1999) can be used to obtain sparse  $\mathbf{C}$  (*Clayton et al.*, 2013; *Buehner et al.*, 2009; *Houtekamer and Mitchell*, 2001). Alternatively, ensemble members can be projected on a complete set of basis functions in spectral space, followed by localization of the correlation between the projection coefficients. A variety of choices for functions and localizations have been used. For example, in *Derber and Bouttier* (1999) and *Berre* (2000) horizontal Fourier basis functions are used and localization is performed by setting correlations between basis functions with different wave numbers to zero. *Buehner and Charron* (2007) also localize in a spectral space spanned by the Fourier basis. However, they replace the Schur product in spectral space with convolution of the ensemble members with *Gaspari and Cohn* (1999) compact piecewise rational functions in physical space. In *Deckmyn and Berre* (2005), two-dimensional (2D) Meyer wavelets are used as basis functions and the coefficients associated with different wavelets are set to zero. Com-

binations of the two aforementioned approaches have also been used: *Buehner* (2012) and *Buehner and Shlyaeva* (2015) band-pass filter the ensemble members and then apply spatial localization to each of the band-pass filtered ensemble members using different localization scales for different bands. The disadvantage of these techniques is that they are nonadaptive: the matrix  $\mathbf{C}$  does not depend on the model state. To overcome this disadvantage, *Bishop and Hodyss* (2007, 2009a,b) constructed  $\mathbf{C}$  from the ensemble sample correlation by repeatedly taking the Schur product with itself. This reduces weak correlations stronger than strong correlations. To correct for the decrease of the strong correlations the result is smoothed. This adaptive method can be combined with the aforementioned nonadaptive methods (*Bishop and Hodyss*, 2011). *Anderson* (2007, 2012) take a different approach that does not require the construction of a localized correlation matrix  $\mathbf{C}$ . Instead the data assimilative correction created by each observation is adjusted for sampling errors stemming from the use of an ensemble of limited size.

Not all of these localization schemes are suitable for implementation in our operational model. Transformation of ensemble members to/from spectral space is problematic near the irregularly shaped coast. For other schemes mentioned above implementation in an ensemble-variational model would be computationally expensive as they require the calculation of matrix-matrix products with matrices of dimensionality of  $O(10^7)$  (*Bishop and Hodyss*, 2009a), require the increase of the ensemble size (*Anderson*, 2007) or application of 4DVAR to each of the  $O(10^4)$  observations separately (*Anderson*, 2007, 2012). For efficient implementation of nonadaptive localization in ensemble covariances in three-dimensional (3D) and four-dimensional (4D) variational data assimilation schemes, *Bishop et al.* (2011) proposed (i) to perform the localization on a coarser grid, (ii) to use a localized correlation matrix in which the elements of the square-root  $\mathbf{C}^{1/2}$  are separable functions of horizontal and vertical coordinates and (iii) assume that  $\mathbf{C}^{1/2}$  is independent of the model variable localized. However, construction of  $\mathbf{C}^{1/2}$  or the use of *Gaspari and Cohn* (1999) was found to be time consuming using computer resources available to us.

To reduce the time and computational cost involved in localization compared to

*Gaspari and Cohn* (1999) and *Bishop et al.* (2011), we have constructed an alternative localization method, called Monte-Carlo (MC) localization. This alternative method is described in detail in section 3.3. In section 3.4 the computational cost of this method for a generic 3D model is compared to that of *Bishop et al.* (2011) and a method that uses  $\mathbf{C}$  constructed using the piecewise 5th-order rational function from *Gaspari and Cohn* (1999). In section 3.5 we will look how well MC localization approximates the true background error covariance in a one-dimension (1D) case with a Gaussian true covariance. Section 3.6 gives an example of MC localization implemented in a 3D oceanographic model. Section 3.7 discusses some of the drawbacks of MC localization and possible directions for improvement and contains a summary of our findings.

### 3.3 Method

In this paper we will assume that localization is applied in a finite-difference model on a structured grid with  $N_x, N_y, N_z$  grid points in three directions ( $x, y, z$ ) respectively, that the spacing in each direction is equidistant and that  $N_r$  quantities are defined at each grid point. In this case the model state can be ordered as a vector of dimension  $N = N_x N_y N_z N_r$ . The support of a model state vector is defined as the set of grid points on which the elements in the vector can be nonzero.

By definition a correlation matrix  $\mathbf{C}$  must be symmetric, nonnegative definite and its diagonal elements must be equal to 1. The first two requirements can be met by defining  $\mathbf{C}$  as the sum of symmetric rank-1 matrices:

$$\mathbf{C} = \sum_{m \in \mathbb{M}} (\boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma})(\boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma})^T. \quad (3.1)$$

Here  $\circ$  denotes the element-by-element multiplication of the Schur product,  ${}^T$  the matrix transpose,  $\boldsymbol{\chi}^{(m)}$  are vectors of dimension  $N$ , while the indices  $m$  are elements of some countable set  $\mathbb{M} \subset \mathbb{N}$  of size  $M$ , i.e.  $m = 1, 2, \dots, M$ . For the moment we only require the vectors  $\boldsymbol{\chi}^{(m)}$  to be unique, but in the following  $\boldsymbol{\chi}^{(m)}$  will be narrowed

down to a specific class of vectors.  $\boldsymbol{\gamma}$  is a  $N$ -dimensional normalization vector that ensures that  $\mathbf{C}$  has ones on its main diagonal. To achieve this, its elements must satisfy

$$\gamma_p = \left( \sum_{m \in \mathbb{M}} (\chi_p^{(m)})^2 \right)^{-1/2}. \quad (3.2)$$

To make  $\mathbf{C}$  a localization matrix, we add the requirement that  $\mathbf{C}_{pq} = 0$  if  $d(p, q)$ , the distance between points  $p$  and  $q$  in the metric  $d$ , is larger than a given localization distance. The distance between grid points  $p$  and  $q$  will be less than or equal to this localization distance if and only if there is a metric ball that has the localization distance as its diameter and contains both points  $p$  and  $q$ . So, to satisfy the localization requirement,  $\chi^{(m)}$  must be a vector with its support in one of these balls.

Let  $\{\mathbf{x}^{(k)}\}_k$ , with  $k = 1, 2, \dots, K$ , be an ensemble of  $N$ -dimensional model state vectors and let  $\bar{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}^{(k)}$  be the ensemble mean. Then the raw, nonlocalized, ensemble sample covariance is

$$\mathbf{B}_{ens} = \frac{1}{K-1} \sum_{k=1}^K (\mathbf{x}^{(k)} - \bar{\mathbf{x}})(\mathbf{x}^{(k)} - \bar{\mathbf{x}})^T. \quad (3.3)$$

Using the square root theorem (*Bishop and Hodyss*, 2009b, eq. 1) the localized estimate  $\mathbf{B}_{ens} \circ \mathbf{C}$  for the background error covariance can be written as:

$$\mathbf{B} = \mathbf{B}_{ens} \circ \mathbf{C} = \frac{1}{K-1} \sum_{k=1}^K \sum_{m \in \mathbb{M}} ((\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \chi^{(m)} \circ \boldsymbol{\gamma}) ((\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \chi^{(m)} \circ \boldsymbol{\gamma})^T. \quad (3.4)$$

This is an  $N \times N$ -dimensional matrix of rank  $M(K-1)$ . Since (3.4) has the same structure as (3.3), the  $N$ -dimensional vectors  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \chi^{(m)} \circ \boldsymbol{\gamma}$  will be referred to as modulated ensemble members (*Bishop and Hodyss*, 2011).

From now on we limit ourselves to  $\chi^{(m)}$  that are step functions on some ball. In particular, if  $L_x, L_y, L_z$  are the localization distances in the x,y,z-directions respectively, we define  $\chi^{(m)}$  as an  $N$ -dimensional vector with elements  $(\chi^{(m)})_p = 1$  if  $L_x^{-1}|x_p - x_m| < \frac{1}{2}$ ,  $L_y^{-1}|y_p - y_m| < \frac{1}{2}$ ,  $L_z^{-1}|z_p - z_m| < \frac{1}{2}$  and zero otherwise. In this

way each grid point gives rise to one unique vector  $\chi^{(m)}$  and hence  $M = N_x N_y N_z$  for this case. From here on we will refer to the  $\mathbf{B}$  constructed with all  $KM$  vectors  $((\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \chi^{(m)} \circ \gamma)$  as  $\mathbf{B}_{full}$ .

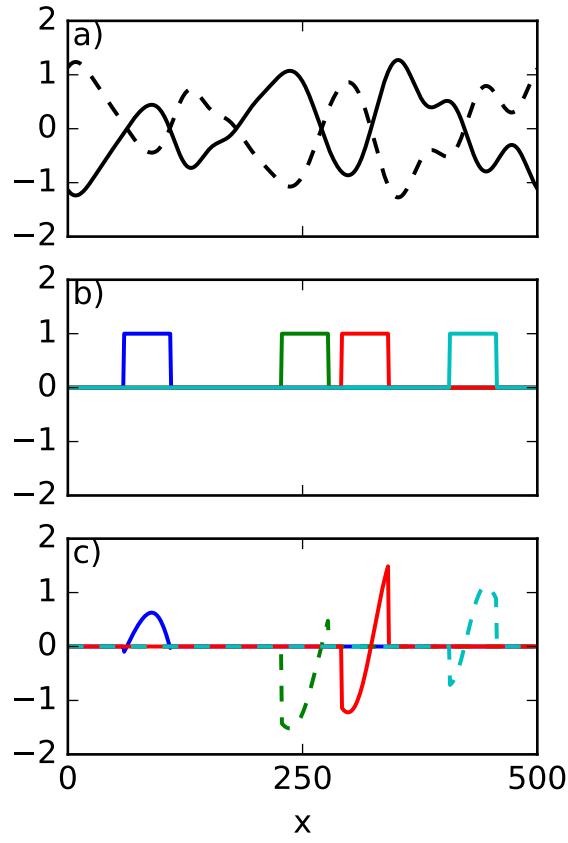


Figure 3.1: 1D example showing the construction of the modulated ensemble members for the case  $L_{MC} = 51$  on a 501 point grid. a) Two raw ensemble members  $\mathbf{x}^{(1)} - \bar{\mathbf{x}}$  (solid) and  $\mathbf{x}^{(2)} - \bar{\mathbf{x}}$  (dashed). b) Vectors  $\chi^{(85)}$  (dark blue),  $\chi^{(253)}$  (green),  $\chi^{(317)}$  (red),  $\chi^{(432)}$  (light blue). c) The modulated ensemble members  $(\mathbf{x}^{(1)} - \bar{\mathbf{x}}) \circ \chi^{(85)} \circ \gamma$  (dark blue),  $(\mathbf{x}^{(2)} - \bar{\mathbf{x}}) \circ \chi^{(253)} \circ \gamma$  (green),  $(\mathbf{x}^{(1)} - \bar{\mathbf{x}}) \circ \chi^{(317)} \circ \gamma$  (red),  $(\mathbf{x}^{(2)} - \bar{\mathbf{x}}) \circ \chi^{(432)} \circ \gamma$  (light blue).

An example of how a modulated ensemble can be constructed using this choice of vectors  $\boldsymbol{\chi}^{(m)}$  is shown in figure 3.1. In this example there is only one model quantity, the model grid has only 1 spatial dimension and it has 501 points with a grid spacing of 1 ( $N = N_x = 501$ ). The compact support has a width of  $L_{MC} = 51$  grid points. For two points that are further than  $L_{MC}$  apart, e.g. point 340 and point 430, it is impossible to find a  $\boldsymbol{\chi}^{(m)}$  in which both these points are nonzero. Consequently, a modulated ensemble member in which these points are both nonzero does not exist and hence  $(\mathbf{B}_{full})_{340,430} = 0$  in this case.

The modulated ensemble is a factor  $N_x N_y N_z$  larger than the raw ensemble. For our oceanographic model  $N_x N_y N_z \approx O(10^7)$  and  $K = 50$  such that calculation of  $\mathbf{B}_{full}$  is considerably more computationally intensive than that of  $\mathbf{B}_{ens}$ . To speed up calculation of the localized ensemble covariance, we propose replacing (3.4) with its Monte-Carlo (MC) approximation.

$$\mathbf{B}_{MC} = \frac{1}{K-1} \sum_{k=1}^K \sum_{m \in \mathbb{M}_k} ((\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma}) ((\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma})^T, \quad (3.5)$$

where  $\mathbb{M}_k$  is a randomly drawn subset of  $\mathbb{M} = \{1, 2, \dots, N_x N_y N_z\}$  of size  $M_{MC}$  and where  $\mathbb{M}_p$  can be different from  $\mathbb{M}_q$  if  $p \neq q$ . So, in this approximation  $\mathbf{B}$  is constructed from  $K M_{MC}$  modulated ensemble members. The elements of  $\boldsymbol{\gamma}$  must be given by

$$\boldsymbol{\gamma}_p = \left( \frac{1}{K} \sum_{k=1}^K \sum_{m \in \mathbb{M}_k} (\boldsymbol{\chi}_p^{(m)})^2 \right)^{-1/2} \quad (3.6)$$

to ensure that (3.4) and (3.5) are the same if no MC approximation is used, i.e. if  $\mathbb{M}_k = \mathbb{M}$  for all  $k$ .

### 3.4 Computational Cost

In 4DVAR the matrix  $\mathbf{B}$  is never explicitly calculated, but always appears in the form of a matrix-vector product. Here we will compare the computational complex-

ity of the matrix-vector computation with the localized background error covariance matrix for three different methods: (i) localization with  $\mathbf{C}$  constructed using *Gaspari and Cohn* (1999) 5th-order rational functions followed by a standard matrix-vector multiplication, (ii) localization using *Bishop et al.* (2011) methodology assuming  $\mathbf{C}^{1/2}$  is readily available, and (iii) MC localization as described above. As in section 3.3 we assume that localization is applied in a finite-difference model with a structured grid that has  $N_x$  points in the horizontal x-direction,  $N_y$  in the horizontal y-direction,  $N_z$  in the vertical z-direction and that  $N_r$  quantities are defined at each grid point, giving the model a total dimensionality of  $N = N_x N_y N_z N_r$ . In all three cases an arbitrary  $N$ -dimensional vector used in the calculation of the matrix-vector product is denoted as  $\mathbf{v}$ . The number of raw ensemble members is  $K$ . The  $O(Z)$  notation is used to indicate that the number of computational operations (addition,multiplication,exponentiation,etc.) is smaller than  $cZ$  as  $Z \rightarrow \infty$ , where  $c > 0$  is a scalar.

The calculation of the ensemble mean  $\bar{\mathbf{x}}$  and the deviations of the ensemble members from the ensemble mean requires  $O(KN)$  operations in all considered methods. When using *Gaspari and Cohn* (1999), it is convenient to rewrite the  $N$ -dimensional vector  $(\mathbf{B}_{ens} \circ \mathbf{C})\mathbf{v}$  as

$$\frac{1}{K-1} \sum_{k=1}^K (\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ (\mathbf{C}((\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \mathbf{v})). \quad (3.7)$$

The calculation of each of the two  $N$ -dimensional Schur products in (3.7) requires  $O(N)$  operations per ensemble member. Calculation of the inner product of the  $i$ th row of the sparse  $N \times N$ -matrix  $\mathbf{C}$  with  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \mathbf{v}$  requires  $L_G$  operations, where  $L_G$  is the number of grid points that lie in the compact support of the 5th-order rational function. The summation of the  $N$ -dimensional terms in (3.7) requires an additional

$O(KN)$  operations. In total, we find that the direct implementation of (3.7) requires

$$O(KN) + O(KN) + O(KNL_G) + O(KN) \sim O(KNL_G) \quad (3.8)$$

operations.

*Bishop et al.* (2011) describes an efficient method for ensemble covariance localization when  $\mathbf{C}^{1/2}$  is readily available and each element is the product of one of  $M_h$  horizontal basis functions with one of  $M_z$  vertical basis functions. First, the  $N$ -dimensional vector  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \mathbf{v}$  needs to be calculated at the expense of  $O(N)$  operations per ensemble member. Next, the different quantities are summed ( $O(N)$  computations per ensemble member) followed by the calculation of the inner-product of the resulting  $N$ -dimensional vector with each vertical basis function ( $O(N_x N_y N_z M_z)$  operations), producing an  $N_x N_y \times M_z$ -matrix. Then the inner-product of each column of this matrix with the horizontal basis functions is obtained ( $O(N_x N_y M_z M_h)$  operations), resulting in a vector of size  $M_h M_z$ . Subsequently, the transpose of the aforementioned operations is applied in reverse order. Finally, the  $K$   $N$ -dimensional vectors obtained in this way are summed together at the expense of  $O(KN)$  operations. In total this will require

$$\begin{aligned} & O(KN) + 2K(O(N) + O(N) + O(N_x N_y N_z M_z) + O(N_x N_y M_z M_h)) \\ & \sim O(KN \frac{M_z}{N_r} \max(1, \frac{M_h}{N_z})) \end{aligned} \quad (3.9)$$

operations.

For MC localization we first need to calculate the  $N$ -dimensional vector  $\boldsymbol{\gamma}$  at the expense of  $O(KM_{MC}L_{MC})$  operations, where  $L_{MC}$  is the number of grid points in the support of one of the vectors  $\boldsymbol{\chi}^{(m)}$ . Calculation of the  $N$ -dimensional vectors  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \boldsymbol{\gamma}$  and  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \boldsymbol{\gamma} \circ \mathbf{v}$  requires an additional  $O(N)$  operations for each ensemble member. The computation of  $\mathbf{B}_{MC}\mathbf{v}$  is then completed by selecting  $M_{MC}$  pieces from each of these vectors, with each piece containing  $L_{MC}N_r$  elements, and adding these parts together. This requires  $O(KM_{MC}L_{MC}N_r)$  operations. Hence the

total number of operations required is

$$\begin{aligned} & O(KN) + O(KM_{MC}L_{MC}) + O(KN) + O(KM_{MC}L_{MC}N_r) \\ & \sim O(KN \max(1, \frac{M_{MC}L_{MC}}{N_x N_y N_z})). \end{aligned} \quad (3.10)$$

A comparison of (3.8) and (3.10) shows that MC-localization is less computationally intensive than *Gaspari and Cohn* (1999) if  $\frac{M_{MC}L_{MC}}{N_x N_y N_z L_G} < 1$ . So, if the compact supports are of comparable size in the two methods ( $L_{MC} \approx L_G$ ) and if the size of  $\mathbb{M}_k$  is chosen to be smaller than the number of grid points, MC-localization is computationally more efficient than using  $\mathbf{C}$  generated with *Gaspari and Cohn* (1999). Comparison of (3.9) and (3.10) shows that MC localization is computationally more efficient than *Bishop et al.* (2011) if  $\frac{M_{MC}L_{MC}N_r}{M_z M_h N_x N_y} < 1$ . No general statement can be made whether this inequality is indeed satisfied, as  $M_z, M_h, M_{MC}$  and  $L_{MC}$  will vary depending on  $\mathbf{C}^{1/2}$  used and the number of modulated ensemble members in the MC approximation, but choices for which the inequality is satisfied do exist. For example, if the case in which no MC approximation is used ( $M_{MC} = N_x N_y N_z$ ) is compared with a case in which localization takes place in spectral space ( $M_h M_z = N_x N_y N_z$ ) as is done e.g. in *Gauthier et al.* (1998) and *Kuhl et al.* (2013) for the horizontal directions, the ratio of the operation counts between MC-localization and *Bishop et al.* (2011) is  $\frac{L_{MC}N_r}{N_x N_y}$  which is smaller than 1 if the horizontal (vertical) localization scale is much smaller than the horizontal (vertical) extent of the model.

### 3.5 MC Localization in a 1D Model

In a strong-constraint 4DVAR system where only the initial conditions are corrected, the data assimilation correction at time  $t_0$  is found by minimizing the following cost function:

$$J = \frac{1}{2} \delta \mathbf{x}(t_0)^T \mathbf{B}^{-1} \delta \mathbf{x}(t_0) + \frac{1}{2} (\mathbf{d} - \mathbf{H} \mathbf{M} \delta \mathbf{x}(t_0))^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H} \mathbf{M} \delta \mathbf{x}(t_0)) \quad (3.11)$$

(Bennett, 1992; Courtier *et al.*, 1994), where the column vector  $\delta\mathbf{x}(t_0)$  is the correction to the modeled ocean state,  $\mathbf{B}$  the localized background error covariance in the initial conditions,  $\mathbf{R}$  the observational error covariance,  $\mathbf{d}$  the innovation vector ( difference between observations made during a time window  $[t_0, t_1]$  and prior model predictions for these observations),  $\mathbf{H}$  the linear operator that samples the model result between  $[t_0, t_1]$  and generates model predictions for the observations,  $\mathbf{M}$  the tangent linear model operator that takes the model perturbation at  $t = t_0$  as input and outputs model results for the period  $[t_0, t_1]$  using a linearized version of the model equations, and  $\mathbf{M}^T$  is its adjoint. Repeated minimizations of (3.11) with updated linearizations  $\mathbf{M}$  and initial conditions are necessary if one wants to account for nonlinearities in the model.  $J$  obtains its minimum for

$$\delta\mathbf{x}(t_0) = \mathbf{B}\mathbf{M}^T\mathbf{H}^T(\mathbf{HMBM}^T\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{d} \quad (3.12)$$

(Courtier, 1997). From (3.12) it follows that when  $\mathbf{B}$  is constructed using (3.4) or (3.5), the model correction  $\delta\mathbf{x}(t_0)$  is a weighted combination of the modulated ensemble members  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma}$ . In this section we will (i) investigate how well  $\mathbf{B}_{MC}$  approximates  $\mathbf{B}_{full}$  as  $M_{MC}$ , the number of elements in the set  $\mathbb{M}_k$ , changes and (ii) analyze the power spectral density of the modulated ensemble members. All experiments in this section use a 1D grid with  $N = 1001$  grid points and one unitless variable to denote the position in space.

We draw the ensemble members  $\mathbf{x}^{(k)}$  for this 1D model by first constructing the Fourier transform, then applying an inverse Fourier transform to it giving a function  $x^{(k)}$  and finally evaluating this function at the grid points. The Fourier transform for ensemble member  $k$  is created as:

$$\mathcal{F}x^{(k)}(\kappa) = \sum_{l=-N_{1/2}}^{N_{1/2}} A_l \left(\frac{2\pi}{N}\right)^{1/2} (2\pi\sigma^2)^{1/4} \exp\left(-\frac{1}{4}\sigma^2\kappa_l^2\right) \delta(\kappa - \kappa_l). \quad (3.13)$$

Here  $\kappa$  is the wave number,  $N_{1/2} = \frac{1}{2}(N - 1)$ ,  $\kappa_l = \frac{2\pi l}{N}$  and

$$A_l = \begin{cases} = A_{Re}(0) & \text{if } l = 0 \\ = \frac{1}{\sqrt{2}}A_{Re}(l) + \frac{i}{\sqrt{2}}A_{Im}(l) & \text{if } l > 0 \\ = A_{-l}^* & \text{if } l < 0 \end{cases} \quad (3.14)$$

where  $i^2 = -1$  and  $A_{Re}(l)$ ,  $A_{Im}(l)$  are drawn from a Gaussian distribution with zero mean and unit variance. Then  $\langle \mathcal{F}x^{(k)}(\kappa) \rangle = 0$ , with  $\langle \cdot \rangle$  denoting the expectation value, and hence the true model error mean  $\langle x^{(k)} \rangle = 0$ . Furthermore, by the Wiener-Kinchin theorem (Engelberg, 2007) the power spectral density  $S(x^{(k)})$  is related to the covariance function  $\rho$  as:

$$S(x^{(k)}) = \langle |\mathcal{F}x^{(k)}|^2(\kappa) \rangle = \sum_{l=-N_{1/2}}^{N_{1/2}} \frac{2\pi}{N} (2\pi\sigma^2)^{1/2} \exp(-\frac{1}{2}\sigma^2\kappa_l^2) \delta(\kappa - \kappa_l) = \mathcal{F}\rho(\kappa) \quad (3.15)$$

Taking the inverse Fourier transform of the right-hand side of (3.15) then shows that the true covariance of the ensemble members created this way is

$$\begin{aligned} \rho(p, q) &\stackrel{def}{=} \langle (\mathbf{x}_p^{(k)} - \langle \mathbf{x}_p^{(k)} \rangle)(\mathbf{x}_q^{(k)} - \langle \mathbf{x}_q^{(k)} \rangle) \rangle \\ &= \sum_{l=-N_{1/2}}^{N_{1/2}} \frac{2\pi}{N} (2\pi\sigma^2)^{1/2} \exp(-\frac{1}{2}\sigma^2\kappa_l^2) \mathcal{F}^{-1}[\delta(\kappa - \kappa_l)](d(p, q)) \\ &= \sum_{l=-N_{1/2}}^{N_{1/2}} \frac{2\pi}{N} \frac{\sigma}{(2\pi)^{1/2}} \exp(-\frac{1}{2}\sigma^2\kappa_l^2) \exp(i\kappa_l d(p, q)) \\ &\approx \int_{-\infty}^{\infty} \frac{\sigma}{(2\pi)^{1/2}} \exp(-\frac{1}{2}\sigma^2\kappa^2) \exp(i\kappa d(p, q)) d\kappa = \exp\left(-\frac{1}{2} \frac{d(p, q)^2}{\sigma^2}\right) \end{aligned} \quad (3.16)$$

where  $d(p, q)$  is the (dimensionless) distance between the grid points  $p$  and  $q$  and the integral approximation in the last line of (3.16) holds if  $\frac{2\pi\sigma}{N} \ll 1$  (necessary to replace the sum with an integral) and  $\pi\sigma \gg 1$  (necessary to allow extension of the integral boundaries to  $\pm\infty$ ).

### 3.5.1 Convergence (Dependence on $M_{MC}$ )

Decrease in  $M_{MC}$  lowers computation cost, but is expected to increase the sampling error in  $\mathbf{B}_{MC}$ . In this subsection we consider the dependence of the relative error  $\|\mathbf{B}_m \mathbf{v} - \mathbf{B}_{full} \mathbf{v}\| / \|\mathbf{B}_{full} \mathbf{v}\|^{-1}$  on  $M_{MC}$  for typical vectors  $\mathbf{v}$ . As inversion (3.12) is performed using an iterative method, e.g. conjugate-gradient,  $\mathbf{B}$  is repeatedly applied to vectors of the form  $\mathbf{M}^T \mathbf{H}^T \mathbf{r}$  where  $\mathbf{r}$  is the residual. In 4DVAR it is assumed that  $\mathbf{r}$  follows a Gaussian distribution with covariance  $\mathbf{HMBM}^T \mathbf{H}^T + \mathbf{R}$  and zero mean (Bennett, 1992) and hence under these assumptions  $\mathbf{M}^T \mathbf{H}^T \mathbf{r}$  itself follows a Gaussian distribution with covariance  $\mathbf{M}^T \mathbf{H}^T (\mathbf{HMBM}^T \mathbf{H}^T + \mathbf{R}) \mathbf{H} \mathbf{M}$  and zero mean. In this paper we will neglect the implementation-dependent intricacies contained in  $\mathbf{M}$ ,  $\mathbf{H}$  and  $\mathbf{R}$  and simply construct the typical vectors  $\mathbf{v}$  by drawing them from a Gaussian distribution with zero mean and with the identity matrix as covariance.

The relative error introduced by the Monte-Carlo approximation has been calculated for different values of  $M_{MC}$ . In our base experiment  $L_{MC} = 101$ , where  $L_{MC}$  is the maximum number of grid points in the support of the  $N$ -dimensional vectors  $\boldsymbol{\chi}^{(m)}$ . We draw a 10-member ensemble ( $K = 10$ ) with covariance (3.16), an  $N$ -dimensional vector  $\mathbf{v}$  from a standard normal distribution and calculate the relative error  $\|\mathbf{B}_m \mathbf{v} - \mathbf{B}_{full} \mathbf{v}\| / \|\mathbf{B}_{full} \mathbf{v}\|^{-1}$ . This procedure is repeated 1000 times. The mean of the relative error is shown in figure 3.2 as a solid black line. A nonlinear least-square fit to the average relative error finds that it scales as  $\sim M_{MC}^{-0.51}$ , close to the  $\sim M_{MC}^{-1/2}$  expected for a Monte-Carlo approximation (Geyer, 1992).

The same experiment has been repeated using  $K = 5$  and  $K = 20$  and the results are also shown in figure 3.2. Nonlinear fits to these lines show that they have similar  $\sim M_{MC}^{-0.5}$  dependence on  $M_{MC}$ , but that for  $K = 5$  the relative error is  $25 \pm 2\%$  larger than for the case  $K = 10$ , while for the case  $K = 20$  the relative error is  $26 \pm 1\%$  smaller. The relation between the mean relative error on one hand and  $K$  and  $M_{MC}$  on the other hand is further explored in figure 3.3. For  $(K, M_{MC}) \in \{5, 10, \dots, 50\} \times \{10, 20, \dots, 100, 150, \dots, 500, 600, \dots, 1000\}$  the relative error  $\|\mathbf{B}_m \mathbf{v} - \mathbf{B}_{full} \mathbf{v}\| / \|\mathbf{B}_{full} \mathbf{v}\|^{-1}$  is calculated for 1000 realizations and then averaged. Each of these averages is represented by one of the  $10 \times 23 = 230$  dots in figure 3.3.

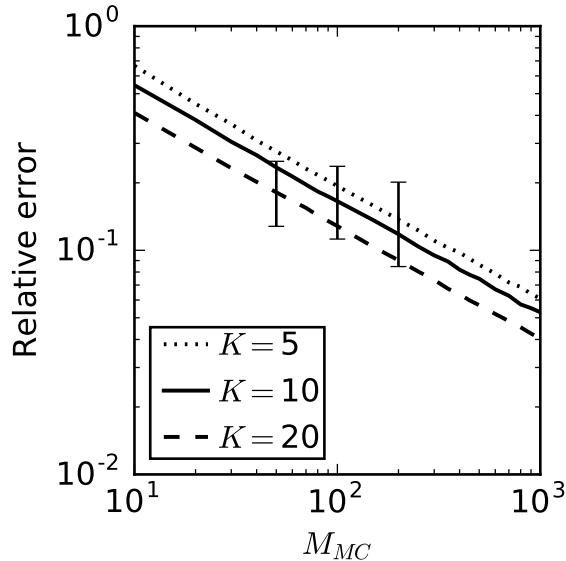


Figure 3.2: The relative error  $\|\mathbf{B}_{full}\mathbf{v} - \mathbf{B}_{MC}\mathbf{v}\| / \|\mathbf{B}_{full}\mathbf{v}\|^{-1}$  as function of  $M_{MC}$  (the size of the sets  $\mathbb{M}_k$  in the MC approximation) and different choices of  $K$  (the number of raw ensemble members); note that  $KM_{MC}$  is the total number of modulated ensemble members. The solid line shows the mean relative error averaged over 1000 different realizations of the raw ensemble, the modulated ensemble and the vector  $\mathbf{v}$ . Error bars indicate the interval around the mean in which 90% of the relative errors fall.

This figure shows that the relative error scales with  $\sim M_{MC}^{-0.5}$  if  $K$  is kept fixed. However, it also shows that the mean relative error can be bounded by a function proportional to  $(KM_{MC})^{-0.5}$ . As  $K$  becomes smaller the mean relative error attains values below the upper bound. This can be explained by the fact that modulated ensemble members created from the same raw ensemble member are better approximations of each other than modulated ensemble members created from different raw ensemble members. For example, vectors  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma}$  and  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \boldsymbol{\chi}^{(m+1)} \circ \boldsymbol{\gamma}$ , where  $m$  and  $m+1$  are neighboring points, are different in at most 2 elements while  $(\mathbf{x}^{(k)} - \bar{\mathbf{x}}) \circ \boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma}$  and  $(\mathbf{x}^{(k+1)} - \bar{\mathbf{x}}) \circ \boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma}$  can be different in as many as  $L_{MC}$  elements. In all cases in this experiment a mean relative error of less than 0.1 is

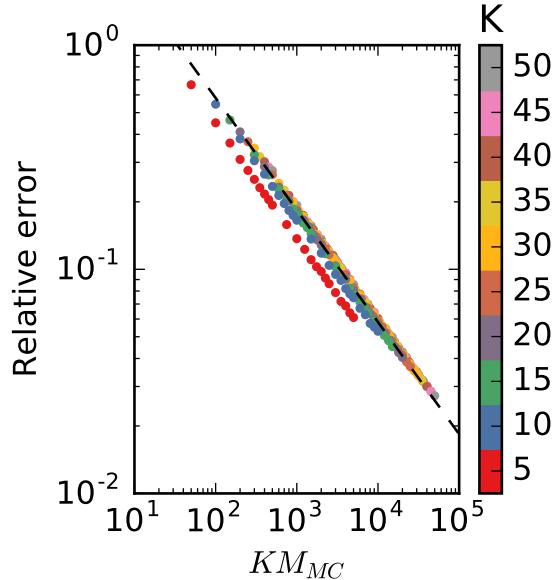


Figure 3.3: Each dot represents the mean relative error  $\|\mathbf{B}_{full}\mathbf{v} - \mathbf{B}_{MC}\mathbf{v}\| / \|\mathbf{B}_{full}\mathbf{v}\|^{-1}$  over 1000 realizations of  $\mathbf{B}_{full}$ ,  $\mathbf{B}_{MC}$  and  $\mathbf{v}$  for a different value of  $K$  and  $M_{MC}$ . Colors show the size of the raw ensemble ( $K$ ). The dashed line follows the relation  $5.8(KM_{MC})^{-1/2}$ .

reached if  $KM_{MC} \geq 3500$ .

So far the elements of the vector  $\mathbf{v}$  were drawn from a normal distribution. If observations only cover part of the domain,  $\mathbf{H}$  will be sparse and it is possible that the vector  $\mathbf{v} \approx \mathbf{M}^T \mathbf{H}^T \mathbf{r}$  mainly consists of zeros. To test the effect of such a structure of  $\mathbf{v}$  on aforementioned results, the previous experiment has been repeated using a 10-member ensemble ( $K = 10$ ) with  $\mathbf{v}$  obtained by drawing from distribution (3.16) and then randomly setting 90% of its elements to zero. The mean relative error still scales with  $\sim M_{MC}^{-0.5}$ , but was  $0.90 \pm 0.02$  times the mean relative error found in the previous experiment. So, in this 1D experiment MC-localization does not behave differently when  $\mathbf{r}$  is sparse.

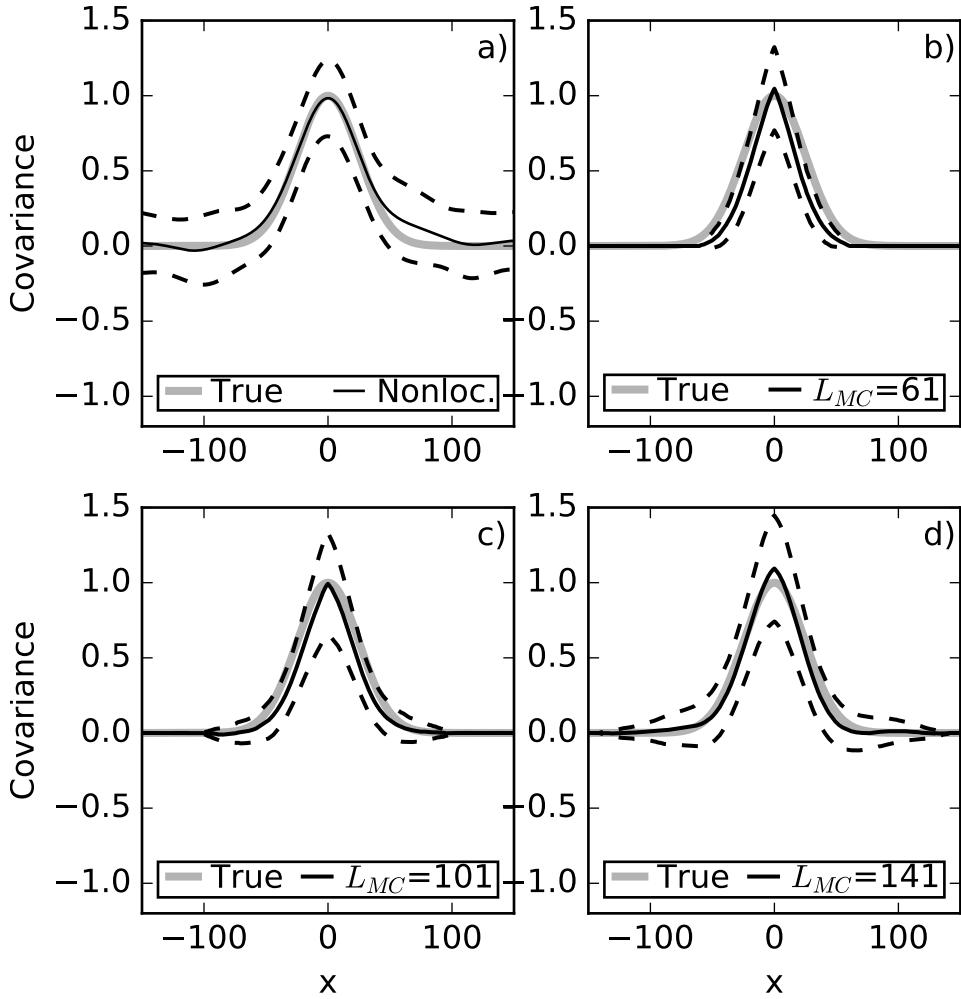


Figure 3.4: The comparison of the true (gray) and ensemble-generated covariances (black) computed with respect to the domain center point: a) mean (solid black line) +/- standard deviation (dashed black lines) of 50 raw ensemble covariances. b) mean (solid black line) +/- standard deviation (dashed black lines) of 50 modulated ensemble covariances with compact support length  $L_{MC} = 61$ . c) as for b) but now  $L_{MC} = 101$ . d) as for b) but with  $L_{MC} = 141$ .

Table 3.1: Proportionality constant  $\alpha$  ( $\pm$  standard deviation) between  $\hat{L}_{MC}$  and  $\sigma$  as function of the number of raw ensemble members  $K$ .

$K$	10	15	20	25	30	40	50
$\alpha$	$3.2 \pm 0.2$	$3.2 \pm 0.14$	$3.6 \pm 0.2$	$4.1 \pm 0.3$	$3.7 \pm 0.2$	$4.9 \pm 0.3$	$4.4 \pm 0.2$

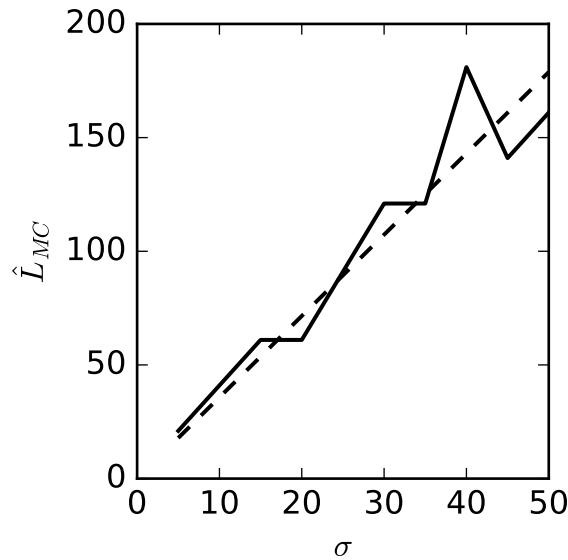


Figure 3.5: The relation between  $\sigma$  as appearing in (3.16) and the value for  $L_{MC}$  for which the RMS error between the true covariance and the localized ensemble covariance is minimal. Dashed line shows the linear fit to the data.

### 3.5.2 Localization Distance ( $L_{MC}$ )

The goal of this subsection is to find  $L_{MC}$ , the number of grid points in the support of  $\chi^{(m)}$ , for which the modulated ensemble covariance best approximates the true covariance. For this purpose we have drawn a 20-member ensemble ( $K = 20$ ) with covariance (3.16), used MC localization with  $M_{MC} = 150$ , thereby creating an ensemble of 3000 modulated members, and calculated the covariance between the point in the center of the domain and its neighboring points. This experiment has been

repeated 50 times for  $L_{MC} = 61, 101, 141$ . For each choice of  $L_{MC}$ , the mean value and standard deviation of these 50 trials are shown in figure 3.4 together with the true covariance and the nonlocalized ensemble covariance. The nonlocalized ensemble covariance (figure 3.4a) does not tend to zero at large distances, but exhibits spurious long distance covariances with a typical size of 0.2 (dashed lines). MC localization is successful at removing these spurious covariances. The figure also shows that on average the width of the localized covariances is smaller than that of the true covariance. This narrowing is most pronounced for  $L_{MC} = 61$ . It is less for  $L_{MC} = 141$ , but the larger standard deviations in the tails show that some of the modulated ensembles generated during the 50 trials have a nonzero covariance where the true covariance is nearly zero. This suggests that there is some optimal value for  $L_{MC}$ , hereafter referred to as  $\hat{L}_{MC}$ , for which the root mean square (RMS) error between the true covariance as given in (3.16) and the localized ensemble covariance is minimal. Repeating this experiment using different values of  $\sigma$  and  $L_{MC}$  with  $KM_{MC} = 4000$  shows that this is indeed the case. The found values for  $\hat{L}_{MC}$  are shown in figure 3.5. A linear fit to the data shows that  $\hat{L}_{MC} = (3.6 \pm 0.2)\sigma$ . In a similar experiment with fixed  $KM_{MC} = 4000$ , but with different  $K$  similar linear relations are found, but with different proportionality constants. As Table 3.1 shows, the proportionality constant remains between 3 – 4 for a small ensemble ( $K \leq 30$ ), but starts increasing when the number of raw ensemble members increases. This is in agreement with the increase in optimal localization scales observed in *Hamill et al.* (2001) and *Kirchgessner et al.* (2014). An experiment with  $K = 20$  and  $M_{MC}$  varying between 100 and 400 gives proportionality coefficients between  $3.4 \pm 0.2$  and  $4.1 \pm 0.2$  with no clear dependence on  $M_{MC}$ .

### 3.5.3 Shape of Covariance

The covariance function in (3.16) is positive. In reality a negative background error covariance might be expected. For example, when the strength of the circulation around an eddy varies in different ensemble members, the background errors in a component of the horizontal velocity field on different sides of the eddy will be nega-

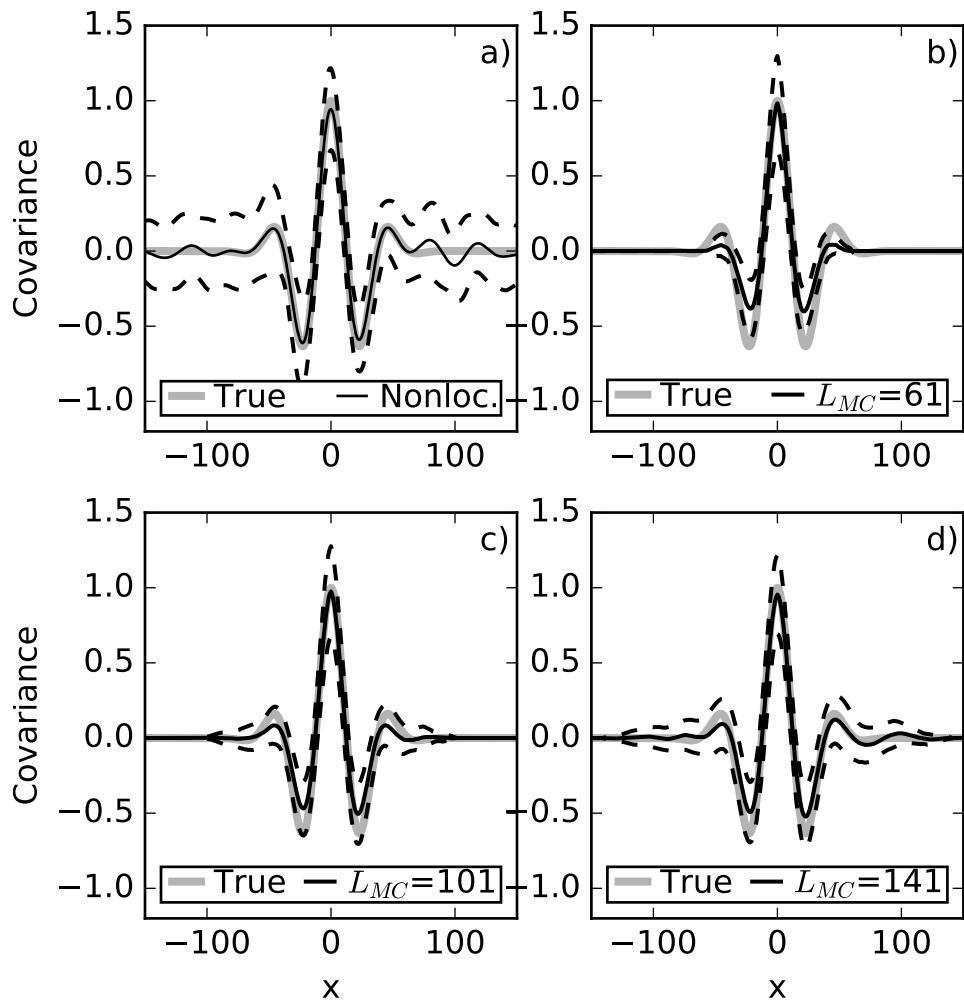


Figure 3.6: As figure 3.4, but now using ensemble members having (3.18) as true covariance function.

tively correlated with each other. To test the effect of MC localization in this case, the foregoing experiment is repeated, but now with raw ensemble members constructed using

$$\mathcal{F}x^{(k)}(\kappa) = \sum_{l=-N_{1/2}}^{N_{1/2}} \left[ A_l \left( \frac{2\pi}{N} \right)^{1/2} (2\pi\sigma^2)^{1/4} \sqrt{\frac{1}{2} \exp(-\frac{1}{2}(\sigma\kappa_l - \pi)^2) + \frac{1}{2} \exp(-\frac{1}{2}(\sigma\kappa_l + \pi)^2)} \delta(\kappa - \kappa_l) \right] \quad (3.17)$$

with  $A_l$  as defined in (3.14). Following a similar procedure as outlined for (3.15) and (3.16) this creates raw ensemble members with an covariance function given by

$$\rho(d(p, q)) \approx \cos\left(\frac{d(p, q)\pi}{\sigma}\right) \exp\left(-\frac{d(p, q)^2}{2\sigma^2}\right). \quad (3.18)$$

Figure 3.6 shows that with a localization width  $L_{MC} = 101$  the mean of the localized covariances follows the true covariance. However, the negative covariances are not as strong as in the true covariance or in the nonlocalized sample covariance. As the distance between grid point  $p$  and grid point  $q$  becomes larger there will be fewer modulated ensemble members that have both points lying in their support. Hence as  $d(p, q)$  becomes larger there will be fewer and fewer terms in (3.4) for which the  $p, q$ th element is nonzero. Consequently  $(\mathbf{B}_{MC})_{p,q}$  will go to zero as well. Although on average the localized covariances are positive for  $x = \pm 50$ , just as the true covariance, figure 3.6 shows that a fraction of localized covariances constructed in this experiment had the wrong sign at this  $x$ . This is however not unique to localization, as the same happens for a fraction of the nonlocalized ensemble covariances.

