Ensemble 4DVAR (En4DVar) data assimilation in a coastal ocean circulation model. Part II: Implementation offshore Oregon-Washington, USA

Ivo Pasmans[a,1], Alexander L. Kurapov[a,b], Jack A. Barth[a,c], P. Michael Kosro[a], R. Kipp Shearman[a]

[a] *College of Earth Ocean & Atmospheric Sciences, Oregon State University, Corvallis, OR, USA*
[b] *NOAA, Silver Spring, MD, USA*
[c] *Marine Studies Initiative, Oregon State University, Corvallis, OR, USA*

Abstract

The ensemble four-dimensional variational (En4DVar) data assimilation (DA) system introduced in Part I (Pasmans and Kurapov, 2019) is tested in the coastal waters offshore Oregon and Washington, U.S. West coast, during the spring and summer of 2011. The background error covariance **B** is derived from the forecast ensemble. Satellite sea-surface temperature (SST), sea-surface height (SSH), and daily-averaged radial surface currents from high-frequency radars (HFRs) are assimilated. The performance of the En4DVar system is compared with a "traditional" 4DVAR system using a static **B**. It is found that the presence of the Columbia River plume has a profound impact on the ensemble **B**. Near the plume front the SST-SSS covariance can be up to a factor 20 larger in magnitude than in the static **B**. This introduces large spatial and temporal variability in the ensemble **B**. The En4DVar system is more successful than the 4DVAR with the static **B** preserving the temperature-salinity properties when compared to glider data.

Email address: `icpasman@uno.edu` (Ivo Pasmans)

[1] Now at the University of New Orleans, New Orleans, LA, USA

The En4DVar system also produces more accurate forecasts and analyses for temperature in the subsurface below 30 m at a buoy location on the continental shelf. In comparisons with other surface and subsurface observations En4DVar shows consistent, albeit not significant, improvement over traditional 4DVAR. Large surface temperature-salinity covariances in combination with the episodic occurrence of large-scale errors in the SST observations lead to erroneous freshening in the centre of the model domain. Adding constraints on the surface salinity corrections based on the prior model reduces this effect.

Keywords: 4DVAR, Data assimilation, Coastal Oceanography, Ensemble Background Error Covariance, River Plume, USA, Oregon

1. Introduction

For the benefit of the local fishing communities, government agencies and other users, the Oregon State University (OSU) coastal ocean forecast system has provided forecasts of temperature, salinity, currents, and other oceanic fields of interest in the Oregon-Washington (OR-WA) coastal area (http://nvs.nanoos.org/Explorer). While the system does produce useful forecasts, we continue exploring ways to improve its performance. In this system, initial conditions for the forecasts are corrected by assimilating surface observations using the 4DVAR data assimilation (DA) algorithm in a series of 3-day windows. This requires specification of the forecast, or 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 ocean state vector, including SSH and three-dimensional fields of the horizontal velocity, temperature, and salinity. The balance operator uses diagnostic relations such as geostrophy, thermal wind balance, the linearised equation of state, and a simplified, linear temperature-salinity relation based on multiyear glider observations. Details on the balance operator used can be found in Appendix B.

Coastal waters in the OR-WA area are very dynamic, with wind-driven currents in excess of 0.5 $\mathrm{m\,s^{-1}}$, strong temperature fronts, geostrophic and ageostrophic baroclinic instabilities, and jets separating from the shelf toward the ocean interior (Koch et al., 2010). The freshwater outflow from the Columbia River creates a shallow plume, hereafter referred to as "the plume", that spreads over a large area (Berdeal et al., 2002; Hickey et al., 2005; Huyer et al., 2005; Liu et al., 2009). The location of the Columbia River plume changes on seasonal and shorter time scales in response to the winds. In summer, due to the predominantly southward winds, the Columbia River plume is transported to the south of the river mouth and offshore, with the coastal upwelling. During periods of wind relaxation the plume is advected toward the coast, freshening the coastal waters off Oregon. In such a dynamic area the utility of a static $\mathbf{B}$ can be limited.

One way to capture the time-varying dynamics in $\mathbf{B}$ is to estimate it from an ensemble of perturbed model runs. This approach has been tested in meteorology with varying results. Kuhl et al. (2013) found that replacing the static $\mathbf{B}$ with one coming partially from an ensemble 3DVAR system in the Naval Research Laboratory Atmospheric Variational Data Assimilation

System-Accelerated Representer (NAVDAS-AR) system reduced background errors. In contrast, running an ensemble Kalman filter to produce $\mathbf{B}$, the Met Office found that a system using a pure ensemble $\mathbf{B}$ resulted in an overall degradation of performance and produced forecasts with larger root-mean square errors (RMSE) compared to a system using a static climatological background covariance (Lorenc and Jardak, 2018). In this study we aim to compare 4DVAR with static $\mathbf{B}$ and ensemble $\mathbf{B}$, both applied to a realistic ocean model. In Part I (Pasmans and Kurapov, 2019), the En4DVar method generating the ensemble $\mathbf{B}$ was described. The OR-WA coastal ocean forecast system was used to illustrate the computational efficiency of the cluster search minimisation 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, 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 recognise one of the general potential dangers using En4DVar for poorly observed fields. By design, an ensemble yields large variances, and hence covariances, in frontal areas. At the same time, if observations are affected by large-scale errors or biases in these areas, unobserved fields can receive a large and erroneous correction.In our case, satellite SST observations are occasionally found to contain large-scale spatial biases. Combined with the large and negative temperature-salinity covariance in the Columbia River plume and along the Oregon coastal upwelling front, assimilation of the biased SST results in unrealistically large corrections to the surface salinity. This issue motivated us to test methodology to constrain the near-surface salinity based on the forecast solution, the so-called salinity constraint (SC) procedure.

This paper is organised as follows: section 2 provides a summary of the ocean model, En4DVar DA, and the 4DVAR DA system. In this section, we also introduce the SC procedure developed to deal with the erroneous corrections to the salinity discovered during this study. Horizontal and vertical spatial patterns of the static and the ensemble $\mathbf{B}$ are compared in section 3. The problems with the salinity field are illustrated in section 4. Analyses and forecasts from the En4DVar system are compared to the model without assimilation and with the standard 4DVAR in section 5. In section 6, our findings are summarised and some other approaches not considered here, e.g., hybrid covariances are briefly discussed.

2. Model experiments

In the OR-WA system, nonlinear forecasts are obtained using the Regional Ocean Modeling System (ROMS) (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 (ζ). The salinity in this paper is reported according to the PSS-78 standard and, following IAPSO (1985) oceanographic standard, is reported as dimensionless quantity. The 4DVAR DA utilises the AVRORA tangent linear and adjoint codes that were developed by our group at OSU and that are not part of the community code developed by the broader ROMS community (Kurapov et al., 2009, 2011; Yu et al., 2012). Model forcing includes wind stress and surface heat flux derived using ROMS's bulk flux formulation (Fairall et al., 2003) and input from the North-American Mesocale model (NAM) (NOAA, 2011a), tides from the Topex database (Egbert and Erofeeva, 2010) and river inputs based on United States Geological Survey discharge measurements (USGS, 2011). Non-tidal lateral boundary conditions are obtained from the HYCOM global model analyses (COAPS, 2015). The data assimilation period runs from 19 April till 1 October 2011, covering a summer upwelling season with the anomalously large Columbia River discharge (Mazzini et al., 2015). Precipitation-evaporation is low in the region in summer and is not included in the model. More details of the model forcing can be found in Pasmans et al. (2019) and Pasmans and Kurapov (2019).

Table 1: Overview of the experiments in this study.

Experiment	DA	$\mathbf{B}$	salinity constrained
No DA	no	-	no
Ens	yes	ensemble	no
Ens-SC	yes	ensemble	yes
Bal	yes	balance operator	yes

Four experiments are discussed in this paper. Experiment *No DA* starts using the interpolated HYCOM analysis as initial condition on 2 January 2011 and is run continuously without DA. All DA experiments use the *No DA* ocean state on 19 April 2011 as the initial condition. In all the DA cases, the model is propagated forward in time as a series of three-day windows. Initial conditions for the control run (cases *Bal*, *Ens*, and *Ens-SC*) and each ensemble member (*Ens*, *Ens-SC*) are updated at the beginning of each window by the assimilation of the GOES satellite sea-surface temperature (SST), daily-averaged radial current observations from high-frequency radars (HFR) and along-track satellite sea-surface height (SSH) altimetry. See Pasmans and Kurapov (2019) for the details on the data sets used. Prior to assimilation the resolution of the SST and HFR observations is reduced to the resolution of the tangent linear and adjoint models by averaging the SST observations over $4 \times 4\,\mathrm{km}$ horizontal cells and radial HFR data over $5 \times 5\,\mathrm{km}$ cells. The 24h-average of the tides based on the harmonical analysis of the *No DA* model is added to the detided absolute dynamic topography SSH observations. In the DA system, this is matched against the 24h-average of the model SSH, with the model and observed means removed along each satellite track. No in-situ observations are assimilated because of inherent problems assimilating very sparse hydrographic profiles (Pasmans et al., 2019). In every assimilation-forecast cycle, 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 as the forecast.

Corrections $\mathbf{x}^{(m)}$ to the ocean state at the beginning of each DA window for both the control run and each of the ensemble members m are calculated by maximising the conditional probability

$$p(\mathbf{x}^{(m)}|\mathbf{d}^{(m)}) \quad \sim \quad p(\mathbf{d}^{(m)}|\mathbf{x}^{(m)})p(\mathbf{x}^{(m)}) \tag{1}$$

with respect to $\mathbf{x}^{(m)}$. Here $\sim$ indicates that the two sides are equal apart from a proportionality constant. The equality in (1) follows from Bayes' theorem, $\mathbf{d}^{(m)}$ is the vector with innovations, i.e. the differences between observations and the nonlinear forecast predictions for the observations, $p(\mathbf{x}^{(m)}|\mathbf{d}^{(m)})$ the probability that the background error is $-\mathbf{x}^{(m)}$ given the innovations $\mathbf{d}^{(m)}$, $p(\mathbf{d}^{(m)}|\mathbf{x}^{(m)})$ the probability that the innovations are $\mathbf{d}^{(m)}$ given that the background error is $-\mathbf{x}^{(m)}$ and $p(\mathbf{x}^{(m)})$ is the a priori probability distribution for the background errors. It is usually assumed that the aforementioned probabilities follow normal distributions, i.e. $p(\mathbf{d}^{(m)}|\mathbf{x}^{(m)}) \sim \exp(-J_{obs}(\mathbf{x}^{(m)}))$, $p(\mathbf{x}^{(m)}) \sim \exp(-J_b(\mathbf{x}^{(m)}))$ with $J_b(\mathbf{x}^{(m)}) = \frac{1}{2}\mathbf{x}^{(m),T}\mathbf{B}^{-1}\mathbf{x}^{(m)}$ and

$$J_{obs}(\mathbf{x}^{(m)}) = \frac{1}{2}(\mathbf{d}^{(m)} - \mathbf{HMx}^{(m)})^T\mathbf{R}^{-1}(\mathbf{d}^{(m)} - \mathbf{HMx}^{(m)})$$

. Here $\mathbf{H}$ is the sampling operator that generates predictions for the different observations from the model output and $\mathbf{M}$ the tangent linear model, i.e. the ROMS model linearised around the forecast. The observation error covariance $\mathbf{R}$ is assumed to be diagonal. In this case, the maximisation of $p(\mathbf{x}^{(m)}|\mathbf{d}^{(m)})$ in (1) is equivalent to the minimisation of the cost function (Courtier et al., 1994; Egbert et al., 1994):

$$J(\mathbf{x}^{(m)}) = J_{obs}(\mathbf{x}^{(m)}) + J_b(\mathbf{x}^{(m)}) \tag{2}$$

In experiments *Ens* and *Ens-SC* an ensemble of 40 forecasts is carried throughout the study period. The control run is computed using the wind forcing without perturbations. The other 39 runs are carried using perturbed winds and perturbed observations (see Pasmans and Kurapov (2019, sec. 4)). The model fields for SSH, salinity, temperature and velocity are averaged over a period beginning 12h prior to the DA window start time and ending 12h after the start time. The time-averaged fields of the 39 runs are then utilised to compute $\mathbf{B}$. Based on the findings in Pasmans and Kurapov (2017) localisation is applied in the horizontal such that zero correlations are imposed on points spaced further than 100 km apart. The wind perturbations are linear combinations of the empirical orthogonal functions (EOFs) derived from the NAM wind fields with their coefficients drawn from normal distributions. Daubechies wavelets are added to these wind perturbations to represent the small-scale errors in the wind field. Distributions of the EOF coefficients are determined by comparison of the model wind field with ASCAT scatterometer data using a Bayesian Hierarchical Model (Pasmans and Kurapov, 2019). All ensemble members start off from the same initial state that is taken from experiment *No DA* on 10 March 2011 and are run till 19 April with perturbed winds, but without DA.

As the ensemble perturbations to the ocean state are generated from the perturbations in the wind forcing using the physics that is contained in the model, it is assumed that the ensemble perturbations are consistent with the ocean dynamics. Therefore, it is expected that the DA corrections are in leading order in geostrophic balance and that the amplitude of the transient solutions that emanate from the DA corrections are sufficiently small that no model blow-ups occur. Indeed, in the control run of experiments *Ens* and *Ens-SC* no blow-ups are encountered. Sporadically, a blow-up is encountered in one of the ensemble members. These blow-ups, however, are not due to the DA corrections, as rerunning the ensemble member with different wind forcing perturbations resolves the problem. The solution $\mathbf{x}^{(m)}$ that minimises the cost-fuction J in (2) for each ensemble is approximated using a numerical iterative solver. In case of the popular Reduced B-conjugate gradient (RBCG) method (Gürol et al., 2014) the approximation to $\mathbf{x}^{(m)}$ would be sought in a subspace that is expanded by 1 dimension in each inner loop iteration. For En4DVar this method was deemed too time- and resource consuming. Instead, the newly developed cluster search method is used to approximate $\mathbf{x}^{(m)}$. In the cluster search method, the search subspace is expanded by $N_s \geq 1$ dimensions per inner loop iteration. If $N_s = 1$ the cluster search method is identical to the RBCG method. For $N_s > 1$ the cluster search still searches for the same solution as ordinary 4DVar, i.e. $\mathbf{x}^{(m)}$ that minimises equation (2) or (3), but it converges faster than RBCG. Full details on the cluster search method can be found in Pasmans and Kurapov (2019, sec. 3). For practical reasons, we use the cluster search method with $N_s = 4$ and 12 inner loop iterations in this study. In experiment *Bal* no ensemble is available. In order to use cluster search, and thus accelerate convergence of the traditional 4DVAR, a low-rank surrogate ensemble is generated as explained in Appendix C.

It is found in case *Ens* that unphysical DA corrections to the sea-surface salinity (SSS) can occur. To reduce this effect, salinity constraints (SC) are added to the 4DVAR cost function (2). These constrains "nudge" the DA correction towards the prior model salinity. We do not want to nudge SSS to the prior model point-by-point since this would suppress changes to the location of the river plume front. Constraining simply the area-averaged SSS does not effectively mitigate the adverse effect of the biased SST, since the domain-averaged SSS can be preserved when salinity is wildly redistributed. Instead, we cover the surface area in the model interior with a hierarchy of boxes or rectangles (Figure 1). The largest box spans the entire domain (area inside the blue edges in Figure 1), excluding the domain edges and also excluding Puget Sound in the northeast corner of our domain. The next level of boxes (areas inside the green edges) is obtained by dividing the largest box (blue edges) into 4. Level 3 boxes (areas bounded by the red edges) are obtained by dividing level 2 boxes (areas bounded by the green edges) by 4 and so on. In order not to impede corrections on the scale of the background error covariance ($R = 25\,\text{km}$) only five levels of boxes are used (the areas bounded by the blue, green, red, cyan edges respectively plus a level not shown in Figure 1). In this way, the smallest boxes have dimensions of 30×65 km. The prior, forecast SSS from a earlier window can then be averaged in each of these boxes at the beginning of the current assimilation window by a sampling operator $\mathbf{H}_S$. These averaged salinity values are treated as additional observations. The computational cost of implementing this scheme in the 4DVAR is negligible. Using the incremental 4DVAR formulation (2), the innovation vector $\mathbf{d}^{(m)}$ corresponding to these synthetic data is zero and an additional term is added to cost function (2). This term provides a penalty on the deviations of the box-averaged SSS from the prior:

$$J_{SC} = \sum_i \frac{[\Delta S_i]^2}{\sigma_{S,i}^2} \tag{3}$$

Here ΔS_i is the DA change in the box-averaged salinity of box i and $\sigma_{S,i}$ some specified variability. We have opted to base these values for $\sigma_{S,i}$ on the natural variance in experiment *No DA* for two reasons. First, 4DVAR corrections are applied to all ensemble members and consequently the large salinity corrections can exaggerate the ensemble spread resulting in unrealistically large estimates of $\sigma_{S,i}^2$ from the ensemble. Second, the additional penalty terms serve solely to limit SSS corrections to climatologically realistic values. Contrary to the *No DA* solution, the 4DVAR ensemble is not available prior to

DA and therefore cannot be used to estimate climatological values. Specifically, the value for $\sigma^2_{S,i}$ has been determined by first calculating the box SSS in *No DA* at the beginning of each window, then taking differences in the box SSS between consecutive windows and after this defining $\sigma^2_{S,i}$ as 5% of the variance in these SSS differences for box i. This yields average values of the standard deviation $\sigma_{S,i}$ of 0.058, 0.12, 0.17, 0.21, and 0.23 for the boxes at levels 1 through 5 respectively. As the period covered in the case *No DA* is limited to one upwelling season, it is possible that in reality the plume front moved beyond the maximal extent of the plume in experiment *No DA*. In that case, the values for $\sigma_{S,i}$ obtained using the aforementioned procedure would underestimate the background error in the SSS. The application of salinity constraints under these conditions would lead to unnecessary local suppression of DA corrections to the SSS.

Even though in practice the salinity constraints are implemented as additional observations, they are no actual observations and J_{SC} depends solely on the background errors in the model SSS. Therefore, in the interpretation of DA based on Bayes' theorem, the addition of J_{SC} to the cost function is equivalent with replacing the a priori background error probability distribution $p(\mathbf{x}^{(m)})$ in (1) with $p(\mathbf{x}^{(m)}) \sim \exp(-J_b(\mathbf{x}^{(m)}))\exp(-\frac{1}{2}\mathbf{x}^{(m),T}\mathbf{H}_S^T\mathbf{\Sigma}^{-2}\mathbf{H}_S\mathbf{x}^{(m)})$, where $\mathbf{\Sigma}$ is a diagonal matrix with the values σ_i on its diagonal. I.e. DA corrections that change surface salinity on scales equal or larger than the scale of the smallest boxes become less probable than they would be without the addition of J_{SC}. In this way, SSS corrections on scales smaller than 20 km are not impacted by the additional penalty, while large scale corrections will result in changes of box-averaged salinity in multiple boxes at multiple levels, thus rapidly increasing J_{SC}.

The DA case in which (3) is added to cost function (2) is referred to as *Ens-SC*. Experiment *Bal* uses the standard 4DVAR with the static, balance operator $\mathbf{B}$ as described in Kurapov et al. (2011) and Appendix B. Experiment *Bal* also uses the SC to enable a fair comparison with *Ens-SC*, the En4DVar experiment with physically acceptable salinity fields.

3. Covariances

In this section, spatial structures in the static, 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 of 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.

Surface maps of the background error correlation between T at a point far offshore, $\mathbf{r}_0 = (127.57\,^\circ\mathrm{W}, 47.00\,^\circ\mathrm{N})$, and fields of T, S, and surface currents are shown in Figure 2. In the balanced $\mathbf{B}$ (Figure 2a,b), the SST-velocity correlation, plotted as a vector field, exhibits an anticyclonic eddy, consistent with lowering of the isopycnal surfaces in the core of such an eddy. Contrary to other implementations of the balance operator in oceanography (Balmaseda et al., 2008, 2013; Weaver et al., 2005) our implementation of the balance operator does not have an unbalanced component. I.e. it is assumed that all background errors can be derived from the background error in the ocean temperature. As a consequence, the static balanced T-S cross-correlation (Figure 2b) is, apart from a sign change, equal to the static balanced T-T correlation (Figure 2a). 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 of this location peaks at 0.41.

We hypothesise that far offshore, away from strong coastal fronts and jets, the ensemble $\mathbf{B}$ computed using daily-averaged member fields yields correlation structures that are overall close to the static, balanced $\mathbf{B}$. To test this, ensemble perturbations from the window ensemble mean are collected from all ensemble members and all windows in experiment *Ens*. These $N \times N_t$ perturbations, with N_t the number of DA windows and $N = 39$ the number of ensemble members, forms a set to which we will refer as the "super-ensemble". The correlation structures obtained from this super-ensemble (Figure 2c,d) are qualitatively similar to the static balanced $\mathbf{B}$ (Figure 2a,b). A more quantitative assessment reveals that the $T(\mathbf{r}_0)$-$S(\mathbf{r}_0)$ correlation is only -0.58, compared to the value of -1 in the balanced $\mathbf{B}$. This is partly due to the fact that mean vertical profiles of T and S are different from window to window, affected by the wind-driven mixing and other conditions. The cross-correlation of $T(\mathbf{r}_0)$ with the surface velocity field is also weaker, with peak magnitudes reaching 0.15-0.21.

Correlation maps for locations near the upwelling front reveal more complicated horizontal structures, particularly if computed for a single assimilation window. Figure 3 shows the surface ensemble error correlation for a point chosen at the inshore edge of the river plume, $S = 31.5$, on 9 July 2011. The fresher, warmer plume water is found to the west and saltier and colder upwelled water to the east of this location. The error in the surface $T(\mathbf{r}_0)$ is strongly correlated with the error in 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 largest is marked by a green dot. The largest T-S correlation in this surface map is also large in magnitude, -0.86.

To analyse the impact of the river plume on the T-T and T-S ensemble error correlations, fifty 200-km meridional sections are selected randomly in each assimilation window. Along each section the ensemble-averaged SSS is determined using the same ensemble of 24h-averaged ocean states that is also used in the calculation of the ensemble-based background covariance. If the minimum ensemble-averaged salinity along the section is larger than 31.5, the section is classified as outside the plume. If the maximum ensemble-averaged 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 using all 39 ensemble members (excluding the control forecast run). The localised ensemble correlations found in this way along the different sections are shown in Figure 4 (grey lines). The average ensemble correlations (dashed lines) are obtained by first removing the ensemble mean from each section and combining the ensemble perturbations from all different sections, spanning all windows into one "super-ensemble". The correlation is then calculated from this super-ensemble. The T-S cross-correlations in individual sections and windows (grey lines) vary widely, e.g., taking either sign outside the plume (Figure 4c). Consequently, the averaged cross-correlation at the central point, -0.25, is only a fraction of the -1.0 cross-correlation used in the balance operator. Within the plume (Figure 4d), the averaged T-S ensemble correlation is closer to the balance operator correlation, being mostly negative and yielding the averaged value of -0.73 at the section centre. Qualitative analysis of the satellite SST imagery and surface model fields suggests that the Columbia River plume water is relatively warmer than the ambient ocean, partly since the river source is warmer and possibly also since strong stratification inhibits downward vertical turbulent heat flux at the base of the river plume. The higher temperature of the plume compared to the surrounding ocean explains the more definitive negative T-S error correlation within the plume compared to areas outside.

In the balanced $\mathbf{B}$, the T-T and T-S horizontal correlations have a Gaussian shape. The correlation is thus 60% and 14% of its peak value at distances $R = 25$ km and $2R = 50$ km, respectively, where R is the horizontal correlation length scales assumed in the balance operator covariance. For the averaged ensemble correlations in Figure 4, we determine the distances from the centre r at which the average correlations have dropped to 60% and 14% of their value at $r = 0$. For the correlation without localisation these distances are found to lie between 14-20 and 40-54 km. While the latter is comparable to the 50 km value in the balance operator, the former is smaller than its balance operator equivalent. So, although both correlations decrease to 14% of their maximum values in about the same distance, the ensemble correlations decrease faster than the Gaussian. This is also directly visible in Figure 4a,b. As expected, the localisation scheme creates an additional reduction in scale: for the localised correlations the above-referenced correlation benchmarks lie between 12-16 and 34-40 km.

Not only are the background error correlations produced by the ensemble different from those produced by the balance operator, but so are the error covariances. Figure 5 shows surface maps of the ensemble error SST variance, at the top, and the T-S covariance, at the bottom. In the T-S covariance maps T and S ensemble perturbations are sampled at the same point. Three windows, 1 May, 30 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 favourable 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.

On average, the ensemble variance in SST (Figure 5a-c) is considerably smaller than the one assumed in the balanced $\mathbf{B}$: the latter is $0.81\,^\circ\mathrm{C}^2$, while the former has a median value of $0.10\,^\circ\mathrm{C}^2$. However, locally, in frontal areas, where SST assimilation will have the largest impact, the ensemble standard deviation can obtain much larger values, up to $1.63\,^\circ\mathrm{C}^2$. These high variances are found, in particular, near the 15°C isotherm (solid black line). Along the coast between 41-48°N, this isotherm is indicative of the location of the cold upwelling front. Similarly (Figure 5d-f), the median of the T-S covariance over all windows is $-0.002\,^\circ\mathrm{C}$, which is two orders of magnitude smaller than the $-0.13\,^\circ\mathrm{C}$ used in the balanced $\mathbf{B}$. Once again, locally the strength of the

SST-SSS ensemble covariance can become a factor 20 larger in magnitude than that of the balanced covariance, reaching $-2.7\,^{\circ}$C. Similarly to the large SST-variances these large SST-SSS covariances can be found at frontal locations, both at the inshore and offshore edge of the river plume. However, in contrast to the SST-variances, they can also be found within the river plume (see Figure 5e,f).

The point-by-point ensemble covariance between SST and the daily-averaged surface current in each direction as well as the surface velocity variance in each direction are calculated for the same dates. Then, for each direction, the absolute value is taken. This generates a (co)variance ellipse for each location comparable to tidal ellipses obtained in the tidal current analysis. The length of the major semi-axes of these ellipses, i.e. the maximum absolute values of the SST-surface current covariance and surface current variance are shown in Figure 6a-c and Figure 6d-f respectively. The areas of the largest SST-velocity covariance match the locations of the largest SST variances (cf. Figure 5a-c). In contrast, the largest surface current variances (Figure 6d-e) are found within the plume. This is notwithstanding that the ensemble variability in the 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 the Ekman depths 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). Hence, the Ekman transport in the plume is distributed over a relatively shallower water column and consequently currents in the plume are more sensitive to perturbations to the wind forcing.

4. The DA impact on the Columbia River Plume volume

In the initial En4DVar runs performed without SC, 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 volume, grid cells for which $S < 31.5$ are only considered. The amount of fresh water contained in each such grid cell, $V_{fresh,ijk}$, is computed using mass conservation:

$$S_{ijk}V_{ijk} = S_{river}V_{fresh,ijk} + S_{ocean}(V_{ijk} - V_{fresh,ijk}), \tag{4}$$

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 7 shows the difference in V_{fresh} from its value on 19 April 2011, the staring 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 model source of fresh water in the plume region. Since 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 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 set by the cumulative river outflow. Experiment *Ens-SC* 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-SC*. However, their magnitude is small enough such that the fresh plume water volume difference in experiment *Ens-SC* stays below the cumulative river outflow. The problems with constraining the fresh plume volume are exclusive to En4DVar: in Experiment *Bal* the change in freshwater volume is comparable to that in experiment *No DA*.

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 8a,b show the SSS field before and after the DA correction on this day. During this period the plume region is fully covered by the available SST data (Figure 8c). Without the additional salinity constraint, the DA correction in SSS expands the plume area to the west and northwest of the plume and decreases the salinity within the plume by approximately 3. Further analysis suggests that these large erroneous freshwater volume changes stretching over the whole plume area are associated with a bias in the assimilated SST observations compared to the forecast SST which is exaggerated by the large magnitude of the T-S covariance in the plume area (Figure 8c). On this day, the *Ens* forecast is colder than the observed SST over the majority of the model domain, on average by $0.5\,^{\circ}$C. If contour lines for the SST-SSS covariance are laid over the SSS DA correction (Figure 8d), it becomes apparent that the instantaneous freshening, introduced by DA, takes place in the area where large SST-SSS covariances overlap with regions in which the SST forecast is colder than the observations.