### 3.5.4 Spectrum

When  $\mathbf{B}_{full}$  is calculated,  $\gamma \circ \chi^{(m)}$  is a boxcar function when grid point  $m$  lies further than  $L_{MC}$  from the domain boundary. As is shown in the appendix, the amplitude of its Fourier transform then scales with a sinc function with a main lobe half-width

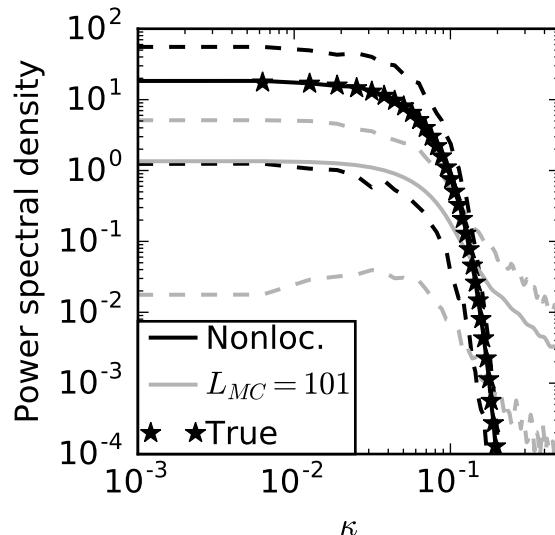


Figure 3.7: Power spectral density as function of the wave number  $\kappa$  for raw ensemble members (black) and for the modulated ensemble members with  $L_{MC} = 101$  and  $M_{MC} = 400$  (grey). Solid lines show the power spectral density averaged over 500 ensemble members. Dashed lines show the 5% and 95% percentiles. The power spectrum of raw ensemble members as given by (3.15) is shown as a line of stars.

proportional to  $L_{MC}^{-1}$ . The Fourier transform of the modulated ensemble member is a convolution of this function and the Fourier transform of the raw ensemble member. Consequently, the Fourier transform of the modulated ensemble member is a smoothed version of that of the raw ensemble member.

When the support of  $\chi^{(m)}$  contains part of the boundary of the domain or when elements  $\gamma$  vary because of the MC approximation, see (3.6), obtaining the spectrum is not as straightforward. To study the spectrum for this case, 10-member ensembles were drawn and localized using MC localization with  $L_{MC} = 101$  and  $M_{MC} = 400$ . Using the MATLAB periodogram routine, the one-sided power spectrum of the non-localized and modulated ensemble members was obtained. These steps were repeated 50 times creating 500 nonlocalized and 500 localized spectra.

The ensemble variance for element  $p$  is given by

$\frac{1}{K-1} \sum_{k=1}^K \sum_{m \in \mathbb{M}_k} (\mathbf{x}^{(k)} - \bar{\mathbf{x}})_p^2 \gamma_p^2 (\chi_p^{(m)})^2$  where  $\gamma$  in (3.2) and (3.6) was chosen such that the variance in the localized ensemble for element  $p$  is exactly (in the case of  $\mathbf{B}_{full}$ ) or approximately (in the case of  $\mathbf{B}_{MC}$ ) equal to the variance of the raw ensemble. For the raw ensemble  $K$  ensemble members contribute to the sum that gives the ensemble variance for element  $p$ , while for the modulated ensemble  $\sum_{k=1}^K \sum_{m \in \mathbb{M}_k} \chi_p^{(m)} = \gamma_p^{-2} K$  members contribute. As usually  $\gamma_p^{-2} > 1$  this means that each modulated ensemble members contributes less to the variance than each raw ensemble member. As the total spectral power is equal to the variance (*Oppenheim and Schafer, 1989*) this implies that the power spectral density of the modulated ensemble members has to be smaller than that of the raw ensemble members. In order to facilitate the comparison of the power spectra on the same scale, the aforementioned 500 localized power spectra have been rescaled by multiplying them with the average ratio of number of modulated members that contribute to the variance over the number raw ensemble members that contribute, i.e. by  $\frac{1}{N} \sum_{p=1}^N \gamma_p^{-2}$ .

The (scaled) mean power spectral density from the 500 selected members as well the 5% and 95% percentiles are shown in figure 3.7 together with the true spectrum from (3.15) as a function of the wave number  $\kappa$ . Comparing the averages in the figure shows that the localization scheme suppresses the power spectral density for wave numbers smaller than  $\kappa < 0.14$ , i.e. wave numbers corresponding to wave lengths longer than 45, or  $\frac{1}{2}L_{MC}$  in our case, by at most 93%.

### 3.6 Example from a 3D-model

The experiments in section 3.5 were all performed on a 1D grid. We have implemented MC localization in our 3D coastal ocean forecasting model described in section 3.2 using an ensemble size of  $K = 50$  and  $KM_{MC} = 750000$  modulated ensemble members. We use the first baroclinic Rossby radius of deformation, which is  $\sim 20$  km for the model region (*Chelton et al., 1998*), as an estimate for  $\sigma$ . This, in combination with values for  $\alpha$  for large ensembles as found in Table 3.1 has let us to set  $L_{MC} = 100$  km in our model. Figure 3.8 shows an example of the the surface temperature-salinity background error covariance in this model. As perturbations in

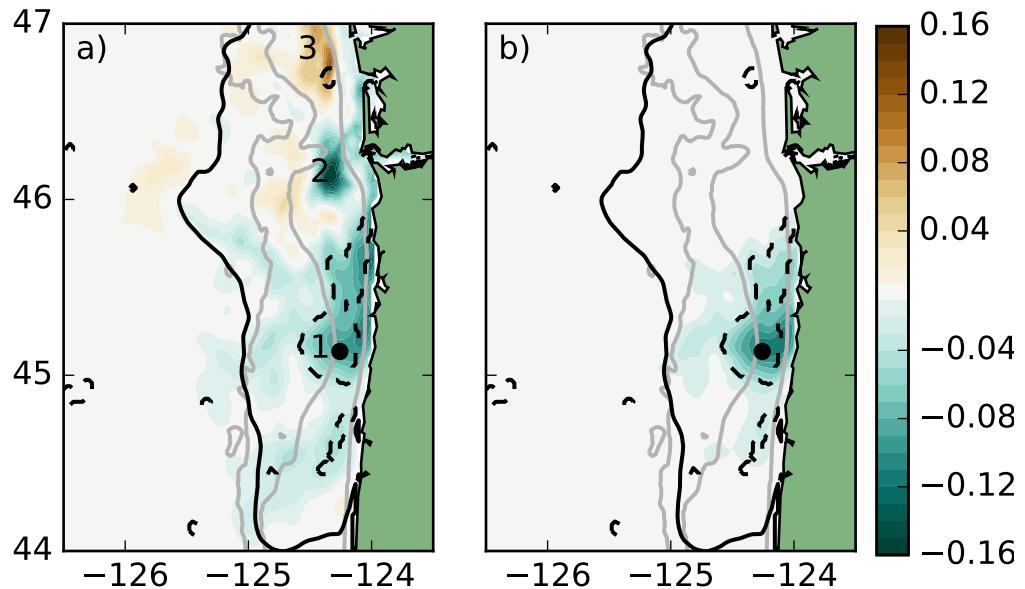


Figure 3.8: Ensemble covariance (in  $10^{-3}\text{C}$ ) on 0000 UTC 4 May 2011 between the sea surface temperature at the location marked by the dot and the sea surface salinity field in the Columbia River plume (black line). Left: for a nonlocalized 50-member ensemble. Right: for the same 50-member ensemble but now using MC localization with  $M_{MC} = 15000$  and  $L_{MC} = 51$  in both zonal and meridional direction corresponding to a localization distance of 100 km. In the area enclosed by the dashed lines the nonlocalized sample correlation is significantly different from zero at the 90%-level. The 200m, 1000m and 2000m isobaths are shown as grey lines.

the flow field alter the advection of river water and as the river outflow is generally fresher and warmer than the ocean this temperature-covariance is expected to be negative in and near the river plume. The figure shows that without localization the covariances with the largest magnitude can be found in the areas marked as 1,2 and 3. For both the shown and an independently generated 50-member ensemble, the p-values of the temperature-salinity correlations have been calculated using MATLAB's `corrcoef` routine. The area within the dashed lines shows where correlations were significantly different from zero at the 90%-confidence level in both ensembles. Apart from a small patch, the covariances in areas 2 and 3 are nonsignificant and

MC localization removes them successfully, while it maintains the nonzero, negative covariances in area 1 that are significant.

### 3.7 Summary and Concluding Discussion

Ensemble estimates of the background model error covariance contain spurious long-distance correlations that can degrade the performance of data assimilation algorithms. For our 4DVAR data assimilation system we sought a method to remove these correlations and that is computationally more efficient than popular localization methods like taking the Schur product of the ensemble sample covariance with  $\mathbf{C}$  generated by *Gaspari and Cohn* (1999) or the algorithm outlined by *Bishop and Hodyss* (2011). With this aim we introduced MC localization which rewrites the localized background error covariance as the sample covariance of a modulated (truncated) ensemble, in which the modulated ensemble members are created by selecting parts of size  $L$  from raw ensemble members. The order estimates of the computational cost show that MC localization is computationally more efficient than using  $\mathbf{C}$  generated by *Gaspari and Cohn* (1999). Whether MC localization is computationally more efficient than using *Bishop and Hodyss* (2011) depends on the details of the implementation used. In particular, computationally efficient MC localization requires that the number of modulated ensemble members that is used in the calculation of the localized background covariance is small. From a test on a 1D grid with 1000 grid points and a background error covariance shaped as a Gaussian it was found that for this case a MC approximation of the localized background covariance can be created using only a limited number of modulated ensemble members. For the case studied we found that the mean relative error introduced by this MC approximation has an upper bound that scales as  $\sim (KM_{MC})^{-1/2}$ , where  $K$  is the number of raw ensemble members and  $M_{MC}$  the number of modulated ensemble members created out of each raw ensemble member. Using this approximation MC localization was successful in removing spurious long-distance covariances in the 1D model as well as in the 3D example.

For our 1D case, we found that the optimal localization width  $L_{MC}$  increases with increasing length scale of the background errors. This relation qualitatively agrees with the findings by *Houtekamer and Mitchell* (2001). They calculated RMS errors for their ensemble-Kalman filter data assimilation system in which localization is performed by taking the Schur-product with  $\mathbf{C}$  generated by *Gaspari and Cohn* (1999) and found that the localization radius for which this error is minimal increases with increasing number of ensemble members. They attributed this to larger distance over which covariances can be computed accurately (*Houtekamer and Mitchell*, 1998). The more accurate computation is possible because the small correlations found between points spaced far apart become distinguishable as the statistical noise level drops with increasing number of ensemble members.

The rescaled power spectrum of the modulated ensemble members was found to be a smoothed version of the spectrum of the nonlocalized ensemble members. In particular, the rescaled power spectrum for the modulated ensemble members contains relatively more power for large wave numbers, while for small wave numbers spectral power is strongly reduced compared to the nonlocalized spectrum. This smoothing was shown to be due to the fact that the Fourier transform of most modulated ensemble members is a convolution of the Fourier transform of the nonlocalized ensemble member with a function that scales with the sinc-function. In this sense the effect of localization on the spectrum of the ensemble members is found to be comparable to that of windowing. This suggests that in future research it might be interesting to study the effects of varying the strength of the smoothing by using functions other than the boxcar for  $\chi^{(m)}$ .

A well-known drawback of the use of localization schemes in data assimilation is their tendency to generate corrections in which the flow is not in geostrophic balance (*Lorenc*, 2003; *Mitchell et al.*, 2002). Though not addressed in this paper, MC localization is expected to produce unbalanced corrections: even if the ensemble perturbations  $\mathbf{x}^{(k)} - \bar{\mathbf{x}}$  are in geostrophic balance, a modulated ensemble member will contain a discontinuity in pressure near the boundary of the boxcar function with no

geostrophic velocity to match it. Ways to mitigate the imbalance, like using smoother functions for  $\chi^{(m)}$  should be addressed in future work.

Another possible topic for future research is the use of multiple localization scales in MC localization. In this paper  $\chi^{(m)}$  was assumed to be independent of index  $m$  marking the location of the support, i.e. the localization distance was assumed to be homogeneous in space. Several inhomogeneous localization schemes have been developed in which  $\mathbf{C}$  is derived from the ensemble (*Bishop and Hodyss*, 2007, 2009a, 2011). It was found that at least in simple systems in which the scale of the error changes in space and in time these schemes outperform nonadaptive localization. Contrary to *Bishop and Hodyss* (2007) and *Bishop and Hodyss* (2009a), MC localization does not provide a way to estimate the correlation length scale. However, when local estimates of correlation scales are available, inhomogeneity is expected to be easily implementable in MC localization by making  $L_{MC}$  a function of  $m$ .

## Chapter 4: Ensemble 4DVAR (En4DVar) Data Assimilation in a Coastal Ocean Circulation Model. Part I: System Design and Performance

### 4.1 Abstract

The ocean state off Oregon-Washington, U.S. West coast, is highly variable in time. Under these conditions the assumption made in traditional 4DVAR that the prior model (background) error covariance is the same in every data assimilation (DA) window can be limiting. The ensemble 4DVAR (En4DVar) DA system has been developed in which the background error covariance is estimated from an ensemble of model runs and is thus time-varying. This part describes the details of this En4DVar system and ensemble statistics verification tests. The main run and 39 ensemble members are forced by perturbed wind fields, which helps maintain the ensemble spread. Error estimates in the wind forcing are obtained using scatterometer observations and a Bayesian Hierarchical Model. DA corrections to the main run and ensemble members are calculated in parallel by the newly developed, Cluster search minimization method. For the realistic coastal ocean application, this method can generate a 30% wall time reduction compared to the restricted B-conjugate gradient method. En4DVar is effective reducing the forecast error in surface temperature, sea-surface height and daily-averaged surface velocity fields compared to a model without DA.

### 4.2 Introduction

Data assimilation (DA) is a procedure, e.g., used in meteorology and oceanography, in which the output of a numerical model is combined with observations to find the most-likely estimate for the true state of the system. To do so, data assimilation algorithms require specification of the error statistics for the model and the observa-

tions. Traditionally, these statistics are assumed to follow a multidimensional normal distribution with zero mean and with a covariance that is static in time, i.e., the same from one assimilation window to the next. An example of such a data-assimilative system is the Oregon State University coastal ocean forecast system for an area offshore of Oregon-Washington (OR-WA), U.S. West coast (*Kurapov et al.*, 2011; *Yu et al.*, 2012). This system applies the 4DVAR algorithm in a series of consecutive 3-day windows. The initial conditions at the beginning of each window are corrected to yield the nonlinear analysis that best fits observations in this window. The simulation is then continued for another 3 days to provide the forecast. Summer dynamics in this region are dominated by the wind-forced upwelling and relaxation and the outflow of the Columbia River (*Hickey et al.*, 2005, 2010; *Huyer*, 1983; *Liu et al.*, 2009). In such a dynamic environment it is unlikely that the stationarity assumption on the error statistics holds. In particular, the temperature-salinity model error covariance will strongly depend on the presence of the river plume.

Over the past three decades, ensemble methods have been developed in meteorology to deal with non-stationarity in the error statistics. In these methods, the covariance for the forecast errors, the background error covariance, is estimated from an ensemble of perturbed model runs. One of the earliest examples of such a method is the ensemble Kalman filter (*Anderson*, 2001; *Bishop et al.*, 2001; *Evensen*, 1994; *Lermusiaux and Robinson*, 1999). Contrary to 4DVAR, the ensemble Kalman filter does not take into account the time-evolution of the background error within the window. Taking this into account has been shown to be beneficial to the accuracy of the forecast (*Lorenc and Rawlins*, 2005; *Lorenc and Jardak*, 2018). Generalization of the ensemble methodology to ensembles of 4DVAR (En4DVar) is nontrivial for three reasons. First, 4DVAR is computationally intensive. In 4DVAR calculation of the DA correction requires minimization of a cost function, or equivalently, solving a linear system with some symmetric and positive definite matrix **A**. This matrix is large and generally not available in explicit form. Instead only the algorithm for a product of **A** and a vector is at hand. An iterative method, e.g., the conjugate gradient algorithm, is used to find an approximate solution of this system. Each iteration requires a

time-consuming forward and backward propagation of the linearized model over the analysis period using the tangent linear (TL) model and its adjoint (ADJ). Second, En4DVar only compounds to this problem as it requires running the computationally intensive 4DVAR algorithm separately for each ensemble member. Third, the ensemble has to be constructed such that its covariance is representative of the background error covariance. For a long time, En4DVar methods were considered impractical for operational forecasting systems.

Recently, solutions have become available to overcome some of those challenges. Instead of applying full 4DVAR to each ensemble member it has become customary to calculate a low-rank approximation to  $\mathbf{A}$ , using e.g. Ritz pairs found by the Lanczos algorithm (*Trefethen and Bau*, 1997). In the Ensemble-Variational Integrated Localized Data Assimilation (EVIL) methodology (*Auligné et al.*, 2016) minimization of the cost function is only carried out for the main, or deterministic, model run. From the Ritz vectors obtained as a by-product from this minimization, a low-rank approximation of  $\mathbf{A}$  is constructed. The inverse of this approximation is then used to solve the linear system for the ensemble members. *Desroziers and Berre* (2012) follow a more conventional approach and apply 4DVAR to each ensemble member to calculate the first DA correction and use the Ritz pairs found during the first DA correction as a preconditioner at later times. *Lorenc et al.* (2017) uses the Mean-Pert method in which 4DVAR is first applied to the ensemble mean followed by application of 4DVAR to the perturbations of the ensemble members from the mean. In this case, the Ritz pairs that were found during the application of 4DVAR to the ensemble mean are reused as preconditioner when the DA correction to the perturbations is calculated.

Advances have also been made to speed up the 4DVAR minimization algorithm. Parallelization is a particularly convenient way to achieve wall time reduction as a parallel computational infrastructure already has to be present to generate the forecasts for the different ensemble members in En4DVar. Parallelization can be applied to the TL and ADJ models by assigning the calculations for different parts of the domain to different processor cores, as well as to the minimization algorithm. The former is currently standard practice, while the latter is still an area of active

research. One currently popular approach is to use an ensemble of concurrently produced nonlinear model runs to generate approximations of the TL and ADJ model. Examples of this approach are 4DEnVar (*Amezcua et al.*, 2017; *Desroziers et al.*, 2014; *Gustafsson and Bojarova*, 2014; *Liu et al.*, 2008; *Tian et al.*, 2017) and the Ensemble Kalman Smoother-4DVAR (*Mandel et al.*, 2016). Localization of these methods is non-trivial. In absence of a TL model to propagate the localized background error covariance forward in time, localization is often assumed to be static in these methods. For limited-size ensembles with non-dense observation networks this can lead to a decrease in forecast performance compared to variational methods that do use TL and ADJ models (*Poterjoy and Zhang*, 2015; *Poterjoy et al.*, 2016). A different approach to parallelization that does not suffer from this problem is taken by *Rao and Sandu* (2016) and *Fisher and Gürol* (2017). They retain the TL and ADJ model but parallelize the 4DVAR minimization algorithm in time. This is done by first dividing the analysis period into separate time intervals. The DA correction is then found by minimizing a cost function that consists of the sum of the 4DVAR cost functions for each interval plus an additional term representing the constraint that the correction should be continuous going from one interval of the another.

An alternative approach, explored in this chapter, is to try to solve the linear system using several search directions in parallel. The current OR-WA system uses the restricted B-conjugate gradient (RBCG) method (*Gürol et al.*, 2014). In this method, an approximation to the solution of the linear system is sought in the low-dimensional Krylov space. In RBCG the dimensionality of the search space is equal to the number of iterations. Several generic iterative solvers have already been developed in which the dimensionality of the search space is higher than this. Consequently, less iterations are necessary to produce a good approximation to the solution. Among these are the Enlarged Krylov space method (*Grigori et al.*, 2016) in which the search space is expanded by multiple directions per iteration simultaneously or the Augmented Krylov space methods where extra search directions are added to the system coming either from an earlier attempt to solve a similar system (*Erhel and Guyomarc'h*, 2000; *Morgan*, 1995), from the eigenvectors of a preconditioner (*Kharchenko and Yeremin*,

1995), or from an attempt to solve the system with a different initial residual (*Chapman and Saad*, 1996). Additional search directions can also lie outside the Krylov subspace. E.g. *Yaremchuk et al.* (2017) uses model-based Empirical Orthogonal Functions (EOFs) to create search directions. Once a general search space is defined, the approximation can be defined as the vector in the search space that has the smallest distance to the exact solution in some appropriately chosen norm (*Brezinski*, 1999).

In this chapter, these latest insights are utilized to construct an En4DVar system and apply it in the OR-WA coastal ocean forecast system. Here we describe two principal, and novel elements of our En4DVar. The first is the cluster search method, used to construct an enlarged search space at each iterations at a price of running a relatively small number of TL-ADJ applications in parallel. The second is the use of a Bayesian Hierarchical Model to estimate the magnitude of the wind forcing perturbations for the ensemble members. Chapter 5 will include a comparison of the ensemble background error covariances produced by the En4DVar system with the current static background error covariance as well as an evaluation of the ability of both systems to provide accurate forecasts.

This chapter is organized as follows: section 4.3 describes the experimental setup and the layout of the En4DVar DA system. Derivation of the cluster search method is presented in section 4.4. The generation of the wind perturbations for the ensemble members is discussed in section 4.5. In section 4.6 it is checked whether the En4DVar statistics are representative of the background errors. A conclusion whether or not the En4DVAR is suitable for DA in the Oregon-Washington region can be found in section 4.7 together with a discussion on how the findings for this En4DVar system compare to other ensemble-variational systems.

### 4.3 En4DVar System

The En4DVar system is tested here using the Oregon-Washington (OR-WA) Coastal Ocean Forecast System. Chapter 2 describes the system, implemented as standard 4DVAR with a static background error covariance, in every detail. Only a short summary is provided here. The nonlinear model dynamics are described by the Regional

Ocean Modeling System (ROMS, [www.myroms.org](http://www.myroms.org)), integrating three-dimensional, fully nonlinear, hydrostatic Boussinesq equations. The model domain is shown in Figure 4.1. The model resolution is approximately 2 km in the horizontal and 40 terrain-following layers in the vertical direction. The computational grid has  $310 \times 522$  points. Non-tidal boundary conditions are taken from the global  $1/12^\circ$  Hycom-NCODA analyses (*COAPS*, 2015). Tidal forcing is added along the open boundaries using tidal components from the TPXO tidal model (*Egbert and Erofeeva*, 2002, 2010). The atmospheric surface forcing is calculated based on the bulk flux algorithm (*Fairall et al.*, 2003) using the 12-km resolution Northern America Mesoscale model (NAM) analysis fields (*NCEP*, 2011). Fresh water discharge from the Columbia River, Fraser River, and 15 small rivers in Puget Sound is also included.

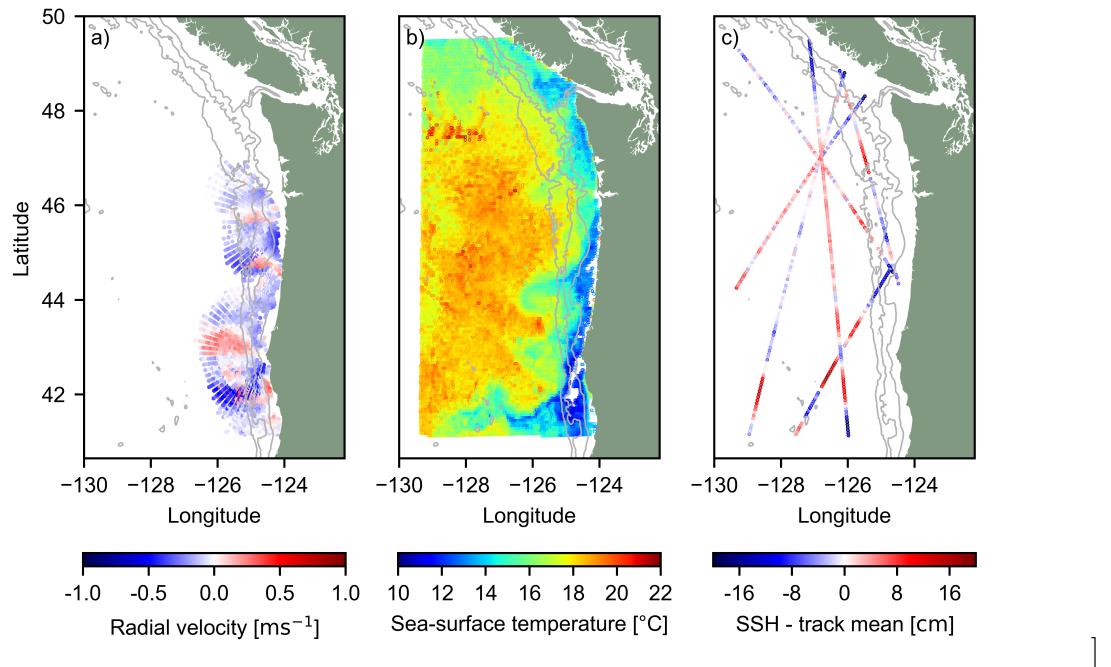


Figure 4.1: Location of (a) HFR, (b) SST and (c) SSH observations in the model for the window starting on 26 August 2011.

While the standard ROMS code includes TL and ADJ models, these are tightly integrated into the code such that implementing the En4DVar directly into ROMS

was too challenging for us as the users. Instead, we utilize the stand-alone TL and ADJ AVRORA codes developed in-house (*Kurapov et al.*, 2009, 2011; *Yu et al.*, 2012) and integrate these with the nonlinear ROMS and other components of the En4DVAR via Linux shell scripts, similarly to how it is done in the present OR-WA operational forecast system. The TL and ADJ runs are performed on a coarser, 4-km resolution model grid and their output is interpolated to and from the 2-km model grid.

All computations are carried out on the COMET cluster with computer allocations made available through the XSEDE framework (*Towns et al.*, 2014). The ROMS and AVRORA TL and ADJ codes are run using message passing interface (MPI) parallelization. In this parallelization the model grid is divided into horizontal tiles and computation in the interior of each tile is performed on a separate processor core. Owing to a relatively small grid size, a small number of  $N_{cores} = 6$  tiles are used for each instance of the ADJ and TL model. More nodes are available to us, and in this chapter we discuss how these can be used to speed up the iterative minimization algorithm used to find the DA corrections by exploring several search directions at every iteration.

In this chapter results from two experiments are compared. In experiment Ens, the En4DVar system is used to run  $M = 40$  instances of the model forward in time. In the discussions below the run with index  $m = 0$  is referred to as the main run, while instances  $m = 1, 2, \dots, M - 1$  are referred to as the ensemble members. Only the ensemble members ( $m > 0$ ) are used to compute the ensemble background error covariance. Since the dynamics in this region are dominated by the wind forcing, we assume that the errors in the wind velocity are the dominant forecast error source. To include this error into the error statistics, the nonlinear forecasts for each ensemble member are run with perturbed wind velocities as described in section 4.5. It will be evaluated in section 4.6 whether the addition of these perturbations alleviates the need for the customary ensemble inflation (*Anderson*, 2001; *Anderson and Anderson*, 1999; *Hamill et al.*, 2001). No wind perturbations are added to the main run and the main run is not utilized in the background error covariance computation. DA results are compared to the results of the free-running model without DA, referred

to as experiment No DA. The latter is spun-up starting on 27 February 2011. The ensemble members are all initialized from the no-DA model output on 10 March 2011 and are propagated forward in time without DA, using the perturbed winds, until 19 April 2011. Both Ens and No DA cases are then run over the period from 19 April 2011 to 1 October 2011.

The set of observations for assimilation includes daily-averaged surface radial currents from high-frequency radars (HFR), alongtrack altimetry, and satellite SST. The data and their associated observational error variances have been described in chapter 2 of this thesis; the only difference here is that the level 2 GOES-11 SST (*Maturi et al.*, 2008) is used instead of the polar-orbiting satellite products. The observational error covariance matrix  $\mathbf{R}$  is diagonal, with the variances estimated as described in chapter 2.

As is schematically shown in Figure 4.2, DA proceeds as a series of 3-day windows. At the beginning of each window, the initial conditions for the main run and each ensemble member are corrected. Then every model starts from the corrected initial conditions and is run forward using the nonlinear ROMS for the length of the assimilation window (3 days) to obtain the analyses and continues for another 3 days to obtain the forecasts.

The prior, or background, initial conditions include the temperature, salinity, sea surface height and horizontal velocity values at the points of the three-dimensional grid at the beginning of the assimilation window. To explain the DA method, we write them as a vector of real numbers of length  $N$ :  $\mathbf{x}_b^{(m)} \in \mathbb{R}^N$ ,  $m = 0, \dots, M - 1$ . The vector containing all observations within the window is denoted as  $\mathbf{y} \in \mathbb{R}^D$ . The innovation vector for each ensemble member is defined as  $\mathbf{d}^{(m)} = \mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(m)}) + \boldsymbol{\epsilon}^{(m)}$ . Here  $\mathcal{M}(\mathbf{x}_b^{(m)})$  is the nonlinear model trajectory started from the initial conditions  $\mathbf{x}_b^{(m)}$  and  $\mathbf{H}$  is the collection of data operators. A perturbation  $\boldsymbol{\epsilon}^{(m)}$  is added to the innovation vector for each ensemble member ( $m = 1, \dots, M - 1$ ). It is drawn from a normal distribution with zero mean and covariance  $\mathbf{R}$  and is added to account for the uncertainty in the analysis introduced by the presence of observational errors (*Burgers et al.*, 1998; *Houtekamer and Mitchell*, 1998).

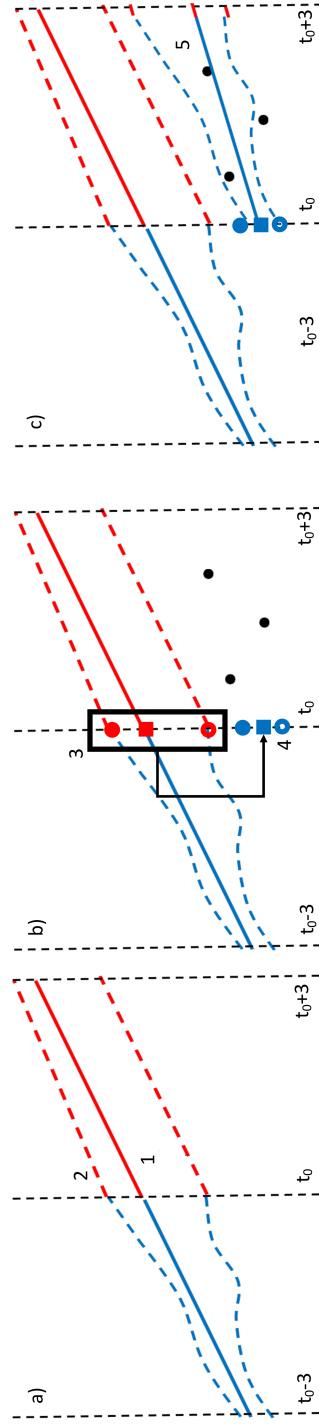


Figure 4.2: Overview of the En4DVar system. Panels a, b, and c show the progression of tasks. In (a), the main run (task 1, solid line) and ensemble members (with the envelope shown as dashed lines, task 2) are run for six days. The first three days are the analyses (blue), the last three days are the forecasts (red). In (b), task 3 is **B** calculation from the ensemble, and task 4 is the calculation of the DA corrections for the main run and the ensemble members using information from the observations (black circles). In (c), task 5 are the new six day model runs started from the corrected ocean states.

The DA correction to the background state  $\mathbf{x}_b^{(m)}$  is denoted as  $\mathbf{x}^{(m)} \in \mathbb{R}^N$ . It is found by minimizing the following cost function for each  $m$  (*Courtier et al.*, 1994):

$$J(\mathbf{x}^{(m)}) = \frac{1}{2}\mathbf{x}^{(m)T}\mathbf{B}^{-1}\mathbf{x}^{(m)} + \frac{1}{2}(\mathbf{d}^{(m)} - \mathbf{H}\mathbf{M}^{(m)}\mathbf{x}^{(m)})^T\mathbf{R}^{-1}(\mathbf{d}^{(m)} - \mathbf{H}\mathbf{M}^{(m)}\mathbf{x}^{(m)}). \quad (4.1)$$

Here,  $\mathbf{M}^{(m)}$  is the TL model, linearized with respect to  $\mathcal{M}(\mathbf{x}_b^{(m)})$ .  $\mathbf{M}^{(m),T}$  will be the ADJ model.  $\mathbf{B}$  is the background error covariance obtained as the sample covariance of the ensemble members with localization as described in chapter 3. To ensure that  $\mathbf{B}$  represents dynamics on relatively slow, subtidal and subinertial temporal scales, each ensemble member is time-averaged over the 24h time interval centered at the beginning of the DA window using the last 12h of the previous analysis and first 12h of the previous forecast. These 24h-averaged fields are used in the sample covariance computation.

The minimizer of (4.1) is sought as a solution of a linear, symmetric and positive-definite system of equations that can take different forms, e.g., depending on whether the solution is sought in the space of size  $N$  or  $D$ , and on how the system is preconditioned. The restricted B-preconditioned conjugate gradient method (RBCG) proved to be an efficient solver in the data space of dimension  $D$  with good convergence (*Gürol et al.*, 2014). It finds approximations of  $\mathbf{x}^{(m)}$  that minimize the cost-function in (4.1) by solving the system

$$(\mathbf{I} + \mathbf{R}^{-1/2}\mathbf{H}\mathbf{M}^{(m)}\mathbf{B}\mathbf{M}^{(m),T}\mathbf{H}^T\mathbf{R}^{-1/2})\hat{\mathbf{x}}^{(m)} \stackrel{\text{def}}{=} \hat{\mathbf{A}}^{(m)}\hat{\mathbf{x}}^{(m)} = \hat{\mathbf{d}}^{(m)} \quad (4.2)$$

where  $\mathbf{x}^{(m)} = \mathbf{B}\mathbf{M}^{(m),T}\mathbf{R}^{-1/2}\hat{\mathbf{x}}^{(m)}$  and  $\hat{\mathbf{d}}^{(m)} = \mathbf{R}^{-1/2}\mathbf{d}^{(m)}$ .

#### 4.4 Cost Function Minimization

In this section we discuss several approaches to finding an approximation to  $\hat{\mathbf{x}}^{(m)}$ , the solution of (4.2), and propose the new, computationally efficient cluster search algorithm.

In RBCG,  $\hat{\mathbf{x}}_i^{(m)}$ , the  $i$ -th iteration approximation to  $\hat{\mathbf{x}}^{(m)}$ , is sought in the low-

dimensional Krylov subspace  $\mathcal{K}_i(\mathbf{d}^{(m)}, \hat{\mathbf{A}}^{(m)})$ , where

$\mathcal{K}_i(\mathbf{z}, \mathbf{A}) = \text{span}(\mathbf{z}, \mathbf{Az}, \mathbf{A}^2\mathbf{z}, \dots, \mathbf{A}^{i-1}\mathbf{z})$ . This search space grows by one dimension per iteration. Let  $I'$  be the number of iterations necessary to bring  $\hat{\mathbf{x}}_i^{(m)}$  within a certain prescribed error margin of  $\hat{\mathbf{x}}^{(m)}$ . Then obtaining  $\hat{\mathbf{x}}_{I'}^{(m)}$  for each  $m$  can require a considerable amount of wall time as well as computational resources. Indeed, every iteration requires the multiplication of  $\hat{\mathbf{A}}^{(m)}$  with a vector. This demands that for each  $m$  a single implementation of the ADJ model over the analysis window is run, followed by application of the background error covariance, and then a TL model application. To carry this out, a total of  $N_{cores} \times M$  processor cores need to be available in parallel (see section 4.4.1).

Faster convergence to the exact solution  $\hat{\mathbf{x}}^{(m)}$ , for each  $m$ , could be achieved by expanding the space in which  $\hat{\mathbf{x}}_i^{(m)}$  is sought with vectors that lie outside  $\mathcal{K}_i(\mathbf{d}^{(m)}, \hat{\mathbf{A}}^{(m)})$ . Such vectors can be generated at no extra computational cost if we, similarly to *Auligné et al.* (2016), make the assumption that

$$\mathbf{M}^{(m)} \approx \mathbf{M}^{(0)} \stackrel{\text{def}}{=} \mathbf{M}, \quad (4.3)$$

and consequently  $\hat{\mathbf{A}}^{(m)} \approx \hat{\mathbf{A}}^{(0)} \stackrel{\text{def}}{=} \hat{\mathbf{A}}$ . Then the solution of (4.2) for different  $m$  can be combined into one system of equations. This will allow to approximate  $\hat{\mathbf{x}}^{(m)}$  with the same target accuracy in  $I < I'$  iterations using the block diagonal CG method (see section 4.4.2). Thanks to approximation (4.3), this requires the same  $N_{cores} \times M$  cores per iteration as an ensemble of regular RBCGs. Given presently available resources, this method is feasible with our relatively small OR-WA system, but it can become prohibitively expensive for larger forecast systems requiring  $N_{cores} = O(10^3)$  (e.g. *Kurapov et al.*, 2017). For these systems, the new cluster search method is introduced (see section 4.4.3). It also depends on the assumption (4.3) and involves  $N_s$  new direction searches at every iteration, where  $1 \leq N_s \ll M$ . These new search directions are generated in parallel, requiring  $N_{cores} \times N_s$  cores to be available simultaneously. It serves as a compromise between RBCG and block diagonal CG. Sections 4.4.1-4.4.3 provide a detailed description of each algorithm.

#### 4.4.1 RBCG

Using RBCG (*Gürol et al.*, 2014)  $\hat{\mathbf{x}}_i^{(m)}$  can be given as the vector that satisfies

$$\hat{\mathbf{x}}_i^{(m)} \in \mathcal{K}_i(\hat{\mathbf{d}}^{(m)}, \hat{\mathbf{A}}^{(m)}) : \|\hat{\mathbf{x}}_i^{(m)} - \hat{\mathbf{x}}^{(m)}\|_E \leq \|\hat{\mathbf{w}} - \hat{\mathbf{x}}^{(m)}\|_E \quad \forall \hat{\mathbf{w}} \in \mathcal{K}_i(\hat{\mathbf{d}}^{(m)}, \hat{\mathbf{A}}^{(m)}) \quad (4.4)$$

I.e., the  $i$ -th approximation to  $\hat{\mathbf{x}}^{(m)}$  can be found as a linear combination of the vectors spanning the  $i$ -th Krylov space that minimizes the solution error in the  $E$ -norm  $\|\mathbf{w}\|_E = (\mathbf{w}^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \mathbf{w})^{1/2}$  with  $\hat{\mathbf{B}}^{(m)} = \hat{\mathbf{A}}^{(m)} - \mathbf{I} = \mathbf{R}^{-1/2} \mathbf{H} \mathbf{M}^{(m)} \mathbf{B} \mathbf{M}^{(m),T} \mathbf{H}^T \mathbf{R}^{-1/2}$ . Then  $\hat{\mathbf{x}}_i^{(m)}$  is uniquely determined as the  $E$ -projection of  $\hat{\mathbf{x}}^{(m)}$  on  $\mathcal{K}_i(\hat{\mathbf{d}}^{(m)}, \hat{\mathbf{A}}^{(m)})$ :

$$\hat{\mathbf{x}}_i^{(m)} = \hat{\mathbf{V}}_i (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \hat{\mathbf{x}}^{(m)} = \hat{\mathbf{V}}_i (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{d}}^{(m)} \quad (4.5)$$

or alternatively,

$$\hat{\mathbf{x}}_i^{(m)} = \hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{d}}^{(m)} \quad (4.6)$$

where the column space of  $\hat{\mathbf{V}}_i$  is equal to  $\mathcal{K}_i(\hat{\mathbf{d}}^{(m)}, \hat{\mathbf{A}}^{(m)})$  and  $\mathbf{T}_i = (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{A}}^{(m)} \hat{\mathbf{V}}_i$ . Here  $\hat{\mathbf{V}}_i$  and  $\mathbf{T}_i$  depend on  $m$  via  $\hat{\mathbf{A}}^{(m)}$  and  $\hat{\mathbf{d}}^{(m)}$ . For the resulting residual  $\hat{\mathbf{r}}_i^{(m)} = \hat{\mathbf{d}}^{(m)} - \hat{\mathbf{A}}^{(m)} \hat{\mathbf{x}}_i^{(m)}$  we get, after inserting (4.5) for  $\hat{\mathbf{x}}_i^{(m)}$ ,

$$\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{r}}_i^{(m)} = \hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} (\hat{\mathbf{d}}^{(m)} - \hat{\mathbf{A}}^{(m)} \hat{\mathbf{x}}_i^{(m)}) = \mathbf{0} \quad (4.7)$$

and we set  $\hat{\mathbf{r}}_0^{(m)} = \hat{\mathbf{d}}^{(m)}$ . I.e. the  $i$ -th residual is by construction  $\hat{\mathbf{B}}^{(m)}$ -orthogonal to  $\hat{\mathbf{V}}_i$ . We refer to the column vectors of  $\hat{\mathbf{V}}_i$  as the search directions. As  $\hat{\mathbf{x}}_i^{(m)}$  is a projection, it is independent of the search directions chosen as long as they span the same space. In RBCG  $i+1$ -th search direction would be chosen to be  $E$ -orthogonal, i.e. conjugate, to  $\hat{\mathbf{V}}_i$ . Here the  $i+1$ -th search direction is chosen to be equal to  $\hat{\mathbf{r}}_i^{(m)}$  which according to (4.7) is  $\hat{\mathbf{B}}^{(m)}$ -orthogonal to  $\hat{\mathbf{V}}_i$ . I.e.  $\hat{\mathbf{V}}_{i+1} = [\hat{\mathbf{V}}_i, \hat{\mathbf{r}}_i^{(m)}]$ . In this case, the E-orthonormalization is contained in  $\mathbf{T}^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}}^{(m)} \hat{\mathbf{V}}_i)^{-1}$  in (4.6). The pseudo-code for this method is included in Table C.1.

#### 4.4.2 Full Parallelization: Block Diagonal Conjugate Gradient Method

If (4.3) is assumed then  $\hat{\mathbf{A}}^{(m)} = \hat{\mathbf{A}}$ ,  $\hat{\mathbf{B}}^{(m)} = \hat{\mathbf{B}}$  and (4.2) for the different  $m$  can be combined into the linear system

$$\hat{\mathbf{A}}\hat{\mathbf{X}} = \hat{\mathbf{D}}, \quad (4.8)$$

where  $\hat{\mathbf{X}} = [\hat{\mathbf{x}}^{(0)}, \hat{\mathbf{x}}^{(1)}, \dots, \hat{\mathbf{x}}^{(M-1)}]$  and  $\hat{\mathbf{D}} = [\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \dots, \hat{\mathbf{d}}^{(M-1)}] \in \mathbb{R}^{D \times M}$ . Similar to (4.6), the  $i$ -th approximation  $\hat{\mathbf{X}}_i$  can be found as

$$\hat{\mathbf{X}}_i = \hat{\mathbf{V}}_i (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{X}} = \hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{D}} \quad (4.9)$$

where  $\mathbf{T}_i = (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i$ .  $\hat{\mathbf{V}}_i = \hat{\mathbf{D}}$  if  $i = 1$  and  $\hat{\mathbf{V}}_i = [\hat{\mathbf{V}}_{i-1}, \hat{\mathbf{D}} - \hat{\mathbf{A}}\hat{\mathbf{X}}_i]$  if  $i > 1$ .  $\mathbf{T}$ ,  $\hat{\mathbf{V}}_i$  are independent of  $m$ . The column space of  $\hat{\mathbf{V}}_i$  is now  $\mathcal{K}_i(\hat{\mathbf{D}}, \hat{\mathbf{A}})$ . The advantage here, compared to RBCG, is that the search space for each  $\hat{\mathbf{x}}_i^{(m)}$ , spanned by  $\hat{\mathbf{V}}_i$ , now has dimension  $i \times M$  instead of  $i$ . The method results in matrices  $\hat{\mathbf{V}}_i$  that are now no longer  $\hat{\mathbf{B}}$ -orthogonal, but  $\hat{\mathbf{B}}$ -block orthogonal: if  $\hat{\mathbf{v}}_p$  and  $\hat{\mathbf{v}}_q$  are two columns of  $\hat{\mathbf{V}}_i$  then  $\hat{\mathbf{v}}_p^T \hat{\mathbf{B}} \hat{\mathbf{v}}_q = 0$  if  $|p - q| \geq M$ , but might be non-zero otherwise. The estimates  $\hat{\mathbf{X}}_i$  retrieved in this way are the same as those found using the block diagonal CG method (O'Leary, 1980) with B-preconditioning. The pseudo-code for the block diagonal CG is given in Table C.2.

#### 4.4.3 Partial Parallelization: Cluster Search Method

In order to expand  $\hat{\mathbf{V}}_{i-1}$  to  $\hat{\mathbf{V}}_i$  in the block diagonal CG method,  $M$  applications of  $\hat{\mathbf{A}}$  to a vector are necessary. This will require  $N_{cores} \times M$  cores and can be prohibitively expensive for large systems. Here we introduce the cluster search method that requires  $N_s \ll M$  concurrent applications of  $\hat{\mathbf{A}}$  to create the expansion. In this case, we still look for a solution to (4.8) with  $\hat{\mathbf{X}}_i$  still given by (4.9) but with  $\hat{\mathbf{V}}_i$  constructed differently. In particular, we focus on  $\hat{\mathbf{x}}^{(0)}$  as it is the main run that will be used to produce the operational forecasts and therefore minimization of the error in  $\hat{\mathbf{x}}_i^{(0)}$  has top priority. To explain how  $\hat{\mathbf{V}}_i$  is constructed, we momentarily assume that the eigendecomposition  $\hat{\mathbf{A}} = \hat{\mathbf{U}}_0 \hat{\Lambda}^2 \hat{\mathbf{U}}_0^T$  with  $\hat{\mathbf{U}}_0^T \hat{\mathbf{U}}_0 = \mathbf{I}$  is available and require that: (i)

$\hat{\mathbf{v}}_p^T \hat{\mathbf{B}} \hat{\mathbf{v}}_q = 0$  if  $|p - q| \geq N_s$  similar to block diagonal CG and (ii) the residual for the main run,  $\hat{\mathbf{r}}_i^{(0)}$ , is in the column space of  $\hat{\mathbf{V}}_{i+1}$  as is the case in RBCG for  $m = 0$ .

Define  $\hat{\mathbf{v}}' = \hat{\mathbf{U}}_0^T \hat{\mathbf{r}}_i^{(0)} = \hat{\mathbf{U}}_0^T \hat{\mathbf{A}}(\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_i^{(0)})$ ,  $\hat{\mathbf{v}}'' = \hat{\mathbf{U}}_0^T \hat{\mathbf{d}}^{(0)} = \hat{\mathbf{U}}_0^T \hat{\mathbf{A}}\hat{\mathbf{x}}^{(0)}$  and  $\hat{\mathbf{U}}_i = \hat{\mathbf{U}}_0 - \hat{\mathbf{A}}\hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0$ . Then

$$\hat{\mathbf{U}}_0 \hat{\mathbf{v}}' = \hat{\mathbf{r}}_i^{(0)} = \hat{\mathbf{r}}_0^{(0)} - \hat{\mathbf{A}}\hat{\mathbf{x}}_i^{(0)} = \hat{\mathbf{r}}_0^{(0)} - \hat{\mathbf{A}}\hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{r}}_0^{(0)} = \hat{\mathbf{U}}_i \hat{\mathbf{v}}'' \quad (4.10)$$

To expand  $\hat{\mathbf{V}}_i$  to  $\hat{\mathbf{V}}_{i+1}$  we look for  $N_s$  new search vectors of the form  $\mathbf{s}^{(n)} = \hat{\mathbf{U}}_i \hat{\Lambda}^2 \mathbf{P}_n \hat{\mathbf{v}}''$  with  $n = 1, 2, \dots, N_s$  and  $\mathbf{P}_n = \sum_{d \in D_n} \hat{\mathbf{e}}_d (\hat{\mathbf{e}}_d^T \hat{\mathbf{v}}'')$  with  $\hat{\mathbf{e}}_d$  the unit vector in direction  $d$ . Here  $D_n$  is a subset of  $\{1, 2, \dots, D\}$  such that the union of  $D_1, D_2, \dots, D_{N_s}$  is  $\{1, 2, \dots, D\}$  and  $D_p$  and  $D_q$  are disjoint if  $p \neq q$ . Consequently,  $\sum_{n=1}^{N_s} \mathbf{P}_n \hat{\mathbf{v}}'' = \hat{\mathbf{v}}''$ . Combined with the equality  $\hat{\mathbf{U}}_0 \hat{\mathbf{v}}' = \hat{\mathbf{U}}_i \hat{\mathbf{v}}''$  in (4.10) this then ensures that  $\hat{\mathbf{r}}_i^{(0)}$  lies within  $\hat{\mathbf{V}}_{i+1}$ . Thus search vectors of the form  $\mathbf{s}^{(n)}$  satisfy requirement (ii). Furthermore,

$$\begin{aligned} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_i &= \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0 - \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i \mathbf{T}_i^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0 \\ &= \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0 - \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i (\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{A}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{U}}_0 = \mathbf{0} \end{aligned} \quad (4.11)$$

This shows that  $\hat{\mathbf{U}}_i$  is  $\hat{\mathbf{B}}$ -orthogonal to  $\hat{\mathbf{V}}_i$  and since the  $N_s$  new search directions in  $\hat{\mathbf{V}}_{i+1}$  are linear combinations of the column vectors of  $\hat{\mathbf{U}}_i$ , they satisfy requirement (i).

For the following we also need to be able to estimate  $\hat{\mathbf{U}}_0^T \hat{\mathbf{A}} \hat{\mathbf{U}}_i$ . For  $i = 0$ ,  $\hat{\mathbf{U}}_i = \hat{\mathbf{U}}_0$  and so  $\hat{\mathbf{U}}_0^T \hat{\mathbf{A}} \hat{\mathbf{U}}_0 = \Lambda^2$ , while for  $i > 0$  exact expressions are not directly available. Instead we observe that the columns of  $\hat{\mathbf{U}}_i$  are the residuals obtained after trying to find a solution in the search space  $\mathbf{V}_i$  to the linear system  $\hat{\mathbf{A}}\mathbf{x}' = \hat{\mathbf{U}}_0$ . This system has the exact solution  $\hat{\mathbf{U}}_0 \Lambda^{-2}$ . Here we make the ad-hoc assumption that these residuals are multiples of the columns of  $\hat{\mathbf{U}}_0$ , i.e.  $\hat{\mathbf{U}}_i \approx \hat{\mathbf{U}}_0 \Xi$  with  $\Xi$  diagonal. In this case,  $\Xi$  can be estimated as  $\Xi^2 = \Xi^T \hat{\mathbf{U}}_0^T \hat{\mathbf{U}}_0 \Xi \approx \hat{\mathbf{U}}_i^T \hat{\mathbf{U}}_i$ .

In RBCG  $\hat{\mathbf{x}}_i^{(0)}$  is defined as the vector in the search space  $\text{span}(\hat{\mathbf{V}}_i)$  that minimizes the error (4.4) in the E-norm. The idea behind cluster search is to find a clustering  $D_1, D_2, \dots, D_{N_s}$  and the associated  $N_s$  new search vectors  $\mathbf{s}^{(n)}$  such that the reduction of the expected error  $\|\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_{i+1}^{(0)}\|_E$  is larger than can be achieved using any other clustering. Using the properties of  $\mathbf{P}_n$ , the estimation  $\hat{\mathbf{U}}_i \approx \hat{\mathbf{U}}_0 \Xi_i$ , and the orthonor-

mality of  $\hat{\mathbf{U}}_0$ , we find that the E-norm of the expected error for  $\hat{\mathbf{x}}_{i+1}^{(0)}$  can be estimated as

$$\|\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_i^{(0)} - \sum_{n=1}^{N_s} \alpha_n \hat{\mathbf{U}}_i \mathbf{P}_n \hat{\mathbf{v}}''\|_E^2 \approx \sum_{n=1}^{N_s} \sum_{d \in D_n} [(1 - \lambda_d^{-2}) v_d'^2 - 2\alpha_n v_d' \lambda_d^2 (1 - \lambda_d^{-2}) v_d'' \xi_d + \alpha_n^2 \lambda_d^4 (1 - \lambda_d^{-2}) v_d''^2 \xi_d^2], \quad (4.12)$$

where  $\lambda_d$  and  $\xi_d$  are the  $d$ -th element on the diagonal of  $\Lambda$  and  $\Xi$ , correspondingly. To find the minimum of this function, we set the derivate of (4.12) as a function of  $\alpha_n$  to zero and get

$$\begin{aligned} \hat{\alpha}_n &= \left( \sum_{d \in D_n} (1 - \lambda_d^{-2}) \lambda_d^4 v_d''^2 \xi_d^2 \frac{v_d'}{v_d'' \xi_d \lambda_d^2} \right) \left( \sum_{d \in D_n} (1 - \lambda_d^{-2}) \lambda_d^4 v_d''^2 \xi_d^2 \right)^{-1} \\ &= \frac{v'}{v'' \xi_d} \frac{1}{\lambda_d^2}, \end{aligned} \quad (4.13)$$

where  $\cdot^n$  denotes the weighted mean over the cluster  $D_n$  with weights  $(1 - \lambda_d^{-2}) \lambda_d^4 v_d''^2 \xi_d^2$ . Inserting  $\hat{\alpha}_n$  from (4.13) back into (4.12) gives that for our guesses of  $\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N_s)}$  the error squared obtains a minimum

$$\begin{aligned} \|\hat{\mathbf{x}}^{(0)} - \hat{\mathbf{x}}_i^{(0)} - \sum_{n=1}^{N_s} \hat{\alpha}_n \hat{\mathbf{U}}_i \mathbf{P}_n \hat{\mathbf{v}}''\|_E^2 &\approx \sum_{n=1}^{N_s} \sum_{d \in D_n} (1 - \lambda_d^{-2}) v_d''^2 \xi_d^2 \lambda_d^4 \left[ \frac{v_d'^2}{v_d''^2 \xi_d^2 \lambda_d^4} - \hat{\alpha}_n^2 \right] \\ &= \sum_{n=1}^{N_s} W_n \text{var}_n \left( \frac{v_d'}{v_d'' \xi_d \lambda_d^2} \right) \end{aligned} \quad (4.15)$$

with  $\text{var}_n$  the weighted variance over cluster  $n$  and  $W_n = \sum_{d \in D_n} (1 - \lambda_d^{-2}) \lambda_d^4 v_d''^2 \xi_d^2$  the normalization coefficient for the  $n$ -th cluster. The K-means clustering algorithm (*MacQueen, 1967*) can now be used to find an approximate clustering that approximately minimizes (4.15). Once K-means produces a clustering,  $\mathbf{s}^{(n)} = \hat{\mathbf{U}}_i \hat{\Lambda}^2 \mathbf{P}_n \hat{\mathbf{v}}''$  are known and  $\hat{\mathbf{V}}_{i+1} = [\hat{\mathbf{V}}_i, \mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N_s)}]$ .

In reality the eigenvalue decomposition of  $\hat{\mathbf{A}}$  is not available. Instead it is used that if  $\mathbf{R}$  and  $\mathbf{B}$  are true estimates of the observational and background error covariance then  $\hat{\mathbf{A}} = \langle \hat{\mathbf{d}} \hat{\mathbf{d}}^T \rangle$  (*Bennett, 1992; Desroziers et al., 2005*). Approximations to the eigenvectors and eigenvalues of  $\hat{\mathbf{A}}$  are then found by calculating the eigenvalue

decomposition of  $\frac{1}{M}[\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \dots, \hat{\mathbf{d}}^{(M-1)}][\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \dots, \hat{\mathbf{d}}^{(M-1)}]^T = \frac{1}{M}\hat{\mathbf{D}}\hat{\mathbf{D}}^T \approx \langle \hat{\mathbf{d}}\hat{\mathbf{d}}^T \rangle$ . The pseudo-code for the cluster search method can be found in Table C.3.

Notice that if  $N_s = 1$  there is only one cluster and requirement (i) ensures that the new search direction is equal to  $\mathbf{r}_i^{(0)}$ . Consequently, the clustering method reverts to RBCG described in section 4.4.1 for  $m = 0$ . If  $N_s = M$ , each column vector of  $\hat{\mathbf{U}}_i$  constitutes its own cluster and hence the new search directions are multiples of the column vectors of  $\hat{\mathbf{U}}_i$ . By construction  $\hat{\mathbf{U}}_0 = \hat{\mathbf{D}}\mathbf{L}$  for some matrix  $\mathbf{L}$ , so  $\text{span}(\hat{\mathbf{V}}_1)$  in the block diagonal CG and cluster search method are equal in the case  $N_s = M$ . In this case

$$\hat{\mathbf{U}}_i = \hat{\mathbf{U}}_0 - \hat{\mathbf{A}}\hat{\mathbf{V}}_i\mathbf{T}_i^{-1}(\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{V}}_i)^{-1}\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{U}}_0 = (\hat{\mathbf{D}} - \hat{\mathbf{A}}\hat{\mathbf{V}}_i\mathbf{T}_i^{-1}(\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{V}}_i)^{-1}\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{D}})\mathbf{L} = (\hat{\mathbf{D}} - \hat{\mathbf{A}}\hat{\mathbf{X}}_i)\mathbf{L} \quad (4.16)$$

with  $(\hat{\mathbf{D}} - \hat{\mathbf{A}}\hat{\mathbf{X}}_i)$  the residuals obtained after the  $i$ -th block CG iteration. Hence if  $\text{span}(\hat{\mathbf{V}}_i)$  is equal in block diagonal CG and cluster search for  $i$  then by (4.16) and  $\hat{\mathbf{V}}_{i+1} = [\hat{\mathbf{V}}_i, \hat{\mathbf{D}} - \hat{\mathbf{A}}\hat{\mathbf{X}}_i] \text{span}(\hat{\mathbf{V}}_{i+1})$  is also equal in both methods.

## 4.5 Wind Perturbations

Surface currents, the strength of the coastal upwelling (*Halpern*, 1976; *Huyer*, 1983), and the location of the fresh water Columbia River plume (*Hickey et al.*, 1998, 2005; *Liu et al.*, 2009) all depend on the wind forcing. Therefore uncertainty in the wind forcing is expected to be the dominant source of model error. To account for the model error in the forecasts, we force the ensemble members ( $m = 1, \dots, M-1$ ) with perturbed wind fields, as described in this section. Later in this manuscript we will verify if adding the wind perturbations helps to avoid ensemble variance shrinking and thus alleviates the need for the ensemble inflation (*Hamill and Whitaker*, 2005; *Li et al.*, 2009; *Whitaker and Hamill*, 2002).

The perturbed wind fields for an ensemble members are generated as

$$\mathbf{w}(t) = \mathbf{w}_{NAM}(t) + \mathbf{w}_L(t) + \mathbf{w}_S(t) \quad (4.17)$$

with  $\mathbf{w}_{NAM}(t) \in \mathbb{R}^{2N_w}$  the vector containing the meridional and zonal wind velocity components from the NAM interpolated to the  $N_w$  ROMS model surface grid points. Fields  $\mathbf{w}_L(t)$  and  $\mathbf{w}_S(t)$  represent the large-scale and small-scale wind perturbations respectively.

For the large-scale perturbations, we use the empirical orthogonal function (EOF) decomposition of the series  $\mathbf{w}'_{NAM}(t) = \mathbf{w}_{NAM}(t) - \langle \mathbf{w}_{NAM} \rangle$ , where the winds are provided every 6 hr from 1 January 2011 00:00 to 31 December 2011 18:00 and  $\langle \mathbf{w}_{NAM} \rangle$  is the mean wind field over this period. After the EOF decomposition the NAM wind field can be written as

$$\mathbf{w}_{NAM}(t) = \langle \mathbf{w}_{NAM} \rangle + \sum_{i=1}^{N_{EOF}} \beta_{L,i}(t) \mathbf{w}_{EOF,i} + \mathbf{w}_\perp(t), \quad (4.18)$$

where  $\mathbf{w}_{EOF,i}$  is the EOF mode associated with the  $i$ -th largest singular value and  $\mathbf{w}_\perp(t)^T \mathbf{w}_{EOF,i} = 0$  for  $i = 1, 2, \dots, N_{EOF}$ . For illustration, the wind fields associated with the two dominant EOFs are shown in Figure 4.3a,b. Here, we use 10 EOFs ( $N_{EOF} = 10$ ) that explain 95% of the variance of  $\mathbf{w}_{NAM}$  in time. Similarly to *Hénaff et al.* (2009) and *Vervatis et al.* (2016), we define the large-scale wind perturbation to be

$$\mathbf{w}_L(t) = \sum_{i=1}^{N_{EOF}} \beta'_{L,i}(t) \mathbf{w}_{EOF,i}. \quad (4.19)$$

Coefficients  $\beta'_{L,i}$  are generated by an AR1-process

$$\beta'_{L,i}(t) = c_\beta \beta'_{L,i}(t - \Delta t) + \sqrt{1 - c_\beta^2} \hat{\sigma}_{EOF,i} \epsilon_{\beta,i}(t). \quad (4.20)$$

Here  $\epsilon_{\beta,i}(t)$  is drawn from a standard normal distribution,  $\Delta t = 6$  h is the output time step of the NAM model and correlation coefficient  $c_\beta = 0.4$  (*Milliff et al.*, 2011).

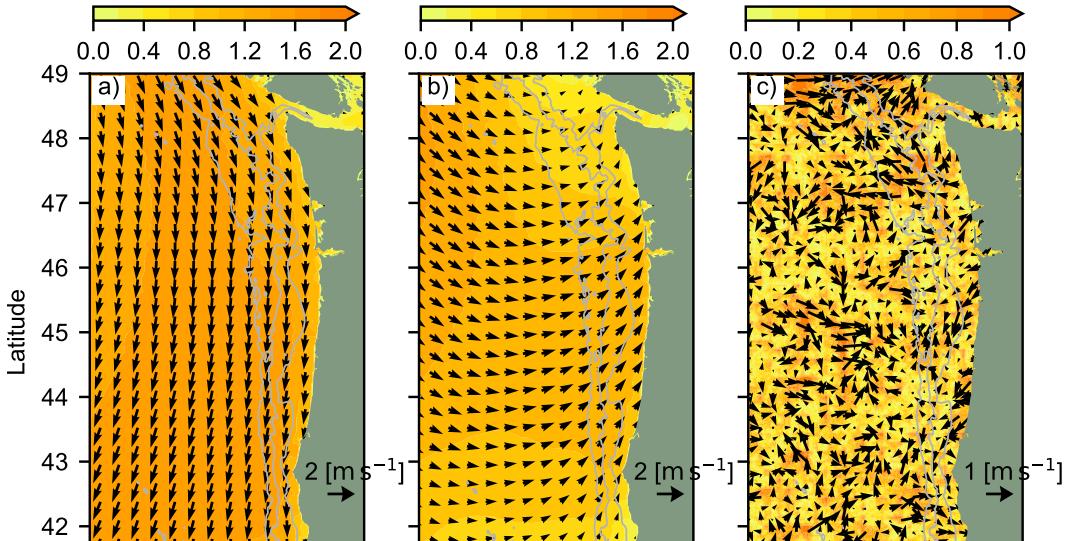


Figure 4.3: (a)  $\hat{\sigma}_{EOF,i} \mathbf{w}_{EOF,i}$ , first EOF of the wind field scaled by its standard deviation in the wind perturbations, (b)  $\hat{\sigma}_{EOF,i} \mathbf{w}_{EOF,i}$  the second EOF of the wind field scaled by its standard deviation in the wind perturbations, (c)  $\mathbf{w}_S$ , example of a small-scale wind error field for one time. Color scale shows the wind error speed of the different wind fields in  $m s^{-1}$ .

To estimate  $\hat{\sigma}_{EOF,i}$ , the standard deviation of the series  $\beta'_{L,i}(t)$ , a Bayesian Hierarchical Model (BHM) is used (see e.g. *Wikle et al.*, 2001; *Milliff et al.*, 2011). The probability distribution for  $\sigma^2_{EOF,i}$  conditioned on the NAM model wind fields, the 25 km ASCAT satellite wind product (*Figa-Saldaa et al.*, 2002; *STAR*, 2011) and NDBC buoy wind observations (*NDBC*, 2016) is estimated (see appendix D) and  $\hat{\sigma}^2_{EOF,i}$  is then defined as the mode of this distribution. The posterior distributions (D.3) for the dominant nine EOF modes are shown together with their a priori distributions (D.2) in Figure 4.4. Also indicated in Figure 4.4 is the percentage of the variance in the NAM wind fields explained by each EOF mode as well as 9% of this variance (dashed lines). This 9%-value is important as it is used in other studies (*Hénaff et al.*, 2009; *Vervatis et al.*, 2016) to estimate  $\hat{\sigma}_{EOF,i}$ . Figure 4.4 shows that the BHM estimate for  $\hat{\sigma}_{EOF,i}$  is higher than the 9% estimate in for all modes except mode one. The difference between the two estimates increases for increasing mode number. For

the higher EOF modes (mode 4 and higher), which represent smaller spatial scales in the wind field (not shown), the 9% estimate severely underestimates the contribution of the mode to the error in the wind fields.

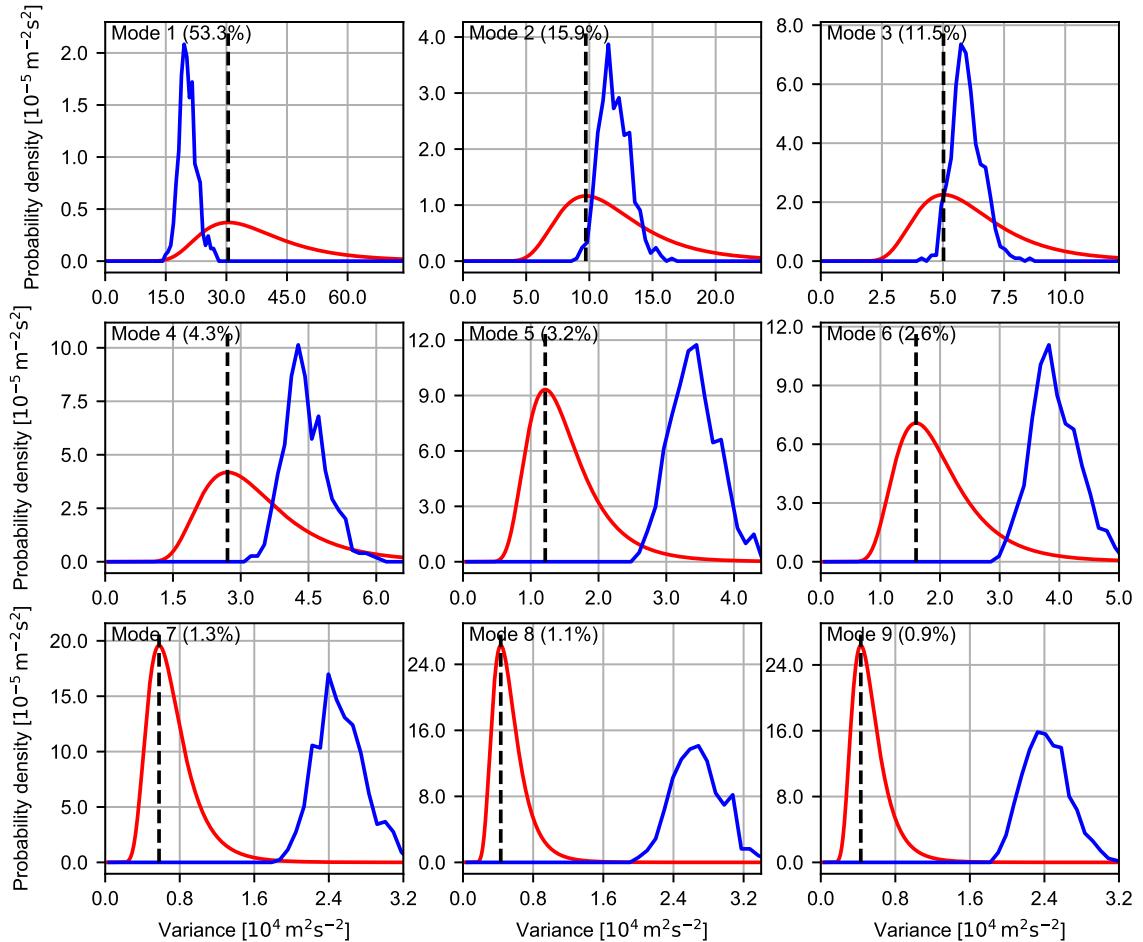


Figure 4.4: A priori (red) and posteriori (blue) probability density distributions for the variance of the EOF coefficients  $\sigma_{EOF,i}^2$ . The variance for which the blue distribution attains its maximum is used as estimate  $\hat{\sigma}_{EOF,i}^2$ . For reference, the time-variance explained by each mode is shown as percentage. Indicated by the dash line is 9% value of the time-variance.

The study of scatterometer wind measurements over the Pacific Ocean has shown that the power spectral density (PSD) of the wind field scales with  $\kappa^{-\hat{\gamma}}$  with  $\kappa$  the wave number and  $\hat{\gamma} \approx 2$  (*Chin et al.*, 1998). The PSD of the wind forcing used is shown in Figure 4.5. In this figure, the solid blue line is the meridional wind velocity PSD computed using the model fields on 8 August 2011 00:00 UTC. It is constructed by first collecting the meridional wind velocities from the 40 longitudes of the model grid that lie between 129°W and 128°W, followed by removal of the mean meridional wind velocity for each longitude, application of the Hamming window, calculation of the PSD for each longitude and averaging in the zonal direction. The power relationship in the power spectrum is correctly reproduced in the wind fields for  $\kappa < 0.3 \text{ rad km}^{-1}$ . As the NAM model has a  $12 \times 12$ -km resolution, smaller scales are absent from the interpolated wind fields and hence the PSD in Figure 4.5 drops below the power relationship for  $\kappa > 0.3 \text{ rad km}^{-1}$ .

As the NAM model cannot represent small-scale errors in the wind field, they are added separately. Following *Wikle et al.* (2001), it is assumed that the small-scale wind errors in (4.17) can be decomposed into Daubechies-2 wavelets (*Cohen et al.*, 1993):

$$\mathbf{w}_S(t) = \gamma_0 \sum_{n=1}^9 \sum_i \gamma_i^{(n)}(t) \psi_i^{(n)} \quad (4.21)$$

with  $\gamma_i^{(n)}(t)$  coming from an AR1-process

$$\gamma_i^{(n)}(t) = c_\gamma \gamma_i^{(n)}(t - \Delta t) + \sqrt{1 - c_\gamma^2} \sigma_\gamma^{(n)} \epsilon_i^{(n)}(t) \quad (4.22)$$

Here  $n$  indicates the level of the wavelet, with the length scale of the wavelets doubling as the level increases with one and  $i$  running over all the wavelets that are available at level  $n$ . Similarly to the large-scale wind field, we use  $c_\gamma = 0.4$ . The advantage of using this wavelet approach is that it yields small-scale wind perturbations that are local in space and are simultaneously constrained to a limited spectral band. By experimentation the standard deviation of  $\sigma_\gamma^{(n)}$  is chosen such that the dependence of the PSD  $S$  on the wave number  $\kappa$  of the total perturbed wind field scales as  $S(\kappa) \sim$

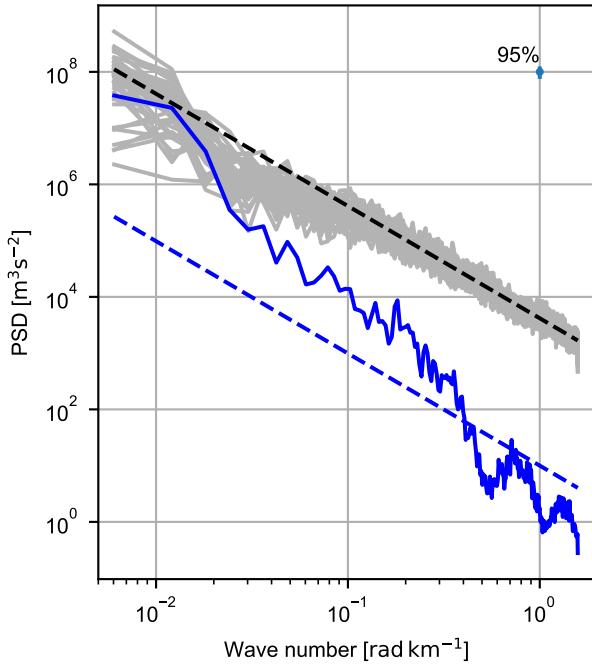


Figure 4.5: Power spectral density of the meridional wind field as a function of the meridional wave number from the interpolated NAM wind field (blue) and the perturbed wind fields of the ensemble members (grey) on 8 August 2011 00:00. Also shown are the fits to the mean of the ensemble power spectral densities (dashed black line) and the NAM power spectral density (dashed blue line) and the 95% confidence interval for the PSD estimates.

$\kappa^{-2}$ . This was achieved by setting  $\sigma_\gamma^{(n)} = \exp[1.3(n - 3)](0.5 - 0.5 \tanh[0.4(n - 4)])$  and picking  $\gamma_0$  such that the variance of the wind speed is  $(0.55 \text{ m s}^{-1})^2$ . The PSDs for the wind fields of the different ensemble members on 8 August 2011 00:00 are shown together with the linear least-square log-log fit to the ensemble mean PSD for  $\kappa > 0.1 \text{ rad km}^{-1}$  in Figure 4.5. The fit confirms that the PSDs have indeed the desired  $PSD \sim \kappa^{-2.0}$  relationship.

## 4.6 Results

In this section we analyze whether the ensemble statistics agree with the assumptions made during the construction of the assimilation system and whether the En4DAR system is indeed capable of delivering forecasts that are an improvement over case No DA.

### 4.6.1 Convergence

First, the effectiveness of the cluster search algorithm using a different number of clusters is compared with that of RBCG and block diagonal CG. Even though the DA correction  $\mathbf{x}^{(m)}$  is only calculated after the last inner loop iteration  $i = I$ , the cost function (4.1) can be calculated for each inner loop iteration  $i$  if the substitution  $\mathbf{x}_i^{(m)} = \mathbf{B}\mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1/2}\hat{\mathbf{x}}_i^{(m)}$  is made in (4.1):

$$J(\mathbf{x}_i^{(m)}) = \frac{1}{2}\hat{\mathbf{x}}_i^{(m),T}\hat{\mathbf{B}}\hat{\mathbf{x}}_i^{(m)} + \frac{1}{2}(\hat{\mathbf{d}}^{(m)} - \hat{\mathbf{B}}\hat{\mathbf{x}}_i^{(m)})^T(\hat{\mathbf{d}}^{(m)} - \hat{\mathbf{B}}\hat{\mathbf{x}}_i^{(m)}). \quad (4.23)$$

Using (4.5) we find that  $\hat{\mathbf{B}}\hat{\mathbf{x}}_i^{(m)} = \hat{\mathbf{B}}\hat{\mathbf{V}}_i(\mathbf{T}_i)^{-1}(\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{V}}_i)^{-1}\hat{\mathbf{V}}_i^T\hat{\mathbf{B}}\hat{\mathbf{d}}^{(m)}$  which is readily available as  $\hat{\mathbf{B}}\hat{\mathbf{V}}_i$  is stored. Using (4.23) the value of the cost function was calculated prior to each inner loop iteration using the RBCG method ( $N_s = 1$ ), block diagonal CG method ( $N_s = 40$ ) and cluster search method with different numbers of clusters ( $1 < N_s < 40$ ). Cost function values in the windows starting on 31 May and 26 August 2011 normalized by the cost function value at the start of the minimization are shown in Figure 4.6a,b. Increase in the number of clusters consistently increases the rate at which the cost function decreases as the function of the inner loop iteration number. To provide a quantitative assessment of the advantage of using several search directions in parallel, the number of iterations necessary to obtain a specified reduction of the cost function is calculated. As a reference, we take the value of the cost function after  $i = 12$  inner loop iterations using cluster search with  $N_s = 4$  (the dashed line in Figure 4.6a,b) as these are the settings used in experiment Ens. Figure 4.6c shows how many more iterations RBCG needs to reach this level compared

to cluster search with  $N_s$  clusters. E.g., RBCG needs  $\sim 30\%$  more iterations than cluster search with  $N_s = 4$  to reach the same amount of cost function reduction. A fit of a 2nd order polynomial to the ratios (dashed black lines in Figure 4.6c) shows that the coefficient for the quadratic term is negative and significantly different from zero at a 95% significance level, thus showing that as the number of clusters increases the additional speed-up obtained from adding another cluster diminishes.

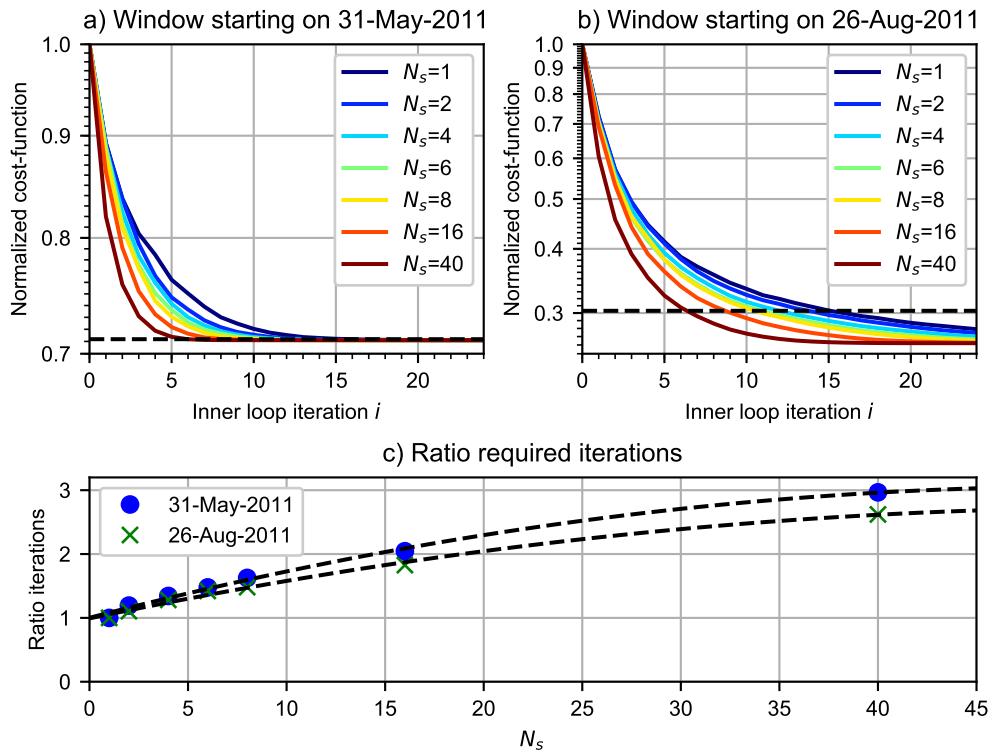


Figure 4.6: Value of the cost function as function of the number of inner loop iterations using different levels of parallelization  $N_s$  for the window starting (a) on 31 May 2011 and (b) on 26 August 2011. For reference, the value of the cost function after 12 inner loop iterations with  $N_s = 4$  is marked by a dashed black line. Panel c shows how many iterations RBCG need to reach this level compared to cluster search with  $N_s$  clusters.

To test to what extent termination of the minimization deteriorates the fit of the analysis to the observations, the error in the analysis predictions normalized by the observational error standard deviations, i.e.  $\mathbf{R}^{-1/2}(\mathbf{d}^{(m)} - \mathbf{HMBM}^T \mathbf{H}^T \mathbf{R}^{-1/2} \hat{\mathbf{x}}_i^{(m)}) = \hat{\mathbf{d}}^{(m)} - \hat{\mathbf{B}} \hat{\mathbf{x}}_i^{(m)}$ , is calculated as a function of the number of inner loop iterations. The normalized root-mean-square error (RMSE) for the three different types of observations used in the main run ( $m = 0$ ) on 26 August 2011 is shown in Figure 4.7. According to the figure, the normalized RMSE has not reached its minimum after  $I = 12$  iterations in any of the shown cases. The use of multiple clusters does however lead to reduction of the normalized RMSE: for the window shown in Figure 4.7, the normalized RMSE in SST, HFR and SSH predictions after  $I = 12$  inner loop iterations with  $N_s = 4$  is respectively 3.0%, 2.8% and 4.4% smaller than if RBCG ( $N_s = 1$ ) were to be used. For 31 May 2011 (not shown) this would be a more moderate 0.7%, 1.8% and 0.6%.

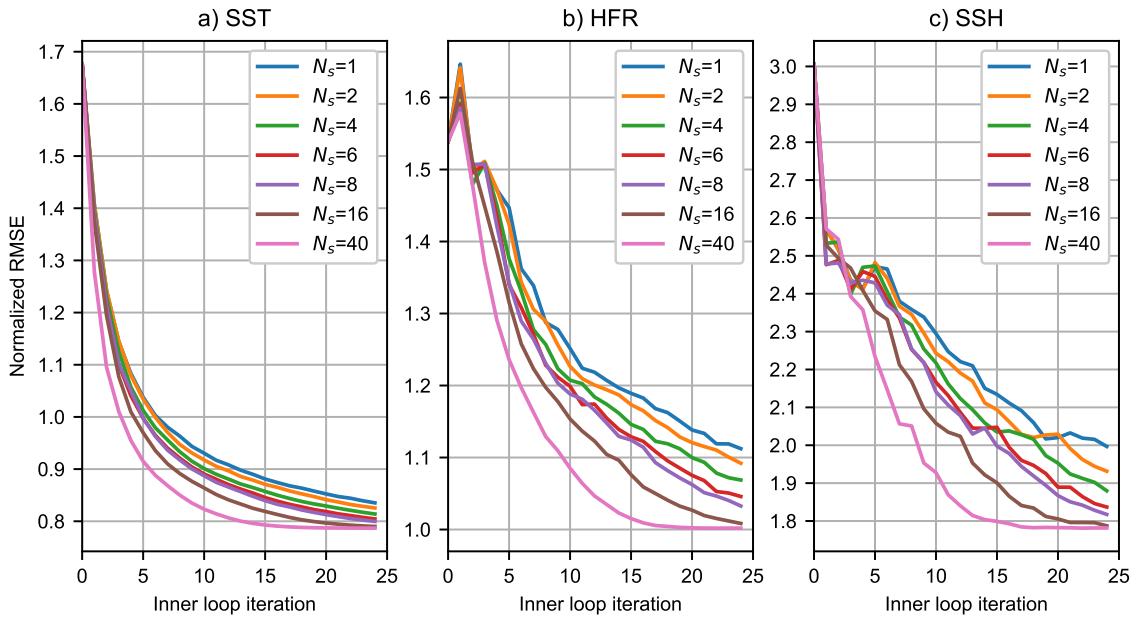


Figure 4.7: Expected normalized RMSE in (a) SST, (b) HFR daily-averaged surface velocity and (c) SSH observations as function of the number of inner loop iterations for the window starting on 26 August 2011.

Figure 4.8 compares differences in the initial condition corrections between RBCG and cluster search with  $N_s = 4$  on 26 August 2011. The plots on the left show the DA correction calculated after  $I = 12$  inner loop iterations for SST, surface velocity and SSH fields with  $N_s = 1$  and the plots on the right show the difference between the DA corrections after 12 inner loop iterations with  $N_s = 4$  and with  $N_s = 1$ . While both methods yield similar large-scale corrections, they differ in details at the scale of geostrophic eddies. The ability of the cluster search method to reduce  $J$  farther than RBCG allows to provide a more detailed change to SSH and as a result to the surface horizontal velocities (where differences reach magnitudes up to  $16 \text{ cm s}^{-1}$ ).

Figure 4.7 showed that the normalized RMSE (data fit) in the main run depends on  $N_s$  and hence on the search space. For the ensemble members additional dependency on the search space comes from the fact that when cluster search is used the right hand side of (4.8) is replaced by its  $\hat{\mathbf{B}}$ -projection on the search space. I.e. in the right-hand side of (4.8),  $\hat{\mathbf{d}}^{(m)}$ , is effectively replaced by  $\hat{\mathbf{V}}_i(\hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{V}}_i)^{-1} \hat{\mathbf{V}}_i^T \hat{\mathbf{B}} \hat{\mathbf{d}}^{(m)}$ . For the main run this is not an issue as by construction  $\hat{\mathbf{d}}^{(0)}$  lies in  $\text{span}(\hat{\mathbf{V}}_1)$ , but for the ensemble members this could result in the systematic elimination of a part of the errors contained in  $\hat{\mathbf{d}}^{(m)}$ . This would lead to DA corrections for the ensemble members that are too small and consequently an ensemble spread that is too large. To test whether this a valid concern, the normalized RMSE for each observation type, i.e. the RMS of those elements of  $\hat{\mathbf{d}}^{(m)}$  associated with one type of observations, is calculated and compared with the RMSE after taking the  $\hat{\mathbf{B}}$ -projection of  $\hat{\mathbf{d}}^{(m)}$  on  $\hat{\mathbf{V}}_I$  with  $I = 12$ . If  $\hat{\mathbf{d}}^{(m)}$  lies completely in  $\hat{\mathbf{V}}_I$ , as is the case for  $m = 0$ , the ratio of the latter over the former is one. The actual ratio in the experiment is calculated for each ensemble member and each window and the lower bound, upper bound and ensemble mean are shown in Figure 4.9. Figure 4.9 shows that, as expected, using the projection can result in the reduction of the RMSE (up to 40%). However, the figure also shows that taking the projection can increase the RMSE. This paradoxical behavior emerges because the projection uses the  $\hat{\mathbf{B}}$ -inner product, while in the calculation of the RMSE involves the normal, Euclidean, inner product. Taking the mean of the ratios over all ensemble members shows that increases in the RMSEs created by the  $\hat{\mathbf{B}}$ -projection

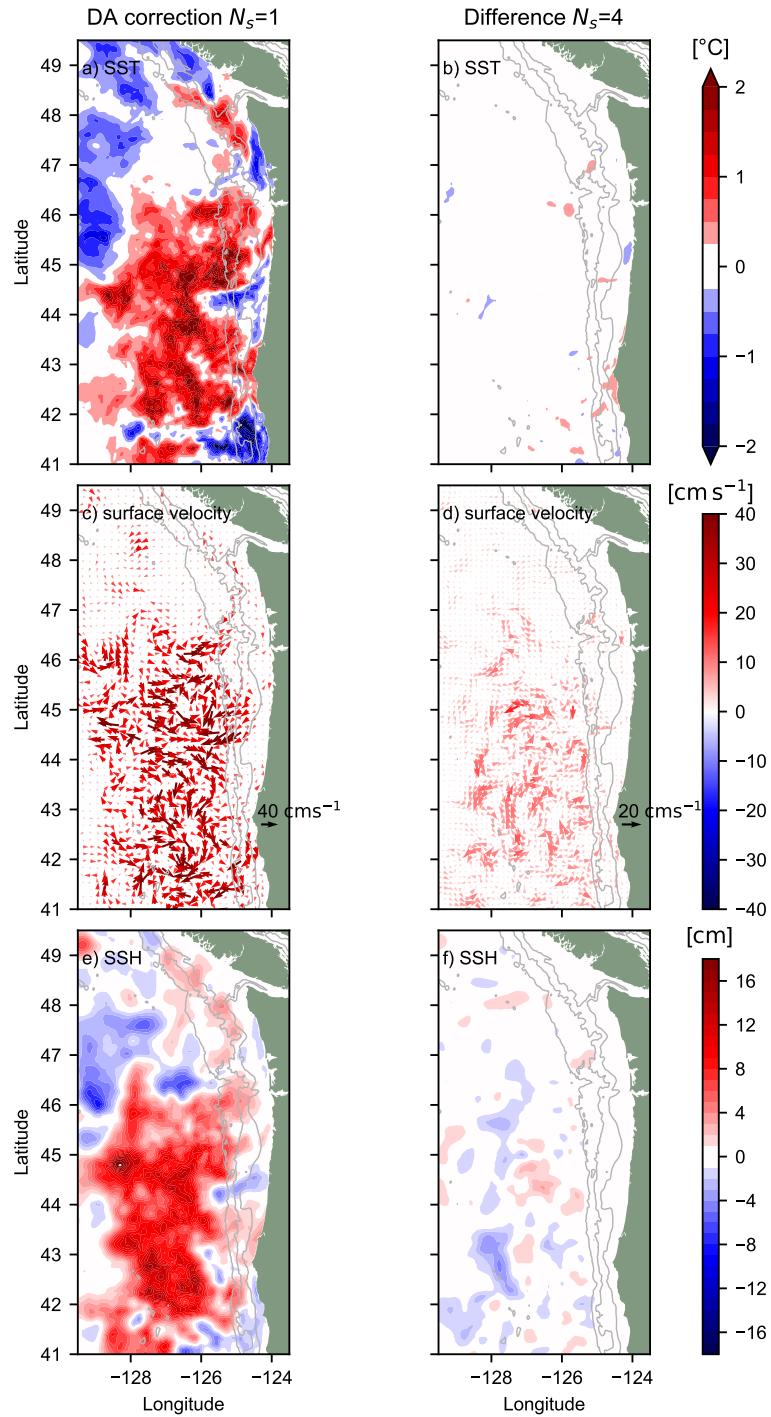


Figure 4.8: DA correction to the surface temperature field (top row), surface velocity field (center row) and sea-surface height (bottom column) based on 12 iterations of the RBCG method (left column) and the differences between this and the DA correction using cluster search with  $N_s = 4$  (right column).

mostly, but not completely, offset the reductions in RMSEs and that the net result is a small decrease in the RMSE of 1.7% for SST, 3.6% for HFR and 2.8% for SSH observations. So, the projection effect might indeed result in overestimation of the error variances by the ensemble, but the effect is small.