In this study, the SC are also applied to experiment *Bal*. Although this study did not investigate the need for salinity constraints in the balance operator case in more depth, the abnormally large SSS corrections observed in experiment *Ens* have never been noticed in the operational OSU coastal forecasting system, which uses a balance operator without salinity contours. So, there seems to be no indication that the salinity constraints in experiment *Bal* are essential to ensure physically realistic corrections to the SSS. A possible explanation for this difference between the balance **B** and ensemble **B** is that the surface temperature-salinity covariances in the balance **B** used ($-0.13 \leq cov_{TS} \leq 0\,°\mathrm{C}$) are an order of magnitude smaller than those that appear in and around the plume in the ensemble-based **B** (ranges in the plume region from $\approx -1.4°\mathrm{C}$ and $-0.1\,°\mathrm{C}$). Consequently, the SSS corrections created with the erroneous SST observations do not become non-physically large when the balance **B** is used.

5. En4DVar versus balance-operator 4DVAR

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.

5.1. Surface

Here model results are compared with surface remote sensing observations. The same data sets that are assimilated are used in this comparison, but with different processing. The hourly GOES SST observation resolution is not reduced. 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) and instead of presenting the along-track altimetry as the 24h-averaged sea level anomaly, it is compared to the instantaneous SSH minus the along-track mean. In this case, model tidal predictions obtained from case *No DA* using Pawlowicz et al. (2002) are added to the detided SSH observations prior to comparison with the model SSH. 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 9 shows the time series of the area-averaged, 3-day time-averaged RMSE for each window. The continuous blue line corresponds to the case *No DA*. Each short line segment is associated with one six-day analysis-forecast cycle. The value at the left side is the RMSE over the first three days of this cycle, i.e. the analysis period, while the value on the right-hand side is the RMSE for the last three days, i.e. the forecast. The opaque bands around the lines represent the 90%-confidence intervals indicating what other RMSEs could have been obtained if the observations had been spread over the domain differently and are constructed using bootstrapping. Details on the bootstrapping method used can be found in Appendix A. Analysis RMSEs are consistently smaller than forecast RMSEs even though for individual

Table 2: The time average of the daily-averaged analysis RMSE (combining analysis days 1, 2, and 3 from each cycle) with 90%-confidence interval, 19 April to 1 October 2011. Smallest averaged RMSE for each observation type is shown in bold.

		Analysis		
	No DA	Ens	Ens-SC	Bal
SST [°C]	1.17	0.78	0.78	**0.78**
	±0.06	±0.02	±0.02	±0.03
$u_{surface}$ [$\mathrm{cm\,s^{-1}}$]	16.4	11.1	**11.0**	11.7
	±1.0	±0.5	±0.4	±0.5
$v_{surface}$ [$\mathrm{cm\,s^{-1}}$]	19.2	**12.5**	12.5	13.1
	±1.0	±0.5	±0.5	±0.5
SSH [cm]	7.1	**5.0**	5.1	5.2
	±0.6	±0.6	±0.6	±0.6

windows the difference is not always significant at the 90% level. Forecast and analysis RMSEs are also predominantly smaller than the RMSEs in *No DA*. Initially, however, improvement in the reduction is not significant at the 90%-level. As the upwelling seasons progresses RMSEs in all experiments increase, but faster in *No DA* than in the other experiments. Eventually, RMSEs in SST, SSH and meridional velocities are significantly smaller in the DA experiment than in *No DA*. Table 2 and Table 3 present a summary of the RMSE, providing in each case the RMS of the daily averaged RMSEs, taking into account analysis or forecast days 1, 2, and 3 from each DA cycle. All DA runs yield RMSEs that are close to each other and all are a significant improvement over case *No DA* at the 90%-level. However, no one DA experiment is significantly better than the others. This is partly a testament

Table 3: The time average of the daily-averaged forecast RMSE (combining forecast days 1, 2, and 3 from each cycle) with 90%-confidence interval, 19 April to 1 October 2011. Smallest forecast RMSE for each observation type is shown in bold. Every three-day forecast period is preceded by a three-day analysis.

		Forecast		
	No DA	Ens	Ens-SC	Bal
SST [°C]	1.17	**0.95**	0.95	0.99
	±0.06	±0.04	±0.04	±0.04
$u_{surface}$ [$\mathrm{cm\,s^{-1}}$]	16.4	**12.6**	12.7	13.2
	±1.0	±0.5	±0.5	±0.6
$v_{surface}$ [$\mathrm{cm\,s^{-1}}$]	19.2	**14.1**	14.3	14.8
	±1.0	±0.6	±0.7	±0.6
SSH [cm]	7.1	**5.6**	5.8	5.8
	±0.6	±0.5	±0.6	±0.5

to a reasonably good presently operating OR-WA ocean forecast system, using the static balanced **B**, and partly the result of extensive observational coverage of the surface.

A more detailed analysis of the model skill against the surface data is done using the Taylor diagrams (Figure 10) where standard deviations normalised by the observational standard deviation and correlations with the observed values are shown for the forecasts and analyses. The crosses mark the extent of the 90%-confidence interval in the correlation and normalised standard deviation. The model skill, here defined as (Taylor, 2001),

$$S = \frac{4(1+\rho)^4}{2(\sigma_{model}\sigma_{obs}^{-1} + \sigma_{obs}\sigma_{model}^{-1})} \tag{5}$$

is indicated by green lines in the diagrams. Measured by skill, experiment *No DA* performs significantly worse than the DA experiments for all four types of surface observations shown. This is mainly because its correlation with the observations is closer to zero than for the DA cases. The analyses in the DA experiments are a significant improvement over the forecast. The ranking of the different DA experiments is the same for the forecasts as for the analyses, e.g. if the forecast of experiment *Bal* has the largest standard deviation of the three forecasts, it also has the largest standard deviation of the three analyses. By any measure the model provides better predictions for SST (Figure 10a) than for any of the other surface observations shown. We attribute this to the fact that large part of the variability of the SST around its mean is caused by the north-south gradient in SST and the seasonal heating of the ocean surface over the model period. These large-scale, long-term processes are captured well by the model dynamics. With the exception of SSH (Figure 10d), for which the standard deviation of the En4DVar is closer to the observational standard deviation than in experiment *Bal*, the DA experiments do not perform significantly different.

5.2. 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-SC* yields improvement over *Bal* 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°N (Erofeev, 2011) (Figure 11). 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., 2002a). During episodes of strong upwelling driven by the southward winds, the glider samples the cold and salty upwelled water with a potential density $> 26.5\,\mathrm{kg\,m^{-3}}$ in the shallower portion of the transect (Austin and Barth, 2002; Barth et al., 2005b; Huyer, 1977; Pasmans et al., 2019). As the glider goes farther offshore, it crosses the Columbia River front, with relatively warmer and fresher water near the surface.

T-*S* diagrams are presented in Figure 12 using glider observations and model solutions. Results for *Ens* are similar to those in *Ens-SC* are therefore not included in Figure 12. Diagrams at the top correspond to the the first half of the study period, 19 April through 29 June 2011; diagrams at the bottom correspond to the second half of the study period, 30 June through 1 October. In the beginning of the upwelling season, observations show a cloud of points corresponding to the shallow river plume ($S < 31.5$) (Figure 12a). The *T*-*S* diagram in experiment *No DA* (Figure 12b) is qualitatively similar to the observed. A problem emerges in case *Bal* (Figure 12c) where a line of points builds up along $S = -\alpha T$ enforced by the very simple choice of the *T*-*S* relation in the balance operator (B.2), literally assuming this linear relation between corrections to S and T. This line is absent in case *No*

DA (Figure 12b) and the verification data (Figure 12a). We mark the top portion of this line ($T > 13.5$°C in Figure 12c) with black dots, and colour their counterparts in every diagram black as well. The low salinities of the black points in the observations (Figure 12a) indicate that they correspond to samples taken while the glider was in the river plume. Qualitatively, *Ens-SC* DA is more successful in reproducing the plume T-S composition during this time period (Figure 12d); in particular the artificial line implied by the specific choice in the balance operator is gone. As the upwelling season progresses, the cold water front and the river plume move farther offshore. Under these conditions the glider samples only inshore of the plume and the observed T-S diagram reverts to a straight line with the slope close to $-\alpha^{-1}$ (Figure 12e). Consequently, the forecasts in experiment *Bal* are able to correctly simulate the observed T-S relation (Figure 12g). 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 12h).

RMSE between the glider data and model analyses or forecasts are calculated separately close to the surface (depth$<$ 22 m) and below the seasonal thermocline (depth$>$ 50 m) and the results are shown in Table 4. Similarly to the RMSE for the surface observations (Table 2 and Table 3) the glider RMSE for the ensemble DA experiments is generally better than for experiment *Bal*, but each difference by itself is not significant at the 90% level. Below 22 m.

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 11). Temperature and salinity are measured by sensors at different

Table 4: The time-averaged RMSE (combining analysis or forecast days 1, 2, and 3 from each cycle) in glider observations for *No DA* and the analyses and forecasts together with their 90%-confidence interval. Model results are compared to the glider observations available between 19 April to 1 October 2011 in the top 22 m and below 50 m depth. Note that salinity is a dimensionless quantity.

		Analysis			*Forecast*		
	No DA	Ens	Ens-SC	Bal	Ens	Ens-SC	Bal
Above 22 m *depth*							
glider T [°C]	1.27	1.31	1.24	1.44	1.25	1.21	1.48
	±0.15	±0.15	±0.19	±0.16	±0.14	±0.21	±0.19
glider S	0.86	0.68	0.71	0.76	0.69	0.71	0.78
	±0.12	±0.11	±0.14	±0.13	±0.10	±0.13	±0.13
Below 50 m *depth*							
glider T [°C]	0.83	0.68	0.79	0.91	0.71	0.82	0.91
	±0.04	±0.03	±0.04	±0.06	±0.04	±0.05	±0.05
glider S	0.20	0.20	0.20	0.23	0.20	0.21	0.23
	±0.02	±0.01	±0.02	±0.02	±0.02	±0.02	±0.02

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 subtitdal 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 13 (temperature) and Figure 14 (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 13a). At the surface, the bias in experiment *Ens* and *Ens-SC* analyses is higher than in experiment *No DA* (Figure 13c,d), though not significantly at the 90%-confidence level. This bias is equal to the average error 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. Below the 20-m depth the bias for experiments *No DA*/*Bal* and experiments *Ens*/*Ens-SC* start to differ significantly with the bias in the latter lying closer to zero below 30 m depth. As a result, RMSE below 30 m is significantly smaller for the En4DVar experiments than in experiment *No DA* and experiment *Bal*. These points hold true for both the analyses and the forecasts.

Near the surface, experiments *Ens* and *Ens-SC* yield NH10 salinity analyses and forecasts with a smaller RMSE and a smaller bias magnitude than the other experiments (Figure 14), esp. when compared to the forecasts and analyses of experiment *Bal*, which are too fresh over the entire depth range for which NH10 salinity measurements are available (Figure 14c,d). However, differences at each depth are not significant at the 90%-level.

We note that the case with the balanced **B** yields worse RMSEs than the *No DA* case both for T and S. Furthermore, the forecast RMSE is smaller than analysis RMSE above the 40 m depth. We can identify three reasons for the lacklustre performance of the standard 4DVAR against the verification shelf mooring data. First, the nearby assimilated satellite SST differ. The RMSE and the time-mean of the difference between satellite and buoy surface temperatures are marked by a "+" in Figure 13a,b and Figure 13c,d respectively. The average SST in *No DA* matches the average buoy observations (Figure 13c,d) and consequently the assimilation of satellite SST observations, which are on average too warm, results in a deterioration of the model performance at the buoy location. Second, in case *Bal* surface velocity corrections must be fully balanced by corrections in the density. When SST is not constrained by the satellite observations, experiment *Bal* can overcorrect the temperature fitting the velocity observations. This results in a higher temperature analysis RMSE in experiment *Bal* than in the other experiments. After the erroneous T correction the forecast RMSE relaxes to be closer to the *No DA* RMSE. Third, in the static **B** errors 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 stream function are defined. For the latter, for the lack of a better guess, zero variance has been assumed. 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* 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 subsurface 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 specific balance operator formulation chosen.

The observed differences in RMSEs between the experiments can partially be explained from the observed biases (13c,d), which are discussed in more detail next. At the surface the bias in *Ens* and *Ens-SC* analyses matches the bias in the SST observations (marked by the '+' in 13c,d). This results in an overall increase in the time-averaged water temperature in the top 70 m during experiments *Ens* and *Ens-SC* compared to experiment *No DA*. As all experiments overestimate the vertical temperature gradient at the buoy location, this overall increase results in experiment *No DA* having a smaller absolute bias than experiments *Ens* and *Ens-SC* in the top 30 m, but a larger absolute bias below it. Consequently, in experiment *No DA* the RMSE is smaller than in experiments *Ens* and *Ens-SC* between 10-30 m depth, but significantly larger below 30 m. The temperature bias in the forecasts and analyses below 30 m in case *Bal* is not significantly different from *No DA*. At the surface the temperature bias in case *Bal* is larger than the other experiments (though not significant at the 90%-confidence level), which cannot be explained by the fact that SST observations are biased (since the same observations are assimilated in all the cases). We find that in case *Bal* the correction to the surface T is stronger than in the other cases as a result of the assimilation of the surface currents, particularly in windows where SST was unavailable. This happens because of stronger coupling between surface velocity and temperature errors in the balanced $\mathbf{B}$, compared to the ensemble $\mathbf{B}$ (see section 3). Depending on the shape, location and magnitude of the velocity errors, the with the velocity corrections associated temperature corrections can result in the appearance of a local bias in the SST field.

While experiments *Ens* and *Ens-SC* perform close to or better than the other experiment at the NH10 location, they do not show good performance against a limited set of Argo float locations (Argo, 2000) farther away from the coast. The locations used in this analysis are shown in Figure 11. Figure 15a and c show the average T and S profiles prior to 20 July 2011, when the profile locations were outside the river plume (SSS$>$ 31.5), and Figure 15b 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-SC* fail to reproduce the strength of the vertical temperature gradient in the thermocline (Figure 15a,b). The deterioration of these forecasts at depths 30-100 m may result from random non-zero correlations between SST and other subsurface fields present in $\mathbf{B}$ as no localisation is applied in the vertical direction. The latter was not applied since the ensemble error variance generally strongly decreases with depth. Also worrisome is the fact that *Ens-SC* and *Ens*, contrary to cases *Bal* and *No DA*, produce salinity forecasts that are too fresh above 50 m (Figure 15c,d). Without salinity constraints, the forecasts in experiment *Ens* produce a river plume with salinity that is 3 units too low at the surface (Figure 15d). Applying salinity constraints is highly beneficial in this case. It reduces the negative bias to about 1. 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 (given the negative bias in the SST data in our particular case). This in agreement with our findings in section 4.