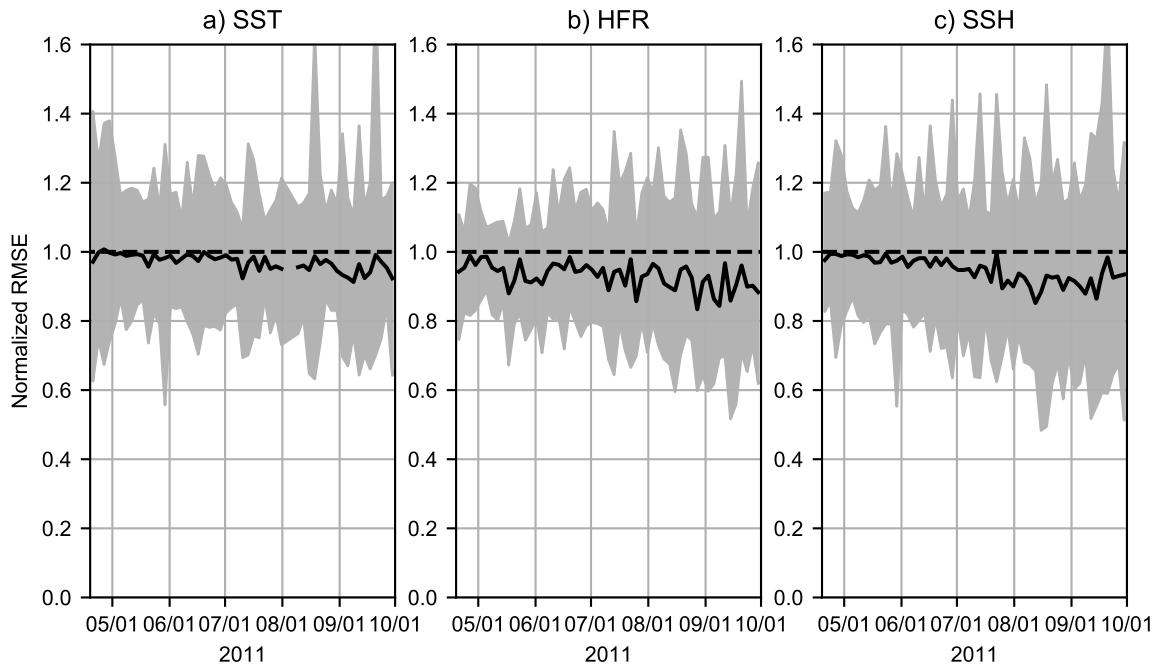


Figure 4.9: The RMS of the normalized forecast errors, as contained in vector  $\hat{\mathbf{d}}$ , after  $\hat{\mathbf{B}}$ -projection expressed as ratio versus the RMS of the total forecast error for (a) SST, (b) HFR and (c) SSH observations for each DA window. Grey area shows the range of this ratio over all ensemble members, while the black solid line marks the ensemble mean of the ratio. Dashed black line marks the value of 1 that would be obtained if the innovation vector  $\hat{\mathbf{d}}$  of the ensemble lies completely in the search space.

#### 4.6.2 Error Reduction

Figure 4.10 shows the time series of the cost function value  $J$  (4.23) for the main run and the ensemble members prior to the first inner loop iteration (Figure 4.10a)

and after the last inner loop iteration (Figure 4.10b). Both at the beginning and end of the minimization, the value of the cost function for the ensemble members is larger than for the main run. This is expected as suboptimal, perturbed winds and observations are used for the ensemble members. Figure 4.10b also displays the 95%-confidence interval based on the  $\frac{1}{2}\chi^2$  distribution with the number of degrees of freedom equal to the number of observations (*Bennett*, 1992), though it is hard to visually distinguish the upper and lower bound. The variability in the minimal value of the cost function of the main run is larger than can be expected based on  $\chi^2$ -statistics: only for 5.5% of the windows lies  $J$  within the 95%-confidence interval. However, there is no indication of systematic under- or overestimation of **R** or **B**: the minimum of  $J$  falls below the 95%-confidence interval as many times as it falls above the upper bound of the interval.

In the construction of the En4DVar system several assumptions were made. E.g. (4.3), the projection effects discussed in section 4.6.1, and that the selection process in cluster search focuses heavily on the main run ( $m = 0$ ). To test whether the system is still effective in correcting the ensemble, the RMSE between the data used in the assimilation and the nonlinear analyses and forecasts is calculated for the ensemble members and shown in Figure 4.11. Each line segment represents the RMSE in the analysis (left point) and in the forecast for the subsequent window (right point). The En4DVar system is effective in reducing the RMSEs: the analysis RMSE for the ensemble members exceeds that in experiment No DA (blue line) in less than 4% of the cases, while forecast RMSEs for the ensemble members exceed No DA forecast RMSEs in less than 27% of the cases. The RMSEs for the ensemble members are consistently larger than those for the main run (green line) as they are forced with perturbed wind fields and corrected with perturbed observations. However, the errors introduced by the perturbations cancel out in the ensemble mean as RMSEs for the ensemble mean lie below those of the ensemble members. In the ensemble Kalman filter the ensemble mean is used as best estimate for the truth (*Evensen*, 1994). Table 4.1 shows that the RMSEs of the main run are on par with those of the ensemble mean, with the exception of the RMSEs in the HFR observations after 14 August 2011

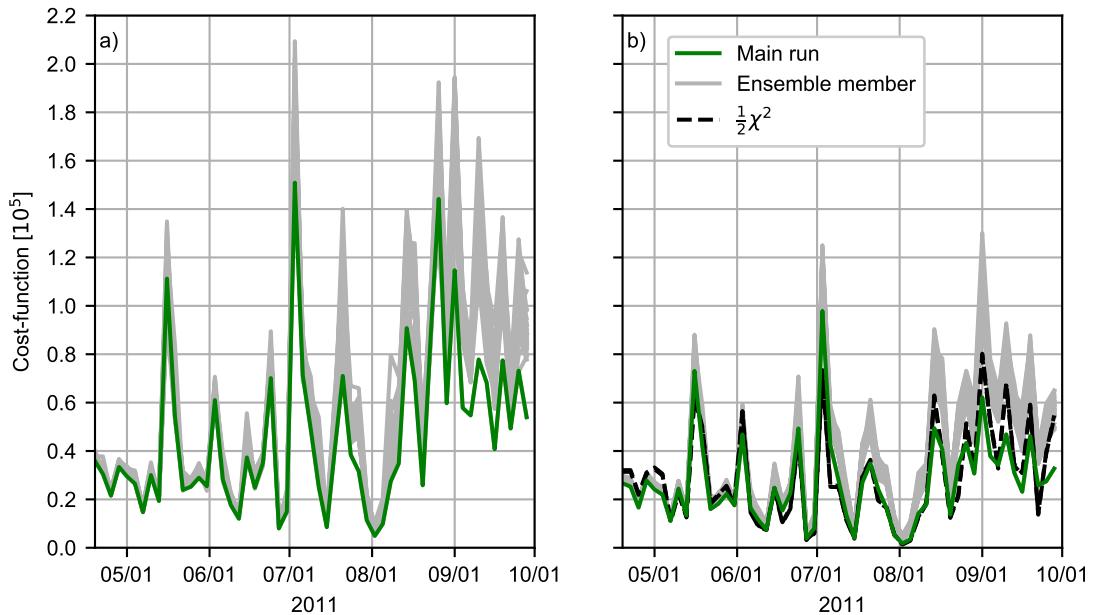


Figure 4.10: (a) Value of the cost function for the main run (green) and ensemble members (grey) prior to cost function minimization. (b) value of the cost function after minimization ( $N_s = 12$ ,  $I = 12$ ). The 95%-confidence interval for the expected value of the cost function is shown as a pair of dashed black lines.

(see Figure 4.11b). Hence, choosing the main run over the ensemble mean to produce the forecasts will have only a limited negative impact on the forecast accuracy.

Figure 4.12 shows the bias in experiment No DA, the ensemble members, the main run and the ensemble mean. Due to removal of the along-track mean in assimilation (see *Kurapov et al. (2011)* and chapter 2) the bias for SSH observations is zero by construction and therefore not shown. The bias in the HFR observation shows a spread around zero for both the ensemble members as well as the main run forecasts. The bias in the forecasts predictions for the SST observations, however, has a negative tendency with particularly large negative biases during the periods 13-16 May, 5-8 June, 21-24 July, 26-29 August, 25-28 September 2011. This results in an overall negative bias over the whole period as shown in Table 4.1. It is indicative for either

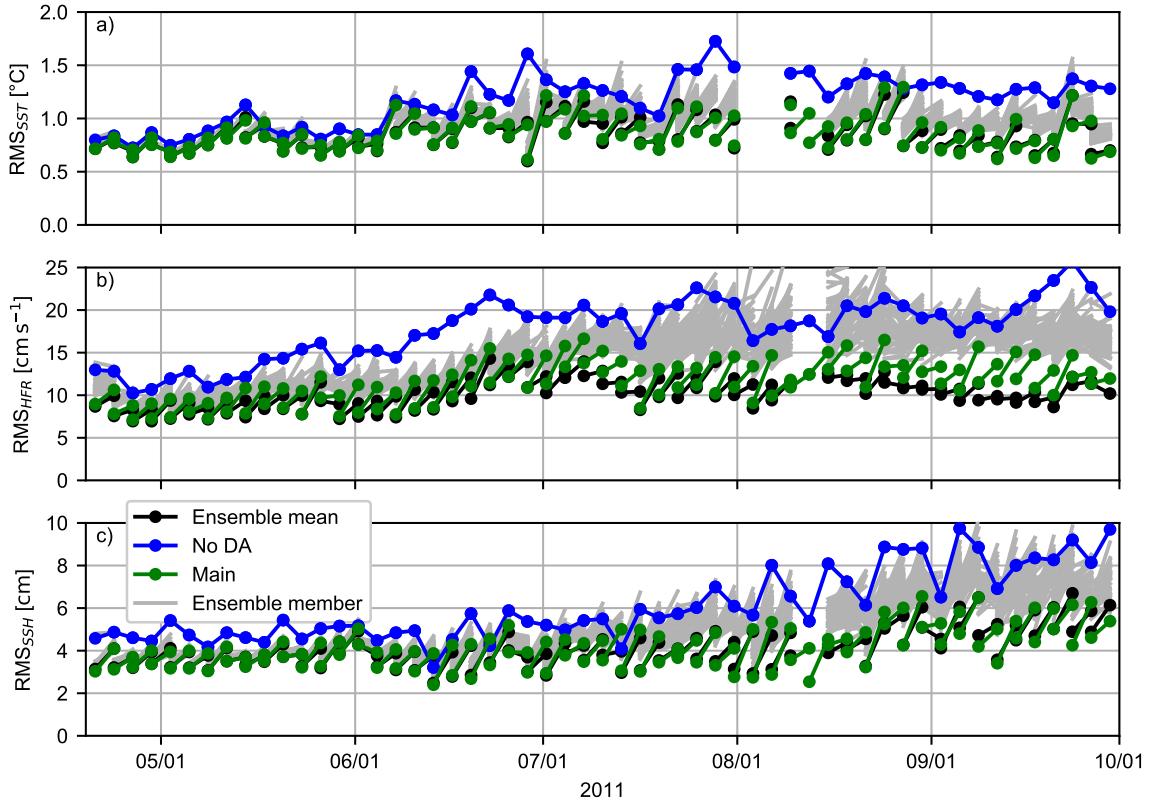


Figure 4.11: RMSE per 3-day window for (a) SST, (b) HFR daily-averaged velocity and (c) SSH observations from the model without DA (blue), the Main run (green), the different ensemble members (grey) and the ensemble average (black). The left side of each line piece marks the RMSE in the analysis, the right side the RMSE in the forecast.

insufficient surface heating in the model or an overestimation of the SST in the satellite observations.

#### 4.6.3 Ensemble Reliability

If the ensemble statistics are truly representative of the background error statistics, the ensemble is said to have a high reliability. The rank diagram is a diagnostic that can be used to test the reliability of the ensemble (*Hamill, 2001*). Figure 4.13

Table 4.1: RMS and the mean of the difference experiment minus observations as shown in Figure 4.11 and 4.12 for the different models and over the period 22 April to 28 September 2011.

	RMSE			Bias	
	SST [°C]	HFR [cm s <sup>-1</sup> ]	SSH [cm]	SST [°C]	HFR [cm s <sup>-1</sup> ]
No DA	1.17	18.3	6.4	-0.18	1.5
Main analysis	0.75	10.5	3.8	-0.04	0.2
Ensemble mean analysis	0.76	9.7	3.9	-0.05	-0.2
Main forecast	0.94	13.2	4.9	-0.07	0.4
Ensemble mean forecast	0.92	11.2	4.8	-0.13	-0.1

shows the rank diagrams for the three different types of observations. To construct these rank diagrams, the following steps are followed: (a) for each ensemble member add a random observation error to the forecast from the ensemble to account for the presence of observational errors, (b) count the number of ensemble forecasts that are lower than the measured value, (c) add 1 to this number to determine the rank of the observation, (d) count the frequency of each rank and divide by the total number of observations to determine the normalized frequency. The frequency of each rank is governed by a Poisson distribution with the mean equal to the number of observations divided by  $M + 1$  if the ensemble is reliable. In particular, the rank diagram should be flat. Comparison of the 95%-confidence intervals based on this distribution after normalization with the rank diagram in Figure 4.13 shows that the reliability of the ensemble is different for different fields. In the rank diagram for SST (Figure 4.13a) there is no distinctive peak. Instead the rank diagram has an upward slope. This is indicative of a negative bias in the ensemble, which is in agreement with Figure 4.12a. For the HFR observations, mid-range ranks are more abundant (see Figure 4.13b). This indicates that either the spread in the ensemble HFR is larger than the variability of the background errors or that the size of the observational errors is overestimated. The opposite is the case for SSH observations.

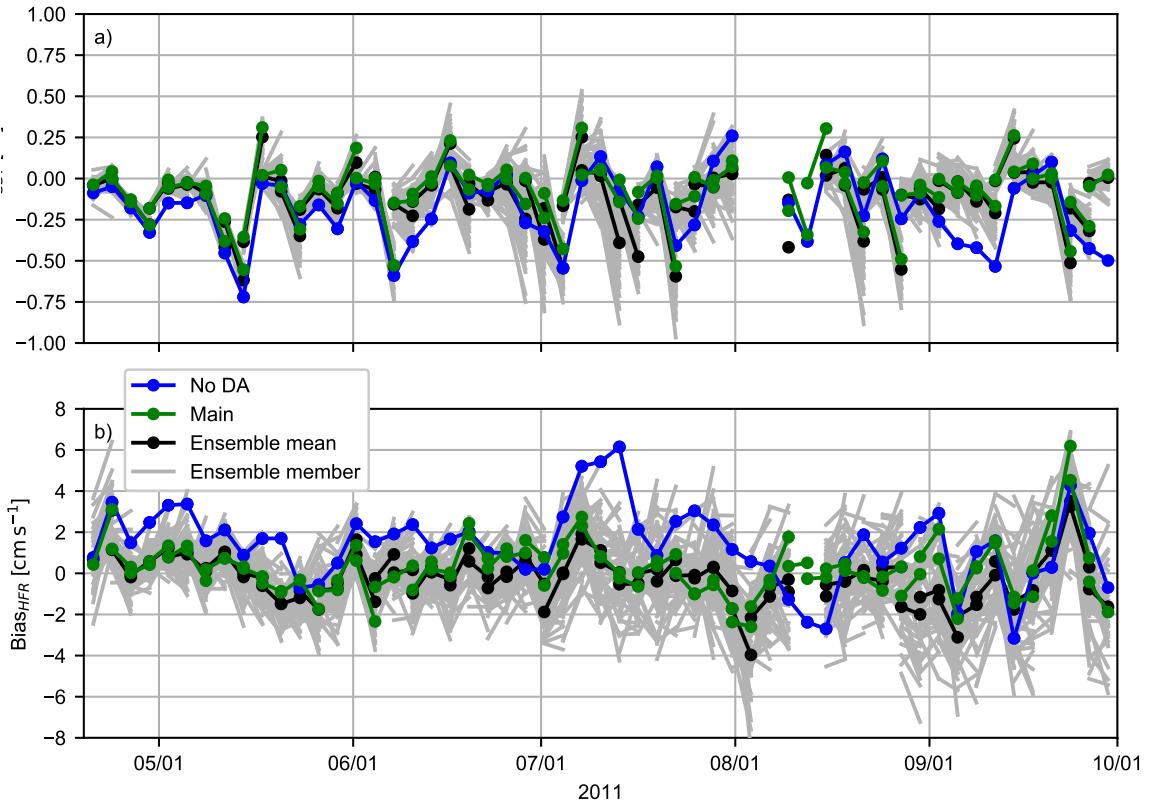


Figure 4.12: The 3-day model bias (model-observations) per window for (a) SST and (b) HFR daily-averaged velocity observations from the model without DA (blue), the main run (green), the different ensemble members (grey) and the ensemble average (black). The left side of each line piece marks the bias in the analysis, the right side the bias in the forecast.

Here the U-shape (Figure 4.13c) implies that the forecast ensemble underestimates the magnitude of the background errors.

Finally, estimates for the background error and observational covariances used in the DA system are compared with estimates obtained from the innovation statistics. The relations between innovation statistics and error variances are given by

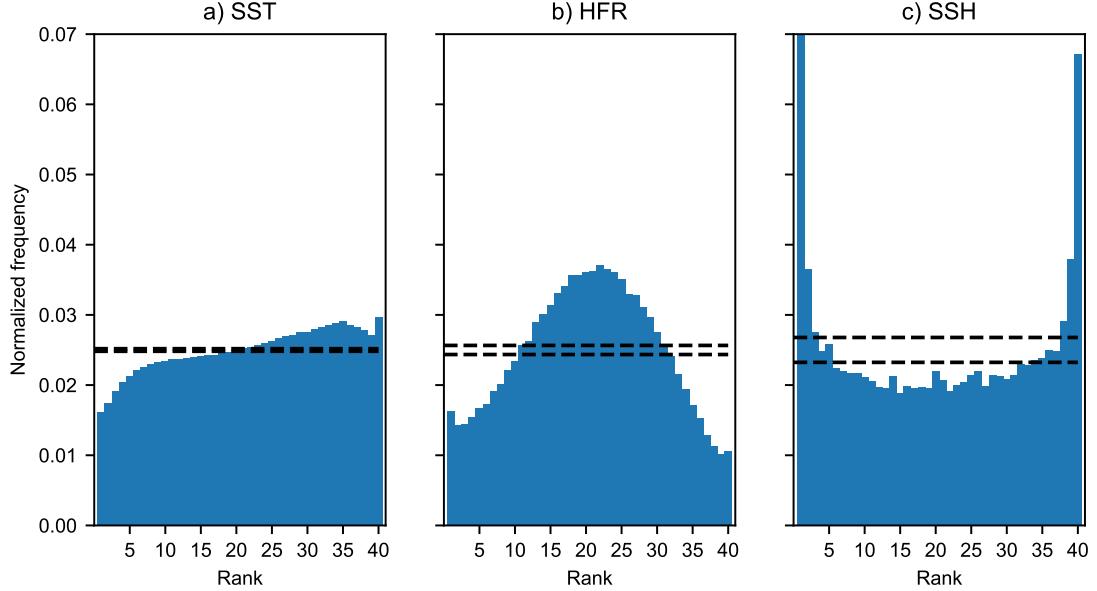


Figure 4.13: Rank of the perturbed (a) SST, (b) HFR daily-averaged surface velocity and (c) SSH observations within the ensemble.

(Desroziers *et al.*, 2005):

$$\left\langle (\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(0)}))^2 \right\rangle_d = (\mathbf{HMBM}^T \mathbf{H}^T + \mathbf{R})_{dd} \quad (4.24)$$

$$\left\langle (\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(0)}))_d (\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(0)}) - \mathbf{HMx}^{(0)})_d \right\rangle = (\mathbf{R})_{dd} \quad (4.25)$$

$$\left\langle (\mathbf{HMx}^{(0)})_d (\mathbf{y} - \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(0)}))_d \right\rangle = (\mathbf{HMBM}^T \mathbf{H})_{dd} \quad (4.26)$$

where  $1 \leq d \leq D$  is the index of the observation and  $\langle \cdot \rangle$  denotes the expectation value,  $(\cdot)_d$  the  $d$ -th element of the vector, and  $(\cdot)_{dd}$  the  $d$ -th element on the diagonal of a matrix. The expectation values on the left-hand side of (4.24), (4.25) and (4.26) are approximated by averaging over all observations of the same type in each window. These estimates are shown as blue lines in Figure 4.14a-c, Figure 4.14d-e and Figure 4.14g-i. An approximation to the right-hand side of (4.25) is obtained by averaging  $(\mathbf{R})_{dd}$  over all the observations of the same type. For the right-hand side of

(4.26) an approximation is obtained by doing the same for

$$\frac{(\mathbf{B}_{ens})_{dd}}{\mathbf{H}\mathcal{M}(\mathbf{x}_b)} \stackrel{def}{=} \frac{(\mathbf{HMBM}^T\mathbf{H}^T)_{dd}}{\frac{1}{M-1} \sum_{m=1}^{M-1} \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(m)})} = \frac{\frac{1}{M-2} \sum_{m=1}^{M-1} (\mathbf{H}\mathcal{M}(\mathbf{x}_b^{(m)}) - \overline{\mathbf{H}\mathcal{M}(\mathbf{x}_b)})_d^2}{\frac{1}{M-1} \sum_{m=1}^{M-1} \mathbf{H}\mathcal{M}(\mathbf{x}_b^{(m)})} \quad (4.27)$$

Error standard deviations based on these estimates are displayed as dashed black lines in Figure 4.14.

It should be noted that relationships (4.25) and (4.26) are only valid under the assumption that  $\mathbf{R}$  and  $\mathbf{B}$  are correctly specified. Consequently, the estimates on the left-hand side of (4.25) and (4.26) should not be interpreted as estimates for the true values of the diagonals of  $\mathbf{R}$  and  $\mathbf{HMBM}^T\mathbf{H}^T$ , but as estimates for the error variances effectively used by the DA system. E.g., a system in which  $\mathbf{B}$  is severely underestimated will produce DA corrections  $\mathbf{x}^{(0)}$  that are near zero. Consequently, the innovation estimate on the left-hand side of (4.26) will still be small.

Figure 4.14g,h,i shows that the ensemble error standard deviation grows over time for all three types of observations and hence that the wind perturbations are sufficient to prevent the ensemble spread from collapsing even without ensemble inflation. For SST the error standard deviation estimates from the innovation statics are in agreement with the specified standard deviations (see Figure 4.14a,d,g). Error standard deviation estimates for HFR observations too are consistent up to 1 July 2011. After 1 July 2011, however, the total error standard deviation estimate is too large (Figure 4.14b). The standard deviations for the observational errors agree (Figure 4.14e), so the overestimation is due to the fact that after 1 July the ensemble background error standard deviation estimate (black line Figure 4.14h) is larger than the standard deviation error estimate that is effectively used by the DA system (blue line in Figure 4.14h). This finding is consistent with the shape of the rank histogram in Figure 4.13b indicating overdispersion in the ensemble. Initially, the total SSH error standard deviation estimate from  $\mathbf{R}$  and  $\mathbf{B}_{ens}$  (black line in Figure 4.14c) underestimates the total SSH error standard deviation (blue line in Figure 4.14c). At this stage one could attribute this to the fact that the standard deviation from  $\mathbf{B}_{ens}$  (black line in Figure 4.14i) is smaller than the observational error estimate used in the DA (black

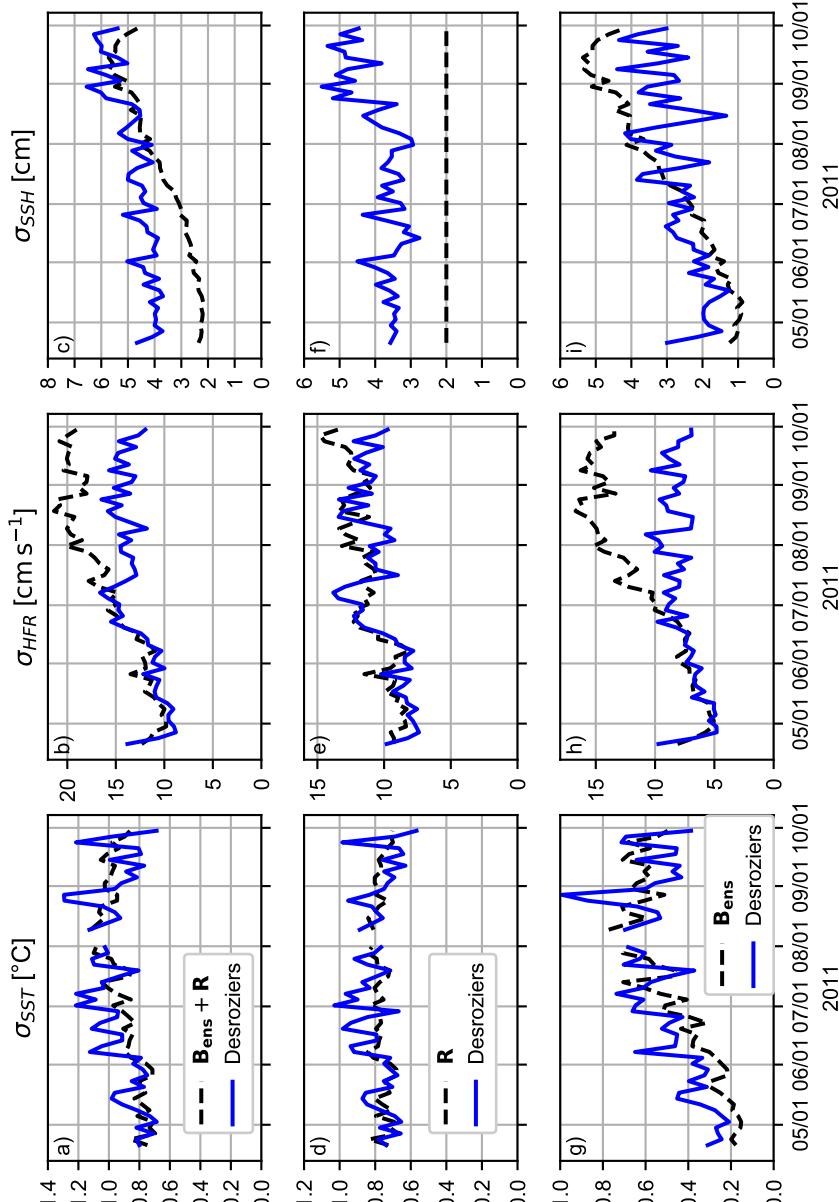


Figure 4.14: Estimates for total error variance (first row), observational error variance (center row) and background error variance (bottom row) based on SST observations (left column), HFR observations (center column) and SSH observations (right column). Blue lines are estimates based on (4.24)-(4.26), while the black lines indicate estimates based on the specified  $\mathbf{R}$  and the ensemble variance.

line in Figure 4.14f) resulting in small SSH DA corrections. As a consequence, the system effectively treats errors as predominantly observational (blue line lies above the black line in Figure 4.14f). However, the standard deviation from  $\mathbf{B}_{ens}$  keeps increasing over time, yet the observational error standard deviation estimate from the innovation statistics (blue line in Figure 4.14f) does not decrease to the specified observational error standard deviation (black line in Figure 4.14f). So, either the SSH observational error variance is underestimated in  $\mathbf{R}$ , or the ensemble is not fully able to resolve the structure of the actual background errors in the main run and treats the background errors as if they were observational errors.

## 4.7 Conclusions and Discussion

The development of ensemble-based 4DVAR systems has been one of the main focus areas for DA development in meteorology. Similarly to numerical weather prediction, there is a rationale to applying ensemble-based 4DVAR systems in oceanic prediction. Ocean shelf circulation is the most obvious potential application. Here the available data sets from multiple platforms sample well the coastal ocean currents and fronts. Utility of a static, balanced operator based  $\mathbf{B}$  can be limiting because of the shallow depth, the presence of the coastline, fronts, and a non-unique relation between temperature and salinity influenced by river plumes. The OR-WA forecast system, used in this study as a testground for En4DVar, is a good example where model error statistics are influenced by high temporal and spatial ocean state variability. Before En4DVar can be applied successfully to the OR-WA or any other coastal system, many technical details must be worked out, as outlined in this manuscript. The newly developed En4DVar systems need to go through statistical tests for self-consistency, using actual observations, which help us understand the system behavior and potential biases in the data.

Critical to the successful implementation of En4DVar for large prediction systems will be the development of time-efficient cost function minimization algorithms that take advantage of the massive parallel computer architectures. The cluster search method developed and tested in this study explores  $N_s$  search directions in parallel

at each inner loop iteration. It was found that using a relatively small number of parallel direction computations  $N_s = 4$  can reduce the wall time by 30% compared to RBCG to achieve the same amount of cost function reduction. However, the increased speed-up provided by the parallelization levels off as  $N_s$  continues to increase. This may suggest that the use of the block diagonal CG (which is equivalent to the cluster method with  $N_s = M$ ), as proposed for future development by *Auligné et al.* (2016); *Lorenc et al.* (2017), will not give much improvement over the cluster search method.

Recently, the idea of 4DVAR minimization parallelization by splitting the assimilation time interval is gaining popularity, so called saddle point methods (*Rao and Sandu*, 2016; *Fisher and Gürol*, 2017). In future studies, it will be interesting to compare the performance of these and the cluster search method. In perspective, we see that the two methods can be combined to deliver the best 4DVAR performance.

Both in this paper as well as in EVIL-S (*Auligné et al.*, 2016) the search space created during the minimization of the main run is used to find the DA corrections for the ensemble members. *Lorenc et al.* (2017) raises concern that the minimum of the ensemble member cost function on this search space is considerably larger than its absolute minimum. I.e. no fully converged solution to  $\mathbf{X}_i$  in (4.8) is obtained. Theoretically, consistent underestimation of the DA corrections could lead the ensemble members to diverge from the best estimate leading to a growing gap between the value of the cost function for the main run and for the ensemble members. Such a divergence has not been discovered in our study. Instead, we find that the cost function values for the ensemble members increase/decrease in tandem with the cost function values of the main run (Figure 4.10).

Slight underestimation of the DA corrections to the ensemble members can be one of the factors that helped us run the ensemble without the usual ensemble inflation (*Anderson*, 2001; *Anderson and Anderson*, 1999; *Hamill et al.*, 2001). Another contribution that increases ensemble spread is the perturbation of the wind forcing fields. For these, we found that the common estimate that the standard deviation of the errors in the wind forcing is 30% of that of standard deviation of the natural time variability (*Hénaff et al.*, 2009; *Palmer et al.*, 2009; *Vervatis et al.*, 2016) is of

the correct order of magnitude, but neglects the scale dependence in the wind errors. Posterior probability distributions for the variance of the wind field error EOF coefficients using ASCAT observations peaked at higher wind variability than would be expected based on the 30% estimate. The difference increased for the higher EOF modes that represent smaller-scale features in the wind field. We attribute this difference to the limited spatial resolution of NAM. This suggests that the assumption that model errors are solely proportional to variability in the model output cannot be justified. This agrees with the findings of *Whitaker and Hamill* (2012) that additive inflation is more suitable to representing model errors than multiplicative inflation.

Even though the En4DVar system was effective in reducing forecast errors compared to the case No DA, the rank diagram analysis suggests that the ensemble fails to represent the background error statistics perfectly: the ensemble overestimates the spread in the surface velocity background errors, while it underestimates the spread in the SSH background errors (see Figure 4.13). However, the latter is not responsible for the relatively DA corrections to the SSH as the ensemble background error variance for SSH is actually larger than what is effectively used by the DA system (Figure 4.14). One possible explanation for this is that the SSH background error cannot be entirely excited by wind perturbations. Although the rank diagrams in Figure 4.13 are not uniform, the bias and the maximum/minimum frequency-ratio of the diagrams is not exceptionally large compared to the rank diagram analyses in ensemble Kalman filter DA systems published earlier (*Cookson-Hills et al.*, 2017; *Fujita et al.*, 2007; *Leeuwenburgh*, 2007; *Meng and Zhang*, 2008).

In summary, in this study we developed an effective En4DVar method and tested it in a dynamically active coastal ocean region off Oregon and Washington, at the U.S. West Coast. The heart of the new method is the cluster search algorithm that allows accelerating the minimization convergence at the expense of using additional available computational cores for parallel direction search. Wind perturbations added

to the ensemble forecasts are assessed using the atmospheric model and wind scatterometer data. Statistical tests of the system are performed using actual observations. In chapter 5 we will discuss whether and where En4DVar yields better quality predictions than the traditional 4DVAR currently used in the operational OR-WA system.

## Chapter 5: Ensemble 4DVAR (En4DVar) Data Assimilation in a Coastal Ocean Circulation Model. Part II: Performance

### 5.1 Abstract

The Ensemble 4DVAR (En4DVar) data assimilation (DA) system introduced in chapter 4 is tested in the coastal waters offshore of Oregon and Washington, U.S. West coast, during the Spring and Summer of 2011. The background error covariance  $\mathbf{B}$  is derived from the forecast ensemble. Satellite sea-surface temperature (SST), sea-surface height (SSH) as well as daily-averaged radial surface currents from high-frequency radars (HFRs) are assimilated. The performance of the En4DVar system is compared with a 4DVAR system using a static  $\mathbf{B}$ . It is found that the presence of the Columbia River plume has a profound impact on the ensemble  $\mathbf{B}$ . Near the plume front the SST-SSS covariance can be up to a factor 20 larger in magnitude than in the static  $\mathbf{B}$ . This introduces large spatial, and since the plume front moves, temporal variability in the ensemble  $\mathbf{B}$ . As a result, the En4DVar was, contrary to the 4DVar system with static  $\mathbf{B}$ , successful in reproducing the temperature-salinity relationship along a glider transect. The En4DVar system also produced more accurate forecasts and analyses for the temperature and salinity at the NH10 buoy location and slightly better forecasts for SST, surface currents and SSH than 4DVAR with a static  $\mathbf{B}$ . However, the strong surface temperature-salinity covariances in combination with the episodic occurrence of large-scale SST biases lead to erroneous freshening in the center of the model domain. Application of the Hierarchical Box Salinity Constraint method, which penalizes DA corrections that change area-averaged surface salinity, was successful in partly suppressing the erroneous freshening.

## 5.2 Introduction

For the benefit of the local fishing communities, government agencies and other users the Oregon State University (OSU) coastal ocean forecast system provides forecasts of temperature, salinity, currents and sea-surface height in the Oregon-Washington (OR-WA) coastal area (<http://nvs.nanoos.org/Explorer>). Initial conditions for the forecasts are corrected by assimilating different types of surface observations using the 4DVAR data assimilation (DA) algorithm in a series of 3-day windows. This requires specification of the forecast, or so called background, error covariance **B**. The **B** currently implemented in the OR-WA system is static, i.e., it does not change from one assimilation window to the next. In this covariance, the balance operator and its adjoint counterpart (*Kurapov et al.*, 2011; *Weaver et al.*, 2005) are used to correlate errors in different components of the state vector, including SSH and three-dimensional fields of the horizontal velocity, temperature, and salinity. The balance operator is based on diagnostic relations including geostrophy, thermal wind balance, the linearized equation of state, and a simplified temperature-salinity relation.

Coastal waters in the OR-WA area are very dynamic, with wind-driven currents in excess of  $0.5 \text{ m s}^{-1}$ , strong temperature fronts, geostrophic and ageostrophic baroclinic instabilities, and jets separating from the shelf toward the ocean interior (*Allen et al.*, 1995; *Barth et al.*, 2000; *Brink*, 1987; *Castelao and Barth*, 2005; *Garvine and Monk*, 1974; *Halpern*, 1976; *Huyer*, 1983; *Huyer et al.*, 2007; *Koch et al.*, 2010; *Kosro*, 2005; *Mooers et al.*, 1976; *Orton and Jay*, 2005). The freshwater outflow from the Columbia River creates a shallow plume that spreads over a large area (*Berdeal et al.*, 2002; *Hickey et al.*, 2005; *Huyer et al.*, 2005; *Liu et al.*, 2009). Its location changes on seasonal and shorter time scales in response to the winds. In summer, due to the predominantly southward winds the resulting coastal upwelling extends the Columbia River plume to the south of the river mouth and offshore. During periods of wind relaxation the plume is advected toward the coast, freshening the coastal waters off Oregon (e.g., see Figure 2.1). On the temperature-salinity diagram, points corresponding to the plume and points corresponding to near-surface waters offshore or inshore of the plume occupy distinctively different parts of the diagram. In such a

dynamic area the use of a static  $\mathbf{B}$  can be limiting.

A  $\mathbf{B}$  estimated from an ensemble of perturbed model runs can take into account temporal and spatial variability in the estimated fields. Two general approaches have been developed to obtain  $\mathbf{B}$  from an ensemble. In the parametric approach (*Berre et al.*, 2015; *Bonavita et al.*, 2012; *Küçükkaraca and Fisher*, 2006) the error correlations are either prescribed or are assumed to be combinations of a limited set of wavelets. The ensemble is only used to estimate the error variances and coefficients for the wavelets. In the other approach, which is used in this study, the  $\mathbf{B}$  is directly estimated from an ensemble.

While ocean forecasters only recently started work on ensemble-variational DA, it has been studied in meteorology for the last two decades. In global models the ensemble covariances have been found to be different from the balance operator based  $\mathbf{B}$  (*Daget et al.*, 2009; *Kuhl et al.*, 2013). For example, *Buehner* (2005) found for a global meteorological model that the ensemble standard deviation exhibits larger spatial variability in the horizontal, that the ensemble covariance is sensitive to the presence of fronts, and that temperature error fields are not predominantly in geostrophic balance with the wind fields.

Results on whether these differences between the ensemble  $\mathbf{B}$  and balance operator  $\mathbf{B}$  are relevant to the forecasts produced by DA systems are ambiguous. In a toy model, all the tested ensemble-based DA systems were able to produce analyses with lower root-mean square error (RMSE) than 4DVAR with a static  $\mathbf{B}$  provided that the ensemble used was sufficiently large and ensemble inflation was applied (*Fairbairn et al.*, 2014). *Caya et al.* (2005) found for a model of a supercell storm that 4DVAR performed best during the first assimilation cycles, but was overtaken by the ensemble Kalman filter after that. However, comparison of 4DVAR with a static  $\mathbf{B}$  versus a 4DVAR with  $\mathbf{B}$  coming entirely from an ensemble Kalman filter showed an overall degradation of performance and an increase in the RMSE in the forecasts for the latter (*Kuhl et al.*, 2013; *Lorenc and Jardak*, 2018).

In chapter 4, the En4DVar method was described. The OR-WA coastal ocean forecast system was used to illustrate the computational efficiency of the cluster search

minimization algorithm and to introduce essential statistical tests of the dynamical ensemble and the resulting time-variable  $\mathbf{B}$ . However, we did not demonstrate if in any regard whether the use of the more computationally demanding En4DVar yields an improvement in the forecast accuracy compared to a traditional 4DVAR system with the static balance operator  $\mathbf{B}$ . This void is filled here.

In the process of this study, we recognize one of the general potential dangers using En4Dvar for poorly observed fields. Contrary to the static balance operator, the ensemble generates a  $\mathbf{B}$  with strong temperature-salinity covariances in the plume and along the upwelling front. While the covariance can correctly represent the error relationship, large-scale biases with respect to satellite SST will create strong and erroneous corrections to the surface salinity field in the plume area, which is not constrained by observations. As part of this study we come up with a methodology to constrain the near-surface salinity based on the forecast solution.

This chapter is organized as follows. Section 5.3 provides a summary of the ocean model, En4DVar DA and the 4DVAR DA system. In this section, we also explain how to mitigate the effect of the SST bias on the surface salinity, using the Hierarchical Box Salinity Constraint (HBSC) method. Horizontal and vertical spatial patterns of the static and the ensemble  $\mathbf{B}$  are compared in section 5.5. Analyses and forecasts from the En4DVar system are compared to the model without assimilation and the standard 4DVAR in section 5.7. In section 5.8, our findings are summarized.

### 5.3 Model experiments

In the OR-WA system, nonlinear forecasts are obtained using the Regional Ocean Modeling System (ROMS) ([www.myroms.org](http://www.myroms.org)) with a 2-km resolution in the horizontal and 40 terrain-following layers in the vertical direction. ROMS solves for temperature ( $T$ ), salinity ( $S$ ), zonal velocity ( $u$ ), meridional velocity ( $v$ ) and sea-surface height ( $\zeta$ ). The 4DVAR DA utilizes the AVRORA tangent linear and adjoint codes developed in-house (*Kurapov et al.*, 2009, 2011; *Yu et al.*, 2012). Model forcing includes wind stress, surface heat flux, tides and river inputs. Details of the model forcing and spin-up are found in chapter 2 and 4.

Four experiments are discussed in this chapter: *No DA*, *Ens*, *Ens-HBSC*, *Bal-HBSC*. All experiments cover the period of 19 April 2011 to 1 October 2011. In experiment No DA, the model is run continuously without DA. In experiment Ens, an ensemble of 40 forecasts is carried throughout the study period. The main run is computed using the wind forcing without perturbations. The other 39 runs are carried using perturbed winds (see section 4.5) and the daily-averaged fields from these 39 runs are utilized to compute  $\mathbf{B}$ , with localization as outlined in chapter 3. Generation of the ensemble and application of the DA corrections to the main run and ensemble members are performed by the En4DVar system introduced in chapter 4. It was found that case Ens can obtain unphysical corrections to the sea-surface salinity (SSS). To reduce this effect, the Hierarchical Box Salinity Constraining (HBSC) procedure is proposed here (section 5.4). Case Ens will refer to En4DVar without HBSC. In experiment Ens-HBSC, DA is performed using En4DVar, HBSC and the same wind perturbations as in experiment Ens. Experiment Bal-HBSC uses the standard 4DVAR with the static, balance operator  $\mathbf{B}$  as described in *Kurapov et al.* (2011) plus the HBSC. For reference, the balance operator is described in the appendix E.

In all three DA experiments the model is propagated forward in time as a series of three-day windows with the first window starting on 19 April 2011 and the last on 28 September 2011. Initial conditions for the main run and each ensemble member are updated at the beginning of each window by assimilation of satellite sea-surface temperature (SST), daily-averaged radial current observations from high-frequency radars (HFR) and along-track satellite sea-surface height (SSH) altimetry with the along-track mean removed. See chapter 4 for the details on the datasets used. The nonlinear ROMS is run for 6 days starting from the corrected initial conditions. The model output over the first three days is referred to as the analysis and the next three days are referred to as the forecast.

The correction  $\mathbf{x}$  to the ocean state at the beginning of the analysis window is calculated by minimizing the cost function (*Courtier et al.*, 1994; *Egbert et al.*, 1994):

$$J(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{B}^{-1} \mathbf{x} + \frac{1}{2} (\mathbf{d} - \mathbf{H}\mathbf{M}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{H}\mathbf{M}\mathbf{x}) \quad (5.1)$$

with  $\mathbf{H}$  the sampling operator that generates predictions for the different observations from the model output,  $\mathbf{M}$  is the tangent linear model, i.e. the ROMS model linearized around the forecast, and  $\mathbf{d}$  is the difference between observations and non-linear forecast predictions for the observations. The observation error covariance  $\mathbf{R}$  is assumed to be diagonal. In experiments Ens and Ens-HBSC  $\mathbf{B}$  is equal to the localized ensemble covariance of the 24h-averaged ensemble fields.

In all DA cases, the cluster search method is used to minimize the cost function (see section 4.4.3). This method searches for a correction to the main run and all the ensemble members in a search space that is expanded by  $N_s$  dimensions per inner loop iteration. Using  $N_s > 1$  accelerates convergence compared to the traditional conjugate gradient method. In our cases  $N_s = 4$  and 12 inner loop iterations are performed. In Bal-HBSC no ensemble is available. In order to use cluster search a low-rank surrogate ensemble has to be generated. Generation is described in appendix F.

## 5.4 Hierarchical Box Salinity Constraint (HBSC)

As was discussed in the introduction and as will be demonstrated in section 5.6, SST assimilation using En4DVar can provide large and erroneous corrections to the near-surface salinity field in the area of the plume. In our case, as will be shown below, it happened in the windows where the SST product had a cold bias compared to SST observations. To mitigate this, the DA correction is “nudged” to the prior model salinity. We did not want to nudge SSS to the prior model point-by-point since this would not accommodate changes to the location of the river plume front. Constraining simply the area-averaged SSS did not mitigate effectively the adverse effect of the biased SST, since the domain-averaged SSS can be preserved when salinity is wildly redistributed. The scheme described here provided a useful constraint.

To start, the surface area in the model interior is presented as an hierarchy of boxes (5.1). The largest box is the entire domain, excluding the edges where the background error variance is reduced to zero and also excluding Puget Sound in the northeast corner of our domain. The next level of boxes is obtained by dividing the largest box into 4. Level 3 boxes are obtained by dividing level 2 boxes by 4 and so

on. In order not to impede corrections on the scale of the background error covariance ( $R = 25$  km) only five levels of boxes are used. In this case, the smallest boxes have dimensions of  $30 \times 65$  km.

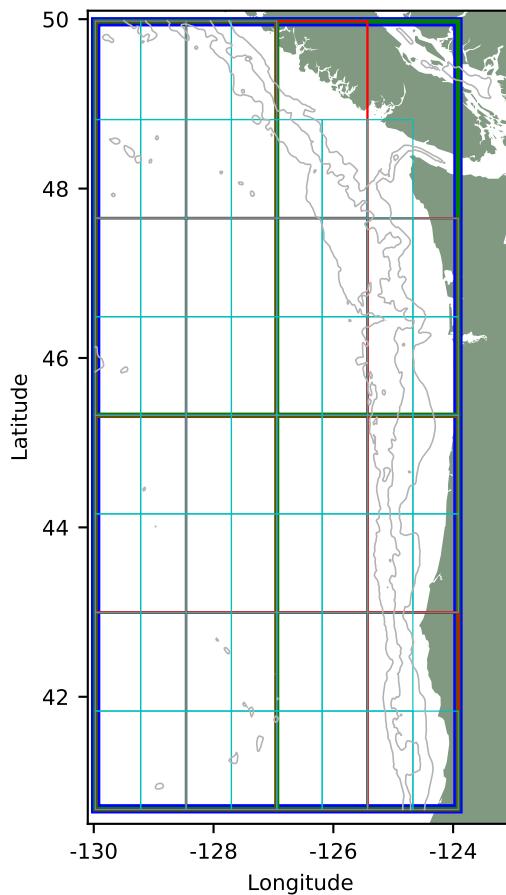


Figure 5.1: Overview of the model domain and the boxes used in HBSC. Boxes at the different levels of the hierarchy are shown with different colors.

The prior (forecast) SSS averaged in each of there areas (boxes) are treated as additional observations. Then, using the incremental 4Dvar formulation (5.1), the

innovation vector  $\mathbf{d}$  corresponding to these “data” is zero and the additional term in the cost function provides a penalty for deviations of the box-averaged SSS from the prior:

$$J_{HBSC} = \sum_i \frac{[\Delta S_i]^2}{\sigma_{S,i}^2} \quad (5.2)$$

Here  $\Delta S_i$  is the DA change in the box-averaged salinity of box  $i$ ,  $\sigma_i^2$  is set to be 5% of the variance in the box-averaged salinity as calculated at beginning of the DA windows in experiment No DA. The average value for  $\sigma_{S,i}^2$  for the boxes at levels 1 through 5 is  $(5.8 \cdot 10^{-2})^2$ ,  $(1.2 \cdot 10^{-1})^2$ ,  $(1.7 \cdot 10^{-1})^2$ ,  $(2.1 \cdot 10^{-1})^2$ ,  $(2.3 \cdot 10^{-1})^2$  respectively. The reasoning behind this scheme is that local corrections to SSS on spatial scales smaller than the smallest box size contribute little to the change in the domain-averaged SSS. Consequently, these corrections are not suppressed. Large-scale corrections, however, will result in changes of box-averaged salinity in boxes at different levels, thus rapidly increasing  $J_{HBSC}$ .

## 5.5 Covariances

In this section, the spatial structures in the balance operator-based and ensemble  $\mathbf{B}$  are compared. First, comparisons will be made for the offshore part of the ocean, where dynamics are governed at the lowest order by geostrophy and where the balance operator covariance is expected to be an accurate approximation for the background error covariance. Then a similar analysis will be obtained for a location near the plume edge. The focus will be on the correlations and covariances with SST as this is the field for which most observations are available and thus, via  $\mathbf{B}$ , heavily impacts the other fields.

Surface maps of the background error correlation between  $T$  at a point far offshore,  $\mathbf{r}_0 = (127.57^\circ\text{W}, 47.00^\circ\text{N})$ , and fields of  $T$ ,  $S$ , and surface currents are shown in Figure 5.2. In the balanced  $\mathbf{B}$  (Figure 5.2a,b), the SST-velocity correlation looks like an anticyclonic eddy, consistent with the balanced diagnostic relations. At the surface,  $T$ - $T$  correlations (Figure 5.2a) and  $T$ - $S$  cross-correlation (Figure 5.2b) are the same apart from a sign change. This consistent with the simple relation between

$T$  and  $S$  in (E.2). In particular, the  $T$ - $S$  correlation at  $\mathbf{r}_0$  is  $-1$ . The correlation between  $T(\mathbf{r}_0)$  and the surface velocity in the vicinity peaks at 0.36-0.41.

We hypothesize that far offshore, away from strong coastal fronts and jets, the ensemble  $\mathbf{B}$  computed using daily-averaged member fields yields correlation structures that overall are close to the balance operator  $\mathbf{B}$ . To test this, ensemble perturbations from the window ensemble mean have been collected from all ensemble members and all windows in experiment Ens. The correlations obtained from this super-ensemble are shown in Figure 5.2c,d. The correlations are qualitatively similar to the result with the balanced  $\mathbf{B}$  shown in Figure 5.2a,b. A more quantitative assessment reveals that the  $T(\mathbf{r}_0)$ - $S(\mathbf{r}_0)$  correlation is only  $-0.58$ , deviating from the value of  $-1$  noted above. This partly due to the fact that mean  $T$  and  $S$  vertical profiles are different from window to window. The cross-correlation of  $T(\mathbf{r}_0)$  with the surface velocity field is also weaker, with peak magnitudes reaching 0.15-0.21.

The situation is different if we focus on the correlation in just one window and for a sample point located in the frontal region. Figure 5.3 shows the surface correlation for 9 July 2011 and for a point chosen on top of the  $S = 31.5$  contour. This contour marks the edge of the Columbia River plume. The fresher, warmer plume water is found to the west and saltier and colder upwelled water is found to the east of this location. Interestingly, surface  $T(\mathbf{r}_0)$  is strongly correlated with the cross-front surface current (correlation magnitudes up to 0.79), reflecting the dominance of the advection mechanism displacing the front. The point where the temperature-velocity correlation is the strongest is marked by a green dot. The strongest  $T$ - $S$  correlation in this surface map is also large in magnitude,  $-0.86$ .

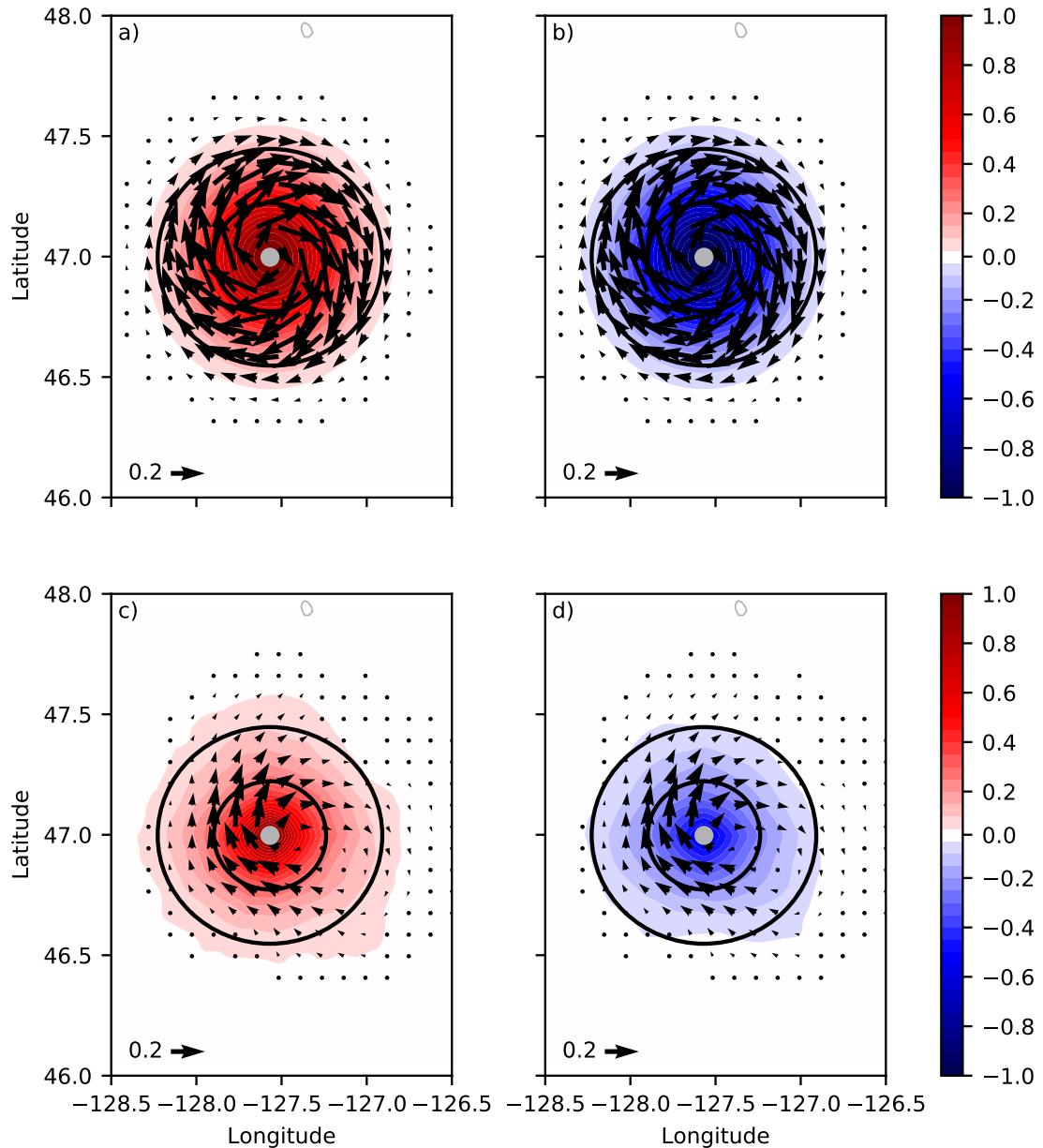


Figure 5.2: Surface maps of the background error correlations in the interior ocean, away from strong fronts. Shown are correlations of  $T$  at the grey dot location and (left) surface  $T$  and (right) surface  $S$ . (Top) balanced  $\mathbf{B}$ , (bottom) ensemble  $\mathbf{B}$ , where correlations are computed using ensemble member perturbations from all the windows.  $T(\mathbf{r}_0)$ -surface velocity correlations are shown as vectors in each panel.

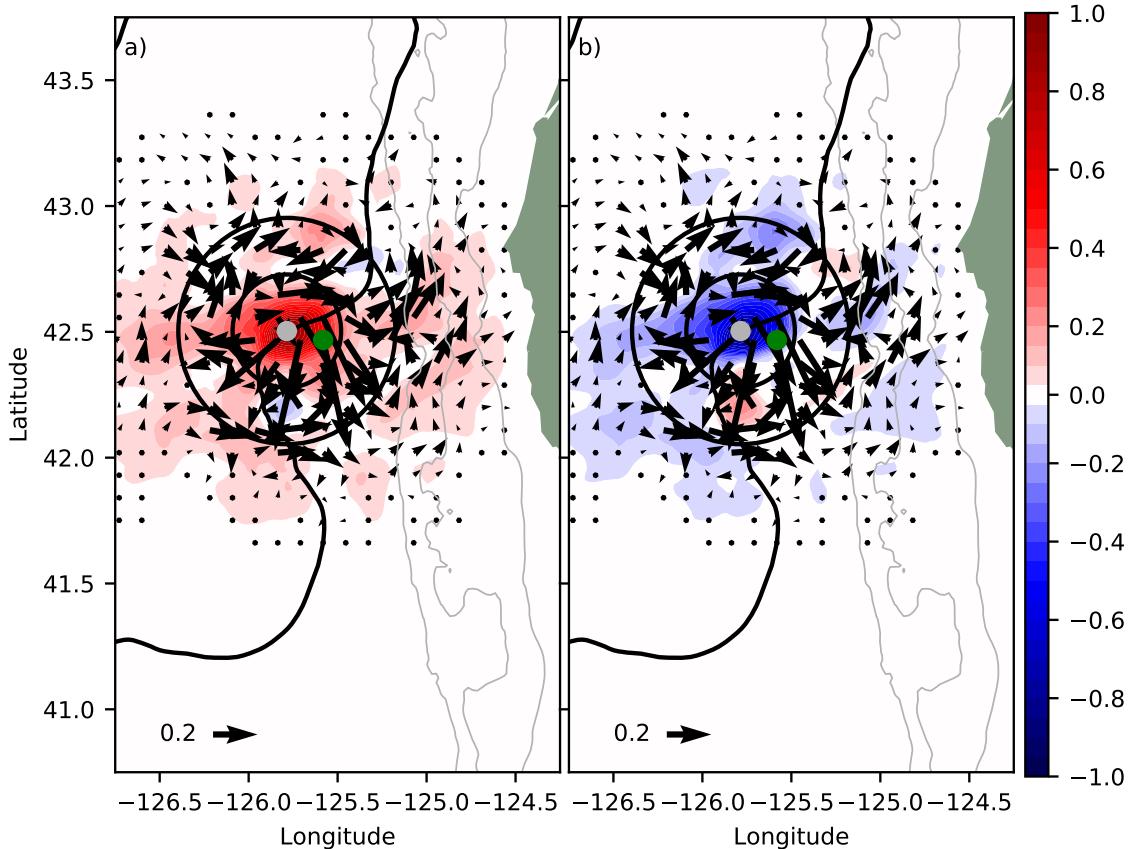


Figure 5.3: Surface maps of ensemble background error correlations for the reference point at the inshore edge of the Columbia River front (black line is  $S = 31.5$ ). Shown are surface maps of the background error correlations of  $T$  at the grey dot location and (a) surface  $T$  and (b) surface  $S$ . The correlations are computed in a single assimilation window, starting on 9 July 2011.  $T(\mathbf{r}_0)$ -surface velocity correlations are shown as vectors in each panel.

To analyze the impact of the river plume on the  $T-T$  and  $T-S$  correlations, 50 200-km meridional sections are selected randomly in each assimilation window. If the minimum salinity along the section is larger than 31.5, the section is classified as outside the plume. If the maximum salinity along the section is lower than 31.5,

the section is classified as inside the plume. Only sections fully inside or outside the plume are retained for analysis here. The ensemble correlation between the SST at the middle point of each section and SST and SSS along the length of each section have been calculated. The localized ensemble correlations from individual assimilation windows along sections falling in either of these two classes are shown in Figure 5.4 (grey lines). Also shown are the average ensemble correlations (dashed lines) obtained by combining the different sections and different windows into one super-ensemble. Again, as above, the ensemble mean is computed and subtracted individually in each window to obtain perturbation fields, from which the correlation is calculated. The  $T$ - $S$  cross-correlations vary widely, e.g., taking either sign outside the plume (Figure 5.4c). Consequently, the averaged cross-correlation at the central point,  $-0.31$ , is only a fraction of the  $-1.0$  cross-correlation used in the balance operator. Within the plume (Figure 5.4d), the averaged  $T$ - $S$  ensemble correlation is closer to the balance operator correlation, being mostly negative and yielding the averaged value of  $-0.83$  at the section center.

For the balanced  $\mathbf{B}$  the  $T$ - $T$  and  $T$ - $S$  horizontal correlations are shaped as a Gaussian. The correlation is thus 60% and 14% of its peak value at distances  $R = 25$  km and  $2R = 50$  km, respectively (here  $R$  is the horizontal correlation length scale assumed in the balance operator covariance). For the averaged ensemble correlations in Figure 5.4, we determine the distances from the center  $r$  at which the average correlations have dropped to 60% and 14% of their value at  $r = 0$ . For the correlation without localization these distances are found to lie between 14-20 and 40-54 km. The latter is comparable to the 50 km value in the balance operator, the former is smaller than its balance operator equivalent. So, although balance operator correlation length scales ( $4R$ ) are in agreement with those found in the ensemble, the ensemble correlations decrease faster than the Gaussian decreases in the balance operator correlation. This is also directly visible in Figure 5.4a,b. The localization scheme creates an additional reduction in scale: for the localized correlations the scales lie respectively between 12-16 and 34-40 km.

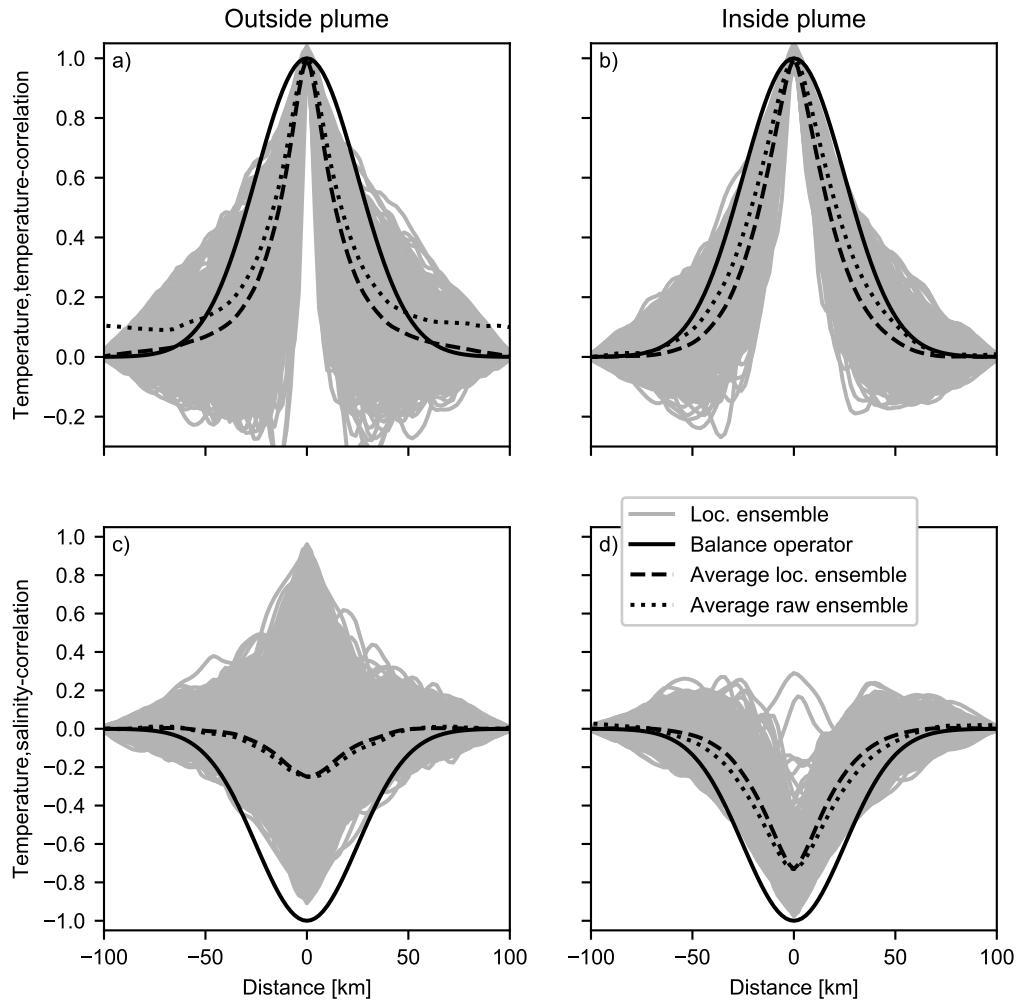


Figure 5.4: Surface (top)  $T$ - $T$  and (bottom)  $T$ - $S$  background error correlations from experiment Ens for randomly selected 200-km meridional sections (left) outside and (right) inside the Columbia River plume. Error ensemble correlations are computed for surface  $T$  at the center point (distance= 0 km) and SST and SSS along the sections. Grey lines: localized ensemble correlation in each window; dashed: averaged correlations, dotted: average correlations for raw (not localized) ensembles; solid black: the balanced  $\mathbf{B}$  correlations.

Not only the background error correlations produced by the ensemble are different from those produced by the balance operator, but so are the error covariances. Figure 5.5 illustrates this showing surface maps of the ensemble error SST variance, at the top, and the  $T\text{-}S$  covariance (where  $T$  and  $S$  errors are sampled at the same point). Three windows, 1 May, 20 July, and 28 September, 2011, are selected to represent three different Columbia River plume geometries. On 1 May the plume stretches over the shelf northward of the river mouth. On 30 July the plume is diverted southward and offshore by the upwelling favorable winds. On 28 September 2011, at the end of the upwelling season, the plume is found in the same area as on 30 July, but the salinity signal is weaker. Also, a new plume is forming northward of the mouth as the winds have reversed to northward.

The ensemble variance in SST (Figure 5.5a-c) is considerably smaller than the one assumed in the balanced **B**: the latter is  $0.81 \text{ }^{\circ}\text{C}^2$ , while the former has a median value of  $0.10 \text{ }^{\circ}\text{C}^2$ . However, locally, in frontal areas, where SST assimilation will have the strongest impact, the ensemble standard deviation can obtain much larger values, up to  $1.63 \text{ }^{\circ}\text{C}^2$ . These high variances are found, in particular, near the  $15^{\circ}\text{C}$  isotherm (solid black line). Along the coast between  $41\text{--}48^{\circ}\text{N}$ , this isotherm is indicative of the location of the cold upwelling front. Similarly (Figure 5.5d-f), the median of the  $T\text{-}S$  covariance over all windows is  $-0.002 \text{ }^{\circ}\text{C}$ , which is two orders of magnitude smaller than the  $-0.13 \text{ }^{\circ}\text{C}$  used in the balanced **B**. However, locally the strength of the SST-SSS covariance can become a factor 20 larger in magnitude than the latter, reaching  $-2.7 \text{ }^{\circ}\text{C}$ . Contrary to the SST-variance in Figure 5.5b,c, these strong SST,SSS-covariances can also be found at the offshore edge of, and within the river plume (see Figure 5.5e,f).

The point-by-point ensemble covariance between SST and the daily-averaged surface currents is calculated for the same dates (Figure 5.6a-c). Since that the current is a vector field, only the maximum covariance at each point is shown. The areas of the largest SST-velocity covariance match the locations of the largest SST variances (cf. Figure 5.5a-c). In contrast, the largest surface current variances (Figure 5.6d-e) are found within the plume. This notwithstanding that the ensemble variability in the

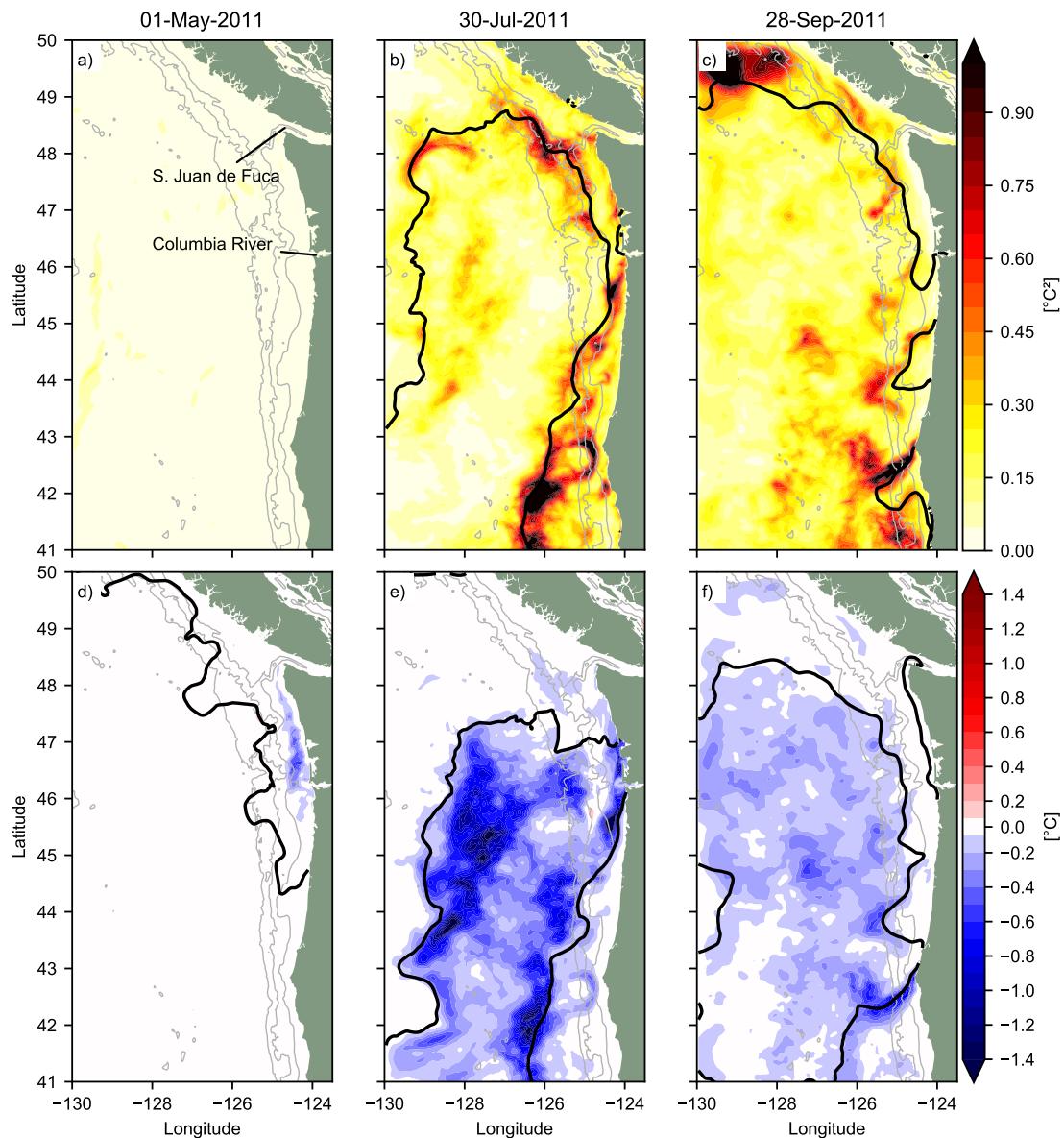


Figure 5.5: Point-by-point surface (a-c) temperature-temperature and (d-f) temperature-salinity ensemble covariance for 1 May (1st column), 30 July (2nd column), 28 September 2011 (3rd column). Solid black lines mark the (a-c) 15  $^{\circ}\text{C}$  isotherm and (d-f) 31.5 isohaline.

wind forcing in this region on these days is actually lower than in the adjacent areas near the coast and the west side of the model domain (not shown). This paradoxical result could be caused by the fact that Ekman depth scales in the plume area are smaller, influenced by the stronger stratification in the shallower mixed layer (*Fong and Geyer, 2001; Gan et al., 2009; McWilliams et al., 2009; Price and Sundermeyer, 1999*). Ekman transport is distributed over a smaller mass of water and consequently currents in the plume are more sensitive to perturbations to the wind forcing.

## 5.6 Columbia River Plume

In the initial En4DVar runs performed without HMSC, we noticed that the extent of the Columbia River plume and generally salinity at the surface occasionally exhibited rapid changes as a result of the DA. To illustrate, we use daily analysis fields and estimate the volume of fresh water in the river plume. To compute this, grid cells for which  $S < 31.5$  are considered only. The amount of fresh water contained in each such grid cell,  $V_{fresh,ijk}$ , is computed using mass conservation:

$$S_{ijk}V_{ijk} = S_{river}V_{fresh,ijk} + S_{ocean}(V_{ijk} - V_{fresh,ijk}), \quad (5.3)$$

where  $S_{river} = 0.3$  is taken as the salinity of the river water,  $S_{ocean} = 32.2$  as the salinity of the oceanic near-surface water, and  $V_{ijk}$  is the grid cell volume. Then  $V_{fresh,ijk}$  is summed over grid cells for which  $S < 31.5$  to obtain the estimate of the total fresh water volume in the river plume  $V_{fresh}$ . Figure 5.7 shows the difference in  $V_{fresh}$  from its value on 19 April 2011, the starting point of all the DA experiments. For reference, we also show the cumulative outflow of the Columbia River since 19 April 2011. As the model forcing does not include evaporation and precipitation, the Columbia River is the only source of fresh water in the plume region. As vertical mixing reduces the fresh water volume within the plume volume (*Hetland, 2005; MacCready et al., 2009*), this cumulative outflow represents an upper bound on  $V_{fresh}$ . However, Figure 5.7 shows that the fresh plume volume in experiment Ens occasionally jumps (e.g. on 3 July, 21 July, 26 August) and exceeds the upper bound

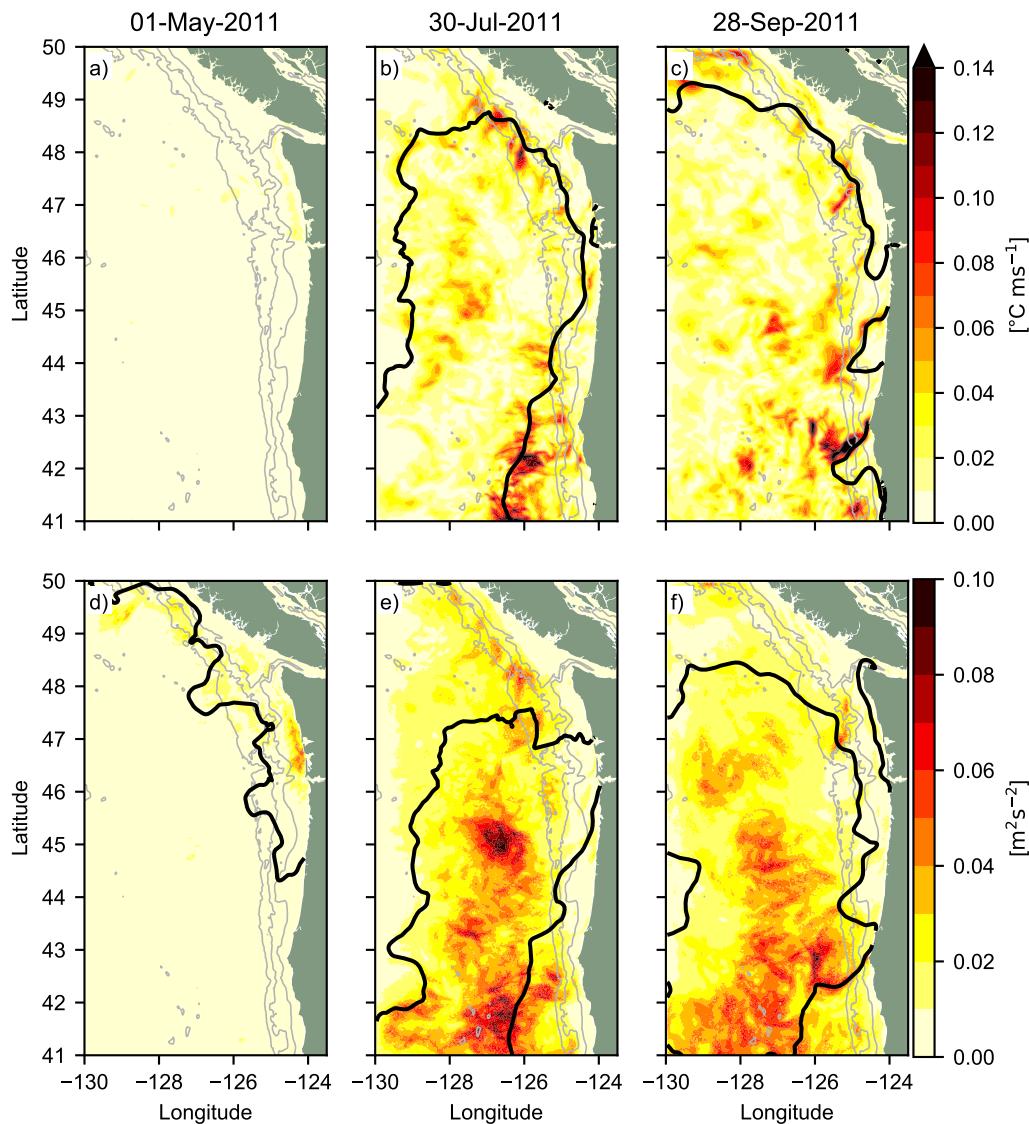


Figure 5.6: (a-c) Maximum point-by-point SST-surface velocity ensemble covariance for 1 May (1st column), 30 July (2nd column), 28 September 2011 (3rd column). (d-f) as (a-c) but now for the ensemble surface velocity variance. Solid black lines mark the (a-c) 15 °C isotherm and (d-f) 31.5 isohaline.

set by the cumulative river outflow. The HBSC is not fully effective in constraining the unphysical instantaneous changes in fresh plume water volume. Concurrently with the jumps in experiment Ens, jumps of smaller amplitude are still noticeable in experiment Ens-HBSC. However, their magnitude is small enough such that the fresh plume water volume difference in experiment Ens-HBSC stays below the cumulative river outflow. The problems with constraining the fresh plume volume are exclusive to ensemble 4DVAR methods: in Experiment Bal-HBSC the change in freshwater volume is comparable to that in experiment No DA.

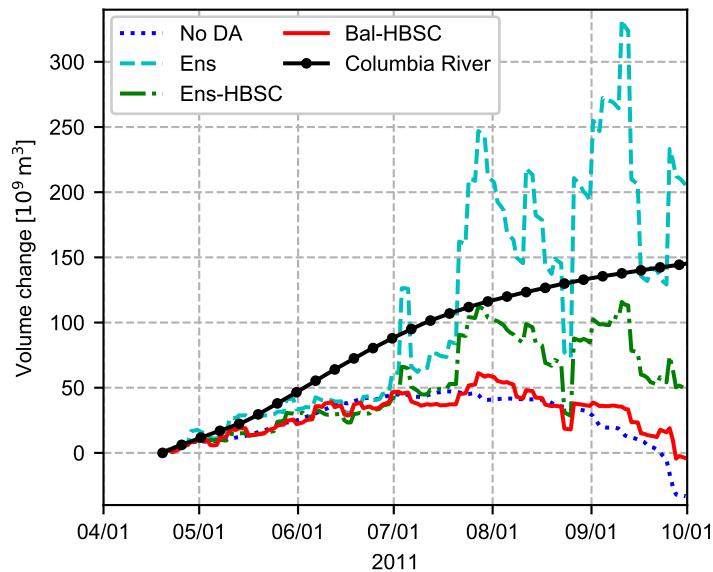


Figure 5.7: Change in fresh water volume in the river plume ( $S < 31.5$ ) since 19 April 2011 in experiment No DA (dark blue) and the analyses from experiments Ens (light blue), Ens-HBSC (green), Bal-HBSC (red). Also shown is the cumulative discharge of the Columbia River since 19 April 2011 (black).

The jumps in fresh plume water volume as shown in experiment Ens are always found at the beginning of the windows and are caused by the instantaneous DA correction. As an example, a particularly large change in fresh plume water volume occurred on 21 July 2011 in experiment Ens. Figure 5.8a,b shows the SSS field before and after the DA correction on this date. Without the additional salinity

constraint, the DA correction expanded the plume area to the west and northwest and decreased its salinity by approximately 3. Further analysis suggests that large erroneous freshwater volume changes are associated with a bias in the forecast SST compared to the assimilated SST data exaggerated by a strong  $T\text{-}S$  covariance in the plume area (Figure 5.8c). On this day, the Ens forecast is colder than the observed SST over the majority of the model domain, on average by  $0.5\text{ }^{\circ}\text{C}$ . If contour lines for the SST,SSS-covariance are overlaid over the SSS DA correction (Figure 5.8d) it is more apparent that the DA-correction related, instantaneous freshening takes place in the area where strong SST-SSS covariances overlap with regions in which the SST forecast is too cold.

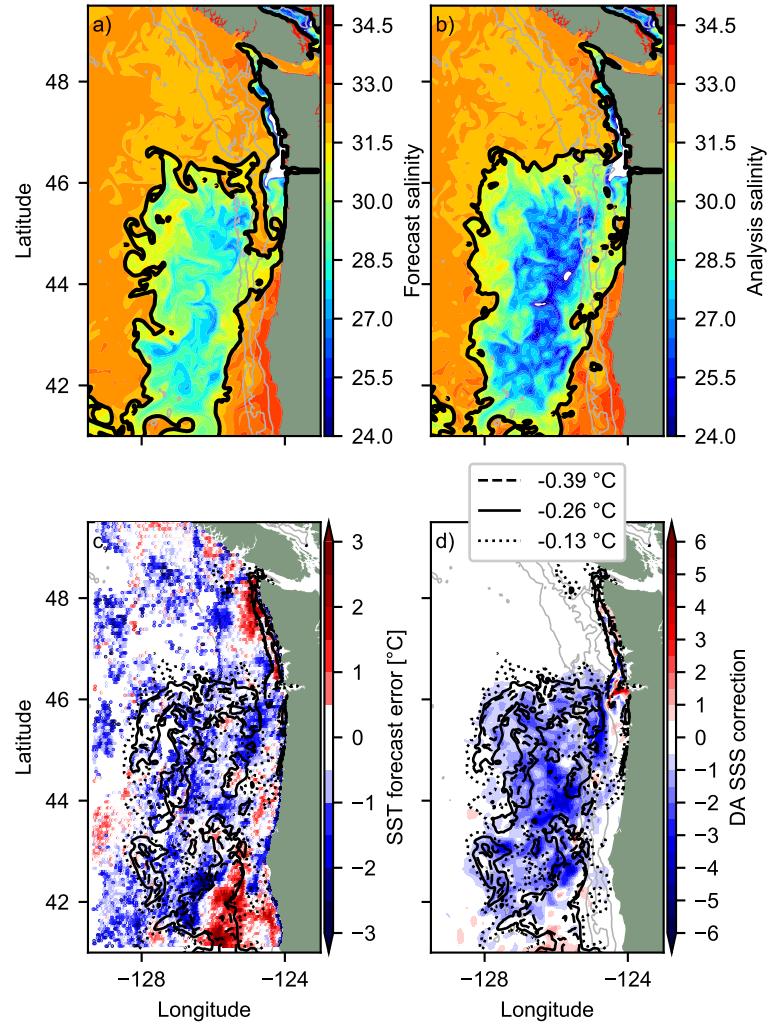


Figure 5.8: The study of DA-induced SSS changes, 21 July 2011: (a) forecast SSS in experiment Ens, prior to DA correction; the black line is  $S = 31.5$ ; (b) analysis SSS, after DA; (c) difference in SST between the Ens forecast and observed SST; and (d) the DA correction to SSS field. In (c) and (d), contours show the point-by-point SST-SSS covariance  $-0.39$ ,  $-0.26$  and  $-0.13$   $^{\circ}\text{C}$ .

## 5.7 Verification

In this section the results from the different experiments are compared to remote and in-situ measurements to find whether En4DVar yields better model analyses and forecasts than the “traditional” 4DVAR with the static covariance currently implemented in the OR-WA forecast system.

### 5.7.1 Surface

Here model results from the experiments are compared with remote surface observations. The same datasets as were assimilated are used in this comparison, but with different processing. The hourly GOES SST observations are not thinned. Instead of using the radial daily-averaged surface velocity HFR components, the HFR maps of the zonal and meridional velocity components on a 6-km grid are used (*Cook and Paduan*, 2001; *Gurgel*, 1994; *Kosro*, 2005). Instead of presenting the along-track altimetry as the 24h-averaged sea level anomaly, it is compared to the instantaneous model SSH minus the along-track mean. Similarly to the SSH used in the assimilation, the model outputs are detided using tidal harmonic constants obtained by the harmonic analysis of the SSH in experiment No DA using *Pawlowicz et al.* (2002). The RMS difference between the measured values for the aforementioned observations and their model predictions will be referred to as the root-mean-square error (RMSE).

Figure 5.9 shows time series of the area-averaged, 3-day time averaged RMSE for each window. The continuous blue line corresponds to the case No DA. Short line segments correspond to analysis-forecast cycles. Analysis RMSEs are consistently smaller than forecast RMSEs and both are smaller than the No DA RMSEs. We note that the gradual increase in RMSE over the study period in all experiments and associate this with the increase in the flow complexity as the upwelling season progresses. Table 5.1 presents a summary of the RMSE averaged over the whole period. All DA runs yield RMSEs that are close to each other and all are an improvement over the case No DA.

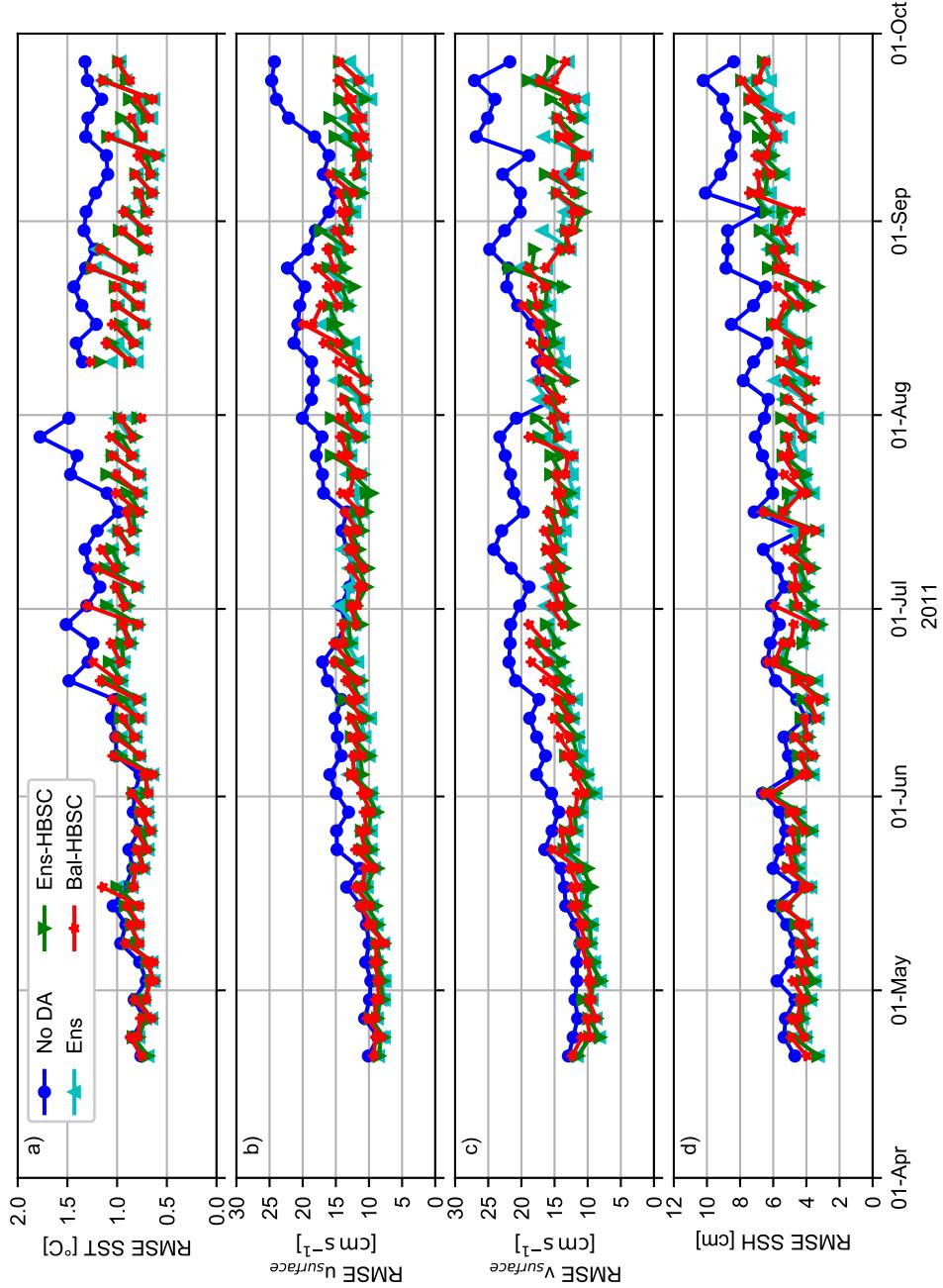


Figure 5.9: Time series of the area-averaged RMSE for (a) SST, (b) daily-averaged zonal velocity, (c) daily-averaged meridional velocity and (d) SSH observations with along-track mean removed. Results from experiment No DA (dark blue), experiment Ens (light blue), experiment Ens-HBSC (green) and Bai-HBSC (red) are shown. Dots at the left (right) side of the DA line segments correspond to the analysis (forecast).