6. Summary and Discussion

In this study the En4DVar DA system described in Pasmans and Kurapov (2019) was tested with the Oregon-Washington coastal ocean circulation model. An ensemble of 4DVAR runs is used to estimate the background error covariance **B**. The results of the En4DVar system are verified against observations that have yet to be assimilated or that will not be assimilated at all 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 forecasts for the fields associated with the assimilated observations (SST, surface velocity, SSH). Time-averaged forecast RMSE in *Ens-SC*, compared to surface data, is $\sim$3-5% smaller, but not significantly different from experiment *Bal* at the 90%-confidence level.

The story is different for fields for which no observations are assimilated like salinity or the subsurface fields. Comparison with glider observations showed that using the ensemble **B** yields corrections that can reproduce the observed spatial structure and temporal variability in the water properties on the T-S plane. 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: it fails to reproduce the observed T-S relationship in the earlier half of 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 $\mathbf{B}$ 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. Similarly, comparison against the mooring T and S profile data not only showed the advantageous effect of using the ensemble error covariance on the subsurface results, but also pointed to yet another problem with the balance operator where improvements must be made: the correction to the depth-integrated transport must be allowed on the shelf. The problems revealed here will encourage future research to re-evaluate the approach used to build the balance operator appropriate for the shelf flows. We must note that although the En4DVar system performs better on many of the skill assessment metrics used, the differences with the 4DVAR using the static $\mathbf{B}$ are small: when each metric is considered on its own, the 90%-confidence intervals of the RMSE in the cases using En4DVar and 4DVAR with the static $\mathbf{B}$ do overlap.

The salinity constraints were 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 balanced $\mathbf{B}$. The salinity constraint was found to suppress, but not completely eliminate the erroneous SSS corrections. Several measures could possibly improve this. First, the weights $\sigma_{S,i}^{-2}$ in (3) 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. Third, currently only the wind velocities are perturbed in the ensemble. The resulting perturbations in the transport and vertical mixing give rise to the large temperature-salinity covariances, particularly in the frontal regions. Adding perturbations to the atmospheric temperature and radiative flux in the ensemble members is expected to increase the SST ensemble variance whilst only weakly impacting the SSS variance. This would reduce the SST-SSS ensemble correlation and yield smaller SSS corrections. We also note that since the bulk flux formulation is utilised for the atmosphere-ocean fluxes, perturbations to the wind velocity already result in perturbations to sensible heat flux and latent heat flux. We expect that the problem of amplification of large and erroneous DA corrections to non-assimilated fields due to the presence of large 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). Furthermore, erroneous observations are not the only possible source of unphysically large corrections to tracer fields. If observational coverage is limited to specific areas, corrections to the tracer field can be local to these areas. Such local corrections would not conserve the total tracer volume in the model. Model domain-wide constraints, as the one proposed in this paper, would ideally help preserve the tracer in the volume-averaged sense.

The benefits and drawbacks of En4DVar listed above have to be weighted against the extra wall-time and computational resources required by the En4DVar compared to the 4DVAR with the balanced **B** . The En4DVar system used in this study requires 16,000 core-hours per 3-day window on an Intel Xeon E5-2680v3 hyper-threaded 2.5 GHz processor (Towns et al., 2014), whilst the balanced **B** 4DVAR only needs 100 core-hours per window. The major part of this 10-fold increase comes from the need to run an ensemble of nonlinear model runs in the En4DVar system. Fortunately, En4DVar allows for significant parallelisation and the required wall-clock time for both systems is 12 hours per window.

In this study the DA systems compared used either a purely static or a purely ensemble **B**. In meteorology it was found that 4DVAR systems in which a combination of the static and ensemble covariances are used (so-called hybrid systems) outperform the purely ensemble or 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. Performance of a DA system using the static **B** will also depend on the formulation details. In the ensemble **B** large 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 **B** (Derber and Bouttier, 1999) can help to produce a more realistic static **B**.

In conclusion, here we have presented the first initial results using Ens4DVAR in a realistic coastal ocean circulation setting. The research on En4DVar must be continued. Although it may seem the En4DVar did not show sizeable improvement over the "traditional" 4DVAR in every aspect, it has helped us to resolve or at least identify several shortcomings of the traditional approach and point to future research directions. Some of these problems could be remedied by a more extensive sampling of the coastal ocean to better constrain the forecasts. E.g., when satellite salinity products become more mature and more suitable for assimilation, there will be less need in adding a constraint on the SSS change based on the forecast model. When wide-swath altimetry (Rodriguez et al., 2017) becomes routinely available, it will provide a much better constraint on the location of the upwelling fronts and eddies than just a few nadir altimeters available for this study. Using data from several satellites for SST will improve coverage within an assimilation window. While research toward the best representation of the background model errors in **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.

7. Acknowledgements

This work was supported by the National Oceanic and Atmospheric Administration (NOAA) Coastal Ocean Modeling Testbed (COMT) [grant number NA13NOS0120139], the NOAA Quantitative Observing System Assessment Program (QOSAP), National Science Foundation (NSF) [grant numbers OCE-0527168 and OCE-0961999], Integrated Ocean Observing System / Northwest Association of Networked Ocean Observing Systems (IOOS/NANOOS) [grant number NA16NOS0120019], the National Aeronautics and Space Administration (NASA) SWOT Science Definition Team project [grant number NNX13AD89G]. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) [allocation TG-OCE160001], which is supported by NSF grant number ACI-1548562. Authors are grateful to Andy Harris for helping with finding the latest GOES SST data. 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. Furthermore, we would like to thank the four anonymous reviewers of this paper for their input.

Appendix A. Bootstrap confidence intervals

Let H be an error statistic defined as

$$H = \frac{1}{|I|} \sum_{i \in I} \epsilon_i^n \tag{A.1}$$

with ϵ_i the error for observation i, I a set of observational indices, $|I|$ the number of indices in I and $n = 1$ if H is the bias and $n = 2$ if H is the RMSE. Define

$$H^{(m)} = \frac{1}{|I|} \sum_{i \in I_m} \epsilon_i^n \tag{A.2}$$

with I_m a set of $|I|$ indices randomly drawn from I, possibly with duplicates. To estimate the 90%-confidence interval, $H^{(m)}$ is calculated for

$m = 1, 2, \ldots, 200$ and the lower and upper limits of the confidence interval are then defined as the 5% and 95%-percentiles of the set $H^{(1)}, H^{(2)}, \ldots, H^{(100)}$ (percentile bootstrap, see Efron (1982, p. 78–80)).

To account for the correlations between the errors not every index in I_m is drawn separately. Instead, for the surface observations, we randomly draw a horizontal position within the model and all observations that are within 50 km of this location and that are within the same window are added to I_m. This process, called moving block bootstrap (Kunsch, 1989), is repeated until $|I_m| = |I|$. Similarly, for the time series we pick a time and all observations that are at the same depth and that lie within a 3 day window around of the selected time are added. This is then repeated until $|I_m| = |I|$.

Appendix B. Balance Operator Covariance

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

$$\begin{aligned}&\langle \epsilon_T(\mathbf{r}_1)\epsilon_T(\mathbf{r}_2)\rangle = \\ &\sigma^2_{bal,T}\exp\left(\tfrac{z}{D_{bal}}\right)\exp\left(-\tfrac{1}{2}\tfrac{(x_1-x_2)^2}{R^2} - \tfrac{1}{2}\tfrac{(y_1-y_2)^2}{R^2} - \tfrac{1}{2}\tfrac{(s_1-s_2)^2}{D_s^2}\right)\end{aligned} \tag{B.1}$$

with ϵ_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

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

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, $\rho_0 = 1025\,\mathrm{kg\,m^{-3}}$ the reference density, $\alpha_T = 1.7\times10^{-4}\,{}^\circ\mathrm{C}^{-1}$, $\beta_S = 7.5\times10^{-4}$. $\alpha = -0.16$ which is based on a linear least-square fit to all in-situ glider temperature and salinity observations made by the OSU glider group on the Oregon shelf between 2006 and 2013. These observations are available at COAS (2017). The vertical length scale of the temperature-temperature covariance $D_{bal} = 100\,\mathrm{m}$, D_s is the vertical scale in s-coordinates and is chosen such that it is 50 m in 3091 m deep water. $R = 25\,\mathrm{km}$, 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 $\sigma_{bal,T} = 0.9\,{}^\circ\mathrm{C}$. 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 (NOAA, 2011b) and predictions from experiment *No DA* over the period 19 April 2011 to 1 October 2011 and then taking the median of these standard deviations.

Appendix C. The ensemble for the balance operator

The cluster search method (Pasmans and Kurapov, 2019) 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-SC* this approximation is constructed using $\hat{\mathbf{A}} \approx \frac{1}{M-1}\mathbf{R}^{-1/2}\mathbf{D}(\mathbf{R}^{-1/2}\mathbf{D})^T$. Here, the columns of $\mathbf{D}$ are the innovation vectors for the control run and the different ensemble members and M is the number of ensemble members (including control run). 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* a different approach is used. The singular value decomposition (SVD) $\mathbf{U\Lambda V}^T$ 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 $\mathbf{\Lambda} \in \mathbb{R}^{(M-1)\times(M-1)}$ diagonal with D the number of observations and $M = 40$, the number of ensemble members including the control run used in experiments *Ens* and *Ens-SC*. The j-th column of $\mathbf{U}$, $\mathbf{u}_j \in \mathbb{R}^D$, and the (j,j)-th element of Λ^2, λ_j^2, then are the j-th eigenvector and eigenvalue respectively of

$$\mathbf{U\Lambda}^2\mathbf{U}^T = \hat{\mathbf{A}}\mathbf{SS}^T\hat{\mathbf{A}}^T \approx (M-1)\hat{\mathbf{A}}\mathbf{I}\hat{\mathbf{A}}^T = (M-1)\hat{\mathbf{A}}^2 \qquad \text{(C.1)}$$

Consequently, the required low-rank estimate for $\mathbf{A}^{1/2}$ is constructed as $(M-1)^{-1/4}\mathbf{U\Lambda}^{1/2}$.

References

Argo, 2000. Argo float data and metadata from Global Data Assembly Centre (Argo GDAC), April 2011 to September 2011, doi:10.17882/42182. SEANOE, Monterey, CA.

Austin, J.A., Barth, J.A., 2002. Variation in the position of the upwelling front on the Oregon shelf. J. Geophys. Res. Oceans 107, 1–15. doi:10.1029/2001JC000858.

Balmaseda, M.A., Mogensen, K., Weaver, A.T., 2013. Evaluation of the ECMWF ocean reanalysis system ORAS4. Q. J. R. Meteor. Soc. 139, 1132–1161. doi:10.1002/qj.2063.

Balmaseda, M.A., Vidard, A., Anderson, D.L., 2008. The ECMWF ocean analysis system: ORA-S3. Mon. Wea. Rev. 136, 3018–3034. doi:10.1175/2008MWR2433.1.

Barth, J.A., Pierce, S.D., Castelao, R.M., 2005a. Time-dependent, wind-driven flow over a shallow midshelf submarine bank. J. Geophys. Res. Oceans 110. doi:10.1029/2004JC002761.

Barth, J.A., Pierce, S.D., Cowles, T.J., 2005b. Mesoscale structure and its seasonal evolution in the northern California Current System. Deep Sea Res. Part 2: Top. Stud. Oceanogr. 52, 5–28. doi:10.1016/j.dsr2.2004.09.026.

Berdeal, I.G., Hickey, B.M., Kawase, M., 2002. Influence of wind stress and ambient flow on a high discharge river plume. J. Geophys. Res. Ocean 107, 3130. doi:10.1029/2001JC000932.

Chelton, D.B., De Szoeke, R.A., Schlax, M.G., El Naggar, K., Siwertz, N., 1998. Geographical variability of the first baroclinic Rossby radius of deformation. J. Phys. Oceanogr. 28, 433–460. doi:10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2.

Ciavatta, S., Torres, R., Martinez-Vicente, V., Smyth, T., DallOlmo, G., Polimene, L., Allen, J.I., 2014. Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling. Prog. Oceanogr. 127, 74–95. doi:10.1016/j.pocean.2014.06.002.

Ciavatta, S., Torres, R., SauxPicart, S., Allen, J.I., 2011. Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? J. Geophys. Res. Oceans 116, C12043. doi:10.1029/2011JC007219.

Clayton, A.M., Lorenc, A.C., Barker, D.M., 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. Q. J. R. Meteor. Soc. 139, 1445–1461. doi:10.1002/qj.2054.

COAPS, 2015. HYCOM + NCODA global 1/12° analysis, Version GLBa0.08, january 2011 to december 2011. Center for Ocean-Atmospheric Prediction Studies, https://hycom.org/data/glba0pt08.

COAS, 2017. Undersea gliders: Valuable new tool in marine research. Available from: https://gliderfs2.coas.oregonstate.edu/gliderweb/docs/COAS50.pdf (Accessed 15 November 2018).

Cook, M.S., Paduan, J.D., 2001. Processing HF radar data using the HFRadarmap software system, in: Paduan, J. (Ed.), Radiowave Oceanography: The First International Workshop, Miami University, Miami, FL. pp. 12–16.

Courtier, P., Thépaut, J.N., Hollingsworth, A., 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. Q. J. R. Meteor. Soc. 120, 1367–1387. doi:10.1002/qj.49712051912.

Dee, D.P., Uppala, S., 2009. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. Q. J. R. Meteorol. Soc. 135, 1830–1841. doi:10.1002/qj.493.

Derber, J., Bouttier, F., 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. Tellus A 51, 195–221. doi:10.1034/j.1600-0870.1999.t01-2-00003.x.

Derber, J.C., Wu, W.S., 1998. The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system. Mon. Weather Rev. 126, 2287–2299. doi:10.1175/1520-0493(1998)126<2287:TUOTCC>2.0.CO;2.

Donlon, C.J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., Wimmer, W., 2012. The operational sea surface temperature and sea ice analysis (OSTIA) system. Remote Sens. Environ. 116, 140–158. doi:10.1016/j.rse.2010.10.017.

Efron, B., 1982. The jackknife, the bootstrap, and other resampling plans. volume 38 of *CBMS-NSF conference series in applied mathematics*. SIAM.

Egbert, G.D., Bennett, A.F., Foreman, M.G.G., 1994. TOPEX/POSEIDON tides estimated using a global inverse model. J. Geophys. Res. 99, 24821–24852. doi:10.1029/94JC01894.

Egbert, G.D., Erofeeva, S.Y., 2010. TPXO atlas, Version 7.2. OSU TOPEX/Poseidon Global Inverse Solution. `http://volkov.oce.orst.edu/tides/global.html`.

Erofeev, A., 2011. Gridded Slocum glider data (1km $\times$ 4m), March 2011 to November 2011. Gliderweb. https://gliderfs2.coas.oregonstate.edu/gliderweb/archive/gridded/2011/.

Fairall, C.W., Bradley, E.F., Hare, J.E., Grachev, A.A., Edson, J.B., 2003. Bulk parameterization of airsea fluxes: Updates and verification for the COARE algorithm. J. Climate 16, 571–591. doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

Fong, D.A., Geyer, W.R., 2001. Response of a river plume during an upwelling favorable wind event. J. Geophys. Res. Oceans 106, 1067–1084. doi:10.1029/2000JC900134.

Fu, L.L., Fukumori, I., Miller, R.N., 1993. Fitting dynamic models to the Geosat sea level observations in the tropical Pacific ocean. Part II: A linear, wind-driven model. J. Phys. Oceanogr. 23, 2162–2181. doi:10.1175/1520-0485(1993)023<2162:FDMTTG>2.0.CO;2.

Gan, J., Li, L., Wang, D., Guo, X., 2009. Interaction of a river plume with coastal upwelling in the northeastern South China Sea. Cont. Shelf Res. 29, 728–740. doi:10.1016/j.csr.2008.12.002.

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, IEEE, Piscataway, NJ. pp. 23–27.

Gürol, S., Weaver, A.T., Moore, A.M., Piacentini, A., Arango, H.G., Gratton, S., 2014. B-preconditioned minimization algorithms for variational data assimilation with the dual formulation. Q.J.R. Meteorol. Soc. 140, 539–556. doi:10.1002/qj.2150.

Hetland, R.D., 2005. Relating river plume structure to vertical mixing. J. Phys. Oceanogr. 35, 1667–1688. doi:10.1175/JPO2774.1.

Hickey, B.M., Geier, S., Kachel, N., MacFadyen, A., 2005. A bi-directional river plume: The Columbia in summer. Cont. Shelf Res. 25, 1631–1656. doi:10.1016/j.csr.2005.04.010.

Huyer, A., 1977. Seasonal variation in temperature, salinity, and density over the continental shelf off Oregon. Limnol. Oceanogr. 22, 442–453. doi:10.4319/lo.1977.22.3.0442.

Huyer, A., Fleischbein, J.H., Keister, J., Kosro, P.M., Perlin, N., Smith, R.L., Wheeler, P.A., 2005. Two coastal upwelling domains in the northern California Current system. J. Mar. Res. 63, 901–929. doi:10.1357/002224005774464238.

IAPSO, 1985. The international system of units (SI) in oceanography, in: Unesco technical papers in marine science, United Nations Educational, Scientific and Cultural Organization. p. 124.

Koch, A.O., Kurapov, A.L., Allen, J.S., 2010. Nearsurface dynamics of a separated jet in the coastal transition zone off Oregon. J. Geophys. Res. 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. J. Geophys. Res. 110, C10S06. doi:10.1029/2004JC002769.

Kuhl, D.D., Rosmond, T.E., Bishop, C.H., McLay, J., Baker, N.L., 2013. Comparison of hybrid ensemble/4DVar and 4DVar within the NAVDAS-AR data assimilation framework. Mon. Weather Rev. 141, 2740–2758. doi:10.1175/MWR-D-12-00182.1.

Kunsch, H.R., 1989. The jackknife and the bootstrap for general stationary observations. Ann. Statist. 17, 1217–1241. doi:10.1214/aos/1176347265.

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

Kurapov, A.L., Egbert, G.D., Allen, J.S., Miller, R.N., 2009. Representer-based analyses in the coastal upwelling system. Dyn. Atmos. Oceans 48, 198–218. doi:10.1016/j.dynatmoce.2008.09.002.

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

Lea, D.J., Drecourt, J.P., Haines, K., Martin, M.J., 2008. Ocean altimeter assimilation with observational- and model-bias correction. Q. J. R. Meteorol. Soc. 134, 1761–1774. doi:10.1002/qj.320.

Lellouche, J.M., Galloudec, O.L., Drévillon, M., Régnier, C., Greiner, E., Garric, G., Ferry, N., Desportes, C., Testut, C.E., Bricaud, C., Bourdallé-Badie, R., Tranchant, B., Benkiran, M., Drillet, Y., Daudin, A., De Nicola, C., 2013. Evaluation of global monitoring and forecasting systems at Mercator Océan. Ocean Sci. 9, 57. doi:10.5194/os-9-57-2013.

Liu, Y., MacCready, P., Hickey, B.M., 2009. Columbia River plume patterns in summer 2004 as revealed by a hindcast coastal ocean circulation model. Geophys. Res. Lett. 36, L02601. doi:10.1029/2008GL036447.

Lorenc, A.C., Jardak, M., 2018. A comparison of hybrid variational data assimilation methods for global NWP. Q. J. R. Meteorol. Soc. , 2748–2760doi:10.1002/qj.3401.

MacCready, P., Banas, N.S., Hickey, B.M., Dever, E.P., Liu, Y., 2009. A model study of tide- and wind-induced mixing in the Columbia River Estuary and plume. Cont. Shelf Res. 29, 278–291. doi:10.1016/j.csr.2008.03.015.

Mazzini, P.L., Risien, C.M., Barth, J.A., Pierce, S.D., Erofeev, A., Dever, E.P., Kosro, P.M., Levine, M.D., Shearman, R.K., Vardaro, M.F., 2015. Anomalous near-surface low-salinity pulses off the central Oregon coast. Sci. Rep. 5, 17145.

McWilliams, J.C., Colas, F., Molemaker, M.J., 2009. Cold filamentary intensification and oceanic surface convergence lines. Geophys. Res. Lett. 36. doi:10.1029/2009GL039402.

NOAA, 2011a. North American Mesoscale Forecast System (NAM) 12 km analysis, january 2011 to december 2011. National Centers for Environmental Information. `https://www.ncei.noaa.gov/thredds/catalog/namanl/catalog.html`.

NOAA, 2011b. Pacific North buoys. National Data Buoy Center. http://www.ndbc.noaa.gov/.

Oke, P.R., Allen, J.S., Miller, R.N., Egbert, G.D., 2002a. A modeling study of the three-dimensional continental shelf circulation off Oregon. Part II: Dynamical analysis. J. Phys. Oceanogr. 32, 1383–1403. doi:10.1175/1520-0485(2002)032<1383:AMSOTT>2.0.CO;2.

Oke, P.R., Allen, J.S., Miller, R.N., Egbert, G.D., Austin, J.A., Barth, J.A., Boyd, T.J., Kosro, P.M., Levine, M.D., 2002b. A modeling study of the three-dimensional continental shelf circulation off Oregon. Part I: Model-data comparisons. J. Phys. Oceanogr. 32, 1360–1382. doi:10.1175/1520-0485(2002)032<1360:AMSOTT>2.0.CO;2.

Pasmans, I., Kurapov, A.L., 2017. A Monte Carlo background covariance localization method for an ensemble-variational assimilation system. Mon. Weather Rev. 145, 4543–4557. doi:10.1175/MWR-D-16-0424.1.

Pasmans, I., Kurapov, A.L., 2019. Ensemble 4DVAR (En4DVar) data assimilation in a coastal ocean circulation model. Part I: Methodology and ensemble statistics. Ocean. Model. 144, 101493. doi:10.1016/j.ocemod.2019.101493.

Pasmans, I., Kurapov, A.L., Barth, J., Ignatov, A., Kosro, P., Shearman, R., 2019. Why gliders appreciate good company: Glider assimilation in the Oregon-Washington coastal ocean 4DVAR system with and without surface observations. J. Geophys. Res. Oceans 124, 750–772. doi:10.1029/2018JC014230.

Pawlowicz, R., Beardsley, B., Lentz, S., 2002. Classical tidal harmonic analysis including error estimates in MATLAB using T_Tide. Computers & Geosciences 28, 929–937. doi:10.1016/S0098-3004(02)00013-4.

Price, J.F., Sundermeyer, M.A., 1999. Stratified Ekman layers. J. Geophys. Res. Oceans 104, 20467–20494. doi:10.1029/1999JC900164.

Rodriguez, E., Fernadez, D., Peral, E., Chen, C., Bleser, J.W., Williams, B., 2017. Satellite altimetry over oceans and land surfaces, in: Stammer, D., Cazenave, A. (Eds.), Wide-Swath Altimetry: A Review, CRC press. pp. 71–112. doi:10.1201/9781315151779-2.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atm. 106, 7183–7192. doi:10.1029/2000JD900719.

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G.D., Roskies, R., Scott, J.R., Wilkins-Diehr, N., 2014. XSEDE: Accelerating scientific discovery. Comp. Sci. & Eng. 16, 62–74. doi:10.1109/MCSE.2014.80.

USGS, 2011. USGS water data for USA , October 2008 to April 2015, `https://waterdata.usgs.gov/usa/nwis/`. U.S. Geological Survey, Reston, VA.

Weaver, A.T., Deltel, C., Machu, E., Ricci, S., Daget, N., 2005. A multivariate balance operator for variational ocean data assimilation. Q. J. R. Meteorol. Soc. 131, 3605–3625. doi:10.1256/qj.05.119.

Yu, P., Kurapov, A.L., Egbert, G.D., Allen, J.S., Kosro, P.M., 2012. Variational assimilation of HF radar surface currents in a coastal ocean model off Oregon. Ocean Model. 49-50, 86–104. doi:10.1016/j.ocemod.2012.03.001.

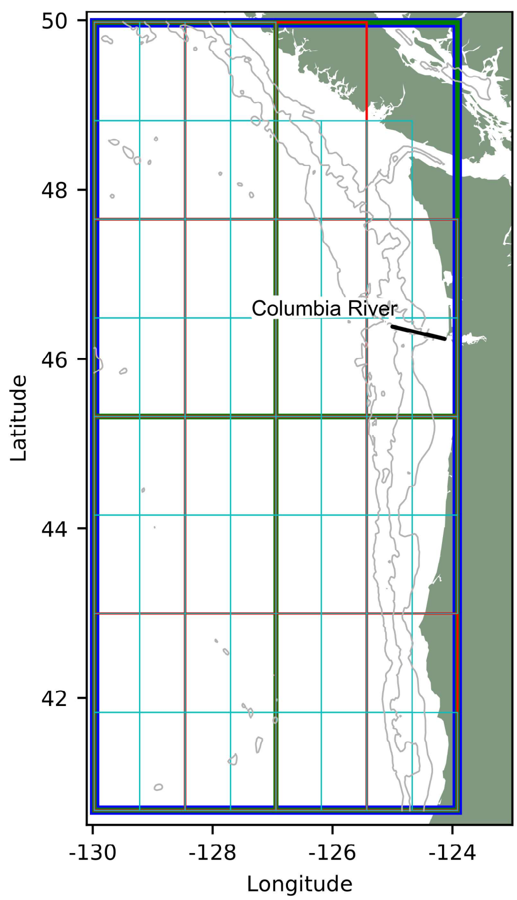

Figure 1: Overview of the model domain and the boxes used for the salinity constraints (SC). Boxes at the different levels of the hierarchy are shown with different colours.

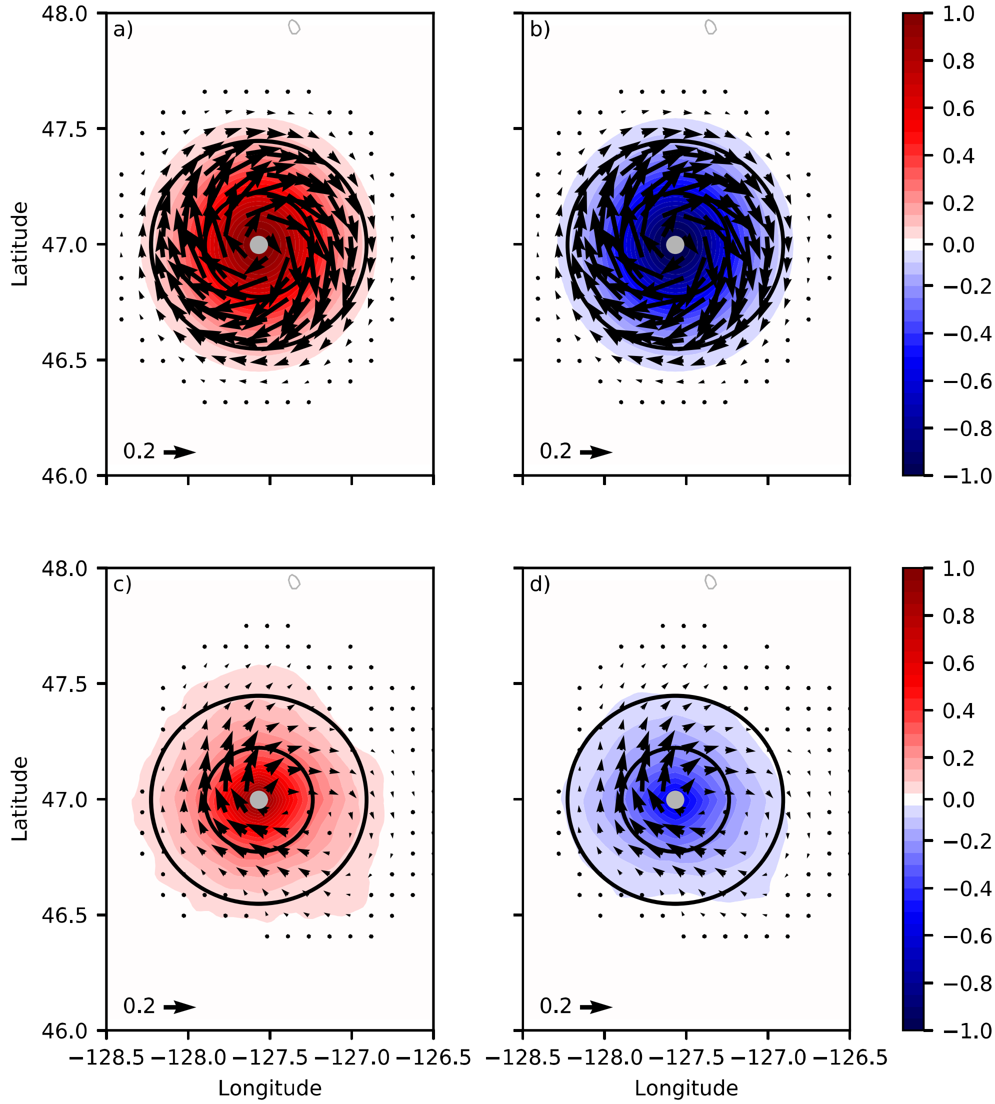

Figure 2: Surface maps of the background error correlations in the interior ocean, away from strong fronts. Shown are correlations of T at the grey dot location $\mathbf{r}_0$ and (left) surface T and (right) surface S. (Top) static balanced $\mathbf{B}$, (bottom) ensemble $\mathbf{B}$, where correlations are computed using ensemble member perturbations from all the windows. $T(\mathbf{r}_0)$-surface velocity correlations are shown as vectors in each panel. Black circles indicate points lying 25 km (inner circle) and 50 km (outer circle) from the centre, corresponding to respectively one and two times the standard deviation of the Gaussian shape function used for the horizontal balance operator correlation function.