Table 5.1: Time-averaged RMSE, 19 April to 1 October 2011. Smallest analysis and forecast RMSE for each observation type are highlighted.

	No DA	Analysis			Forecast		
		Ens	Ens -HBSC	Bal -HBSC	Ens	Ens -HBSC	Bal -HBSC
SST [°C]	1.15	<b>0.77</b>	0.77	0.78	<b>0.92</b>	0.93	0.96
$u_{surface}$ [cm s <sup>-1</sup> ]	15.9	11.0	<b>10.8</b>	11.5	<b>12.4</b>	12.4	13.0
$v_{surface}$ [cm s <sup>-1</sup> ]	18.9	<b>12.3</b>	12.4	13.0	<b>13.9</b>	14.1	14.6
SSH [cm]	6.8	<b>4.7</b>	4.7	4.9	<b>5.3</b>	5.4	5.5

In Figure 5.10, the left column shows maps of the time-averaged RMSE for case No DA, the center (right) column shows the difference in RMSE between Bal-HBSC and No DA (Ens-HBSC and Bal-HBSC) forecasts. In case Bal-HBSC, the largest reduction in SST RMSE is found along the coast and the southern boundary (Figure 5.10b). Even though the background error variance in SST is on average smaller in the ensemble **B** than in the balanced **B**, experiment Ens-HBSC produces forecasts with a SST RMSE that differs little from that in case Bal-HBSC (Fig. 5.10c). In Bal-HBSC, forecast SSH RMSE gets reduced compared to experiment No DA throughout the model domain (Fig. 5.10d and e). However, scattered between areas of reduction are areas in which SSH RMSE increases with respect to No DA (Figure 5.10e). I.e. there are small scales in the background errors that are not removed in Bal-HBSC. En4DVar is not able to resolve these small-scales either (Figure 5.10f). None of the DA systems is able to reproduce the sharp turn of the coastal current west that was observed south of Cape Blanco, at 42°N (Figure 5.11a). Instead, in the average over the entire study period, the DA systems produce southwestern surface velocities over the region south of Cape Blanco (Figure 5.11c and d). Still, this is better than the sharp frontal southward current in case No DA (Figure 5.11b): the surface velocity RMSE in experiment Bal-HBSC is smaller than that in experiment No DA for most of the region for which surface velocity observations are available (Figure 5.10g and h). From Figure 5.10i, however, it cannot be concluded whether Ens-HBSC performs better or worse than Bal-HBSC in terms of the surface velocity.

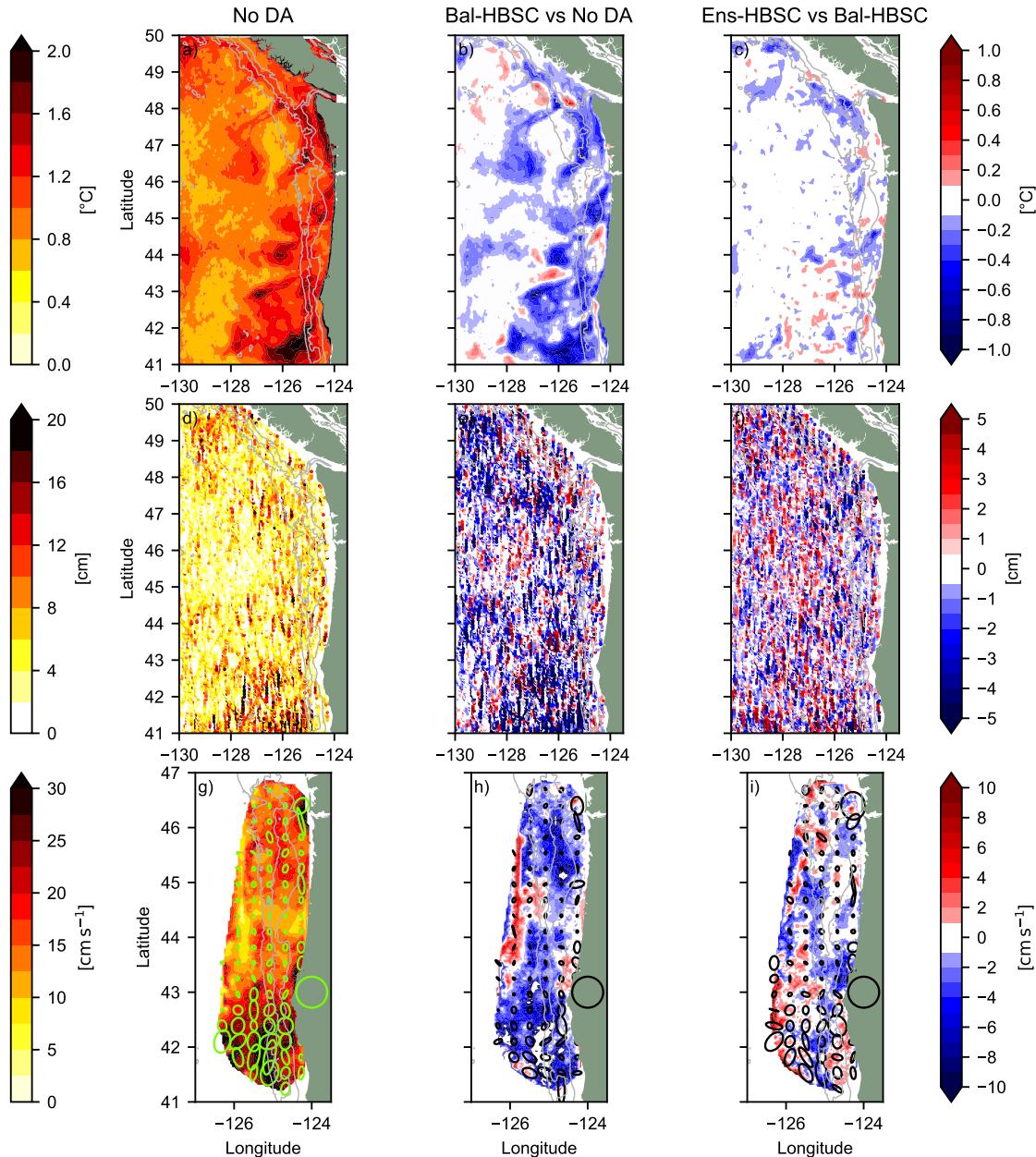


Figure 5.10: Surface maps of time-averaged RMSE. (Top to bottom): SST, SSH, and velocity. (Left) Case no DA RMSE; (center) Bal-HBSC forecast RMSE minus No DA RMSE; (right) Ens-HBSC forecast RMSE minus Bal-HBSC forecast RMSE. Ellipses in the bottom row indicate the second moment (RMSE squared) of the errors in the surface velocities. The scale circle indicates a RMSE squared of  $0.1 \text{ m}^2 \text{s}^{-2}$ . The 200, 1000 and 2000 m isobath contours are shown as grey lines.

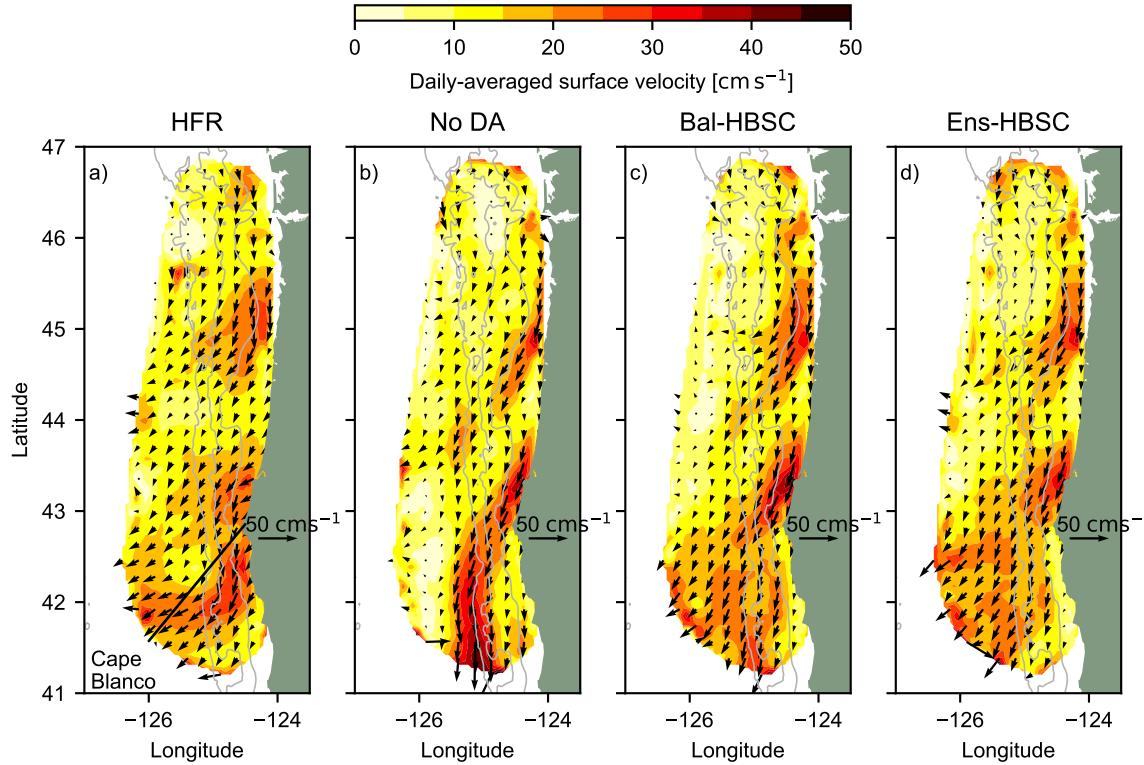


Figure 5.11: Average surface currents from (a) HFR observations, (b) experiment No DA and from the forecasts produced by experiments (c) Bal-HBSC and (d) Ens-HBSC over the period 19 April to 1 October 2011. The 200, 1000 and 2000 m isobaths are shown as grey lines.

The difference in RMSE between experiment Bal-HBSC on one hand and experiments Ens, Ens-HBSC on the other was found to be small in section 5.7.1. This is partly a testament to a reasonably good presently operating OR-WA ocean forecast system, using the static balanced  $\mathbf{B}$ , and partly the result of extensive observational coverage of the surface.

### 5.7.2 Subsurface

In our tests, subsurface observations are not assimilated but used only for verification. In the following we compare model results with independent subsurface observations

to see if using Ens-HBCS yields improvement over Bal-HBSC in any way. Special attention will be paid to salinity observations as this field is not assimilated at all.

On a regular basis temperature and salinity measurements are made by gliders, low power autonomous underwater vehicles, in cross-shore sections along the Newport line, near  $44.65^{\circ}\text{N}$  (*Erofeev*, 2015) (Figure 5.12). On their tracks, Slocum gliders repeatedly descend to depths of 200 m and return to the surface. A single transect (either west-to-east or east-to-west) takes about three days. The transects are located in a region with dynamics that are challenging to model: it partly runs over the continental shelf (depth  $\lesssim 200$  m) just south of the point where the southward coastal current separates from the shore (*Barth et al.*, 2005a; *Kosro*, 2005; *Kurapov et al.*, 2005; *Oke et al.*, 2002c). During episodes of strong upwelling driven by the southward winds, the glider samples the cold and salty upwelling water with a potential density  $> 26.5 \text{ kg m}^{-3}$  in the shallower portion of the transect (*Austin and Barth*, 2002; *Barth et al.*, 2005b; *Huyer*, 1977; *Oke et al.*, 2002b,c). As the glider goes farther offshore, it crosses the Columbia River front, with relatively warmer and fresher water (see chapter 2).

$T$ - $S$  diagrams are presented in Figure 5.13. Plots at the top (bottom) correspond to the first half of the study period, 19 April through 29 June 2011 (the second half of the study period, 30 June through 1 October). In the beginning of the upwelling season (Figure 5.13a) observations show a cloud of points corresponding to the shallow river plume ( $S < 31.5$ ). The  $T$ - $S$  relation in experiment No DA (Figure 5.13b) is qualitatively similar to the observed. A problem emerges in case Bal-HBSC (Figure 5.13c), where a line of points is built along  $T = -\alpha^{-1}S$ . This line is absent in case No DA (Figure 5.13b) and the verification data (Figure 5.13a). The line is enforced by the very simple choice of the  $T$ - $S$  relation in the balance operator, literally assuming for the errors that  $\delta S = -\alpha\delta T$  (E.2). We mark the top portion of this line ( $T > 13.5^{\circ}\text{C}$  in Figure 5.13c) with black dots, and color their counterparts in every plot black as well. The low salinities of the black points in the observations (Figure 5.13a) indicate that the black points correspond to samples taken while the glider was in the river plume. Qualitatively, Ens-HBSC DA is successful in repro-

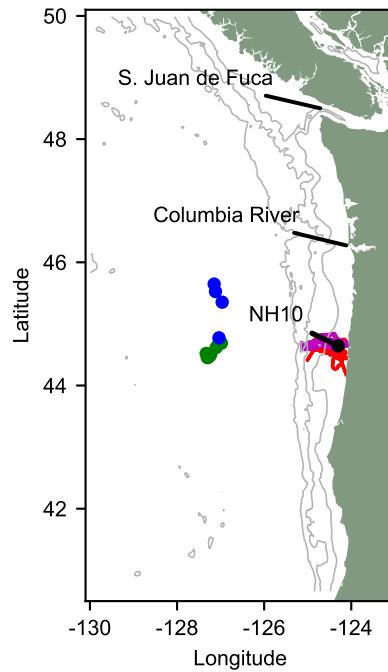


Figure 5.12: Location of in-situ observations. Shown are glider positions prior to 30 June 2011 (red), after 30 June 2011 (purple), Argo floats outside the plume (green) and inside the river plume (blue) as well as the location of the NH10 buoy (black). The 200, 1000 and 2000 m isobath contours are shown as grey.

ducing the plume  $T$ - $S$  relation during this time period (Figure 5.13d); in particular the artificial line implied by the specific choice in the balance operator is gone. After 30 June 2011 northerly, upwelling favorable winds continue to persist and the plume moves farther offshore. Under these conditions the glider samples only inshore of the plume and the observed  $T$ - $S$  relation reverts to a straight line with the slope close to  $-\alpha^{-1}$  (Figure 5.13e). Consequently, the forecasts in experiment Bal-HBSC are able to correctly simulate the observed  $T$ - $S$  relation (Figure 5.13g). The ensemble covariance is sensitive to adapt to the new background oceanic conditions and also yields the correct  $T$ - $S$  diagram along the glider section (Figure 5.13h).

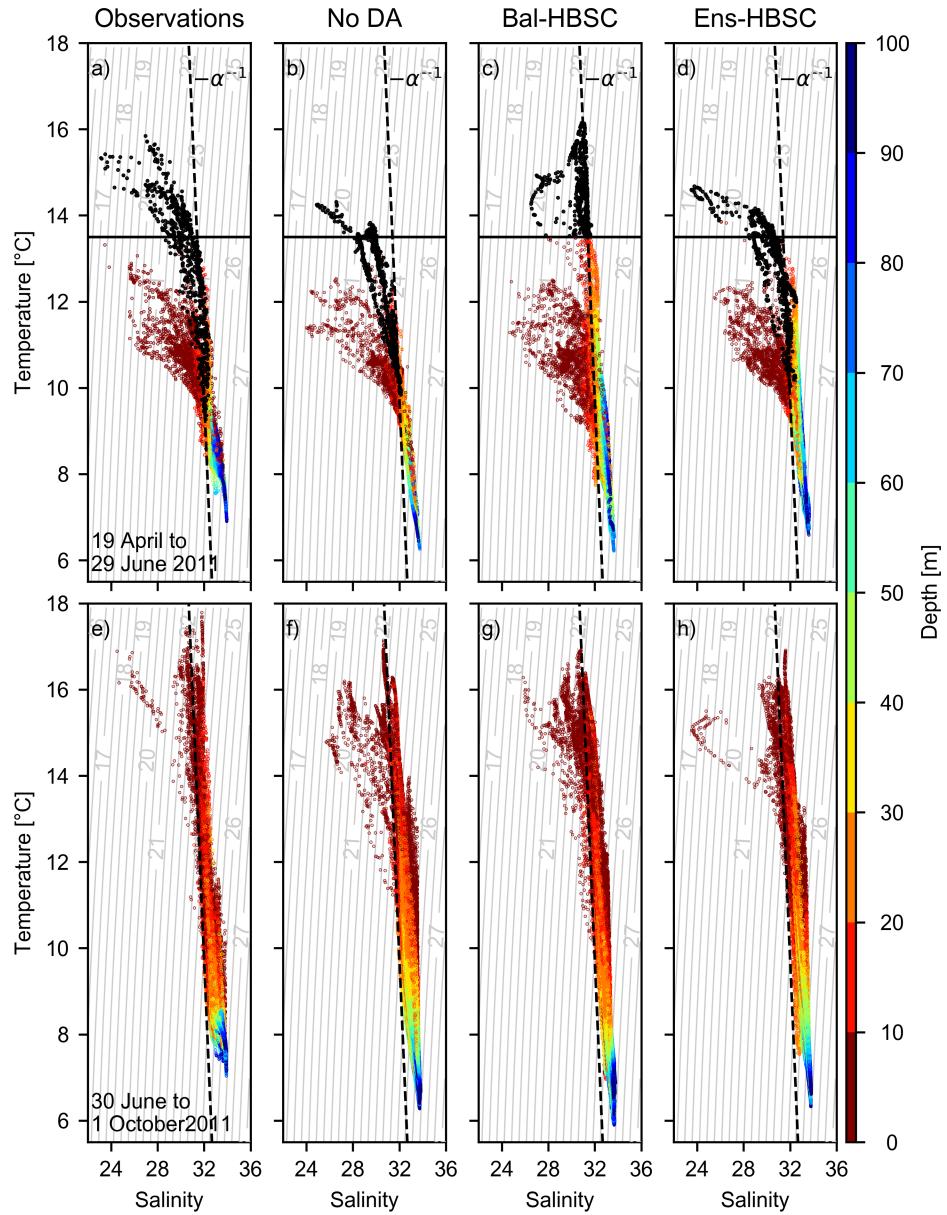


Figure 5.13: Glider-observed and model forecast  $T$ - $S$  diagrams along the glider transect near  $44.65^{\circ}\text{N}$ . (Top) 19 April 00:00 to 30 June 00:00 2011, (bottom) 30 June 00:00 through 1 October 2011 00:00. (Left to right): observed, No DA, Bal-HMSC, Ens-HMSC. Colors indicate the depth at which the observations are taken. Black solid line is  $T = 13.5^{\circ}\text{C}$  and the black dashed line corresponds to the assumed  $\delta T = -\alpha^{-1}\delta S$  relation in the balance operator  $\mathbf{B}$  (E.2).

Mooring NH10 is located on the shelf 10 nautical miles offshore of the Oregon coast at the Newport line, anchored at 81 m below the surface (see Figure 5.12). Temperature and salinity are measured by sensors at different depths. Hourly-averaged measurements are compared with model output from the different experiments. The differences are then filtered with a 24h Bartlett filter as we focus on subtidal time scales. Next, we will discuss the vertical profiles of the time-averaged RMSE and the mean of the differences (i.e., the bias) between the model outputs (analyses and forecasts) and the buoy observations, shown in Figure 5.14 (temperature) and Figure 5.15 (salinity).

The largest temperature RMSE in experiment No DA is found in the top 10 m where surface heating, the river plume, and coastal upwelling all contribute to model uncertainty in  $T$ . Below 10 m the No DA RMSE decreases sharply (Figure 5.14a). However, the magnitude of the model bias increases with depth (Figure 5.14c) as deep waters are colder in the model than in the observations. At the surface, the bias in experiment Ens and Ens-HBSC analyses is higher than in experiment No DA (Figure 5.14c,d) and is equal to the bias in the GOES SST observations compared to the moored temperature measurements (see the “+” mark in the figure). Hence the increase in the bias results from the DA correctly fitting erroneous SST observations (Figure 5.15c). Notwithstanding the larger bias, the RMSE at the surface is comparable to that in case No DA indicating that experiments Ens and Ens-HBSC better capture the variability in the SST. Below the surface, both Ens and Ens-HMSC outperform Bal-HBSC throughout the entire water column in terms of RMSE. These points hold true for both the analyses and forecasts.

Near the surface, experiments Ens and Ens-HBSC yield NH10 salinity analyses and forecast with smaller RMSE and smaller bias magnitude than the other experiments (Fig 5.15). Especially when compared to the forecasts and analyses of experiment Bal-HBSC, which are too fresh over the entire depth range for which NH10 salinity measurements are available (Figure 5.15c,d). However, differences in RMSE and bias become smaller when going further down the water column.

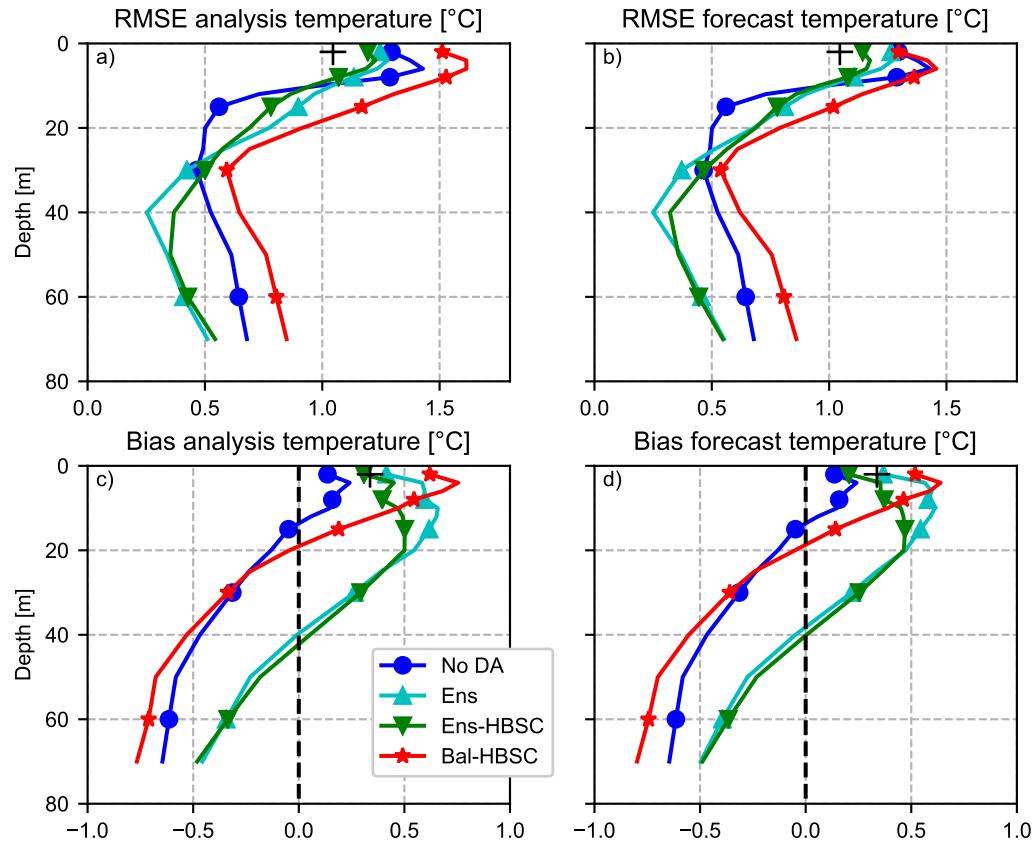


Figure 5.14: Vertical profiles of the time-averaged, 19 April 2011 to 1 October 2011 (Top) RMSE and (bottom) bias in NH10 temperature. (left) model analyses (right) forecasts. Experiments: No DA (dark blue), Ens (light blue), Ens-HBSC (green) and Bal-HBSC (red) over the period 19 April 2011 till 1 October 2011. The “+” symbol marks the RMSE and bias between the GOES SST observations in the 6 km radius around the mooring and the mooring temperature at 2 m depth (i.e. GOES minus mooring temperature).

We note that the case with the balance operator based  $\mathbf{B}$  yields worse RMSEs than the No DA case both for  $T$  and  $S$ . Cases Ens and Ens-HBSC provide improvement over Bal-HBSC and more importantly No DA. The reason for the lackluster performance of the standard 4DVAR setup against the verification shelf mooring data was tracked back to the details of the balance operator covariance (see appendix E). Er-

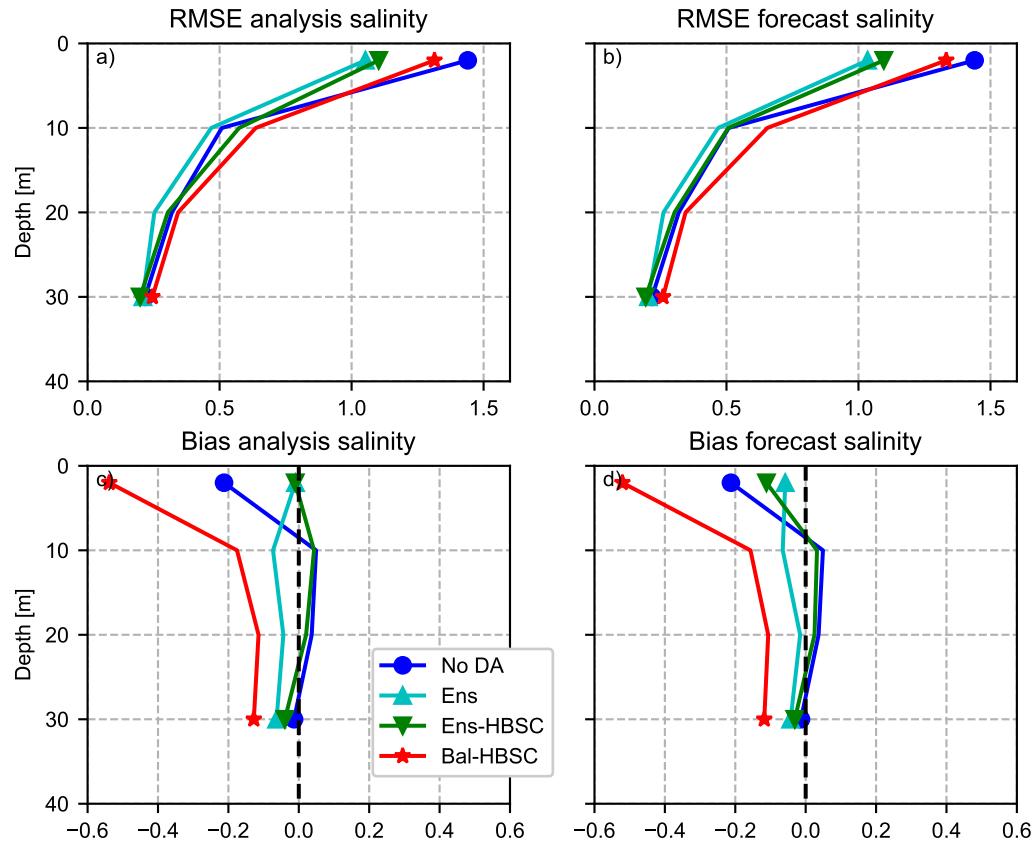


Figure 5.15: As Figure 5.14, but now for salinity.

rors in all the components of the ROMS state vector (SSH, horizontal velocity,  $T$ , and  $S$ ) can be derived if the error statistics for  $T$  and the depth-integrated transport streamfunction are defined. For the latter, for the lack of a better guess, we assumed zero variance. While the assumption that the depth-integrated transport is not changed by DA is reasonable over deep water, it is limiting at the shelf where changes in the surface currents correlate with changes in the depth-averaged current (e.g. *Oke et al.*, 2002b). In case Bal-HBSC, e.g., increase in the alongshore surface current by the DA correction has to be compensated by the decreasing subsurface current which adversely influences variability in the temperature and salinity. In the En4DVar cases, we do not have to make an explicit assumption about the error in the

depth-averaged flow. The ensemble provides the covariance that yields better performance and helps to reveal yet another limitation of the balance operator formulation on the shelf.

While experiments Ens and Ens-HBSC provide the most accurate predictions of all experiments for temperature and salinity at the NH10 location, this is not the case when the model runs are compared to a limited set of Argo floats (*Argo*, 2000) farther away from the coast. Argo floats locations used in this analysis are shown in Figure 5.12. Figure 5.16a and c shows the average  $T$  and  $S$  profiles prior to 20 July 2011, when the profile locations were outside the river plume ( $\text{SSS} > 31.5$ ), and Figure 5.16b and d show the average profile after 20 July 2011 when each of the profiles sampled was located inside the river plume. The forecasts from experiments Ens and Ens-HBSC fail to reproduce the strength of the vertical temperature gradient in the thermocline (Figure 5.16a,b). The deterioration of the Ens, Ens-HBSC forecasts at depths 30-100 m might be the results of random non-zero correlations between SST and other subsurface field which are present in the  $\mathbf{B}$  as no localization is applied in the vertical direction. The latter was not applied since the ensemble error variance generally decreases with depth.

While all experiments correctly reproduce the average salinity profile below 75 m (Figure 5.16c,d) with similar accuracy, experiment Ens-HBSC and Ens produce salinity forecasts that are too fresh above 50 m. Without HBSC constraining the corrections to SSS, the forecasts in experiment Ens produce a river plume with a salinity that is too low at the surface (Figure 5.16d). Applying HBSC is highly beneficial in this case. It increases salinity in the forecasts bringing it closer to the observations. However, it does not completely eliminate the tendency of the ensemble DA system to overestimate the magnitude of the SSS corrections in and around the plume. This in agreement with our findings in section 5.6.

## 5.8 Discussion and Conclusions

In this study the En4DVar DA system described in chapter 4 was tested with the Oregon-Washington coastal ocean circulation model. In this system DA is carried out

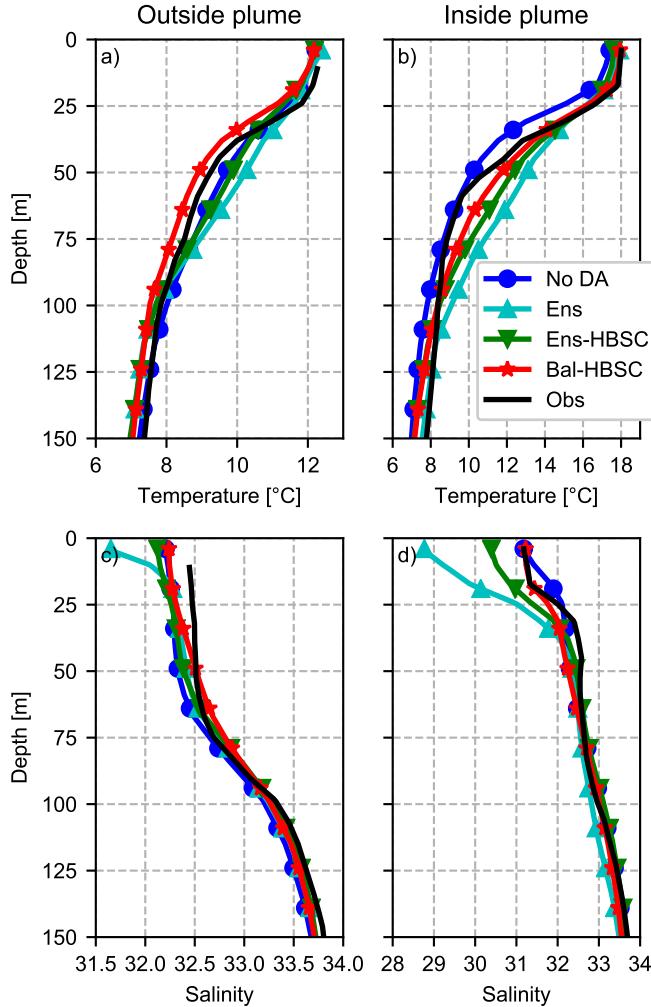


Figure 5.16: Time-average temperature-depth profiles (top row) and salinity-depth profiles (bottom row) from ARGO observations (black) and forecasts from experiment No DA (blue), Ens (light blue), Ens-HBSC (green) and Bal-HBSC (red). Separate average profiles are shown based on profiles taken while the float was outside the river plume (a,c) and while the float was in or beneath the river plume (b,d). See Figure 5.12 for the location of the ARGO floats.

using an ensemble of 4DVAR runs to estimate the background error covariance **B**. The Columbia River plume is a dominant feature in the model region. Its variability

creates a **B** that varies from one assimilation window to the next. The results of the En4DVar system are verified against independent observations and compared with forecasts from a 4DVAR system using a static, balanced **B**.

The ensemble **B** yields  $T$ - $S$  error correlations and covariances that are different from those assumed in the static **B**. In particular, ensemble  $T$ - $S$  error correlations and covariances are weak outside the plume. In and near the plume, however, the magnitude can be a factor 20 larger than in the static **B**. These differences between the static and ensemble **B** have little impact on the accuracy of the surface forecasts produced. Forecast RMSE in Ens-HBSC, compared to the formally future surface data, are only slightly ( $\sim$ 2-4%) smaller than in experiment Bal-HBSC.

Comparison with glider observations showed that using the ensemble **B** yields corrections that can reproduce the spatial structure and temporal variability in the  $T$ - $S$  space. This comparison helped us reveal one of the deficiencies of the presently used balance operator in which a very simple  $T$ - $S$  relation is assumed. While it produces reasonable results later in the study period, when the model and observed points are all on one line, it fails to reproduce the observed  $T$ - $S$  relationship earlier in the study period, when water of different origins (including the river plume) is present in the area sampled by the glider. The reason our balance operator **B** used such a simple  $T$ - $S$  relation is that more complicated approaches, e.g., using the forecast state to provide a more realistic and spatially varying  $T$ - $S$  relation, resulted in unstable results. The problem revealed here is an encouragement to re-evaluate the approach used to build the balance operator in future applications. Similarly, comparison against the mooring  $T$  and  $S$  profile data not only showed the advantage of using the ensemble error covariance, but also pointed to yet another problem with the balance operator where improvements must be made: the correction to the depth-integrated transport on the shelf.

The Hierarchical Box Salinity Constraint (HBSC) was added as an additional term to the penalty function to inhibit large and erroneous variations in the surface salinity caused by the assimilation of the biased SST data, amplified by the large ensemble  $T$ - $S$  covariance. Such an issue is specific only to the En4DVar and did not present a

problem in the case with the balance operator  $\mathbf{B}$ . The HBSC was found to suppress, but not completely eliminated the erroneous SSS corrections. Several measures could possibly improve this. First, the weights  $\sigma_{S,i}^{-2}$  in (5.2) can be increased, especially for the larger boxes. Second, a bias correction scheme (e.g. *Dee and Uppala*, 2009; *Derber and Wu*, 1998; *Donlon et al.*, 2012; *Lea et al.*, 2008; *Lellouche et al.*, 2013) that removes large-scale, systematic errors in the observations, can be tried. However, the SST outliers were not persistently present, but only manifested themselves during a select number of times during the window. This suggest outliers due, e.g., errors in cloud masking, error dependence on the sun zenith angle, and the difference between the skin temperature, which is measured by the satellite, and the surface temperature, which is used in the model. Future studies will demonstrate if advanced outlier algorithms (e.g. *Anderson and Jrvinen*, 1999; *Ingleby and Lorenc*, 1993; *Tavolato and Isaksen*, 2015) are capable of dealing with this. We expect that the problem of amplification of large and erroneous DA corrections to unassimilated fields due to strong ensemble error covariances will not be limited to salinity. For example, it may also concern poorly constrained biochemical fields in coupled biogeochemical model applications (*Ciavatta et al.*, 2011, 2014).

In this study only DA systems that use either a purely static or a purely ensemble  $\mathbf{B}$  have been compared. In meteorology it was found that 4DVAR systems in which a combination of a static and ensemble covariances are used outperform these pure ensemble or pure static systems (*Clayton et al.*, 2013; *Kuhl et al.*, 2013; *Lorenc and Jardak*, 2018). Apart from finding a suitable weighting, there are no impediments to implementing such a hybrid scheme in the En4DVar system described here in the future. Performance of a DA system using the static  $\mathbf{B}$  will also depend on the formulation details. In the ensemble  $\mathbf{B}$  strong covariances are located near the fronts and this spatial and temporal variability could possibly be reproduced in the static covariance without running the ensemble by estimating the point-by-point covariance at each point from the spatial variability around the point (*Fu et al.*, 1993). Furthermore, the assumption that the  $T$ - $S$  correlation is -1 was found to contradict the  $T$ - $S$  ensemble correlation. This suggests that the addition of an unbalanced, univariate

part to the balance operator  $\mathbf{B}$  (*Derber and Boutrier, 1999*) can help to produce a more realistic static  $\mathbf{B}$ .

While research toward the best representation of the background model errors in  $\mathbf{B}$  must be continued, without doubt, at the coastal and regional scale we operate in the data-hungry environment. Any future efforts to sample surface and subsurface fields at an ever improving spatial and temporal resolution are key to improved prediction.

## Chapter 6: Conclusions

The Oregon-Washington (OR-WA) coastal ocean forecast system currently implemented at Oregon State University assimilates satellite sea-surface temperature (SST), sea-surface height (SSH) and daily-averaged high-frequency radar surface velocities (HFR) using 4DVAR. The background error covariance matrix **B** is based on the balance operator (*Kurapov et al.*, 2011; *Weaver et al.*, 2005) and is static. In this thesis, two possible improvements to the OR-WA system to overcome several current limitations have been tested: (i) the assimilation of subsurface in-situ glider temperature and salinity observations and (ii) the estimation of the background error covariance **B** from an ensemble of 4DVAR cycles to account for the evolution of the background error statistics over time. The impact of these advancements on the analyses and forecasts produced by the system has been evaluated. Particular attention has been paid to the role of the background error covariance.

### 6.1 Conclusions

- *Conclusion I: assimilation of glider observations in coastal ocean models, as well as any other in-situ profiles, has to be done with care.* It is hypothesized that differences between the background and subsurface glider observations can be indicative of a large-scale error in the depth of the isopycnal surfaces or smaller, more local errors due to eddies in the model. Horizontal scales in **B** are representative of the eddies. Assimilation of the glider observations with this **B** can possibly result in scale aliasing (*Ooyama*, 1987) and as a result create unphysical corrections. In chapter 2, it is shown that indeed the assimilation of glider observations by themselves results in unphysical eddies that deteriorate the analyses and forecasts. Combined assimilation of the glider observations with surface observations, especially SST observations with their extensive spa-

tial coverage, is effective in suppressing these unphysical corrections. However, no evidence was found that combined assimilation can force the DA system to make large-scale (region-scale) subsurface corrections.

- *Conclusion II: En4DVar is operationally feasible with only a moderate amount of computational resources provided computationally efficient algorithms are used.* The En4DVar system contains several components that can take a considerable amount of computational resources and time to execute. Among these are the forward propagation of the ensemble members using the nonlinear ocean model, the localization of the ensemble covariance, and the iterative minimization of the 4DVAR cost function. For the last two components, efficient algorithms were developed to enable execution using the moderate computational resources available to us. In chapter 3 it has been shown that a combination of modulated ensemble members from *Bishop and Hodyss* (2011) and a Monte-Carlo approximation can speed up the localization. The cluster search method in chapter 4 does not only allow to calculate the DA correction for all ensemble members simultaneously but also reduces the wall time needed to minimize the cost function.
- *Conclusion III: the ensemble  $\mathbf{B}$  is variable in space and time.* According to the results in chapter 5, the magnitude of the ensemble covariances is generally 1-2 orders of magnitude smaller than the ones assumed in the static  $\mathbf{B}$ , except in the river plume and frontal regions where covariances, especially the temperature-salinity covariance, can be up to a factor 20 stronger than in the static  $\mathbf{B}$ . As the shape of the plume and the upwelling front change over time, this introduces substantial time variability in the ensemble  $\mathbf{B}$ . The inclusion of this variability in the 4DVAR resulted in slightly better forecasts for the SST, SSH and HFR observations and a more accurate representation of the time-varying temperature-salinity properties along the Newport glider line.
- *Conclusion IV: strong covariances generated by En4DVar in areas of high variability can have overly strong and erroneous impact on unobserved fields.* In

chapter 5 it was shown that due to a combination of strong surface temperature-salinity covariances in the river plume area and large-scale errors with respect to satellite SST observations, strong erroneous DA corrections are applied to the unobserved salinity field. Penalizing DA corrections to the forecast model (near-)surface salinity field in the cost function was found to be a partially successful method for reducing the erroneous salinity corrections.

## 6.2 Outlook

The foregoing conclusions suggest that the following two issues need to be addressed in future work: (i) the development of ocean DA systems that can provide DA corrections at a wide range scales and (ii) the development of a robust ensemble ocean DA system insensitive to outliers or biases in the assimilated observations. *Li et al.* (2015a) perform separate 4DVARs for the large-scale and small-scale error fields. In this case, the small-scale errors are separated from the large-scale errors by spatial averaging the background-observation differences. As an alternative, the 4DVAR cost function minimization could be carried out using a multigrid method similar to *Lau- rent et al.* (2015). In this method, the minimization of the 4DVAR cost function is carried out sequentially on model grids with different resolutions. This would not only eliminate the need to specify a separate  $\mathbf{B}$  for the small and large scales as in *Li et al.* (2015a), but it also has the potential to speed up the cost function minimization as part of the minimization is carried out on coarser model grids.

In this study, we tried to suppress the unphysical corrections to the surface salinity caused by the errors in the SST observations by adding penalty terms to the cost function. A preferential approach would be to eliminate the erroneous SST observations altogether. However, this will not be easy. The erroneous SST observations are not permanently present and are not incidental outlier points. Studies into methods that provide robust estimates of poorly observed fields must be continued.

## Bibliography

- Adams, K. A., J. A. Barth, and C. Francis (2013), Temporal variability of nearbottom dissolved oxygen during upwelling off central Oregon, *Journal of Geophysical Research: Oceans*, 118(10), 4839–4854, doi:10.1002/jgrc.20361.
- Allen, J. S., P. A. Newberger, and J. Federiuk (1995), Upwelling circulation on the Oregon continental shelf. Part I: Response to idealized forcing, *Journal of Physical Oceanography*, 25(8), 1843–1866, doi:10.1175/1520-0485(1995)025<1843:UCOTOC>2.0.CO;2.
- Amezcuia, J., M. Goodliff, and P. J. V. Leeuwen (2017), A weak-constraint 4DEnsembleVar. Part I: Formulation and simple model experiments, *Tellus A: Dynamic Meteorology and Oceanography*, 69(1), 1271564, doi:10.1080/16000870.2016.1271564.
- Anderson, E., and H. Jrvinen (1999), Variational quality control, *Quarterly Journal of the Royal Meteorological Society*, 125(554), 697–722, doi:10.1002/qj.4971255416.
- Anderson, J. L. (2001), An ensemble adjustment Kalman filter for data assimilation, *Monthly Weather Review*, 129(12), 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- Anderson, J. L. (2007), Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, *Journal of Physics D: Applied Physics*, 230(12), 99–111, doi:10.1016/j.physd.2006.02.011.
- Anderson, J. L. (2012), Localization and sampling error correction in ensemble Kalman filter data assimilation, *Monthly Weather Review*, 140(7), 2359–2371, doi:10.1175/MWR-D-11-00013.1.
- Anderson, J. L., and S. L. Anderson (1999), A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, *Monthly Weather Review*, 127(12), 2741–2758, doi:10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2.

- Argo (2000), Argo float data and metadata from Global Data Assembly Centre (Argo GDAC), april 2011 to september 2011, doi.org/10.17882/42182, SEANOE, Monterey, CA.
- Auligné, T., B. Ménétrier, A. C. Lorenc, and M. Buehner (2016), Ensemble-variational integrated localized data assimilation, *Monthly Weather Review*, 144(10), 3677–3696, doi:10.1175/MWR-D-15-0252.1.
- Austin, J. A., and J. A. Barth (2002), Variation in the position of the upwelling front on the Oregon shelf, *Journal of Geophysical Research: Oceans*, 107(C11), 1–15, doi:doi.org/10.1029/2001JC000858.
- Banas, N. S., L. Conway-Cranos, D. A. Sutherland, P. MacCready, P. Kiffney, and M. Plummer (2015), Patterns of river influence and connectivity among subbasins of Puget Sound, with application to bacterial and nutrient loading, *Estuaries and Coasts*, 38(3), 735–753, doi:10.1007/s12237-014-9853-y.
- Barth, J. A., S. D. Pierce, and R. L. Smith (2000), A separating coastal upwelling jet at Cape Blanco, Oregon and its connection to the California Current System, *Deep Sea Research Part II: Topical Studies in Oceanography*, 47(5), 783–810, doi:10.1016/S0967-0645(99)00127-7.
- Barth, J. A., S. D. Pierce, and R. M. Castelao (2005a), Time-dependent, wind-driven flow over a shallow midshelf submarine bank, *Journal of Geophysical Research: Oceans*, 110(C10), doi:10.1029/2004JC002761.
- Barth, J. A., S. D. Pierce, and T. J. Cowles (2005b), Mesoscale structure and its seasonal evolution in the northern California Current System, *Deep Sea Research Part II: Topical Studies in Oceanography*, 52(1), 5–28, doi:10.1016/j.dsr2.2004.09.026.
- Bennett, A. F. (1992), *Inverse Methods of the Ocean and Atmosphere*, Cambridge University Press, Cambridge.
- Berdeal, I. G., B. M. Hickey, and M. Kawase (2002), Influence of wind stress and ambient flow on a high discharge river plume, *Journal of Geophysical Research: Oceans*, 107(C9), 13–1–13–24, doi:10.1029/2001JC000932.
- Berre, L. (2000), Estimation of synoptic and mesoscale forecast error covariances in a limited-area model, *Monthly Weather Review*, 128(3), 644–667, doi:10.1175/1520-0493(2000)128<0644:EOSAMF>2.0.CO;2.