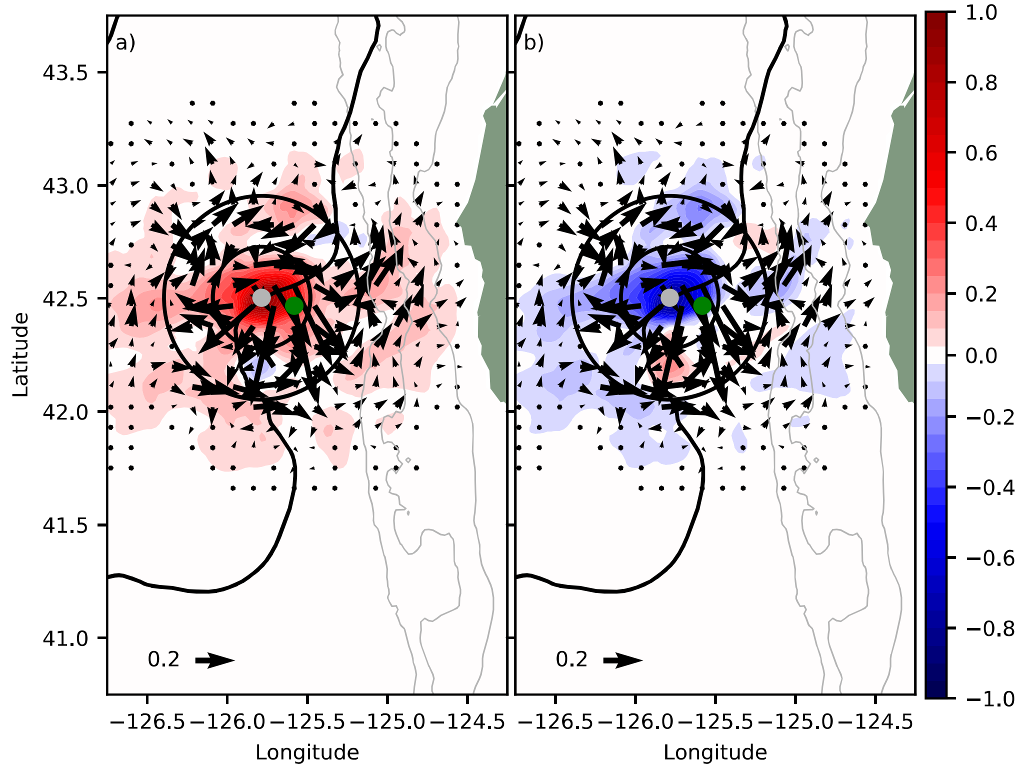

Figure 3: Surface maps of ensemble background error correlations for the reference point at the inshore edge of the Columbia River front (black line is $S = 31.5$). Shown are surface maps of the background error correlations of T at the grey dot location $\mathbf{r}_0$ and (a) surface T and (b) surface 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. Black circles indicate points lying 25 km (inner circle) and 50 km (outer circle) from the centre, corresponding to respectively one and two times the standard deviation of the Gaussian shape function used for the horizontal balance operator correlation function.

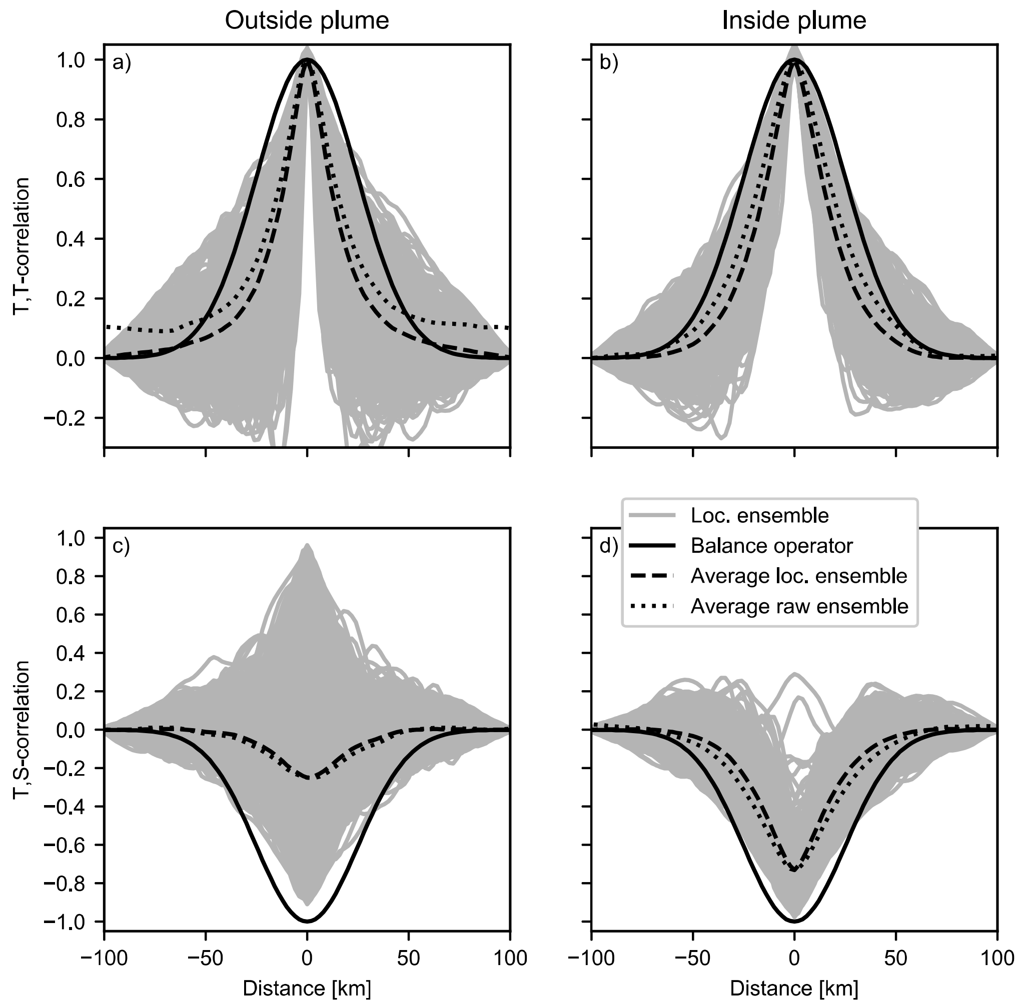

Figure 4: Surface (top) T-T and (bottom) T-S background error correlations from experiment *Ens* for randomly selected 200-km meridional sections (left) outside and (right) inside the Columbia River plume. Error ensemble correlations are computed for surface T at the centre point (distance= 0 km) and SST and SSS along the sections. Grey lines: localised ensemble correlation in each window; dashed: averaged correlations, dotted: average correlations for raw (not localised) ensembles; solid black: the balanced $\mathbf{B}$ correlations.

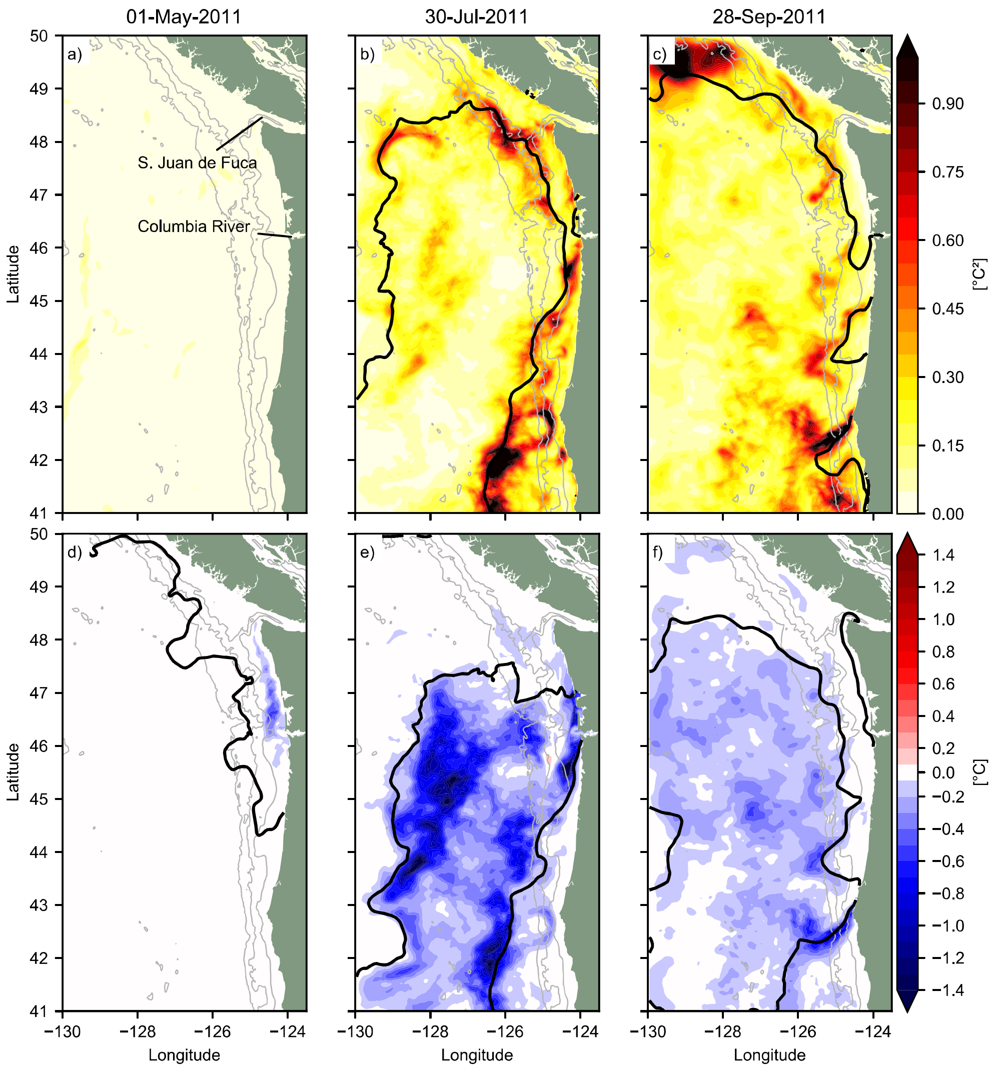

Figure 5: Point-by-point surface (a-c) temperature variance and (d-f) temperature-salinity ensemble covariance for 1 May (1st column), 30 July (2nd column), 28 September 2011 (3rd column). Solid black lines mark the (a-c) 15 °C isotherm and (d-f) 31.5 isohaline in the forecast solution. Note that salinity is a dimensionless quantity.

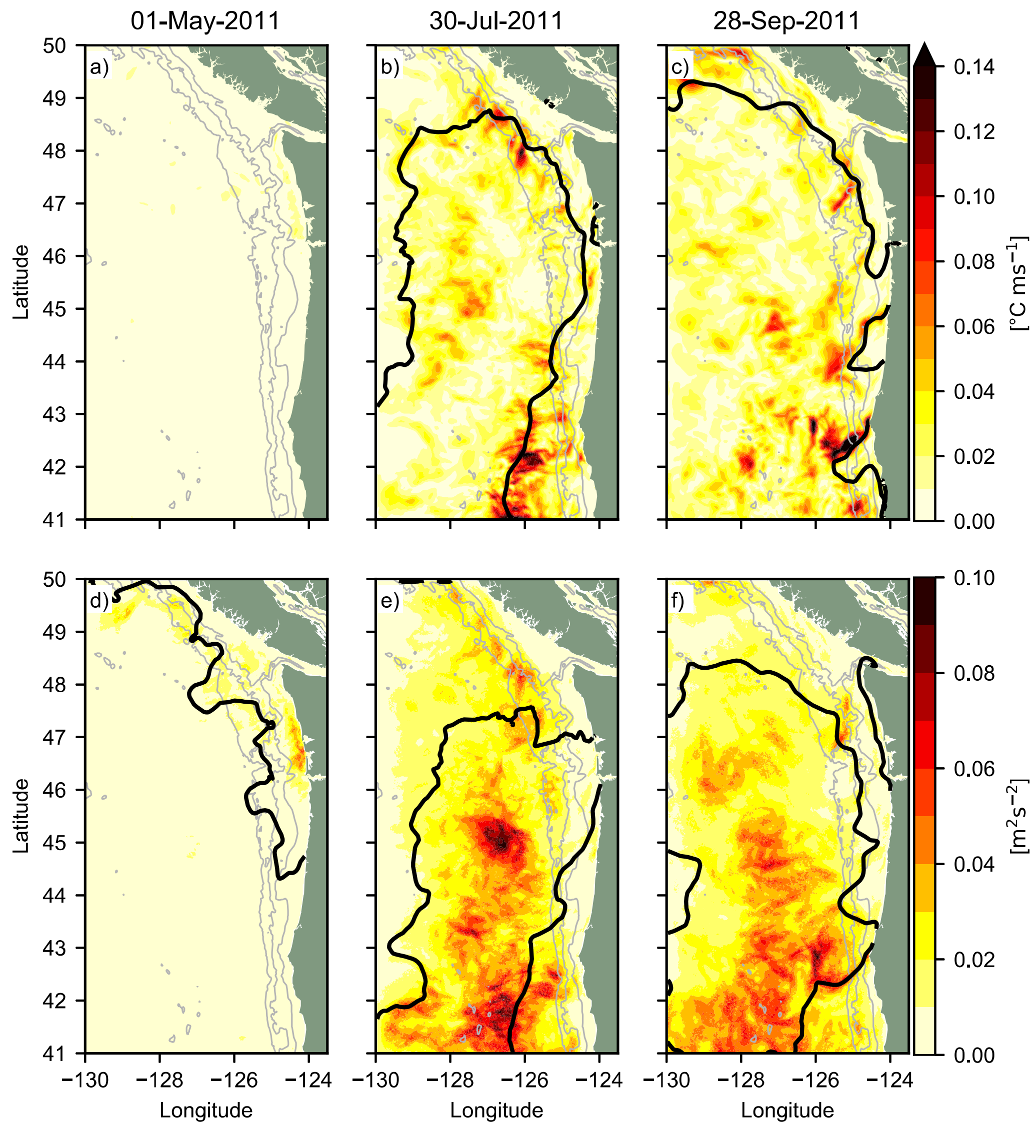

Figure 6: (a-c) Maximum point-by-point SST-surface velocity ensemble covariance for 1 May (1st column), 30 July (2nd column), 28 September 2011 (3rd column). (d-f) as (a-c) but now for the ensemble surface velocity variance. Solid black lines mark the (a-c) 15 °C isotherm and (d-f) 31.5 isohaline.

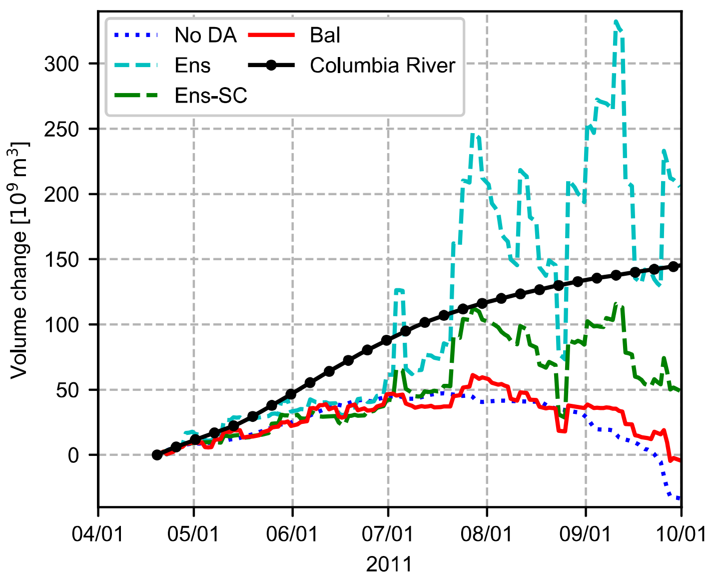

Figure 7: Change in fresh water volume in the river plume ($S < 31.5$) since 19 April 2011 in experiment *No DA* (dark blue) and the analyses from experiments *Ens* (light blue), *Ens-SC* (green), *Bal* (red). Also shown is the cumulative discharge of the Columbia River since 19 April 2011 (black).

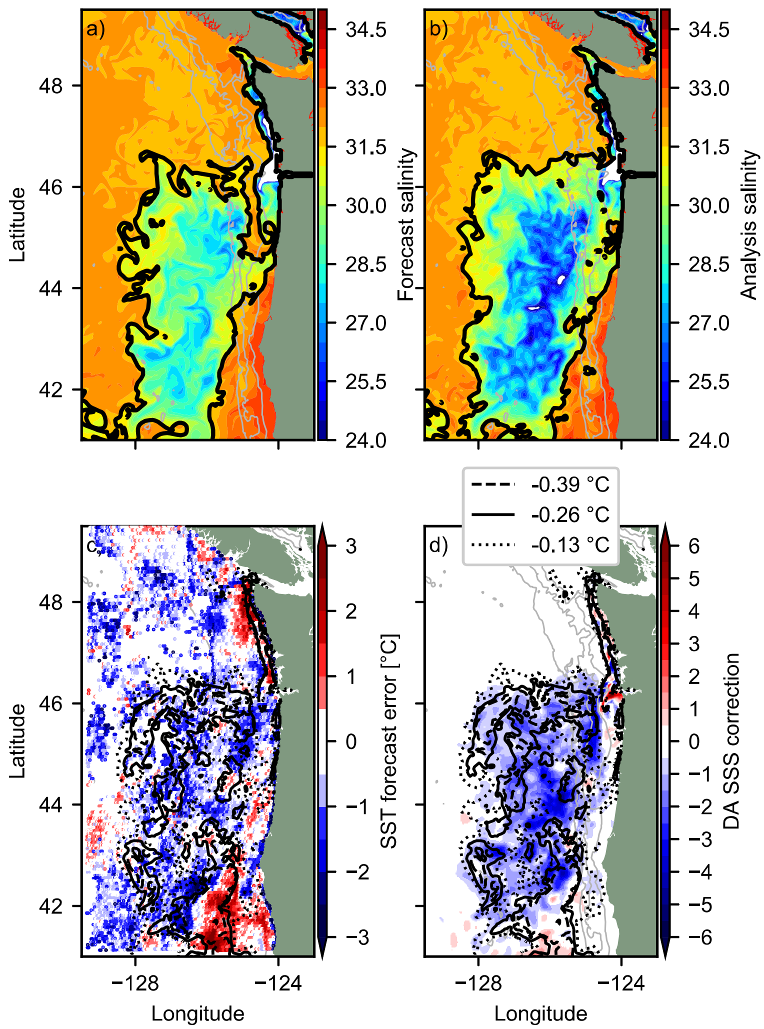

Figure 8: The study of DA-induced SSS changes, 21 July 2011: (a) forecast SSS in experiment *Ens*, prior to DA correction; the black line is $S = 31.5$; (b) analysis SSS, after DA; (c) difference in SST between the *Ens* forecast and observed SST; and (d) the DA correction to the SSS field. In (c) and (d), contours show the point-by-point SST-SSS covariance -0.39, -0.26 and -0.13 °C. Note that salinity is a dimensionless quantity.

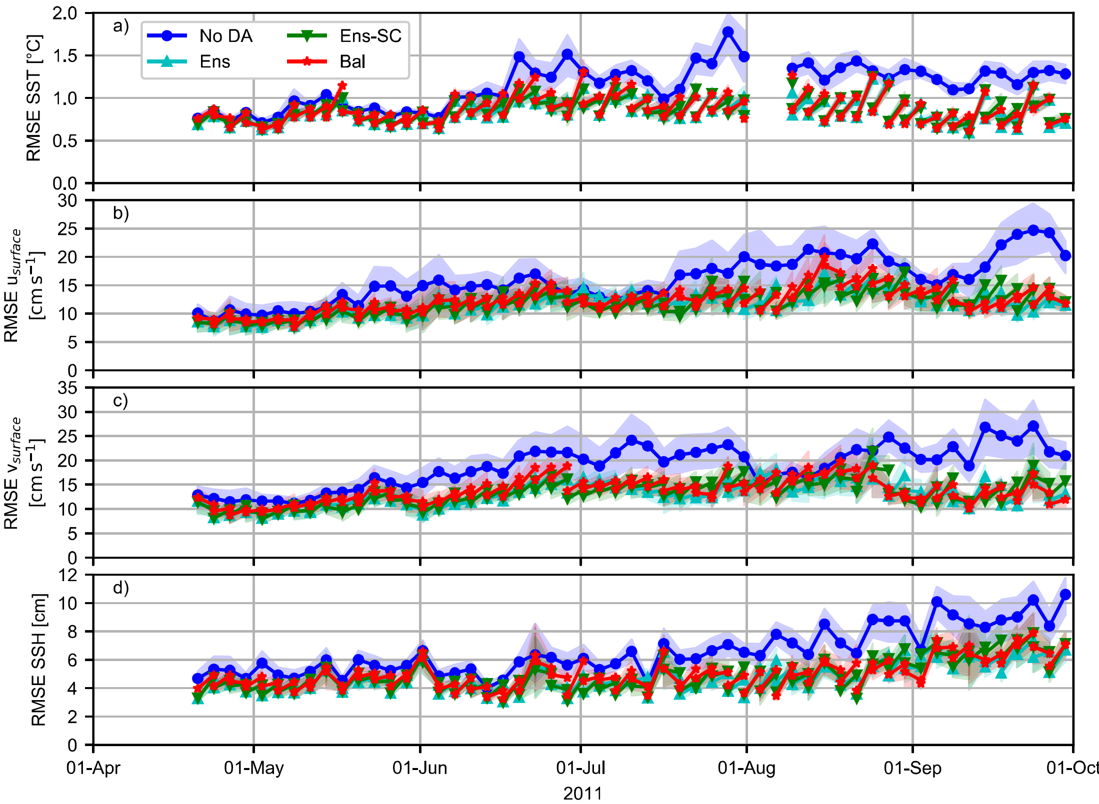

Figure 9: Time series of the area-averaged RMSE for (a) SST, (b) daily-averaged zonal velocity, (c) daily-averaged meridional velocity and (d) SSH observations with along-track mean removed. Results from experiment *No DA* (dark blue), experiment *Ens* (light blue), experiment *Ens-SC* (green) and *Bal* (red) are shown. Dots at the left (right) side of the DA line segments correspond to the analysis (forecast). The opaque area around the lines marks the 90%-confidence interval.

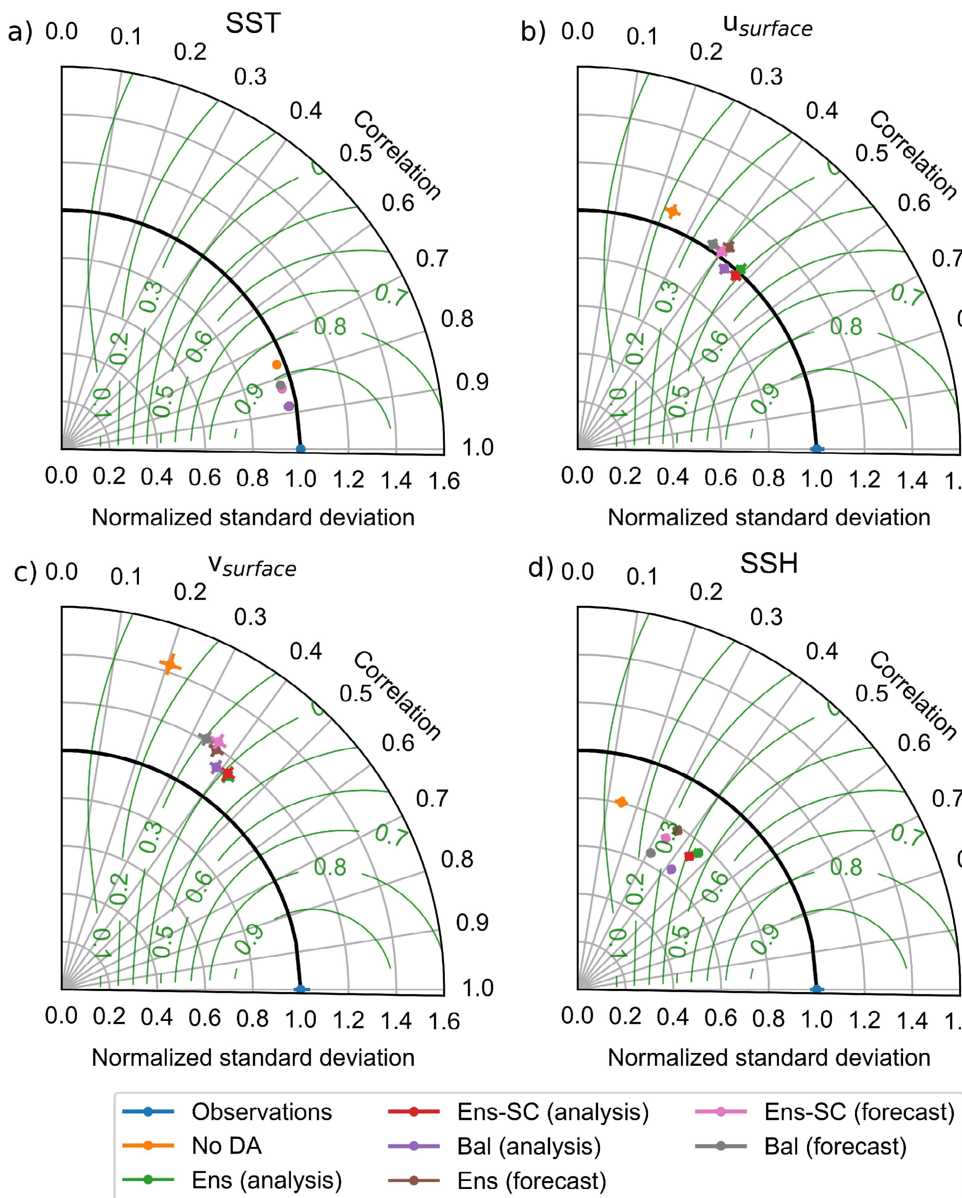

Figure 10: Taylor diagrams showing the standard deviation in the observations and the model predictions for the observations, versus the correlations between the observations and the model predictions in the different experiments. The standard deviations are normalised by the observation standard deviation. (a) SST, (b) zonal surface velocity, (c) meridional surface velocity, (d) SSH deviations from the track mean. Green lines mark contours of equal skill (5).