- Berre, L., H. Varella, and G. Desroziers (2015), Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France, *Quarterly Journal of the Royal Meteorological Society*, 141(692), 2803–2812, doi:10.1002/qj.2565.
- Bishop, C. H., and D. Hodyss (2007), Flow-adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 133(629), 2029–2044, doi:10.1002/qj.169.
- Bishop, C. H., and D. Hodyss (2009a), Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models, *Tellus A: Dynamic Meteorology and Oceanography*, 61(1), 84–96, doi:10.1111/j.1600-0870.2008.00371.x.
- Bishop, C. H., and D. Hodyss (2009b), Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere, *Tellus A: Dynamic Meteorology and Oceanography*, 61(1), 97–111, doi:10.1111/j.1600-0870.2008.00372.x.
- Bishop, C. H., and D. Hodyss (2011), Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation, *Monthly Weather Review*, 139(4), 1241–1255, doi:10.1175/2010MWR3403.1.
- Bishop, C. H., B. J. Etherton, and S. J. Majumdar (2001), Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects, *Monthly Weather Review*, 129(3), 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.
- Bishop, C. H., D. Hodyss, P. Steinle, H. Sims, A. M. Clayton, A. C. Lorenc, D. M. Barker, and M. Buehner (2011), Efficient ensemble covariance localization in variational data assimilation, *Monthly Weather Review*, 139(2), 573–580, doi:10.1175/2010MWR3405.1.
- Bocquet, M., C. A. Pires, and L. Wu (2010), Beyond Gaussian statistical modeling in geophysical data assimilation, *Monthly Weather Review*, 138(8), 2997–3023, doi:10.1175/2010MWR3164.1.
- Bonavita, M., L. Isaksen, and E. Hólm (2012), On the use of EDA background error variances in the ECMWF 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, 138, 1540–1559, doi:10.1002/qj.1899.
- Brezinski, C. (1999), Multiparameter descent methods, *Numerical Linear Algebra with Applications*, 296(1), 113–141, doi:10.1016/S0024-3795(99)00112-3.

- Brink, K. H. (1987), Upwelling fronts: Implications and unknowns, *South African Journal of Marine Science*, 5(1), 3–9, doi:10.2989/025776187784522315.
- Buehner, M. (2005), Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting, *Quarterly Journal of the Royal Meteorological Society*, 131, 1013–1043, doi:10.1256/qj.04.15.
- Buehner, M. (2012), Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation, *Monthly Weather Review*, 140(2), 617–636, doi:10.1175/MWR-D-10-05052.1.
- Buehner, M., and M. Charron (2007), Spectral and spatial localization of background-error correlations for data assimilation, *Quarterly Journal of Royal Meteorological Society*, 133(624), 615–630, doi:10.1002/qj.50.
- Buehner, M., and A. Shlyaeva (2015), Scale-dependent background-error covariance localisation, *Tellus A: Dynamic Meteorology and Oceanography*, 67(1), doi:10.3402/tellusa.v67.28027.
- Buehner, M., P. L. Houtekamer, C. Charette, H. L. Mitchell, and B. He (2009), Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations, *Monthly Weather Review*, 138(5), 1567–1586, doi:10.1175/2009MWR3158.1.
- Burgers, G., P. Jan van Leeuwen, and G. Evensen (1998), Analysis scheme in the ensemble Kalman filter, *Monthly Weather Review*, 126(6), 1719–1724, doi:10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.
- Burnett, W., S. Harper, R. Preller, G. A. Jacobs, and K. LaCroix (2014), Overview of operational ocean forecasting in the U.S. Navy: Past, present and future, *Tech. Rep. NRL/JA/7300-14-2038*, Naval Research Laboratory.
- Casella, G., and E. I. George (1992), Explaining the Gibbs Sampler, *The American Statistician*, 46(3), 167–174, doi:10.2307/2685208.
- Castelao, R. M., and J. A. Barth (2005), Coastal ocean response to summer upwelling favorable winds in a region of alongshore bottom topography variations off Oregon, *Journal of Geophysical Research: Oceans*, 110(C10), C10S04, doi:10.1029/2004JC002409.

- Caya, A., J. Sun, and C. Snyder (2005), A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation, *Monthly Weather Review*, 133(11), 3081–3094, doi:10.1175/MWR3021.1.
- Chapman, A., and Y. Saad (1996), Deflated and augmented Krylov subspace techniques, *Numerical Linear Algebra with Applications*, 4(1), 43–66, doi:10.1002/(SICI)1099-1506(199701/02)4:1<3C43::AID-NLA99>3E3.0.CO;2-Z.
- Chapman, D. C. (1985), Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model, *Journal of Physical Oceanography*, 15(8), 1060–1075, doi:10.1175/1520-0485(1985)015<1060:NTOCSO>2.0.CO;2.
- Chelton, D. B., R. A. De Szoeke, M. G. Schlax, K. El Naggar, and N. Siwertz (1998), Geographical variability of the first baroclinic Rossby radius of deformation, *Journal of Physical Oceanography*, 28(3), 433–460, doi:10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2.
- Chin, T. M., R. F. Milliff, and W. G. Large (1998), Basin-scale, high-wavenumber sea surface wind fields from a multiresolution analysis of scatterometer data, *Journal of Atmospheric and Oceanic Technology*, 15(3), 741–763, doi:10.1175/1520-0426(1998)015<0741:BSHWSS>2.0.CO;2.
- Chua, B. S., and A. F. Bennett (2001), An inverse ocean modeling system, *Ocean Modelling*, 3(3), 137–165, doi:10.1016/S1463-5003(01)00006-3.
- Ciavatta, S., R. Torres, S. SauxPicart, and J. I. Allen (2011), Can ocean color assimilation improve biogeochemical hindcasts in shelf seas?, *Journal of Geophysical Research: Oceans*, 116(C12), doi:10.1029/2011JC007219.
- Ciavatta, S., R. Torres, V. Martinez-Vicente, T. Smyth, G. DallOlmo, L. Polimene, and J. I. Allen (2014), Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling, *Progress in Oceanography*, 127, 74–95, doi:10.1016/j.pocean.2014.06.002.
- Clayton, A. M., A. C. Lorenc, and D. M. Barker (2013), Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, *Quarterly Journal of the Royal Meteorological Society*, 139(675), 1445–1461, doi:10.1002/qj.2054.

- COAPS (2015), HYCOM + NCODA global 1/12° analysis, Version GLBa0.08, january 2011 to december 2011, <https://hycom.org/data/glba0pt08>, Center for Ocean-Atmospheric Prediction Studies, Tallahassee, FL.
- COAS (2016), Undersea gliders: Valuable new tool in marine research, Available from: <https://gliderfs2.coas.oregonstate.edu/gliderweb/docs/COAS50.pdf> (Accessed 15 November 2018).
- Cohen, A., I. Daubechies, and P. Vial (1993), Wavelets on the interval and fast wavelet transforms, *Applied and computational harmonic analysis*, 1(1), 54–81, doi:10.1006/acha.1993.1005.
- Connolly, T. P., B. M. Hickey, S. L. Geier, and W. P. Cochlan (2010), Processes influencing seasonal hypoxia in the northern California Current System, *Journal of Geophysical Research*, 115, C03021, doi:10.1029/2009JC005283.
- Cook, M. S., and J. D. Paduan (2001), Processing HF radar data using the HFRadarmap software system, in *Radiowave Oceanography: The First International Workshop*, edited by J. Paduan, pp. 12–16, Miami University, Miami, FL.
- Cookson-Hills, P., D. J. Kirshbaum, M. Surcel, J. G. Doyle, L. Fillion, D. Jacques, and S.-J. Baek (2017), Verification of 24-h quantitative precipitation forecasts over the Pacific Northwest from a high-resolution ensemble Kalman filter system, *Weather Forecasting*, 32(3), 1185–1208, doi:10.1175/WAF-D-16-0180.1.
- Courtier, P. (1997), Dual formulation of four-dimensional variational assimilation, *Quarterly Journal of the Royal Meteorological Society*, 123, 2449–2461, doi:10.1002/qj.49712354414.
- Courtier, P., J.-N. Thépaut, and A. Hollingsworth (1994), A strategy for operational implementation of 4D-Var, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society*, 120(519), 1367–1387, doi:10.1002/qj.49712051912.
- Daget, N., A. T. Weaver, and M. A. Balmaseda (2009), Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean, *Quarterly Journal of the Royal Meteorological Society*, 135(641), 1071–1094, doi:10.1002/qj.412.
- Daley, R. (1981), Normal mode initialization, *Reviews of Geophysics*, 19(3), 450–468, doi:10.1029/RG019i003p00450.

- Deckmyn, A., and L. Berre (2005), A wavelet approach to representing background error covariances in a limited-area model, *Monthly Weather Review*, 133(5), 1279–1294, doi:10.1175/MWR2929.1.
- Dee, D. P., and S. Uppala (2009), Variational bias correction of satellite radiance data in the ERA-Interim reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 135(644), 1830–1841, doi:10.1002/qj.493.
- Derber, J., and F. Bouttier (1999), A reformulation of the background error covariance in the ECMWF global data assimilation system, *Tellus A: Dynamic Meteorology and Oceanography*, 51(2), 195–221, doi:10.1034/j.1600-0870.1999.t01-2-00003.x.
- Derber, J. C., and W.-S. Wu (1998), The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system, *Monthly Weather Review*, 126(8), 2287–2299, doi:10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2.
- Desroziers, G., and L. Berre (2012), Accelerating and parallelizing minimizations in ensemble and deterministic variational assimilations, *Quarterly Journal of the Royal Meteorological Society*, 138(667), 1599–1610, doi:10.1002/qj.1886.
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis-error statistics in observation space, *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3385–3396, doi:10.1256/qj.05.108.
- Desroziers, G., J. Camino, and L. Berre (2014), 4DEnVar: Link with 4D state formulation of variational assimilation and different possible implementations, *Quarterly Journal of the Royal Meteorological Society*, 140(684), 2097–2110, doi:10.1002/qj.2325.
- Dettmering, D., and W. Bosch (2010), Global calibration of Jason-2 by multi-mission crossover analysis, *Marine Geodesy*, 33(sup1), 150–161, doi:10.1080/01490419.2010.487779.
- Dimet, F.-X. L., and O. Talagrand (1986), Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, *Tellus A: Dynamic Meteorology and Oceanography*, 38(2), 97–110, doi:10.3402/tellusa.v38i2.11706.
- Dobricic, S., N. Pinardi, P. Testor, and U. Send (2010), Impact of data assimilation of glider observations in the Ionian Sea (Eastern Mediterranean), *Dynamics of Atmospheres and Oceans*, 50(1), 78–92, doi:10.1016/j.dynatmoce.2010.01.001.

- Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer (2012), The operational sea surface temperature and sea ice analysis (OSTIA) system, *Remote Sensing of Environment*, 116, 140–158, doi:10.1016/j.rse.2010.10.017.
- Egbert, G. D., and S. Y. Erofeeva (2002), Efficient inverse modeling of barotropic ocean tides, *Journal of Atmospheric and Oceanic Technology*, 19(2), 183–204, doi: 10.1175/1520-0426(2002)019<0183:EIMOBO>2.0.CO;2.
- Egbert, G. D., and S. Y. Erofeeva (2010), TPXO atlas, Version 7.2, <http://volkov.oce.orst.edu/tides/global.html>, Oregon State University, Corvallis, OR.
- Egbert, G. D., A. F. Bennett, and M. G. G. Foreman (1994), TOPEX/POSEIDON tides estimated using a global inverse model, *Journal of Geophysical Research*, 99(C12), 24821–24852.
- Engelberg, S. (2007), *Random Signals and Noise: A Mathematical Introduction*, CRC Press, Boca-Rotan, FL.
- Environment Canada (2015), Historical hydrometric data, <https://wateroffice.ec.gc.ca>, Environment Canada, Gatineau, QC (Updated daily).
- Erhel, J., and F. Guyomarc'h (2000), An augmented conjugate gradient method for solving consecutive symmetric positive definite linear systems, *SIAM Journal of Matrix Analysis & Applications*, 21(4), 1279–1299, doi:10.1137/S0895479897330194.
- Erofeev, A. (2015), Gridded Slocum glider data ( 1km × 4m ), March 2011 to November 2011, <https://gliderfs2.coas.oregonstate.edu/gliderweb/archive/gridded/2011/>, Oregon State University, Corvallis, OR.
- Erofeeva, L. (2018), The Oregon Washington coastal ocean forecast system: The interactive viewer, Available from: <http://ingria.coas.oregonstate.edu/rtdavow/index.html> (Accessed 17 April 2018).
- Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, 99(5), doi:10.1029/94JC00572.
- Evensen, G., and P. J. van Leeuwen (2000), An ensemble Kalman smoother for nonlinear dynamics, *Monthly Weather Review*, 128(6), 1852–1867, doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2.

- Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson (2003), Bulk parameterization of airsea fluxes: Updates and verification for the COARE algorithm, *Journal of Climate*, 16(4), 571–591, doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.
- Fairbairn, D., S. R. Pring, A. C. Lorenc, and I. Roulstone (2014), A comparison of 4DVar with ensemble data assimilation methods, *Quarterly Journal of the Royal Meteorological Society*, 140, 281–294, doi:10.1002/qj.2135.
- Figa-Saldaa, J., J. J. W. Wilson, E. Attema, R. Gelsthorpe, M. R. Drinkwater, and A. Stoffelen (2002), The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: A follow on for European wind scatterometers, *Canadian Journal of Remote Sensing*, 28(3), 404–412, doi:10.5589/m02-035.
- Fisher, M., and E. Andersson (2001), Developments in 4D-Var and Kalman filtering, *Technical memorandum no. 347*, ECMWF, Reading.
- Fisher, M., and S. Gürol (2017), Parallelization in the time dimension of four-dimensional variational data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 143(703), 1136–1147, doi:10.1002/qj.2997.
- Flather, R. A. (1976), A tidal model of the northwest European continental shelf, *Mémoires de la Société Royale des Sciences Liège*, 10, 141–164.
- Fong, D. A., and W. R. Geyer (2001), Response of a river plume during an upwelling favorable wind event, *Journal of Geophysical Research: Oceans*, 106(C1), 1067–1084, doi:10.1029/2000JC900134.
- Fu, L.-L., and E. Rodriguez (2004), High-resolution measurement of ocean surface topography by radar interferometry for oceanographic and geophysical applications, in *The State of the Planet: Frontiers and Challenges in Geophysics*, edited by R. S. J. Sparks and C. J. Hawkesworth, pp. 209–224, American Geophysical Union, Washington, DC, doi:10.1029/150GM17.
- Fu, L.-L., I. Fukumori, and R. N. Miller (1993), Fitting dynamic models to the Geosat sea level observations in the tropical Pacific ocean. Part II: A linear, wind-driven model, *Journal of Physical Oceanography*, 23(10), 2162–2181, doi:10.1175/1520-0485(1993)023<2162:FDMTTG>2.0.CO;2.

- Fujita, T., D. J. Stensrud, and D. C. Dowell (2007), Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties, *Monthly Weather Review*, 135(5), 1846–1868, doi:10.1175/MWR3391.1.
- Gan, J., L. Li, D. Wang, and X. Guo (2009), Interaction of a river plume with coastal upwelling in the northeastern South China Sea, *Continental Shelf Research*, 29(4), 728–740, doi:10.1016/j.csr.2008.12.002.
- Garvine, R. W., and J. D. Monk (1974), Frontal structure of a river plume, *Journal of Geophysical Research*, 79(15), 2251–2259, doi:10.1029/JC079i015p02251.
- Gaspari, G., and S. E. Cohn (1999), Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757, doi:10.1002/qj.49712555417.
- Gauthier, P., M. Buehner, and L. Fillion (1998), Background-error statistics modelling in a 3D variational data assimilation scheme: Estimation and impact on the analyses, in *Proceedings of the ECMWF Workshop on Diagnosis of Data Assimilation Systems*, Shinfield Park, Reading, pp. 131–145, ECMWF, Reading.
- Geyer, C. J. (1992), Practical Markov chain Monte Carlo, *Statistical Science*, 7(4), 473–483.
- Giddings, S. N., P. MacCready, B. M. Hickey, N. S. Banas, K. A. Davis, S. A. Siedlecki, V. L. Trainer, R. M. Kudela, N. A. Pelland, and T. P. Connolly (2014), Hindcasts of potential harmful algal bloom transport pathways on the Pacific Northwest coast, *Journal of Geophysical Research: Oceans*, 119(4), 2439–2461, doi:10.1002/2013JC009622.
- Grigori, L., S. Moufawad, and F. Nataf (2016), Enlarged Krylov subspace conjugate gradient methods for reducing communication, *SIAM Journal of Matrix Analysis & Applications*, 37(2), 744–773, doi:10.1137/140989492.
- Gurgel, K.-W. (1994), Shipborne measurement of surface current fields by HF radar, in *Proceedings of the OCEANS'94. 'Oceans Engineering for Today's Technology and Tomorrow's Preservation'*, Brest, France, vol. 3, pp. 23–27, IEEE, Piscataway, NJ.
- Gürol, S., A. T. Weaver, A. M. Moore, A. Piacentini, H. G. Arango, and S. Gratton (2014), B-preconditioned minimization algorithms for variational data assimilation with the dual formulation, *Quarterly Journal of the Royal Meteorological Society*, 140(679), 539–556, doi:10.1002/qj.2150.

- Gustafsson, N., and J. Bojarova (2014), Four-dimensional ensemble variational (4D-En-Var) data assimilation for the High Resolution Limited Area Model (HIRLAM), *Nonlinear Processes in Geophysics*, 21(4), 745–762, doi:10.5194/npg-21-745-2014.
- Halpern, D. (1976), Structure of a coastal upwelling event observed off Oregon during July 1973, *Deep Sea Research and Oceanographic Abstracts*, 23(6), 495–508, doi: 10.1016/0011-7471(76)90861-5.
- Hamill, T. M. (2001), Interpretation of rank histograms for verifying ensemble forecasts, *Monthly Weather Review*, 129(3), 550–560, doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.
- Hamill, T. M., and J. S. Whitaker (2005), Accounting for the error due to unresolved scales in ensemble data assimilation: A comparison of different approaches, *Monthly Weather Review*, 133(11), 3132–3147, doi:10.1175/MWR3020.1.
- Hamill, T. M., J. S. Whitaker, and C. Snyder (2001), Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Monthly Weather Review*, 129(11), 2776–2790, doi:10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2, 00494.
- Hénaff, M. L., P. D. Mey, and P. Marsaleix (2009), Assessment of observational networks with the representer matrix spectra method-application to a 3D coastal model of the Bay of Biscay, *Ocean Dynamics*, 59(1), 3–20, doi: 10.1007/s10236-008-0144-7.
- Hetland, R. D. (2005), Relating river plume structure to vertical mixing, *Journal of Physical Oceanography*, 35(9), 1667–1688, doi:10.1175/JPO2774.1.
- Hickey, B. M., L. J. Pietrafesa, D. A. Jay, and W. C. Boicourt (1998), The Columbia River plume study: Subtidal variability in the velocity and salinity fields, *Journal of Geophysical Research: Oceans*, 103(C5), 10339–10368.
- Hickey, B. M., S. Geier, N. Kachel, and A. MacFadyen (2005), A bi-directional river plume: The Columbia in summer, *Continental Shelf Research*, 25(14), 1631–1656, doi:10.1016/j.csr.2005.04.010.
- Hickey, B. M., R. M. Kudela, J. D. Nash, K. W. Bruland, W. T. Peterson, P. MacCready, E. J. Lessard, D. A. Jay, N. S. Banas, A. M. Baptista, E. P. Dever, P. M. Kosro, L. K. Kilcher, A. R. HornerDevine, E. D. Zaron, R. M. McCabe, J. O. Peterson, P. M. Orton, J. Pan, and M. C. Lohan (2010), River influences on shelf

ecosystems: Introduction and synthesis, *Journal of Geophysical Research: Oceans*, 115(C2), doi:10.1029/2009JC005452.

Houtekamer, P. L., and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman filter technique, *Monthly Weather Review*, 126(3), 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.O;2.

Houtekamer, P. L., and H. L. Mitchell (2001), A sequential ensemble Kalman filter for atmospheric data assimilation, *Monthly Weather Review*, 129(1), 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Huyer, A. (1977), Seasonal variation in temperature, salinity, and density over the continental shelf off Oregon, *Limnology and Oceanography*, 22(3), 442–453, doi:10.4319/lo.1977.22.3.0442.

Huyer, A. (1983), Coastal upwelling in the California Current System, *Progress in Oceanography*, 12(3), 259–284, doi:10.1016/0079-6611(83)90010-1.

Huyer, A., J. H. Fleischbein, J. Keister, P. M. Kosro, N. Perlin, R. L. Smith, and P. A. Wheeler (2005), Two coastal upwelling domains in the northern California Current system, *Journal of Marine Research*, 63(5), 901–929.

Huyer, A., P. A. Wheeler, P. T. Strub, R. L. Smith, R. Letelier, and P. M. Kosro (2007), The Newport line off Oregon Studies in the North East Pacific, *Progress in Oceanography*, 75(2), 126–160, doi:10.1016/j.pocean.2007.08.003.

Ignatov, A., X. Zhou, B. Petrenko, X. Liang, Y. Kihai, P. Dash, J. Stroup, J. Sapper, and P. DiGiacomo (2016), AVHRR GAC SST reanalysis version 1 (RAN1), *Remote Sensing*, 8(4), 315, doi:10.3390/rs8040315, Data available at [https://coastwatch.noaa.gov/cw\\_html/sst\\_avhrr\\_gac.html](https://coastwatch.noaa.gov/cw_html/sst_avhrr_gac.html).

Ingleby, N. B., and A. C. Lorenc (1993), Bayesian quality control using multivariate normal distributions, *Quarterly Journal of the Royal Meteorological Society*, 119(513), 1195–1225, doi:10.1002/qj.49711951316.

Jochens, A. E., T. C. Malone, R. P. Stumpf, B. M. Hickey, M. Carter, R. Morrison, J. Dyble, B. Jones, and V. L. Trainer (2010), Integrated Ocean Observing System in support of forecasting harmful algal blooms, *Marine Technology Society Journal*, 44, doi:10.4031/MTSJ.44.6.16.

- Jones, E. M., P. R. Oke, F. Rizwi, and L. M. Murray (2012), Assimilation of glider and mooring data into a coastal ocean model, *Ocean Modelling*, 47, 1–13, doi: 10.1016/j.ocemod.2011.12.009.
- Kharchenko, S. A., and A. Y. Yeremin (1995), Eigenvalue translation based preconditioners for the GMRES(k) method, *Numerical Linear Algebra with Applications*, 2(1), 51–77, doi:10.1002/nla.1680020105.
- Kim, S. Y., P. M. Kosro, and A. L. Kurapov (2014), Evaluation of directly wind-coherent near-inertial surface currents off Oregon using a statistical parameterization and analytical and numerical models, *Journal of Geophysical Research: Oceans*, 119(10), 6631–6654, doi:10.1002/2014JC010115.
- Kirchgessner, P., L. Nerger, and A. Bunse-Gerstner (2014), On the choice of an optimal localization radius in ensemble Kalman filter methods, *Monthly Weather Review*, 142(6), 2165–2175, doi:10.1175/MWR-D-13-00246.1.
- Koch, A. O., A. L. Kurapov, and J. S. Allen (2010), Nearsurface dynamics of a separated jet in the coastal transition zone off Oregon, *Journal of Geophysical Research: Oceans*, 115, C08020, doi:10.1029/2009JC005704.
- Kosro, P. M. (2005), On the spatial structure of coastal circulation off Newport, Oregon, during spring and summer 2001 in a region of varying shelf width, *Journal of Geophysical Research*, 110, C10S06, doi:10.1029/2004JC002769.
- Kosro, P. M. (2017), HFRnet: HF radar national network production TDS, april 2011 to october 2011, [http://hfrnet-tds.ucsd.edu/thredds/HFRADAR\\_USWC\\_hourly\\_RTV.html](http://hfrnet-tds.ucsd.edu/thredds/HFRADAR_USWC_hourly_RTV.html), Scipps Institution of Oceanography, San Diego, CA.
- Küçükkaraca, E., and M. Fisher (2006), Use of analysis ensembles in estimating flow-dependent background error variances, *Technical memorandum no. 492*, ECMWF, Reading.
- Kuhl, D. D., T. E. Rosmond, C. H. Bishop, J. McLay, and N. L. Baker (2013), Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework, *Monthly Weather Review*, 141(8), 2740–2758, doi:10.1175/MWR-D-12-00182.1.
- Kurapov, A. L., J. S. Allen, R. N. Miller, and G. D. Egbert (1999), Generalized inverse for baroclinic coastal flows, in *Proceedings of the 3rd Conference on Coastal*

*Atmospheric and Oceanic Prediction and Processes*, pp. 3–5, American Meteorology Society, Boston, MA.

Kurapov, A. L., G. D. Egbert, R. N. Miller, and J. S. Allen (2002), Data assimilation in a baroclinic coastal ocean model: Ensemble statistics and comparison of methods, *Monthly Weather Review*, 130(4), 1009–1025, doi:10.1175/1520-0493(2002)130<1009:DAIABC>2.0.CO;2.

Kurapov, A. L., G. D. Egbert, J. S. Allen, R. N. Miller, S. Y. Erofeeva, and P. M. Kosro (2003), The M2 internal tide off Oregon: Inferences from data assimilation, *Journal of Physical Oceanography*, 33(8), 1733–1757, doi:10.1175/1520-0485(2003)033<1733:TMITOO>2.0.CO;2.

Kurapov, A. L., J. S. Allen, G. D. Egbert, R. N. Miller, P. M. Kosro, M. D. Levine, T. Boyd, and J. A. Barth (2005), Assimilation of moored velocity data in a model of coastal wind-driven circulation off Oregon: Multivariate capabilities, *Journal of Geophysical Research: Oceans*, 110(C10), doi:10.1029/2004JC002493.

Kurapov, A. L., G. D. Egbert, J. S. Allen, and R. N. Miller (2009), Representer-based analyses in the coastal upwelling system, *Dynamics of Atmospheres and Oceans*, 48(13), 198–218, doi:10.1016/j.dynatmoce.2008.09.002.

Kurapov, A. L., D. Foley, P. T. Strub, G. D. Egbert, and J. S. Allen (2011), Variational assimilation of satellite observations in a coastal ocean model off Oregon, *Journal of Geophysical Research: Oceans*, 116, C05006, doi:10.1029/2010JC006909.

Kurapov, A. L., N. A. Pelland, and D. L. Rudnick (2017), Seasonal and interannual variability in along-slope oceanic properties off the US West Coast: Inferences from a high-resolution regional model, *Journal of Geophysical Research: Oceans*, 122(7), 5237–5259, doi:10.1002/2017JC012721.

Labroue, S., F. Boy, N. Picot, M. Urvoy, and M. Ablain (2012), First quality assessment of the Cryosat-2 altimetric system over ocean, *Advances in Space Research*, 50(8), 1030–1045, doi:10.1016/j.asr.2011.11.018.

Laurent, D., N. Emilie, S. Ehouarn, F. Le Dimet, and A. Vidard (2015), Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems, *Quarterly Journal of the Royal Meteorological Society*, 142(694), 515–528, doi:10.1002/qj.2676.

- Lea, D. J., J.-P. Drecourt, K. Haines, and M. J. Martin (2008), Ocean altimeter assimilation with observational- and model-bias correction, *Quarterly Journal of the Royal Meteorological Society*, 134(636), 1761–1774, doi:10.1002/qj.320.
- Leeuwenburgh, O. (2007), Validation of an EnKF system for OGCM initialization assimilating temperature, salinity, and surface height measurements, *Monthly Weather Review*, 135(1), 125–139, doi:10.1175/MWR3272.1.
- Lellouche, J.-M., O. L. Galloudec, M. Dréville, C. Régnier, E. Greiner, G. Garric, N. Ferry, C. Desportes, C.-E. Testut, C. Bricaud, R. Bourdallé-Badie, B. Tranchant, M. Benkiran, Y. Drillet, A. Daudin, and C. De Nicola (2013), Evaluation of global monitoring and forecasting systems at Mercator Océan, *Ocean Science*, 9(1), 57, doi:10.5194/os-9-57-2013.
- Lermusiaux, P. F. J., and A. R. Robinson (1999), Data assimilation via error subspace statistical estimation. Part I: Theory and schemes, *Monthly Weather Review*, 127(7), 1385–1407, doi:10.1175/1520-0493(1999)127<1385:DAVESS>2.0.CO;2.
- Lewis, J. M., and J. C. Derber (1985), The use of adjoint equations to solve a variational adjustment problem with advective constraints, *Tellus A: Dynamic Meteorology and Oceanography*, 37A(4), 309–322, doi:10.1111/j.1600-0870.1985.tb00430.x.
- Li, H., E. Kalnay, T. Miyoshi, and C. M. Danforth (2009), Accounting for model errors in ensemble data assimilation, *Monthly Weather Review*, 137(10), 3407–3419, doi:10.1175/2009MWR2766.1.
- Li, Z., Y. Chao, J. D. Farrara, and J. C. McWilliams (2013), Impacts of distinct observations during the 2009 Prince William Sound field experiment: A data assimilation study, *Continental Shelf Research*, 63, S209–S222, doi:10.1016/j.csr.2012.06.018.
- Li, Z., J. C. McWilliams, K. Ide, and J. D. Farrara (2015a), A multiscale variational data assimilation scheme: Formulation and illustration, *Monthly Weather Review*, 143(9), 3804–3822, doi:10.1175/MWR-D-14-00384.1.
- Li, Z., J. C. McWilliams, K. Ide, and J. D. Farrara (2015b), Coastal ocean data assimilation using a multi-scale three-dimensional variational scheme, *Ocean Dynamics*, 65(7), 1001–1015, doi:10.1007/s10236-015-0850-x.
- Liu, C., Q. Xiao, and B. Wang (2008), An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test, *Monthly Weather Review*, 136(9), 3363–3373, doi:10.1175/2008MWR2312.1.

- Liu, Y., P. MacCready, and B. M. Hickey (2009), Columbia River plume patterns in summer 2004 as revealed by a hindcast coastal ocean circulation model, *Geophysical Research Letters*, *36*, L02601, doi:10.1029/2008GL036447.
- Liu, Y., R. H. Weisberg, C. Hu, and L. Zheng (2011), Tracking the Deepwater Horizon oil spill: A modeling perspective, *Eos*, *92*(6), 45–46, doi:10.1029/2011EO060001.
- Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Quarterly Journal of the Royal Meteorological Society*, *112*(474), 1177–1194, doi:10.1002/qj.49711247414.
- Lorenc, A. C. (2003), The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, *129*, 3183–3203.
- Lorenc, A. C., and M. Jardak (2018), A comparison of hybrid variational data assimilation methods for global NWP, *Quarterly Journal of the Royal Meteorological Society*, pp. 2748–2760, doi:10.1002/qj.3401.
- Lorenc, A. C., and F. Rawlins (2005), Why does 4D-Var beat 3D-Var?, *Quarterly Journal of the Royal Meteorological Society*, *131*(613), 3247–3257, doi:10.1256/qj.05.85.
- Lorenc, A. C., M. Jardak, T. Payne, N. E. Bowler, and M. A. Wlasak (2017), Computing an ensemble of variational data assimilations using its mean and perturbations, *Quarterly Journal of the Royal Meteorological Society*, *143*(703), 798–805, doi:10.1002/qj.2965.
- Lynch, P., and X.-Y. Huang (1992), Initialization of the HIRLAM model using a digital filter, *Monthly Weather Review*, *120*(6), 1019–1034, doi:10.1175/1520-0493(1992)120<1019:IOTHMU>2.0.CO;2.
- MacCready, P., and S. N. Giddings (2016), The mechanical energy budget of a regional ocean model, *Journal of Physical Oceanography*, *46*(9), 2719–2733, doi:10.1175/JPO-D-16-0086.1.
- MacCready, P., N. S. Banas, B. M. Hickey, E. P. Dever, and Y. Liu (2009), A model study of tide- and wind-induced mixing in the Columbia River Estuary and plume, *Continental Shelf Research*, *29*(1), 278–291, doi:10.1016/j.csr.2008.03.015.

- MacQueen, J. (1967), Some methods for classification and analysis of multivariate observations, in *Proceedings of the Berkeley symposium on mathematics and statistics and probability, 5th*, vol. 1, pp. 281–297, Oakland, CA, USA.
- Mandel, J., E. Bergou, S. Gürol, S. Gratton, and I. Kasanický (2016), Hybrid Levenberg-Marquardt and weak-constraint ensemble Kalman smoother method, *Nonlinear Processes in Geophysics*, 23(2), 59–73, doi:10.5194/npg-23-59-2016.
- Marchesiello, P., J. C. McWilliams, and A. Shchepetkin (2001), Open boundary conditions for long-term integration of regional oceanic models, *Ocean Modelling*, 3(1), 1–20, doi:10.1016/S1463-5003(00)00013-5.
- Matthews, D., B. S. Powell, and I. Janeković (2012), Analysis of four-dimensional variational state estimation of the Hawaiian waters, *Journal Geophysical Research: Oceans*, 117, C03013, doi:10.1029/2011JC007575.
- Maturi, E., A. Harris, J. Mittaz, C. Merchant, B. Potash, W. Meng, and J. Sapper (2008), NOAA’s sea surface temperature products from operational geostationary satellites, *Bulletin of the American Meteorological Society*, 89(12), 1877–1888, doi: 10.1175/2008BAMS2528.1.
- Mazzini, P. L. F., J. A. Barth, R. K. Shearman, and A. Erofeev (2014), Buoyancy-driven coastal currents off Oregon during Fall and Winter, *Journal of Physical Oceanography*, 44(11), 2854–2876, doi:10.1175/JPO-D-14-0012.1.
- McWilliams, J. C., F. Colas, and M. J. Molemaker (2009), Cold filamentary intensification and oceanic surface convergence lines, *Geophysical Research Letters*, 36(18), doi:10.1029/2009GL039402.
- Melet, A., J. Verron, and J.-M. Brankart (2012), Potential outcomes of glider data assimilation in the Solomon Sea: Control of the water mass properties and parameter estimation, *Journal of Marine Systems*, 94, 232–246, doi:10.1016/j.jmarsys.2011.12.003.
- Mellor, G. L., and T. Yamada (1982), Development of a turbulence closure model for geophysical fluid problems, *Reviews of Geophysics*, 20(4), 851–875, doi:10.1029/RG020i004p00851.
- Meng, Z., and F. Zhang (2008), Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part IV: Comparison with 3DVAR in a

month-long experiment, *Monthly Weather Review*, 136(10), 3671–3682, doi:10.1175/2008MWR2270.1.

Milliff, R. F., A. Bonazzi, C. K. Wikle, N. Pinardi, and L. M. Berliner (2011), Ocean ensemble forecasting. Part I: Ensemble Mediterranean winds from a Bayesian hierarchical model, *Quarterly Journal of the Royal Meteorological Society*, 137(657), 858–878, doi:10.1002/qj.767.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin (2002), Ensemble size, balance, and model-error representation in an ensemble Kalman filter, *Monthly Weather Review*, 130(11), 2791–2808, doi:10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2.

Mooers, C. N. K., C. A. Collins, and R. L. Smith (1976), The dynamic structure of the frontal zone in the coastal upwelling region off Oregon, *Journal of Physical Oceanography*, 6(1), 3–21, doi:10.1175/1520-0485(1976)006<0003:TDSOTF>2.0.CO;2.

Moore, A. M., H. G. Arango, G. Broquet, B. S. Powell, A. T. Weaver, and J. Zavala-Garay (2011), The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems: Part I - System overview and formulation, *Progress in Oceanography*, 91(1), 34–49, doi:10.1016/j.pocean.2011.05.004.

Morgan, R. (1995), A restarted GMRES method augmented with eigenvectors, *SIAM Journal of Matrix Analysis & Applications*, 16(4), 1154–1171, doi:10.1137/S0895479893253975.

NANOOS (2018), NVS: Data explorer, Available from: <http://nvs.nanoos.org/> Explorer (Updated daily).

Navon, I. M., and R. de Villiers (1983), Combined penalty multiplier optimization methods to enforce integral invariants conservation, *Monthly Weather Review*, 111(6), 1228–1243, doi:10.1175/1520-0493(1983)111<1228:CPMOMT>2.0.CO;2.

NCEP (2003), U.S. coastal relief model (CRM). Vol. 7 and 8, <https://www.ngdc.noaa.gov/mgg/coastal/crm.html>, National Centers for Environmental Information, NOAA, Boulder, CO.

NCEP (2006), 2-minute gridded global relief data (ETOPO2v2), version 2, <https://www.ngdc.noaa.gov/mgg/global/etopo2.html>, National Geophysical Data Center, NOAA, Boulder, CO, doi:10.7289/V5J1012Q.

- NCEP (2011), North American Mesoscale Forecast System (NAM) 12 km analysis, january 2011 to december 2011, <https://www.ncei.noaa.gov/thredds/catalog/namanl/catalog.html>, National Centers for Environmental Information, NOAA, Asheville, NC.
- NDBC (2016), National Data Buoy Center, January 2011 to December 2012, <http://www.ndbc.noaa.gov/>, National Weather Service, NOAA, Stennis Space Center, MS. (Updated daily).
- Nescio (1911), De uitvreter, *De Gids*, 75, 45–74, (In Dutch).
- Ngodock, H., and M. Carrier (2014a), A 4DVAR system for the Navy Coastal Ocean Model. Part I: System description and assimilation of synthetic observations in Monterey Bay, *Monthly Weather Review*, 142(6), 2085–2107, doi: 10.1175/MWR-D-13-00221.1.
- Ngodock, H., and M. Carrier (2014b), A 4DVAR system for the Navy Coastal Ocean Model. Part II: Strong and weak constraint assimilation experiments with real observations in Monterey Bay, *Monthly Weather Review*, 142(6), 2108–2117, doi: 10.1175/MWR-D-13-00220.1.
- Oke, P. R., J. S. Allen, R. N. Miller, G. D. Egbert, and P. M. Kosro (2002a), Assimilation of surface velocity data into a primitive equation coastal ocean model, *Journal of Geophysical Research: Oceans*, 107(C9), 3122, doi:10.1029/2000JC000511.
- Oke, P. R., J. S. Allen, R. N. Miller, G. D. Egbert, J. A. Austin, J. A. Barth, T. J. Boyd, P. M. Kosro, and M. D. Levine (2002b), A modeling study of the three-dimensional continental shelf circulation off Oregon. Part I: Model-data comparisons, *Journal of Physical Oceanography*, 32(5), 1360–1382, doi: 10.1175/1520-0485(2002)032<1360:AMSOTT>2.0.CO;2.
- Oke, P. R., J. S. Allen, R. N. Miller, and G. D. Egbert (2002c), A modeling study of the three-dimensional continental shelf circulation off Oregon. Part II: Dynamical analysis, *Journal of Physical Oceanography*, 32(5), 1383–1403, doi: 10.1175/1520-0485(2002)032<1383:AMSOTT>2.0.CO;2.
- O’Leary, D. P. (1980), The block conjugate gradient algorithm and related methods, *Linear Algebra and its Applications*, 29(Supplement C), 293–322, doi:10.1016/0024-3795(80)90247-5.

- OOI (2018), Mobile assets (CE05MOAS), Available from: <http://oceanobservatories.org/site/ce05moas/> (Accessed 17 April 2018).
- Ooyama, K. V. (1987), Scale-controlled objective analysis, *Monthly Weather Review*, 115(10), 2479–2506, doi:10.1175/1520-0493(1987)115<2479:SCOA>2.0.CO;2.
- Oppenheim, A. V., and R. W. Schafer (1989), *Discrete-Time signal processing*, Prentice Hall, Upper Saddle River, NJ.
- Orton, P. M., and D. A. Jay (2005), Observations at the tidal plume front of a high-volume river outflow, *Geophysical Research Letters*, 32(11), doi:10.1029/2005GL022372.
- Osborne, J. J., A. L. Kurapov, G. D. Egbert, and P. M. Kosro (2011), Spatial and temporal variability of the M2 internal tide generation and propagation on the Oregon shelf, *Journal of Physical Oceanography*, 41(11), 2037–2062, doi:10.1175/JPO-D-11-02.1.
- Palmer, T., R. Buizza, D.-R. F., T. Jung, M. Leutbecher, G. Shutts, M. Steinheimer, and A. Weisheimer (2009), Stochastic parametrization and model uncertainty, *Technical memorandum no. 598*, ECMWF, Reading.
- Pan, C., L. Zheng, R. H. Weisberg, Y. Liu, and C. E. Lembke (2014), Comparisons of different ensemble schemes for glider data assimilation on West Florida Shelf, *Ocean Modelling*, 81, 13–24, doi:10.1016/j.ocemod.2014.06.005.
- Pasmans, I., and A. L. Kurapov (2017), A Monte Carlo background covariance localization method for an ensemble variational assimilation system, *Monthly Weather Review*, 145(11), 4543–4557, doi:10.1175/MWR-D-16-0424.1.
- Pawlowicz, R., B. Beardsley, and S. Lentz (2002), Classical tidal harmonic analysis including error estimates in MATLAB using T\_Tide, *Computers & Geosciences*, 28(8), 929–937, doi:10.1016/S0098-3004(02)00013-4.
- Poterjoy, J., and F. Zhang (2015), Systematic comparison of four-dimensional data assimilation methods with and without the tangent linear model using hybrid background error covariance: E4DVar versus 4DEnVar, *Monthly Weather Review*, 143(5), 1601–1621, doi:10.1175/MWR-D-14-00224.1.
- Poterjoy, J., F. Zhang, J. Poterjoy, and F. Zhang (2016), Comparison of hybrid four-dimensional data assimilation methods with and without the tangent linear and

- adjoint models for predicting the life cycle of hurricane Karl (2010), doi:10.1175/MWR-D-15-0116.1.
- Price, J. F., and M. A. Sundermeyer (1999), Stratified Ekman layers, *Journal of Geophysical Research: Oceans*, 104(C9), 20467–20494, doi:10.1029/1999JC900164.
- Purser, R. J. (1984), A new approach to the optimal assimilation of meteorological data by iterative Bayesian analysis, in *Conference on Weather Forecasting and Analysis, 10th, Clearwater Beach, FL*, pp. 102–105, American Meteorology Society, Boston, MA.
- Rao, V., and A. Sandu (2016), A time-parallel approach to strong-constraint four-dimensional variational data assimilation, *Journal of Computational Physics*, 313, 583–593, doi:10.1016/j.jcp.2016.02.040.
- Richardson, L. F. (1922), *Weather prediction by numerical process*, Cambridge University Press, London.
- Saldías, G. S., R. K. Shearman, J. A. Barth, and N. Tufillaro (2016), Optics of the offshore Columbia River plume from glider observations and satellite imagery, *Journal of Geophysical Research: Oceans*, 121(4), 2367–2384, doi:10.1002/2015JC011431.
- Sasaki, Y. (1970a), Some basic formalisms in numerical variational analysis, *Monthly Weather Review*, 98(12), 875–883, doi:10.1175/1520-0493(1970)098<0875:SBFINV>2.3.CO;2.
- Sasaki, Y. (1970b), Numerical variational analysis formulated under the constraints as determined by longwave equations and a low-pass filter, *Monthly Weather Review*, 98(12), 884–898, doi:10.1175/1520-0493(1970)098<0884:NVAFUT>2.3.CO;2.
- Scharroo, R., E. W. Leuliette, J. L. Lillibridge, D. Byrne, M. C. Naeije, and G. T. Mitchum (2013), RADS: Consistent multi-mission products, in *Proceedings of the Symposium on 20 Years of Progress in Radar Altimetry, Venice, 20–28 September 2012*, vol. 20, p. 4, ESA Communications, Neuilly-sur-Seine/Noordwijk, Data available from: <http://rads.tudelft.nl/rads/rads.shtml>.
- Shchepetkin, A. F., and J. C. McWilliams (2003), A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate, *Journal of Geophysical Research*, 108(C3), 3090, doi:10.1029/2001JC001047.