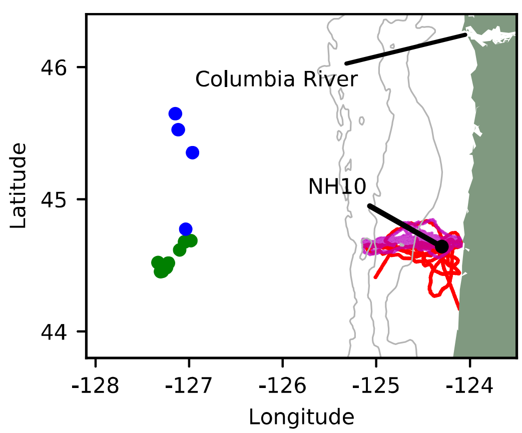

Figure 11: Location of the in-situ observations. Shown are glider positions prior to 30 June 2011 (red), after 30 June 2011 (purple), Argo floats outside the plume (green) and inside the river plume (blue) as well as the location of the NH10 buoy (black). The 200, 1000 and 2000 m isobath contours are displayed as grey lines.

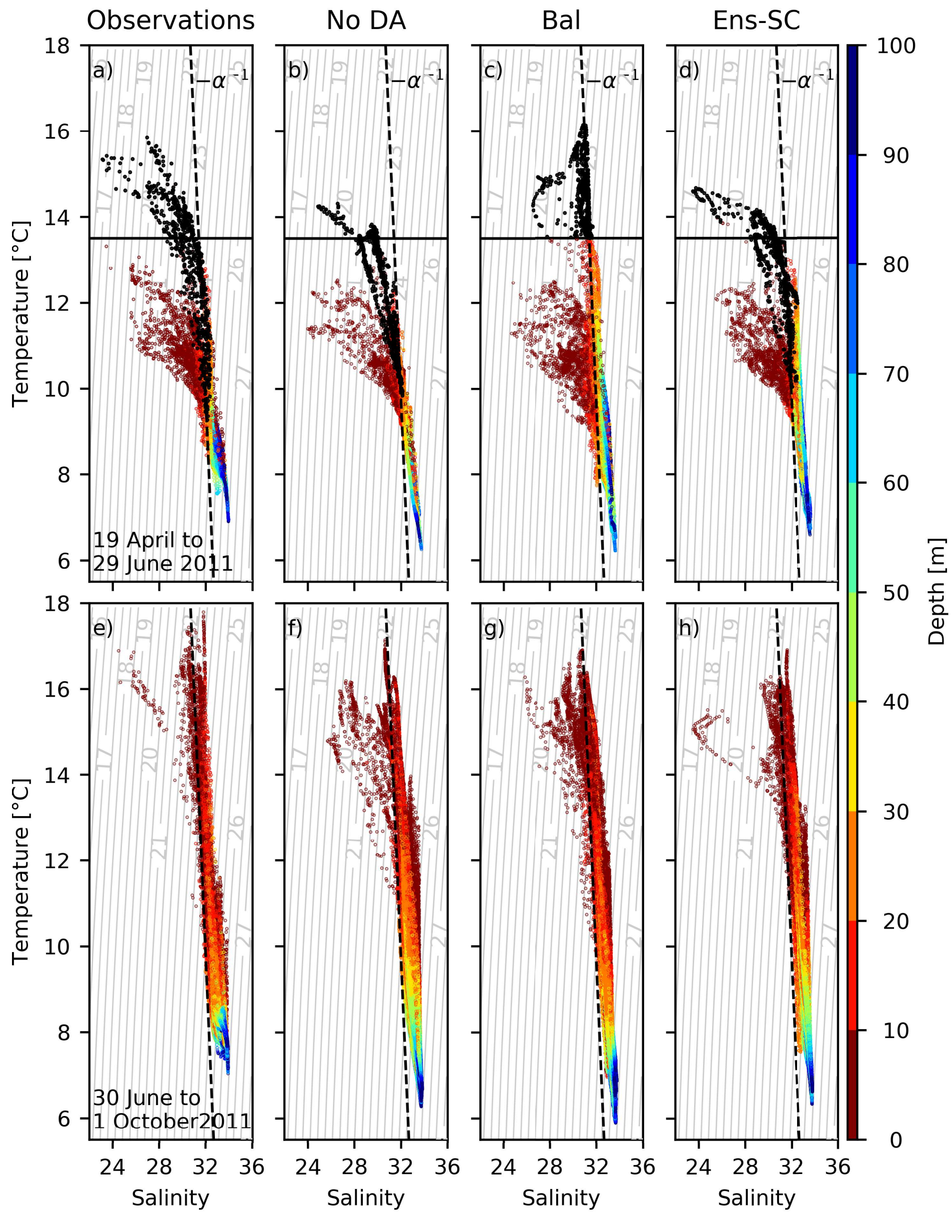

Figure 12: Glider-observed and model forecast T-S diagrams along the glider transect near 44.65°N. (Top) 19 April 00:00 to 30 June 00:00 2011, (bottom) 30 June 00:00 through 1 October 2011 00:00. (Left to right): observed, *No DA*, Bal, *Ens-SC*. Colours indicate the depth at which the observations are taken. Black solid line is $T = 13.5$°C and the black dashed line shows the slope in the relation $\delta T = -\alpha^{-1}\delta S$ used in the balance operator (B.2). Black dots indicate where points corresponding to $T > 13.5$°C in experiment *Bal* show up in all the experiments. Note that salinity is a dimensionless quantity.

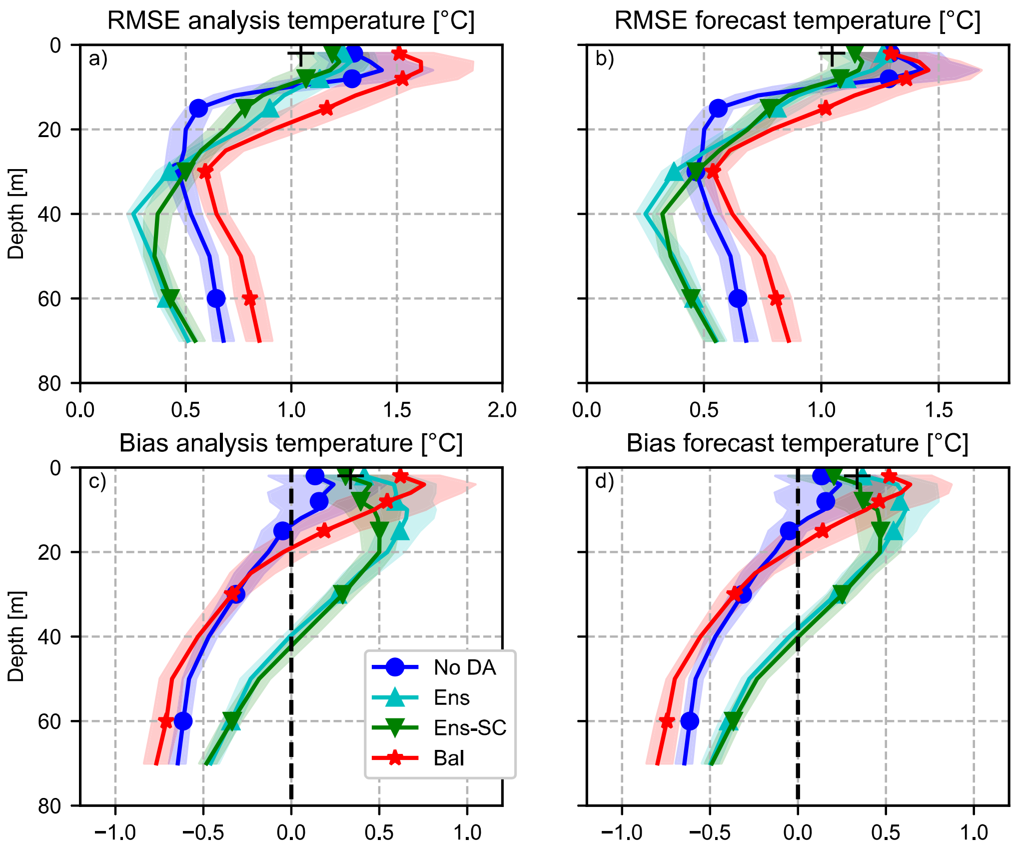

Figure 13: Vertical profiles of the time-averaged, 19 April 2011 to 1 October 2011 (Top) RMSE and (bottom) bias in NH10 temperature. (Left) model analyses, (right) forecasts. Experiments: *No DA* (dark blue), *Ens* (light blue), *Ens-SC* (green) and *Bal* (red) over the period 19 April 2011 till 1 October 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).

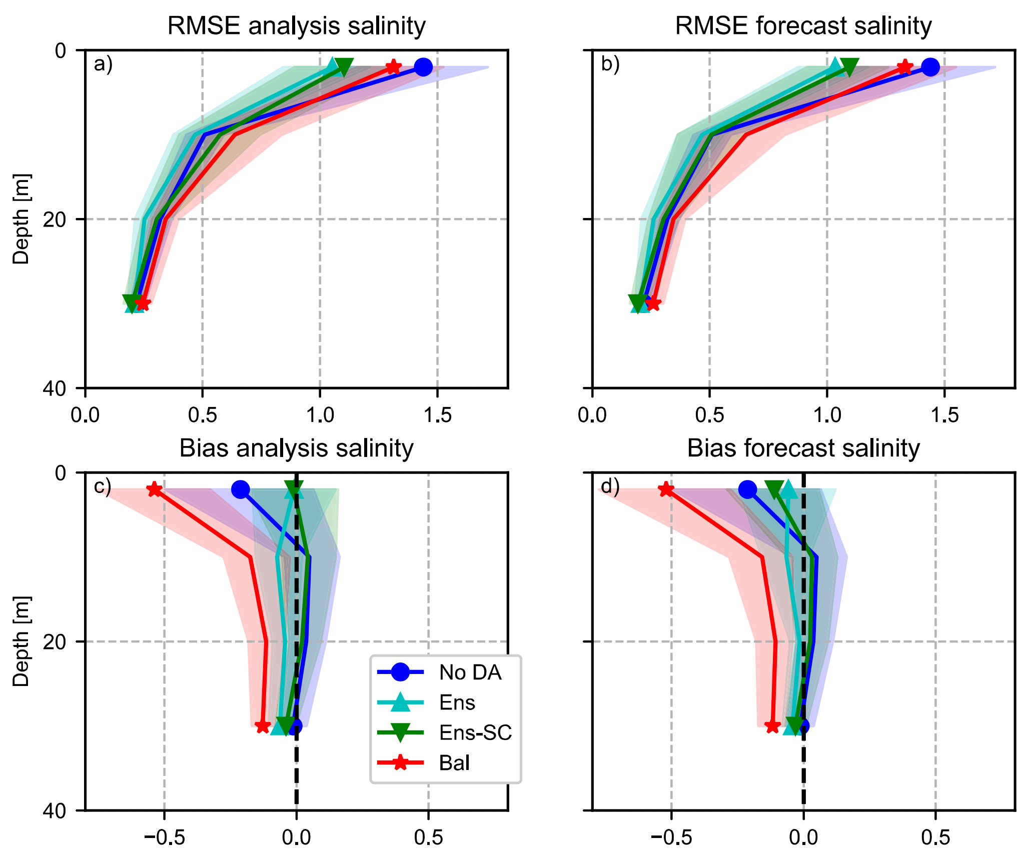

Figure 14: As Figure 13, but now for salinity.

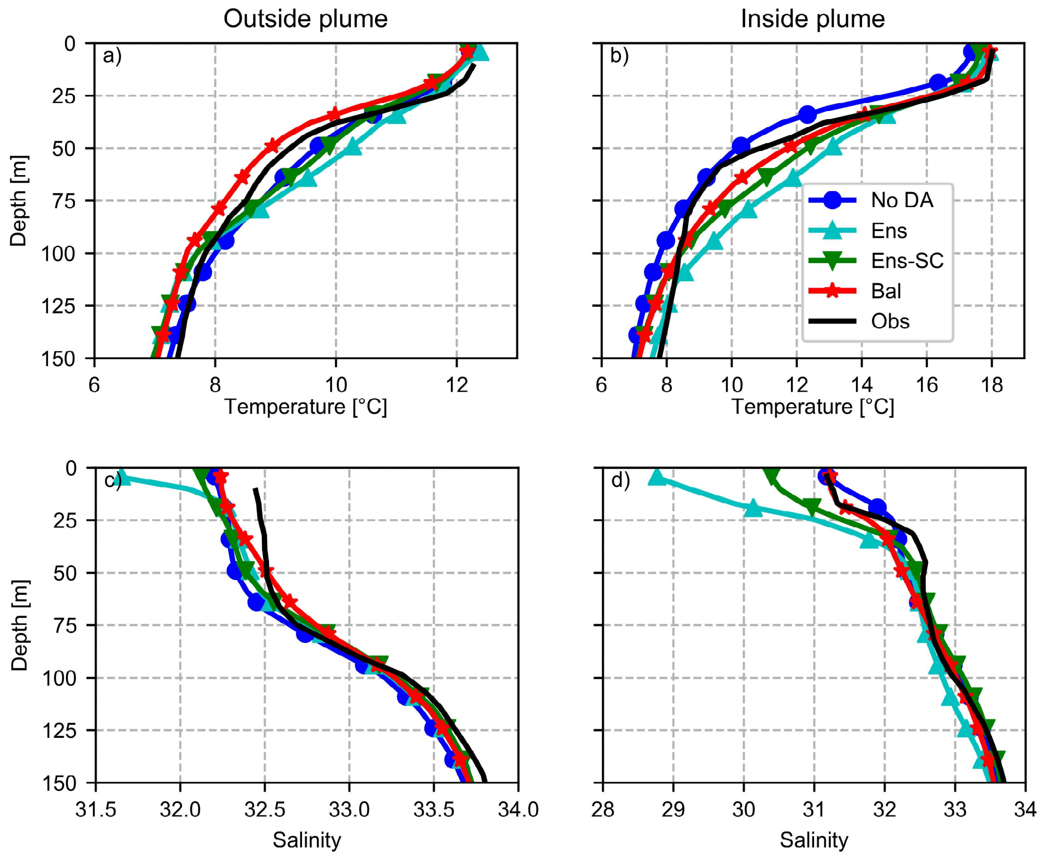

Figure 15: Time-averaged temperature-depth profiles (top row) and salinity-depth profiles (bottom row) from ARGO observations (black) and forecasts from experiment *No DA* (blue), *Ens* (light blue), *Ens-SC* (green) and *Bal* (red). Separate average profiles are shown based on profiles taken while the float was outside the river plume (a,c) and while the float was in or beneath the river plume (b,d). See Figure 11 for the location of the ARGO floats.