- Shchepetkin, A. F., and J. C. McWilliams (2005), The Regional Oceanic Modeling System (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model, *Ocean Modelling*, 9(4), 347–404, doi:10.1016/j.ocemod.2004.08.002.
- Shulman, I., C. Rowley, S. Anderson, S. DeRada, J. Kindle, P. Martin, J. Doyle, J. Cummings, S. Ramp, F. Chavez, D. Fratantoni, and R. Davis (2009), Impact of glider data assimilation on the Monterey Bay model, *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(35), 188–198, doi:10.1016/j.dsr2.2008.08.003.
- Siedlecki, S. A., N. S. Banas, K. A. Davis, S. Giddings, B. M. Hickey, P. MacCready, T. Connolly, and S. Geier (2015), Seasonal and interannual oxygen variability on the Washington and Oregon continental shelves, *Journal of Geophysical Research: Oceans*, 120(2), 608–633, doi:10.1002/2014JC010254.
- Sikirić, M. D., I. Janečković, and M. Kuzmić (2009), A new approach to bathymetry smoothing in sigma-coordinate ocean models, *Ocean Modelling*, 29(2), 128–136, doi:10.1016/j.ocemod.2009.03.009.
- STAR (2011), ASCAT METOPA-A, April 2011 to September 2011, <https://manati.star.nesdis.noaa.gov/datasets/ASCATData.php/>, National Environmental Satellite, Data and Information Service, NOAA, Asheville, NC.
- Strichartz, R. (1994), *A guide to distribution theory and Fourier transforms*, CRC press, Boca Raton, FL.
- Tavolato, C., and L. Isaksen (2015), On the use of a Huber norm for observation quality control in the ECMWF 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, 141(690), 1514–1527, doi:10.1002/qj.2440.
- Thépaut, J.-N., P. Courtier, G. Belaud, and G. Lemaître (1996), Dynamical structure functions in a four-dimensional variational assimilation: A case study, *Quarterly Journal of the Royal Meteorological Society*, 122(530), 535–561, doi:10.1002/qj.49712253012.
- Tian, X., H. Zhang, X. Feng, and Y. Xie (2017), Nonlinear least square En4DVar to 4DEnVar methods for data assimilation: Formulation, analysis, and preliminary evaluation, *Monthly Weather Review*, 146(1), 77–93, doi:10.1175/MWR-D-17-0050.1.

- Towns, J., T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr (2014), XSEDE: Accelerating scientific discovery, *Computing in Science & Engineering*, 16(5), 62–74, doi:10.1109/MCSE.2014.80.
- Trefethen, L. N., and D. Bau (1997), *Numerical linear algebra*, SIAM, Philadelphia, PA.
- USGS (2015), USGS water data for USA , October 2008 to April 2015, <https://waterdata.usgs.gov/usa/nwis/>, U.S. Geological Survey, Reston, VA.
- Verspeek, J., A. Verhoef, and A. Stoffelen (2013), ASCAT-B NWP ocean calibration and validation, *Technical report saf/osi/cdop2/knmi/tec/rp/199*, KNMI, De Bilt.
- Vervatis, V., C. E. Testut, P. De Mey, N. Ayoub, J. Chanut, and G. Quattrochi (2016), Data assimilative twin-experiment in a high-resolution Bay of Biscay configuration: 4DEnOI based on stochastic modeling of the wind forcing, *Ocean Modelling*, 100, 1–19, doi:10.1016/j.ocemod.2016.01.003.
- Weaver, A. T., C. Deltel, E. Machu, S. Ricci, and N. Daget (2005), A multivariate balance operator for variational ocean data assimilation, *Quarterly Journal of the Royal Meteorological Society*, 131(613), 3605–3625, doi:10.1256/qj.05.119.
- Whitaker, J. S., and T. M. Hamill (2002), Ensemble data assimilation without perturbed observations, *Monthly Weather Review*, 130(7), 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.
- Whitaker, J. S., and T. M. Hamill (2012), Evaluating methods to account for system errors in ensemble data assimilation, *Monthly Weather Review*, 140(9), 3078–3089, doi:10.1175/MWR-D-11-00276.1.
- Wikle, C. K., R. F. Milliff, D. Nychka, and L. M. Berliner (2001), Spatiotemporal hierarchical Bayesian modeling tropical ocean surface winds, *Journal of the American Statistical Association*, 96(454), 382–397, doi:10.1198/016214501753168109.
- Yaremchuk, M., P. Martin, and C. Beattie (2017), A hybrid approach to generating search subspaces in dynamically constrained 4-dimensional data assimilation, *Ocean Modelling*, 117, 41–51, doi:10.1016/j.ocemod.2017.08.003.
- Yu, P., A. L. Kurapov, G. D. Egbert, J. S. Allen, and P. M. Kosro (2012), Variational assimilation of HF radar surface currents in a coastal ocean model off Oregon, *Ocean Modelling*, 49-50, 86–104, doi:10.1016/j.ocemod.2012.03.001.

Zhang, W. G., J. L. Wilkin, and H. G. Arango (2010), Towards an integrated observation and modeling system in the New York Bight using variational methods. Part I: 4DVAR data assimilation, *Ocean Modelling*, 35(3), 119–133, doi: 10.1016/j.ocemod.2010.08.003.

## APPENDICES

## Appendix A: Mixing Algorithm

For a vertical column let  $\rho_i$  be the potential density in the  $i$ -th layer and  $h_i$  the depth of the  $i$ -th layer with  $i = 1$  being the bottom layer and  $i = N_z$  being the top layer. The MATLAB pseudo-code used to mix the potential density into a stable profile whilst preserving the depth-averaged potential density is as follows:

```

for  $i = [N_z : -1 : 2]$ 
    for  $j = [i - 1 : -1 : 1]$ 
        if  $\rho_j < \rho_i + \delta(i - j)$ 
             $\bar{\rho} = (\sum_{k=j}^i h_k \rho_k + \delta h_k(k - i))(\sum_{k=j}^i h_k)^{-1}$ 
             $\rho_{[i:j]} = \bar{\rho} + \delta(i - [j : i])$ 
        end if
    end for
end for

```

Here  $\delta$  is the minimum density difference between two layers. For purely practical reasons  $\delta$  is set to  $10^{-6} \text{ kg m}^{-3}$  such that the density profile becomes a strictly monotonically decreasing function of the vertical coordinate  $z$  and can therefore be used as a vertical coordinate itself.

## Appendix B: Fourier Transform of the Modulated Ensemble Members

As in section 3.5 consider a uniform grid with spatial coordinate  $\xi$  and grid spacing  $d\xi$ . To simplify the derivation of the Fourier transform of the modulated ensemble member  $\mathbf{x}^{(k)} \circ \boldsymbol{\chi}^{(m)} \circ \boldsymbol{\gamma}$  we extend  $\boldsymbol{\chi}^{(m)}$ ,  $\boldsymbol{\gamma}$  to all  $\xi \in \mathbb{R}$  by defining functions  $\chi^{(m)}(\xi)$ ,  $\gamma(\xi)$  with  $\chi^{(m)}(\xi) = 1$  if  $|\xi - \xi_m| \leq \frac{1}{2}L_{MC}$ ,  $\gamma(\xi) = \left(\frac{1}{K} \sum_{k=1}^K \sum_{m \in \mathbb{M}} \chi^{(m)}(\xi)^2\right)^{-1/2} = \left(\sum_{m=1}^M \chi^{(m)}(\xi)^2\right)^{-1/2}$ . Then  $\chi^{(m)}(\xi_p) = \boldsymbol{\chi}_p^{(m)}$  and  $\gamma(\xi_p) = \boldsymbol{\gamma}_p$ .

If  $|\xi - \xi_1|, |\xi - \xi_N| \geq \frac{1}{2}L_{MC}$  then there are  $\lfloor \frac{L_{MC}}{d\xi} \rfloor + 1$  points for which  $|\xi - \xi_p| < \frac{1}{2}L_{MC}$  with  $\lfloor \cdot \rfloor$  the floor function. Hence,  $\gamma(\xi) = (\lfloor \frac{L_{MC}}{d\xi} \rfloor + 1)^{-1/2} \stackrel{\text{def}}{=} c_\gamma$  for such  $\xi$ . If  $m$  is such that  $|\xi_m - \xi_1|, |\xi_m - \xi_N| > L_{MC}$   $\chi^{(m)}(\xi) = 0$  if  $|\xi - \xi_1|, |\xi - \xi_N| < \frac{1}{2}L_{MC}$  and hence  $\chi^{(m)}(\xi)\gamma(\xi) = c_\gamma\chi^{(m)}(\xi)$  for such  $m$ . In this case the Fourier transform of  $\chi^{(m)} \circ \gamma$  is given by

$$\begin{aligned} \mathcal{F}(\chi^{(m)} \circ \gamma)(\kappa) &= \int_{\mathbb{R}} c_\gamma \chi^{(m)}(\xi) \exp(-i\kappa\xi) d\xi = c_\gamma \int_{\xi_m - \frac{1}{2}L_{MC}}^{\xi_m + \frac{1}{2}L_{MC}} \exp(-i\kappa\xi) d\xi \\ &= c_\gamma \exp(-i\kappa\xi_m) \frac{\exp(-i\kappa\xi)}{-i\kappa} \Big|_{\xi=-\frac{1}{2}L_{MC}}^{\frac{1}{2}L_{MC}} \\ &= c_\gamma L_{MC} \exp(-i\kappa\xi_m) \text{sinc}\left(\frac{1}{2}L_{MC}\kappa\right) \end{aligned} \quad (\text{B.1})$$

Provided the integrals exist the Fourier transform of the modulated ensemble

member  $x^{(k)}(\xi)$  then follows from the convolution theorem:

$$\begin{aligned}
& \mathcal{F}(x^{(k)} \circ \chi^{(m)} \circ \gamma)(\kappa) \\
&= \int_{\mathbb{R}} x^{(k)}(\xi)(\chi^{(m)} \circ \gamma)(\xi) \exp(-i\kappa\xi) d\xi \\
&= \int_{\mathbb{R}} x^{(k)}(\xi) \left( \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}(\chi^{(m)} \circ \gamma)(\kappa') \exp(i\kappa'\xi) d\kappa' \right) \exp(-i\kappa\xi) d\xi \\
&= \frac{1}{2\pi} \iint_{\mathbb{R}^2} x^{(k)}(\xi) \exp(-i(\kappa - \kappa')\xi) \mathcal{F}(\chi^{(m)} \circ \gamma)(\kappa') d\xi d\kappa' \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \mathcal{F}x^{(k)}(\kappa - \kappa') \mathcal{F}(\chi^{(m)} \circ \gamma)(\kappa') d\kappa' \\
&= \frac{1}{2\pi} \mathcal{F}x^{(k)} * \mathcal{F}(\chi^{(m)} \circ \gamma)(\kappa)
\end{aligned} \tag{B.2}$$

Here the fifth line of (B.2) shows that the Fourier transform of the modulated ensemble member  $x^{(k)} \circ \chi^{(m)} \circ \gamma$  at the wave number  $\kappa$  is a weighted combination of values of the Fourier transform at nearby wave numbers  $\kappa - \kappa'$  where the amplitude of the nearby Fourier component is scaled with  $\text{sinc}(\frac{1}{2}L_{MC}\kappa')$  and the phase is shifted by  $\kappa'\xi_m$ . Notice that for  $x^{(k)}$  constructed using (3.13) or (3.17)  $\mathcal{F}x^{(k)}$  consists of a finite number of delta-functions and hence the integrals in (B.2) will be properly defined.

On a discrete grid the value of the modulated ensemble member is only sampled on the grid points and hence the discrete modulated ensemble member can be written as  $x^{(k)} \circ \chi^{(m)} \circ \gamma \circ \text{III}(\xi)$  with

$$\text{III}_{d\xi}(\xi) = \sum_{l=-\infty}^{\infty} \delta(\xi - ld\xi) \tag{B.3}$$

the Dirac Comb or sampling function which has Fourier transform (*Strichartz*, 1994)

$$\mathcal{F}\text{III}_{d\xi}(\kappa) = \frac{2\pi}{d\xi} \text{III}_{\frac{2\pi}{d\xi}}(\kappa) = \frac{2\pi}{d\xi} \sum_{l=-\infty}^{\infty} \delta(\kappa - l\kappa_{sample}) \tag{B.4}$$

Here  $\kappa_{sample} = \frac{2\pi}{d\xi}$  is the sampling wave number. Now, using once again the con-

volution theorem the Fourier transform of the discrete modulated ensemble member becomes

$$\begin{aligned}
\mathcal{F}(x^{(k)} \circ \chi^{(m)} \circ \gamma \circ \text{III}_{d\xi})(\kappa) &= \frac{1}{2\pi} \mathcal{F}(x^{(k)} \circ \chi^{(m)} \circ \gamma) * \mathcal{F}\text{III}_{d\xi}(\kappa) \\
&= \frac{1}{d\xi} \sum_{l=-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{F}(x^{(k)} \circ \chi^{(m)} \circ \gamma)(\kappa') \delta(\kappa - l\kappa_{sample} - \kappa') d\kappa' \\
&= \frac{1}{d\xi} \sum_{l=-\infty}^{\infty} \mathcal{F}(x^{(k)} \circ \chi^{(m)} \circ \gamma)(\kappa - l\kappa_{sample}) \\
&= \frac{c_{\gamma} L_{MC}}{2\pi d\xi} \int_{-\infty}^{\infty} \left[ \sum_{l=-\infty}^{\infty} \mathcal{F}x^{(k)}(\kappa - \kappa' - l\kappa_{sample}) \right] \exp(-i\kappa' \xi_m) \text{sinc}\left(\frac{1}{2}L_{MC}\kappa'\right) d\kappa'
\end{aligned} \tag{B.5}$$

Comparison of (B.2) and (B.5) shows that discretization gives rise to aliasing:  $\mathcal{F}x^{(k)}(\kappa - \kappa')$  has been replaced by  $\sum_{l=-\infty}^{\infty} \mathcal{F}x^{(k)}(\kappa - \kappa' - l\kappa_{sample})$ . Consequently, for the discrete modulated ensemble member the Fourier transform at wave number  $\kappa$  is a weighted combination dominated by Fourier components with wave numbers close to  $\kappa$  and wave numbers differing from  $\kappa$  by a multiple of  $\kappa_{sample}$ . In particular, if  $x^{(k)}$  is generated using (3.13) or (3.17) the discrete modulated ensemble member also has nonzero Fourier components for  $|\kappa| > |\kappa_{\pm N_{1/2}}|$ . As a consequence, Fourier components of the discrete modulated ensemble member for  $\kappa \sim \kappa_{N_{1/2}}$  ( $\kappa \sim \kappa_{-N_{1/2}}$ ) are not only combinations of the Fourier components of  $x^{(k)}$  with wave numbers near  $\lesssim \kappa_{N_{1/2}}$  ( $\gtrsim \kappa_{-N_{1/2}}$ ), but also with wave numbers near  $\kappa_{-N_{1/2}}$  ( $\kappa_{N_{1/2}}$ ).

## Appendix C: Pseudocode Minimization Algorithms

Table C.1: Pseudocode for RBCG using (4.6) for ensemble member  $m$ .

$\hat{\mathbf{V}}_0 = [ ]$ ; $\hat{\mathbf{W}}_0 = [ ]$ ; $\hat{\mathbf{r}} = \hat{\mathbf{d}}^{(m)}$
<b>for</b> $i = [1 : I]$
$\hat{\mathbf{V}}_i = [\hat{\mathbf{V}}_{i-1}, \hat{\mathbf{r}}]$
Use AVRORA to calculate $\hat{\mathbf{W}}_i = [\hat{\mathbf{W}}_{i-1}, \hat{\mathbf{B}}\hat{\mathbf{r}}^{(m)}]$
$\hat{\mathbf{A}}\hat{\mathbf{V}}_i = \hat{\mathbf{W}}_i + \hat{\mathbf{V}}_i$
$\mathbf{T} = (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T (\hat{\mathbf{A}}\hat{\mathbf{V}}_i)$
$\mathbf{p} = \mathbf{T}^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T \hat{\mathbf{d}}^{(m)}$
$\hat{\mathbf{x}}^{(m)} = \hat{\mathbf{V}}_i \mathbf{p}$ and $\hat{\mathbf{B}}\hat{\mathbf{x}}^{(m)} = \hat{\mathbf{W}}_i \mathbf{p}$
$\hat{\mathbf{r}}^{(m)} = \hat{\mathbf{d}}^{(m)} - \hat{\mathbf{B}}\hat{\mathbf{x}}^{(m)} - \hat{\mathbf{x}}^{(m)}$
<b>end for</b>
Use AVRORA to calculate correction $\mathbf{x}^{(m)} = \mathbf{B}\mathbf{M}^T \mathbf{R}^{-1/2} \hat{\mathbf{x}}^{(m)}$

Table C.2: Pseudocode for block diagonal CG with B-preconditioning using (4.6).

---

```

Initialize  $\hat{\mathbf{V}}_0 = []$  and  $\hat{\mathbf{W}}_0 = []$ 
Set  $\mathbf{D} = [\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \dots, \hat{\mathbf{d}}^{(M-1)}]$  and  $\mathbf{S} = \hat{\mathbf{D}}$ 
for  $i = [1 : I]$ 
   $\hat{\mathbf{V}}_i = [\hat{\mathbf{V}}_{i-1}, \mathbf{S}]$ 
  parfor  $m = [1 : M]$ 
    Take  $\hat{\mathbf{r}}$  to be the  $m$ -th column vector of  $\mathbf{S}$ 
    Use AVRORA to calculate  $\mathbf{w}^{(m-1)} = \hat{\mathbf{B}}\hat{\mathbf{r}}$ 
  end parfor
   $\hat{\mathbf{W}}_i = [\hat{\mathbf{W}}_{i-1}, \mathbf{w}^{(0)}, \mathbf{w}^{(1)}, \dots, \mathbf{w}^{(M-1)}]$ 
   $\hat{\mathbf{A}}\hat{\mathbf{V}}_i = \hat{\mathbf{W}}_i + \hat{\mathbf{V}}_i$ 
   $\mathbf{T} = (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T (\hat{\mathbf{A}}\hat{\mathbf{V}}_i)$ 
   $\mathbf{P} = \mathbf{T}^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T \hat{\mathbf{D}}$ 
   $\hat{\mathbf{X}} = \hat{\mathbf{V}}_i \mathbf{P}$  and  $\hat{\mathbf{B}}\hat{\mathbf{X}} = \hat{\mathbf{W}}_i \mathbf{P}$ 
   $\mathbf{S} = \hat{\mathbf{D}} - \hat{\mathbf{B}}\hat{\mathbf{X}} - \hat{\mathbf{X}}$ 
end for
parfor  $m = [0 : M - 1]$ 
  Let  $\hat{\mathbf{x}}^{(m)}$  be the  $m + 1$ -th column of  $\hat{\mathbf{X}}$ 
  Use AVRORA to calculate correction  $\mathbf{x}^{(m)} = \mathbf{B}\mathbf{M}^T \mathbf{R}^{-1/2} \hat{\mathbf{x}}^{(m)}$ 
end parfor

```

---

Table C.3: Pseudocode for the cluster search method.  $\mathbf{e}_j$  is a unit vector in direction  $j$ .

---

```

Initialize  $\hat{\mathbf{V}}_0 = []$  and  $\hat{\mathbf{W}}_0 = []$ 
Set  $\mathbf{D} = [\hat{\mathbf{d}}^{(0)}, \hat{\mathbf{d}}^{(1)}, \dots, \hat{\mathbf{d}}^{(M)}]$  and  $\mathbf{r} = \hat{\mathbf{d}}^{(0)}$ 
Use SVD to decompose  $\frac{1}{\sqrt{M}}\mathbf{D} = \hat{\mathbf{U}}\Lambda\mathbf{Z}^T$ . Discard  $\mathbf{Z}$ 
Set  $\hat{\mathbf{U}}' = \hat{\mathbf{U}}$ 
Calculate  $\hat{\mathbf{v}}'' = \hat{\mathbf{U}}'^T \hat{\mathbf{d}}^{(0)}$ 
for  $i = [1 : I]$ 
  for  $m = [1, 2, \dots, M]$ 
     $\xi_m = (\hat{\mathbf{U}}\mathbf{e}_m)^T (\hat{\mathbf{U}}'\mathbf{e}_m)$ 
  Continues on the next page.

```

---

Table C.3 (Continued): Pseudocode for the cluster search method.  $\mathbf{e}_j$  is a unit vector.

```

end for
 $\hat{\mathbf{v}}' = \hat{\mathbf{U}}^T \mathbf{r}$ 
Get  $D_1, D_2, \dots, D_{N_s} = \text{func\_cluster}(\Lambda, \mathbf{v}', \mathbf{v}'', \xi)$ 
for  $n = [1 : N_s]$ 
 $\mathbf{s}^{(n)} = \sum_{j \in D_n} (\hat{\mathbf{U}}' \mathbf{e}_j) (\mathbf{e}_j^T \mathbf{v}'')$ 
end for
 $\hat{\mathbf{V}}_i = [\hat{\mathbf{V}}_{i-1}, \mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N_s)}]$ 
parfor  $n = [1 : N_s]$ 
    Use AVRORA to calculate  $\mathbf{w}^{(n)} = \hat{\mathbf{B}} \mathbf{s}^{(n)}$ 
end parfor
 $\hat{\mathbf{W}}_i = [\hat{\mathbf{W}}_{i-1}, \mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \dots, \mathbf{w}^{(N_s)}]$ 
 $\hat{\mathbf{A}}\hat{\mathbf{V}}_i = \hat{\mathbf{W}}_i + \hat{\mathbf{V}}_i$ 
 $\mathbf{T} = (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T (\hat{\mathbf{A}}\hat{\mathbf{V}}_i)$ 
 $\mathbf{P} = \mathbf{T}^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T \hat{\mathbf{D}}$  and  $\mathbf{Q} = \mathbf{T}^{-1} (\hat{\mathbf{V}}_i^T \hat{\mathbf{W}}_i)^{-1} \hat{\mathbf{W}}_i^T \hat{\mathbf{U}}$ 
 $\hat{\mathbf{X}} = \hat{\mathbf{V}}_i \mathbf{P}$  and  $\hat{\mathbf{B}}\hat{\mathbf{X}} = \hat{\mathbf{W}}_i \mathbf{P}$ 
 $\mathbf{U}' = \hat{\mathbf{U}} - \hat{\mathbf{W}}_i \mathbf{Q} - \hat{\mathbf{V}}_i \mathbf{Q}$ 
 $\mathbf{r} = \mathbf{d}^{(0)} - (\hat{\mathbf{B}}\hat{\mathbf{X}} + \hat{\mathbf{X}}) \mathbf{e}_1$ 
end for


---


parfor  $m = [0 : M - 1]$ 
    Let  $\hat{\mathbf{x}}^{(m)}$  be the  $m + 1$ -th column of  $\hat{\mathbf{X}}$ 
    Use AVRORA to calculate correction  $\mathbf{x}^{(m)} = \mathbf{B} \mathbf{M}^T \mathbf{R}^{-1/2} \hat{\mathbf{x}}^{(m)}$ 
end parfor


---


function  $[D_1, D_2, \dots, D_{N_s}] = \text{func\_cluster}(\Lambda, \hat{\mathbf{v}}', \hat{\mathbf{v}}'', \xi)$ 
    Calculate weight  $w_m = v''_m \xi_m^2 \lambda_m^4 (1 - \lambda_m^{-2})$  with  $\lambda_m = (\Lambda)_{mm}$ ,  $\xi_m = \mathbf{e}_m^T \xi$ , and  $v''_m = \mathbf{e}_m^T \mathbf{v}''$ .
    Set  $z_m = \frac{v'_m}{v''_m \xi_m \lambda_m^2}$  with  $v'_m = \mathbf{e}_m^T \mathbf{v}'$ 
    for  $m = [1 : N_s]$ 
         $\bar{z}_m = \min(z) + \frac{1}{N_s} (m - \frac{1}{2}) (\max(z) - \min(z))$ 
    Continues on the next page.

```

Table C.3 (Continued): Pseudocode for the cluster search method.  $\mathbf{e}_j$  is a unit vector.

```

end for
for  $j = [1 : 1000]$ 
    (Re)initialize  $D_1, D_2, \dots, D_{N_s} = []$ 
    for  $m = [1 : M]$ 
        Find  $l$  that minimizes  $|\bar{z}_l - z_m|$ 
         $D_l = [D_l, m]$ 
    end for
    for  $m = [1 : N_s]$ 
         $\bar{z}_m = (\sum_{l \in D_m} w_m z_m)(\sum_{l \in D_m} w_m)^{-1}$ 
    end for
end function

```

---

## Appendix D: Estimation Conditional Distribution $\sigma_{EOF,i}$

In this appendix the procedure for estimating the probability distributions for  $\sigma_{EOF,i}$  given the ASCAT wind fields  $\hat{\mathbf{w}}_{obs}$  is described. These probability distributions are denoted by  $p(\sigma_{EOF,i}|\hat{\mathbf{w}}_{obs})$ . For this purpose the ASCAT observations, which are provided as a daily aggregate taken at day  $j$  are compared with the NAM wind fields at 00:00, 06:00, 12:00, 18:00 and 24:00 UTC of that day. The hour for which the root-mean square (RMS) difference between the ASCAT observations and NAM model is smallest is assigned to the ASCAT observations and denoted as  $t_j$ . The ASCAT observations are then collected into a vector  $\mathbf{w}_{obs}(t_j)$  for each time  $t_j$ . Subsequently, the wind velocities at five NDBC wind buoys (numbers 46089, 46015, 46050, 46029, 46041) measured at time  $t_j$  are added to  $\mathbf{w}_{obs}(t_j)$ . The linear operator that maps the wind velocities at time  $t_j$  from the model grid to the ASCAT and buoy observation locations is given by  $\mathbf{H}_{t_j}$ . Following the BHM approach the wind fields can be written as

$$\begin{aligned}\mathbf{w}_{obs}(t_j) &= \mathbf{H}_{t_j} \mathbf{w}_{true}(t_j) + \boldsymbol{\epsilon}_{obs}(t_j) \quad (\text{data stage}) \\ \mathbf{w}_{true}(t_j) &= \mathbf{w}_{NAM}(t_j) + \boldsymbol{\epsilon}_S(t_j) + \sum_{i=1}^{N_{EOF}} \mathbf{w}_{EOF,i} \boldsymbol{\epsilon}_{EOF,i}(t_j) \quad (\text{process stage}) \\ \sigma_{obs}^2, \sigma_S^2, \sigma_{EOF,i}^2 &\quad (\text{parameter stage})\end{aligned}\quad (\text{D.1})$$

with  $\mathbf{w}_{true}(t_j) \in \mathbb{R}^{2N_w}$  the unknown true wind field at time  $t_j$ ,  $\boldsymbol{\epsilon}_{obs}(t_j)$  the measurement error in the ASCAT/NDBC buoy wind observations,  $\boldsymbol{\epsilon}_S(t_j) \in \mathbb{R}^{2N_w}$  the error in the small-scale wind field and  $\mathbf{w}_{EOF,i} \boldsymbol{\epsilon}_{EOF,i}(t_j)$  the contribution of the  $i$ -th EOF to the large-scale error in the wind field. In the absence of knowledge about the wind observations, a priori distributions of  $\boldsymbol{\epsilon}_{obs}(t_j)$ ,  $\boldsymbol{\epsilon}_S(t_j)$ ,  $\boldsymbol{\epsilon}_{EOF,i}(t_j)$ ,  $\sigma_S^2$  and  $\sigma_{EOF,i}^2$  are

assumed to be:

$$\begin{aligned}\boldsymbol{\epsilon}_{obs}(t_j) &\sim N(\boldsymbol{\epsilon}_{obs}(t_j); \mathbf{0}, \sigma_{obs}^2 \mathbf{I}) \\ \boldsymbol{\epsilon}_S(t_j) &\sim N(\boldsymbol{\epsilon}_S(t_j); \mathbf{0}, \sigma_S^2 \mathbf{I}) \\ \epsilon_{i,EOF}(t_j) &\sim N(\epsilon_{i,EOF}(t_j); 0, \sigma_{EOF,i}^2) \\ \sigma_S^2 &\sim IG(\sigma_S^2; a_S, b_S) \\ \sigma_{EOF,i}^2 &\sim IG(\sigma_{EOF,i}^2; a_{EOF,i}, b_{EOF,i})\end{aligned}\tag{D.2}$$

with  $N(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C}) = (2\pi)^{-\frac{1}{2}D} \det(\mathbf{C})^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu}))$  the normal distribution with mean  $\boldsymbol{\mu} \in \mathbb{R}^D$  and covariance  $\mathbf{C}$ ,

$IG(x; a, b) = \Gamma(a)^{-1} b^a x^{-a-1} \exp(-\frac{b}{x})$  the inverse gamma distribution and  $\sim$  denotes that a value or vector is drawn from the given distribution. Here  $\det(\cdot)$  denotes the determinant and  $\Gamma(\cdot)$  the gamma function. Any spatial structure in the small-scale errors  $\boldsymbol{\epsilon}_S(t_j)$  is neglected. Based on ASCAT validation (*Verspeek et al.*, 2013)  $\sigma_{obs}$  is set to  $0.7 \text{ ms}^{-1}$ . We pick  $a_{EOF,i} = \frac{1}{20}N_t$ ,  $b_{EOF,i} = 0.1a_{EOF,i} \text{var}(\beta_i)$  with  $N_t$  the number of days on which ASCAT observations are available and  $\text{var}(\beta_i)$  the variance of the coefficient  $\beta_i$  in (4.18). This gives an a priori distribution for  $\sigma_{EOF,i}^2$  with mode  $\frac{b_{EOF,i}}{1+a_{EOF,i}} \approx 0.1 \text{ var}(\beta_i)$ . These values were chosen such that this mode corresponds to the  $\hat{\sigma}_{EOF,i}^2 = 0.09 \text{ var}(\beta_{L,i})$  estimate used by *Hénaff et al.* (2009) and *Vervatis et al.* (2016). Similarly,  $a_S$  and  $b_S$  are chosen to be  $a_S = \frac{2}{20}N_t N_w$  and  $b_S = \sigma_{obs}^2 a_S$  given the a priori distribution of  $\sigma_S^2$  with a mode of approximately  $\sigma_{obs}^2$ .

The conditional probability distribution for  $\sigma_{EOF,i}^2$  is given by

$$\begin{aligned}p(\sigma_{EOF,i}^2 | \underline{w}_{obs}) &= \\ \int p(\underline{\boldsymbol{\epsilon}}_S, \underline{\boldsymbol{\epsilon}}_{EOF}, \underline{\boldsymbol{\epsilon}}_{obs}, \sigma_{EOF,1}^2, \dots, \sigma_{EOF,N_{EOF}}^2, \sigma_S^2 | \underline{w}_{obs}) d\underline{\boldsymbol{\epsilon}}_S \\ \times d\underline{\boldsymbol{\epsilon}}_{EOF} d\underline{\boldsymbol{\epsilon}}_{obs} \prod_{j=1, j \neq i}^{N_{EOF}} d\sigma_{EOF,j}^2 d\sigma_S^2\end{aligned}\tag{D.3}$$

with the underbar denoting the concatenation of vectors taken at different times into one vector, e.g.  $\underline{\boldsymbol{\epsilon}}_S = [\boldsymbol{\epsilon}(t_1); \boldsymbol{\epsilon}(t_2); \dots; \boldsymbol{\epsilon}(t_{N_t})]$ . The integral on the right-hand side is approximated by drawing 500 samples of

$\mathbf{s} = (\underline{\boldsymbol{\epsilon}}_S, \underline{\boldsymbol{\epsilon}}_{EOF}, \sigma_{\beta_1}^2, \dots, \sigma_{\beta_{N_{EOF}}}^2, \sigma_S^2)$  from the distribution  $p(\mathbf{s})$  followed by the creation of a normalized histogram of  $\sigma_{EOF,i}^2$  from these samples. The sampling is carried

out using a Gibbs sampler (*Casella and George*, 1992) and consists of sequentially drawing components of  $\mathbf{s}$  under the condition that the other components remain constant. I.e. a new sample  $\mathbf{s}' = (\underline{\epsilon}'_S, \underline{\epsilon}'_{EOF}, \sigma'^2_{EOF,1}, \dots, \sigma'^2_{EOF,N_{EOF}}, \sigma'^2_S)$  is constructed from the previous sample

$\mathbf{s} = (\underline{\epsilon}_S, \underline{\epsilon}_{EOF}, \sigma^2_{EOF,1}, \dots, \sigma^2_{EOF,N_{EOF}}, \sigma^2_S)$  by sequentially drawing

1. for each  $t_j$

$$\begin{aligned}
\underline{\epsilon}'_{EOF}(t_j) &\sim p(\underline{\epsilon}'_{EOF}(t_j) | \mathbf{w}_{obs}, \underline{\epsilon}_S, \Sigma^2_{EOF}, \sigma^2_S) \\
&\sim p(\mathbf{w}_{obs}(t_j) | \underline{\epsilon}'_{EOF}(t_j), \underline{\epsilon}_S(t_j), \Sigma^2_{EOF}, \sigma^2_S) \\
&\times p(\underline{\epsilon}'_{EOF}(t_j) | \underline{\epsilon}_S(t_j), \Sigma^2_{EOF}, \sigma^2_S) \\
&\sim N(\mathbf{w}_{obs}(t_j) - \mathbf{H}_{t_j} \mathbf{w}_{NAM}(t_j) - \mathbf{H}_{t_j} \mathbf{W} \underline{\epsilon}'_{EOF}(t_j) - \mathbf{H}_{t_j} \underline{\epsilon}_S(t_j); \mathbf{0}, \sigma^2_{obs} \mathbf{I}) \\
&\times N(\underline{\epsilon}'_{EOF}(t_j); \mathbf{0}, \Sigma^2_{EOF}) \\
&\sim N(\underline{\epsilon}'_{EOF}(t_j); \sigma^{-2}_{obs} \mathbf{C} \mathbf{H}_{t_j}^T [\mathbf{w}_{obs}(t_j) - \mathbf{H}_{t_j} \mathbf{w}_{NAM}(t_j) - \mathbf{H}_{t_j} \underline{\epsilon}_S(t_j)], \mathbf{C}) \\
&\text{with } \mathbf{C}^{-1} = \sigma^{-2}_{obs} \mathbf{W}^T \mathbf{H}_{t_j}^T \mathbf{H}_{t_j} \mathbf{W} + \Sigma^{-2}_{EOF}
\end{aligned} \tag{D.4}$$

2. for each  $t_j$

$$\begin{aligned}
\underline{\epsilon}'_S(t_j) &\sim p(\underline{\epsilon}'_S(t_j) | \mathbf{w}_{obs}, \underline{\epsilon}'_{EOF}, \Sigma^2_{EOF}, \sigma^2_S) \\
&\sim p(\mathbf{w}_{obs}(t_j) | \underline{\epsilon}'_S(t_j), \underline{\epsilon}'_{EOF}(t_j), \Sigma^2_{EOF}, \sigma^2_S) \\
&\times p(\underline{\epsilon}'_S(t_j) | \underline{\epsilon}'_{EOF}(t_j), \Sigma^2_{EOF}, \sigma^2_S) \\
&\sim N(\mathbf{w}_{obs}(t_j) - \mathbf{H}_{t_j} \mathbf{w}_{NAM}(t_j) - \mathbf{H}_{t_j} \mathbf{W} \underline{\epsilon}'_{EOF}(t_j) - \mathbf{H}_{t_j} \underline{\epsilon}'_S(t_j); \mathbf{0}, \sigma^2_{obs} \mathbf{I}) \\
&\times N(\underline{\epsilon}'_S(t_j); \mathbf{0}, \sigma^2_S \mathbf{I}) \\
&\sim N(\underline{\epsilon}'_S(t_j); \sigma^{-2}_{obs} \mathbf{C} \mathbf{H}_{t_j}^T [\mathbf{w}_{obs}(t_j) - \mathbf{H}_{t_j} \mathbf{w}_{NAM}(t_j) - \mathbf{H}_{t_j} \mathbf{W} \underline{\epsilon}'_{EOF}(t_j)], \mathbf{C}) \\
&\text{with } \mathbf{C}^{-1} = \sigma^{-2}_{obs} \mathbf{H}_{t_j}^T \mathbf{H}_{t_j} + \sigma^{-2}_S \mathbf{I}
\end{aligned} \tag{D.5}$$

3. for each  $i = 1, 2, \dots, N_{EOF}$ :

$$\begin{aligned}
\sigma'_{EOF,i}^2 &\sim p(\sigma'_{EOF,i}^2 | \underline{\mathbf{w}}_{obs}, \underline{\boldsymbol{\epsilon}}'_{EOF}, \underline{\boldsymbol{\epsilon}}'_S, \sigma_S^2) \\
&\sim p(\underline{\boldsymbol{\epsilon}}'_{EOF,i} | \sigma'_{EOF,i}^2) p(\sigma'_{EOF,i}^2) \\
&\sim N(\underline{\boldsymbol{\epsilon}}'_{EOF,i} | \mathbf{0}, \sigma'_{EOF,i}^2) IG(\sigma'_{EOF,i}^2 | a_{EOF,i}, b_{EOF,i}) \\
&\sim IG(\sigma'_{EOF,i}^2 | a_{EOF,i} + \frac{1}{2}N_t, b_{EOF,i} + \frac{1}{2}\sum_j \epsilon_{EOF,i}^2(t_j))
\end{aligned} \tag{D.6}$$

4.

$$\begin{aligned}
\sigma_S'^2 &\sim p(\sigma_S'^2 | \underline{\mathbf{w}}_{obs}, \underline{\boldsymbol{\epsilon}}'_{EOF}, \underline{\boldsymbol{\epsilon}}'_S, \Sigma'_{EOF}) \\
&\sim p(\underline{\boldsymbol{\epsilon}}'_S | \sigma_S'^2) p(\sigma_S'^2) \\
&\sim N(\underline{\boldsymbol{\epsilon}}'_S | \mathbf{0}, \sigma_S'^2 \mathbf{I}) IG(\sigma_S'^2 | a_S, b_S) \\
&\sim IG(\sigma_S'^2 | a_S + N_t N_w, b_S + \frac{1}{2}\sum_j \|\underline{\boldsymbol{\epsilon}}_S^2(t_j)\|^2)
\end{aligned} \tag{D.7}$$

with  $\boldsymbol{\Sigma}$  the diagonal matrix having  $\sigma_{EOF,1}, \dots, \sigma_{EOF,N_{EOF}}$  on its diagonal and  $\mathbf{W}$  the matrix having  $\mathbf{w}_{EOF,i}$  as its  $i$ -th column. In the second lines of (D.4)-(D.7) Bayes' theorem has been used. Samples generated with the Gibbs sampler are not necessarily uncorrelated. In order to remove these correlations 10000 samples are generated with the Gibbs sampler, but only every 20th sample is retained. Calculation of the correlation in  $\epsilon_{EOF,i}(t_j)$  between subsequent retained samples indicates that the retained samples are uncorrelated: only in 5.7% of the cases is the time-correlation of  $\epsilon_{EOF,i}(t_j)$  different from zero at the 95%-significance level.

## Appendix E: Balance Operator Covariance

Temperature-temperature background error covariance in the balance operator is specified as

$$\langle \epsilon_T(\mathbf{r}_1) \epsilon_T(\mathbf{r}_2) \rangle = \sigma_{bal,T}^2 \exp\left(\frac{z}{D_{bal}}\right) \exp\left(-\frac{1}{2} \frac{(x_1-x_2)^2}{R^2} - \frac{1}{2} \frac{(y_1-y_2)^2}{R^2} - \frac{1}{2} \frac{(s_1-s_2)^2}{D_s^2}\right) \quad (\text{E.1})$$

with  $\epsilon_T$  the background error in the temperature field and  $\langle \cdot \rangle$  the expectation value. All other covariances,  $\langle \epsilon_T(\mathbf{r}_1) \epsilon_S(\mathbf{r}_2) \rangle$ ,  $\langle \epsilon_S(\mathbf{r}_1) \epsilon_S(\mathbf{r}_2) \rangle$ , etc., can be derived from this using the assumptions that

$$\begin{aligned} \epsilon_S(\mathbf{r}_1) &= -\alpha \epsilon_T(\mathbf{r}_1) \\ \epsilon_\rho(\mathbf{r}_1) &= -\alpha_T \rho_0 \epsilon_T(\mathbf{r}_1) + \beta_S \rho_0 \epsilon_S(\mathbf{r}_1) \\ \frac{\partial}{\partial z} \epsilon_v(\mathbf{r}_1) &= -\frac{g}{f \rho_0} \frac{\partial}{\partial x} \epsilon_\rho(\mathbf{r}_1) + g \frac{\partial}{\partial x} \epsilon_\zeta(\mathbf{r}_1) \\ \frac{\partial}{\partial z} \epsilon_u(\mathbf{r}_1) &= \frac{g}{f \rho_0} \frac{\partial}{\partial y} \epsilon_\rho(\mathbf{r}_1) - g \frac{\partial}{\partial y} \epsilon_\zeta(\mathbf{r}_1) \\ (0, 0) &= \int_{-H}^{\zeta} (\epsilon_u(\mathbf{r}_1), \epsilon_v(\mathbf{r}_1)) dz \end{aligned} \quad (\text{E.2})$$

with  $x$ ,  $y$  the horizontal coordinates,  $z$  the vertical Cartesian coordinate increasing in the upward direction,  $s$  the vertical s-coordinate  $\epsilon_T(\mathbf{r}_1)$ ,  $\epsilon_S(\mathbf{r}_1)$ ,  $\epsilon_u(\mathbf{r}_1)$ ,  $\epsilon_v(\mathbf{r}_1)$ ,  $\epsilon_\zeta(\mathbf{r}_1)$  the background error at location  $\mathbf{r}_1$  in the temperature, salinity, zonal velocity, meridional velocity and SSH field respectively,  $H$  the water depth,  $f$  is the Coriolis parameter,  $g$  the gravitational acceleration,  $\rho_0 = 1025 \text{ kg m}^{-3}$  the reference density,  $\alpha_T = 1.7 \times 10^{-4} \text{ }^\circ\text{C}^{-1}$ ,  $\beta_S = 7.5 \times 10^{-4}$ . The vertical length scale of the temperature-temperate covariance  $D_{bal} = 100 \text{ m}$ ,  $D_s$  is the vertical scale in s-coordinates and is chosen such that it is 50 m in 3091 m deep water.  $R = 25 \text{ km}$ , which is equal to the Rossby radius of deformation for the first baroclinic mode in this region (*Chelton et al.*, 1998). The background error standard deviation for the temperature is set to  $\sigma_{bal,T} = 0.9 \text{ }^\circ\text{C}$ . It was determined by calculating the standard deviations of

the difference between observed daily-averaged temperatures at National Data Buoy Center buoys 46015, 46022, 46027, 46029, 46041, 46050, 46087, 46088, 46089, 46094, 46211, 46229, 46243, 46244, 46248 (*NDBC*, 2016) and predictions from experiment No DA over the period 19 April 2011 to 1 October 2011 and taking the median of these standard deviations.

## Appendix F: Ensemble for the Balance Operator

The cluster search method requires a low-rank approximation of  $\hat{\mathbf{A}}^{1/2}$  with  $\hat{\mathbf{A}} = \mathbf{R}^{-1/2} \mathbf{HMBM}^T \mathbf{H}^T \mathbf{R}^{-1/2} + \mathbf{I}$ . In experiment Ens and Ens-HBSC this approximation is constructed using  $\hat{\mathbf{A}} \approx \frac{1}{M-1} \mathbf{D}\mathbf{D}^T$  (see section 4.4.3). Here, the columns of  $\mathbf{D}$  are the innovation vectors for the main run and the different ensemble members. I.e. the columns are the differences between the observations and the forecasts for those observations. As no ensemble members are available in experiment Bal-HBSC, a different approach is used. The singular value decomposition (SVD)  $\mathbf{U}\Lambda\mathbf{V}^T$  of  $\hat{\mathbf{A}}\mathbf{S}$  is calculated. Here the elements of  $\mathbf{S} \in \mathbb{R}^{D \times (M-1)}$  are drawn from a standard normal distribution,  $\mathbf{U} \in \mathbb{R}^{D \times M}$ ,  $\mathbf{V} \in \mathbb{R}^{M \times M}$  are orthonormal and  $\Lambda \in \mathbb{R}^{(M-1) \times (M-1)}$  diagonal with  $D$  the number of observations and  $M = 40$  the number of ensemble members plus main run used in experiments Ens and Ens-HBSC. The  $j$ -th column of  $\mathbf{U}$ ,  $\mathbf{u}_j \in \mathbb{R}^D$ , and the  $(j,j)$ -th element of  $\Lambda^2$ ,  $\lambda_j^2$ , then are the  $j$ -th eigenvector and eigenvalue respectively of

$$\mathbf{U}\Lambda^2\mathbf{U}^T = \hat{\mathbf{A}}\mathbf{S}\mathbf{S}^T\hat{\mathbf{A}}^T \approx (M-1)\hat{\mathbf{A}}\mathbf{I}\hat{\mathbf{A}}^T = (M-1)\hat{\mathbf{A}}^2 \quad (\text{F.1})$$

The required low-rank estimate for  $\mathbf{A}^{1/2}$  is then constructed as  $M^{-1/4}\mathbf{U}\Lambda^{1/2}$ .

